2009 International Fuzzy Systems Association World Congress
2009 European Society for Fuzzy Logic and Technology Conference

Proceedings

Lisbon, Portugal, 20th - 24th July, 2009

Edited by J. P. Carvalho, D. Dubois, U. Kaymak and J. M. C. Sousa

IFSA/EUSFLAT 2009

INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

ERASMUS UNIVERSITEIT ROTTERDAM

idmec IST engenharia mecânica

inesc id lisboa

2009 International Fuzzy Systems Association WORLD CONGRESS
2009 European Society for Fuzzy Logic and Technology CONFERENCE

Proceedings

Lisbon, Portugal, 20th - 24th July, 2009

Edited by J. P. Carvalho, D. Dubois, U. Kaymak and J. M. C. Sousa

IFSA/EUSFLAT 2009

The Belém Tower (in Portuguese "Torre de Belém") is a fortified tower located in the Belém district of Lisbon, Portugal. The Belém Tower was built in the early sixteenth century in the Portuguese late-Gothic style, the Manueline, to commemorate Vasco da Gama's expedition (discovered the see way to India). This defensive, yet elegant construction has become one of the symbols of the city, a memorial to Portuguese power as it was during the Age of Discovery, which started in the 15th Century and continuing into the 17th Century. In 1983, it was classified, together with the nearby Mosteiro dos Jerónimos (Jerónimos Monastery), as a UNESCO World Heritage Site.

Cover design: Susana M. Vieira.

Proceedings of the 2009 International Fuzzy Systems Association World Congress and 2009 European Society for Fuzzy Logic and Technology Conference (IFSA–EUSFLAT 2009).

ISBN: 978-989-95079-6-8

This publication was supported by the Program "Fundo de Apoio à Comunidade Científica" from the "Fundação para a Ciência e a Tecnologia" (FCT).

Table of Contents

An Approach to Interval-Valued R-Implications and Automorphisms

Benjamín C. Bedregal[1] Graçaliz P. Dimuro, Renata H. S. Reiser[2]

1.Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte
Campus Universitário, 59072-970 Natal, Brazil
2.Programa de Pós-Graduação em Informática, Universidade Católica de Pelotas
Rua Felix da Cunha, 96010-000 Pelotas, Brazil
Email: bedregal@dimap.ufrn.br, {liz,reiser}@ucpel.tche.br

Abstract— The aim of this work is to introduce an approach for interval-valued R-implications, which satisfy some analogous properties of R-implications. We show that the best interval representation of an R-implication that is obtained from a left continuous t-norm coincides with the interval-valued R-implication obtained from the best interval representation of such t-norm, whenever this is an inclusion monotonic interval function. This provides, under this condition, a nice characterization for the best interval representation of an R-implication, which is also an interval-valued R-implication. We also introduce interval-valued automorphisms as the best interval representations of automorphisms. It is shown that interval automorphisms act on interval R-implications, generating other interval R-implications.

Keywords— Interval-valued fuzzy logic, R-implications, automorphisms, interval representations.

1 Introduction

Interval analysis [1] has been playing an important role in the modeling of the uncertainty and the errors that occur in numerical computations, allowing the development of computational models, methods and tools for the automatic error analysis of numerical algorithms in digital computers, with application in technological and scientific computation.[1]

On the other hand, fuzzy set theory [3] has provided a more generical and complete mathematical model of uncertainty and vagueness, which has been considered as the oldest and most widely reported component of present-day soft computing, helping with the design of flexible information processing systems, with application in, e.g., control systems, decision making, expert systems, pattern recognition. [4]

However, there may be the case that we do not have precise knowledge about the membership function that should be taken in a certain application. These considerations have led to some extensions of fuzzy sets, such us type-2 fuzzy sets [5], which incorporate uncertainty about the membership function into fuzzy set theory. Type-2 fuzzy sets has been largely applied since the works of Jerry Mendel in the 90's [6].

Interval-valued fuzzy sets are a particular case of type-2 fuzzy sets, where the membership degree of each element of the fuzzy set is given by a closed subinterval of the unit interval $[0, 1]$, allowing to deal not only with vagueness (lack of sharp class boundaries), but also with the uncertainty (lack of information), intuitively [7, 8]. Interval-valued fuzzy sets were introduced independently by Zadeh [5] and other authors (e.g., [9, 10, 11, 12]) in the 70's. Among several papers connecting these areas (see, e.g., [13, 14, 15, 16]), we adopted Bedregal and Takahashi's work [17, 18].

Fuzzy implications [19, 20] play an important role in fuzzy logic. They are usually derived from t-norms and t-conorms in several ways, e.g, S-implications, R-implications, QL-implications and D-implications.[2] The importance of fuzzy implications is not only because they are used in representing "If ... then" rules in fuzzy systems, but also because their use in performing inferences in approximate reasoning and fuzzy control. This is the main reason for searching many different models to perform this kind of fuzzy connectives. In particular, R-implications [21], which are based on left continuous t-norms, are used in the modelling of fuzzy rules that satisfy natural properties of implications (see Sect. 4), leading to the residuation property and to an important family of fuzzy logics [22]. On the other hand, the use of automorphisms [19] allows to modify implications preserving their fundamental properties, and then preserving their axiomatic.

The aim of this work is to introduce an approach for interval-valued R-implications as interval generalizations for R-implications that satisfy some analogous properties of R-implications. We show that the best interval representation of an R-implication that is obtained from a left continuous t-norm coincides with the interval-valued R-implication obtained from the best interval representation of such t-norm, whenever this is an inclusion monotonic interval function. Then, it is possible to provide, under this condition, a nice characterization for the best interval representation of an R-implication, which is also an interval-valued R-implication. We also introduce interval-valued automorphisms [16] as the best interval representations of automorphisms, showing that interval automorphisms act on interval R-implications, generating other interval R-implications. Commutative diagrams are used to analyze the relationship between the process for obtaining R-implications from t-norms and the process for obtaining interval-valued R-implications from interval-valued t-norms and those canonical interval constructions. Also, we show that the use of automorphism over R-implications and the use of interval-valued automorphisms over interval-valued

[1]For a survey on applications of Interval Mathematics, see, e.g., [2] and http://www.cs.utep.edu/interval-comp/.

[2]There are also some other methods for the generation of fuzzy implications (see, e.g, [20]).

ISBN: 978-989-95079-6-8

R-implications also comute when the interval constructor is applied.

The paper is organized as follows. In Sect. 2, the main concepts related to interval representations are presented. Section 3 considers interval-valued fuzzy t-norms. Fuzzy implications and R-implications, together with their properties, are presented in Sect. 4. Interval-valued fuzzy implications and R-implications are introduced in Sect. 5, and interval-valued automorphism in Sect. 6, as well as the main related results of the paper. Section 6 is the Conclusion and final remarks.

2 Interval Representations

Consider the real unit interval $U = [0, 1] \subseteq \mathbb{R}$ and let $\mathbb{U}$ be the set of subintervals of U, that is, $\mathbb{U} = \{[a, b] \mid 0 \leq a \leq b \leq 1\}$. The interval set has two projections $l, r : \mathbb{U} \to U$, defined by $l([a, b]) = a$ and $r([a, b]) = b$, respectively. For $X \in \mathbb{U}$, $l(X)$ and $r(X)$ are also denoted by $\underline{X}$ and $\overline{X}$, respectively.

The partial orders that are considered in this paper are the inclusion relation and the component-wise *Kulisch-Miranker order* (also called *product order*), defined by:

$$\forall X, Y \in \mathbb{U} : X \leq Y \Leftrightarrow \underline{X} \leq \underline{Y} \wedge \overline{X} \leq \overline{Y}. \quad (1)$$

An interval $X \in \mathbb{U}$ is said to be an interval representation of a real number α if $\alpha \in X$. Considering two interval representations X and Y of a real number α, X is said to be a better representation of α than Y if $X \subseteq Y$. This notion can be easily extended for tuples of n intervals $\overrightarrow{X} = (X_1, \ldots, X_n)$.

Definition 1 *A function $F : \mathbb{U}^n \longrightarrow \mathbb{U}$ is an* interval representation *of a function $f : U^n \longrightarrow U$ if, for each $\overrightarrow{X} \in \mathbb{U}^n$ and $\overrightarrow{x} \in \overrightarrow{X}$, $f(\overrightarrow{x}) \in F(\overrightarrow{X})$ [23].*[3]

Definition 2 *Let $F : \mathbb{U}^n \longrightarrow \mathbb{U}$ and $G : \mathbb{U}^n \longrightarrow \mathbb{U}$ be two interval representations of the function $f : U^n \longrightarrow U$. F is a* better interval representation *of f than G, denoted by $G \sqsubseteq F$, if, for each $\overrightarrow{X} \in \mathbb{U}^n$, the inclusion $F(\overrightarrow{X}) \subseteq G(\overrightarrow{X})$ holds.*

Definition 3 *For each real function $f : U^n \longrightarrow U$, the interval function $\widehat{f} : \mathbb{U}^n \longrightarrow \mathbb{U}$, defined by*

$$\widehat{f}(\overrightarrow{X}) = \left[\inf\{f(\overrightarrow{x}) \mid \overrightarrow{x} \in \overrightarrow{X}\}, \sup\{f(\overrightarrow{x}) \mid \overrightarrow{x} \in \overrightarrow{X}\}\right], \quad (2)$$

is called the best interval representation of *f [23].*[4]

The interval function $\widehat{f}$ is well defined and for any other interval representation F of f, $F \sqsubseteq \widehat{f}$. The interval function $\widehat{f}$ returns an interval that is narrower than any other interval representation of f. Thus, $\widehat{f}$ presents the *optimality property* of interval algorithms mentioned by Hickey et al. [24], when it is seen as an algorithm to compute a real function f.

Notice that if f is continuous in the usual sense, then for each $\overrightarrow{X} \in \mathbb{U}^n$, the interval function $\widehat{f}$ applied to $\overrightarrow{X}$ coincides with the image of f when applied to $\overrightarrow{X}$, that is, $\widehat{f}(\overrightarrow{X}) = f(\overrightarrow{X})$, where $f(\overrightarrow{X}) = \{f(\overrightarrow{x}) \mid \overrightarrow{x} \in \overrightarrow{X}\}$.

In this paper we will take into consideration the following notions of continuity for interval functions:

- *Moore* continuity: It is obtained as an extension of the continuity in the reals, considering the metric $d(X, Y) = \max(|\ \underline{X} - \underline{Y}\ |, |\ \overline{X} - \overline{Y}\ |)$ defined over $\mathbb{U}$.
- *Scott* continuity: The set $\mathbb{U}$ with reverse inclusion order is defined as a continuous domain [25], and a function $f : (\mathbb{U}, \supseteq) \to (\mathbb{U}, \supseteq)$ is Scott continuous if it is monotonic and preserves the least upper bound of directed sets [25].

The main result in [23] can be adapted to our context, considering U^n instead of $\mathbb{R}$, as shown in the following:

Theorem 4 *Let $f : U^n \longrightarrow U$ be a function. The following statements are equivalent: (i) f is continuous; (ii) $\widehat{f}$ is Scott continuous; (iii) $\widehat{f}$ is Moore continuous.*

3 Interval t-norms

A *triangular norm* (t-norm, for short) is a function $T : U^2 \to U$ that is commutative, associative, monotonic and has 1 as neutral element [26]. In the following definition, a natural extension of the t-norm notion for $\mathbb{U}$ is considered, following the approach introduced in [17].

Definition 5 *A function $\mathbb{T} : \mathbb{U}^2 \to \mathbb{U}$ is an* interval t-norm *if it is commutative, associative, monotonic with respect to the product and inclusion order and $[1, 1]$ is a neutral element.*

An interval t-norm may be considered as an interval representation of a t-norm. This generalization fits with the fuzzy principle, which means that the interval membership degree may be thought as an approximation of the exact degree.

See [18] for the proofs of the next two propositions in this section. The following result shows how an interval t-norm can be constructed from two given t-norms.

Proposition 6 *A function $\mathbb{T} : \mathbb{U}^2 \to \mathbb{U}$ is an interval t-norm if and only if there exist t-norms $\underline{\mathbb{T}}$ and $\overline{\mathbb{T}}$ with $\underline{\mathbb{T}} \leq \overline{\mathbb{T}}$ and*

$$\mathbb{T}(X, Y) = [\underline{\mathbb{T}}(\underline{X}, \underline{Y}), \overline{\mathbb{T}}(\overline{X}, \overline{Y})]. \quad (3)$$

The next proposition states that the best interval representation of a t-norm is an interval t-norm. It also shows that the interval representation of a t-norm coincides with the interval construction of Prop. 6 when both t-norms are the same.

Proposition 7 *Let T be a t-norm. Then $\widehat{T} : \mathbb{U}^2 \to \mathbb{U}$ is an interval t-norm, such that*

$$\widehat{T}(X, Y) = [T(\underline{X}, \underline{Y}), T(\overline{X}, \overline{Y})]. \quad (4)$$

4 Fuzzy Implications and R-Implications

Several definitions for fuzzy implication together with related properties have been given (see, e.g., [21, 19, 20]). The unique consensus in these definitions is that the fuzzy implication should present the same behavior of the classical implication for the crisp case. Thus, a function $I : U^2 \longrightarrow U$ is a *fuzzy implication* if it satisfies the minimal boundary conditions:

$$I(1, 1) = I(0, 1) = I(0, 0) = 1 \text{ and } I(1, 0) = 0.$$

Several reasonable properties may be required for fuzzy implications. The properties considered in this paper are:

I1 : If $y \leq z$ then $I(x, y) \leq I(x, z)$; (2-place monotonicity)

I2 : $I(x, I(y, z)) = I(y, I(x, z))$ (exchange principle);

I3 : $I(x, y) = 1$ if and only if $x \leq y$;

I4 : $\lim_{n\to\infty} I(x, y_n) = I(x, \lim_{n\to\infty} y_n)$ (right-continuity);

I5 : If $x \leq z$ then $I(x, y) \geq I(z, y)$; (1-place antitonicity)

Proposition 8 *[19, Lemma 1 (xi)] If I is a fuzzy implication satisfying* **I1**, **I2** *and* **I3**, *then I also satisfies* **I5**.

[3] Notice that the concept of interval representation is different from interval extension and natural extension. [1, page 21]

[4] Notice that $\widehat{f}$ is the interval hull of the range of f.

ISBN: 978-989-95079-6-8

Considering a t-norm T, the equation

$$I_T(x,y) = \sup\{z \in [0,1] \mid T(x,z) \leq y\}, \quad (5)$$

defines a fuzzy implication, called *R-implication* or *residuum* of T [21]. R-implications arise from the notion of residuum in Intuitionistic Logic [21] or, equivalently, from the notion of residue in the theory of lattice-ordered semigroups [27]. Observe that an R-implication is well-defined only if the t-norm is left-continuous[5] [19]. This justifies the name "residuum of T", since an R-implication satisfies the residuation condition when the underlying t-norm is left continuous:

$$T(x,z) \leq y \text{ if and only if } I_T(x,y) \geq z. \quad (6)$$

Moreover, a t-norm T is left-continuous if and only if it satisfies the residuation condition [28].

The main results relating R-implications and the properties **I1 – I5** are presented in the following.

Theorem 9 *[29, Theorem 1.14] A fuzzy implication $I : U^2 \to U$ is an R-implication with a left-continuous underlying t-norm if and only if I satisfies the properties* **I1–I4**.

Since, by Prop. 8, **I5** follows directly from **I1 – I3**, then R-implications with left-continuous underlying t-norms satisfy the properties **I1 – I5**.

5 Interval-valued Fuzzy R-Implications

The minimal properties required for fuzzy implications can be naturally extended to consider interval fuzzy degrees, by using degenerate intervals. Thus, a function $\mathbb{I} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is an *interval fuzzy implication* if the following conditions hold:

$$\mathbb{I}([1,1],[1,1]) = \mathbb{I}([0,0],[0,0]) = \mathbb{I}([0,0],[1,1]) = [1,1]; \quad (7)$$

$$\mathbb{I}([1,1],[0,0]) = [0,0]. \quad (8)$$

Other properties (see Sect. 4) can be naturally extended:

$\mathbb{I}$**1** : If $Y \leq Z$ then $\mathbb{I}(X,Y) \leq \mathbb{I}(X,Z)$;

$\mathbb{I}$**2** : $\mathbb{I}(X,\mathbb{I}(Y,Z)) = \mathbb{I}(Y,\mathbb{I}(X,Z))$;

$\mathbb{I}$3**a** : If $\mathbb{I}(X,Y) = [1,1]$ then $X \leq Y$;

$\mathbb{I}$3**b** : If $\overline{X} \leq \underline{Y}$ then $\mathbb{I}(X,Y) = [1,1]$;

$\mathbb{I}$**4** : $\mathbb{I}_Y(X) = \mathbb{I}(X,Y)$ is (Moore, Scott) continuous;

$\mathbb{I}$**5** : If $X \leq Z$ then $\mathbb{I}(X,Y) \geq \mathbb{I}(Z,Y)$,

It is always possible to obtain an interval fuzzy implication from any fuzzy implication canonically. Interval fuzzy implications also meet the optimality property and preserve the same properties satisfied by fuzzy implications. The proof of the following proposition is straightforward.

Proposition 10 *If I is a fuzzy implication then $\widehat{I}$ is an interval fuzzy implication. [18]*

The next theorem states that the best interval representation of a fuzzy implication preserves, in some sense, the properties **I1–I5** listed in Sect. 4.

Theorem 11 *Let I be a fuzzy implication satisfying* **I1**. *If I satisfies a property* **Ik**, *for $k = 1,\ldots,5$, then $\widehat{I}$ satisfies* $\mathbb{I}$**k**.

Proof:

[5] A t-norm T is said to be left-continuous whenever $\lim_{n\to\infty} T(x_n,y) = T(\lim_{n\to\infty} x_n, y)$. [26]

$\mathbb{I}$**1**: If I satisfies **I1**, then $\widehat{I}(X,Y) = [\inf\{I(x,\underline{Y})|x \in X\}, \sup\{I(x,\overline{Y})|x \in X\}]$ and $\widehat{I}(X,Z) = [\inf\{I(x,\underline{Z})|x \in X\}, \sup\{I(x,\overline{Z})|x \in X\}]$. Thus, if $Y \leq Z$ then, for each $x \in X$, it holds that $I(x,\underline{Y}) \leq I(x,\underline{Z})$ and $I(x,\overline{Y}) \leq I(x,\overline{Z})$. It follows that $\inf\{I(x,\underline{Y})|x \in X\} \leq \inf\{I(x,\underline{Z}) \mid x \in X\}$, $\sup\{I(x,\overline{Y})|x \in X\} \leq \sup\{I(x,\overline{Z})|x \in X\}$, and, thus, $\widehat{I}(X,Y) \leq \widehat{I}(X,Z)$.

$\mathbb{I}$**2**: If I satisfies **I2**, then, for each $x,y,z \in U$, one has that $I(x,I(y,z)) = I(y,I(x,z))$. Then, since I satisfies **I1**:

$$\begin{aligned}
&\widehat{I}(X,\widehat{I}(Y,Z))\\
&= \widehat{I}(X,[\inf\{I(y,z)) \mid y \in Y, z \in Z\},\\
&\qquad \sup\{I(y,z)) \mid y \in Y, z \in Z\}])\\
&= \widehat{I}(X,[\inf\{I(y,\underline{Z}))|y \in Y,\},\sup\{I(y,\overline{Z}))|y \in Y\}])\\
&= [\inf\{I(x,\inf\{I(y,\underline{Z}) \mid y \in Y\}) \mid x \in X\},\\
&\quad \sup\{I(x,\sup\{I(y,\overline{Z}) \mid y \in Y\}) \mid x \in X\}]\\
&= [\inf\{I(x,I(y,\underline{Z})) \mid y \in Y, x \in X\},\\
&\qquad \sup\{I(x,I(y,\overline{Z})) \mid y \in Y, x \in X\}]\\
&= [\inf\{I(y,I(x,\underline{Z})) \mid y \in Y, x \in X\},\\
&\qquad \sup\{I(y,I(x,\overline{Z})) \mid y \in Y, x \in X\}]\\
&= \widehat{I}(Y,\widehat{I}(X,Z))
\end{aligned}$$

$\mathbb{I}$3a: One has that if $\widehat{I}(X,Y) = [1,1]$ then $\inf\{I(x,y)|x \in X, y \in Y\} = 1 = \sup\{I(x,y)|x \in X, y \in Y\}$. However, this is only possible if $\{I(x,y)|x \in X, y \in Y\} = \{1\}$, and, thus, $I(x,y) = 1$, for each $x \in X, y \in Y$. Since I satisfies **I3**, then, for each $x \in X, y \in Y$, it is valid that $x \leq y$. This is only possible if $\overline{X} \leq \underline{Y}$, that is, $X \leq Y$.

$\mathbb{I}$3b: If $\overline{X} \leq \underline{Y}$ then, for each $x \in X$ and $y \in Y$, it is valid that $x \leq y$. So, one has that $\{I(x,y)|x \in X, y \in Y\} = \{1\}$, $\inf\{I(x,y)|x \in X, y \in Y\} = 1 = \sup\{I(x,y)|x \in X, y \in Y\}$. It follows that $\widehat{I}(X,Y) = [1,1]$.

$\mathbb{I}$**4**: For each $x \in X$, let $I_x : U \longrightarrow U$ be defined by $I_x(y) = I(x,y)$. Thus, I is right-continuous if and only if, for each $x \in X$, I_x is continuous. If I_x is continuous then, by Theorem 4, $\widehat{I}_x$ is Scott (Moore) continuous. Thus:

$$\begin{aligned}
&\widehat{I}(X,Y)\\
&= [\inf\{I(x,y)|x \in X, y \in Y\}, \sup\{I(x,y)|x \in X, y \in Y\}]\\
&= [\inf\{\inf\{I(x,y)|y \in Y\} \mid x \in X\},\\
&\qquad \sup\{\sup\{I(x,y)|y \in Y\}|x \in X\}]\\
&= \bigcup_{x\in X} [\inf\{I(x,y)|y \in Y\}, \sup\{I(x,y)|y \in Y\}] = \bigcup_{x\in X} \widehat{I}_x(Y)
\end{aligned}$$

and then, considering that I_x is (topologically) continuous and the union preserves continuity, one concludes that $\widehat{I}$ is also Scott (Moore) continuous.

$\mathbb{I}$**5**: If I satisfies **I5**, then $\widehat{I}(X,Y) = [\inf\{I(\overline{X},y)|y \in Y\}, \sup\{I(\underline{X},y)|y \in Y\}]$ and $\widehat{I}(Z,Y) = [\inf\{I(\overline{Z},y)|y \in Y\}, \sup\{I(\underline{Z},y)|y \in Y\}]$. Thus, if $X \leq Z$ then, for each $y \in Y$, it holds that $I(\underline{X},y) \leq I(\underline{Z},y)$ and $I(\overline{X},y) \leq I(\overline{Z},y)$. It follows that $\inf\{I(\overline{Z},y)|y \in Y\} \leq \inf\{I(\overline{X},y) \mid y \in Y\}$ and $\sup\{I(\underline{Z},y)|y \in Y\} \leq \inf\{I(\underline{X},y)|y \in Y\}$, and, thus, $\widehat{I}(X,Y) \geq \widehat{I}(Z,Y)$.

□

The next corollary indicates that the best interval representation of a fuzzy implication satisfying **I1–I3** satisfies the properties $\mathbb{I}$**1**–$\mathbb{I}$**3** and $\mathbb{I}$**5**.

Corollary 12 *Let $I : U^2 \longrightarrow U$ be a fuzzy implication satisfying* **I1**, **I2** *and* **I3**. *Then $\widehat{I}$ satisfies* $\mathbb{I}$**1**–$\mathbb{I}$**3** *and* $\mathbb{I}$**5**.

ISBN: 978-989-95079-6-8

Proof: It follows from Prop. 8 and Theorem 11. □

Next proposition provides, for the best interval representation of a fuzzy implication satisfying the properties **I1**, **I2** and **I3**, a more concrete and simpler characterization than Eq. (2).

Proposition 13 *Let $I : U^2 \longrightarrow U$ be a fuzzy implication satisfying the properties* **I1**, **I2** *and* **I3**. *Then a characterization of $\widehat{I}$ can be obtained as*

$$\widehat{I}(X,Y) = [I(\overline{X},\underline{Y}), I(\underline{X},\overline{Y})]. \quad (9)$$

Proof: By Prop. 8, I also satisfies **I5** and, therefore, it holds that $\widehat{I}(X,Y) = [I(\overline{X},\underline{Y}), I(\underline{X},\overline{Y})]$ (see also [18]). □

Theorem 14 *Let $\mathbb{I} : \mathbb{U}^2 \to \mathbb{U}$ be an interval fuzzy implication. If $\mathbb{I}$ satisfies $\mathbb{I}$**1** and $\mathbb{I}$**5** and is inclusion monotonic, then $\mathbb{I}$ is representable by the functions $\underline{\mathbb{I}}(x,y) = l(\mathbb{I}([x,x],[y,y]))$ and $\overline{\mathbb{I}} = r(\mathbb{I}([x,x],[y,y]))$, that is, $\mathbb{I}(X,Y) = [\underline{\mathbb{I}}(\overline{X},\underline{Y}), \overline{\mathbb{I}}(\underline{X},\overline{Y})]$.*

Proof: Given $X, Y \in \mathbb{U}$, since $[\overline{X},\overline{X}] \subseteq X, [\underline{Y},\underline{Y}] \subseteq Y$, then, by the $\subseteq$-monotonicity of $\mathbb{I}$, we have that $\mathbb{I}([\overline{X},\overline{X}],[\underline{Y},\underline{Y}]) \subseteq \mathbb{I}(X,Y)$, and, thus, $l(\mathbb{I}([\overline{X},\overline{X}],[\underline{Y},\underline{Y}])) \geq l(\mathbb{I}(X,Y))$. On the other hand, since $X \leq [\overline{X},\overline{X}]$ and $[\underline{Y},\underline{Y}] \leq Y$, then, by $\mathbb{I}$**1** and $\mathbb{I}$**5**, it holds that $\mathbb{I}([\overline{X},\overline{X}],[\underline{Y},\underline{Y}]) \leq \mathbb{I}(X,Y)$, and, thus, $l(\mathbb{I}([\overline{X},\overline{X}],[\underline{Y},\underline{Y}])) \leq l(\mathbb{I}(X,Y))$. It follows that $l(\mathbb{I}([\overline{X},\overline{X}],[\underline{Y},\underline{Y}])) = l(\mathbb{I}(X,Y))$. Analogously, since $[\underline{X},\underline{X}] \subseteq X$ and $[\overline{Y},\overline{Y}] \subseteq Y$, then, by the inclusion monotonicity of $\mathbb{I}$, we have that $\mathbb{I}([\underline{X},\underline{X}],[\overline{Y},\overline{Y}]) \subseteq \mathbb{I}(X,Y)$, and, thus, $r(\mathbb{I}([\underline{X},\underline{X}],[\overline{Y},\overline{Y}])) \leq r(\mathbb{I}(X,Y))$. On the other hand, since $[\underline{X},\underline{X}] \leq X$ and $Y \leq [\overline{Y},\overline{Y}]$, then, by $\mathbb{I}$**1** and $\mathbb{I}$**5**, it holds that $\mathbb{I}(X,Y) \leq \mathbb{I}([\underline{X},\underline{X}],[\overline{Y},\overline{Y}])$, and, thus, $r(\mathbb{I}([\underline{X},\underline{X}],[\overline{Y},\overline{Y}])) \geq r(\mathbb{I}(X,Y))$. One has that $r(\mathbb{I}([\underline{X},\underline{X}],[\overline{Y},\overline{Y}])) = r(\mathbb{I}(X,Y))$, and then $\mathbb{I}(X,Y) = [\underline{\mathbb{I}}(\overline{X},\underline{Y}), \overline{\mathbb{I}}(\underline{X},\overline{Y})]$. □

Definition 15 *An interval fuzzy implication $\mathbb{I}$ is an* interval R-implication *if there is an interval t-norm $\mathbb{T}$ such that*

$$\mathbb{I} = \mathbb{I}_{\mathbb{T}}(X,Y) = \sup\{Z \in \mathbb{U} | \mathbb{T}(X,Z) \leq Y\}. \quad (10)$$

Analogously, given an interval fuzzy implication $\mathbb{I}$, one has that $\mathbb{T}_{\mathbb{I}}(X,Y) = \inf\{Z \in \mathbb{U} \mid \mathbb{I}(X,Z) \geq Y\}$.

In Eq. (10), the supremum is determined considering the product order, and, thus, it results from the supremum considering the usual order on the real numbers (the interval endpoints).

The following theorem provides necessary conditions for an interval implication to be an interval R-implication.

Theorem 16 *Let $\mathbb{T}$ be a (Moore, Scott) left-continuous interval t-norm. Then $\mathbb{I}_{\mathbb{T}}$ satisfies $\mathbb{I}$**1**, and $\mathbb{I}$**3**–$\mathbb{I}$**5**.*

Proof: It follows that :

$\mathbb{I}$**1**: If $Y \leq Z$ and $\mathbb{T}(X,Z') \leq Y$ then it holds that $\mathbb{T}(X,Z') \leq Z$. It follows that $\{Z' \in \mathbb{U} \mid \mathbb{T}(X,Z') \leq Y\} \subseteq \{Z' \in \mathbb{U} \mid \mathbb{T}(X,Z') \leq Z\}$, and, thus, $\mathbb{I}_{\mathbb{T}}(X,Y) \leq \mathbb{I}_{\mathbb{T}}(X,Z)$.

$\mathbb{I}$**3**: $\mathbb{I}_{\mathbb{T}}(X,Y)=[1,1] \Leftrightarrow \sup\{Z \in \mathbb{U} | \mathbb{T}(X,Z) \leq Y\}=[1,1] \Leftrightarrow \{Z \in \mathbb{U} | \mathbb{T}(X,Z) \leq Y\}=\mathbb{U} \Leftrightarrow X=\mathbb{T}(X,[1,1]) \leq Y$.

$\mathbb{I}$**4**: It follows from the Moore (Scott) left-continuity of $\mathbb{T}$.

$\mathbb{I}$**5**: If $X \leq X'$ then $\mathbb{T}(X,Z) \leq Y$ implies that $\mathbb{T}(X',Z) \leq Y$. Thus, one has that $\{Z \in \mathbb{U} \mid \mathbb{T}(X',Z) \leq Y\} \subseteq \{Z \in \mathbb{U} \mid \mathbb{T}(X,Z) \leq Y\}$ and then it holds that $\sup\{Z \in \mathbb{U} \mid \mathbb{T}(X',Z) \leq Y\} \leq \sup\{Z \in \mathbb{U} \mid \mathbb{T}(X,Z) \leq Y\}$. It follows that $\mathbb{I}_{\mathbb{T}}(X',Y) \leq \mathbb{I}_{\mathbb{T}}(X,Y)$.

□

In [21, Lemma 5.16], it was proved that $T = T_{I_T}$, where $T_{I_T}(x,y) = \inf\{z \in U \mid I_T(x,z) \geq y\}$. The following two lemmas and proposition shows that this result can be extended for interval R-implications.

Lemma 17 *Let T be a left-continuous t-norm and $\mathbb{T}$ be a Moore left-continuous interval t-norm. If $\widehat{I_T} = \mathbb{I}_{\mathbb{T}}$ then $\mathbb{I}_{\mathbb{T}}([x,x],Z) \geq [y,y]$ if and only if $\mathbb{I}_{\mathbb{T}}([x,x],[\underline{Z},\underline{Z}]) \geq [y,y]$.*

Proof: It follows that: $\mathbb{I}_{\mathbb{T}}([x,x],Z) \geq [y,y] \Leftrightarrow \widehat{I_T}([x,x],Z) \geq [y,y] \Leftrightarrow \{I_T(x,z) \mid z \in Z\} \geq [y,y] \Leftrightarrow I_T(x,\underline{Z}) \geq y \Leftrightarrow \widehat{I_T}([x,x],[\underline{Z},\underline{Z}]) \geq [y,y] \Leftrightarrow \mathbb{I}_{\mathbb{T}}([x,x],[\underline{Z},\underline{Z}]) \geq [y,y]$. □

Lemma 18 *Let T be a left-continuous t-norm and $\mathbb{T}$ be a Moore left-continuous interval t-norm. If $\widehat{I_T} = \mathbb{I}_{\mathbb{T}}$ then it holds that $\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}([x,x],[y,y]) = \mathbb{T}([x,x],[y,y])$.*

Proof: It follows that:

$$\begin{aligned}
&\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}([x,x],[y,y])\\
&= \inf\{Z \in \mathbb{U} \mid \mathbb{I}_{\mathbb{T}}([x,x],Z) \geq [y,y]\}\\
&= \inf\{[\underline{Z},\underline{Z}] \in \mathbb{U} \mid \mathbb{I}_{\mathbb{T}}([x,x],[\underline{Z},\underline{Z}]) \geq [y,y]\} \text{ by Lemma 17}\\
&= \inf\{[\underline{Z},\underline{Z}] \in \mathbb{U} | \sup\{Z' \in \mathbb{U} | \mathbb{T}([x,x],Z') \leq [\underline{Z},\underline{Z}]\} \geq [y,y]\}\\
&= \inf\{[\underline{Z},\underline{Z}] \in \mathbb{U} \mid \sup\{[\overline{Z'},\overline{Z'}] \in \mathbb{U} \mid \mathbb{T}([x,x],[\overline{Z'},\overline{Z'}])\\
&\qquad \leq [\underline{Z},\underline{Z}]\} \geq [y,y]\} \text{ by } \subseteq\text{-monotonicity of } \mathbb{T}\\
&= \inf\{[\underline{Z},\underline{Z}] \in \mathbb{U} \mid I_{\mathbb{T}}(x,\underline{Z}) \leq y \wedge I_{\overline{\mathbb{T}}}(x,\underline{Z}) \leq y\}\\
&= [T_{I_{\mathbb{T}}}(x,y), T_{I_{\overline{\mathbb{T}}}}(x,y)]=[\underline{\mathbb{T}}(x,y), \overline{\mathbb{T}}(x,y)]=\mathbb{T}([x,x],[y,y])
\end{aligned}$$

□

Proposition 19 *Let T be a left-continuous and $\mathbb{T}$ be a Moore left-continuous interval t-norms. If $\widehat{I_T} = \mathbb{I}_{\mathbb{T}}$ then $\mathbb{T}_{\mathbb{I}_{\mathbb{T}}} = \mathbb{T}$.*

Proof: It follows that:

$$\begin{aligned}
&\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}(X,Y)\\
&= [\underline{\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}}(\underline{X},\underline{Y}), \overline{\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}}(\overline{X},\overline{Y})]\\
&= [l(\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}([\underline{X},\underline{X}],[\underline{Y},\underline{Y}]), r(\mathbb{T}_{\mathbb{I}_{\mathbb{T}}}([\overline{X},\overline{X}],[\overline{Y},\overline{Y}])\\
&= [l(\mathbb{T}([\underline{X},\underline{X}],[\underline{Y},\underline{Y}])), r(\mathbb{T}([\overline{X},\overline{X}],[\overline{Y},\overline{Y}]))] \text{ by Lemma 18}\\
&= [\underline{\mathbb{T}}([\underline{X},\underline{X}],[\underline{Y},\underline{Y}]), \overline{\mathbb{T}}([\overline{X},\overline{X}],[\overline{Y},\overline{Y}])] = \mathbb{T}(X,Y)
\end{aligned}$$

□

Proposition 20 *Let T be a left-continuous t-norm and $\mathbb{T}$ be a Moore left-continuous interval t-norm. If $\widehat{I_T} = \mathbb{I}_{\mathbb{T}}$ then $\mathbb{T}$ is an interval representation of T.*

Proof: It holds that

$$\begin{aligned}
&\mathbb{T}([x,x],[y,y])\\
&= \mathbb{T}_{\mathbb{I}_{\mathbb{T}}}([x,x],[y,y]) \text{by Prop. 19}\\
&= \inf\{Z \in \mathbb{U} \mid \mathbb{I}_{\mathbb{T}}([x,x],Z) \geq [y,y]\}\\
&= \inf\{Z \in \mathbb{U} \mid \widehat{I_T}([x,x],Z) \geq [y,y]\} \text{ by hypothesis}\\
&= \inf\{Z \in \mathbb{U} \mid I_T(x,z) \geq y, \forall z \in Z\}\\
&= \inf\{Z \in \mathbb{U} \mid I_T(x,\underline{Z}) \geq y\} \text{ by } \mathbf{I1} \text{ and Theorem 9}\\
&= \inf\{[\underline{Z},\underline{Z}] \in \mathbb{U} \mid I_T(x,\underline{Z}) \geq y\}\\
&= [\inf\{\underline{Z} \in U \mid I_T(x,\underline{Z}) \geq y\}, \inf\{\underline{Z} \in U \mid I_T(x,\underline{Z}) \geq y\}]\\
&= [T_{I_T}(x,y), T_{I_T}(x,y)] = [T(x,y), T(x,y)].
\end{aligned}$$

By the $\subseteq$-monotonicity of $\mathbb{T}$, we have that $\mathbb{T}$ represents T. □

Theorem 21 *If T is a left continuous t-norm, then $\mathbb{I}_{\widehat{T}} \subseteq \widehat{I_T}$.*

Proof: It follows that:

$$\begin{aligned}
&\mathbb{I}_{\widehat{T}}(X,Y)\\
&= \sup\{Z \in \mathbb{U} | \widehat{T}(X,Z) \leq Y\} \text{ by Eq. (10)}\\
&= \sup\{[\underline{Z},\overline{Z}] \in \mathbb{U} | [T(\underline{X},\underline{Z}), T(\overline{X},\overline{Z})] \leq [\underline{Y},\overline{Y}]\} \text{ by Eq. (4)}\\
&= \sup\{[\underline{Z},\overline{Z}] \in \mathbb{U} | T(\underline{X},\underline{Z}) \leq \underline{Y} \wedge T(\overline{X},\overline{Z}) \leq \overline{Y}\} \text{ by Eq. (1)}\\
&= [\min\{\sup\{\underline{Z} \in U | T(\underline{X},\underline{Z}) \leq \underline{Y}\},\\
&\quad \sup\{\overline{Z} \in U | T(\overline{X},\overline{Z}) \leq \overline{Y}\}\}, \sup\{\overline{Z} \in U | T(\overline{X},\overline{Z}) \leq \overline{Y}\}]\\
&= [\min\{I_T(\underline{X},\underline{Y}), I_T(\overline{X},\overline{Y})\}, I_T(\overline{X},\overline{Y})] \text{ by Eq. (5)}\\
&\subseteq [I_T(\overline{X},\underline{Y}), I_T(\underline{X},\overline{Y})] = \widehat{I_T}(X,Y) \text{ by Theorem 9 and Eq. (9)}
\end{aligned}$$

ISBN: 978-989-95079-6-8

□

The following results states a sufficient condition to guarantee the desired equivalence $\mathbb{I}_{\widehat{T}} = \widehat{I_T}$.

Lemma 22 *Let T be left-continuous t-norm. Then, it holds that $\underline{\mathbb{I}_{\widehat{T}}} = I_T = \overline{\mathbb{I}_{\widehat{T}}}$.*

Proof: It follows that

$$\begin{aligned}\underline{\mathbb{I}_{\widehat{T}}}(x,y) &= l(\mathbb{I}_{\widehat{T}}([x,x],[y,y]))\\ &= l(\sup\{Z \in \mathbb{U} \mid \widehat{T}([x,x],Z) \leq [y,y]\})\\ &= l(\sup\{[\overline{Z},\overline{Z}] \in \mathbb{U} \mid \widehat{T}([x,x],[\overline{Z},\overline{Z}]) \leq [y,y]\})\\ &\qquad \text{by } \leq \text{ and } \subseteq\text{-monotonicity}\\ &= \sup\{\overline{Z} \in U \mid T(x,\overline{Z}) \leq y\} = I_T(x,y).\end{aligned}$$

The proof of the other equality is analogous. □

Proposition 23 *If T is left-continuous t-norm then $\widehat{T}$ is a Moore left-continuous interval t-norm.*

Proof: It is analogous to the proof of Theorem 11 ($\mathbb{I}$**4**). □

Theorem 24 *Let T be a left-continuous t-norm. If $\mathbb{I}_{\widehat{T}}$ is $\subseteq$-monotonic then*

$$\mathbb{I}_{\widehat{T}} = \widehat{I_T}. \tag{11}$$

Proof: By Theorem 9 and Prop. 8, I_T satisfies **I1** and **I5** and $\widehat{I_T}(X,Y) = [\inf\{I_T(x,y) \mid x \in X \wedge y \in Y\}, \sup\{I_T(x,y) \mid x \in X \wedge y \in Y\}] = [I_T(\overline{X},\underline{Y}), I_T(\underline{X},\overline{Y})]$. On the other hand, since $\mathbb{I}_{\widehat{T}}$ is $\subseteq$-monotonic, T is (Moore, Scott) left-continuous (by Prop. 23), and, by Theorem 16, it satisfies $\mathbb{I}$**1** and $\mathbb{I}$**5**. By Theorem 14, it follows that $\mathbb{I}_{\widehat{T}}(X,Y) = [\underline{\mathbb{I}_{\widehat{T}}}(\overline{X},\underline{Y}), \overline{\mathbb{I}_{\widehat{T}}}(\underline{X},\overline{Y})]$. Then, by Lemma 22, one has that $\mathbb{I}_{\widehat{T}}(X,Y) = [I_T(\overline{X},\underline{Y}), I_T(\underline{X},\overline{Y})]$. Therefore, it holds that $\mathbb{I}_{\widehat{T}}(X,Y) = \widehat{I_T}(X,Y)$. □

The above results state the commutativity of the diagram of Fig. 1, where $\mathcal{C}(T)$ ($\mathcal{C}(\mathbb{T})$) is the class of left-continuous (interval) t-norms and $\mathcal{C}(I)$ ($\mathcal{C}(\mathbb{I})$) is the class of ($\subseteq$-monotonic interval) R-implications.

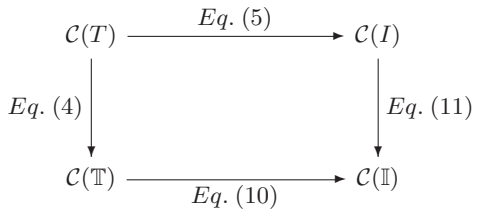


Figure 1: Commutative diagram relating R-implications with interval R-implications

6 Interval-valued Automorphisms

Definition 25 *[30] $\rho : U \longrightarrow U$ is an* automorphism *if it is bijective and monotonic.*[6] *The* action *of an automorphism ρ on a function $f : U^2 \to U$, denoted by f^ρ, is defined as $f^\rho(x,y) = \rho^{-1}(f(\rho(x),\rho(y)))$.*

As it is well known, if T is a t-norm then T^ρ is also a t-norm, and if I is an R-implication then I^ρ is also an R-implication.

$\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ is an *interval automorphism* if it is bijective and monotonic with respect to the product order [16]. The set of interval automorphisms is denoted by $Aut(\mathbb{U})$.

The next theorem shows that each interval automorphism can be constructed from an automorphism.

[6] In [19], $\rho : U \longrightarrow U$ is an automorphism if it is a continuous and strictly increasing function such that $\rho(0) = 0$ and $\rho(1) = 1$.

Theorem 26 *[16, Theorems 2 and 3] $\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ is an interval automorphism if and only if there exists an automorphism $\rho : U \longrightarrow U$ such that $\varrho(X) = [\rho(\underline{X}), \rho(\overline{X})]$.*

Clearly, if $\rho : U \to U$ is an automorphism then $\widehat{\rho}$ can be obtained as $\widehat{\rho}(X) = [\rho(\underline{X}), \rho(\overline{X})]$. Therefore, interval automorphisms are the best interval representations of automorphisms.

Remark 1 *Let ϱ be an interval automorphism. Then: (i) ϱ^{-1} is an interval automorphism; (ii) $X \leq Y \Leftrightarrow \varrho(X) \leq \varrho(Y)$.*

Definition 27 *The* action *of an interval automorphism ϱ on an interval function $\mathbb{F}:\mathbb{U}^2 \longrightarrow \mathbb{U}$, denoted by $\mathbb{F}^\varrho$, is defined as*

$$\mathbb{F}^\varrho(X,Y) = \varrho^{-1}(\mathbb{F}(\varrho(X),\varrho(Y))). \tag{12}$$

In the following, we show how interval automorphisms act on interval t-norms and interval R-implications.

Proposition 28 *[17, Theorem 6.1] Let $\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ be an interval automorphism and $\mathbb{T} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ be an interval t-norm. Then the mapping $\mathbb{T}^\varrho : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is an interval t-norm.*

Theorem 29 *Let $\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ be an interval automorphism and $\mathbb{T} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ be an interval t-norm. Then the mapping $\mathbb{I}_{\mathbb{T}}^\varrho : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is defined by $\mathbb{I}_{\mathbb{T}}^\varrho(X,Y) = \mathbb{I}_{\mathbb{T}^\varrho}(X,Y)$.*

Proof: Since ϱ is bijective and ϱ is monotonic, it follows that:

$$\begin{aligned}&\mathbb{I}_{\mathbb{T}}^\varrho(X,Y)\\ &= \varrho^{-1}(\mathbb{I}_{\mathbb{T}}(\varrho(X),\varrho(Y))) \text{ by Eq. (12)}\\ &= \varrho^{-1}(\sup\{Z \in \mathbb{U} | \mathbb{T}(\varrho(X),Z) \leq \varrho(Y)\}) \text{ by Eq. (10)}\\ &= \varrho^{-1}(\sup\{\varrho(Z') \in \mathbb{U} | \mathbb{T}(\varrho(X),\varrho(Z')) \leq \varrho(Y)\})\\ &= \sup\{\varrho^{-1}(\varrho(Z')) \in \mathbb{U} | \mathbb{T}(\varrho(X),\varrho(Z')) \leq \varrho(Y)\}\\ &= \sup\{Z' \in \mathbb{U} | \varrho^{-1}(\mathbb{T}(\varrho(X),\varrho(Z'))) \leq \varrho^{-1}(\varrho(Y))\} \text{ by Rem. 1}\\ &= \sup\{Z' \in \mathbb{U} | \varrho^{-1}(\mathbb{T}(\varrho(X),\varrho(Z'))) \leq Y\} \text{ by Rem. 1}\\ &= (\sup\{Z \in \mathbb{U} | \mathbb{T}^\varrho(X,Z) \leq Y\}) \text{ by Eq. (12)}\\ &= \mathbb{I}_{\mathbb{T}^\varrho}(X,Y) \text{ by Eq. (10)} .\end{aligned}$$

□

Corollary 30 *Let $\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ be an interval automorphism and $\mathbb{T} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ be an interval t-norm. Then the mapping $\mathbb{I}_{\mathbb{T}}^\varrho : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is an interval R-implication.*

Theorem 29 and Cor. 30 state that to apply an interval automorphism to an interval R-implication is the same that to apply it to an interval t-norm, and then to obtain an interval R-implication. Whenever an interval R-implication is submitted to an interval automorphism, a new interval R-implication is generated, which means that interval automorphisms may be applied in order to generate new interval R-implications. So, the commutative diagram pictured in Fig. 2 holds.

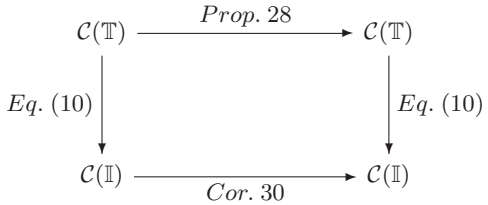


Figure 2: Commutative diagram relating interval t-norms, interval R-implications and interval automorphisms

ISBN: 978-989-95079-6-8

7 Conclusion and Final Remarks

The results presented in this paper extend our previous work (e.g., [17, 31]). We show that interval R-implications satisfy some analogous properties of R-implications. The best interval representation of an R-implication that is obtained from a left continuous t-norm coincides with the interval-valued R-implication obtained from the best interval representation of such t-norm, whenever this is a $\subseteq$-monotonic interval function. This provided, under this condition, a nice characterization for the best interval representation of an R-implication. Interval automorphisms are presented as best interval representations of automorphisms, showing that interval automorphisms act on interval R-implications, generating other interval R-implications.

Future work will consider the analysis of other important properties of interval-valued R-implications, in other to obtain a stronger relation between R-implications and interval-valued R-implications, establishing under which conditions interval R-implications are representable in the sense of [32].

Acknowledgment

This work is supported by CNPq (473201/07-0, 307185/07-9, 307879/06-2). We are grateful to the referees for their valuable suggestions, and also to Xiaohong Zhang for his comments.

References

[1] R.E. Moore. *Methods and Applications of Interval Analysis*. SIAM, Philadelphia, 1979.

[2] R. B. Keafort and V. Kreinovich, editors. *Applications of Interval Computations*. Kluwer, Boston, 1996.

[3] L. A. Zadeh. Fuzzy sets. *Information and Control*, pages 338–353, 1965.

[4] S. Mitra and S. K. Pal. Fuzzy sets in pattern recognition and machine intelligence. *Fuzzy Sets and Systems*, 156:381–386, 2005.

[5] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning - I. *Information Sciences*, 6:199–249, 1975.

[6] J. M. Mendel. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall, Upper Saddle River, 2001.

[7] W.A. Lodwick. Preface. *Reliable Computing*, 10(4):247–248, 2004.

[8] D. Dubois and H. Prade. Interval-valued fuzzy sets, possibility theory and imprecise probability. In *Proc. Intl. Conf. on Fuzzy Logic and Technology*, pages 314–319, Barcelona, 2005.

[9] I. Grattan-Guiness. Fuzzy membership mapped onto interval and many-valued quantities. *Z. Math. Logik. Grundladen Math.*, 22:149–160, 1975.

[10] K.U. Jahn. Intervall-wertige mengen. *Math. Nach.*, 68:115–132, 1975.

[11] H. T. Nguyen, V. Kreinovich, and Q. Zuo. Interval-valued degrees of belief: applications of interval computations to expert systems and intelligent control. *Intl. Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 5(3):317–358, 1997.

[12] R. Sambuc. *Fonctions ϕ-floues. Application l'aide au diagnostic en pathologie thyroidienne*. PhD thesis, Univ. Marseille, Marseille, 1975.

[13] R. E. Moore and W. Lodwick. Interval analysis and fuzzy set theory. *Fuzzy Sets and Systems*, 135(1):5–9, 2003.

[14] G. Cornelis, G. Deschrijver, and E. E. Kerre. Advances and challenges in interval-valued fuzzy logic. *Fuzzy Sets and Systems*, 157:622–627, 2006.

[15] D. Dubois and H. Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems*, 12:87–101, 1991.

[16] M. Gehrke, C. Walker, and E. Walker. Some comments on interval valued fuzzy sets. *International Journal of Intelligent Systems*, 11:751–759, 1996.

[17] B. C. Bedregal and A. Takahashi. The best interval representation of t-norms and automorphisms. *Fuzzy Sets and Systems*, 157(24):3220–3230, 2006.

[18] B. C. Bedregal and A. Takahashi. Interval valued versions of t-conorms, fuzzy negations and fuzzy implications. In *Proc. of the IEEE Intl. Conf. on Fuzzy Systems, Vancouver, 2006*, pages 553–559. IEEE, 2006.

[19] H. Bustince, P. Burilo, and F. Soria. Automorphism, negations and implication operators. *Fuzzy Sets and Systems*, 134:209–229, 2003.

[20] R.R. Yager. On some new classes of implication operators and their role in approximate reasoning. *Information Sciences*, 167:193–216, 2004.

[21] M. Baczynski and B. Jayaramb. (S,N)- and R-implications: A state-of-the-art survey. *Fuzzy Sets and Systems*, 159:1836–1859, 2008.

[22] S. Gottwald and P. Hajek. Triangular norm-based mathematical fuzzy logics. In E. P. Klement and R. Mesiar, editors, *Logical, Algebraic, Analytic and Probabilistic Aspects of Triangular Norms*, pages 275–299. Elsevier, Amsterdam, 2005.

[23] R. H. N. Santiago, B. C. Bedregal, and B. M. Acióly. Formal aspects of correctness and optimality in interval computations. *Formal Aspects of Computing*, 18(2):231–243, 2006.

[24] T. Hickey, Q. Ju, and M. Emdem. Interval arithmetic: from principles to implementation. *Journal of the ACM*, 48(5):1038–1068, 2001.

[25] G. Gierz, K. Hofmann, K. Keimel, J. D. Lawson, M. Mislove, and D. Scott. *Continuous Lattices and Domains*. Cambridge Press, Cambridge, 2003.

[26] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer, Dordrecht, 2000.

[27] J. C. Fodor. Contrapositive symmetry of fuzzy implications. *Fuzzy Sets and Systems*, 69:141–156, 1995.

[28] J. C. Fodor. Left-continuous t-norms in fuzzy logics: an overview. *Journal of Applied Sciences at Budapest Tech Hungary*, 1(2):141–156, 2004.

[29] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer, Dordrecht, 1994.

[30] E.P. Klement and M. Navara. A survey on different triangular norm-based fuzzy logics. *Fuzzy Sets and Systems*, 101:241–251, 1999.

[31] R. H. S. Reiser, G. P. Dimuro, B.C. Bedregal, and R.H.N. Santiago. Interval valued QL-implications. In D. Leivant and R. Queiroz, editors, *Logic, Language, Information and Computation*, number 4576 in LNCS, pages 307–321. Springer, Berlin, 2007.

[32] G. Deschrijver. A representation of t-norms in interval-valued l-fuzzy set theory. *Fuzzy Sets and Systems*, 159, 2008.

ISBN: 978-989-95079-6-8

Recognition and Teaching of Robot Skills by Fuzzy Time-Modeling

Rainer Palm, Bourhane Kadmiry, Boyko Iliev, and Dimiter Driankov
AASS, Dept. of Technology Orebro University SE-70182 Orebro, Sweden
Emails: rub.palm@t-online.de, {bourhane.kadmiry, boyko.iliev, dimiter.driankov}@aass.oru.se

Abstract – Robot skills are low-level motion and/or grasping capabilities that constitute the basic building blocks from which tasks are built. Teaching and recognition of such skills can be done by Programming-by-Demonstration approach. A human operator demonstrates certain skills while his motions are recorded by a data-capturing device and modeled in our case via fuzzy clustering and Takagi-Sugeno modeling technique. The resulting skill models use the time as input and the operator's actions and reactions as outputs. Given a test skill by the human operator the robot control system recognizes the individual phases of skills and generates the type of skill shown by the operator.

Keywords – Fuzzy modeling, time clustering, robot skills, Programming-by-Demonstration

1 Introduction

Robot skills are low-level motion and/or grasping capabilities that constitute the basic building blocks from which tasks are built. One major challenge in robot programming is to be able to program skills in an easy and fast manner with high accuracy. *Programming by Demonstration (PbD)* (PbD) is one such approach that has the afore mentioned properties. In PbD a human who demonstrates the skills is equipped with data-capturing devices (e.g., data glove, cameras, haptic devices etc.). The demonstrator performs a skill while the robot captures the associated motion data, analyzes it and generates a robot-centered model of the demonstrated skill, that is a corresponding robot skill. Once the robot has acquired a number of robot skills from demonstrations it is able to recognize a demonstrated human skill as one of its already available robot skills. In a final step, when a task is demonstrated then the robot recognizes the robot skills that constitute it and thus creates a program consisting of these skills. This approach can be used not only for industrial robots but also in the fields of prosthetics, humanoid service robots, remote control and teleoperation in hasardous and dangerous environments, and last but not least in the entertainment industry. However, such applications are relatively few so far due to the lack of appropriate sensor systems and some unsolved problems with the man-robot interaction. Selected skills are

- contour following
- assembly (peg-in-hole insertion)
- handling of objects
- grasping of objects.

Different techniques for recognition of skills have been applied for PbD. For the manipulation domain Morrow and Khosla describe the construction of a library of robot capabilities by analysis and identification of tasks using a camera and a force-torque sensor [1]. Their approach is to develop a sensorimotor layer which integrates sensing into the robot programming primitives. In [2] Kaiser and Dillmann describe a neural net approach for the initial skill learning and reinforcement learning, skill refinement and adaptation. Experimental results were shown by an insertion example and a door-opening experiment. In the context of task learning Geib et. al. proposed an approach to integrating high-level artificial intelligence planning technology with low-level robotic control [3]. Kwun Han and M. Veloso describe an automated recognition of the behavior of robots using HMMs to represent and recognize strategic behaviors of robotic agents [4]. In the field of recognition of robot behaviors the following publications are important: Zoellner et. al. [5] use a data glove with integrated tactile sensors for behavior recognition which is based on support vector machines (SVM). Ekvall and Kragic [6] apply Hidden Markov Models (HMM) and address the PbD-problem using the arm trajectory as an additional feature for grasp classification. Li et. al. [7] use the singular value decomposition (SVD) for the generation of feature vectors of human grasps and support vector machines (SVM) which are applied to the classification problem. A fuzzy logic approach for gesture recognition was published by Bimber [8]. The method is applied to 6 d.o.f. (degrees of freedom) trajectories of a human arm but cannot cope with more than 6 dimensions. Despite the advanced state of the art, the cited methods on dynamic classification do not consider the evolution of a robot behavior in time and space. This appears to be a disadvantage because neither the estimation of the occurence of a specific skill nor the problem of segmentation can be solved in a general way. The method presented in this paper tries to overcome some of these drawbacks by modeling trajectories in time and space. Its advantage over other methods like HMM or Gaussian Mixture Models is discussed in [9] and [10]. To solve a 'fuzzy' task like skill recognition it turns out that fuzzy modeling is a suitable approach to deal with. The main focus of this method is *programming by demonstration* and *recognition of human skills* using fuzzy clustering and Takagi-Sugeno (TS) fuzzy modeling (see also [11] and [12]). The paper is organized as follows: In Section 2 a general approach to learning skills by partitioning them into phases is discussed. Section 3 deals with fuzzy time-modeling and the segmentation principle. Section 4 describes the recognition of phases and the decision process for the classification of skills. Section 5 presents simulations and experimental results. The final Section 6 draws some conclusions and directions for future work.

ISBN: 978-989-95079-6-8

2 Programming of robot skills by human demonstration

Programming of robot skills requires the building of a library of models of skills being taught or trained by a human demonstrator. In a next step newly demonstrated skills lead to test models which are then compared with the training models (skills) of the library. By such a comparison the robot is able to recognize these newly demonstrated skills. Finally a robot task including the recognized skills can automatically be generated (see Fig. 1). For the training phase two main tasks are needed to perform: *segmentation* of human demonstrations into skill phases and *phase modeling* of the skill phases. *Segmentation* means a partition of the data record into a sequence of episodes, where each one contains a single skill phase. For the test phase three main tasks are needed to perform: *segmentation* of the human test demonstrations, *phase recognition*, and *skill classification*. *Phase recognition* means to recognize the phases performed in each episode. The third task is to connect the recognized skill phases in such a way that a full skill can be identified.

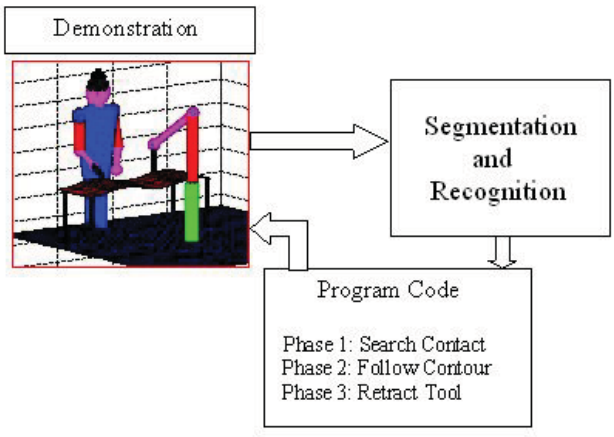


Figure 1: Learning skills from human demonstrations

3 Fuzzy time-modeling and the segmentation principle

Let a skill be partitioned into a sequence of phases as described above. Each phase starts and ends with a discrete event coming either from a discrete sensor or from some appropriate preprocessing of continuous sensor signals. The structure of a skill can be described most appropriately by a hybrid automaton in which nodes represent continuous phases and arcs the discrete transitions (switches) between them. Figure 2 shows an example of such a hybrid automaton. The hybrid process is event-controlled (see [13]) and is assumed to be stable both within the individual phases and with respect to the switching behavior between them. On the other hand the purpose of segmentation is to identify the discrete instants of time for the switches to occur in order to cut the whole skill into phases during demonstration.

The following subsection deals with fuzzy time-modeling in general that is used both for phase modeling and for segmentation. In the same context the training of time cluster models using new data is described. After that the segmentation procedure is presented.

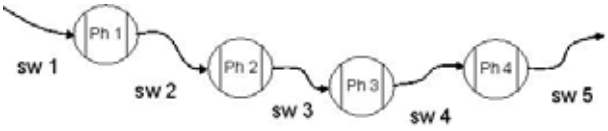


Figure 2: Hybrid automaton of a skill

3.1 Fuzzy time-modeling

Let us for the time being only concentrate on the modeling of a skill phase. The recognition of a skill phase is achieved by a model that reflects the *behavior of the robot end-effector in time* during the episode considered. Each demonstration is repeated several times to collect enough samples of every particular skill phase. From those data, models for each individual phase are developed using fuzzy clustering and Takagi-Sugeno fuzzy modeling ([14, 11]). We consider time instants as model inputs and end-effector coordinates as model outputs. Define the end-effector coordinate by

$$\mathbf{x}(t) = \mathbf{f}(t) \tag{1}$$

where $\mathbf{x}(t) \in R^3, \mathbf{f} \in R^3$, and $t \in R^+$. Further linearize (1) at selected time points t_i

$$\mathbf{x}(t) = \mathbf{x}(t_i) + \frac{\Delta \mathbf{f}(t)}{\Delta t}|_{t_i} \cdot (t - t_i) \tag{2}$$

which is a linear equation in t,

$$\mathbf{x}(t) = \mathbf{A}_i \cdot t + \mathbf{d}_i \tag{3}$$

where $\mathbf{A}_i = \frac{\Delta \mathbf{f}(t)}{\Delta t}|_{t_i} \in R^3$ and $\mathbf{d}_i = \mathbf{x}(t_i) - \frac{\Delta \mathbf{f}(t)}{\Delta t}|_{t_i} \cdot t_i \in R^3$. Using (3) as a local linear model one can express (1) in terms of a Takagi-Sugeno fuzzy model [15]

$$\mathbf{x}(t) = \sum_{i=1}^{c} w_i(t) \cdot (\mathbf{A}_i \cdot t + \mathbf{d}_i) \tag{4}$$

$w_i(t) \in [0, 1]$ is the degree of membership of a time point t to a cluster with the cluster center t_i, c is the number of clusters, and $\sum_{i=1}^{c} w_i(t) = 1$.

Let $\mathbf{x} = [x_1, x_2, x_3]^T$ be the 3 end-effector cordinates and t the time. The general clustering and modeling steps are described as follows

- Select an appropriate number of local linear models (data clusters) c
- Find c cluster centers $(t_i, x_{1i}, x_{2i}, x_{3i})$, $i = 1...c$, in the product space of the data quadruples (t, x_1, x_2, x_3) by Fuzzy-c-elliptotype clustering
- Find the corresponding fuzzy regions in the space of input data (t) by projection of the clusters of the product space into Gustafson-Kessel clusters (GK) within the input space [16]
- Calculate c local linear (affine) models (4) using the GK clusters from step 2.

ISBN: 978-989-95079-6-8

The membership degree $w_i(t)$ of an input data point t in an input cluster C_i is then calculated by

$$w_i(t) = \frac{1}{\sum_{j=1}^{c} \left(\frac{(t-t_i)^T M_{i_{pro}} (t-t_i)}{(t-t_j)^T M_{j_{pro}} (t-t_j)} \right)^{\frac{1}{\tilde{m}_{proj}-1}}} \quad (5)$$

The projected cluster centers t_i and induced matrices $M_{i_{pro}}$ define the input clusters C_i ($i = 1 \dots c$). The parameter $\tilde{m}_{pro} > 1$ determines the fuzziness of an individual cluster. The spheres in Fig. 3 covering the trajectories represent the local fuzzy models. The stripes along the time coordinate represent projections of the local fuzzy models onto the time line.

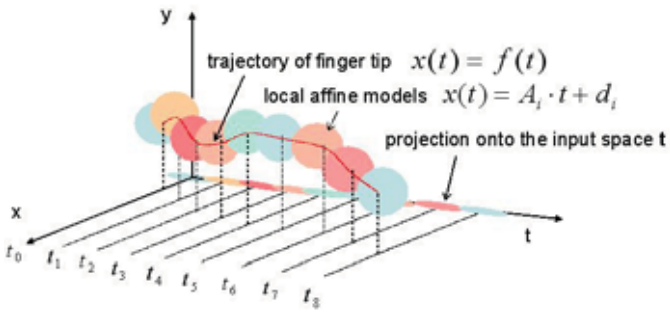


Figure 3: Time-clustering principle for the end-effector and its motion in (x,y)

3.1.1 *Training of time cluster models using new data*

A skill model can be built in several ways

- A single user trains the model by repeating the same skill n times
- m users train the model by repeating the same skill n times

The 1st model is generated by the time sequences $[(t_1, t_2, ..., t_N)_1 ... (t_1, t_2, ..., t_M)_n]$ and the output (end-effector position) sequences $[(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)_1 ... (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M)_n]$.

The 2nd model is generated by the time sequences $[((t_1, t_2, ..., t_N)_1^1 ... (t_1, t_2, ..., t_M)_n^1) \; ... \; ((t_1, t_2, ..., t_N)_1^m ... (t_1, t_2, ..., t_M)_n^m)]$ and the output sequences $[((\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)_1^1 ... (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M)_n^1) \; ... \; ((\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)_1^m ... (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M)_n^m)]$ where m is the number of users in the training process, N, M are lengths of time sequences where $N \approx M$.

Once a particular skill model has been generated it might be necessary to take new data into account. These data may originate from different human operators to cover several ways of performing the same type of skill. Let for simplicity the old model be built by a time sequence $[t_1, t_2, ..., t_N]$ and a respective output sequence $[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$. The old model is then represented by the input cluster centers t_i and the output cluster centers $\mathbf{x}_i$ ($i = 1...c$). It is also described by the parameters $\mathbf{A}_i$ and $\mathbf{d}_i$ of the local linear models. Let $[\tilde{t}_1, \tilde{t}_2, ..., \tilde{t}_M]$, $[\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, ..., \tilde{\mathbf{x}}_M]$ be new training data. A new model can be built by "chaining" old and new training data leading for the time sequences to $[t_1, t_2, ..., t_N, \tilde{t}_1, \tilde{t}_2, ..., \tilde{t}_M]$, and for the output sequences to $[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N, \tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, ..., \tilde{\mathbf{x}}_M]$. The result is a model that involves properties of the old model and the new data. If the old sequence of data is not available, a corresponding sequence can be generated by running the old model with the time instants $[t_1, t_2, ..., t_N]$ as inputs and the end-effector positions $[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$ as outputs.

3.2 *Segmentation principle*

Let for simplicity the signals of a skill be the end-effector coordinates $\mathbf{x}(t) \in R^3$ and the forces $\mathbf{f}(t) \in R^3$ in the end-effector. In order to generate the 'events' determining the time bounds for the phases, $\mathbf{x}(t)$ is differentiated twice and $\mathbf{f}(t)$ only ones by time. The absolute values of the resulting vectors are collected in

$$\mathbf{X}(t) = [|\ddot{\mathbf{x}}(t)|^T, |\dot{\mathbf{f}}(t)|^T]^T. \quad (6)$$

For a segmentation we need the time-discrete case

$$\tilde{\mathbf{X}} = [\mathbf{X}(t_1) ... \mathbf{X}(t_n)] \in R^{6 \times n}. \quad (7)$$

Further define a vector of bounds $\mathbf{B} > 0 \in R^6$ above which $\mathbf{X}(t_i)$ are counted as 'events'. Then a vector $\mathbf{I} = [I_1 ... I_k ... I_m]^T$ is generated where I_k are discrete time stamps t_i for which at least one component of $\mathbf{X}(t_i)$ lies above the corresponding component of the vector of bounds $\mathbf{B}$.

$$I_k = t_i \quad if \quad \mathbf{X}(t_i) > \mathbf{B}. \quad (8)$$

The next step is to select the number of time clusters c and find the clusters by time clustering for the data $\mathbf{Y} = ([\mathbf{X}(I_1); I_1] ... [\mathbf{X}(I_k); I_k] ... [\mathbf{X}(I_m); I_m])$. $\mathbf{Y}$ is a combination of 'events' $\mathbf{X}(t_i)$ and their corresponding time instants I_k. Once the time clusters are found the skill could be cut at these time intants into phases. However, for complicated skills the number of phases c might not be known in advance. Therefore we choose a higher number c and merge those cluster centers who are located close to each other into one.

4 Recognition of robot skills

In this section the recognition of phases or sub-skills ist discussed first. In a second step on the basis of this knowledge - number and type of phases - the construction of the full skill is presented.

4.1 *Recognition of phases using the distance between fuzzy clusters*

Let the model of each phase have the same number of clusters $i = 1...c$ so that each duration T_l ($l = 1...L$) of the l-th phase is divided into $c - 1$ time intervals Δt_i, $i = 2...c$ of the same length. Let the phases be executed in an environment comparable with the modeled phase in order to avoid calibration and re-scaling procedures. Furthermore let

$$V_{model\,l} = [\mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_c]_{model,l}$$
$$\mathbf{X}_i = [x, y, z, f_x, f_y, f_z]_i^T$$

where matrix $V_{model\,l}$ includes the output cluster centers $\mathbf{X}_i$ for the l-th phase model.

A model of the phase to be classified is built by the matrix

$$V_{test} = [\mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_c]_{test,l} \quad (9)$$

ISBN: 978-989-95079-6-8

A decision about which phase is present is made by applying the Euclidean matrix norm

$$N_l = ||V_{model_l} - V_{test}|| \quad (10)$$

Once the unknown phase is classified to the phase model with the smallest norm $min(N_l), l = 1...L$ then the recognition of the phase is finished.

4.2 Recognition of skills using phase models

Once the phases of a test skill are recognized (identified) one should be able to recognize the skill as a whole and finally to reconstruct the hybrid automaton that represents the skill (see Fig. 2). For this purpose a list of possible skills and their phases should be produced. In the following we will discuss the following robot skills
- handling
- contour following
- assembly

The corresponding phases can be found in table 1. Figures 4,5, and 6 show the correspondence between the skills and their individual phases.

Table 1: Skills and phases

Phases Skills:	handling	contour	assembly
1. Grasp object	x		x
2. Free motion	x		
3. Search contact	x	x	x
4. Keep contact	x		
5. Follow with contact	x	x	x
6. Peg-in hole			x
7. release object/contact	x	x	x

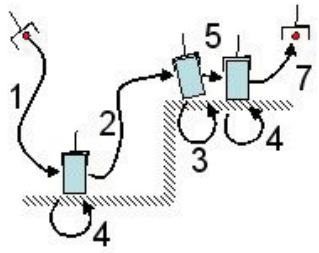


Figure 4: Handling skill

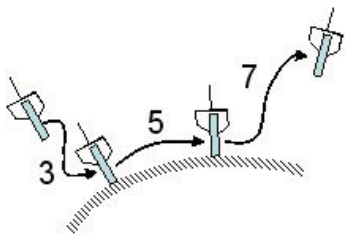


Figure 5: Contour following skill

5 Experiments and simulations

In this section an experimental evaluation of the recognition of phases will be presented. The experimental platform comprises a data glove with diodes mounted at the fingertips and links (see Fig. 7). A system of 4 stereo cameras takes records

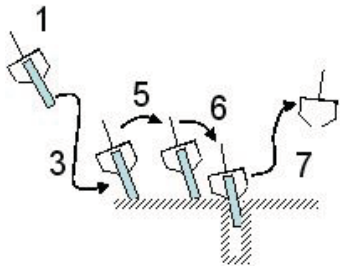


Figure 6: Assembly skill

of the positions of the diodes so that the position of the hand and its fingers can be tracked. In addition, tactile sensors are mounted at each finger tip in order to detect the contact between the fingertips and an object or a surface, respectively. In the experiment only the tip of the index finger is tracked in order to identify the contour performed by this finger. The experiments described here cover only contour following examples but using different contours running at different speeds. Each example consists of three phases: the approach phase, the contour following phase, and the retract phase. The experiment starts with the index fingertip being in contact with a defined start location at a distance from the contour. Then it follows the approach phase without any contact to an object ending at the begin of the contour to be followed. In the contour following phase the index finger follows the contour while the contact is preserved until the end of the contour is reached and the retract phase starts. During the retract phase there is no contact until the index finger reaches the start location again. The experiments can be divided into 3 groups:

A Straight lines
- 1: slow speed (see Fig. 8)
- 2: fast speed
- 3: ramp downhill slow
- 4: ramp downhill fast
- 5: ramp uphill slow
- 6: ramp uphill fast

B Meander
- 7: meander slow (see Fig. 9)
- 8: meander fast

C Loops
- 9: loop slow 1 (see Fig. 10)
- 10: loop slow 2
- 11: loop fast 1

Three modeling examples are shown in Figs. 8 - 10. The blue curves represent the modeled phases whereas the red curves represent the original data. Each phase of a skill is modeled by 15 cluster centers. The crosses depict the cluster centers. It can be observed that the modeling/approximation quality of the fuzzy time models is excellent. The partition of the skill into phases has been done by means of the forces applied to the tip of the index finger. Figure 11 shows the time plots for the meander experiment 7. By means of the force f applied to the fingertip and its derivative df a segmen-

ISBN: 978-989-95079-6-8

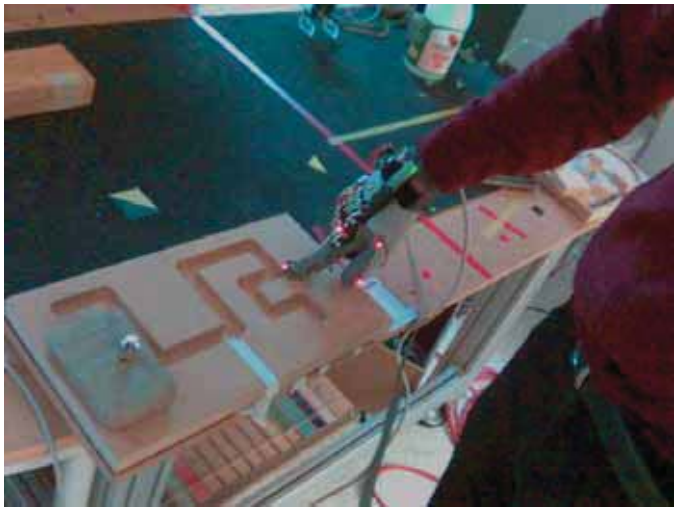


Figure 7: Experiment with a meader-like contour

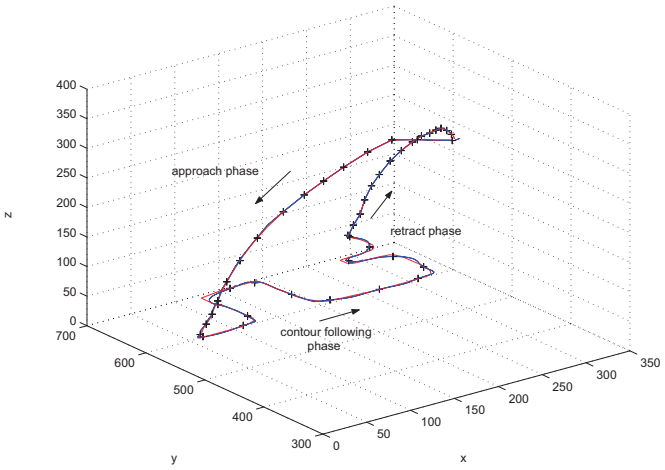


Figure 9: contour following, meander

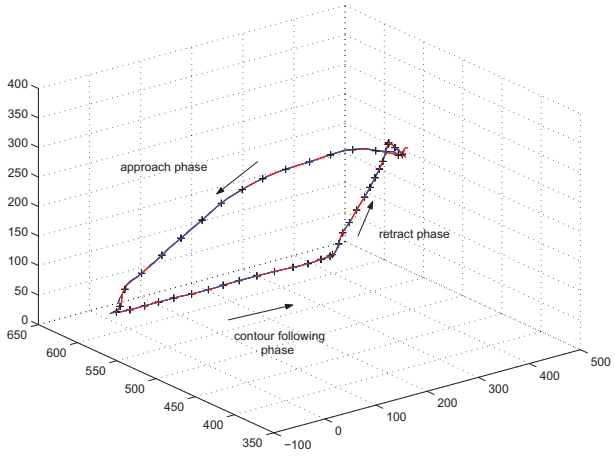


Figure 8: contour following, line

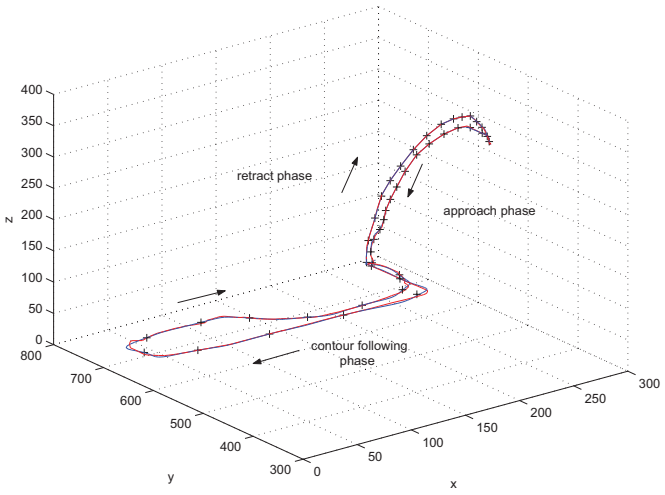


Figure 10: contour following, loop

tation can be done very easily because of the distinct derivatives of the force signals *df*. The recognition of skills has been done by comparing each skill with all other skills using the method described in the last section. Since we only deal with contour following experiments the norms over the whole skill have been taken into account instead of considering the phases seperately. The results are shown in Table 2. To explain the matrix $M(i,j)$ $(i,j = 1...11)$ in Table 2, let us consider experiment 3, *ramp downhill slow*, as an example (3rd row), $M(3,i)$ $(i = 1...11)$. Compared to itself the norm of the differences between model and test skill is zero, $M(3,3) = 0$. The next higher norm difference can be observed for experiment 4, *ramp downhill fast*, $M(3,4) = 0.3$. This corrresponds completely with the idea that a similarity of trajectories should lead to small norm differences. Going through all 11 experiments it turns out that almost all contour following skills can be indentified. One exception is experiment 6 where either skill 1 or 2 are identified instead of skill 5 as expected, $M(6,1) \wedge M(6,2) < M(6,5)$. It can also be noticed that the groups "A: Straight lines", $M(i,j)$ $(i,j = 1...6)$, "B: Meander" , $M(k,l)$ $(k,l = 7,8)$, and "C: Loops" , $M(m,n)$ $(m,n = 9...11)$, can be significantly distinguished from each other. Furthermore, one can see that, from the recognition point of view, the groups A and B are more related with each other than B and C or A and C, respectively.

6 Conclusions

In this paper modeling and recognition/identification of robot skills using fuzzy time modeling have been presented. Recognition of skills is a part of the teach-in process and the Programming-by-Demonstration (PbD) of robot tasks by a human operator. In this paper the focus is directed to the skills "handling", "contour following", and "assembly". In this context a partitioning of skills into phases by segmentation and their modeling using fuzzy time clustering is discussed. The recognition of phases and skills is done by comparing fuzzy time clusters of model skills and test skills. The experimental platform consists of a data glove with diodes mounted at the fingertips and links. 4 stereo cameras track the positions of the hand and its fingers. Tactile sensors are mounted at each fingertip to detect the contact between the fingertips and objects. In the experiments different groups of contour following skills are presented. In 11 experiments several contours like "straight line", "meander", or "loop", respectively,

ISBN: 978-989-95079-6-8

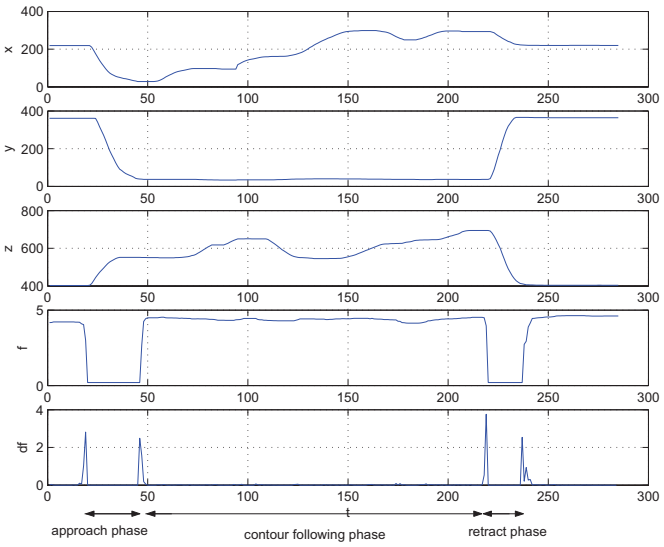


Figure 11: contour following, loop, time plots

Table 2: identification results

skill	1	2	3	4	5	6	7	8	9	10	11
1	0	0.2	0.6	0.5	0.6	0.4	0.7	0.9	1.8	1.8	1.8
2	0.2	0	0.7	0.6	0.6	0.5	0.7	0.9	1.8	1.9	1.8
3	0.5	0.7	0	0.3	0.8	0.6	0.9	1.0	1.9	1.9	1.8
4	0.5	0.6	0.3	0	0.8	0.6	0.8	0.9	1.8	1.9	1.8
5	0.6	0.6	0.8	0.8	0	0.5	0.9	1.0	1.9	2.0	1.9
6	0.4	0.5	0.6	0.6	0.5	0	0.8	1.0	1.8	2.0	1.9
7	0.7	0.7	0.9	0.8	0.9	0.8	0	0.4	1.4	1.4	1.4
8	0.9	0.9	1.0	0.9	1.0	1.0	0.4	0	1.5	1.4	1.3
9	1.8	1.8	1.9	1.8	1.9	1.8	1.4	1.5	0	0.6	0.4
10	1.8	1.9	1.9	1.9	2.0	1.9	1.4	1.4	0.6	0	0.5
11	1.8	1.8	1.8	1.8	1.9	1.9	1.4	1.4	0.4	0.5	0

are performed by the index fingertip. The segmentation has been done by differentiation of the tactile forces. The modeling and recognition results were very good to excellent. In the future more experiments will be done for handling and assembly tasks. The recognition of skills using fuzzy time modeling and hybrid automata will be further developed. A further focus will be the integration of high-level AI planning techniques with low-level robotic control.

References

[1] J. D. Morrow and P. K. Khosla. Manipulation task primitives for composing robot skills. *Proceedings of the 1997 IEEE Intemational Conference on Robotics and Automation, Albuquerque, New Mexico*, pages 3354–3359, 1997.

[2] M. Kaiser and R. Dillmann. Building elementary robot skills from human demonstration. *Proceedings of the 1996 IEEE Intemational Conference on Robotics and Automation, Minneapolis, Minnesota*, 1996.

[3] C. Geib, K. Mourao, R. Petrick, N. Pugeault, M. Steedman, N. Krueger, and F. Woergoetter. Object action complexes as an interface for planning and robot control. In *Proc. IEEE RAS Int Conf. Humanoid Robots(Genova)*, Genova, Dec. 4-6 2006. IEEE, IEEE.

[4] Kwun Han and M. Veloso. Automated robot behavior recognition. *Proceedings of the 1999 ijcai Conference*, 1999.

[5] R. Zoellner, O. Rogalla, R. Dillmann, and J.M. Zoellner. Dynamic grasp recognition within the framework of programming by demonstration. In *Proceedings Robot and Human Interactive Communication,10th IEEE International Workshop*, Bordeaux, Paris, France, September 18-21 2001. IEEE.

[6] S. Ekvall and D. Kragic. Grasp recognition for programming by demonstration. In *Proceedings of International conference on robotics and automation, ICRA 2005*, Barcelona, Spain, 2005.

[7] Ch. Li, L. Khan, and B. Prabhaharan. Real-time classification of variable length multi-attribute motions. *Knowledge and Information Systems*, DOI 10.1007/s10115-005-0223-8, 2005.

[8] O. Bimber. Continuous 6d gesture recognition: A fuzzy-logic approach. *Proceedings of 7-th International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media*, WSCG'99:24–30, 1999.

[9] R. Palm and B. Iliev. Grasp recognition by time clustering, fuzzy modeling, and hidden markov models (hmm) - a comparative study. In *Proceedings FUZZ-IEEE 2008 - IEEE International Conference on Fuzzy Systems*, Hong Kong, July 1-5 2008. IEEE.

[10] Z. Ju, H. Liu, X.Y. Zhu, and Y.L. Xiong. Dynamic grasp recognition using time clustering, gaussian mixture models and hidden markov models. In *Proceedings Intern. Conf. on Intell. Robots and Appl.s, ICIRA-2008*, Wuhan,China, Oct. 15-17 2008. Springer.

[11] R. Palm and B. Iliev. Learning of grasp behaviors for an artificial hand by time clustering and takagi-sugeno modeling. In *Proceedings FUZZ-IEEE 2006 - IEEE International Conference on Fuzzy Systems*, Vancouver, BC, Canada, July 16-21 2006. IEEE.

[12] R. Palm and B. Iliev. Segmentation and recognition of human grasps for programming-by-demonstration using time clustering and takagi-sugeno modeling. In *Proceedings FUZZ-IEEE 2007 - IEEE International Conference on Fuzzy Systems*, London, UK, July 23-26 2007. IEEE.

[13] M.S.Branicky, V.S.Borkar, and S.K.Mitter. A unified framework for hybrid control: Background, model, and theory. *Proceedings of the 11th Intern. Conference on Analysis ond Optimization of Systems. INRIA. France*, June 15-17, 1994.

[14] R. Palm and Ch. Stutz. Open loop dynamic trajectory generator for a fuzzy gain scheduler. *Engineering Applications of Artificial Intelligence*, Vol. 16:213–225, 2003.

[15] T. Takagi and M. Sugeno. Identification of systems and its applications to modeling and control. *IEEE Trans. on Syst., Man, and Cyb.*, Vol. SMC-15. No.1:116–132, January/February 1985.

[16] D.E. Gustafson and W.C. Kessel. Fuzzy clustering with a fuzzy covariance matrix. *Proceedings of the 1979 IEEE CDC*, pages 761–766, 1979.

ISBN: 978-989-95079-6-8

A Consensus Model for Group Decision Making with Incomplete Unbalanced Fuzzy Linguistic Preference Relations

F.J. Cabrerizo[1] J.M. Moreno[2] E. Herrera-Viedma[3]

1.Department of Software Engineering and Computer Systems, Distance Learning University of Spain
28040 Madrid, Spain
2.Department of Information and Communication Engineering, University of Murcia
30100 Murcia, Spain
3.Department of Computer Science and Artificial Intelligence, University of Granada
18071 Granada, Spain
Email: cabrerizo@issi.uned.es, jmmoreno@um.es, viedma@decsai.ugr.es

***Abstract*— *We present a consensus model for group decision making with unbalanced fuzzy linguistic preference relations, i.e., assuming that the preferences are assessed on linguistic term sets whose terms are not symmetrically and uniformly distributed. This consensus model can manage incomplete information situations, that is, situations where the experts do not give all the preference values that they are usually requested. In addition, both consistency and consensus measures are used and it allows to achieve consistent solutions with a great level of agreement.*

***Keywords*—** Group decision making, consensus, unbalanced fuzzy linguistic preference relations, incomplete information, consistency.

1 Introduction

A Group Decision Making (GDM) problem is usually understood as a decision problem which consists in finding the best alternative(s) from a set of feasible alternatives according to the preferences provided by a group of experts characterized by their experience and knowledge. To do this, experts have to express their preferences by means of a set of evaluations over the set of alternatives. In this contribution, we assume that experts use preference relations [1, 2, 3] in an unbalanced fuzzy linguistic context [4, 5] (see Fig. 1).

Figure 1: Example of an unbalanced fuzzy linguistic term set.

In these problems, a difficulty that has to be addressed is the lack of information. Since each expert has his/her own experience concerning the problem being studied, there may be cases where an expert would not be able to express the preference degree between two or more of the available alternatives. This may be due to an expert not possessing a precise or sufficient level of knowledge of part of the problem, or because that expert is unable to discriminate the degree to which some options are better than others.

To solve GDM problems, the experts are faced by applying two processes before obtaining a final solution [6, 7]: *the consensus process* and *the selection process*. The consensus process is defined as a dynamic and iterative group discussion process, coordinated by a moderator helping experts to bring their opinions closer. If the consensus level is lower than a specified threshold, the moderator would urge experts to discuss their opinions further in an effort to bring them closer. Otherwise, the moderator would apply the selection process which consists in obtaining the final solution to the problem from the opinions expressed by the experts. Clearly, it is preferable that the experts achieve a great agreement among their opinions before applying the selection process and, therefore, we focus on the consensus process.

The aim of this paper is to present a consensus model to deal with GDM problems in which experts use incomplete unbalanced fuzzy linguistic preference relations (FLPRs) to provide their preferences. We use two kinds of consensus measures to guide the consensus reaching process, *consensus degrees*, which evaluate the agreement of all the experts, and *proximity measures*, which evaluate the agreement between the experts' individual opinions and the group opinion. However, this consensus model will not only be based on consensus measures but also on consistency measures. To compute them, first, all missing values are estimated using an estimation procedure based on the Tanino's consistency principle [3]. Both consistency and consensus measures are used to design a feedback mechanism, and, in such a way, we substitute the actions of the moderator and give advice to the experts on how they should change and complete their opinions to obtain a solution with a high consensus degree (making experts' opinions closer).

The rest of the paper is set out as follows. Section 2 deals with the preliminaries necessary to develop our consensus model. In Section 3, the consensus model for GDM problems with incomplete unbalanced FLPRs is presented. Finally, some concluding remarks are pointed out in Section 4.

2 Preliminaries

2.1 Methodology to Manage Unbalanced Fuzzy Linguistic Information

To manage unbalanced fuzzy linguistic information, we propose a methodology similar to those proposed in [4, 5]. This methodology is based on the transformation of the unbalanced fuzzy linguistic information in a *Linguistic Hierarchy (LH)* [8], which is the linguistic representation domain that allows us to develop comparison and combination processes of unbalanced fuzzy linguistic information.

A LH is a set of levels, where each level represents a linguistic term set with different granularity from the remaining

ISBN: 978-989-95079-6-8

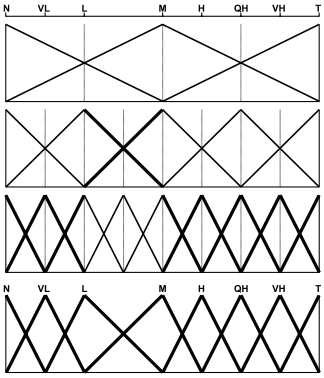


Figure 2: Representation for an unbalanced linguistic fuzzy term set.

levels of the hierarchy. Each level is denoted as $l(t, n(t))$, where t is a number indicating the level of the hierarchy, and $n(t)$ is the granularity of the linguistic term set of t. Then, a LH can be defined as the union of all levels t: $LH = \bigcup_t l(t, n(t))$. Given a LH, we denote as $\mathcal{S}^{n(t)}$ the linguistic term set of LH corresponding to the level t of LH characterized by a cardinality $n(t)$: $\mathcal{S}^{n(t)} = \{s_0^{n(t)}, \ldots, s_{n(t)-1}^{n(t)}\}$.

The procedure to represent unbalanced fuzzy linguistic information presents the following steps:

1. Find a level t^- of LH to represent the subset of linguistic terms $\mathcal{S}_{un}^L$ on the left of the mid linguistic term of unbalanced fuzzy linguistic term set $\mathcal{S}_{un}$.
2. Find a level t^+ of LH to represent the subset of linguistic terms $\mathcal{S}_{un}^R$ on the right of the mid linguistic term of $\mathcal{S}_{un}$.
3. Represent the mid term of $\mathcal{S}_{un}$ using the mid terms of the levels t^- and t^+.

If there does not exist a level t^- or t^+ in LH to represent $\mathcal{S}_{un}^L$ or $\mathcal{S}_{un}^R$, respectively, then the procedure applies the following recursive algorithm, which is defined, in this case, assuming that there does not exist t^-, as it happens with the unbalanced fuzzy linguistic term set given in Fig. 1:

1. Represent $\mathcal{S}_{un}^L$:
   (a) Identify the mid term of $\mathcal{S}_{un}^L$, called $\mathcal{S}_{mid}^L$.
   (b) Find a level t_2^- of the left sets of LH^L to represent the left term subset of $\mathcal{S}_{un}^L$, where LH^L represents the left part of LH.
   (c) Find a level t_2^+ of the right sets of LH^L to represent the right term subset of $\mathcal{S}_{un}^L$.
   (d) Represent the mid term $\mathcal{S}_{mid}^L$ using the levels t_2^- and t_2^+.
2. Find a level t^+ of LH to represent the subset of linguistic terms $\mathcal{S}_{un}^R$.
3. Represent the mid term of $\mathcal{S}_{un}$ using the levels t_2^+ and t^+.

For example, applying this algorithm, the representation of the unbalanced fuzzy linguistic term set $\mathcal{S}_{un} = \{N, VL, L, M, H, QH, VH, T\}$ shown in Fig. 1, using a linguistic hierarchy LH, would be as it is shown in Fig. 2.

To operate with the linguistic information in LH, the 2-tuple fuzzy linguistic model [9] is used.

Definition 2.1. Let $S = \{s_0, \ldots, s_g\}$ be a linguistic term set and $\beta \in [0, g]$ a value representing the result of a symbolic aggregation operation, then the 2-tuple that expresses the equivalent information to β is obtained with the following function $\Delta\colon [0, g] \longrightarrow S \times [-0.5, 0.5)$:

$$(\beta) = (s_i, \alpha), \text{ with } \begin{cases} s_i, & i = round(\beta) \\ \alpha = \beta - i, & \alpha \in [-0.5, 0.5), \end{cases} \quad (1)$$

where $round(\cdot)$ is the usual round operation, s_i has the closest index label to "β", and "α" is the value of the symbolic translation. In addition, for all Δ, there exists Δ^{-1}, defined as $\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta$.

Finally, transformation functions between labels from different levels to make processes of computing with words in multigranular linguistic information contexts without loss of information were defined in [8].

Definition 2.2. [8] Let $LH = \bigcup_t l(t, n(t))$ be a linguistic hierarchy whose linguistic term sets are denoted as $\mathcal{S}^{n(t)} = \{s_0^{n(t)}, \ldots, s_{n(t)-1}^{n(t)}\}$, and let us consider the 2-tuple fuzzy linguistic representation. The transformation function from a linguistic label in level t to a label in level t' is defined as $TF_{t'}^t : l(t, n(t)) \longrightarrow l(t', n(t'))$ such that

$$TF_{t'}^t(s_i^{n(t)}, \alpha^{n(t)}) = \Delta_{t'}\left(\frac{\Delta_t^{-1}(s_i^{n(t)}, \alpha^{n(t)}) \cdot (n(t') - 1)}{n(t) - 1}\right). \quad (2)$$

2.2 Incomplete Unbalanced FLPRs

In this paper, we deal with GDM problems where the experts e_h express their preferences relations $P^h = (p_{ik}^h)$ on the set of alternatives X using an unbalanced linguistic fuzzy term set, $\mathcal{S}_{un} = \{s_0, \ldots, s_{mid}, \ldots, s_g\}$, which has a minimum label, called s_0, a maximum label, called s_g, and the remaining labels are non-uniformly and non-symmetrically distributed around the central one, called s_{mid} (Fig. 1). Therefore, $p_{ik}^h \in \mathcal{S}_{un}$ represents the preference of alternative x_i over alternative x_k for the experts e_h assessed on the unbalanced fuzzy linguistic term set $\mathcal{S}_{un}$.

Definition 2.3. An unbalanced FLPR P^h on a set of alternatives X is characterized by a membership function $\mu_{P^h} : X \times X \longrightarrow \mathcal{S}_{un}$. If it is not possible to give the preference degree for every pair of alternatives, we have an incomplete unbalanced FLPR.

When cardinality of X is small, the preference relation may be conveniently represented by a $n \times n$ matrix $P^h = (p_{ik}^h)$, being $p_{ik}^h = \mu_{P^h}(x_i, x_k), \forall i, k \in \{1, \ldots, n\}$ and $p_{ik}^h \in \mathcal{S}_{un}$.

2.3 Consistency Measures

For GDM problems with preference relations, some properties are usually assumed desirable to avoid contradictions within the preferences expressed by the experts, that is, to avoid inconsistent opinions. One of them is the *additive transitivity*,

ISBN: 978-989-95079-6-8

$$cp_{ik}^h = TF_t^{t'}(\Delta_{t'}\left(\frac{\sum_{j=1;i\neq k\neq j}^{n}(\Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j1})) + \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j2})) + \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j3})))}{3(n-2)}\right)). \quad (3)$$

$$cp_{ik}^h = TF_t^{t'}(\Delta_{t'}\left(\frac{\sum_{j\in H_{ik}^{h1}}\Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j1})) + \sum_{j\in H_{ik}^{h2}}\Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j2})) + \sum_{j\in H_{ik}^{h3}}\Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j3}))}{(\#H_{ik}^{h1} + \#H_{ik}^{h2} + \#H_{ik}^{h3})}\right)). \quad (4)$$

which was defined for fuzzy preference relations [1, 3] as:

$$(p_{ij}^h - 0.5) + (p_{jk}^h - 0.5) = (p_{ik}^h - 0.5),\ \forall i,j,k \in \{1,\ldots,n\}. \quad (5)$$

In the case of an unbalanced fuzzy linguistic context, previously to carry out any computation task, we have to choose a level $t' \in \{t^-, t_2^-, t^+, t_2^+\}$, such that $n(t') = max\{n(t^-), n(t_2^-), n(t^+), n(t_2^+)\}$. Then, once a result is obtained, it is transformed to the correspondent level $t \in \{t^-, t_2^-, t^+, t_2^+\}$ by means of $TF_t^{t'}$ for expressing the result in the unbalanced fuzzy linguistic term set $\mathcal{S}_{un}$. In this way, the unbalanced fuzzy linguistic additive transitivity for unbalanced FLPRs is defined as:

$$\begin{array}{l} TF_t^{t'}(\Delta_{t'}[(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ij}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid})))+ \\ (\Delta_{t'}^{-1}(TF_{t'}^t(p_{jk}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid})))]) = \\ TF_t^{t'}(\Delta_{t'}[(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid})))]), \end{array} \quad (6)$$

being $p_{ij}^h = (s_v^{n(t)}, \alpha_1)$, $t \in \{t^-, t_2^-, t^+, t_2^+\}$, $p_{jk}^h = (s_w^{n(t)}, \alpha_2)$, $t \in \{t^-, t_2^-, t^+, t_2^+\}$, $p_{ik}^h = (s_z^{n(t)}, \alpha_3)$, $t \in \{t^-, t_2^-, t^+, t_2^+\}$, s_{mid} is the mid term of $\mathcal{S}_{un}$ and $t' \in \{t^-, t_2^-, t^+, t_2^+\}$.

Expression (6) can be rewritten as:

$$\begin{array}{l} p_{ik}^h = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ij}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(p_{jk}^h)) - \\ \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0)))),\ \ \forall i,j,k \in \{1,\ldots,n\}. \end{array} \quad (7)$$

Equation (7) can be used to calculate an estimated value of a preference degree p_{ik}^h $(i \neq k)$ using an intermediate alternative x_j in three different ways:

1. From $p_{ik}^h = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ij}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(p_{jk}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0))))$, we obtain:

$$\begin{array}{l} (cp_{ik}^h)^{j1} = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ij}^h)) + \\ \Delta_{t'}^{-1}(TF_{t'}^t(p_{jk}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0)))). \end{array} \quad (8)$$

2. From $p_{jk}^h = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ji}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0))))$, we obtain:

$$\begin{array}{l} (cp_{ik}^h)^{j2} = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{jk}^h)) - \\ \Delta_{t'}^{-1}(TF_{t'}^t(p_{ji}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0)))). \end{array} \quad (9)$$

3. From $p_{ij}^h = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(p_{kj}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0))))$, we obtain:

$$\begin{array}{l} (cp_{ik}^h)^{j3} = TF_t^{t'}(\Delta_{t'}(\Delta_{t'}^{-1}(TF_{t'}^t(p_{ij}^h)) - \\ \Delta_{t'}^{-1}(TF_{t'}^t(p_{kj}^h)) + \Delta_{t'}^{-1}(TF_{t'}^t(s_{mid}, 0)))). \end{array} \quad (10)$$

The overall estimated value cp_{ik}^h of p_{ik}^h is obtained as the average of all possible $(cp_{ik}^h)^{j1}$, $(cp_{ik}^h)^{j2}$ and $(cp_{ik}^h)^{j3}$ values as shown in (3).

When the information provided is completely consistent, then $(cp_{ik}^h)^{jl} = p_{ik}^h$, $\forall j,l$. The error between a preference value and its estimated one in $[0,1]$ is defined as follows:

$$\varepsilon p_{ik}^h = \frac{|\Delta_{t'}^{-1}(TF_{t'}^t(cp_{ik}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^h))|}{n(t') - 1}. \quad (11)$$

We should point out that some estimated values of an incomplete unbalanced FLPR could lie outside the $\mathcal{S}_{un}$, i.e., we may have $cp_{ik}^h < s_0$ or $cp_{ik}^h > s_g$. In order to normalize the expression domains, the following function is used:

$$f(cp_{ik}^h) = \begin{cases} s_0, & \text{if } cp_{ik}^h < s_0 \\ s_g, & \text{if } cp_{ik}^h > s_g \\ cp_{ik}^h, & \text{otherwise.} \end{cases} \quad (12)$$

Thus, it can be used to define the consistency level between the preference degree p_{ik}^h and the rest of the preference values of the unbalanced FLPR as follows:

$$cl_{ik}^h = 1 - \varepsilon p_{ik}^h. \quad (13)$$

Easily, we can define the consistency measures for particular alternatives and for the whole unbalanced FLPR.

Definition 2.4. The consistency measure, $cl_i^h \in [0,1]$, associated to a particular alternative x_i of an unbalanced FLPR P^h is defined as:

$$cl_i^h = \frac{\sum_{k=1;i\neq k}^{n}(cl_{ik}^h + cl_{ki}^h)}{2(n-1)}. \quad (14)$$

Definition 2.5. The consistency level, $cl^h \in [0,1]$, of an unbalanced FLPR P^h is defined as follows:

$$cl^h = \frac{\sum_{i=1}^{n} cl_i^h}{n}. \quad (15)$$

When working with an incomplete unbalanced FLPR, (3) cannot be used to obtain the estimate of a known preference value. In this case, the following sets can be defined [10]:

$$\begin{array}{l} A = \{(i,j) \mid i,j \in \{1,\ldots,n\} \wedge i \neq j\} \\ MV^h = \{(i,j) \in A \mid p_{ij}^h \text{ is unknown}\} \\ EV^h = A \setminus MV^h \\ H_{ik}^{h1} = \{j \neq i,k \mid (i,j),(j,k) \in EV^h\} \\ H_{ik}^{h2} = \{j \neq i,k \mid (j,i),(j,k) \in EV^h\} \\ H_{ik}^{h3} = \{j \neq i,k \mid (i,j),(k,j) \in EV^h\} \\ EV_i^h = \{(a,b) \mid (a,b) \in EV^h \wedge (a = i \vee b = i)\}, \end{array} \quad (16)$$

Then, the estimated value of a particular preference degree p_{ik}^h $((i,k) \in EV^h)$ can be calculated as shown in (4) assuming $(\#H_{ik}^{h1} + \#H_{ik}^{h2} + \#H_{ik}^{h3}) \neq 0$.

ISBN: 978-989-95079-6-8

function estimate_p(h,i,k)

1) $(cp_{ik}^h)^1 = (s_0, 0),\ (cp_{ik}^h)^2 = (s_0, 0),\ (cp_{ik}^h)^3 = (s_0, 0)$

2) if $\#H_{ik}^{h1} \neq 0$, then $(cp_{ik}^h)^1 = TF_t^{t'}(\Delta_{t'}(\sum_{j \in H_{ik}^{h1}} \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j1}))))$

3) if $\#H_{ik}^{h2} \neq 0$, then $(cp_{ik}^h)^2 = TF_t^{t'}(\Delta_{t'}(\sum_{j \in H_{ik}^{h2}} \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j2}))))$

4) if $\#H_{ik}^{h3} \neq 0$, then $(cp_{ik}^h)^3 = TF_t^{t'}(\Delta_{t'}(\sum_{j \in H_{ik}^{h3}} \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^{j3}))))$

5) Calculate $cp_{ik}^h = TF_t^{t'}(\Delta_{t'}\left(\frac{\Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^1)) + \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^2)) + \Delta_{t'}^{-1}(TF_{t'}^t((cp_{ik}^h)^3))}{(\#H_{ik}^{h1} + \#H_{ik}^{h2} + \#H_{ik}^{h3})}\right))$

end function

2.4 Estimation Procedure of Missing Values

The procedure estimates missing information in an expert's incomplete unbalanced FLPR using only the preference values provided by that particular expert. It is designed using (4) and estimates missing information values by means of two different tasks:

1. **Elements to be estimated in step *t* of the procedure:**

$$EMV_t^h = \{(i,k) \in MV^h \setminus \bigcup_{l=0}^{t-1} EMV_l^h \mid i \neq k \wedge \exists j \in \{H_{ik}^{h1} \cup H_{ik}^{h2} \cup H_{ik}^{h3}\}\}, \tag{17}$$

   and $EMV_0^h = \emptyset$ (by definition).

2. **Estimation of a particular missing value:** In order to estimate a particular value, the function $estimate\_p(h,i,k)$ at the top of the page is used.

3 Consensus model

In this section, we present a consensus model for GDM problems where experts provide their preferences using incomplete unbalanced FLPRs. To solve GDM problems with this kind of preference relations, firstly, it is necessary to deal with the missing values [10]. The previous consistency based procedure of missing values allows us to measure the consistency levels of each expert. This consistency information is used in this section to propose a consensus model based not only on consensus criteria but also on consistency criteria. We consider that both criteria are important to guide the consensus process in an incomplete decision framework. In such a way, we get that experts change their opinions toward agreement positions in a consistent way, which is desirable to achieve a consistent and consensus solution. The main characteristics of the proposed consensus model are the following:

- It is designed to guide the consensus process of incomplete unbalanced fuzzy linguistic GDM problems.
- It uses a consistency based procedure to calculate the incomplete unbalanced fuzzy linguistic information.
- It is based both consensus criteria and consistency criteria. The proposed consensus model is designed with the aim of obtaining the maximum possible consensus level while trying to achieve a high level of consistency in experts' preferences.
- A feedback mechanism is defined using the above criteria. It substitutes the moderator's actions, avoiding the possible subjectivity that he/she can introduce, and gives advice to the experts to find out the changes they need to make in their opinions to obtain a solution with certain consensus and consistency degrees simultaneously.

In particular, the consensus model develops its activity in five phases that will be described in further detail in the following subsections: 1) computing missing information, 2) computing consistency measures, 3) computing consensus measures, 4) controlling the consistency/consensus state, and 5) feedback mechanism.

3.1 Computing Missing Information

In this first step, for each incomplete unbalanced FLPR P^h, we obtain its corresponding complete unbalanced FLPR $\bar{P}^h$ using the estimation procedure described in Section 2.4.

3.2 Computing Consistency Measures

To compute consistency measures, first, for each $\bar{P}^h$, we compute its corresponding unbalanced FLPR $CP^h = (cp_{ik}^h)$ according to expression (3). Second, we apply (13)-(15) to $(\bar{P}^h, CP^h)$ $(\forall h)$ to compute the consistency measures $CL^h = (cl_{ik}^h)$, cl_i^h, cl^h, $\forall i,k \in \{1,\ldots,n\}$. Finally, we define a global consistency measure among all experts to control the global consistency situation as follows:

$$CL = \frac{\sum_{h=1}^m cl^h}{m}. \tag{18}$$

3.3 Computing Consensus Measures

As in [6, 10], we compute two different kinds of measures: consensus degrees and proximity measures. Consensus degrees are used to measure the actual level of consensus in the process, while the proximity measures give information about how close to the collective solution every expert is. These measures are given on three different levels for a preference relation: pairs of alternatives, alternatives and relation. It will allow us to find out the consensus state of the process at different levels. For example, we will be able to identify which experts are close to the consensus solution, or in which alternatives the experts are having more trouble to reach consensus.

3.3.1 Consensus Degrees

For each pair of experts (e_h, e_l) $(h = 1,\ldots,m-1,\ l = h+1,\ldots,m)$, a similarity matrix, $SM^{hl} = (sm_{ik}^{hl})$, is defined, where

$$sm_{ik}^{hl} = 1 - \frac{|\Delta_{t'}^{-1}(TF_{t'}^t(\bar{p}_{ik}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(\bar{p}_{ik}^l))|}{n(t') - 1}, \tag{19}$$

being $\bar{p}_{ik}^h = (s_v^{n(t)}, \alpha_1),\ t \in \{t^-, t_2^-, t^+, t_2^+\},\ \bar{p}_{ik}^l = (s_w^{n(t)}, \alpha_2), t \in \{t^-, t_2^-, t^+, t_2^+\}$, and $t' \in \{t^-, t_2^-, t^+, t_2^+\}$.

ISBN: 978-989-95079-6-8

Then, a consensus matrix, $CM = (cm_{ik})$, is calculated by aggregating all the similarity matrices using the arithmetic mean as the aggregation function ϕ:

$$cm_{ik} = \phi(sm_{ik}^{hl},\ h = 1, \ldots, m-1,\ l = h+1, \ldots, m). \quad (20)$$

Once the consensus matrix, CM, is computed, we proceed to calculate the consensus degrees at the three different levels:

1. **Level 1.** *Consensus degree on pairs of alternatives,* cop_{ik}. It measures the consensus degree amongst all the experts on the pair of alternatives (x_i, x_k).

$$cop_{ik} = cm_{ik};\ \forall\, i, k = 1, \ldots, n \wedge i \neq k. \quad (21)$$

2. **Level 2.** *Consensus degree on alternatives,* ca_i. It measures the consensus degree amongst all the experts on the alternative x_i.

$$ca_i = \frac{\sum_{k=1;k\neq i}^{n} (cop_{ik} + cop_{ki})}{2(n-1)}. \quad (22)$$

3. **Level 3.** *Consensus degree on the relation,* cr. It measures the global consensus degree amongst all the experts and is used to control the consensus situation.

$$cr = \frac{\sum_{l=1}^{n} ca_i}{n}. \quad (23)$$

3.3.2 Proximity Measures

These measures evaluate the agreement between the individual experts' opinions and the group opinion. To compute them for each expert, we need to obtain the collective unbalanced FLPR, $P^c = (p_{ik}^c)$, which summarizes preferences given by all the experts and is calculated by means of the aggregation of the set of individual unbalanced FLPRs $\{\bar{P^1}, \ldots, \bar{P^m}\}$. In this way, to obtain P^c we use the unbalanced fuzzy linguistic version of an IOWA operator [11, 12], which uses both consensus and consistency criteria as inducing variable. Thus, we obtain each collective unbalanced fuzzy linguistic preference degree p_{ik}^c according to the most consistent and consensual individual unbalanced fuzzy linguistic preference degrees.

Thus, to obtain each p_{ik}^c according to the most consistent and consensual individual unbalanced fuzzy linguistic preference degrees, we propose to use an unbalanced fuzzy linguistic IOWA operator with the consistency/consensus values, $\{z_{ik}^1, z_{ik}^2, \ldots, z_{ik}^m\}$, as the values of the order inducing variable, i.e.,

$$\begin{aligned} p_{ik}^c = \Phi_W(\langle z_{ik}^1, \bar{p}_{ik}^1\rangle, \ldots, \langle z_{ik}^m, \bar{p}_{ik}^m\rangle) = \\ TF_t^{t'}(\Delta_{t'}(\textstyle\sum_{h=1}^{m} w_h \cdot \Delta_{t'}^{-1}(TF_{t'}^t(\bar{p}_{ik}^{\sigma(h)})))), \end{aligned} \quad (24)$$

where

- σ is a permutation of $\{1, \ldots, m\}$ such that $z_{ik}^{\sigma(h)} \geq z_{ik}^{\sigma(h+1)}$, $\forall h = 1, \ldots, m-1$, i.e., $\langle z_{ik}^{\sigma(h)}, \bar{p}_{ik}^{\sigma(h)}\rangle$ is the 2-tuple with $z_{ik}^{\sigma(h)}$ the h-th highest value in the set $\{z_{ik}^1 \ldots, z_{ik}^m\}$;
- the weighting vector is computed according to the following expression:

$$w_h = Q\left(\frac{\sum_{j=1}^{h} z_{ik}^{\sigma(j)}}{T}\right) - Q\left(\frac{\sum_{j=1}^{h-1} z_{ik}^{\sigma(j)}}{T}\right), \quad (25)$$

  with $T = \sum_{j=1}^{m} z_{ik}^j$;
- and the set of values of the inducing variable $\{z_{ik}^1, \ldots, z_{ik}^m\}$ are computed as follows:

$$z_{ik}^h = (1-\delta)\cdot cl_{ik}^h + \delta \cdot co_{ik}^h, \quad (26)$$

  being co_{ik}^h the consensus measure for the preference value $\bar{p}_{ik}^h$ and $\delta \in [0,1]$ a parameter to control the weight of both consistency and consensus criteria in the inducing variable. Usually $\delta > 0.5$ will be used to give more importance to the consensus criterion. We should note that in our framework, each value co_{ik}^h used to calculate $\{z_{ik}^1, \ldots, z_{ik}^m\}$ is defined as follows:

$$co_{ik}^h = \frac{\sum_{l=h+1}^{n} sm_{ik}^{hl} + \sum_{l=1}^{h-1} sm_{ik}^{lh}}{n-1}. \quad (27)$$

Once we have computed P^c, we can compute the proximity measures in each level of an unbalanced FLPR.

1. **Level 1.** *Proximity measure on pairs of alternatives,* pp_{ik}^h. The proximity measure of an expert e_h on a pair of alternatives (x_i, x_k) to the group's one is calculated as:

$$pp_{ik}^h = 1 - \frac{\left|\Delta_{t'}^{-1}(TF_{t'}^t(\bar{p}_{ik}^h)) - \Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^c))\right|}{n(t')-1}. \quad (28)$$

   being $\bar{p}_{ik}^h = (s_v^{n(t)}, \alpha_1)$, $t \in \{t^-, t_2^-, t^+, t_2^+\}$, $p_{ik}^c = (s_w^{n(t)}, \alpha_2)$, $t \in \{t^-, t_2^-, t^+, t_2^+\}$, and $t' \in \{t^-, t_2^-, t^+, t_2^+\}$.

2. **Level 2.** *Proximity measure on alternatives,* pa_i^h. The proximity measure of an expert e_h on an alternative x_i to the group's one is calculated as follows:

$$pa_i^h = \frac{\sum_{k=1;k\neq i}^{n} (pp_{ik}^h + pp_{ki}^h)}{2(n-1)}. \quad (29)$$

3. **Level 3.** *Proximity measure on the relation,* pr^h. The proximity measure of an expert e_h on his/her unbalanced FLPR to the group's one is calculated as follows:

$$pr^h = \frac{\sum_{l=1}^{n} pa_i^h}{n}. \quad (30)$$

3.4 Controlling Consistency/Consensus State

The consistency/consensus state control process will be used to decide when the feedback mechanism should be applied to give advice to the experts or when the consensus reaching process has to come to an end. It should take into account both the consensus and consistency measures. To do that, we use a measure or level of satisfaction, called *consistenty/consensus level* (CCL) [10], which is used as a control parameter:

$$CCL = (1-\delta)\cdot CL + \delta \cdot cr, \quad (31)$$

with δ the same value used in [7]. When CCL satisfies a minimum threshold value $\gamma \in [0,1]$, the consensus reaching process finishes and the selection process can be applied. To avoid that the consensus process does not converge, a maximum number of consensus rounds is incorporated.

ISBN: 978-989-95079-6-8

3.5 Feedback Mechanism

The feedback mechanism generates personalized advice to the experts according to the consistency and consensus criteria. It helps experts to change their preferences and to complete their missing values. This activity is carried out in two steps:

1. **Identification of the preference values.** We must identify preference values that are contributing less to reach a high consensus/consistency state. To do that, we define set APS that contains 3-tuples (h, i, k) symbolizing preference degrees p_{ik}^h that should be changed because they affect badly to that consistency/consensus state.

   (a) *Identification of experts.* We identify the set of experts EXPCH that should receive advice on how to change some of their preference values.

   $$EXCPCH = \{h \mid (1-\delta) \cdot cl^h + \delta \cdot pr^h < \gamma\}. \quad (32)$$

   (b) *Identification of alternatives.* We identify the alternatives ALT that the above experts should consider to change.

   $$ALT = \{(h,i) \mid h \in EXPCH \wedge (1-\delta) \cdot cl_i^h + \delta \cdot pa_i^h < \gamma\}. \quad (33)$$

   (c) *Identification of pairs of alternatives.* Finally, we identify preference values for every alternative and expert $(x_i; e_h \mid (h,i) \in ALT)$ that should be changed according to their proximity and consistency measures on the pairs of alternatives, i.e.,

   $$APS = \{(h,i,k) \mid (h,i) \in ALT \wedge (1-\delta) \cdot cl_{ik}^h + \delta \cdot pp_{ik}^h < \gamma\}. \quad (34)$$

   Additionally, the feedback process must provide rules for missing preference values. To do so, it has to take into account in APS all missing values that were not provided by the experts, i.e.,

   $$APS' = APS \cup \{(h,i,k) \mid p_{ik}^h \in MV_h\}. \quad (35)$$

2. **Generation of advice.** In this step, the feedback mechanism generates personalized recommendations to help the experts to change their preferences. These recommendations are based on easy recommendation rules that will not only tell the experts which preference values they should change, but will also provide them with particular values for each preference to reach a higher consistency/consensus state.

   The new preference degree of alternatives x_i over alternative x_k to recommend to the expert e_h, rp_{ik}^h, is calculated as the following weighted average of the preference value cp_{ik}^h and the collective preference value p_{ik}^c:

   $$rp_{ik}^h = TF_t^{t'}(\Delta_{t'}((1-\delta) \cdot \Delta_{t'}^{-1}(TF_{t'}^t(cp_{ik}^h)) + \delta \cdot \Delta_{t'}^{-1}(TF_{t'}^t(p_{ik}^c)))). \quad (36)$$

   As previously mentioned, with $\delta > 0.5$, the consensus model leads the experts towards a consensus solution rather than towards an increase on their own consistency levels.

   Finally, we should distinguish two cases:

   (a) $\forall (h,i,k) \in APS'$, if $p_{ik}^h \in EV_h$, the recommendation generated for the expert e_h is: "You should change your preference value (i,k) to a value close to rp_{ik}^h".

   (b) $\forall (h,i,k) \in APS'$, if $p_{ik}^h \in MV_h$, the recommendation generated for the expert e_h is: "You should provide a value for (i,k) close to rp_{ik}^h".

4 Concluding Remarks

In this paper, we have proposed a model of consensus for GDM with incomplete unbalanced fuzzy linguistic information. It uses two different kinds of measures to guide the consensus reaching process, consistency and consensus measures, and generates advice to experts in a discriminate way. As a consequence, this model allows us to achieve consistent and consensus solutions. In addition, it supports the consensus process automatically, without moderator.

Acknowledgment

This paper has been developed with the financing of SAINFOWEB project (TIC00602), Feder Funds in FUZZYLING project (TIN2007-61079) and PETRI project (PET2007-0460).

References

[1] E. Herrera-Viedma et al. Some issues on consistency of fuzzy preference relations. *European Journal of Operational Research*, 154(1):98–109, 2004.

[2] Th.L. Saaty. *The Analytic Hierarchy Process*. MacGraw-Hill, New York, 1980.

[3] T. Tanino. Fuzzy preference orderings in group decision making. *Fuzzy Sets and Systems*, 12:117–131, 1984.

[4] F. Herrera et al. A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE Transactions on Fuzzy Systems*, 16(2):354–370, 2008.

[5] E. Herrera-Viedma and A.G. López-Herrera. A model of information retrieval system with unbalanced fuzzy linguistic information. *International Journal of Intelligent Systems*, 22(11):1197–1214, 2007.

[6] F. Herrera et al. A model of consensus in group decision making under linguistic assessments. *Fuzzy Sets and Systems*, 78(1):73–87, 1996.

[7] J. Kacprzyk et al. Group decision making and consensus under fuzzy preferences and fuzzy majority. *Fuzzy Sets and Systems*, 49(1):21–31, 1992.

[8] F. Herrera and L. Martínez. A model based on linguistic 2-tuples for dealing with multigranularity hierarchical linguistic contexts in multiexpert decision-making. *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, 31(2):227–234, 2001.

[9] F. Herrera and L. Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems*, 8(6):746–752, 2000.

[10] E. Herrera-Viedma et al. A consensus model for group decision making with incomplete fuzzy preference relations. *IEEE Transactions on Fuzzy Systems*, 15(5):863–877, 2007.

[11] R.R. Yager. Induced aggregation operators. *Fuzzy Sets and Systems*, 137(1):59–69, 2003.

[12] R.R. Yager and D.P. Filev. Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(2):141–150, 1999.

ISBN: 978-989-95079-6-8

Real-time segmentation of moving objects in H.264 compressed domain with dynamic design of fuzzy sets

C. Solana-Cipres, L. Rodriguez-Benitez[1] J. Moreno-Garcia[2] L. Jimenez[1] G. Fernandez-Escribano[3]

1.Escuela Superior de Informatica, Universidad de Castilla-La Mancha
Ciudad Real, Spain
2.Escuela Universitaria de Ingenieria Tecnica Industrial, Universidad de Castilla-La Mancha
Toledo, Spain
3.Instituto de Investigacion en Informatica, Universidad de Castilla-La Mancha
Albacete, Spain
Email: {cayetanoj.solana, luis.rodriguez, luis.jimenez, juan.moreno, gerardo.fernandez}@uclm.es

Abstract— *This paper presents a real-time segmentation algorithm to obtain moving objects from the H.264 compressed domain. The proposed segmentation works with very little information and is based on two features of the H.264 compressed video: motion vectors associated to the macroblocks and decision modes. The algorithm uses fuzzy logic and allows to describe position, velocity and size of the detected regions in a comprehensive way, so the proposed approach works with low level information but manages highly comprehensive linguistic concepts. The performance of the algorithm is improved using dynamic design of fuzzy sets that avoids merge and split problems. Experimental results for several traffic scenes demonstrate the real-time performance and the encouraging results in diverse situations.*

Keywords— Moving object detection, image segmentation, H.264 advanced video coding, dynamic fuzzy sets.

1 Introduction

In Computer Vision, moving object segmentation is referred to the process of identifying and partitioning of the meaningful regions present in video sequences. Classical techniques like *background subtraction* or *temporal differencing* use low level information about each pixel. These algorithms have a good performance and can identify with accuracy the border of the regions, but the most of them can not work in real-time because they have to decode and process each pixel of each frame. So fast algorithms to segment moving objects usually work directly on compressed video. The proposed algorithm uses only the information related to the motion and works with groups of pixels called *macroblocks*.

The proposed segmentation algorithm is designed for the H.264 advanced video coding standard. This standard has a better compression ratio than the MPEG compressed video family standards and provides the mechanisms for coding video that are optimized for compression efficiency and aims to meet the needs of practical multimedia communication applications. This new standard has more choices of coding parameters and strategies than MPEG-4. This work is based on previous algorithms for MPEG [1, 2], but the new proposal is faster and can fulfill the requeriment of real-time applications. Another contribution of this work is the improvement of the algorithm's robustness. This improvement can be obtained because the proposed technique can adapt itself to the scenario and the objects present in it in a dynamic way. More concretely, the fuzzy sets describing the linguistic variables are updated dynamically and this fact allows to reduce the merge and split ratios.

The paper is organized as follows. Section 2 describes the approaches to the H.264 segmentation algorithms. In the next section, some features of the H.264 video standard are presented. Fuzzy linguistic concepts are shown in Section 4. Later, in Section 5 it is analyzed the proposed method of moving object segmentation. The experimental results in several video traffic scenes are presented in Section 6. Finally, conclusions and future works are described in Section 7.

2 Related Work

The main problem related to the compressed video processing is the shortage of information and the uncertainty that it implies. In this segmentation method luminance and chrominance values are not used, but information related to the motion compensation between frames is needed. This fact implies the increase of the uncertainty related to the input information. In this noisy environment, fuzzy logic presents itself as the adequate theoretical medium to reduce the negative influence of the uncertainty.

Fuzzy logic has been applied in many tasks in computer vision [3, 4] because of its potential and its link with the natural language. It allows to translate video information in linguistic representations capable to be processed through fuzzy techniques. Approximate reasoning is used in this work to carry out a segmentation algorithm of moving objects that obtains conceptual representations which describe position, velocity and size of the regions detected in a comprehensive way.

The segmentation algorithms works directly over H.264 compressed domain video and exploit two features of macroblock: motion vectors and DCT coefficients. Zeng *et al.* [5] present an algorithm that employs a block-based Markov Random Field (MRF) model to segment moving objects from the sparse motion vector field obtained directly from the H.264 bitstream. A later approximation has been proposed in Liu *et al.* [6], where a real-time spatiotemporal segmentation is presented. In this case, spatial segmentation only exploits the motion vector field extracted from the H.264 compressed video. Regions are classified using the block residuals of global motion compensation and the projection is exploited for interframe tracking of video objects.

ISBN: 978-989-95079-6-8

3 Features of the H.264 Advanced Video Coding

H.264 [7], also known as MPEG-4 Part 10, is a standard for video compression developed jointly between the *Motion Picture Expert Group* (MPEG) and the *Video Coding Experts Group* (VCEG). Richardson explains deeply in [8] the features of the H.264 compressed domain, but in this section only the H.264 concepts used into the segmentation algorithm are described. This standard provides mechanisms for video coding that are optimized for a better compression efficiency and aims to meet the multimedia communication applications.

The pictures in a video sequence may be encoded in several ways. The goal of the different kinds of frames is to exploit the diverse class of redundant information that it is possible to find in a video sequence. Three of the most common kinds of pictures available in H.264, as well as in other multimedia standards are **Intra pictures** (*I Type*), **Predicted pictures** (*P Type*) and **Bi-directional or Interpolated pictures** (*B Type*).

The most relevant concept to this work is the **macroblock** (MB), which is the basic unit in which an image is divided into. It contains the information of a 16x16 pixels region and there are two types depending on the encoding: **Intra macroblock**, in which Intra-prediction algorithms are applied directly to exploit the spatial redundancy according to the H.264 standard, and **Inter macroblock**, in which motion compensation is used to exploit temporal redundancy from a reference macroblock (earlier, later or a combination of both).

H.264 uses *block-based motion compensation*, the same principle adopted by every major coding standard since H.261. This motion compensation is done through the redundant information between consecutive frames looking for a pattern that captures the kind of movement between pictures. This pattern is represented as a **motion vector**, which defines a distance and a direction and has two dimensions: *right_x* and *down_x*. Important differences from earlier standards include the support for a range of block sizes and the use of multiple reference frames to improve the performance of the coding.

H.264 supports motion compensation block sizes ranging from 16x16 to 4x4 samples. Each macroblock may be split up in 4 ways: 16x16, 16x8, 8x16 or 8x8. Each of the subdivided regions is a **macroblock partition**. If the 8x8 mode is chosen, each of the four 8x8 macroblock partitions within the macroblock may be further split in 4 ways: 8x8, 8x4, 4x8 or 4x4 (known as **sub-macroblock partitions**). These partitions and sub-partitions give rise to a large number of possible combinations within each macroblock. This method of partitioning macroblocks into motion compensated sub-blocks of varying size is known as *tree structured motion compensation*.

Since each macroblock and sub-macroblock partition has a motion vector associated, a macroblock has from 0 to 16 pairs of *motion vectors*. For example, an Intra macroblock has not any motion vector, an Inter macroblock partitioned into two 16x8 blocks has two pairs of motion vectors and an Inter macroblock partitioned into four 8x8 blocks -each of them partitioned into 4x4 sub-macroblocks- has 16 pairs of motion vectors ($4 * 4 = 16$).

Since H.264 allows block-based motion compensation, the *decision mode* and the *block size* give additional information about the motion degree on a specific region of a frame (Fig. 1). Thus the more partitions into a macroblock, the more implicit motion ratio over that region. The H.264 standard supports the **decision modes** illustrated in Table 1, where can also be seen the average frequency. This information is relevant to the segmentation algorithm.

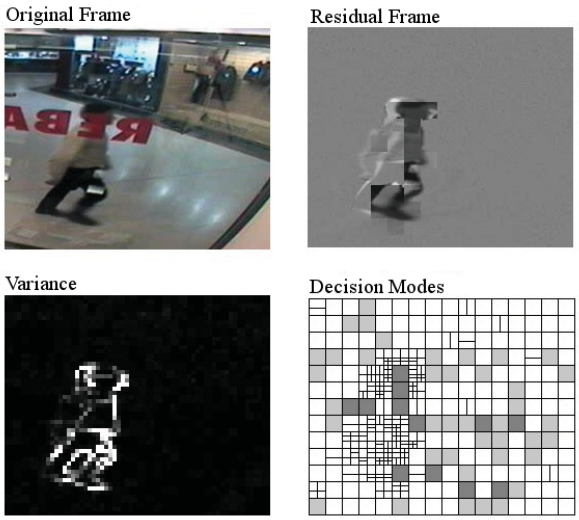


Figure 1: Processing a frame.

Table 1: Adopted nomenclature for decision modes.

Id	Macroblock type	Frequency
0	Skip, without added information	19,6 %
1	One 16x16 partition	50,6 %
2	Two 16x8 partitions	3,3 %
3	Two 8x16 partitions	2,2 %
4	Four 8x8 partitions	10,7 %
5	Two 8x4 sub-partitions in an 8x8 partition	1,9 %
6	Two 4x8 sub-partitions in an 8x8 partition	1,5 %
7	Four 4x4 sub-partitions in an 8x8 partition	0,5 %
8	Four 8x8 partitions without sub-partitions	0,4 %
9	4x4 Intra-frame macroblock	4,8 %
10	16x16 Intra-frame macroblock	0,9 %
11	Without using in an standard encoding	0,0 %
12	Without using in an standard encoding	0,0 %
13	8x8 Intra-frame macroblock	3,7 %

4 Fuzzy Linguistic Concepts

The motion vectors of an H.264 video flow are imprecise approximations to the real *Optical Flow Fields* [9]. In H.264 frames, the areas more sensitive to contain wrong motion vectors [10] are those with imperceptible variation of luminance values and those corresponding to the boundary objects. In cases like that, it is complex to the motion compensation algorithms to find correspondences between macroblocks in consecutive frames. The most compressed-domain techniques for image manipulation apply any filter to the motion vectors field to avoid the anomalous values as consequence of the mistakes introduced by the encoders. The use of the filters implies the growth of the information used in the algorithm and this kind of operations use to be expensive from a computational point of view; thus, these procedures are inappropriate for real-time algorithms.

ISBN: 978-989-95079-6-8

The related uncertainty justifies the use of the fuzzy logic in the H.264 compressed domain. The algorithm uses approximate reasoning to compact the information related to the video flow, so values not much affected by the encoding error will have similar fuzzy values. Besides, given that the goal of this work is to obtain the linguistic description of the moving regions present in a video scene, it is necessary to define the set of linguistic elements used in the description of the regions.

The proposed algorithm uses four **linguistic variables** [11] to represent the horizontal position (HP), vertical position (VP), horizontal velocity (HV) and vertical velocity (VV). One of the critical points of the system is the definition of the membership function for each variable because it depends on the scenario under observation. For example, Fig. 2 illustrates the linguistic labels belonging to the linguistic variable HP used as initial values in this study. These values can be adapted to obtain the optimal design for the specific scenario.

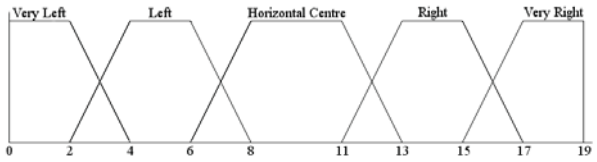


Figure 2: *Horizontal Position* linguistic labels.

A *linguistic motion vector* is the fuzzy representation of a motion vector belonging to a macroblock. It represents a linguistic description of the motion of a macroblock between consecutive frames and is defined as:

$$LMV =< FN, I_{HP}(mb\_row), I_{VP}(mb\_col), \\ , I_{HV}(right\_x), I_{VV}(down\_x) > \quad (1)$$

where the first element identifies the frame number where the motion vector is and the next four elements are four linguistic intervals. A **linguistic interval** is an ordered set of consecutive pairs of linguistic labels ($\{Left : 0,5 ; Horizontal\ Centre : 0,5\}$). The I_{PH} and I_{PV} intervals represent the column and row of the macroblock. The I_{HV} e I_{VV} intervals are obtained from the *right_x* and *down_x* values of the motion vector.

A **valid linguistic motion vector** is a linguistic motion vector that contains relevant information about the direction and velocity of an object, i.e., at least one of the two magnitudes concerning the velocity (I_{HV} and I_{VV}) of the vector is distinct of the label *No Motion*.

A *linguistic blob* is an 7-tuple composed by one or more conceptually similar motion vectors that represents in a linguistic way a region (*blob*) in a frame:

$$LB =< FN, Size, MBs, I_{HP}, I_{VP}, I_{HV}, I_{VV} > \quad (2)$$

where FN is the frame number in which is located the blob, $Size$ is the number of linguistic motion vectors grouped in the blob, MBs is a list of macroblocks belongs to the blob and the last four elements are the linguistic intervals that represent the position and velocity of the blob.

5 Segmentation Algorithm

In this work, the real displacement between frames is computed by using a fuzzification procedure working on motion vectors, neither DCT coefficients information nor statistical filtering is needed; so the algorithm requires very little of data; from a frame size of 320x240 pixels and about 35-40 KB, the algorithm only needs about 2-4 KB and around 18 KB per frame for a 640x480 video sequence with a frame size of 120-140 KB.

Fig. 3 shows the overview of the proposed system. The H.264 compressed video streams are the input of the algorithm. The first stage is the decoding of the video streaming with the goal of extracting the motion vectors and the decision modes. In the next stage, a k-neighbor algorithm is applied to the mode decision matrix to obtain the motion vectors to take into account. After that, in the fuzzification stage, motion vectors are converted into linguistic motion vectors in order to obtain the fuzzy representation of them. Then the noisy information is removed and the valid linguistic motion vectors are obtained. The algorithm rules out the motion vectors with low values of *right_x* and *down_x* because they could be noisy and they do not provide information about moving objects. Next stage is the segmentation, i.e., the valid vectors are grouped into linguistic blobs, each of them could be identified as a moving object of the video scene. Finally, the linguistic blobs are filtered to delete noisy ones. Note that the result of the algorithm is not a set of objects detected in a video sequence, but a set of linguistic blobs or regions which form the previous step to the identification of the objects. In a parallel process to the segmentation, the fuzzy sets defined over each linguistic variable are modified in a dynamic way to adapt the algorithm to a specific scenario, as it is explained in Section 5.3. Thus, the linguistic labels are modified in accordance with the size of the blobs identified in the scene and this fact has influence on the performance of the algorithm.

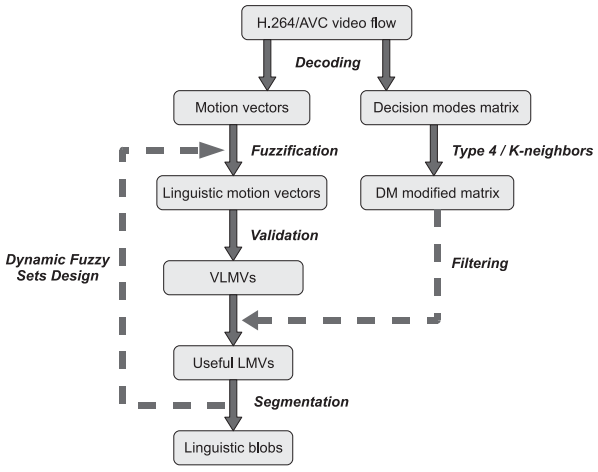


Figure 3: Overview of the proposed system.

5.1 Decision Modes Selection

The segmentation algorithm has to choose the valid motion vectors to obtain the right blobs. It can be selected between two ways of decision modes selection. In the first one, the algorithm takes into account only motion vectors belonging to a **macroblock of type 4** (four 8x8 partitions) because it has

ISBN: 978-989-95079-6-8

been observed experimentally that the moving objects are usually coded with this decision mode (Fig. 4(a)). In the second one, it is obtained the decision modes matrix corresponding to a frame and it is applied a **k-neighbor algorithm** to obtain the implicit motion degree in a region of the frame. The decision modes has been named in increasing motion degree, thus the higher is the k-neighbor macroblock value, the more is the movement inside the macroblock (Fig. 4(b)). In this case, the segmentation procedure works only with motion vector values greater than a predefined threshold U_{kn} in the k-neighbor algorithm. In Section 6 the results using both decision modes selection are analyzed.

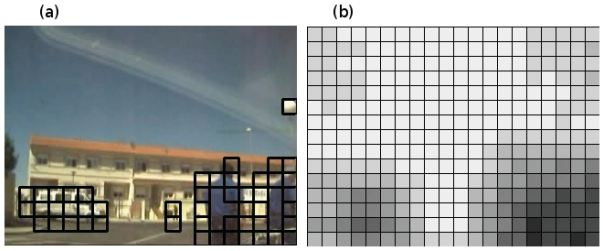


Figure 4: Decision modes selection: (a) type 4 macroblocks in a frame, (b) decision modes matrix with k-neighbor algorithm.

5.2 Total distance and weighted aggregation

After fuzzification of motion vectors, the proposed system clasifies the valid linguistic motion vectors into linguistic blobs through a *clustering* algorithm. If there is a linguistic motion vector suitable to belong to a blob, it is included into the linguistic blob and is modified the representation of this one according to a **weighted aggregation**. The similarity between a vector and a blob is calculated with a distance measure based in the numbering order of the defined linguistic labels over a variable. The distance between two linguistic intervals, the **total distance** between a linguistic motion vector and a blob and the weighted aggregation are explained deeply in [1, 2].

5.3 Dynamic Design

The segmentation algorithm proposed presents the typical problems of any segmentation system: **merge** (two or more objects are identified as one; it could happen with little objects) and **split** (one object is identified as two or more; it could happen with big objects). One of the reasons of these problems come from the static design of the linguistic variables used in the algorithm, i.e., their values remain unaltered for different kinds of objects. Since the definition of the total distance and the weighted aggregation, this segmentation system has a rigid behaviour and cannot adapt itself to the nature of the objects. Besides, the initial version of the algorithm is strongly dependent of the expert knowledge because the fuzzy sets defined over each linguistic variable are manually established and their values are invariables. Since the application is designed for real traffic scenarios, the linguistic variables are initially configured for the most frequently appeared object in that domain, i.e., the car. However, there could be a merge problem if an object little than a car (for example, a person) appears and there could be an split problem if a big object like a truck appears on the video. Then, if the number and the values of the fuzzy sets are selected in a wrong way, the result of the algorithm will be poor.

For this reason, this algorithm introduces the dynamic design of the linguistic labels to adapt them to each specific scenario, where the *dynamic design of a linguistic variable* X is understood as the adaptation and modification of the values of the fuzzy sets belonging to X during the execution of a fuzzy procedure. Thus, the segmentation procedure allows to give itself some feedback to improve its results. The algorithm has an initial configuration, but it could change the values of the fuzzy sets in function of the obtained results taking into consideration the size and shape of the obtained blobs in previous frames. So the automatic changes allow to improve the quality of the algorithm because they have positive influence on the merge and split rates, as shown in Fig 5.

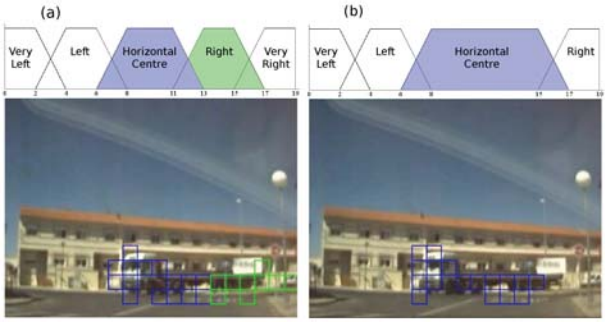


Figure 5: Dynamic fuzzy sets avoid the merge problem.

The dynamic design of the fuzzy sets is based on some common characteristics of the fuzzy systems like the **support set**, the **height** of a fuzzy set and the **alpha cut threshold** [12]. Being X a linguistic variable, A a fuzzy set of X, D the domain of X and α the threshold of the alpha cut set:

$$support(A) = \{x \in D : \mu_A(x) > 0\} \quad (3)$$

$$height(A) = max(\mu_A(x)) \quad (4)$$

$$alpha\_cut_\alpha(A) = \{x \in D : \mu_A(x) \geq \alpha\} \quad (5)$$

The linguistic variables whose values are updated dynamically are HP and VP. Each variable X has a set of fuzzy sets $\{A_1, A_2, ..., A_{size}\}$ over his domain D_{HP} or D_{VP}. Only normal fuzzy sets are used, so the height of them is always equal to one. The algorithm is ready to work with both triangular and trapezoidal fuzzy sets. The support of each fuzzy set A is $support(A) = alpha\_cut_1(A) + 4$, except for the first and the last sets (A_1 and A_{size}) of a linguistic variable, where $support(A) = alpha\_cut_1(A) + 2$.

The dynamic design of the fuzzy sets is executed at the end of the segmentation algorithm, as shown in Fig. 3. It is only processed to the biggest blob detected in the frame, if there is at least one blob detected in said frame. Being the macroblocks $MBs = \{MB_1, MB_2, ..., MB_m\}$ belonging to the biggest blob LB, it can be calculated the **rectangular size** of the blob by:

$$height(LB) = row(MB_m) - row(MB_1) + 1 \quad (6)$$

ISBN: 978-989-95079-6-8

$$width(LB) = max_i(col(MB_i)) - min_i(col(MB_i)) + 1 \quad (7)$$

where col is the column of the macroblock, max is he maximum of the values and min is the minimum value.

The system calculates the new **number** of the new fuzzy sets defined to the linguistic variables HP and VP in function of the size of the biggest blob in the frame. Being W and H the width and the height in macroblocks of each frame, the number of the redisegned linguistic labels can be calculated as $size(HP) = W/width(LB)$ and $size(VP) = H/height(LB)$. The **names** of the linguistic labels of HP and VP are got from two configuration files that have a correspondence between the number of labels and the names of them. After that, if trapezoidal fuzzy sets are used, it is necessary to calculate the support and the alpha-level set with threshold 1 to define the **shape** of the corresponding fuzzy sets. Since $size(HP)$ and $size(VP)$, it can be calculated the size of the alpha-level set with threshold value 1 with the same size to all the fuzzy sets for each linguistic variable. Being A_{HP} a fuzzy set of HP and A_{VP} a fuzzy set of VP, the size of the alpha cut set is calculated as:

$$size(alpha\_cut_1(A_{HP})) = \frac{W - 1 - 2 * (size(HP) - 1)}{size(HP)} \quad (8)$$

$$size(alpha\_cut_1(A_{VP})) = \frac{H - 1 - 2 * (size(VP) - 1)}{size(VP)} \quad (9)$$

For example, being the fuzzy sets of the linguistic variable HP shown in Fig. 2, if the algorithm finds a full region in a frame which a width of 6 macroblocks ($width(LB) = 6$), the module which updates the values of the linguistic variable HP will change its values. First, it calculates the number of the labels as $size(HP) = W/width(LB) = 20/7 \backsimeq 3$ and reads the names of them in the configuration file (LE, HC and RI). Then, the size of the alpha cut set and the support of each set are calculated as:

$$size(alpha\_cut_1(A_{HP})) = \tfrac{W-1-2*(size(HP)-1)}{size(HP)} =$$

$$= \frac{20 - 1 - 2 * (3 - 1)}{3} = 5 \quad (10)$$

$$support(A_1) = alpha\_cut_1(A) + 2 = 5 + 2 = 7 \quad (11)$$

$$support(A_2) = alpha\_cut_1(A) + 4 = 5 + 4 = 9 \quad (12)$$

The module can redefine the values of the linguistic variable HP with this information and obtains the new values, which are shown in Fig. 6.

Figure 6: New *Horizontal Position* linguistic labels.

6 Experimental Results

In this work a supervised method to evaluate the segmentation quality is used. The segmentation approach has been evaluated on several video sequences compressed using the H.264 encoder of JM 14.0. The encoder configuration has been set with *baseline profile* and low complexity mode on RD Optimization. The video tested resolution varies from 320x240 to 640x480 pixels. The specific configuration parameters of the algorithm (U_{mv}, U_{kn} and the minimum blob size) change depending on the application scenario. Fig. 7 shows three snapshots of three different video sequences tested where the detected blobs are drawn over the frames.

As it has been explained in Section 5.1, the algorithm can be executed in two modes depending of the decision modes selection: only *type 4* macroblocks or *k-neighbor algorithm* with a threshold. Table 2 shows the differences between both modes in several experiments (320x240 pixel resolution) using the same configuration parameters. It can be seen that one of them is faster but its performance is worse. Thus the choice will depend on the application field.

Table 2: Decision modes selection.

Measurement	**Type 4**	**K-neighbor**
Detection possibility	73 %	81 %
Precision	78 %	89 %
Processing time	16,2 ms.	18,6 ms.
Segmentation size	148 bits/frame	103 bits/frame

The hardware used in the experiments is an Intel Pentium IV Core 2 Duo T7100 1.8 GHz with 2 GB RAM memory. Table 3 shows average processing time depending on the frame resolution. Since a 320x240 frame is processed in around 20 ms. and a 640x480 one in around 38 ms., it can be ensured that the segmentation works in real-time, i.e., at least 25 fps.

Table 3: Average processing time.

Frame resolution (pixels)	**Processing time**
320x240	20 ms.
640x240	26 ms.
640x480	38 ms.

Table 4 shows the improvement of the results using dynamic design of the fuzzy sets. Two measurements are used to describe the segmentation performance. The first measurement is *detection possibility* that is defined as the percentage of the real objects detected. The second one is *precision* that measures the percentage of the detected objects corresponding to real objects. Then, high detection possibility means less miss-segmentation and high precision means less false segmentation. Besides, two measurements are defined to analyze the *processing time* and the *segmentation size*, i.e., the temporal and spatial requirements of the algorithm.

7 Conclusions

This work presents a real-time segmentation approach of moving objects with the next features. First of all, the algorithm is designed for H.264/AVC compressed domain due to its excellent compression efficiency and its widespread field of multimedia applications. The algorithm can work in real-time and

ISBN: 978-989-95079-6-8

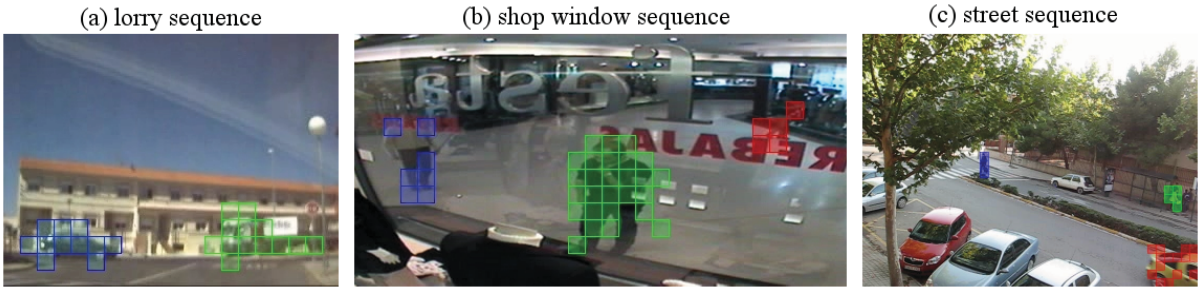


Figure 7: Snapshots of three video sequences with the detected blobs drawn.

Table 4: Improvement using dynamic design.

Measurement	Lorry sequence		Shop window sequence		Street sequence	
	Static	Dynamic	Static	Dynamic	Static	Dynamic
Resolution	320x240		640x240		640x480	
Ground truth	353	353	808	808	1355	1355
Detected blobs	378	362	649	684	1156	1138
True positives	293	302	631	665	996	1077
False positives	85	60	18	19	160	61
False negatives	60	51	177	143	359	278
Detection possibility	82 %	85 %	78 %	82 %	74 %	80 %
Precision	78 %	83 %	97 %	97 %	86 %	95 %
Processing time	17,4 ms.	19,7 ms.	25,5 ms.	32,7 ms.	37,2 ms.	41,1 ms.
Segmentation size	184 KB.	173 KB.	101 KB.	110 KB.	173 KB.	164 KB.

requires very little of data because it uses only the motion vectors field and the decision modes to carry out the segmentation. The proposed approach is based on fuzzy logic to detect the moving objects; fuzzy logic allows to avoid the noise inherent to the encoding process and to obtain conceptual representations that describe the regions detected in a comprehensive way. Finally, the segmentation performance is robust because it can adapt itself to the application scenario in a dynamic way updating in real-time the values of the fuzzy sets. This feature allows to improve the merge and split rates.

Acknowledgment

This work was supported by the Ministry of Science and Technology of Spain under CONSOLIDER Project CSD2006-46 and CICYT Project TIN2006-15516-C04-02 and by the Council of Science and Technology of Castilla-La Mancha under Projects PAI06-0106, PII2I09-0052-3440 and PII1C09-0137-6488.

References

[1] L. Rodriguez-Benitez, J. Moreno, J.J. Castro-Schez, J. Albusac, and L. Jimenez. An approximate reasoning technique for segmentation on compressed MPEG video. In *2nd International Conference on Computer Vision Theory and Applications, VISAPP'07*, pages 184–191, 2007.

[2] L. Rodriguez-Benitez, J. Moreno, J.J. Castro-Schez, J. Albusac, and L. Jimenez. Automatic objects behaviour recognition from compressed video domain. *Image and Vision Computing*, 2008. In Press, Accepted Manuscript, Available online 15 July 2008.

[3] P. Matsakis, J.M. Keller, L. Wendling, J. Marjamaa, and O.Sajhputera. Linguistic description of relative positions in images. *IEEE Transactions on systems, man and cybernetics*, 31(4):573–588, 2001.

[4] O. Sajhputera, P. Matsakis, J.M. Keller, and R. Bondougula. Linguistic descriptions for an object in motion. *Proceedings of the NAFIPS, New Orleans, Louisiana, USA*, pages 243–248, 2002.

[5] W. Zeng, J. Du, W. Gao, and Q. Huang. Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model. *Real-Time Imaging*, (11):290–299, 2005.

[6] Z. Liu, Y. Luy, and Z. Zhang. Real-time spatiotemporal segmentation of video objects in the H.264 compressed domain. *Journal of Visual Communication and Image Representation*, (18):275–290, 2007.

[7] ITU-T Recommendation H.264. Advanced video coding for generic audiovisual services, May 2003.

[8] I. E. G. Richardson. *H.264 and MPEG-4 Video Compression*. John Wiley & Sons Ltd, New Jersey, 2003.

[9] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[10] T. Yoshida, H. Katoh, and Y. Sakai. Block matching motion estimation using block integration based on reliability metric. *Proceedings of the International Conference on Image Processing (ICIP '97)*, 2:152–155, 1997.

[11] L.A. Zadeh. The concept of linguistic variable and its application to approximate reasoning. *Information Sciences*, 8:199–249, 1975.

[12] E. Cox. *The fuzzy systems handbook*. AP Professional, 2nd edition, 1998.

ISBN: 978-989-95079-6-8

On the Benefits of Representing Music Objects as Fuzzy Sets

Klaas Bosteels[1,2] Elias Pampalk[2] Etienne E. Kerre[1]

1. Fuzziness and Uncertainty Modelling Research Group
Dept. of Applied Mathematics and Computer Science, Ghent University, Ghent, Belgium
2. Last.fm Ltd., London, UK
Email: klaas.bosteels@ugent.be, elias@last.fm, etienne.kerre@ugent.be

***Abstract*— *Over the past few years, fuzzy set theory has become a well-established mathematical theory that has been applied in many practical applications. In this paper, we elaborate on the use of fuzzy sets in the context of music information retrieval. First, we demonstrate that music objects like artists and songs can straightforwardly be represented as fuzzy sets, and then we explain why it is beneficial to represent them in this way. In particular, we illustrate the potential of this fuzzy approach by describing how we relied on it to build the popular "Multi Tag Search" demonstration on* `http://playground.last.fm`.

***Keywords*—** Music information retrieval, fuzzy sets, fuzzy comparison measures.

1 Introduction

The way in which the objects are represented is an important aspect of most music information retrieval applications. Simple vectors are used in many cases, but certain more advanced representations like statistical distributions are popular too. In this paper, we advocate for the usage of fuzzy sets for this purpose. As we will explain in detail, representing a music object as a fuzzy set in some universe U, i.e., as a $U \to [0, 1]$ mapping that associates a degree of membership with each u from U, is a natural and mathematically sound approach that offers many possibilities and advantages. To enforce our claims, we will describe how we relied on fuzzy set theory for building the "Multi Tag Search" demonstration on `http://playground.last.fm`, the section of the CBS-owned music community website Last.fm[1] where new ideas and experimental technologies are showcased.

2 Representing music objects as fuzzy sets

There are three main ways to obtain descriptions for music objects:

1. Hiring music experts to manually annotate the objects.
2. Assembling a large community of music enthusiasts and then tapping into the wisdom of crowds.
3. Applying signal processing techniques to generate descriptions automatically.

Hence, song descriptions can come from experts, communities, or computers. In each case, it is often possible to convert them to fuzzy sets.

2.1 From experts

When music experts are hired to generate descriptions, they are usually asked to provide values for a predefined list of attributes. For instance, an expert could be used to assign values to the attributes "genre" and "complexity" for a particular song. Since the latter attribute is numerical, it can straightforwardly be converted to a gradual element of a fuzzy set. For example, if a song is 80% complex, then we would give "complexity" a membership degree of 0.8 in the fuzzy set that corresponds with this song. In case of a categorical attribute like "genre", however, we cannot interpret its values as membership degrees. Nevertheless, it is still possible to convert such attributes, namely, by regarding the values as binary attributes. For example, we can express that a song is an alternative rock song by assigning the membership degree 1 to "alternative rock" and setting the membership degrees of all other possible genres to 0. In fact, it might even be more appropriate to directly present the different genre values as separate numerical attributes to the expert, since a song often belongs to several genres to some extent. The expert can then express that, e.g., a particular song is mainly alternative rock, but also incorporates some notable blues and jazz influences. The well-known online radio station Pandora[2] uses 150-500 so called "genes" to describe songs [1, 2], where each gene is a numerical attribute that takes a value from the interval $[0, 5]$. Hence, the descriptions generated by Pandora's music experts could easily be interpreted as fuzzy sets.

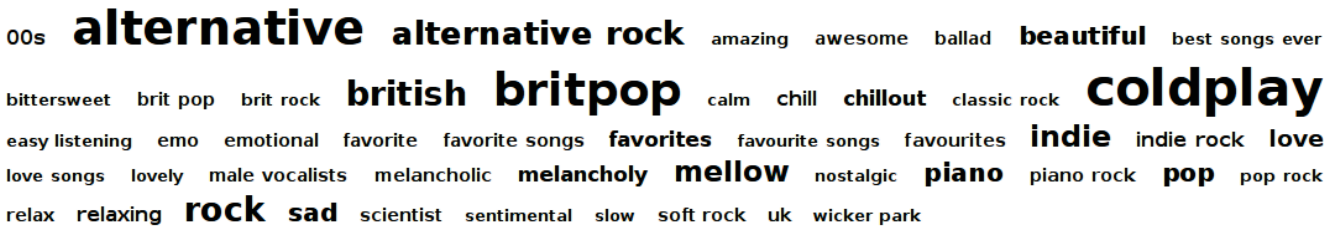
2.2 From communities

The data that powers Last.fm's recommendation engine is mainly generated by a technique called "scrobbling". When a user listens to a song, the name of this song is sent to Last.fm and added to his or her profile. By counting how many times each user listens to each song, implicit user ratings can be obtained from these "scrobbles". Since such ratings can easily be interpreted as membership degrees, they can be used to associate a fuzzy set of users with each song. For instance, the users that listened the most to the song in question can be given membership degree 1, and the membership degrees of the remaining users can be determined by dividing their play counts by the greatest play count, e.g., if there are only three users and their play counts for a particular song are 10, 5, and 1, then the membership degrees of these users in the fuzzy set for this song would be 1, 0.5, and 0.1, respectively. Comparing such fuzzy sets by means of the cosine similarity measure

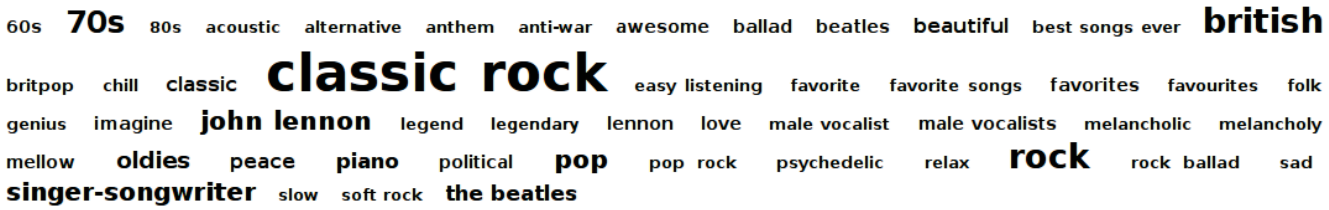
[1] http://www.last.fm

[2] http://www.pandora.com

ISBN: 978-989-95079-6-8



(a) Tag cloud for "The Scientist" by "Coldplay".



(b) Tag cloud for "Imagine" by "John Lennon".

Figure 1: Two examples of tag clouds.

is equivalent to the basic "item-to-item collaborative filtering" recommendation algorithm [3].

Last.fm also uses social tagging [4] as an additional community-driven way to generate descriptions. Social tags are individual keywords or phrases that people associated with objects, usually for organizational purposes. When the tags applied by a large community are combined, rich and complex descriptions emerge. A popular way to visualize such a description is by means of a tag cloud, i.e., an alphabetically ordered list of tags in which the font size is proportional to the number of people that associated the tag in question with the object. Fig. 1 shows two examples. By interpreting the font sizes in a tag cloud as membership degrees, we obtain a fuzzy set of tags that describes the corresponding object.

2.3 From computers

Using algorithms that analyze audio signals, song descriptions can be obtained in a fully automatic way [5]. However, these descriptions are usually very low-level and difficult to interpret. A possible solution for this problem is to teach computers how to add high-level annotations to songs by feeding the automatically generated low-level descriptions into machine learning algorithms that were trained on descriptions from experts or communities [6, 7], or by developing algorithms that propagate annotations by analysing the low-level descriptions [8]. Although it is certainly not impossible to convert the low-level descriptions to fuzzy sets [9], it makes more sense to apply such a conversion to the high-level ones obtained by incorporating machine learning techniques. Since these automatically generated high-level descriptions are intended to resemble the descriptions generated by humans, they can be converted to fuzzy sets as explained above.

3 The Benefits

In the 40 years since Zadeh introduced the concept of a fuzzy set [10], tens of thousands of papers and plenty of books have been written on this seminal topic, and more than 20 international journals were established in the corresponding field. The general advantage of representing songs as fuzzy sets is that we can rely on this vast amount of research for dealing with such representations. In this section, we describe a few more concrete benefits that ensue from this general one.

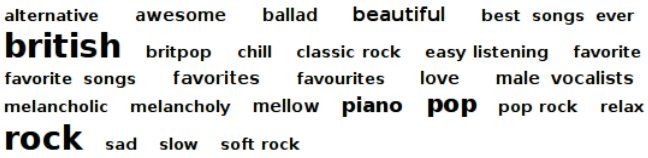


Figure 2: Intersection of the tag clouds shown in Fig. 1.

3.1 Fuzzy aggregation operators

Fuzzy aggregation operators are mathematical objects that aggregate multiple membership degrees into one. Formally, a fuzzy aggregation operator $\mathcal{H}$ of arity n, with $n \in \mathbb{N} \setminus \{0\}$, is an increasing $[0,1]^n \to [0,1]$ mapping that satisfies $\mathcal{H}(0,0,\ldots,0) = 0$ and $\mathcal{H}(1,1,\ldots,1) = 1$. Triangular norms [11], Quasi-arithmetic means [12], Sugeno and Choquet integrals [13, 14], and ordered weighted averaging (OWA) operators [15] are just a few examples of fuzzy aggregation operators that could be valuable for combining fuzzy descriptions of music objects. We refer to [16] for an introductionary overview of the extensively studied field of fuzzy aggregation operators. Fig. 2 provides a very simple example. It shows the result of applying the minimum operator to compute the intersection of the tag clouds in Fig. 1, which could be used as an intuitive visualization of the similarities between the songs that correspond with these tag clouds.

3.2 Fuzzy comparison measures

Expressions like, e.g.,

$$\frac{|A \cap B|}{|A \cup B|}, \qquad \frac{2\,|A \cap B|}{|A| + |B|}, \qquad \text{and} \qquad \frac{|A \cap B|}{\sqrt{|A|\,|B|}}$$

are commonly used to determine the similarity between two (ordinary) sets A and B. By defining the cardinality of a fuzzy set A in U as $|A| = \sum_{u \in U} A(u)$ and generalizing the classical set-theoretic operations to fuzzy sets, such expressions can

ISBN: 978-989-95079-6-8

be extended to similarity measures for fuzzy sets [17]. More generally, a measure that compares crisp sets can often be extended to fuzzy sets. We call such a measure a fuzzy comparison measure. A wide plethora of fuzzy comparison measures can be found in the literature, and the properties of many of these measures have been studied extensively. For instance, the measures M for which $1 - M$ satisfies the triangle inequality have been characterized for a broad family of fuzzy similarity measures [18], which can be very useful in practice since the triangle inequality allows to reduce the required number of comparisons [19]. Another family of fuzzy comparison measures that has practical value is the one developed in [20, 21, 22], since it can be used to systematically construct measures that satisfy certain properties. We rely on this family in the practical example described in Section 4.

3.3 Fuzzy relations

A (binary) fuzzy relation on U is a $U \times U \rightarrow [0, 1]$ mapping that associates a membership degree with each pair of elements from U, i.e., a fuzzy set in $U \times U$. Since fuzzy comparison measures are usually fuzzy relations on the (ordinary) set of all possible fuzzy sets, the approach described in the previous subsection has the additional advantage that the obtained similarities are also membership degrees, which implies that fuzzy aggregation operators can be used to combine them. For example, a fuzzy relation that models the similarities between the songs in the considered collection is a prerequisite of the general framework for defining dynamic playlist generation heuristics introduced in [23], because it relies on fuzzy set theory to make the definitions systematic, formal, and intuitively clear. Since music similarity measures based on a fuzzy similarity measure are fuzzy relations already, we do not have to apply any normalization procedures on them in order to be able to use this framework.

4 Practical example

In this section, we describe a practical example that illustrates the potential of representing songs as fuzzy sets. More precisely, we explain how we relied on fuzzy set theory to build the "Multi Tag Search", a web application that allows users to find music by supplying one or more tags. Fig. 3 shows the first results returned by this application when popular artists that match with the tags "british" and "rock" are requested. Any set of keywords or phrases can be used to search for either songs or artists, and the options "up-and-coming" and "free downloads" are also available, in addition to "popular".

4.1 Basic method

Putting it in one sentence, the "Multi Tag Search" ranks all relevant music objects from Last.fm's database by representing the query as a fuzzy set of tags and comparing this fuzzy set with the fuzzy tag set associated with each object. More formally, the ranking score for object i is given by:

$$score(\widetilde{Q}, \widetilde{O}_i) = w(i)\, M(\widetilde{Q}, \widetilde{O}_i) \tag{1}$$

with $\widetilde{Q}$ and $\widetilde{O}_i$ the fuzzy sets corresponding to the query and object i, respectively, M a fuzzy comparison measure, and w a function that associates a real number with each object index. The weighting function w allows us to push certain objects up

Figure 3: Screenshot of the "Multi Tag Search".

or down in the generated ranking. It is mainly used to implement the "popular", "up-and-coming", and "free downloads" search options, but we also utilize it for some small optimizations like penalizing objects to which only very few tags have been applied.

We used the following general definitions of the fuzzy tag sets $\widetilde{Q}$ and $\widetilde{O}_i$:

$$\widetilde{Q}(t) = \begin{cases} \dfrac{q(t)}{\max_{u \in T}\, q(u)} & \text{when } t \in Q \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$\widetilde{O}_i(t) = \begin{cases} \dfrac{o_i(t)}{\max_{u \in T, j \in I}\, o_j(u)} & \text{when } t \in O_i \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

for all $t \in T$ and $i \in I$, with T the universe of all possible tags, I the set of all object indexes, Q the set of query tags, O_i the set consisting of all tags that have been applied to object i, and q and o_i, for each $i \in I$, a $T \rightarrow \mathbb{R}$ mapping. For the values of the parameters q and o_i in these definitions, we drew some inspiration from text retrieval, namely, we put $q = idf$ and $o_i = ntf_i$, with idf the so called "inverse document frequency", and ntf_i the "normalized term frequency" [24]:

$$idf(t) = \log\left(\frac{\text{total number of objects}}{\text{number of objects tagged with } t}\right) \tag{4}$$

$$ntf_i(t) = \frac{\text{number of taggings with } t \text{ for object } i}{\text{total number of taggings for object } i} \tag{5}$$

for each $t \in T$ and $i \in I$. Hence, tags that are not frequently used get a high membership degree in $\widetilde{Q}$, and the membership degree of the tags in $\widetilde{O}_i$ is proportional to the number of times they have been applied to object i. Many other term-weighting schemes can be found in the text retrieval literature [25], but using the inverse document frequency for the query and the normalized term frequency for the objects appeared to work very well for Last.fm's tag data.

ISBN: 978-989-95079-6-8

4.2 Constructing a suitable comparison measure

High-quality tag data is by far the most important requirement for a good tag-based search engine. Improving the quality of the underlying data is generally more worthwhile than trying to find a better comparison measure, and which comparison measure is used exacly becomes less important when the tag data is of high quality. "Worry about the data first before you worry about the algorithm," as Google's director of research Peter Norvig puts it [26]. Nevertheless, it is of course important to use a suitable comparison measure that has the right properties, but it makes sense to choose the simplest possible measure that satisfies these properties, especially when you have to deal with millions of objects and users.

For the "Multi Tag Search", we wanted the fuzzy comparison measure M to satisfy the following properties:

$$M(A,B) = 0 \Leftarrow \operatorname{supp} A \cap \operatorname{supp} B = \emptyset \qquad \text{(exclusive)}$$
$$M(A,B) = 0 \Rightarrow \operatorname{supp} A \cap \operatorname{supp} B = \emptyset \qquad \text{(coexclusive)}$$
$$M(A,B) = M(A/_{\operatorname{supp} A}, B/_{\operatorname{supp} A}) \qquad \text{(left-restrictable)}$$

for all A and B from the class $\mathcal{F}(U)$ of all fuzzy sets in U, with $\operatorname{supp} A$ the support of $A \in \mathcal{F}(U)$, i.e., $\operatorname{supp} A = \{u \in U \mid A(u) > 0\}$, and $A/_V$, for $A \in \mathcal{F}(U)$ and $V \subseteq U$, the restriction of A to V, i.e., the $V \to [0,1]$ mapping that associates $A(v)$ with each $v \in V$. Putting it in words, we want the value returned by M to be 0 if and only if none of the query tags have been applied to the object in question, and we do not want the returned value to depend on the membership degrees of the tags that are not part of the query. The latter is particularly important, mainly because it is advantageous from a computational point of view, but also because it ensures that we cannot obtain different values for objects with fuzzy tags sets in which each query tag always has the same membership degree. For instance, if we would use the well-known cosine similarity measure, which is not left-restrictable, then the query $Q = \{$"british", "rock"$\}$ would lead to a different value for the fuzzy sets obtained by directly interpreting the font sizes in the tag clouds shown in Fig.1 as membership degrees, while there really is no reason to prefer one over the other in this case since the font sizes for "british" and "rock" are exactly the same in both clouds. In the final ranking, one of the corresponding songs would have to be put before the other of course, but we want that decision to only depend on the value returned by w.

The family of fuzzy comparison measures introduced in [21], which includes the popular cosine similarity measure, makes it very easy to systematically construct fuzzy comparison measures that satisfy given properties. All members of this family are instances of a triparametric general form, and several properties can be ensured by imposing constraints on the parameters $\mathcal{H}$, φ_1, and φ_2 of this general form, where $\mathcal{H}$ is a binary fuzzy aggregation operator, i.e., an increasing $[0,1]^2 \to [0,1]$ mapping such that $\mathcal{H}(0,0) = 0$ and $\mathcal{H}(1,1) = 1$, and φ_1 and φ_2 are two $[0,1]^3 \to \mathbb{R}$ mappings.[3] As value of the first parameter $\mathcal{H}$, we chose the product t-norm T_{P} given by $T_{\mathrm{P}}(x,y) = x \cdot y$, for all $x, y \in [0,1]$, because it is a very simple operator that is interactive [27], i.e., a modification of x or y always implies an alteration of $T_{\mathrm{P}}(x,y)$ if $x \neq 0$ and $y \neq 0$. The general form from [21] then reduces to:

$$M(A,B) = \frac{\varphi_1(||A \cap_{T_{\mathrm{P}}} A||, ||B \cap_{T_{\mathrm{P}}} B||, ||A \cap_{T_{\mathrm{P}}} B||)}{\varphi_2(||A \cap_{T_{\mathrm{P}}} A||, ||B \cap_{T_{\mathrm{P}}} B||, ||A \cap_{T_{\mathrm{P}}} B||)} \quad (6)$$

for all $A, B \in \mathcal{F}(U)$, with $||.||$ the $\mathcal{F}(U) \to [0,1]$ mapping given by $||A|| = |A|/|U|$ for all $A \in \mathcal{F}(U)$, and $\cap_{T_{\mathrm{P}}}$ an infix notation for the pointwise extension of T_{P}, i.e., $(A \cap_{T_{\mathrm{P}}} B)(u) = T_{\mathrm{P}}(A(u), B(u))$ for all $u \in U$, with A and B fuzzy sets in U. Using the proofs provided in [21], it can easily be verified that this $\mathcal{F}(U) \times \mathcal{F}(U) \to \mathbb{R}$ mapping is an exclusive, coexclusive, and left-restrictable fuzzy comparison measure when the formulas

$$z \leq \sqrt{x \cdot y} \implies 0 \leq \varphi_1(x,y,z) \leq \varphi_2(x,y,z)$$
$$\varphi_1(x,y,0) = 0 \qquad \varphi_1(x,u,z) = \varphi_1(x,v,z)$$
$$\varphi_1(\dot{x},\dot{y},\dot{z}) > 0 \qquad \varphi_2(x,u,z) = \varphi_2(x,v,z)$$

are all satisfied for every $x, y, z, u, v \in [0,1]$ and $\dot{x}, \dot{y}, \dot{z} \in]0,1]$. The simplest solution that we could find for this is $\varphi_1(x,y,z) = z$ and $\varphi_2(x,y,z) = 1$, for all $x, y, z \in [0,1]$, which leads to:

$$M(A,B) = \frac{||A \cap_{T_{\mathrm{P}}} B||}{1} = \frac{|A \cap_{T_{\mathrm{P}}} B|}{|U|} \quad (7)$$

for all $A, B \in \mathcal{F}(U)$, i.e., a fuzzification of the Russell and Rao coefficient [28]. We then have

$$\begin{aligned} M(\widetilde{Q}, \widetilde{O}_i) &= M(\widetilde{Q}/_{\operatorname{supp}\widetilde{Q}}, \widetilde{O}_i/_{\operatorname{supp}\widetilde{Q}}) && (M \text{ left-restr.}) \\ &= M(\widetilde{Q}/_Q, \widetilde{O}_i/_Q) && (\operatorname{supp}\widetilde{Q} = Q) \\ &= \frac{1}{|Q|} \sum_{t \in Q} \widetilde{Q}(t)\, \widetilde{O}_i(t) && (\text{definition } M) \\ &\propto \sum_{t \in Q} idf(t)\, ntf_i(t) && (\text{def. } \widetilde{Q} \text{ and } \widetilde{O}_i) \end{aligned}$$

and thus:

$$score(\widetilde{Q}, \widetilde{O}_i) \propto w(i) \sum_{t \in Q} idf(t)\, ntf_i(t) \quad (8)$$

Hence, we merely need to compute $w(i) \sum_{t \in Q} idf(t)\, ntf_i(t)$ to determine the ranking of object i for query Q.

4.3 User feedback

Last.fm's Playground was launched in May 2008, with the "Multi Tag Search" as one of the initial demonstrations. Since then, this demonstration has rather consistently been the most popular one over a period of several months. It also seemed to be more "sticky" than the other initial demonstrations, since it was the only one for which the number of page views increased after launch. Moreover, we received substantially more positive direct user feedback for it than any of the other initial demonstrations. Hence, the users seemed to like the "Multi Tag Search" quite a lot, which suggests that it must at least work reasonably well. The great variety in the supplied queries confirms this observation. We received about 25 thousand unique queries so far, and, as illustrated by Table 1, several different types of tags occur in these queries.

[3] We required φ_1 and φ_2 to be increasing in their first and second argument in [21], but since the proofs for the constraints considered in this paper do not rely on this prerequisite, we can omit it here.

ISBN: 978-989-95079-6-8

Table 1: Results of a manual classification of the 250 most frequently occurring query tags.

Type	Frequency	Examples
Genre	51%	rock, pop
Style	26%	experimental
Locale	7%	british, detroit
Instrument	5%	piano, guitar
Mood	4%	sad, romantic
Time	3%	90s, oldies
Opinion	2%	beautiful, love
Other	2%	free, cover

5 Conclusion

We have shown that fuzzy set theory has a lot to offer to the music information retrieval community. By representing music objects as fuzzy sets, the possibility arises to exploit the rich mathematical toolset provided by this theory, including fuzzy aggregation operators, fuzzy comparison measures, and fuzzy relations. As a practical example, we explained how we employed fuzzy set theory to build the popular "Multi Tag Search" demonstration on Last.fm's Playground. Using previous work on fuzzy comparison measures, we were able to systematically construct a suitable comparison measure for this tag-based search engine, instead of merely relying on intuition and trial-and-error. Rather than arbitrarily choosing one of the many comparison measures available in the literature and checking if it works as expected, we started from the desired properties and systematically constructed a simple and computationally efficient measure that satisfies these properties. Since different applications have different requirements, such a systematic way of constructing comparison measures can be very valuable for building applications in which objects need to be compared, as is commonly the case in music information retrieval as well as in certain other domains.

Acknowledgement

The first author thanks the Fund for Scientific Research–Flanders (FWO) for funding this research.

References

[1] W. Glaser, T. Westergren, J. Stearns, and J. Kraft. Consumer item matching method and system. US Patent Number 7003515, 2006.

[2] D. Turnbull, L. Barrington, and G. Lanckriet. Five approaches to collecting tags for music. In *Proceedings of the ISMIR International Conference on Music Information Retrieval*, 2008.

[3] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7:76–80, 2003.

[4] P. Lamere. Social tagging and music information retrieval. *Journal of New Music Research*. To appear.

[5] E. Pampalk. *Computational models of music similarity and their application in music information retrieval*. PhD thesis, Vienna University of Technology, 2006.

[6] D. Eck, P. Lamere, T. Bertin-Mahieux, and S. Green. Automatic generation of social tags for music recommendation. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in neural information processing systems 20*. MIT Press, 2008.

[7] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech, and Language Processing*, 16:467–476, 2008.

[8] M. Sordo, C. Lauier, and O. Celma. Annotating music collections: How content-based similarity helps to propagate labels. In *Proceedings of the ISMIR International Conference of Music Information Retrieval*, 2007.

[9] K. Bosteels and E. Kerre. Fuzzy Audio Similarity Measures Based on Spectrum Histograms and Fluctuation Patterns. In A.-E. Hassanien, A. Abraham, and J. Kacprzyk, editors, *Computational Intelligence in Multimedia Processing: Recent Advances*. Springer, 2008.

[10] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[11] E. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Publishers, 2000.

[12] J. Aczel. On mean values. *Bulletin of the American Mathematical Society*, 54:392–400, 1948.

[13] M. Sugeno. *Theory of fuzzy integrals and its application*. PhD thesis, Tokyo Institute of Technology, 1974.

[14] G. Choquet. Theory of capacities. *Annales de L'Institut Fourier*, 5:131–295, 1953.

[15] R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18:183–190, 1988.

[16] M. Detyniecki. Numerical aggregation operators: State of the art. In *International Summer School on Aggregation Operators and their Applications*, 2001.

[17] B. De Baets and H. De Meyer. Transitivity-preserving fuzzification schemes for cardinality-based similarity measures. *European Journal of Operational Research*, 160(3):726–740, 2005.

[18] S. Janssens. *Bell inequalities in cardinality-based similarity measurement*. PhD thesis, Ghent University, 2006.

[19] J. Barros, J. French, W. Martin, P. Kelly, and T. Cannon. Using the triangle inequality to reduce the number of comparisons required for similarity-based retrieval. In *Storage and Retrieval for Image and Video Databases*, 1996.

[20] K. Bosteels and E. Kerre. A biparametric family of cardinality-based fuzzy similarity measures. *New Mathematics and Natural Computation*, 3:307–319, 2007.

[21] K. Bosteels and E. Kerre. A triparametric family of cardinality-based fuzzy similarity measures. *Fuzzy Sets and Systems*, 158:2466–2479, 2007.

[22] K. Bosteels and E. Kerre. On a reflexivity-preserving family of cardinality-based fuzzy comparison measures. *Information Sciences*. To appear.

[23] K. Bosteels and E. Kerre. Fuzzy heuristics for dynamic playlist generation. Submitted.

[24] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24:513–525, 1988.

[25] H. Blanken, A. de Vries, H. Blok, and L. Feng, editors. *Multimedia retrieval*. Springer, 2007.

[26] P. Norvig. Theorizing from data: Avoiding the capital mistake. CITRIS Distinguished Speaker Series, UC Berkeley, 2006.

[27] D. Dubois and H. Prade. *Fuzzy sets and systems: Theory and applications*. Academic Press, 1980.

[28] B. Sarker. The resemblance coefficients in group technology: A survey and comparative study of relational metrics. *Computers and Industrial Engineering*, 30:103–116, 1996.

ISBN: 978-989-95079-6-8

Type–II Fuzzy Possibilistic C-Mean Clustering

M.H. Fazel Zarandi[1], M. Zarinbal[1], I.B. Turksen[2,3]

1 Department of Industrial Engineering, Amirkabir University of Technology, P.O. Box 15875-4413, Tehran, Iran
2 Department of Industrial Engineering, TOBB Economy and Technology University, Sogutozo Cad. No:43, Sogutozo, Ankara 06560, Turkey
3 Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ont., Canada M5S 3G8
Email: zarandi@aut.ac.ir, mzarinbal@aut.ac.ir, turksen@mie.utoronto.ca

Abstract—*Fuzzy clustering is well known as a robust and efficient way to reduce computation cost to obtain the better results. In the literature, many robust fuzzy clustering models have been presented such as Fuzzy C-Mean (FCM) and Possibilistic C-Mean (PCM), where these methods are Type-I Fuzzy clustering. Type-II Fuzzy sets, on the other hand, can provide better performance than Type-I Fuzzy sets, especially when many uncertainties are presented in real data. The focus of this paper is to design a new Type-II Fuzzy clustering method based on Krishnapuram and Keller PCM. The proposed method is capable to cluster Type-II fuzzy data and can obtain the better number of clusters (c) and degree of fuzziness (m) by using Type-II Kwon validity index. In the proposed method, two kind of distance measurements, Euclidean and Mahalanobis are examined. The results show that the proposed model, which uses Mahalanobis distance based on Gustafson and Kessel approach is more accurate and can efficiently handle uncertainties.*

Keywords— *Type-II Fuzzy Logic; Possibilistic C-Mean (PCM); Mahalanobis Distance; Cluster Validity Index;*

1. Introduction

Clustering is an important method in data mining, decision-making, image segmentation, pattern classification, and etc. Fuzzy clustering can obtain not only the belonging status of objects but also how much the objects belong to the clusters. In the last 30 years, many fuzzy clustering models for crisp data have been presented such as Fuzzy K-Means and Fuzzy C-Mean (FCM) [1]. FCM is a popular clustering method, but its memberships do not always correspond well to the degrees of belonging and it may be inaccurate in a noisy environment [2]. To relieve these weaknesses, Krishnapuram and Keller presented a Possibilistic C-Mean (PCM) approach [3]. In addition, in real data, there exist many uncertainties and vaguenesses, which Type-I Fuzzy sets are not able to directly model as their membership functions are crisp. On the other hand, as Type-II membership functions are fuzzy, they are able to model uncertainties more appropriately [2, 4]. Therefore, Type-II Fuzzy Logic Systems have the potential to provide better performance than Type-I [5].

In the case of combining Type-II Fuzzy logic with clustering methods, the data can be clustered more appropriately and more accurately. The focus of this paper is to design a new Type-II Fuzzy clustering method based on Krishnapuram and Keller PCM. The proposed method is capable to cluster Type-II fuzzy data and can obtain the better number of clusters (c) and degree of fuzziness (m) by using Type-II Kwon validity index.

The rest of this paper is organized as follows: The clustering methods are reviewed in Section 2. Section 3 presents the historical review of Type-II Fuzzy Logic. Section 4 is dedicated to the proposed method and Section 5 presents the experimental results. Finally, conclusions are presented in Section 6.

2. Clustering Methods

The general philosophy of clustering is to divide the initial set into homogenous groups [6] and to reduce the data [1]. Clustering is useful in several exploratory decision-making, machine learning, data mining, image segmentation, and pattern classification [7]. In literature, most of the clustering methods can be classified into two types: Crisp clustering and Fuzzy clustering. Crisp clustering assigns each data to a single cluster and ignores the possibility that these data may also belongs to other clusters [8]. However, as the boundaries between clusters could not be defined precisely, some of the data could belong to more than one cluster with different positive degrees of memberships [6]. This clustering method considers each cluster as a fuzzy set and the membership function measures the degree of belonging of each feature in a cluster. So, each feature may be assigned to multiple clusters with some degree of belonging [8]. Two important applied models of fuzzy clustering, Fuzzy C-Means, and Possibilistic C-Means are described as follows:

Fuzzy C-Means (FCM): Fuzzy C-Means clustering model can be defined as follows [9]:

$$\min\left\{J(x,\mu,c)=\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}^{m}\,d_{ij}^{2}\right\} \quad (1)$$

$$\text{ST:}\begin{cases}0<\sum_{j=1}^{N}\mu_{ij}<N \quad \forall i\in(1,2,\dots,c) & (2)\\ \sum_{i=1}^{c}\mu_{ij}=1 \qquad \forall j\in(1,2,\dots,N) & (3)\end{cases}$$

where, μ_{ij} is the degree of belonging of the j^{th} data to the i^{th} cluster, d_{ij} is the distance between the j^{th} data and the i^{th} cluster center, m is the degree of fuzziness, c is the number of clusters, and N is the number of the data.

Although FCM is a very good clustering method, it has some disadvantages: the obtained solution may not be a desirable solution and the FCM performance might be inadequate, specially when the data set is contaminated by noise. In addition, the membership values show degrees of

ISBN: 978-989-95079-6-8

probabilities of sharing [10] and not only depend on the distance of that point to the cluster center, but also on the distance from the other cluster centers [11]. In addition, when the norm used in the FCM method is different to the Euclidean, introducing restrictions is necessary, e.g., Gustafson and Kessel [12] and Windham limit the volume of the groups using fuzzy covariance and scatter matrices, respectively [13].

Possibilistic C-Mean (PCM): To improve the FCM weaknesses, Krishnapuram and Keller created a possibilistic approach, which uses a possibilistic type of membership function to describe the degree of belonging. It is desirable that the memberships for representative feature points be as high as possible and unrepresentative points have low membership. The objective function, which satisfies the requirements, is formulated as follows [3]:

$$\min\left\{J_m(x,\mu,c)=\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}^{m}\,d_{ij}^{2}+\sum_{i=1}^{c}\eta_i\sum_{j=1}^{N}(1-\mu_{ij})^{m}\right\} \quad (4)$$

where, d_{ij} is the distance between the j^{th} data and the i^{th} cluster center, μ_{ij} is the degree of belonging of the j^{th} data to the i^{th} cluster, m is the degree of fuzziness, η_i is a suitable positive number, c is the number of the clusters, and N is number of the data. μ_{ij} can be obtained by using (5) [3]:

$$\mu_{ij}=\frac{1}{1+\left(\frac{d_{ij}^{2}}{\eta_i}\right)^{\frac{1}{m-1}}} \quad (5)$$

where, d_{ij} is the distance between the j^{th} data and the i^{th} cluster center, μ_{ij} is the degree of belonging of the j^{th} data to the i^{th} cluster, m is the degree of fuzziness, η_i is a suitable positive numbers. The value of η_i determines the distance at which the membership value of a point in a cluster becomes 0.5. In practice, (6) is used to obtained η_i values. The value of η_i can be fixed or changed in each iteration by changing the values of μ_{ij} and d_{ij}, but the care must be exercised, since it may lead to instabilities [3]:

$$\eta_i=\frac{\sum_{j=1}^{N}\mu_{ij}^{m}\,d_{ij}^{2}}{\sum_{j=1}^{N}\mu_{ij}^{m}} \quad (6)$$

The PCM is more robust in the presence of noise, in finding valid clusters, and in giving a robust estimate of the centers [14].

Updating the membership values depends on the distance measurements [11].The Euclidean and Mahalanobis distance are two common ones. The Euclidean distance works well when a data set is compact or isolated [7] and Mahalanobis distance takes into account the correlation in the data by using the inverse of the variance-covariance matrix of data set which could be defines as follows [15]:

$$\mathrm{D}=\sum_{\mathrm{i,j=1}}^{\mathrm{i,j=p}}A_{ij}\,(x_i-y_i)(x_i-y_j) \quad (7)$$

$$A_{ij}=\rho_{ij}\,\sigma_i\sigma_j \quad (8)$$

where, x_i and y_i are the mean values of two different sets of parameters, X and Y. σ_i^2 are the respective variances and ρ_{ij} is the coefficient of correlation between the i^{th} and j^{th} variates. Gustafson and Kessel proposed a new approach based on Mahalanobis distance, which enables the detection of ellipsoidal clusters. Their approach focused on the case where the matrix A is different for each cluster [12].

Satisfying the underlying assumptions, such as shape and number, is another important issue in clustering methods, which could be obtained by validation indices. Xie & Beni's (XB) and Kwon are two common validity indices [1]. Xie and Beni defined a cluster validity, (9), which aims to quantify the ratio of the total variation within clusters and the separation of the clusters [1]:

$$XB(c)=\frac{\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^{m}\left\|x_j-v_i\right\|^{2}}{N\min\limits_{i\neq j}\left\|v_i-v_j\right\|^{2}} \quad (9)$$

where, μ_{ij} is the degree of belonging of the j^{th} data to the i^{th} cluster, v_j is the center of the j^{th} cluster, m is the degree of fuzziness, c is the number of clusters, and N is number of the data. The optimal number of clusters should minimize the value of the index [1]. However, in practice, when $c\rightarrow N\Rightarrow XB\rightarrow 0$ and it usually does not generate appropriate clusters. The $V_k(c)$, (10), was proposed by Kwon based on the improvement of the XB index [16];

$$V_k(c)=\frac{\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^{m}\left\|x_j-v_i\right\|^{2}+\frac{1}{c}\sum_{i=1}^{c}\|v_i-\bar{v}\|^{2}}{\min\limits_{i\neq j}\left\|v_i-v_j\right\|^{2}} \quad (10)$$

where, μ_{ij} is the degree of belonging of the j^{th} data to the i^{th} cluster, v_j is center of the j^{th} cluster, $\bar{v}$ is the mean of the cluster centers, m is the degree of fuzziness, c is the number of clusters, and N is number of the data. To assess the effectiveness of clustering method, the smaller the V_k, the better the performance [16].

All of the clustering methods and validation indices, mentioned above, are based on Type-I fuzzy set. However, in real world, there exists many uncertainties, which Type-I fuzzy could not model them. Type-II fuzzy set, on the other hand, is able to successfully model these uncertainties [4].

3. Type-II Fuzzy Clustering

The concept of a Type-II fuzzy set, was introduced by Zadeh as an extension of Type-I fuzzy set [17]. A Type-II fuzzy set is characterized by fuzzy membership function, i.e., the membership grade for each element of this set is a fuzzy set in interval [0,1]. Such sets can be used in situations where there are uncertainties about the membership values [18]. Type-II fuzzy logic is applied in many clustering methods e.g., [19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. There are essentially two types of Type-II fuzziness: Interval-Valued Type-II and generalized Type-II. Interval-Valued Type-II fuzzy is a special Type-II fuzzy, where the upper and lower bounds of membership are crisp and the spread of membership distribution is ignored with the assumption that membership values between upper and lower values are

ISBN: 978-989-95079-6-8

uniformly distributed or scattered with a membership value of 1 on the $\mu(\mu(x))$ axis (Figure 1.a). Generalized Type-II fuzzy identifies upper and lower membership values as well as the spread of membership values between these bounds either probabilistically or fuzzily. That is there is a probabilistic or possibilistic distribution of membership values that are between upper and lower bound of membership values in the $\mu(\mu(x))$ axis (Figure 1.b) [29].

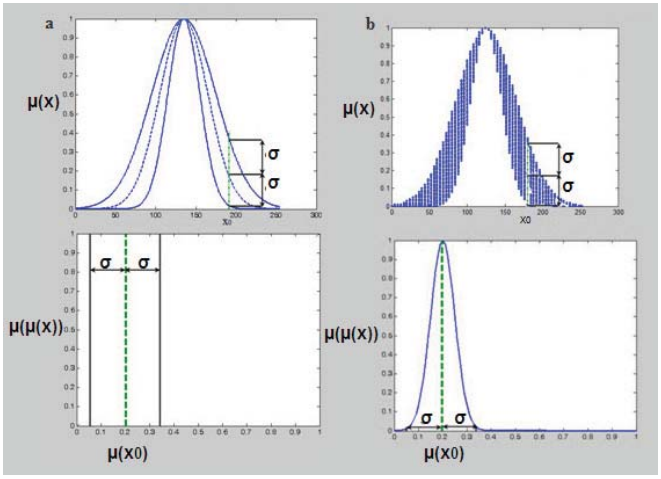


Figure 1:(a) Interval Valued Type-II (b) Generalized Type-II

4. Proposed Type-II PCM Method

Considering the growing application areas of Type-II fuzzy logic, designing a Type-II clustering method is essential. Several researchers designed a Type-II fuzzy clustering method based on FCM but FCM itself has some weaknesses, which make some of the developed methods ineffective in situations in which the data set is contaminated by noise, the norm used is different from the Euclidean, or the pixels on an input data are highly correlated. PCM could improves these weaknesses.

The proposed method is the extension of Krishnapuram and Keller Possibilistic C-Mean (PCM). Here, the membership functions are Type-II Fuzzy, the distance is assumed to be Euclidean and Mahalanobis and Type-II Kwon validity index is used to find the optimal degree of fuzziness (*m)* and number of clusters (*c)*. The proposed Type-II PCM model is as follows:

$$J_m(x,\tilde{\mu},c) = min\left[\sum_{i=1}^{c}\sum_{j=1}^{N}\tilde{\mu}_{ij}^{m} D_{ij} + \sum_{i=1}^{c}\eta_i\sum_{j=1}^{N}\left(1-\tilde{\mu}_{ij}\right)^m\right] \quad (11)$$

$$S.T:\begin{cases} 0 < \sum_{j=1}^{N}\tilde{\mu}_{ij} < N & (12) \\ \tilde{\mu}_{ij} \in [0,1] \quad \forall i,j & (13) \\ \max\tilde{\mu}_{ij} > 0 \quad \forall j & (14) \end{cases}$$

where, $\tilde{\mu}_{ij}$ is Type-II membership for the i^{th} data in the j^{th} cluster, D_{ij} is the Mahalanobis distance of the i^{th} data to the j^{th} cluster's center, η_i is positive numbers, c is the number of the clusters, and N is the number of input data. The first term make the distance to the cluster's center be as low as possible and the second term make the membership values in a cluster to be as large as possible. The membership values for data in each cluster must lie in the interval [0,1], and their sum are restricted to be smaller than the number of input data, as shown in (12), (13), and (14).

Minimizing $J_m(x,\tilde{\mu},c)$ with respect to $\tilde{\mu}_{ij}$ is equivalent to minimizing the individual objective function defined in (15) with respect to $\tilde{\mu}_{ij}$ (provided that the resulting solution lies in the interval [0,l]).

$$J_m^{ij}(x,\tilde{\mu},c) = \tilde{\mu}_{ij}^{m} D_{ij} + \eta_i\left(1-\tilde{\mu}_{ij}\right)^m \quad (15)$$

Differentiating (15) with respect to $\tilde{\mu}_{ij}$ and setting it to 0, leads to (16) which satisfies (12), (13), and (14).

$$\tilde{\mu}_{ij} = \frac{1}{1+\left(\frac{D_{ij}}{\eta_i}\right)^{\frac{1}{m-1}}} \quad \forall i = 1,\dots,c \quad (16)$$

$\tilde{\mu}_{ij}$ is updated in each iteration and depends on the D_{ij} and η_i. As mentioned in [3], the value of η_i determines the distance at which the membership value of a point in a cluster becomes 0.5. In general, it is desirable that η_i relate to i^{th} cluster and be of the order of D_{ij} [3].

$$\eta_i = \frac{\sum_{j=1}^{N}\tilde{\mu}_{ij}^{m} D_{ij}}{\sum_{j=1}^{N}\tilde{\mu}_{ij}^{m}} \quad \forall i = 1,\dots,c \quad (17)$$

where, D_{ij} is the distance measure and number of clusters (c) and degree of fuzziness (m) are unknown. Since the parameter $\eta_i$is independent of the relative location of the clusters, the membership value $\tilde{\mu}_{ij}$ depends only on the distance of a point to the cluster centre. Hence, the membership of a point in a cluster is determined solely by how far a point is from the centre and is not coupled with its location with respect to other clusters [11].

The clustering method needs a validation index to define the number of clusters (c) and degree of fuzziness (m), which are used in (15), (16), and (17). Therefore a Type-II Kwon Index based on Kwon index is designed, which is represented by (18):

$$\tilde{V}_k(c) = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N}\tilde{\mu}_{ij}^{m}\left\|x_j-\tilde{v}_i\right\|^2 + \frac{1}{c}\sum_{i=1}^{c}\|v_i-\bar{\tilde{v}}\|^2}{\min_{i\neq j}\left\|\tilde{v}_i-\tilde{v}_j\right\|^2} \quad (18)$$

where, $\tilde{\mu}_{ij}$ is Type-II possibilistic membership values for the i^{th} data in the j^{th} cluster, $\tilde{v}_i$ is the i^{th} center of cluster, $\bar{\tilde{v}}$ is the mean of centers, N is the number of input data, c is the number of the classes and m is the degree of fuzziness. The first term in the numerator denotes the compactness by the sum of square distances within clusters and the second term denotes the separation between clusters, while denominator denotes the minimum separation between clusters, so the smaller the $\tilde{V}_k(c)$, the better the performance.

In sum, the steps of the proposed clustering method are described bellow and are shown in Figure 2.

Step 1: Define the initial parameters including:

- Maximum iteration of the method (R)
- Number of the clusters (*c=2* is the initial value)

ISBN: 978-989-95079-6-8

- Degree of fuzziness (m=1.5 is the initial value)
- Primary membership functions ($\tilde{\mu}_{ij}{}^{0}$) (Note that these membership functions are Type-II)

Step 2: Estimate η_i by using (17).

Step 3: Calculate the membership functions for each data in each cluster ($\tilde{\mu}_{ij}{}^{r}$) by using (16).

Step 4: If the difference between two membership functions for each data is bigger than the threshold, defined by user, ($\left|\tilde{\mu}_{ij}{}^{r} - \tilde{\mu}_{ij}{}^{r-1}\right| > \varepsilon$) go to step 4.1 and 4.2. On the other hand, go to step 4.3.

Step 4.1: $r = r + 1$

Step 4.2: Recalculate $\tilde{\mu}_{ij}{}^{r}$

Step 4.3: Compute the Kwon index ($\tilde{V}_k$).

Step 5: If the difference between two Kwon indexes for each membership functions is bigger than the threshold, ($\left|\tilde{V}_k{}^{r} - Vkr{-}1{>}\varepsilon\right)$, go to step 5.1. On the other hand, go to step 5.2.

Step 5.1: Increase degree of fuzziness, m and run the method for another iteration.

Step 5.2: If the number of clusters are smaller than the number of data, ($c < N$), go to step 5.2.1. On the other hand, go to step 5.2.2.

Step 5.2.1: Run the method for another iteration

Step 5.2.2: Returns the value of *c, m*, and $\tilde{\mu}_{ij}{}^{r}$.

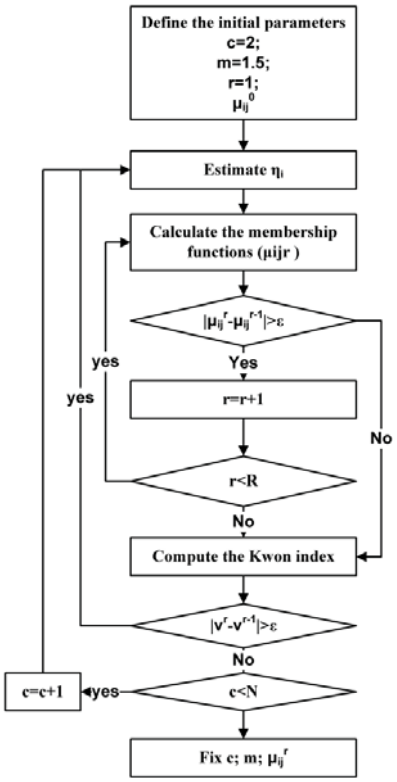


Figure 2: Type-II PCM Algorithm

5. Expand and Compare

In order to show the behavior of the proposed method, an image is used as input data. There may have been many uncertainties in images such as uncertainties caused by projecting a 3D object into a 2D image or digitizing analog pictures into digital images, and the uncertainty related to boundaries and non-homogeneous regions. Therefore, Type-II fuzzy logic can provide better performance than Type-I, this is shown by generating two models based on Type-I and Type-II Possibilistic C-Mean (PCM), each has two kind of distance measure, Euclidean and Mahalanobis. The Kwon Validity index is used to validate the results (m and c) as shown in Figure 2 and table 1, 2, 3 and 4. The results show that Type-II PCM using Mahalanobis distance can obtain better values for degree of fuzziness and number of clusters, which both are used for calculating the membership functions.

Table 1 shows the results of Kwon Validity index for Type-I PCM and Table 2 shows the results for Type-II PCM. In both of these tables Euclidean distance is used as the distance function. The elements of the tables are number of clusters (c) and degree of fuzziness (m) as input variables and the Kwon index values ($\tilde{V}_k$) as results, e.g., for m=2.7 and c=3, (2.7,3), the Kwon index for Type-I PCM is 4.5696 and is 750.34 for Type-II PCM. For m=3.3 and c=2, (3.3,2), the Kwon index could not be defined in both Type-I and Type-II PCM.

Table 3 shows the results of Kwon Validity index for Type-I PCM and Table 4 shows the results for Type-II PCM. In both of these tables Mahalanobis distance are used. The elements of the tables are number of clusters *(c)* and degree of fuzziness *(m)* as an input variables and the Kwon index values as results, i.e., for m=4.1 and c=3, (4.1,3), the Kwon index is 1.88 for Type-I PCM and is 33.686 for Type-II PCM.

ISBN: 978-989-95079-6-8

Table 1- Kwon Values for Type-I PCM with Euclidean Distance

c\m	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5
2	0.250	0.25	0.25	0.2500	0.2501	0.2509	0.2515	0.2662	0.2758	NaN	NaN
3	3.734	3.958	3.956	3.9877	4.0142	4.0762	4.5696	5.7768	6.889	NaN	NaN
4	5.95	5.944	6.05	6.3488	6.4236	2.464	987.99	7.0394	NaNi	NaN	NaN
5	5.351	5.367	5.474	5.7481	5.818	5.4528	991.14	11769	NaNi	NaN	NaN

Table 2- Kwon Values for Type-II PCM with Euclidean Distance

c\m	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5
2	57.82	38.36	14.53	14.436	30.184	6.2069	2.6636	1.5969	2.1884	NaN	NaN
3	514.7	113.3	92.05	37.547	484.97	201.14	750.34	392.91	10.381	NaN	NaN
4	9.17E+05	2.62E+05	8654.5	3.78E+05	1979.5	92.563	1854.1	331.16	NaN	NaN	NaN
5	1.01E+06	77060	6454	1534.8	386.57	130.98	9276.8	1240.2	NaN	NaN	NaN

Table 3- Kwon Values for Type-I PCM with Mahalanobis Distance

c\m	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1
2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.27	0.28	0.28	0.28	0.28	0.27	0.27	0.27	0.27	0.27	0.26	0.26
3	3.73	3.96	3.95	3.99	4.01	4.08	4.57	5.78	6.88	7.56	2.53	1.02	0.98	1.88	3.76	8.04	16.6	37.5	84.2
4	5.95	5.94	6.05	6.35	6.42	2.46	988	7.04	10.44	32.1	28.6	46.4	117	351	716	3080	5262	12599	28180
5	5.35	5.37	5.47	5.75	5.82	5.45	991	11769	587	282	56.5	49.3	121	356	716	3061	5235	12625	28521

Table 4- Kwon Values for Type-II PCM with Mahalanobis Distance

c\m	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1
2	57.8	38.3	14.5	14.4	30.1	5.75	2.69	1.60	2.74	4.42	6.2	9.09	9.26	9.54	9.71	9.23	11.2	8.58	8.96
3	515	113	92.1	37.5	485	345	600	481	1043	29.5	11.1	9.24	13.9	33.7	76.2	232	399	853	2318
4	9.17E+05	2.62E+05	8654.5	3.78E+05	1979.5	92.563	192.2	417.1	220.20	61.774	177.69	539.7	132.5	490.3	774.6	500.61	570.76	1.41E+05	4.21E+05
5	1.01E+06	77060	6454.5	1534.8	386.57	130.98	956.4	480.6	230.29	885.7	278.64	1.03E+05	6.50E+05	1.93E+06	1.26E+07	2.32E+07	4.90E+07	6.78E+07	1.11E+08

By comparing Type-I and Type-II PCM, the following conclusions can be obtained:

- In the same distance function cases, Kwon index is ascending for Type-I PCM, and its procedure is not clear for Type-II PCM case. Therefore, in Type-I PCM in all conditions the optimum *(m,c)* pairs are (1.5,2), (1.5,3), (1.5,4), (1.5,5), which may not be good results. However, in Type-II PCM the optimum *(m,c)* pairs are (2.9,2), (3.7,3), (3.3,4), (2.5,5), which seems to be good results. Figures 3 and 4 show the Kwon index results for *c=2* measured by Mahalanobis and Euclidean distance.

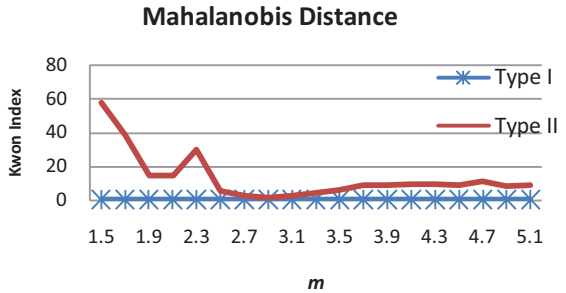


Figure 3: Kwon Index Result for c=2 and Mahalanobis Distance (based on Table 3 and 4)

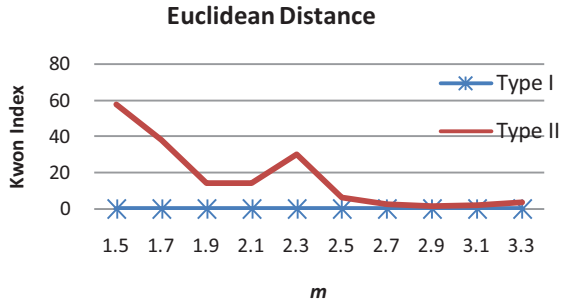


Figure 4- Kwon Index Result for c=2 and Euclidean Distance (based on Table 3 and 4)

- In the same type of fuzzy logic (Type-I or Type-II) case, for m<3.1 using two different distance functions did not show many differences for Kwon index results. However, for m>3.1, the Kwon index could be calculated by Mahalanobis but it is not defined for Euclidean distance, as shown in Figures 5.

ISBN: 978-989-95079-6-8

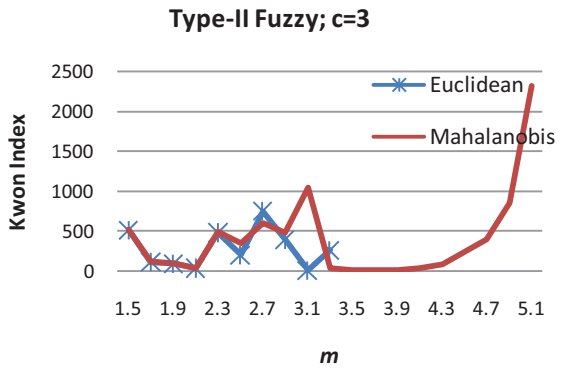


Figure 5: Kwon Index for c=3 and Different Distance Functions (based on Table 2 and 4)

6. Conclusions

This paper has presented a Type-II Possibilistic C-Mean (PCM) method for clustering purposes. The results of the proposed method are compared with Type-I PCM using an image as an input data and two kind of distance functions, Euclidean and Mahalanobis. The results show that Type-II PCM using Mahalanobis distance can provide better values for degree of fuzziness and number of clusters, which both are used in calculating the membership functions. Therefore the proposed clustering method is more accurate, can provide better performance and can handle uncertainties that exist in the data efficiently.

References

[1] J.V. Oliveira and W. Pedrycz, *Advances in Fuzzy Clustering and Its Applications*. John Wiley & Sons Ltd., 2007.

[2] J.M. Mendel and R. John, Type 2 Fuzzy Sets Made Simple. *IEEE Transactions On Fuzzy Systems*, 10(2):117-127, 2002.

[3] R. Krishnapuram and J.M. Keller, A Possibilistic Approach to Clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98-110, 1993.

[4] R. Seising (Ed.), *Views on Fuzzy Sets and Systems from Different Perspectives*. Springer-Verlag, 2009.

[5] J.M. Mendel et al., Interval Type 2 Fuzzy Logic Systems Made Simple. *IEEE Transactions on Fuzzy Systems,* 14(6):808-821, 2006.

[6] E. Nasibov and G. Ulutagay, A New Unsupervised Approach for Fuzzy Clustering. *Fuzzy Sets and Systems*, Article in Press.

[7] A.K Jain et al., Data clustering: A review. *ACM Computing Surveys*, 31(3):264-323, 1999.

[8] M. Menard and M. Eboueya Extreme Physical Information and Objective Function in Fuzzy Clustering. *Fuzzy Sets and Systems*, 128(3):285-303, 2002.

[9] I.B. Turksen, *An Ontological and Epistemological Perspective of Fuzzy Set Theory*. Elsevier Inc., 2006.

[10] H. Frigui, R. Krishnapuram, A Robust Competitive Clustering Algorithm with Applications in Computer Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):450-465 1999.

[11] K.P. Detroja et al., A Possibilistic Clustering Approach to Novel Fault Detection and Isolation. *Journal of Process Control,* 16(10):1055-1073, 2006.

[12] D.E. Gustafson and W.C. Kessel, Fuzzy Clustering with Fuzzy Covariance Matrix. *Proc. IEEE CDC, San Diego, CA,* 761-766, 1979.

[13] A. Flores-Sintas et al., A Local Geometrical Properties Application to Fuzzy Clustering. *Fuzzy Sets and Systems*, 100(3): 245-256, 1998.

[14] O. Nasraoui and R. Krishnapuram, Crisp Interpretations of Fuzzy and Possibilistic Clustering Algorithms. *In Proceedings of the 3rd European Congress on Intelligent Techniques and Soft Computing*, 1312-1318, 1995.

[15] P.C. Mahalanobis, On the generalized distance in statistics. *Proceedings National Institute of Science*, 2, 1936.

[16] C. Duo et al., An Adaptive Cluster Validity Index for the Fuzzy C-Means. *International Journal of Computer Science and Network Security*, 7(2)146-156, 2007.

[17] L.A. Zadeh, The Concept of a Linguistic Variable and its Applications to Approximate Reasoning-I. *Information Sciences*, 8:199-249, 1975.

[18] O. Castillo and P. Melin, *Type-2 Fuzzy Logic: Theory and Applications*. Springer-Verlag, 2008.

[19] A. Celikyilmaz and I.B. Turksen, Enhanced Type 2 Fuzzy System Models with Improved Fuzzy Functions. *Annual Conference of the North American Fuzzy Information Processing Society*, 140-145, 2007.

[20] B.I. Choi and F.C. Rhee, Interval Type-2 Fuzzy Membership Function Generation Methods for Pattern Recognition. *Information Sciences, Article in Press*, 2008.

[21] C. Hwang and F.C. Rhee, An interval type-2 Fuzzy C-Spherical Shells algorithm. *IEEE International Conference on Fuzzy Systems*, 2:1117-1122, 2004.

[22] C. Hwang, F.C. Rhee, Uncertain Fuzzy Clustering: Interval Type-2 Fuzzy Approach To C-Means. *IEEE Transactions on Fuzzy Systems,* 15(1):107-120, 2007.

[23] D.C. Lin and M.S. Yang, A Similarity Measure between Type 2 Fuzzy Sets with its Application to Clustering. *4th International Conference on Fuzzy Systems and Knowledge Discovery*, 1:726-731, 2007.

[24] F.C. Rhee, Uncertain Fuzzy Clustering: Insights and Recommendations. *IEEE Computational Intelligence Magazine*, 2(1):44-56, 2007.

[25] F.C. Rhee and C. Hwang, A Type-2 Fuzzy C-Means Clustering Algorithm. *Annual Conference of the North American Fuzzy Information Processing Society*, 4:1926-1929, 2001.

[26] O. Uncu and I.B. Turksen, Discrete Interval Type 2 Fuzzy System Models using Uncertainty in Learning Parameters. *IEEE Transactions on Fuzzy Systems*, 15(1):90-106, 2006.

[27] W.B. Zhang et al., Rules Extraction of Interval Type-2 Fuzzy Logic System based on Fuzzy C-Means Clustering. *4th International Conference on Fuzzy Systems and Knowledge Discovery*, 2:256-260, 2007.

[28] W.B. Zhang and W.J. Liu, IFCM: Fuzzy Clustering for Rule Extraction of Interval Type-2 Fuzzy Logic System. *Proceedings of 46th IEEE Conference on Decision and Control,* 5318-5322, 2007.

[29] I.B. Turksen, Type 2 Representation and Reasoning for CWW. *Fuzzy Sets and Systems*, 127(1):17-36, 2002.

ISBN: 978-989-95079-6-8

Opposite Fuzzy Sets with Applications in Image Processing

Hamid R. Tizhoosh
Pattern Analysis and Machine Intelligence Laboratory
Systems Design Engineering, University of Waterloo
200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada
Email: tizhoosh@uwaterloo.ca

Abstract— *Diverse forms of the concept of opposition are already existent in philosophy, linguistics, psychology and physics. The interplay between entities and opposite entities is apparently fundamental for balance maintenance in almost a universal manner. However, it seems that we have failed to incorporate oppositional thinking in engineering, mathematics and computer science. Especially, the set theory in general, and fuzzy set theory in particular, do not offer a formal framework to incorporate opposition in inference engines. Considering sets along with their opposites can establish a new computing scheme with a wide range of applications. In this work, preliminary definitions for opposite fuzzy sets will be established. The underlaying idea of opposition-based computing is the simultaneous consideration of guess and opposite guess, and estimate and opposite estimate, in order to accelerate learning, search and optimization. To demonstrate the applicability and usefulness of opposite fuzzy sets, a new image segmentation algorithm will be proposed as well.*

Keywords— Fuzzy sets, opposition, opposite fuzzy sets, antonym, antonymy, complement

1 Motivation

Wherever we observe the physical world and the human society, some notions of opposition exist. The intensity of *oppositeness* and the interplay between entities and opposite entities may vary greatly by closer observation. However, the antipodality, duality, complementarity, antithetics, polarity, in one word, opposition is, in different forms and at various intensity levels, present almost everywhere.

Strikingly, the human communication, expressed verbally and in written form, is without opposition unimaginable. How often we hear "on the one hand" followed by "and on the other hand" when humans weigh opposite alternatives, or how frequently we make use of antonyms: cold vs. hot, short vs. tall, true vs. false etc., or how impressively we make a point by using sayings such as "actions speak louder than words" (activity vs. passivity), "All for one and one for all" (union of opposites), "the pen is mightier than the sword" (spirit versus might). In spite of the virtual omnipresence of opposition, there is no formal framework in logic and set theory to integrate oppositional concepts. This preliminary work aims at providing elementary definitions and thoughts in this very direction.

2 What is Opposition?

Opposition is related to entities, objects or their abstractions of the same nature that are completely different in some manner. For instance, *cold* and *hot* both describe a certain temperature perception (are of the same kind), however, are completely different since they are located at opposite spots of temperature scale. Transition from one entity to its opposite is understandably accompanied with rapid and fundamental changes. Social revolutions, for instance, mainly aim at attaining opposite circumstances, dictatorship vs. democracy, by initiating sudden transitions. On the other hand, in many cases the balance due to interplay between entities and opposite entities is more dominant than rapid transition from one state to its opposite. These may be taken as initial thoughts for the purpose to improve algorithms – by smart switching between estimates and opposite estimates. However, in many other cases, this is the coexistence, the synergy of things and opposite things, and not their antagonism, that creates changes and enhances the system state.

In natural language, opposition can be detected at different levels [1]. For instance, directional opposition (north-south, up-down, left-right), adjectival opposition (ugly-handsome, long-short, high-low), and prefix opposition (thesis vs. antithesis, revolution vs. counter-revolution, direction vs. opposite direction) are all different levels of linguistic opposition. Further, one can distinguish complements (mutually exclusive properties: dead-alive, true-false), antonyms (two corresponding points or ranges of a scale: long-short, hot-cold), directional converses (two directions along an axis: east-west, up-down), and relational converses (the relative positions of two entities on opposite sides: above-below, teacher-pupil) [1]. The human communication without utilizing linguistic opposition is unimaginable.

Many examples also exist in the philosophy which deal with opposition. The concept of **yin** and **yang** originates in ancient Chinese philosophy and metaphysics, which describes two primal *opposing* but complementary forces found in all things in the universe. Also in the dualism delivered by Avesta, the book of Zoroastrians, the everlasting fight between *Ahura* (God=good) and *Ahriman* (evil) defines the worldlook. The **Table of Opposites** of Pythagoras, delivered by Aristotle, has his own list of opposites: finite-infinite, odd-even, one-many, right -left, rest-motion, straight-crooked, light-darkness, good-evil, and square-oblong. Many philosophers have focused on *one vs. many* as one of the most fundamental questions in philosophy.

In physics, *antiparticles* are subatomic particles having the same mass as one of the particles of ordinary matter but with opposite electric charge and magnetic moment. The positron (positive electron), hence, is the antiparticle of the negative electron (negatron). As another physical example, we define *electric polarization* as slight relative shift of positive and negative electric charge in opposite directions within an insulator, or dielectric, induced by an external electric field. Polarization occurs when an electric field distorts the negative cloud

ISBN: 978-989-95079-6-8

of electrons around positive atomic nuclei in a direction opposite the field[1]. And we could go back in time to the Newton's Mechanics with its principal of *action* and *reaction* which is a major example for discovering and exploiting oppositional behavior.

In mathematics we have several concepts employing some notions of opposition. For instance, the bisection method for solving equations makes use of *positive* vs. *negative* sign change in order to shrink the search interval. In probability theory, the probability of *contrary* situation is given by $1 - p$ if the initial event occurs with probability p. In Monte Carlo simulation, *antithetic* random numbers are used to reduce the variance, and in category theory there is the concept of *duality*.

These examples show that opposition plays a central rule in many fields ranging from linguistics to physics. In spite of all examples, however, it should be mentioned that understanding and defining opposition may not be trivial in some cases.

3 Opposition-Based Computing

The concept of opposition-based computing (OBC) [2, 4] is based on a quite simple idea: Whenever we are looking for a solution, we should consider the opposite solution as well. This may appear too simple, however, since opposition is almost universal, it can be understood and implemented in many different ways such that the application fields and potential impacts are immense. The idea of opposition-based computing is not entirely new in machine intelligence. Numerous works have treated and employed oppositional concepts in different ways and under different labels [4].

Past research on oppositional concepts has not been conducted under a unified framework. As such, a universal view of opposition-based learning should regard and treat entities and their opposites in a much more general framework, particularly when for application in the context of machine intelligence schemes. Among others, we should not forget that opposition considers both qualitative and abstract entities, and not only numbers.

In a preliminary paper [2], OBC-extensions of neural nets (with opposite weights), genetic algorithms (with anti-chromosomes) and reinforcement agents (with opposite actions) were provided, with experimental results indicating modest improvements in learning/searching speed. More detailed reports, on improvement of the learning behavior of evolutionary algorithms, reinforcement learning, ant colonies and neural nets have been reported as well [3, 5, 6, 7, 9, 10]. Among other benefits of incorporating of OBC into existing learn and search algorithms, total computational time has decreased by 1.6 times for differential evolution (DE) by incorporating anti-chromosomes. (This becomes even more significant when we recall that DE is generally considered a fast algorithm to begin with).

OBC may be regarded as a new philosophy for integration of a-priori knowledge (interplay between things and opposite things during search and learning procedures). However, this is a very special type of a-priori knowledge. Among other potential benefits, working with opposite entities is not accompanied with uncertainties, if the initial entities and their evaluation/fitness are known. In contrast, any work based on similarity and dissimilarity is always to a certain degree uncertain. The main question, however, is how we should find/define the opposites for every given problem. This work will attempt to find some answers within fuzzy set theory.

[1] Encyclopdia Britannica

4 Oppositional Ideas in Fuzzy Systems

Opposite fuzzy sets have been a part of fuzzy systems from the beginning. Fuzzy inferencing is without the concept of opposition virtually impossible. Linguistic terms are generally considered as pairs such as "short-tall", "cold-warm", "positive-negative" etc., something that is so natural and customary that we have overlooked its significance and failed to develop a more systematic paradigm to exploit its true potential.

Some explicit notions of employing opposition can be found in previous works published on fuzzy sets and their applications. One may immediately say that the concept of "negation" is the natural translation of opposition. A set A, describing a set of objects, is negated, $\neg A$, to quantify "not A". However, this is a general misconception. Negation cannot represent opposition, for it is too general to illustrate the oppositeness of attributes. For instance, "not very young" encompasses the entire spectrum of "young" to "very old", whereas the opposite of "very young" is clearly "very old". Most likely Roland Yager was the first one who used the term "opposition" in relation to fuzzy sets [11]. In literature on fuzzy systems, the opposition has been scarcely studied, and that only with respect to *antonyms* and their significance from a pure linguistic perspective [12, 13]. As we will see, antonyms are what we call in this paper "type-I opposite fuzzy sets".

5 Opposite Fuzzy Sets

In this section we attempt to establish a generic framework for opposite fuzzy sets. For all following definitions we consider a general class of membership functions (MFs) $\mu(x) = f(x; \mathbf{a}, \delta_1, \delta_2)$ where the vector $\mathbf{a}$ denotes the points with $\mu(a_i) = 1$, and δ_1 and δ_2 are the function widths to the left and right, respectively (Fig. 1).

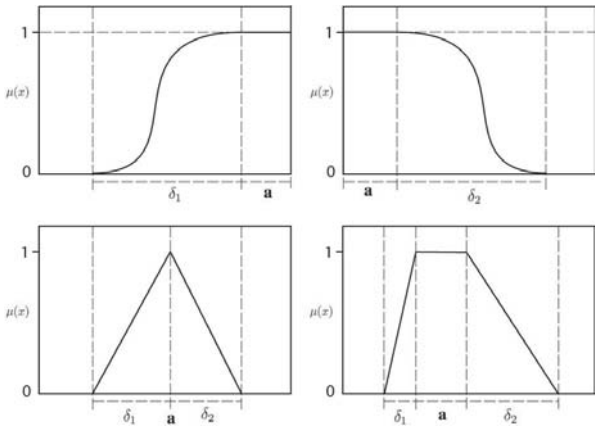


Figure 1: General MFs $\mu(x) = f(x; \mathbf{a}, \delta_1, \delta_2)$

Let X be the universe of discourse and $x \in X$ the elements, objects to be classified.

Definition (Fuzzy Set) - A fuzzy set $A \subset X$ with membership function $\mu_A(x)$ is defined as

$$A = \{(x, \mu_A(x)) | x \in X, \mu(x) \in [0, 1]\} \quad (1)$$

ISBN: 978-989-95079-6-8

The membership function is given as $\mu_A(x) = f(x; \mathbf{a}, \delta)$ where $\mu_A(\mathbf{a}) = 1 \quad \forall a_i \in \mathbf{a}$ and δ is the somatic parameter that changes the shape of the membership function.

Definition (Opposite Fuzzy Set) - Given a fuzzy set $A \subset X$, the opposite fuzzy set $\breve{A} \subset X$ with membership function $\mu_{\breve{A}}(x)$ is defined as

$$\breve{A} = \{(x, \mu_{\breve{A}}(x)) | x \in X, \mu_{\breve{A}}(x) \in [0,1]\} \tag{2}$$

where $\mu_{\breve{A}}(x) = f(x; \breve{\mathbf{a}}, \breve{\delta})$.

The vector $\mathbf{a} = [a_1, a_2, \dots]$ and its opposite vector $\breve{\mathbf{a}} = [\breve{a}_1, \breve{a}_2, \dots]$ represent the points on the universe of discourse with $\mu(a_i) = \mu(\breve{a}_i) = 1$; $\breve{\delta}$ will cause the opposite shape modification compared to original δ.

Example - If a Gaussian, bell curve, function is employed to represent $\mu_A(x) = f(x; \mathbf{a}, \delta)$, then $a = c$ and $\delta = \sigma$. The opposite membership function $\mu_{\breve{A}}(x) = f(x; \breve{\mathbf{a}}, \breve{\delta})$ will be defined with $\breve{a} = \breve{c}$ and $\breve{\delta} = \breve{\sigma}$

Since both elements of $\mathbf{a}$ and δ are real numbers, the definition of opposite fuzzy sets comes down to definition of the opposite of any arbitrary real number a.

Opposite numbers are generally defined for the set of irrational number $\mathbb{Z}$ on the interval $(-\infty, +\infty)$. For each $x \in \mathbb{Z}$ the opposite number $\breve{x}$ is defined by $\breve{x} = -x$. This definition, however, lacks generality. Following, we introduce more comprehensive definitions of opposite numbers [4].

Definition (Type-I Opposite Points[2]) - Let $P = (a_1, a_2, \cdots, a_n)$ be a point in an n-dimensional space, where $a_i \in [X^i_{min}, X^i_{max}] \in \mathbb{R}$. The type-I opposite point $\breve{P} = (\breve{a}_1, \breve{a}_2, \cdots, \breve{a}_n)$ is then completely defined where

$$\breve{a}_i = X^i_{max} + X^i_{min} - a_i. \tag{3}$$

For $n = 1$ following special cases exist:

$$\begin{array}{rcll} \breve{a} & = & -a & \text{for} \quad X_{max} = -X_{min}, \\ \breve{a} & = & 1 - a & \text{for} \quad X_{min} = 0 \quad \& \quad X_{max} = 1, \\ \breve{a} & = & a & \text{for} \quad a = \frac{X_{max} + X_{min}}{2}. \end{array} \tag{4}$$

Definition (Type I Opposite Fuzzy Sets) - The set $\breve{A}$ with membership function $\mu_{\breve{A}}(x) = f(x; \breve{a}, \breve{\delta})$ is type I opposite of the set A with membership function $\mu_A(x) = f(x; a, \delta)$ if $\breve{a}$ and $\breve{\delta}$ are type I opposites of a and δ, respectively.

Definition (Type-I Super-Opposite Points) - Let $P = (a_1, a_2, \cdots, a_n)$ be a point in n-dimensional space with $a_i \in [X^i_{min}, X^i_{max}]$ and its opposite point $\breve{P} = (\breve{a}_1, \breve{a}_2, \cdots, \breve{a}_n)$. Then all points $\breve{P}^s$ are type-I super-opposite of P when $d(\breve{P}^s, P) > d(\breve{P}, P)$ where $d(\cdot, \cdot)$ denotes a metric such as the Euclidian distance.

For $n = 1$ the type-I super-opposite of $a \in [X_{min}, X_{max}]$ can be given as

$$\breve{a}^s \in \begin{cases} [X_{min}, \breve{a}) & \text{for} \quad a > (X_{min} + X_{max})/2 \\ [X_{min}, X_{max}] - \{a\} & \text{for} \quad a = (X_{min} + X_{max})/2 \\ (\breve{a}, X_{max}] & \text{for} \quad a < (X_{min} + X_{max})/2 \end{cases}$$

[2]The terms "type I" and "type II" opposition have been taken from previously published works on OBC [4] and should not be confused with type-1 and type-2 fuzzy sets. Type-I opposites reflect linear opposition whereas type-II opposites constitute nonlinear opposition.

In other words, for $a = \frac{X_{min} + X_{max}}{2}$ the entire interval except a becomes the super-opposite of a. This means that for $a \to \frac{X_{min} + X_{max}}{2}$ the opposition generally loses its benefit against random guesses, since any random guess will be a super-opposite of the initial guess.

Definition (Type I Super-Opposite Fuzzy Sets) - The set $\breve{A}^S$ with membership function $\mu_{\breve{A}^S}(x) = f(x; \breve{a}^S, \breve{\delta}^S)$ is type I super-opposite of the set A with membership function $\mu_A(x) = f(x; a, \delta)$ if $\breve{a}^S$ and $\breve{\delta}^S$ are type I super-opposites of a and δ, respectively.

Definition (Type-I Quasi-Opposite Points) - Let $P = (a_1, a_2, \cdots, a_n)$ be a point in n dimensional space with $a_i \in [X^i_{min} + X^i_{max}]$ and its opposite point $\breve{P} = (\breve{a}_1, \breve{a}_2, \cdots, \breve{a}_n)$. Then all points $\breve{P}^q$ are type-I quasi-opposite of P when $d(\breve{P}^q, \breve{P}) = d(\breve{P}^s, \breve{P})$ and $d(\breve{P}^q, P) < d(\breve{P}^s, P)$ for any super-opposite $\breve{P}^s$.

For $n = 1$ the type-I quasi-opposite of $a \in [X_{min}, X_{max}]$ can be given as

$$\breve{a}^q \in \begin{cases} (\breve{a}, 2\breve{a}] & \text{for} \quad a > (X_{min} + X_{max})/2 \\ [X_{min}, X_{max}] - \{a\} & \text{for} \quad a = (X_{min} + X_{max})/2 \\ [2\breve{a} - X_{max}, \breve{a}) & \text{for} \quad a < (X_{min} + X_{max})/2 \end{cases}$$

Definition (Type I Quasi-Opposite Fuzzy Sets) - The set $\breve{A}^q$ with membership function $\mu_{\breve{A}^q}(x) = f(x; \breve{a}^q, \breve{\delta}^q)$ is type I quasi-opposite of the set A with membership function $\mu_A(x) = f(x; a, \delta)$ if $\breve{a}^q$ and $\breve{\delta}^q$ are type I quasi-opposites of a and δ, respectively.

Figure 2 illustrates type-I super- and quasi-opposition for one dimensional cases. Figure 3 shows a fuzzy subset A, its opposite $\breve{A}$, quasi- and super-opposite.

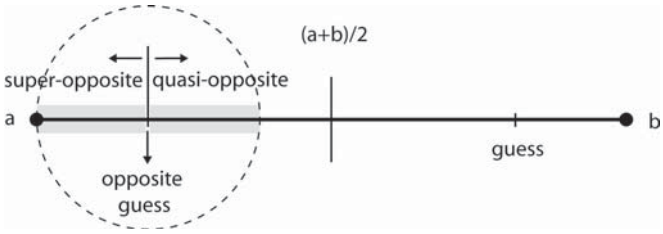


Figure 2: Definition of super- and quasi-opposition based on the distance of the opposite guess from the interval boundary.

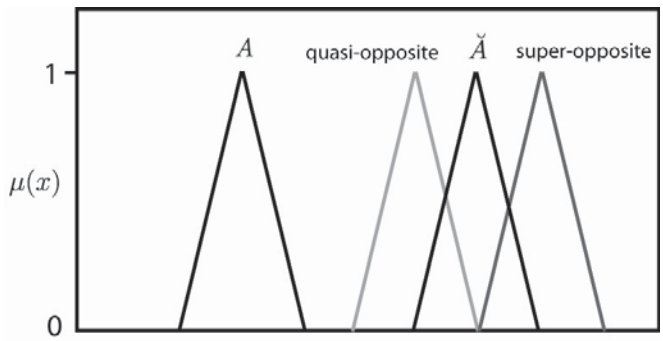


Figure 3: A fuzzy set and its opposites

Obviously type I opposition is a straightforward and a simple understanding of opposition. Generally, only linear,

ISBN: 978-989-95079-6-8

almost-linear and symmetric systems can benefit from type I opposition. In order to deal with complex and nonlinear $X-Y$-relationships we have to access higher levels of opposition.

Theorem - The entropy (fuzziness) of the fuzzy set A, denoted with $E(A)$ is equal to the entropy (fuzziness) of its type I opposite set $\breve{A}_I$: $E(A) = E(\breve{A}_I)$

Proof - since the membership function of $\breve{A}$ is just a reflection and translation of A, then $E(A) = E(\breve{A})$. □

Definition (Type-II Opposite Points) - Let $y = f(x_1, x_2, \cdots, x_n) \in \mathbb{R}$ be an arbitrary function with $y \in [y_{\min}, y_{\max}]$. For every point $P = (a_1, a_2, \cdots, a_n)$ the type-II opposite point $\breve{P} = (\breve{a}_1, \breve{a}_2, \cdots, \breve{a}_n)$ is defined by

$$\breve{a}_i = \{x| \quad \breve{y} = y_{\min} + y_{\max} - y\}. \tag{5}$$

This definition assumes that the function $f(x_1, x_2, \cdots, x_n)$ is not known but $y_{\min}$ and $y_{\max}$ are given or can be estimated. Alternatively, the temporal type-II opposite can be calculated according to

$$\breve{a}_i(t) = \left\{ \; x \; \middle| \quad \breve{y}(t) = \min_{j=1,\ldots,t} y_j + \max_{j=1,\ldots,t} y_j - y(t) \; \right\}. \tag{6}$$

If only the evaluation function $h(X)$ is available, then we may define a temporal *degree of opposition* $\breve{\Phi}$ for any two variables a_1 and a_2:

$$\breve{\Phi}(a_1, a_2, t) = \frac{|h(a_1) - h(a_2)|}{\max\limits_{j=1,\ldots,t} h(a_j) - \min\limits_{j=1,\ldots,t} h(a_j)} \in [0, 1]. \tag{7}$$

The type-II opposite $\breve{a}_i$ of any variable a_i can then be given as

$$\breve{a}_i = a_j|_{\breve{\Phi}(a_i,a_j,t)=\max\limits_k \breve{\Phi}(a_i,a_k,t)}. \tag{8}$$

Definition (Type II Opposite Fuzzy Sets) - The set $\breve{A}_{II}$ with membership function $\mu_{\breve{A}_{II}}(x) = f(x; \breve{a}, \breve{\delta})$ is type II opposite of the set A with membership function $\mu_A(x) = f(x; a, \delta)$ if $\breve{a}$ and $\breve{\delta}$ are type II opposites of a and δ, respectively.

Definition - Degree of Oppositeness for Type I Opposition For two real numbers $a, b \in X$ bounded in $[X_{min}, X_{max}]$, the degree of oppositeness can be given as

$$\breve{\varphi}_I(a, b) = \left(\frac{|a - b|}{X_{max} - X_{min}} \right)^{\beta} \tag{9}$$

where $\beta \in (0, 1]$ controls the opposition intensity.

Definition - Degree of Oppositeness for Type II Opposition Assuming the evaluation function $h(\cdot)$ is given, then for two real numbers $a, b \in X$, the degree of oppositeness can be given as

$$\breve{\varphi}_{II}(a, b) = \left(\frac{|h(a) - h(b)|}{h_{max} - h_{min}} \right)^{\beta} \tag{10}$$

where $\beta \in (0, 1]$ controls the opposition intensity. If the bounds of the evaluation function are not known, then a temporal type-II opposition may be calculated:

$$\breve{\varphi}^{(t)}_{II}(a, b) = \left(\frac{|h(a) - h(b)|}{h^{(t)}_{max} - h^{(t)}_{min}} \right)^{\beta} \tag{11}$$

Based on the introduced definitions, following observations can be made:

- Opposition, in general, can only be determined if the system parameters are bounded and *opposite* has a clear relationship with the finite interval, in which those parameters are defined.
- Type-I opposition delivers the real opposite (type-II) only for linear or quasi-linear functions. Type-I opposition can, however, be more easily calculated.
- Type-II opposition can, in absence of any information about the function $f(\cdot)$, only be calculated if an evaluation function $g(\cdot)$ (error, fitness, reward, cost etc.) is available.

Now that we have basic definitions available we can define the opposition-based computing.

> **Opposition-Based Computing (OBC) -** For any system behaving according a (unknown) function $y = f(x_1, x_2, \cdots, x_n)$ with an evaluation function $h(x_1, x_2, \cdots, x_n)$ with higher/lower values being desirable, for any estimate vector $X = (x_1, x_2, \cdots, x_n)$ to estimate the input-output relationship we calculate its opposite $\breve{X} = (\breve{x}_1, \breve{x}_2, \cdots, \breve{x}_n)$ and consider $\max(h(X), h(\breve{X}))$. The learning/search will continue with X if $h(X) \gtrless h(\breve{X})$ otherwise with $\breve{X}$.

The opposition can be calculated using the aforementioned definitions and based on the specific domain knowledge.

The definition of OBC is making two fundamental assumptions: 1) one of estimate and counter-estimate is always closer to the solution, and 2) considering the opposition is more beneficial than generating independent random solutions and taking the best among them. (this assumption has already been proved [8]).

The aforementioned definition of OBC is primarily meant to be used in the model-free context of machine intelligence, where an evaluation function such as error, cost, fitness, reward etc. is being used within multiple iterations/epochs/generations/episodes in order to find some optimal values. For pure fuzzy systems we may rely on the following definition.

> **Opposition-Based Fuzzy Inference Systems (OFIS) -** An inference mechanism which systematically uses fuzzy sets along with their opposites to construct membership functions and rules and/or carry out the inference is an opposition-based fuzzy inference system.

6 Applications

As mentioned before, we have virtually been using opposite fuzzy sets since Zadeh's first paper. However, due to lack of necessary formalism and due to investigative studies on the nature of opposition in its diverse forms, we have not exploited their potentials.

Most likely, one major application for opposite fuzzy sets is generation of membership functions (MFs) for an inference system. From the beginning days of fuzzy sets, determination of membership functions has been a challenge. Oppositional

ISBN: 978-989-95079-6-8

relationships can be considered to generate membership functions. Table 1 provides a simple algorithm for this purpose. The underlaying idea is as follows: Assuming we are looking for 3 membership functions, and assuming we can fix the position of one of them namely the set A. Then if we automatically determine the opposite fuzzy set $\breve{A}$, the third (middle) membership function, which quantifies "everything else" can be given by $\neg(A \cup \breve{A})$ (see Figure 4).

Table 1: Oppositional generation of membership functions

1.	Determine n to generate $2n+1$ membership functions
2.	For $i = 1 : n$
3.	$\mu_{A_i} = f(x; a_i, \delta_i)$
4.	end
5.	For $i = n+2 : 2n+1$
6.	Determine $\breve{a}_i$ subject to the constraint $\breve{a}_i < \breve{a}_{i+1}$
7.	Determine $\breve{\delta}_i$
8.	$\mu_{\breve{A}_i} = f(x; \breve{a}_i, \breve{\delta}_i)$
9.	end
10.	Determine the middle MF based on $\neg(A \cup \breve{A})$

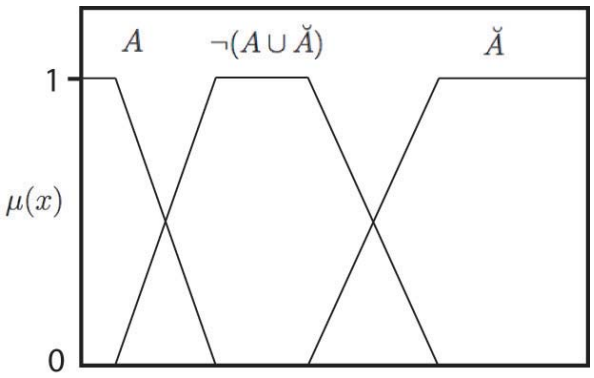


Figure 4: The set A, its opposite $\breve{A}$ and "everything else"

Processing of digital images is certainly an interesting application to verify the applicability of opposite fuzzy sets. The task of *image segmentation* is the most significant part in many vision-based application whereas segmentation is equivalent to classification of each pixel to different classes (segments). Here, based on proposed algorithm in Table 1 we can demonstrate how opposite fuzzy sets can help to segment a digital image.

Let us assume that we are interested in extracting only one object of interest in the image. The rest, regardless how many other objects, are irrelevant. Further assume we have the knowledge that the object of interest is *dark*. Then an oppositional fuzzy rule-based approach to object segmentation can be implemented as described in Table 2. As apparent from line 7 and 8 in this algorithm, we assume that the entropy of A and $\breve{A}$ are equal, which is, according to the theorem in section 5, only proved for type-I opposites. However, experimental results will show that this is a reasonable assumption, which still needs to be theoretically solidified.

Table 2: An OFIS for image segmentation

1. Define the set $A = \{$ *dark pixels* $\}$ with $\mu_A(x)$
2. Calculate the entropy of $A : E(A)$
3. Initialize the opposite membership $\breve{\mu}_A(x)$ at $\breve{a} = L-1$
4. For $i = L-1 : -1 : a+\delta$
5. Define opposite fuzzy set $\breve{A}_i$ at $\breve{a} = i$ (Fig. 5)
6. Calculate the entropy of $\breve{A}_i : E(\breve{A}_i)$
7. Calculate the entropy difference $d_i = \lvert E(\breve{A}_i) - E(A) \rvert$
8. Find the position j with $d_j = \min_i d_i$
9. Define the final opposite fuzzy set at $\breve{a} = j$
10. Define the middle membership function via $\neg(A \cup \breve{A})$
11. (OPTION 1) Create a fuzzy segmentation map via
$g' = \frac{\mu_A(g)+\mu_{\neg(A\cup\breve{A})}(g)\times(L-1)/2+\mu_{\breve{A}}(g)\times(L-1)}{\mu_A(g)+\mu_{\neg(A\cup\breve{A})}(g)+\mu_{\breve{A}}(g)}$
13. (OPTION 2) Threshold the image with $T = (a+\breve{a})/2$
14. $g' = 1$ if $g <= T$, and $g' = 0$ otherwise

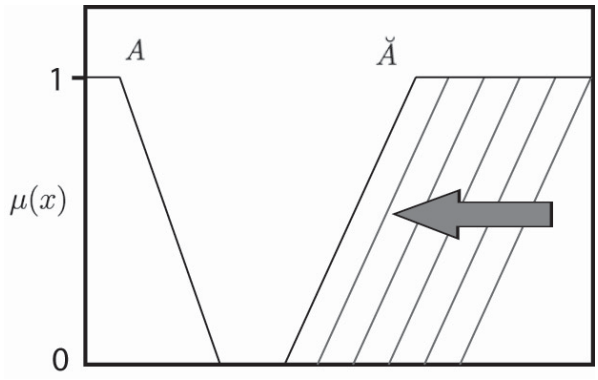


Figure 5: Moving an initial set toward A to find the opposite set $\breve{A}$

The results of this approach are presented in Figure 6. Needless to say that this simple algorithm should only demonstrate the usefulness of opposite fuzzy sets for a significant application. Most likely more sophisticated segmentation algorithms can be developed by incorporating oppositional thinking.

Breast ultrasound images have been used to verify the performance of this method (Fig. 6). The first two scan contain anechoic (dark) breast cysts. These are relatively easy to segment with almost any thresholding technique. The other two contain breast masses that are no cyst and should be examined on malignancy. These cases are challenging and cannot be easily segmented.

7 Conclusions

In this paper, we attempted to establish a formalism for opposite fuzzy sets and introduce some definitions. Preliminary results for a simple image segmentation method were provided as well. Opposition-based fuzzy inference systems (OFISs) exploit domain knowledge and seem to have real-world applications. Many questions, however, remain open with respect to the nature of opposition and how it can be embedded into existing fuzzy systems.

ISBN: 978-989-95079-6-8

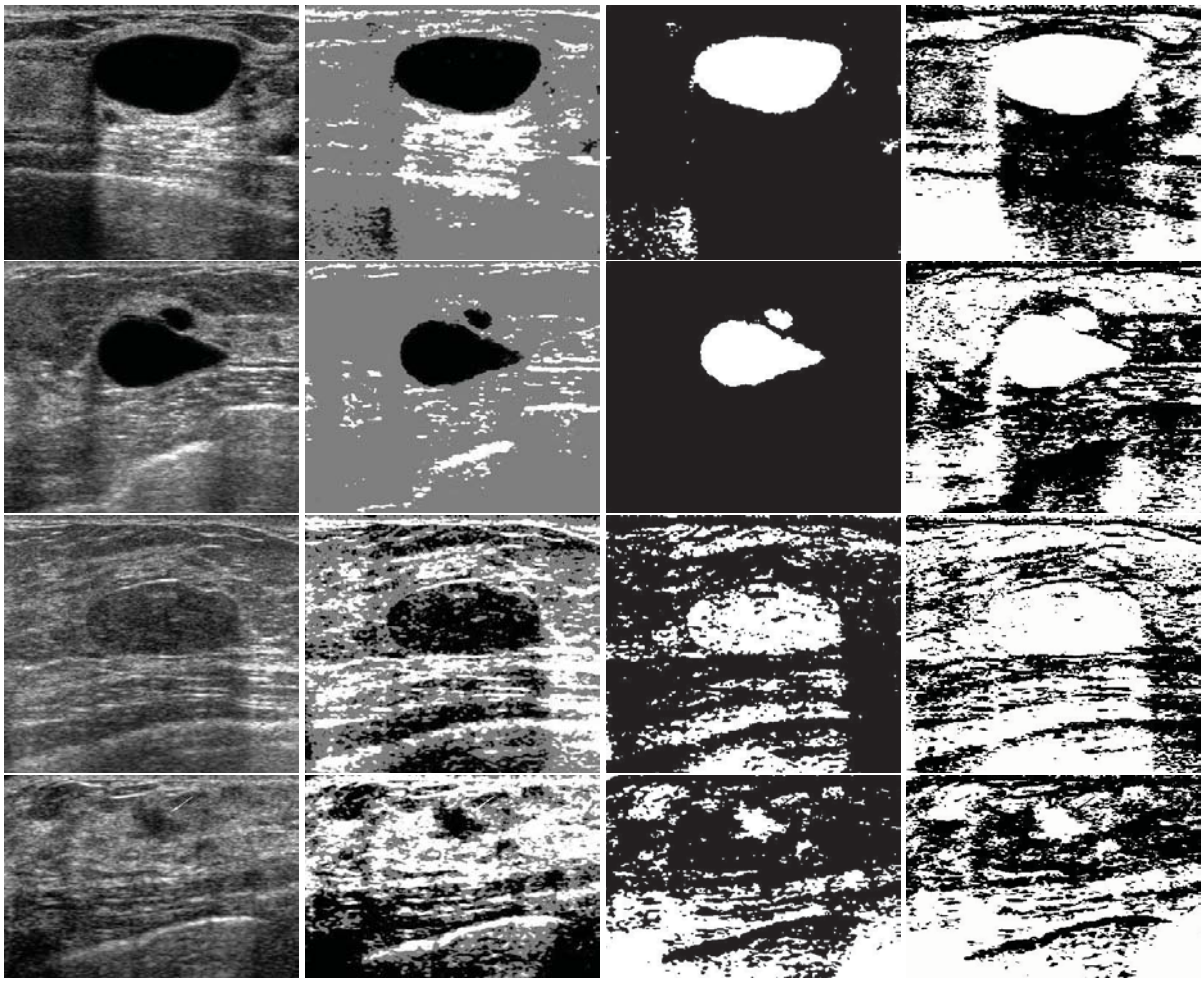


Figure 6: From left: first column: original breast ultrasound scans, second column: fuzzy segmentation with $A, \neg(A \cup \breve{A})$ and $\breve{A}$ (OPTION 1), third column: thresholding with $T = (a + \breve{a})/2$ (OPTION 2). The calculated values from top to bottom: $a = \{1, 1, 64, 62\}$ and $\breve{a} = \{143, 159, 127, 117\}$. The last column depicts the results by Otsu method for comparison [Source of images: Philips]

Acknowledgements - This work was supported by the *Natural Sciences and Engineering Research Council of Canada* (NSERC).

References

[1] Cruse, D.A., Lexical Semantics. Cambridge Univ. Press, 1986.

[2] H.R. Tizhoosh, *Opposition-based learning: a new scheme for machine intelligence*. International conference on computational intelligence for modeling, control and automation, Vienna, Austria, vol. I, pp. 695-701, 2005.

[3] H.R. Tizhoosh, *Opposition-Based Reinforcement Learning*, Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), vol. 10, no. 5, pp. 578-585, 2006.

[4] Tizhoosh, H.R. , Ventresca, M., *Oppositional Concepts in Computational Intelligence*. Springer, 2008.

[5] Ventresca, M., Tizhoosh, H.R., *A diversity maintaining population-based incremental learning algorithm*. Information Sciences 178(21), pp. 4038–405, 2008.

[6] Malisia, A., Tizhoosh, H.R., *Applying Opposition-Based Ideas to the Ant Colony system*. IEEE Symposium Series on Computational Intelligence (IEEE SSCI), pp. 182-189, 2007

[7] Montgomery, J., Randall, M., *Anti-pheromone as a tool for better exploration of search space*. In Dorigo et al. (Eds.) ANTS2002, LNCS 2463, Springer, pp. 100-110, 2002

[8] Rahnamayan, S., Tizhoosh, H.R., Salama, M.M.A., *Opposition versus randomness in soft computing techniques*. Applied Soft Computing 8(2), pp. 906-918, 2008

[9] Rahnamayan, S., Tizhoosh, H.R., Salama, M.M.A., *Opposition-Based Differential Evolution*. IEEE Transactions on Evolutionary Computation, vol. 12, issue 1, pp. 64–79, 2007

[10] Wang, R., Zhang, X., *Particle swarm optimization with opposite particles*., proceedings of 4th Mexican international conference on artificial intelligence, Monterrey, Mexico, November 14-18, 2005, pp. 633-640, 2005

[11] R.R. Yager, *Opposites and Measures of Extremism in Concepts and Constructs*. International Journal of Man-Machine Studies 18(4), pp. 307–322, 1983.

[12] A.R. de Soto, E. Trillas, *On Antonym and Negate in Fuzzy Logic*. International Journal of Intelligent Systems, vol. 14, 295-303, 1999.

[13] V. Novak, *Antonyms and linguistic quantifiers in fuzzy logic*. Fuzzy Sets and Systems 124, pp. 335–351, 2001.

ISBN: 978-989-95079-6-8

Genetic Cooperative-Competitive Fuzzy Rule Based Learning Method using Genetic Programming for Highly Imbalanced Data-Sets

Alberto Fernández[1] Francisco J. Berlanga[2] María J. del Jesus[3] Francisco Herrera[1]

1.Department of Computer Science and Artificial Intelligence, University of Granada
Granada, Spain
2.Department of Computer Science and Systems Engineering, University of Zaragoza
Zaragoza, Spain
3.Department of Computer Science, University of Jaén
Jaén, Spain
Email: alberto@decsai.ugr.es, berlanga@unizar.es, mjjesus@ujaen.es, herrera@decsai.ugr.es

Abstract— *Classification in imbalanced domains is an important problem in Data Mining. We refer to imbalanced classification when data presents many examples from one class and few from the other class, and the less representative class is the one which has more interest from the point of view of the learning task. The aim of this work is to study the behaviour of the GP-COACH algorithm in the scenario of data-sets with high imbalance, analysing both the performance and the interpretability of the obtained fuzzy models. To develop the experimental study we will compare this approach with a well-known fuzzy rule learning algorithm, the Chi et al.'s method, and an algorithm of reference in the field of imbalanced data-sets, the C4.5 decision tree.*

Keywords— Fuzzy Rule-Based Classification Systems, Genetic Fuzzy Systems, Genetic Programming, Imbalanced Data-Sets, Interpretability

1 Introduction

In the area of Data Mining, real world classification problems present some features that can diminish the accuracy of Machine Learning algorithms, such as the presence of noise or missing values, or the imbalanced distribution of classes.

Specifically, the problem of imbalanced data-sets has been considered as one of the emergent challenges in Data Mining [1]. This situation occurs when one class is represented by a large number of examples (known as negative class), whereas the other is represented by only a few (positive class).

Our objective is to develop an empirical analysis in the context of imbalance classification for binary data-sets when the class imbalance ratio is high. In this study, we will make use of Fuzzy Rule Based Classification Systems (FRBCSs), a very useful tool in the ambit of Machine Learning, since they provide a very interpretable model for the end user [2].

We will employ a novel approach, GP-COACH (Genetic Programming-based evolutionary algorithm for the learning of COmpact and ACcurate FRBCS) [3], that learns disjunctive normal form (DNF) fuzzy rules (generated by means of a context-free grammar) and obtains very interpretable FRBCSs, with few rules and conditions per rule, with a high-generalization capability.

We want to analyse whether this model is accurate for data-sets with high imbalance in contrast with an FRBCS, the Chi et al.'s approach [4] and with C4.5 [5], a decision tree algorithm that has been used as a reference in the imbalanced data-sets field [6, 7]. We will also focus on the tradeoff between accuracy and interpretability [8] for the final obtained models. We will employ the Area Under the Curve (AUC) metric [9] to compute the classification performance, whereas we will measure the interpretability of the system by means of the number of rules in the system.

We have selected a large collection of data-sets with high imbalance from UCI repository [10] for developing our empirical analysis. In order to deal with the problem of imbalanced data-sets we will make use of a preprocessing technique, the "Synthetic Minority Over-sampling Technique" (SMOTE) [11], to balance the distribution of training examples in both classes. In this manner, we will analyse the positive synergy between the GP-COACH model and the SMOTE preprocessing technique for dealing with imbalanced data-sets. Furthermore, we will perform a statistical study using non-parametric tests [12, 13, 14] to find significant differences among the obtained results.

This contribution is organized as follows. First, Section 2 introduces the problem of imbalanced data-sets, describing its features, how to deal with this problem and the metric we have employed in this context. Next, in Section 3 we present the GP-COACH algorithm, explaining in detail the characteristics of this novel approach. Section 4 contains the experimental study for GP-COACH, Chi et al.'s and C4.5 algorithms regarding performance and interpretability. Finally, Section 5 summarizes and concludes the work.

2 Imbalanced Data-Sets in Classification

Learning from imbalanced data is an important topic that has recently appeared in the Machine Learning community [15, 16, 17]. The significance of this problem consists in its presence in most of the real domains of classification, such as fraud detection [18], risk management [19] and medical applications [20] among others.

This problem occurs when the number of instances of one class is much lower than the instances of the other classes. In this situation, the class of interest is often the one with the smaller number of examples, whereas the other class(es) represent(s) the counterpart of that concept and, in that manner, include(s) a high amount of data.

Standard classifier algorithms have a bias towards the majority class, since the rules that predicts the larger number of

ISBN: 978-989-95079-6-8

examples are positively weighted during the learning process in favour of the accuracy metric. Consequently, the instances that belongs to the minority class are misclassified more often than those belonging to the majority class [21]. Other important issue of this type of problem is the small disjuncts that can be found in the data-set [22] and the difficulty of most learning algorithms to detect those regions. Furthermore, the main handicap on imbalanced data-sets is the overlapping between the examples of the positive and the negative class [23]. These facts are depicted in Fig. 1.a and 1.b.

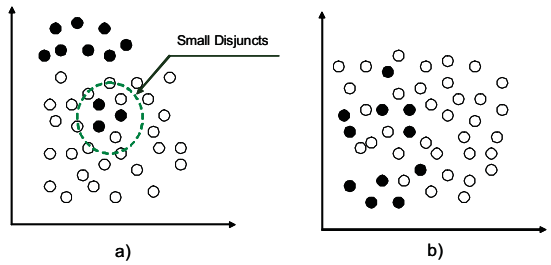


Figure 1: Example of the imbalance between classes: a) small disjuncts b) overlapping between classes

In this contribution we will focus on the data-sets with a higher degree of imbalance, using the imbalance ratio (IR) [24] as a threshold to categorize the different imbalanced scenarios. This measure is defined as the ratio of the number of instances of the majority class and the minority class.

In our previous work on this topic [25], we analysed the cooperation of some preprocessing methods with FRBCSs, showing a good behaviour for the oversampling methods, specially in the case of the SMOTE methodology [11]. According to this, we will employ in this contribution the SMOTE algorithm in order to deal with the problem of imbalanced data-sets.

In short, its main idea is to form new minority class examples by interpolating between several minority class examples that lie together. Thus, the overfitting problem is avoided and causes the decision boundaries for the minority class to spread further into the majority class space.

Regarding the empirical measure, instead of using accuracy, a more correct metric is considered. This is due to the fact that accuracy can lead to erroneous conclusions, since it does not take into account the proportion of examples for each class. Because of this, in this work we use the AUC metric [9], which can be defined as

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \tag{1}$$

where TP_{rate} is the percentage of positive cases correctly classified as belonging to the positive class and FP_{rate} is the percentage of negative cases misclassified as belonging to the positive class.

3 GP-COACH Algorithm

In this work we will make use of GP-COACH, a new FRBCS proposal [3]. The main features of this approach are listed below:

- It uses a context-free grammar that allows the learning of DNF fuzzy rules and the absence of some input features.
- It follows the *cooperative-competitive* approach, that is, it encodes a single rule per individual and the rule base (RB) is formed by the whole population. That makes necessary the use two different fitness functions in GP-COACH:
  - On the one hand, a local fitness function that evaluates the goodness of each one of the different rules in the population of individuals. From now on, we will refer it simply as *fitness function*.
  - On the other hand, a global fitness function that evaluates the goodness of a whole population of individuals (a rule set). From now on, we will refer it as *global fitness score*.

    This last fitness function has been introduced in GP-COACH in order to obtain the best rule set generated during the evolutionary process.
- It includes a mechanism to promote the diversity into the population, in order to avoid that all individuals converge to the same area of search space. Specifically, it uses the *Token Competition* diversity mechanism which makes rules compete among themselves during the evolutionary process, deleting irrelevant rules, and thus giving out a smaller number of rules that present a high-generalization capability.
- Finally, GP-COACH uses a two level hierarchical inference process because it learns rule sets containing two different types of rules: *primary rules*, which are strong and general rules generated by the genetic operators, and *secondary rules*, which are weaker and more specific rules, generated after the token competition procedure to increase the diversity in the population.

In the following subsections we will explain each one of these components and we will describe the way of working of this algorithm.

3.1 Context-free grammar for learning DNF fuzzy rules

GP-COACH learns DNF fuzzy rules:

$$R_k: \quad If\ X_1\ is\ \hat{A}_{k1}\ and\ \ldots\ and\ X_{n_v}\ is\ \hat{A}_{kn_v} \quad then\ Class\ is\ C_k\ with\ RW_k \tag{2}$$

where each input variable X_i takes as a value a set of linguistic terms or labels $\hat{A}_{ki} = \left\{ L_i^1\ or\ \ldots\ or\ L_i^{l_i} \right\}$ joined by a disjunctive operator, $(Class)$ is one of the class labels and RW_k is the rule weight [26]. We use triangular membership functions as antecedent fuzzy sets.

In GP-COACH, these DNF fuzzy rules are generated according to the production rules of a context-free grammar. In Table 1, an example of the grammar for a classification problem with two features (X_1, X_2), three linguistic labels per feature (*Low*, *Medium*, *High*) and three classes (C_1, C_2, C_3) is shown.

3.2 Evaluating an individual: Fitness function

Each one of the individuals in the population is evaluated according to a fitness function based on the estimation of:

ISBN: 978-989-95079-6-8

Table 1: Grammar example

$Start \longrightarrow [If],\ antec,\ [then],\ conseq,\ [.].$
$antec \longrightarrow descriptor1,\ [and],\ descriptor2.$
$descriptor1 \longrightarrow [any].$
$descriptor1 \longrightarrow [X_1\ is]\ label.$
$descriptor2 \longrightarrow [any].$
$descriptor2 \longrightarrow [X_2\ is]\ label.$
$label \longrightarrow \{member(?a,\ [L,\ M,\ H,\ L\ or\ M,\ L\ or\ H,$ $M\ or\ H,\ L\ or\ M\ or\ H])\},\ [?a].$
$conseq \longrightarrow [Class\ is]\ descriptorClass.$
$descriptorClass \longrightarrow \{member(?a,\ [C_1,\ C_2,\ C_3])\},\ [?a].$

- *Confidence*, which measures the accuracy of an individual, that is, the confidence of the consequent to be true if the antecedent is verified

$$Conf(R_k) = \frac{\mu_{tp}}{(\mu_{tp} + \mu_{fp})} \tag{3}$$

- *Support*, which measures the coverage of the knowledge represented in the individual

$$Supp(R_k) = \frac{\mu_{tp}}{N_{C_k}} \tag{4}$$

where μ_{tp} and μ_{fp} are the sums of the matching degrees for true and false positives, and N_{C_k} is the number of examples belonging to the class indicated in the consequent of the individual (R_k).

Both measures are combined to make up the fitness function in the following way

$$raw\_fitness = (\alpha * Conf) + ((1 - \alpha) * Supp) \tag{5}$$

where α parameter allow us to give more importance to any of these both measures.

3.3 Evaluating a population: Global fitness score

Global fitness score measure is defined as follows:

$$Global\_fitness = \begin{array}{l} (w_1 * AUC_{Tra}) + (w_2 * \overline{\#V}) + \\ (w_3 * \overline{\#C}) + (w_4 * \overline{\#R}) \end{array} \tag{6}$$

The global fitness score measure is formed by other four measures: a) AUC_{Tra}, the normalized value of the AUC metric for the training examples, b) *#V*, the normalized value of the number of variables per individual (rule) in the population, c) *#C*, the normalized value of the number of labels (or conditions) per individual, and d) *#R*, the normalized value of the number of rules in the population. We must remark that in the previous formula we employ the complement of these values.

Furthermore, we have included some weights (w_i) to give more importance to any of these four measures.

3.4 Token competition: Maintaining the diversity of the population

The idea of this mechanism is that each example in the training set can provide a resource called "token", for which all chromosomes in the population will compete to capture. If an individual (i.e. a rule) can match the example, it sets a flag to indicate that the token is seized preventing other weaker individuals to get this token.

The priority of receiving tokens is determined by the strength of the individuals. The individuals with higher fitness scores will exploit their niches by seizing as many tokens as they can. The other ones entering the same niches will have their strength decreased because they cannot compete with the stronger ones. This is done introducing a penalization in the fitness score of each individual. This penalization is based on the number of tokens that each individual has seized:

$$Penalized\_fitness = raw\_fitness * \frac{count}{ideal} \tag{7}$$

where *raw_fitness* is the fitness score obtained from the evaluation function, *count* is the number of tokens that the individual actually seized and *ideal* is the total number of tokens that it can take, which is equal to the number of examples that the individual matches.

As a result of the token competition, there exist individuals that cannot grab any token. These individuals are considered as irrelevant, and they can be eliminated from the population due to all of their examples are covered by other stronger individuals.

3.5 Secondary rules: Improving population diversity

Once the token competition mechanism has finished, it is possible that some training examples remain uncovered. The generation of new specific rules covering these examples improves the diversity in the population, and helps the evolutionary process to easily find stronger and more general rules covering these examples.

Therefore, GP-COACH learns rule sets having two different types of fuzzy rules: A core of strong and general rules (*primary rules*) that covers most of the examples, and a small set of weaker and more specific rules (*secondary rules*) that are only taken into account if there not exist any primary rule matching with some of the examples. This two level hierarchical inference process allows GP-COACH to obtain rule sets having a better interpretability-accuracy trade-off.

3.6 Genetic operators

GP-COACH makes use of four different genetic operators to generate new individuals during the evolutionary process:

1. *Crossover*: A part in the first parent is randomly selected and exchanged by another part, randomly selected, in the second one.
2. *Mutation*: It operates on label sets level. A variable in the rule is randomly chosen and then one of the next three different actions is carried out:
   (a) A new label is added to the label set.
   (b) A label is removed from the label set.
   (c) A label in the label set is exchanged by another one not included in it.
3. *Insertion*: It looks for all the variables in the rule and adds another different one that has not been included in it with a linguistic label set randomly chosen, although it must have at least one label and it must be different from the "any" set (see Table 1).
4. *Dropping Condition*: It randomly selects one variable in the rule and then turns it into "any". The label set associated with this variable is also removed.

We must remark that we generate only one child by using the genetic operators described above.

ISBN: 978-989-95079-6-8

3.7 Description of the Algorithm

GP-COACH algorithm begins creating a random initial population according to the rules in the context-free grammar. Each individual in this population is then evaluated. After that, the initial population is kept as the best evolved population and its global fitness score is calculated. Then, the initial population is copied to the current population and the evolutionary process begins:

1. An offspring population, with the same size than the current one, is created. Parents are selected by using the binary tournament selection mechanism and children are created by using one of the four genetic operators. The genetic operator selection is done in a probabilistic way. Specifically, the probabilities of the four genetic operator are added in a single measure $P = P_c + P_m + P_i + P_{dp}$ (where P_c, P_m, P_i and P_{dp} represent the crossover, mutation, insertion and dropping condition probabilities, respectively) and then a random value $u \in [0, P]$ is obtained to choose the genetic operator to be applied.

2. Once the offspring population is created, it is joined to the current population, creating a new population whose size is double the current population size. Individuals in this new population are sorted according to their fitness and Token Competition mechanism is applied. Secondaries rules are created if some examples remain uncovered.

3. Global fitness score measure is then calculated for this new population. We check whether this new fitness is better than the one stored for the best population, updating the best population and fitness if necessary. In any case, the new population is copied as the current population in order to be able to apply the evolutionary process again.

The evolutionary process ends when the stop condition is verified (i.e. number of evaluations). Then, the population kept as the best one is returned as a solution to the problem and GP-COACH finishes.

4 Experimental Study

In this study, our aim is to analyse the behaviour of the GP-COACH approach in the context of data-sets with high imbalance. We will compare the performance of this method against one state-of-the-art FRBCS algorithm, the Chi et al.'s approach, and the C4.5 decision tree, employing a large collection of imbalanced data-sets.

Specifically, we have considered twenty-two data-sets from UCI repository [10] with different IR, as shown in Table 2, where we denote the number of examples (#Ex.), number of attributes (#Atts.), class name of each class (minority and majority), class attribute distribution and IR. This table is in ascending order according to the IR. Data-sets with more than two classes have been modified by taking one against the others or by contrasting one class with another.

In order to reduce the effect of imbalance, we will employ the SMOTE preprocessing method [11] for all our experiments, considering only the 1-nearest neighbour to generate the synthetic samples, and balancing both classes to the 50% distribution.

In the remaining of this section, we will first present the experimental framework and all the parameters employed in this study and then we will show the results and all the statistical study for the GP-COACH approach.

4.1 Experimental Framework

To develop the different experiments we consider a *5-fold cross-validation model*, i.e., 5 random partitions of data with a 20%, and the combination of 4 of them (80%) as training and the remaining one as test. Since GP-COACH is a probabilistic method, we perform three executions per partition with different random seeds. For each data-set we consider the average results of the five partitions per three executions. Furthermore, Wilcoxon's Signed-Ranks Test [27] is used for statistical comparison of our empirical results. In all cases the level of confidence (α) will be set at 0.05.

The configuration for the FRBCSs approaches, GP-COACH and Chi et al.'s, is presented in Table 3. This parameter selection has been carried out according to the results achieved by GP-COACH and Chi et al.'s in our former studies in [3] and [25] respectively.

Table 3: Configuration for the FRBCS approaches

Parameter	GP-COACH	Chi et al.'s
Conjunction operator	Minimum T-norm	Product T-norm
Rule Weight	Certainty Factor	Penalized Certainty Factor
Fuzzy Reasoning Method	Additive Combination	Winning Rule
Number of Labels	5 Labels	5 Labels

The specific parameters setting for the GP-COACH algorithm, is listed below:

- Number of evaluations: 20000 evaluations.
- Initial Population Size: 200 individuals.
- α: 0.7.
- Crossover Probability P_c: 0.5.
- Mutation Probability P_m: 0.2.
- Dropping Probability P_{dp}: 0.15.
- Insertion Probability P_i: 0.15.
- Tournament Size: 2.
- w_1: 0.8, w_2 = w_3: 0.05, w_4: 0.1.

4.2 Analysis of the GP-COACH Behaviour in Data-sets with High Imbalance

The main objective in this study is to analyse the behaviour of the GP-COACH approach in the context of data-sets with high imbalance. According to this, Table 4 shows the results in performance (using the AUC metric) for GP-COACH and the algorithms employed for comparison, that is, Chi et al.'s learning method and C4.5.

We observe that the performance obtained by GP-COACH is higher than the one for Chi et al.'s and C4.5. This situation is represented statistically by means of a Wilcoxon test (Table 5) which shows a higher ranking in both cases for the GP-COACH algorithm.

Furthermore, when we compare the results for the fuzzy methods in each data-set, we observe that, as the IR increases, GP-COACH achieves a better performance than Chi et al.'s method. In this manner, we may state that GP-COACH is a very robust method according to the IR.

ISBN: 978-989-95079-6-8

Table 2: Summary Description for Imbalanced Data-Sets.

Data-set	#Ex.	#Atts.	Class (min.; maj.)	%Class(min., maj.)	IR
Data-sets with High Imbalance (IR higher than 9)					
Yeast2vs4	514	8	(cyt; me2)	(9.92, 90.08)	9.08
Yeast05679vs4	528	8	(me2; mit,me3,exc,vac,erl)	(9.66, 90.34)	9.35
Vowel0	988	13	(hid; remainder)	(9.01, 90.99)	10.10
Glass016vs2	192	9	(ve-win-float-proc; build-win-float-proc, build-win-non_float-proc,headlamps)	(8.89, 91.11)	10.29
Glass2	214	9	(Ve-win-float-proc; remainder)	(8.78, 91.22)	10.39
Ecoli4	336	7	(om; remainder)	(6.74, 93.26)	13.84
Yeast1vs7	459	8	(nuc; vac)	(6.72, 93.28)	13.87
Shuttle0vs4	1829	9	(Rad Flow; Bypass)	(6.72, 93.28)	13.87
Glass4	214	9	(containers; remainder)	(6.07, 93.93)	15.47
Page-blocks13vs2	472	10	(graphic; horiz.line,picture)	(5.93, 94.07)	15.85
Abalone9vs18	731	8	(18; 9)	(5.65, 94.25)	16.68
Glass016vs5	184	9	(tableware; build-win-float-proc, build-win-non_float-proc,headlamps)	(4.89, 95.11)	19.44
Shuttle2vs4	129	9	(Fpv Open; Bypass)	(4.65, 95.35)	20.5
Yeast1458vs7	693	8	(vac; nuc,me2,me3,pox)	(4.33, 95.67)	22.10
Glass5	214	9	(tableware; remainder)	(4.20, 95.80)	22.81
Yeast2vs8	482	8	(pox; cyt)	(4.15, 95.85)	23.10
Yeast4	1484	8	(me2; remainder)	(3.43, 96.57)	28.41
Yeast1289vs7	947	8	(vac; nuc,cyt,pox,erl)	(3.17, 96.83)	30.56
Yeast5	1484	8	(me1; remainder)	(2.96, 97.04)	32.78
Ecoli0137vs26	281	7	(pp,imL; cp,im,imU,imS)	(2.49, 97.51)	39.15
Yeast6	1484	8	(exc; remainder)	(2.49, 97.51)	39.15
Abalone19	4174	8	(19; remainder)	(0.77, 99.23)	128.87

Table 4: Detailed results in performance (AUC metric) for GP-COACH, Chi et al.'s learning method and C4.5

Dataset	GP-COACH		Chi et al.'s		C4.5	
	AUC_{Tr}	AUC_{Tst}	AUC_{Tr}	AUC_{Tst}	AUC_{Tr}	AUC_{Tst}
Yeast2vs4	83.46 ± 2.01	78.26 ± 5.92	90.51 ± 1.43	86.85 ± 6.68	98.14 ± 0.88	85.88 ± 8.78
Yeast05679vs4	82.76 ± 1.25	79.92 ± 5.54	87.97 ± 0.65	76.42 ± 6.17	95.26 ± 0.94	76.02 ± 9.36
Vowel0	96.74 ± 0.76	92.20 ± 5.80	99.64 ± 0.19	97.89 ± 1.83	99.67 ± 0.48	94.94 ± 4.95
Glass016vs2	74.57 ± 2.98	59.35 ± 17.29	76.16 ± 2.11	60.02 ± 8.41	97.16 ± 1.86	60.62 ± 12.66
Glass2	78.45 ± 2.82	66.11 ± 13.76	75.50 ± 1.80	52.06 ± 11.20	95.71 ± 1.51	54.24 ± 14.01
Ecoli4	97.74 ± 0.66	91.91 ± 6.04	98.14 ± 0.65	92.30 ± 8.13	97.69 ± 1.96	83.10 ± 9.90
Yeast1vs7	76.44 ± 2.17	64.03 ± 8.50	84.08 ± 2.14	65.24 ± 10.47	93.51 ± 2.20	70.03 ± 1.46
Shuttle0vs4	99.88 ± 0.08	99.99 ± 0.04	100.0 ± 0.00	98.72 ± 1.17	99.99 ± 0.02	99.97 ± 0.07
Glass4	95.94 ± 1.87	86.94 ± 14.19	98.88 ± 0.56	82.85 ± 10.20	98.44 ± 2.29	85.08 ± 9.35
Page-Blocks13vs4	98.03 ± 0.82	96.39 ± 4.31	98.71 ± 0.23	93.41 ± 8.53	99.75 ± 0.21	99.55 ± 0.47
Abalone9vs18	77.91 ± 2.31	74.27 ± 7.15	71.22 ± 3.09	67.44 ± 9.88	95.31 ± 4.44	62.15 ± 4.96
Glass016vs5	95.83 ± 1.45	94.29 ± 8.21	98.43 ± 0.41	84.86 ± 21.91	99.21 ± 0.47	81.29 ± 24.44
Shuttle2vs4	97.36 ± 3.34	97.43 ± 3.78	100.0 ± 0.00	88.38 ± 21.60	99.90 ± 0.23	99.17 ± 1.86
Yeast1458vs7	66.36 ± 1.61	58.21 ± 8.47	81.83 ± 1.70	59.32 ± 7.68	91.58 ± 2.78	53.67 ± 2.09
Glass5	98.11 ± 1.01	78.05 ± 24.24	98.78 ± 0.48	74.63 ± 20.52	99.76 ± 0.40	88.29 ± 13.31
Yeast2vs8	80.97 ± 3.28	78.77 ± 9.37	83.46 ± 1.68	80.66 ± 6.94	91.25 ± 1.84	80.66 ± 11.22
Yeast4	84.88 ± 1.42	81.95 ± 4.08	87.96 ± 1.54	83.25 ± 2.39	91.01 ± 2.64	70.04 ± 5.65
Yeast1289vs7	72.17 ± 2.75	66.26 ± 10.42	80.03 ± 2.33	70.27 ± 3.75	94.65 ± 1.13	68.32 ± 6.16
Yeast5	95.78 ± 0.68	93.53 ± 2.74	95.43 ± 0.54	93.72 ± 2.72	97.77 ± 1.45	92.33 ± 4.72
Ecoli0137vs26	90.13 ± 2.42	81.00 ± 18.30	96.85 ± 1.59	68.80 ± 22.87	96.78 ± 3.28	81.36 ± 21.68
Yeast6	90.43 ± 1.68	86.67 ± 7.92	89.60 ± 2.00	88.20 ± 8.55	92.42 ± 3.54	82.80 ± 12.77
Abalone19	74.00 ± 3.70	68.45 ± 8.38	77.19 ± 2.49	67.48 ± 10.77	85.44 ± 2.49	52.02 ± 4.41
Mean	86.72 ± 1.87	**80.64 ± 8.84**	89.56 ± 1.25	78.76 ± 9.65	95.93 ± 1.68	78.25 ± 8.38

Table 5: Wilcoxon test to compare GP-COACH with Chi et al.'s approach and C4.5 according to their performance. R^+ corresponds to GP-COACH and R^- to Chi or C4.5

Comparison	R^+	R^-	Hypothesis ($\alpha = 0.05$)	p-value
GP-COACH vs. Chi	161.0	92.0	Not Rejected	0.263
GP-COACH vs. C4.5	162.0	91.0	Not Rejected	0.249

Regarding interpretability of the obtained models, we must stress that GP-COACH is designed to obtain few DNF rules to describe the concept accurately. Table 6 shows the average number of rules for the three algorithms considered in this study, together with the associated standard deviation. The results from this table clearly shows that GP-COACH is the most interpretable model.

In Table 7 we show an example of an RB generated with the GP-COACH algorithm for the "shuttle2vs4" data-set. We can observe that the problem is described using few rules and only three of nine variables, enhancing the readability for the end-user.

5 Conclusions

In this contribution, we have studied the behaviour of GP-COACH in the context of data-sets with high imbalance.

Our results have shown the good performance achieved by this approach in contrast with the Chi et al.'s method, a well-known fuzzy rule learning algorithm, and C4.5, an algorithm of reference in the area of imbalanced data-sets.

Furthermore, we have compared the interpretability of the obtained models by means of the size of the rule set, concluding that GP-COACH employs a more compact RB in comparison with Chi et al.'s and C4.5 algorithms.

We must remark that GP-COACH is a good methodology for imbalanced data-sets, since it obtains very good classification results according to the AUC measure, by using a very interpretable model with few linguistic fuzzy rules.

Acknowledgment

This work had been supported by the Spanish Ministry of Science and Technology under Projects TIN2008-06681-C06-

01 and TIN2008-06681-C06-02.

Table 6: Detailed results table in interpretability (number of rules) for GP-COACH, Chi et al.'s learning method and C4.5

Data-set	GP-COACH	Chi et al.'s	C4.5
Yeast2vs4	6.60 ± 0.74	164.80 ± 3.83	20.40 ± 4.56
Yeast05679vs4	9.87 ± 2.17	191.20 ± 9.12	30.00 ± 6.63
Vowel0	5.60 ± 1.12	694.60 ± 8.32	11.80 ± 2.17
Glass016vs2	6.60 ± 2.23	65.20 ± 5.67	15.20 ± 1.48
Glass2	5.53 ± 1.96	73.20 ± 2.28	16.80 ± 2.49
Ecoli4	5.67 ± 0.90	116.80 ± 6.42	9.20 ± 2.17
Yeast1vs7	8.80 ± 2.01	156.80 ± 6.46	33.20 ± 6.10
Shuttle0vs4	2.60 ± 0.63	79.20 ± 7.56	2.80 ± 1.10
Glass4	6.33 ± 1.59	98.40 ± 10.06	7.40 ± 2.88
Page-Blocks13vs4	5.07 ± 0.96	164.00 ± 7.78	7.00 ± 1.22
Abalone9vs18	9.20 ± 2.24	93.60 ± 5.03	47.60 ± 1.14
Glass016vs5	5.13 ± 0.74	99.60 ± 8.62	10.00 ± 2.83
Shuttle2vs4	3.60 ± 0.99	33.00 ± 4.95	4.00 ± 0.00
Yeast1289vs7	8.93 ± 1.91	160.80 ± 2.68	58.40 ± 6.35
Glass5	4.93 ± 0.70	90.80 ± 6.87	7.00 ± 1.00
Yeast2vs8	5.73 ± 1.53	110.00 ± 2.65	14.60 ± 2.07
Yeast4	8.07 ± 1.22	197.20 ± 7.01	40.40 ± 3.78
Yeast1458vs7	7.27 ± 1.22	181.00 ± 8.69	47.80 ± 6.38
Yeast5	3.47 ± 0.52	206.60 ± 5.32	11.60 ± 2.07
Ecoli0137vs26	5.13 ± 0.74	168.60 ± 5.41	8.00 ± 1.41
Yeast6	5.73 ± 1.49	198.80 ± 6.14	21.60 ± 6.47
Abalone19	6.80 ± 1.15	180.20 ± 7.40	69.20 ± 3.70
Mean	**6.21 ± 1.31**	160.20 ± 6.29	22.45 ± 3.09

Table 7: Example of a DNF RB extracted using GP-COACH for the shuttle2vs4 data-set

Rule1: IF X_7 is $(L_4|L_5)$ THEN Class = negative with RW = 1
Rule2: IF X_1 is $(L_3|L_4|L_5)$ AND X_7 is $(L_1|L_2)$
THEN Class = positive with RW = 0.934089
Rule3: IF X_7 is $(L_2|L_4)$ AND X_8 is L_4
THEN Class = positive with RW = 0.495765
Rule4: IF X_7 is $(L_3|L_5)$ THEN Class = positive
with RW = 0.562457

References

[1] Q. Yang and X. Wu. 10 challenging problems in data mining research. *International Journal of Information Technology and Decision Making*, 5(4):597–604, 2006.

[2] H. Ishibuchi, T. Nakashima, and M. Nii. *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining*. Springer-Verlag, 2004.

[3] F.J. Berlanga, M.J. del Jesus, and F. Herrera. A novel genetic cooperative-competitive fuzzy rule based learning method using genetic programming for high dimensional problems. In *3rd International Workshop on Genetic and Evolving Fuzzy Systems (GEFS08)*, pages 101–106, 2008.

[4] Z. Chi, H. Yan, and T. Pham. *Fuzzy algorithms with applications to image processing and pattern recognition*. World Scientific, 1996.

[5] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo-California, 1993.

[6] A. Estabrooks, T. Jo, and N. Japkowicz. A multiple resampling method for learning from imbalanced data sets. *Computational Intelligence*, 20(1):18–36, 2004.

[7] C.T. Su and Y.H. Hsiao. An evaluation of the robustness of MTS for imbalanced data. *IEEE Transactions on Knowledge Data Engineering*, 19(10):1321–1332, 2007.

[8] J. Casillas, F. Herrera, R. Pérez, M.J. del Jesus, and Pedro Villar. Special issue on genetic fuzzy systems and the interpretability-accuracy trade-off - editorial. *International Journal of Approximate Reasoning*, 44(1):1–3, 2007.

[9] J. Huang and C. X. Ling. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3):299–310, 2005.

[10] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007. University of California, Irvine, School of Information and Computer Sciences. URL `http://www.ics.uci.edu/~mlearn/MLRepository.html`

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligent Research*, 16:321–357, 2002.

[12] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

[13] S. García and F. Herrera An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. *Journal of Machine Learning Research*, 9:2607–2624, 2008.

[14] S. García, A. Fernández, J. Luengo and F. Herrera A Study of Statistical Techniques and Performance Measures for Genetics-Based Machine Learning: Accuracy and Interpretability. *Soft Computing*, 13:10 959–977, 2009.

[15] N. V. Chawla, N. Japkowicz, and A. Kolcz. Editorial: special issue on learning from imbalanced data sets. *SIGKDD Explorations*, 6(1):1–6, 2004.

[16] R. Barandela, J.S. Sánchez, V. García, and E. Rangel. Strategies for learning in class imbalance problems. *Pattern Recognition*, 36(3):849–851, 2003.

[17] M.C. Chen, L.S. Chen, C.C. Hsu, and W.R. Zeng. An information granulation based data mining approach for classifying imbalanced data. *Information Sciences*, 178(16):3214–3227, 2008.

[18] T. Fawcett and F. J. Provost. Adaptive fraud detection. *Data Mining and Knowledge Discovery*, 1(3):291–316, 1997.

[19] Y. M. Huang, C. M. Hung, and H. C. Jiau. Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem. *Nonlinear Analysis: Real World Applications*, 7(4):720–747, 2006.

[20] M.A. Mazurowski, P.A. Habas, J.M. Zurada, J.Y. Lo, J.A. Baker, and G.D. Tourassi. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Networks*, 21(2-3):427–436, 2008.

[21] G. M. Weiss. Mining with rarity: a unifying framework. *SIGKDD Explorations*, 6(1):7–19, 2004.

[22] G.M. Weiss and F. Provost. Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research*, 19:315–354, 2003.

[23] V. García, R.A. Mollineda, and J. S. Sánchez. On the k-NN performance in a challenging scenario of imbalance and overlapping. *Pattern Analysis Applications*, 11(3–4):269–280, 2008.

[24] A. Orriols-Puig and E. Bernadó-Mansilla. Evolutionary rule-based systems for imbalanced datasets. *Soft Computing*, 13(3):213–225, 2009.

[25] A. Fernández, S. García, M.J. del Jesus, and F. Herrera. A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets and Systems*, 159(18):2378–2398, 2008.

[26] H. Ishibuchi and T. Yamamoto, Rule Weight Specification in Fuzzy Rule-Based Classification Systems. *IEEE Transactions on Fuzzy Systems*, 13(4):428-435, 2005.

[27] D. Sheskin. *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC, second edition, 2006.

ISBN: 978-989-95079-6-8

New Fuzzy Color Clustering Algorithm Based on *hsl* Similarity

Vasile Pătraşcu

Department of Informatics Technology, Tarom Company
Bucharest, Romania
Email: patrascu.v@gmail.com

Abstract—*In this paper, one presents a fuzzy color clustering algorithm that is based on a new measure of similarity. This new measure of color similarity is defined on a perceptual color system called hsl.*

Keywords— Color similarity, fuzzy color clustering, *hsl* perceptual color system.

1 Introduction

The color clustering algorithms play an important role in image analysis. The obtained results are very dependent on the coordinate system and similarity/dissimilarity functions used for color separation [1], [4], [5] and [6]. In this paper, one proposes the use of a perceptual system hsl. In the framework of this system, a new measure is proposed for the color similarity. In the following sections, the paper is thus organized: section 2 presents the system hsl for color representation based on new measures of luminosity L and saturation S; section 3 presents distances and similarities for each component of the hsl system; section 4 presents the new color similarity; section 5 presents the color clustering algorithm based on the proposed similarity; section 6 presents the experimental results and conclusions are in section 7.

2 The perceptual color system *hsl*

The most part of the color images are represented by the RGB color system. Image analysis is done in the perceptual coordinate systems. One of them is the hSL system where h is the hue and an angular value; S is the saturation and lastly, L is the luminosity. In this paper, we will use the following formulae for the definition of the hSL components:

$$\begin{cases} L = \dfrac{M}{1+M-m} \\ S = \dfrac{2\cdot(M-m)}{1+|M-0.5|+|m-0.5|} \\ h = \text{atan2}\left(\dfrac{B-G}{\sqrt{2}}, \dfrac{2R-B-G}{\sqrt{6}}\right) \end{cases} \tag{1}$$

where:

$$M = \max(R,G,B)$$
$$m = \min(R,G,B)$$

We suppose that $R,G,B \in [0,1]$. According to the considered formulae for the definition of the components hSL, it results that $L,S \in [0,1]$ and $h \in (-\pi,\pi]$. In order to have a unitary consideration for the three color components, hSL, we will scale the luminosity L and the saturation S from $[0,1]$ to $\left[0,\dfrac{\pi}{2}\right]$ by formulae:

$$\begin{cases} l = \dfrac{\pi}{2}\cdot L \\ s = \dfrac{\pi}{2}\cdot S \end{cases} \tag{2}$$

Thus, any color q from RGB space will have the following representation in the hsl space:

$$\begin{cases} l = \dfrac{\pi}{2}\cdot\dfrac{M}{1+M-m} \\ s = \pi\cdot\dfrac{M-m}{1+|M-0.5|+|m-0.5|} \\ h = \text{atan2}\left(\dfrac{B-G}{\sqrt{2}}, \dfrac{2R-B-G}{\sqrt{6}}\right) \end{cases} \tag{3}$$

3 Distances and similarities for the values *hsl*

We know that for two angular values α,β a good measure for distance is based on the *sin* function, namely:

$$d(\alpha,\beta) = \left|\sin\left(\frac{\alpha-\beta}{2}\right)\right| \tag{4}$$

Thus, for the hue, the distance is given by the function:

$$d_h(h_1,h_2) = \left|\sin\left(\frac{h_1-h_2}{2}\right)\right| \tag{5}$$

Using the formula (4) for the saturation s and luminosity l we obtain their values in the interval $\left[0,\dfrac{1}{\sqrt{2}}\right]$, because $l,s \in \left[0,\dfrac{\pi}{2}\right]$. Consequently, we will multiply the formula (4) with the factor $\sqrt{2}$ and thus we will obtain the values in

ISBN: 978-989-95079-6-8

the whole interval $[0,1]$. In this way, the distances for luminosity and saturation will be defined by:

$$d_l(l_1,l_2)=\sqrt{2}\cdot\left|\sin\left(\frac{l_1-l_2}{2}\right)\right| \tag{6}$$

$$d_s(s_1,s_2)=\sqrt{2}\cdot\left|\sin\left(\frac{s_1-s_2}{2}\right)\right| \tag{7}$$

In this paper, we will take into account that the square of distance is a good measure of dissimilarity and the negation of dissimilarity is a good measure of similarity. In other words, if d is the distance and σ is the similarity, then the following relation exists between them:

$$\sigma=1-d^2 \tag{8}$$

Having the distance definitions, we will then define the similarity function for luminosity and saturation,

$$\begin{cases}\sigma_l(l_1,l_2)=1-d_l^2(l_1,l_2)=\cos(l_1-l_2)\\ \sigma_s(s_1,s_2)=1-d_s^2(s_1,s_2)=\cos(s_1-s_2)\end{cases} \tag{9}$$

and next for the hue:

$$\sigma_h(h_1,h_2)=1-d_h^2(h_1,h_2)=\cos^2\left(\frac{h_1-h_2}{2}\right) \tag{10}$$

4 The *hsl* color similarity

We will initially define the chromaticity c and the achromaticity a for a color having as saturation the value s:

$$\begin{cases}c=\sin(s)\\ a=\cos(s)\end{cases} \tag{11}$$

The following relation between these two defined parameters is obvious:

$$a^2+c^2=1 \tag{12}$$

We will consider two colors q_1,q_2 defined by parameters $(h_1,s_1,l_1),(h_2,s_2,l_2)$. We will add to these parameters the chromaticity coefficients c_1,c_2 and achromaticity coefficients a_1,a_2, computed by formulae (11). Now, we define the similarity between two colors by formula:

$$\sigma(q_1,q_2)=c_1c_2\cdot\sigma_h(h_1,h_2)+a_1a_2\cdot\sigma_l(l_1,l_2) \tag{13}$$

Seeing the formula (13), one remarks that this new color similarity measure is a linear and adaptive combination, between the hues similarity and luminosities similarity. Thus for two chromatic colors, the main component is given by the hues similarity and for two achromatic colors the main component is given by the luminosities similarity. We can see that, when a color is chromatic and the other is achromatic, the similarity has small values. This is nothing else than the two considered colors are quite non-similar. Also, we can see that in formula (13), the saturations similarity does not appear directly, but it exists somewhere in background. The formula (13) has the following equivalent form:

$$\begin{aligned}\sigma(q_1,q_2)=\sin(s_1)\cdot\sin(s_2)\cdot\cos^2\left(\frac{h_1-h_2}{2}\right)+\\ +\cos(s_1)\cdot\cos(s_2)\cdot\cos(l_1-l_2)\end{aligned} \tag{14}$$

Taking into account that:

$$\begin{cases}\cos^2\left(\frac{h_1-h_2}{2}\right)\le 1\\ \cos(l_1-l_2)\le 1\end{cases} \tag{15}$$

it results from (14) this inequality:

$$\sigma(q_1,q_2)\le\sin(s_1)\cdot\sin(s_2)+\cos(s_1)\cdot\cos(s_2)$$

namely

$$\sigma(q_1,q_2)\le\cos(s_1-s_2)$$

and finally we obtain that:

$$\sigma(q_1,q_2)\le\sigma_s(s_1,s_2) \tag{16}$$

The formula (16) shows that if the saturations of the two colors have a low similarity then the two colors have a low similarity too and if the two colors are strongly similar then their saturations are strongly similar as well. Also, from (15) and (16) it results the following implications:

$$\sigma(q_1,q_2)=1\ \Rightarrow\ \begin{cases}\sigma_h(h_1,h_2)=1\\ \sigma_l(l_1,l_2)=1\\ \sigma_s(s_1,s_2)=1\end{cases}\ \Rightarrow\ \begin{cases}h_1=h_2\\ l_1=l_2\\ s_1=s_2\end{cases}$$

Using the formula (8) we can define the distance between two colors in the *hsl* space as:

$$d(q_1,q_2)=\sqrt{1-\sigma(q_1,q_2)} \tag{17}$$

We can state that formula (17) defines a metric because it verifies the following three metric properties:

- $d(q_1,q_2)=0\ \Leftrightarrow\ q_1=q_2$
- $d(q_1,q_2)=d(q_2,q_1)$
- $d(q_1,q_2)+d(q_2,q_3)\ge d(q_1,q_3)$

A detailed proof of these three properties is not a subject of this paper and thus not considered here.

ISBN: 978-989-95079-6-8

5 The fuzzy color clustering algorithm

We consider n colors $q_1, q_2, ..., q_n$ that must be separated into k clusters. Each cluster i is characterized by the membership coefficients $w_{i1}, w_{i2}, ..., w_{in}$ for the considered n colors and the cluster center defined by the color $\bar{q}_i = (\bar{h}_i, \bar{s}_i, \bar{l}_i)$. For color clustering we will construct an algorithm that is similar to the fuzzy c-means algorithm [1]. We will consider the following objective function:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} w_{ij}^{\alpha} \cdot \sigma(\bar{q}_i, q_j) \tag{18}$$

where α is a fuzzification-defuzzification parameter and also, $\alpha \in (0,1)$. If α approaches 1, then the fuzzy algorithm approaches a crisp one. The objective function J (18) must be maximized. In order to have fuzzy partitions, we must add the following conditions, for $j = 1,2,...,n$:

$$w_{1j} + w_{2j} + ... + w_{kj} = 1 \tag{19}$$

In this case, considering for the condition (19) the Lagrange multipliers $\lambda_1, \lambda_2, ..., \lambda_n$, the objective function (18) becomes:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} w_{ij}^{\alpha} \cdot \sigma(\bar{q}_i, q_j) + \sum_{j=1}^{n} \lambda_j \cdot \left(\sum_{i=1}^{k} w_{ij} - 1 \right) \tag{20}$$

The maximum value of the objective function J (20) results from the following conditions:

- $\forall i \in [1,k], \forall j \in [1,n]$,

$$\frac{\partial J}{\partial w_{ij}} = 0 \tag{21}$$

It results:

$$w_{ij} = \frac{\left(\sigma(\bar{q}_i, q_j)\right)^{\frac{1}{1-\alpha}}}{\sum_{m=1}^{k} \left(\sigma(\bar{q}_m, q_j)\right)^{\frac{1}{1-\alpha}}} \tag{22}$$

- $\forall i \in [1,k]$,

$$\frac{\partial J}{\partial \bar{h}_i} = 0 \tag{23}$$

It results:

$$\bar{h}_i = \operatorname{atan2}\left(\frac{\sum_{j=1}^{n} u_{ij} \sin(h_j)}{\sum_{j=1}^{n} u_{ij}}, \frac{\sum_{j=1}^{n} u_{ij} \cos(h_j)}{\sum_{j=1}^{n} u_{ij}} \right) \tag{24}$$

where

$$u_{ij} = w_{ij}^{\alpha} \cdot c_j \tag{25}$$

- $\forall i \in [1,k]$,

$$\frac{\partial J}{\partial \bar{l}_i} = 0 \tag{26}$$

It results:

$$\bar{l}_i = \operatorname{atan2}\left(\frac{\sum_{j=1}^{n} v_{ij} \sin(l_j)}{\sum_{j=1}^{n} v_{ij}}, \frac{\sum_{j=1}^{n} v_{ij} \cos(l_j)}{\sum_{j=1}^{n} v_{ij}} \right) \tag{27}$$

where

$$v_{ij} = w_{ij}^{\alpha} \cdot a_j \tag{28}$$

- $\forall i \in [1,k]$,

$$\frac{\partial J}{\partial \bar{s}_i} = 0 \tag{29}$$

It results:

$$\bar{s}_i = \operatorname{atan2}\left(\frac{\sum_{j=1}^{n} x_{ij} \sin(s_j)}{\sum_{j=1}^{n} w_{ij}^{\alpha}}, \frac{\sum_{j=1}^{n} y_{ij} \cos(s_j)}{\sum_{j=1}^{n} w_{ij}^{\alpha}} \right) \tag{30}$$

where

$$\begin{cases} x_{ij} = w_{ij}^{\alpha} \cdot \sigma_h(\bar{h}_i, h_j) \\ y_{ij} = w_{ij}^{\alpha} \cdot \sigma_l(\bar{l}_i, l_j) \end{cases} \tag{31}$$

6 Experimental results

The proposed algorithm was applied to the image „flower” (Fig. 1) and the clustered image can be seen in Fig. 2. For comparison, the fuzzy c-means algorithm [1], [3], [8] was applied to image „flower”, in the coordinate systems RGB [3], [9], Lab [7], Luv [3], [9], hSL [9], I1I2I3 [2], [3] and the results can be seen in Figs. 3, 4, 5, 6 and 7.

Figure 1: The image „flower”.

ISBN: 978-989-95079-6-8

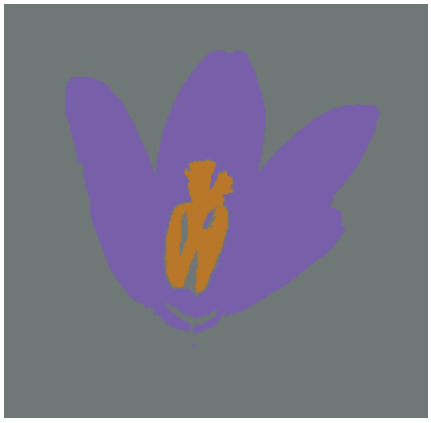


Figure 2: The image „flower” clustered with the proposed algorithm and *hsl*.

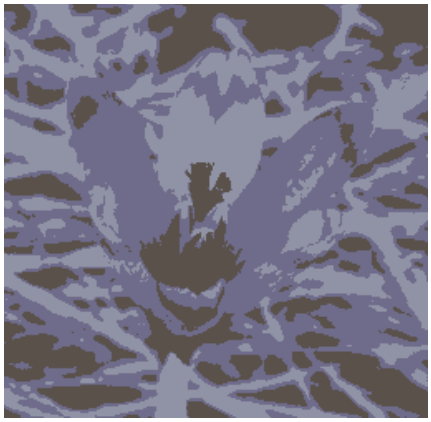


Figure 3: The image „flower” clustered with FCM algorithm and *RGB*.

Figure 4: The image „flower” clustered with FCM algorithm and *Lab*.

Figure 5: The image „flower” clustered with FCM algorithm and *Luv*.

Figure 6: The image „flower” clustered with FCM algorithm and *hSL*.

Figure 7: The image „flower” clustered with FCM algorithm and *I1I2I3*.

7 Conclusions

In this paper, a new measure for color similarity was defined. This new measure was defined in a new perceptual system *hsl* . Using this measure, an algorithm for color clustering was constructed.

In the section dedicated to the experimental results, one can see the advantage of using this new color similarity measure.

References

[1] J.C. Bezdek. Pattern Recognition with Fuzzy Objective Functions. New York: Plenum Press, 1981.

[2] Y.I. Ohta, T. Kanade and T. Sakai. Color information for region segmentation, *Computer and Image Processing*, vol. 13, pp.222-241, 1980.

[3] Y.W. Lim and S.U. Lee. On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques, *Pattern Recognition*, vol. 23, no.9 , pp.935-952, 1990.

ISBN: 978-989-95079-6-8

[4] V. Pătraşcu. A generalization of fuzzy c-means algorithm using a new dissimilarity function, The 11th International Fuzzy Systems Association World Congres, IFSA 2005, *Fuzzy Logic, Soft Computing and Computational Intelligence*, Vol II, pp. 591-596, Beijing China July, 28-31,2005.

[5] V. Pătraşcu. A Generalization of Gustafson-Kessel Algorithm Using a New Constraint Parameter*, In Proceedings of the 4th Conference of the European Society for Fuzzy Logic and Technology and 11th Rencontres Francophones sur la Logique Floue et ses Applications*, EUSFLAT-LFA 2005, pp.1250-1255, Barcelona, Spain 7-9 September, 2005.

[6] V. Pătraşcu. Fuzzy Image Segmentation Based on Triangular Function and Its n-dimensional Extension, in the volume *Soft Computing in Image Processing. Recent Advances, Series: Studies in Fuzziness and Soft Computing*, pp 187-208, Vol. 210 , (Eds.) M. Nachtegael, D. Van der Weken, E. E. Kerre, W. Philips, ISBN: 3-540-38232-1, Springer-Verlag, 2007.

[7] S. Tominga. A colour classification method for color images using a uniform color space, *Proc. 10th Int. Conf. On Pattern Recognition*, vol. I, Atlantic City, new Jersey, pp. 803-807, 16-21 June 1990.

[8] M. Trivedi and J.C. Bezdek. Low-level segmentation of aerial images with fuzzy clustering, *IEEE Trans. On Systems, Man, and Cybernetics*, vol. 16, no. 4, pp.589-598, 1986.

[9] S.E. Umbaugh, R.H. Moss, W.V. Stoecker and G.A. Hance. Automatic colour segmentation algorithms with application to skin tumor feature identification, IEEE Engineering in Medcine and Biology, vol. 12, no. 3, pp. 75-82, 1993.

ISBN: 978-989-95079-6-8

Multi-Dimensional Fuzzy Transforms for Attribute Dependencies

Ferdinando Di Martino[1,2] Vincenzo Loia[2] Salvatore Sessa[1]

[1]Dipartimento di Costruzioni e Metodi Matematici in Architettura, Università degli Studi di Napoli Federico II, 80134 Napoli, Italy
[2] Dipartimento di Matematica e Informatica, Università degli Studi di Salerno, 84084 Salerno, Italy
Email: fdimarti@unina.it, loia@unisa.it, sessa@unina.it

Abstract *— We explore attribute dependencies in the datasets by using direct and inverse fuzzy transforms. Our algorithm optimizes the fuzzy partitions of the universe of the attributes and moreover establishes if the set of the data points is sufficiently dense with respect to the chosen partitions: two specific regression indexes measure the reliability of our model. The known "El Nino" dataset is the basis of our experiments, whose results are consistent with the regression analysis made with the same data.*

Keywords — attribute dependence, fuzzy transform, index of determinacy, regression.

1 Introduction

Regression is a well known statistical supervised technique used in data mining (cfr. [5, 6, 7, 8, 9]. The dependency between numerical attributes is studied via an equation of regression: the dependent attribute is modeled as a function of the independent attributes or predictors. We present a fast regression algorithm in which we use the fuzzy transforms (for short, F-transforms) for exploring numerical attribute dependencies in datasets. A *direct fuzzy transform* [1, 11, 12, 13, 14] gives a correspondence between a set of continuous functions defined on the interval [a, b] and a set of n-dimensional vectors defined suitably. An *inverse fuzzy transform* gives the converse correspondence, that is an n-dimensional vector is transformed into a continuous function which approximates the original function up to a small quantity ε. The method based on the F-transforms has been used in other topics like image processing [2, 3, 4, 12], geology [11].
In [10] the attribute dependencies in data analysis are established via F-transforms: the dependency of an attribute X_z from the attributes $X_1,\ldots,X_k$ is made by setting $X_z = H(X_1,\ldots,X_k)$, where H: $[a_1,b_1] \times\ldots\times [a_k,b_k] \rightarrow [a_z,b_z]$ is a continuous function and $[a_i,b_i]$ is the domain of X_i, i= 1,…,k. The approximating function $H_{F,n}$ is obtained via the *discrete F-transforms* and the difference between H and $H_{F,n}$ at the their common points is evaluated by means of the index of determinacy. We devote particular attention to the solution of the following problems:

a) to find the best fuzzy partition of the attribute domains in the construction of the direct F-transform. We observe that the index of determinacy calculated with coarse grained partitions is less than that one calculated with finer partitions;

b) our method analyzes if the set of the assigned data points is sufficiently dense w. r. t. the chosen partitions.
Our algorithm makes a fuzzy partition of $[a_i,b_i]$, i=1,…,k, in n fuzzy sets and establishes a threshold value for the indexes of determinacy after some training tests. The attribute dependency is found if the index overcomes the threshold value, otherwise we pass to a partition of $[a_i,b_i]$ in n+1 fuzzy sets by checking additionally if the set of the assigned data points is sufficiently dense w.r.t. the chosen partition. Of course if this condition is violated, then the process is stopped.
In Section 2 we give the definitions of F-transforms in k (≥2) variables. In Section 3 we present our process, in Section 4 a sample simulation shows how to find the threshold value for the indexes of determinacy. The known dataset "El Nino" is the basis of our tests contained in Section 5. Section 6 concludes this paper.

2 Multi-dimensional F-transforms

In accordance to [12], let $n \geq 2$ and $x_1, x_2, \ldots, x_n$ be points of [a,b], called nodes, such that $x_1 = a < x_2 <\ldots< x_n = b$. The fuzzy sets $A_1,\ldots,A_n : [a,b] \rightarrow [0,1]$, called basic functions, form a *fuzzy partition* of [a,b] if the following hold:

(1) $A_i(x_i) =1$ for every i =1,2,…,n;

(2) $A_i(x) = 0$ if $x \notin (x_{i-1},x_{i+1})$ for every i = 2,…,n;

(3) $A_i(x)$ is a continuous function on [a,b] for every i = 1,2,…,n;

(4) $A_i(x)$ strictly increases on $[x_{i-1}, x_i]$ for every i = 2, …, n and strictly decreases on $[x_i,x_{i+1}]$ for every i = 1,…, n-1;

(5) $\sum_{i=1}^{n} A_i(x)=1$ for every $x \in [a,b]$.

The partition $\{A_1(x),\ldots,A_n(x)\}$ is said *uniform* if moreover the following hold:
(6) $n \geq 3$ and the nodes are equidistant, that is $x_i = a+ h\cdot(i-1)$ for every i = 1, 2, …, n where h = (b-a)/(n-1);

(7) $A_i(x_i – x)=A_i(x_i + x)$ for every $x \in [0,h]$ and i = 2,…, n-1;

ISBN: 978-989-95079-6-8

(8) $A_{i+1}(x) = A_i(x-h)$ for every $x \in [x_i, x_{i+1}]$ and i= 1,…, n-1.

By limiting ourselves to the discrete case (cfr. [12] in the case of a continuous function), let f be a real function defined on assigned points $p_1,...,p_m$ of [a,b]. Let P be the set of these points and suppose that it is *sufficiently dense* w. r. t. the fixed partition $\{A_1, A_2, \ldots, A_n\}$, that is for every i = 1,…,n there exists an index $j \in \{1,\ldots,m\}$ such that $A_i(p_j) > 0$. Thus we define the numerical vector $[F_1, F_2, ..., F_n]$ as the *direct F-transform* of f w.r.t. the basic functions $\{A_1, A_2, \ldots, A_n\}$, where each F_i is given by

$$F_i = \frac{\sum_{j=1}^{m} f(p_j) A_i(p_j)}{\sum_{j=1}^{m} A_i(p_j)} \quad (1)$$

for every i=1,…,n. Then we define the *inverse F-transform* of f w. r. t. $\{A_1, A_2, \ldots, A_n\}$ by setting for every j = 1,…,m:

$$f_{F,n}(p_j) = \sum_{i=1}^{n} F_i A_i(p_j) \quad (2)$$

Then the following theorem [12] holds:

Theorem 1. Let f(x) be a real function assigned on a set $P = \{p_1,..., p_m\} \subseteq [a,b]$. Then for every $\varepsilon > 0$, there exist an integer $n(\varepsilon)$ and a related fuzzy partition $\{A_1, A_2, \ldots, A_{n(\varepsilon)}\}$ of [a,b] such that P is sufficiently dense w. r. t. $\{A_1, A_2, \ldots, A_{n(\varepsilon)}\}$ and for every $p_j \in [a, b]$, j = 1,…,m, the following inequality holds:

$$|f(p_j) - f_{F,n(\varepsilon)}(p_j)| < \varepsilon \quad (3)$$

As in [15], we extend these definitions to the multi-dimensional case. Let k be given intervals $[a_i,b_i]$ (i=1,…,k) and $[a_1,b_1]\times[a_2,b_2]\times\ldots\times[a_k,b_k]$ be our universe of discourse. Let $x_{11}, x_{12},...., x_{1n_1} \in [a_1,b_1], \ldots, x_{k1}, x_{k2},...., x_{kn_k} \in [a_k,b_k]$ be $n_1+\ldots+n_k$ assigned points, called nodes, such that $x_{i1} = a_i < x_{i2} < \ldots < x_{in_i} = b_i$ for every i = 1,…,k. Let $\{A_{i1}, A_{i2},...., A_{in_i}\}$ be a fuzzy partition of $[a_i,b_i]$ for every i = 1,…,k and $f(x_1,x_{2,\ldots,}x_k)$ be a real function by assuming determined values in m points $p_j = (p_{j1}, p_{j2},\ldots,p_{jk}) \in [a_1,b_1]\times[a_2,b_2]\times\ldots\times[a_k,b_k]$ for every j=1,…,m. We say that the set $P=\{(p_{11}, p_{12}, \ldots, p_{1k}), (p_{21}, p_{22}, \ldots, p_{2k}),\ldots,(p_{m1}, p_{m2}, \ldots,p_{mk})\}$ is *sufficiently dense* w. r. t. the chosen partitions $\{A_{11}, A_{12},..., A_{1n_1}\},\ldots,\{A_{k1}, A_{k2},..., A_{kn_k}\}$ if for each k-tuple $\{h_1,\ldots,h_k\} \in \{1,\ldots,n_1\}\times\ldots\times\{1,\ldots,n_k\}$ there exists a point $p_j = (p_{j1},p_{j2},\ldots,p_{jk}) \in P$, $j \in \{1,\ldots,m\}$, such that $A_{1h_1}(p_{j1}) \cdot A_{2h_2}(p_{j2}) \cdot \ldots \cdot A_{kh_K}(p_{jk}) > 0$. Then we can define the $(h_1,h_2,\ldots,h_k)$-th component $F_{h_1h_2\ldots h_K}$ of the *direct F-transform* of f w.r.t. the basic functions $\{A_{11}, A_{12},..., A_{1n_1}\}, \ldots, \{A_{k1}, A_{k2},..., A_{kn_k}\}$ as

$$F_{h_1h_2\ldots h_K} = \frac{\sum_{j=1}^{m} f(p_{j1}, p_{j2},\ldots p_{jk}) \cdot A_{1h_1}(p_{j1}) \cdot A_{2h_2}(p_{j2}) \cdot \ldots \cdot A_{kh_K}(p_{jk})}{\sum_{j=1}^{m} A_{1h_1}(p_{j1}) \cdot A_{2h_2}(p_{j2}) \cdot \ldots \cdot A_{kh_K}(p_{jk})} \quad (4)$$

The *inverse F-transform* of f w.r.t. the basic functions $\{A_{11}, A_{12},..., A_{1n_1}\}, \ldots, \{A_{k1}, A_{k2},..., A_{kn_K}\}$ is the following function defined for each point $p_j=(p_{j1}, p_{j2},\ldots,p_{jk}) \in [a_1,b_1]\times\ldots\times[a_k,b_k]$ as

$$f^F_{n_1n_2\ldots n_K}(p_{j1}, p_{j2},..., p_{j,k}) = \sum_{h_1=1}^{n_1}\sum_{h_2=1}^{n_2}\ldots\sum_{h_K=1}^{n_k} F_{h_1h_2\ldots h_K} \cdot A_{1h_1}(p_{j1}) \cdot \ldots \cdot A_{kh_K}(p_{jk}) \quad (5)$$

for every j=1,…,m. The following extension of Theorem 1 holds:

Theorem 2. Let $f(x_1,x_{2,\ldots,}x_K)$ be a function assigned on the set of points $P = \{(p_{11},p_{12}, \ldots,p_{1k}) ,(p_{21}, p_{22}, \ldots, p_{2k}),\ldots,(p_{m1}, p_{m2}, \ldots,p_{mk})\} \subseteq [a_1,b_1]\times[a_2,b_2]\times\ldots\times[a_k,b_k]$. Then for every $\varepsilon > 0$, there exist k integers $n_1(\varepsilon),\ldots, n_k(\varepsilon)$ and fuzzy partitions $\{A_{11}, A_{12},..., A_{1n_1(\varepsilon)}\},\ldots,\{A_{k1}, A_{k2},..., A_{kn_k(\varepsilon)}\}$ such that the set P is sufficiently dense w.r.t. the above fuzzy partitions and for every $p_j = (p_{j1}, p_{j2},\ldots,p_{jk}) \in P$, j=1,…,m, the following inequality holds:

$$\left| f(p_{j1}, p_{j2},..., p_{jk}) - f^F_{n_1(\varepsilon)n_2(\varepsilon)\cdots n_k(\varepsilon)}(p_{j1}, p_{j2},..., p_{jk}) \right| < \varepsilon \quad (6)$$

The proof follows the same lines of the similar Theorem 1.

3 Attribute dependencies

We modify slightly the algorithm of [10], where the data are represented in the following matrix:

	X_1	.	X_i	.	X_r
O_1	p_{11}	.	p_{1i}	.	p_{1r}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
O_j	p_{j1}	.	p_{ji}	.	p_{jr}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
O_m	p_{m1}	.	p_{mi}	.	p_{mr}

being $X_1,\ldots,X_i,\ldots,X_r$ the attributes, $O_1, \ldots,O_j, \ldots,O_m$ the objects and p_{ji} the value of the attribute X_i. By setting $a_i = \min\{p_{1i},\ldots,p_{mi}\}$ and $b_i = \max\{p_{1i},\ldots,p_{mi}\}$, we define the interval $[a_i,b_i]$ of the given values of the attribute X_i. We can consider the attribute X_z, $z \in \{1,\ldots, r\}$, dependent from k independent attributes $X_1, \ldots, X_k$, (with $k \le r < m$ and $z \notin \{1,\ldots,k\}$) via a function H defined as

ISBN: 978-989-95079-6-8

$$X_z = H(X_1,...,X_k) \tag{7}$$

We use the F-transforms method in the following process:

1) We assume $n_1 = n_2 = \ldots = n_k = n$ without loss of generality and we refer to the model (7). After normalization in [0,1] of the values of the independent and dependent attributes, we start with n = 3.

2) A uniform fuzzy partition $\{ A_{i1}, A_{i2}, ..., A_{in_i} \}$ is assigned in each interval $[a_i,b_i]$ by setting for every i = 1,…, k and j = 2,…, k-1:

$$A_{i1}(x) = \begin{cases} 0.5 \cdot (1 + \cos\frac{\pi}{h_i}(x - x_{i1})) & \text{if } x \in [x_{i1}, x_{i2}] \\ 0 & \text{otherwise} \end{cases}$$

$$A_{ij}(x) = \begin{cases} 0.5 \cdot (1 + \cos\frac{\pi}{h_i}(x - x_{ij})) & \text{if } x \in [x_{i(j-1)}, x_{i(j+1)}] \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$A_{in}(x) = \begin{cases} 0.5 \cdot (1 + \cos\frac{\pi}{h_i}(x - x_{in})) & \text{if } x \in [x_{i(n-1)}, x_{in}] \\ 0 & \text{otherwise} \end{cases}$$

where $h_i = (b_i - a_i)/(n - 1)$ and $x_{ij} = a_i + h_i \cdot (j-1)$.

3) For each k-tuple $\{h_1,\ldots,h_k\} \in \{1,\ldots,n\}^k$ we see that exists at least an object O_j with values $(p_{j1}, p_{j2}, \ldots p_{jk})$ such that $A_{1h_1}(p_{j1}) \cdot A_{2h_2}(p_{j2}) \cdot \ldots \cdot A_{kh_K}(p_{jk}) > 0$. If this condition is not verified, the set $P = \{(p_{j1},p_{j2},\ldots,p_{jk}), j=1,\ldots,m$ is not sufficiently dense w. r. t. the basic functions (8) and hence the process is stopped.

4) For simplicity, we put $H(p_{j1},p_{j2},\ldots,p_{jk}) = p_{jz}$ for every j=1,2,…,m. In accordance to (4), the $(h_1,h_2,\ldots,h_k)$-th component $F_{h_1h_2\ldots h_K}$ of the direct F-transform of H is defined as

$$F_{h_1h_2..h_k} = \frac{\sum_{j=1}^{m} p_{jz} \cdot A_{1h_1}(p_{j1}) \cdot \ldots \cdot A_{kh_K}(p_{jk})}{\sum_{j=1}^{m} A_{1h_1}(p_{j1}) \cdot \ldots \cdot A_{kh_K}(p_{jk})} \tag{9}$$

Then, in accordance to (5), the inverse F-transform $H^F_{n_1n_2\ldots n_k}$ of H is defined as

$$H^F_{n_1n_2\ldots n_k}(p_{j1}, p_{j2},\ldots p_{jk}) = \sum_{h_1=1}^{n_1}\sum_{h_2=1}^{n_2}\ldots\sum_{h_k=1}^{n_k} F_{h_1h_2\ldots h_K} \cdot A_{1h_1}(p_{j1}) \cdot \ldots \cdot A_{kh_K}(p_{jk}) \tag{10}$$

5) The difference between $H^F_{n_1n_2\ldots n_k}$ and H (or X_z) is estimated in the points $(p_{j1},p_{j2},\ldots,p_{jm})$, j=1,…,m, via the following statistical indexes of determinacy:

$$r_c^2 = \frac{\sigma^2_{H_F}}{\sigma^2_{X_Z}} = \frac{\sum_{j=1}^{m}\left(H^F_{n_1n_2\ldots n_k}(p_{j1}, p_{j2},\ldots p_{jk}) - \hat{p}_z\right)^2}{\sum_{j=1}^{m}\left(p_{jz} - \hat{p}_z\right)^2} \tag{11}$$

$$r'^2_c = 1 - \left[\left(1 - r_c^2\right) \cdot \frac{m-1}{m-k-1}\right] \tag{12}$$

where $\hat{p}_z$ is the mean of the values of X_z. We also look at a threshold value α determined experimentally: if r_c^2 and r'^2_c are less or equal to α, then we consider correct our model (7) and stop the process, otherwise we set n ← n+1 and go back to 2). We fix the threshold α at a confidence value.

Strictly speaking, Theorem 2 is unable to evaluate the correctness of our model (7) because it does not give a method of calculation of the k indexes $n_1(\varepsilon),\ldots, n_k(\varepsilon)$ and of the basic functions w. r. t. which the set $\{p_{j1}, p_{j2}, \ldots, p_{jk}$, j=1,…,m} is sufficiently dense in such a way

$$\left|H(p_{j1},p_{j2},\ldots p_{jk}) - H^F_{n_1(\varepsilon)\ldots n_k(\varepsilon)}(p_{j1},p_{j2},\ldots p_{jk})\right| < \varepsilon \tag{13}$$

In other words, the substitution of H obtained with $H^F_{n_1(\varepsilon)\ldots n_k(\varepsilon)}$ (up to an arbitrary quantity ε) is possible by Theorem 2, but we use the statistical index (12) from a practical point of view. Indeed it allows additionally to compare various regression models that intend to explain the same dependent variable from a different number k of explanatory variables.

4 A sample simulation

The calculation of the threshold value α for the indexes of determinacy (11) and (12) is made with training tests. As sample example, we present the case of a dataset with 1000 records and two fields, that is m=1000, r=2, k=1, X_1 = "Area of a circle", X_2 = "Radius of a circle", $a_1 = 0$, $b_1 = \pi$, $a_2 = 0$ and $b_2 = 1$. Hence the values of the attributes are $p_{j1} \in [0,\pi]$ and $p_{j2} \in [0,1]$ for every j = 1, …, 1000. If z = 2, we have $X_2 = H(X_1) = (X_1/\pi)^{1/2}$. In the interval [0, π] we consider a maximum number of 500 basic functions of the type (8) since we have seen that the set $P = \{p_{j1} : j = 1,\ldots,1000\}$ is not sufficiently dense w. r. t. the chosen partitions for $n \geq 500$. Table 1 contains the values of the indexes r_c^2 and r'^2_c for several values of n and it is plain that a good choice is α = 0.9 for $n \leq 7$.

ISBN: 978-989-95079-6-8

Table 1. Indexes of determinacy

n	r_c^2	r'^2_c
3	0.587000	0.588000
5	0.826898	0.827725
7	0.900207	0.901108
10	0.943000	0.944000
15	0.969749	0.970719
25	0.986085	0.987072
50	0.995084	0.996080
300	0.999000	1.000000
400	0.999000	1.000000
500	0.999000	1.000000

For sake of completeness, we show the plots in Figure 1 and Figure 2 of the values of X_2 and H_n^F for n = 3 and n = 300, respectively.

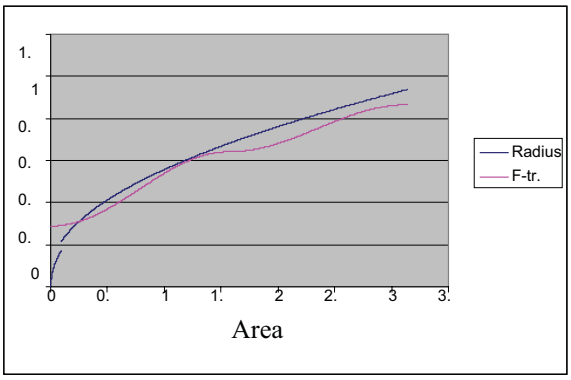


Figure 1. Graph of X_2 and H_3^F.

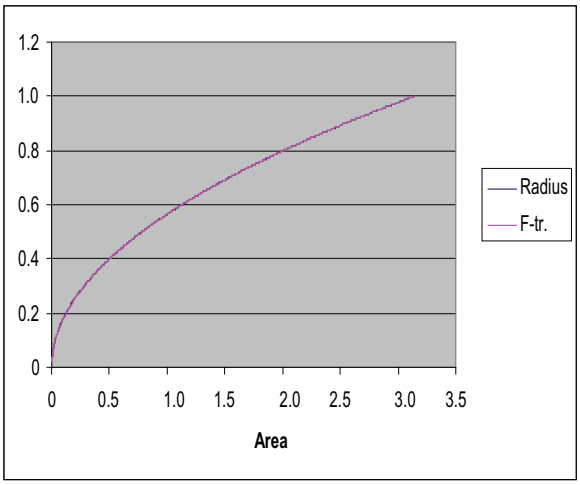


Figure 2. Graph of X_2 and H_{300}^F.

5 A complete experiment

We have downloaded from the known dataset "El Nino" (http://kdd.ics.uci.edu/databases/el_nino/el_nino.data.html) the oceanographic and surface meteorological data measured from a series of buoys positioned throughout the equatorial ocean Pacific. The data are formed from the following attributes:

- X_1 = date,
- X_2 = latitude,
- X_3 = longitude,
- X_4 = zonal winds (west < 0, east > 0),
- X_5 = meridional winds (south < 0, north > 0),
- X_6 = relative humidity,
- X_7 = air temperature in degrees Celsius,
- X_8 = sea surface temperature in degrees Celsius.

The data were measured from the buoys in various locations since 1980. The regression analysis establishes that significant relationships between the variables were not observed except for $X_8=H(X_7)$ and $X_8=H(X_1,X_7)$ which are of linear type like we will show in the sequel.
The training tests give $\alpha = 0.8$ as a threshold value. For single attribute dependencies, Table 2 shows the possible maximal number n_{max} of basic functions which can be used for deducing H_n^F. If n_{max} is overcome, the set of the attribute values is not sufficiently dense w. r. t. the chosen partitions. Moreover the absence of n_{max} in some entry of Table 2 means that the dependence model is not analyzable with the F-transforms method.

Table 2. n_{max} usable to explore single dependencies

	Date	Latit.	Lon.	Zon. wind	Mer. wind	Rel. hum.	Air temp	Sea sur. tem..
Date		13	3	14	48	5	14	3
Latitude	25		3	14	48	5	21	3
Longit.	25	13		14	48	5	21	3
Zonal winds	25	13	3		48	4	21	3
Merid. winds	25	13	3	14			21	3
Relative humid.	25	13		10	48		20	3
Air temper.	25	13	3	7	33			3
Sea sur. temper.	25	13	3	7	33	3	21	

The most meaningful model is $X_8=H(X_7)$ as Figure 3 shows: the plot is obtained for n = 10 and moreover we have $r_c^2 = r'^2_c = 0.83$. In the analysis of multiple attribute dependencies, Table 3 shows that only the model $X_8 = H(X_1,X_7)$ has the indexes of determinacy greater than $\alpha = 0.8$. Figure 4 shows the related 3D graph.

6 Conclusions

The F-transforms analyze all the possible dependencies between attributes in the datasets via a function H whose inverse F-transform is inside the value of the indexes of determinacy compared with a threshold parameter α. Our approach has the advantage to control always that the set of the attribute values is sufficiently dense w.r.t. the basic functions forming an uniform partition of the interval of context related to each attribute, otherwise the algorithm is stopped.

ISBN: 978-989-95079-6-8

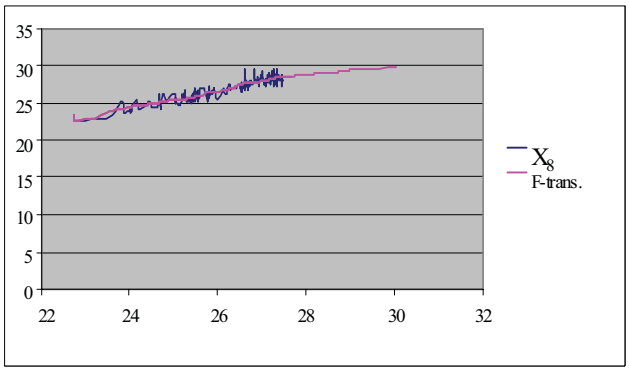


Figure 3. Graph of $X_8 = H(X_7)$ and H_{10}^F.

Table 3. Best values of indexes of determinacy for X_8

X_z	$X_1,\ldots, X_k$	r_c^2	r'^2_c
X_8	X_1, X_7	0.827	0.828
X_8	X_6, X_7	0.341	0.338
X_8	X_4, X_5	0.266	0.266
X_8	X_1, X_6, X_8	0.110	0.111

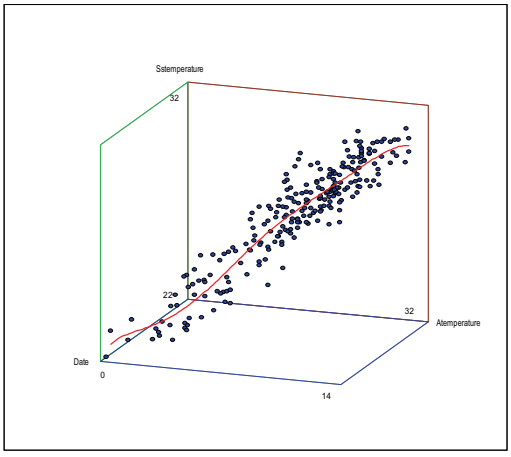


Figure 4. Plot 3D of $X_8 = H(X_1,X_7)$.

This method shall be integrated in future works with data mining classification and with results of regression data, such as decision tree learning and association mining algorithms.

References

[1] M. Daňková and M. Štěpnička, Fuzzy transform as an additive normal form, *Fuzzy Sets and Systems* 157 (8) (2006) 1024–1035.

[2] F. Di Martino, V. Loia, I. Perfilieva and S. Sessa, An image coding/ decoding method based on direct and inverse fuzzy transforms, *Internat. Journal of Approximate Reasoning* 48 (2008) 110–131.

[3] F.Di Martino, V. Loia and S. Sessa, Fuzzy transforms for compression and decompression of color videos, *Information Sciences,* submitted.

[4] F. Di Martino and S. Sessa, Compression and decompression of images with discrete fuzzy transforms, *Information Sciences* 177 (2007) 2349–2362.

[5] N.R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons Inc., New York, 1998.

[6] J. Han and M. Kamber, *Data Mining: Concepts and Techniques* (2nd Edition), Morgan-Kaufman, New York, 2006.

[7] R.A. Johnson and D.W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall International, London, 1992.

[8] Y. S. Lee and S. J. Yen, Classification based on attribute dependency, Proceedings of 6th Internat. Conference DaWaK' 04 (Zaragoza, 1-3/9/04) *Lecture Notes in Computer Sciences*, Vol. 5192, 2004, 259 – 268.

[9] S. Mitra, S.K. Pal and P. Mitra, Data mining in soft computing framework: a survey, *IEEE Transactions on Neural Networks* 13 (1) (2002) 3–14.

[10] I. Perfilieva, V. Novák and A. Dvořák, Fuzzy transforms in the analysis of data, *Internat. Journal of Approximate Reasoning* 48 (2008) 36–46.

[11] I. Perfilieva, Fuzzy transforms: application to reef growth problem, in: R. B. Demicco, G. J. Klir (eds.), *Fuzzy Logic in Geology*, Academic Press, Amsterdam, 2003, pp. 275–300.

[12] I. Perfilieva, Fuzzy transforms: theory and applications, *Fuzzy Sets and Systems* 157 (8) (2006) 993–1023.

[13] I. Perfilieva, Fuzzy transforms: a challenge to conventional transforms, In: P.W. Hawkes (Ed.), *Advances in Images and Electron Physics*, vol. 147, Elsevier Academic Press, San Diego, 2007, pp. 137–196.

[14] M. Štěpnička and R. Valášek, Fuzzy transforms for functions with two variables, in: J. Ramik, V. Novák (eds.), *Methods for Decision Support in Environment with Uncertainty-Applications in Economics*, *Business and Engineering*, IRAFM, Ostrava, Czech Republic, 2003, pp. 96–102.

[15] M. Štěpnička and R. Valášek, Numerical solution of partial differential equations with help of fuzzy transform, in: *Proc. Int. Conf. on FUZZ – IEEE*, Reno, 2005, pp. 1104–1109.

ISBN: 978-989-95079-6-8

Interval-valued Fuzzy propositions. An application to the L-Fuzzy contexts with absent values

C.Alcalde[1] A. Burusco, R. Fuentes-González[2]

1.Escuela Universitaria Politécnica. Departamento de Matemática Aplicada.
Universidad del País Vasco. Plaza de Europa 1, 20018 San Sebastián, Spain.
2.Departamento de Automática y Computación.
Universidad Pública de Navarra, Campus de Arrosadía, 31006 Pamplona, Spain.
Email: c.alcalde@ehu.es, {burusco,rfuentes}@unavarra.es

Abstract— *In order to obtain the information from an L-Fuzzy context, the complete relation between the objects and the attributes is needed. However, the contexts that model many situations have absent values.*
To solve this problem, at the beginning of the paper we remind the interval-valued linguistic variable definition and, later, we propose an extension of the fuzzy propositions to the interval-valued case.
In the second part, we apply these ideas to the problem of replacement of the absent values in an interval-valued L-Fuzzy Context. The use of interval-valued fuzzy propositions simplifies the process.

Keywords— Interval-valued fuzzy propositions, Interval-valued linguistic variables, Interval-valued L-Fuzzy Contexts.

1 Introduction

The first extension of the Formal Concept Analysis to the fuzzy case, the L-Fuzzy Concept Theory, is due to A. Burusco and R. Fuentes-González and was published in 1994 [1].

Later, to extract knowledge from a more general table, we defined the interval-valued L-Fuzzy context as an extension to the interval-valued case of the L-Fuzzy context defined in the L-Fuzzy Concept Theory [2]:

An *interval-valued L-Fuzzy context* is a tuple $(\mathcal{J}[L], X, Y, R)$, with X and Y two finite sets (of *objects* and *attributes*), R an interval-valued L-Fuzzy relation between X and Y and $\mathcal{J}[L]$ the set of the closed intervals of a lattice L with the usual order.

To replace the absent values, we have already developed a theory based on implications between attributes [3] that has given good results when has been applied to the Technique [4]. In these papers, in order to replace the absent values, we have used some labels that allow to obtain implications with high values of support and confidence.The good behavior of these labels, suggested us the application of the linguistic variables to our work [5].

In this work, we are going to define fuzzy propositions for the interval-valued case (they do not exist in the Literature of the subject) and later we will try to use them to replace the absent values in contexts instead of using implications between attributes. We will finally analyze the advantages of the method.

2 Linguistic variables defined in the set of closed intervals of $[0, 1]$

Let $(V, T(V), U, G, M)$ be a linguistic variable (Zadeh [6]) defined in the set $U = [0, 1]$, whose values or labels $T(V)$ are associated with the generalized trapezoidal fuzzy numbers defined by Yao and Lin [7]. We have extended the linguistic variable definition to the interval-valued case [5]:

Definition 1 *Taking as a departure point a linguistic variable V, we define an interval-valued linguistic variable associated with V as the linguistic variable $\mathbf{V}$ defined in the set $\mathcal{J}[U]$ of the closed intervals of U, characterized by the tuple $(\mathbf{V}, T(V), \mathcal{J}[U], G, M)$, where the compatibility function with each label $t \in T(V)$ is $\boldsymbol{c}_t : \mathcal{J}[U] \to \mathcal{J}[0, 1]$.*
Notation. *We denote the compatibility of the value $[\alpha, \beta] \in \mathcal{J}[U]$ with the label t by:*

$$\boldsymbol{c}_t([\alpha, \beta]) = [\alpha, \beta]_t = \\ = \begin{cases} [\min\{\alpha_t, \beta_t\}, \max\{\alpha_t, \beta_t\}] & \text{if } \alpha \geq b \text{ or } \beta \leq a \\ [\min\{\alpha_t, \beta_t\}, 1] & \text{in other case} \end{cases}$$

where a and b are the values that define the generalized trapezoidal fuzzy number associated with the label $t \in T(V)$.

With this definition we will try to represent the interval in which x_t have its values when $x \in [\alpha, \beta]$.

The defined interval-valued linguistic variable $\mathbf{V}$ is an extension of the linguistic variable V to the set of closed intervals of U.

Some properties of these interval-valued linguistic variables have been proven in [5].

3 Interval-valued fuzzy propositions

Before studying the interval-valued fuzzy propositions, we are going to remind the main aspects about fuzzy propositions.

3.1 Fuzzy propositions

As can be seen in [8], the main difference between classical and fuzzy propositions is that the truth value of the second ones belongs to the interval [0,1]. Moreover, four types of fuzzy propositions are defined by Klir [8]:

Unconditional and unqualified propositions: $p : \Phi$ *is* A, with Φ a variable and A a predicate attributed to the variable and represented by a fuzzy set. In other words, p : 'Φ *is* A' *is true*. The truth value of the proposition $p_\Phi : \Phi = \phi$ *is* A, is obtained by $T_r(p_\phi) = A(\phi), \forall \phi \in \Phi$.

Unconditional and qualified propositions: The difference with the previous ones is that the proposition has a qualified truth value that modifies the truth value. They are

ISBN: 978-989-95079-6-8

characterized by the form p : 'Φ *is* A*' is* S, with S qualified truth value.
Examples of truth qualifiers are expressions as *true*, *very true*, *fairly true*, *false*, *very false* or *fairly false* (See Fig. 1).

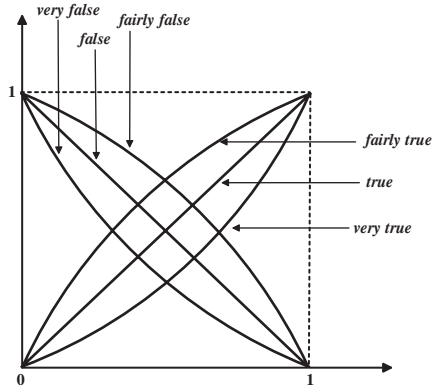


Figure 1: Some examples of truth qualifiers

Conditional and unqualified propositions: These propositions are expressed by p : *If* Φ *is* A*, then* Ψ *is* B, with Φ and Ψ variables and A and B predicates represented by fuzzy sets. That is, p : '*If* Φ *is* A*, then* Ψ *is* B*' is true*. The truth value of the proposition is calculated using an implication operator:

$$T_r(p_{\phi,\psi}) = I[A(\phi), B(\psi)]$$

Conditional and qualified propositions: p : '*If* Φ *is* A*, then* Ψ *is* B*' is* S, with S a qualified truth value. The truth value of the proposition is obtained by

$$T_{rS}(p_{\phi,\psi}) = S[T_r(p_{\phi,\psi})]$$

3.1.1 Fuzzy quantifiers

The two quantifiers of the predicate logic are *all* and *exists*. Fuzzy quantifiers are a tool for symbolizing quantified statements that minimizes the lost of information forced by the choice of quantifier.

The fuzzy propositions of any of the types introduced in the previous section may be quantified by a suitable fuzzy quantifier. In general, fuzzy quantifiers are fuzzy numbers which take part in the various propositional forms and affect the degrees of truth of specific fuzzy propositions. Each fuzzy quantifier expresses an approximate number of elements or an approximate proportion of elements in a given universal set that claims to satisfy a given property.

There are two types of fuzzy quantifiers: The absolute quantifiers expressed by fuzzy numbers defined on the set of real numbers or on the set of integers: *about a dozen*, *at most about 10*, *at least about 100...* And the relative quantifiers expressed by fuzzy numbers defined on [0,1]: *most*, *almost all*, *about half*, *about 20%*. These relative quantifiers are the most interesting for our study. So, we are going to see how the fuzzy propositions with these fuzzy quantifiers are.

Fuzzy propositions with quantifiers of the second kind have the general form ([8]):

p : *Among z's in* Z *such that* $\nu_1(z)$ *is* F_1 *there are* Q *z's in* Z *such that* $\nu_2(z)$ *is* F_2

Any propositions p can be expressed in a simplified form:

$$p' : QE_1's \text{ are } E_2's$$

where Q is the used quantifier, E_1 and E_2 are the fuzzy sets on Z defined by $E_1(z) = F_1(\nu_1(z))$, $E_2(z) = F_2(\nu_2(z)), \forall z \in Z$, with F_1 and F_2 fuzzy sets.

We may rewrite this proposition in the form:

$$p' : W \text{ is } Q$$

with W a variable that represents the degree of subsethood of E_2 in E_1, that is:

$$W = \frac{|E_1 \cap E_2|}{|E_1|}$$

where $|\quad|$ represents the cardinality of a set.

Using the standard fuzzy intersection, we obtain:

$$W = \frac{\sum_{z \in Z} \min\{F_1(\nu_1(z)), F_2(\nu_2(z))\}}{\sum_{z \in Z} F_1(\nu_1(z))}$$

for any E_1 and E_2.

Then, the truth value is $T_r(p) = T_r(p') = Q(W)$.

3.1.2 Linguistic hedges

Linguistic hedges are special linguistic terms by which other linguistic terms are modified: *Very*, *more or less*, *fairly* or *extremely*. They can be used for modifying fuzzy predicates, fuzzy truth values and fuzzy quantifiers.

Any linguistic hedge may be interpreted as an unary function h in [0,1]. For example, *Very* ($h(a) = a^2$), *fairly* ($h(a) = \sqrt{a}$). These functions h are also said to be modifiers or qualifiers.

3.2 Interval-valued fuzzy propositions

There are some situations where we work with interval-valued fuzzy variables and fuzzy propositions and we think that the truth value of an interval-valued fuzzy proposition must be also an interval in order to model in a more suitable way the represented data. An example of this situation can be seen in next section.

With this aim, we are going to study the interval-valued fuzzy propositions.
Then, we define an interval-valued fuzzy proposition with fuzzy quantifiers of the second kind as:

$$\mathbf{p}' : QE_1{}'s \text{ are } E_2{}'s$$

where Q is the quantifier, and E_1,$E_2 \in \mathcal{J}[L]^X$ are interval-valued fuzzy sets.

As in the previous section, we can rewrite this proposition as:

$$\mathbf{p}' : W \text{ is } Q$$

ISBN: 978-989-95079-6-8

with W a variable that represents the idea of degree of subsethood of E_2 in E_1.

Using the standard fuzzy intersection, we can express this degree as:

$$W = \left[\frac{\sum_{x\in X}\min\{l_{E_1}(x), l_{E_2}(x)\}}{\sum_{x\in X} u_{E_1}(x)}, \frac{\sum_{x\in X}\min\{u_{E_1}(x), u_{E_2}(x)\}}{\sum_{x\in X} l_{E_1}(x)} \wedge 1\right]$$

where u_{E_1}, u_{E_2}, l_{E_1} y l_{E_2} are the upper and lower bounds of the interval-valued observations of E_1 and E_2.

Then, we define the truth value of the proposition $\mathbf{p}'$ as

$$\mathbf{T_r}(\mathbf{p}') = Q(W) = \\ = [\min\{Q(l_W), Q(u_W)\}, \max\{Q(l_W), Q(u_W)\}]$$

where l_W and u_W are the lower and upper bounds of W respectively.

4 Use of the linguistic variables and fuzzy propositions to replace the absent values in Contexts

4.1 Interval-valued L-Fuzzy Contexts with absent values

To extract knowledge from a table with incomplete and ambiguous information, we use the interval-valued L-Fuzzy context as an extension of the L-Fuzzy context given by an implication operator [1, 2] to the interval-valued case.

Definition 2 *Let $\mathcal{J}[L]$ be the set of the closed intervals of a lattice L with the usual order. An* interval-valued L-Fuzzy context *is a tuple $(\mathcal{J}[L], X, Y, R)$, with X and Y two finite sets (of* objects *and* attributes*) and $R \in \mathcal{J}[L]^{X\times Y}$ an interval-valued L-Fuzzy relation between X and Y.*

We are going to see an example of an interval-valued L-Fuzzy context to understand the problem.

Example 1 *In Table 1 we have collected the amounts of the different ingredients (I_i) of a certain food item produced under several trademarks (M_j). As can be seen, there are two unknown values. The table values have been normalized and belong to $\mathcal{J}[L]$, the set of intervals of the lattice $L = \{0, 0.1, 0.2, \ldots 0.9, 1\}$.*
This is an example of an interval-valued L-Fuzzy context $(\mathcal{J}[L], X, Y, R)$, where the set of objects X is the set of trademarks or brands, the set of attributes Y is the set of ingredients and the relation R between both is represented in the table.

In the papers developed to study contexts with absent values [3, 4] we have analyzed the implications between attributes associated with labels to obtain good values to replace the absent ones.

In those papers, we represent these absent values by the interval [0,1] and analyze the implications between attributes with the intention of obtaining good values to reduce, as much as possible, the width of that interval.

The support and confidence definitions of an interval-valued implication are based on the idea of association rules [9].

Definition 3 *Let $(\mathcal{J}[L], X, Y, R)$ be an interval-valued L-Fuzzy context. Given the attribute sets $B, C \in \mathcal{J}[L]^Y$, we define the* support *of the implication $B \Rightarrow C$, $\mathrm{supp}(B \Rightarrow C)$, as the interval:*

$$\left[\frac{\sum_{x\in X} l_{(B\cup C)_2}(x)}{|X|}, \frac{\sum_{x\in X} u_{(B\cup C)_2}(x)}{|X|}\right]$$

where the membership function of the derived set B_2 is $B_2(x) = [l_{B_2}(x), u_{B_2}(x)]$ and represents the percentage of objects that share the attributes of B and C.

The confidence *of the implication, $\mathrm{conf}(B \Rightarrow C)$, is given by the interval:*

$$\left[\frac{\sum_{x\in X} l_{(B\cup C)_2}(x)}{\sum_{x\in X} u_{B_2}(x)}, \frac{\sum_{x\in X} u_{(B\cup C)_2}(x)}{\sum_{x\in X} l_{B_2}(x)} \wedge 1\right]$$

and represents the percentage of objects that verify the implication, that is, the percentage of objects that having the attributes of B to a certain degree also have those of C to the same degree.

One kind of these implications that we have used are the implications associated with labels [3]. They are denoted by $y_{i_r} \Longrightarrow y_{j_s}$ where y_i and y_j are attributes of the context and r, s labels. In the calculation process, a function G_s (based on the distance between intervals) to measure the degree of similarity between the different labels and the attributes is needed. In addition, once the implications with highest support and confidence have been chosen, the absent values are replaced by intervals, in general, with less amplitude than [0,1].

In this work, we are going to approach all this process, in an easy way, using interval-valued linguistic variables and fuzzy propositions.

The implication between attributes $y_{i_r} \Longrightarrow y_{j_s}$ can be considered as a quantified fuzzy proposition since the expression:

$\mathbf{p}$: *Most of those that have y_i(high, medium, low...) also have y_j (high, medium, low...)*

can be interpreted in the sense of section 3.2:

The quantifier Q is *most* and E_1,E_2 are interval-valued linguistic variables resulting of applying the linguistic variable (with values *high, medium, low...*) to the columns of the context. The quantifier Q will be applied to both of the bounds of the intervals. In this way, we can calculate the truth value of the proposition:

$$\mathbf{T_r}(\mathbf{p}) = Q(W) = \\ = [\min\{Q(l_W), Q(u_W)\}, \max\{Q(l_W), Q(u_W)\}]$$

4.2 Substitution of the absent values using the compatibility function of an interval-valued linguistic variable and the interval-valued fuzzy propositions.

In the study of contexts with absent values, we begin replacing these absent values by [0,1]. Next, we are going to define an interval-valued linguistic variable $(\mathbf{V}, T(V), \mathcal{J}[U], G, M)$, and use the interval-valued fuzzy propositions to estimate the

ISBN: 978-989-95079-6-8

Table 1: Amounts of the different ingredients under several trademarks.

R	$I1$	$I2$	$I3$	$I4$	$I5$	$I6$
$M1$	$[0.8, 1]$	$[0.1, 0.1]$	$[0.6, 0.6]$	$[0.2, 0.2]$	$[0.4, 0.4]$	$[0.4, 0.6]$
$M2$	$[0.8, 0.8]$		$[0.4, 0.6]$	$[0.4, 0.4]$	$[0.2, 0.2]$	$[0.6, 0.6]$
$M3$	[1,1]	[0,0]	$[0.2, 0.2]$	$[0.5, 0.5]$	$[0.4, 0.4]$	$[0.4, 0.4]$
$M4$	$[0.9, 0.9]$	$[0.2, 0.2]$	$[0, 0]$	$[0.4, 0.4]$	$[0.1, 0.2]$	
$M5$	$[0.8, 0.8]$	$[0, 0.2]$	$[0.4, 0.4]$	$[0.2, 0.2]$	$[0.2, 0.4]$	$[0.2, 0.2]$

absent values $R(x_i, y_j)$ of the interval-valued L-Fuzzy context by intervals with less amplitude than that of [0,1].

The compatibility function of the interval-valued linguistic variable will be used to establish the new columns in the table that measure the similarity between some attributes and the values or labels $t \in T(V)$ of the linguistic variable.

Example 2 *In the previous context, the two absent values ($R(M2, I2)$ and $R(M4, I6)$) have been replaced by the interval [0,1].*

We are going to define a linguistic variable ***V****, with labels* $T(V)$ =*{high, medium, low}.*

Now, we take the label t =medium $\in T(V)$ *represented by Fig. 2,*

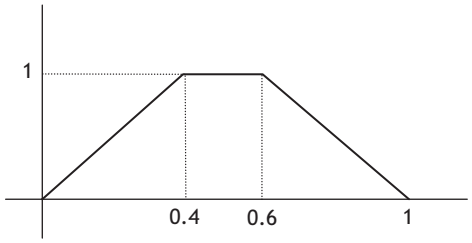


Figure 2: L-Fuzzy set associated with the label t

where the compatibility of the value $x \in X = [0, 1]$ *with the label* t =medium *has the following expression:*

$$x_t = \begin{cases} \dfrac{x}{0.4} & \text{if } x \leq 0.4 \\ 1 & \text{if } 0.4 \leq x \leq 0.6 \\ \dfrac{1-x}{1-0.6} & \text{if } 0.6 \leq x \end{cases}$$

and the interval-valued compatibility is defined, $\forall [\alpha, \beta] \in \mathcal{J}[X]$, *by:*

$$[\alpha, \beta]_t = \begin{cases} [\min\{\alpha_t, \beta_t\}, \max\{\alpha_t, \beta_t\}] & \text{if } \alpha \geq 0.6 \text{ or } \beta \leq 0.4 \\ [\min\{\alpha_t, \beta_t\}, 1] & \text{in other case} \end{cases}$$

From this point, we can extend the table with a new column $I6_{medium}$ *where the compatibility of the attribute* $I6$ *with the label* $t = medium$ *is valued, obtaining Table 2.*

Remark. Sometimes, the obtained values do not belong to the lattice $L = \{0, 0.1, 0.2, \ldots 0.9, 1\}$. Then, we will round the fractions to the nearest element of L.

One of the advantages of the use of linguistic variables to establish the compatibility between the values of the relation R and the labels is the maintenance of the unknowledge. It is possible to observe this property in the example where $R(M4, I6) = [0, 1] = \widehat{R}(M4, I6_{medium})$ and it is certain for any label $t \in T(V)$, as it was proven in [5].

On the other hand, the values here obtained are not very different from which turned out applying the techniques developed in [3] which can be seen in the Table 3. Nevertheless, we consider that the use of linguistic variables solves the problem in a simpler and more natural way.

Table 3: Values obtained from a previous method.

	$I6_{medium}$
$M1$	$[0.8, 1]$
$M2$	$[0.8, 1]$
$M3$	$[0.8, 1]$
$M4$	$[0.4, 1]$
$M5$	$[0.6, 0.8]$

Once extended the table, the interval-valued fuzzy propositions related to the new attribute $I6_{medium}$ can be calculated to find those that have highest values. In this case, this is the proposition $I1 \Rightarrow I6_{medium}$:

$\mathbf{p}$: *Most of those that have* $I1$ *also have* $I6_{medium}$

Then, Q =*most*, $E_1 = I1$, $E_2 = I6_{\text{medium}}$, where

$$W = \left[\frac{3.1}{4.5}, \frac{4.2}{4.3} \wedge 1\right] = [0.68, 0.97]$$

Now, if we take the quantifier Q defined by ([10]):

$$Q(x) = \begin{cases} 1.25x & \text{if } x \leq 0.8 \\ 1 & \text{in other case} \end{cases}$$

and we apply to both bounds of the interval, then we have:

$$Q([0.68, 0.97]) = [Q(0.68), Q(0.97)] = [0.85, 1]$$

that is the truth value of the proposition $\mathbf{p}$.

The obtained results are similar to those with implications between attributes but, in this case, the process is easier.

We can conclude that, in many cases (at least 85%), the ingredient $I1$ is associated with a medium amount of ingredient $I6$.

Given the high percentage of cases in which the implication is valid, we suppose that it also holds for brand $M4$. Therefore we expect that:

$$\widehat{R}(M4, I6_{medium}) \geq \widehat{R}(M4, I1) = [0.9, 0.9]$$

ISBN: 978-989-95079-6-8

Table 2: New relation $\widehat{R}$.

$\widehat{R}$	$I1$	$I2$	$I3$	$I4$	$I5$	$I6$	$I6_{medium}$
$M1$	$[0.8, 1]$	$[0.1, 0.1]$	$[0.6, 0.6]$	$[0.2, 0.2]$	$[0.4, 0.4]$	$[0.4, 0.6]$	$[1, 1]$
$M2$	$[0.8, 0.8]$	$[0, 1]$	$[0.4, 0.6]$	$[0.4, 0.4]$	$[0.2, 0.2]$	$[0.6, 0.6]$	$[1, 1]$
$M3$	[1,1]	[0,0]	$[0.2, 0.2]$	$[0.5, 0.5]$	$[0.4, 0.4]$	$[0.4, 0.4]$	$[1, 1]$
$M4$	$[0.9, 0.9]$	$[0.2, 0.2]$	$[0, 0]$	$[0.4, 0.4]$	$[0.1, 0.2]$	$[0, 1]$	$[0, 1]$
$M5$	$[0.8, 0.8]$	$[0, 0.2]$	$[0.4, 0.4]$	$[0.2, 0.2]$	$[0.2, 0.4]$	$[0.2, 0.2]$	$[0.5, 0.5]$

The election of the attributes and labels to obtain interval-valued fuzzy propositions with high truth values is very important because we will have then good values of substitution for the absent ones.

At this point, we have obtained the value $\widehat{R}(M4, I6_{medium}) = [0.9, 0.9]$. Now, using this value we want to get the best value to replace the initial absent value $R(M4, I6)$. To do this, denoting by $R(x_i, y_j) = [\alpha, \beta]$ the unknown value, we will use the following proposition associated with compatibility functions to obtain this value:

Proposition 1 *$\forall x \in [0, 1]$ and $[\alpha, \beta] \in \mathcal{J}[0, 1]$ it is verified that*

$$[\alpha, \beta]_t \subseteq [x, 1] \Leftrightarrow [\alpha, \beta] \subseteq [ax, 1 - x(1 - b)]$$

where $a, b \in [0, 1]$ are the values that appear in the fuzzy set associated with the label $t \in T(V)$.

Proof

$\Leftarrow$) Let us suppose that $[\alpha, \beta] \subseteq [ax, 1 - x(1 - b)]$.
In [5], we have proven that:

If $[\alpha, \beta] \subseteq [ax, 1 - x(1 - b)]$ holds,

then

$$[\alpha, \beta]_t \subseteq [ax, 1 - x(1 - b)]_t$$

Moreover, for all $x \in [0, 1]$, it is verified

$$\left.\begin{array}{l} ax \leq a \\ 1 - x(1 - b) \geq b \end{array}\right\} \Rightarrow$$

Therefore,

$$\begin{aligned} [\alpha, \beta]_t \subseteq [ax, 1 - x(1 - b)]_t = \\ = [\min\{(ax)_t, (1 - x(1 - b))_t\}, 1] = \\ = \left[\min\left\{\frac{ax}{a}, \frac{1 - (1 - x(1 - b))}{1 - b}\right\}, 1\right] = [x, 1] \end{aligned}$$

$\Rightarrow$) By reduction to the absurd, suppose that

$$[\alpha, \beta] \not\subseteq [ax, 1 - x(1 - b)]$$

then we have $\alpha < ax$ or $\beta > 1 - x(1 - b)$.

- If $\alpha < ax < a$, then:

$$\alpha_t = \frac{\alpha}{a} < \frac{ax}{a} = x \Rightarrow \min\{\alpha_t, \beta_t\} < x$$

- If $\beta > 1 - x(1 - b) > b$, then:

$$\beta_t = \frac{1 - \beta}{1 - b} < \frac{1 - (1 - x(1 - b))}{1 - b} = x \Rightarrow$$
$$\Rightarrow \min\{\alpha_t, \beta_t\} < x$$

Therefore $[\alpha, \beta]_t \not\subseteq [x, 1]$.

Remark. Really, we are only interested in the implication

$$[\alpha, \beta]_t \subseteq [x, 1] \Rightarrow [\alpha, \beta] \subseteq [ax, 1 - x(1 - b)]$$

To replace the absent value, we will take the widest interval verifying the condition, that is, $[ax, 1 - x(1 - b)]$.

As $[a, b] \subseteq [ax, 1 - x(1 - b)]$, we have to avoid to take labels with a wide amplitude of the interval $[a, b]$ because, in other case, the interval $[ax, 1 - x(1 - b)]$ will be also wide.

Example 3 *Returning to the example, if we apply the previous proposition associated with the label t=medium (with the values $a = 0.4$ and $b = 0.6$), then we obtain*

$$R(M_4, I_6) = [0.36, 0.64]$$

that will be the interval with which we will replace the absent value (also it will be necessary here to round the results to the nearest element of L). The final result given in [3] for this same absent value $R(M_4, I_6)$ was $[0.5, 0.5]$.

In both cases we have a fuzzy value distant from 0 and 1.

Finally, the other unknown value, $R(M2, I2)$ can be replaced using interval-valued linguistic variables and fuzzy propositions by the interval [0,0.2]. In this case, the proposition with the highest value is

p : *Most of those that have $I1_{high}$ also have $I2_{low}$*

5 Conclusions and future work

The use of linguistic variables and fuzzy propositions in the field of contexts with absent values supposes a simplification of the process of replacement of these values. The old method, based on implication between attributes, needs the calculus of functions to measure the similarity between attributes and labels that makes difficult the process.

In future works we will analyze its reliability.

Acknowledgment

This work has been partially supported by the Research Group "Intelligent Systems and Energy" of the University of the Basque Country, under grant GIU 07/45 and by the Research Project of the Government of Navarra (Resolution 2031 of 2008).

ISBN: 978-989-95079-6-8

References

[1] A. Burusco and R. Fuentes-González. The study of the L-fuzzy concept lattice. *Mathware and Soft Computing*, 1(3):209–218, 1994.

[2] A. Burusco and R. Fuentes-González. The study of the interval-valued contexts. *Fuzzy Sets and Systems*, 121(3):69–82, 2001.

[3] C. Alcalde, A. Burusco, R. Fuentes-González, and I. Zubia. Treatment of L-fuzzy contexts with absent values. *Information Sciences*, 179:1–15, 2009.

[4] C. Alcalde, A. Burusco, R. Fuentes-González, and I. Zubia. An application of the L-fuzzy concept theory to the short-circuit detection. *IEEE International Conference on Fuzzy Systems*, pages 547–555, 2007.

[5] C. Alcalde, A. Burusco, and R. Fuentes-González. Interval-valued linguistic variables. *The 8th International FLINS Conference on Computational Intelligence in Decision and Control*, pages 217–222, 2008.

[6] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences*, 8:199–249, 1975.

[7] J.S. Yao and Lin F.T. Constructing a fuzzy flow-shop sequencing model based on statistical data. *International Journal of Approximate Reasoning*, 29(3):215–234, 2002.

[8] G.J. Klir, U.H. St.Clair, and B. Yuan. *Fuzzy Set Theory. Foundations and Applications*. Prentice Hall, 1997.

[9] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD International Conference on Management of Data.*, pages 207–216, 1993.

[10] J. Kacprzyk, M. Fedrizzi, and H. Nurmi. *Fuzzy Logic with Linguistic quantifiers in group decision making. In: An introduction to Fuzzy Logic applications in intelligent systems*. Kluwer Academic Publishers, 1992.

ISBN: 978-989-95079-6-8

On the structure of the k-aditive fuzzy measures

Elías F. Combarro[1] Pedro Miranda[2]

1.Artificial Intelligence Center, University of Oviedo, Spain
2.Department of Statistics and O. R. , Complutense University of Madrid, Spain
Email: elias@aic.uniovi.es, pmiranda@mat.ucm.es

***Abstract*—** The family of k-additive measures has been introduced as a midterm between probabilities and general fuzzy measures and finds a wide number of applications in practice. However, its structure is different from other families of fuzzy measures and is certainly more complex (for instance, its vertices are not always $\{0,1\}$-valued), so it has not been yet fully studied.

In this paper we present some results concerning the extreme points of the k-additive fuzzy measures. We give a characterization of these vertices as well as an algorithm to compute them. We show some examples of the results of this algorithm and provide lower bounds on the number of vertices of the $n-1$-additive measures, proving that it grows much faster than the number of vertices of the general fuzzy measures. This suggests that k-additive measures might not be a good choice in modeling certain decision problems when the value of k is high but not equal to n.

***Keywords*—** Fuzzy measures, k-additive measures, polytope vertices.

1 Introduction

Fuzzy measures (also known as capacities or non-additive measures) are a generalization of probability distributions. More concretely, they are measures in which the additivity axiom has been relaxed to a monotonicity condition. This extension is needed in many practical situations, in which additivity is too restrictive. Fuzzy measures have proved themselves to be a powerful tool in many different fields, as Decision Theory ([1], [2]), Game Theory ([3], [4]), and many others (see, for example, [5],[6]).

However, despite all these advantages, the practical application of fuzzy measures is limited by the increased complexity of the measure. If we have a finite space of cardinality n, only $n-1$ values are needed in order to completely determine a probability, while 2^n-2 coefficients are needed to define a fuzzy measure on the same referential. This exponential growth is the actual *Achilles' heel* of fuzzy measures and translates in big complexity when identifying the fuzzy measure modelling a situation.

With the aim of reducing this complexity several of subfamilies have been defined. In these families some extra restrictions are added in order to decrease the number of coefficients but trying to keep the modelling capabilities of the measures. Examples of subfamilies include the λ-measures [7], the k-intolerant measures [8, 9], the p-symmetric measures [10], the decomposable measures [11], etc. In this paper we will focus on the complexity of k-additive measures (see Definition 9 below) introduced by Grabisch in [4].

Consider a situation that can be modeled by a k-additive measure (possibly through Choquet integral [12]); to this extent, an axiomatic characterization of such a model can been found in [13]. Next step is the identification of the measure. Suppose that sample information is available. Using Decision Theory terminology, this information consists in a collection of objects whose overall score and valuation on each criterium are known. We assume that this information is numerical; otherwise, we should previously apply a tool to transform ordinal into numerical data, such as MACBETH [14] or TOMASO [15]. The desired measure (that might not be unique) is the one that best fits the data.

If the considered proximity criterion is the squared error, different techniques exist to solve the problem [16, 17], leading to a problem whose complexity is very high. In a previous work [18], we have proposed a learning method based on genetic algorithms [19]. In this method, the crossover operator is the convex combination (as it will become clearer below, the set of fuzzy measures being at most k-additive forms a convex polytope). This operator is simple and natural and allows to decrease the computational cost. However, it has the drawback that the search regions are embedded one in other. This can be overcome with a suitable mutation operator, but even in that case, the search region is determined by the initial population. Thus, this initial population must be wide enough to include the solution and, obviously, the best option is to select the set of vertices of the polytope.

In this paper we study this set of vertices for the family of k-additive fuzzy measures. The results in the paper seem to mean that the number of k-additive vertices is even greater than the number of vertices for general fuzzy measures. The rest of the paper is organized as follows. In Section 2 we briefly introduce definitions and results that will be used throughout the paper. In Section 3 we provide an algorithm to determine the vertices of the polytope of k-additive (at most) measures from those of the general fuzzy measures and show some examples of the results of the computation of the algorithm. Section 4 is devoted to the study of lower bounds on the number of vertices when $k=n-1$. Finally, in Section 5 some conclusions are drawn and ideas for future work are presented.

2 Basic concepts and previous results

Consider a finite referential set $X=\{1,...,n\}$ of n elements or *criteria*. Let us denote by $\mathcal{P}(X)$ the set of subsets of X. Subsets of X are denoted $A,B,...$ and also by $A_1,A_2,....$

Definition 1 *[12, 20, 21] A* **fuzzy measure**, **non-additive measure** *or* **capacity** *over X is a function $\mu:\mathcal{P}(X)\to[0,1]$ satisfying*

- $\mu(\emptyset)=0$, $\mu(X)=1$ *(boundary conditions).*

ISBN: 978-989-95079-6-8

- $\forall A, B \in \mathcal{P}(X)$, *if* $A \subseteq B$, *then* $\mu(A) \leq \mu(B)$ *(monotonicity).*

We will denote the set of all fuzzy measures over X by $\mathcal{FM}(X)$. Remark that $\mathcal{FM}(X)$ is a bounded convex polyhedron in $\mathbb{R}^{2^n-2}$ (or $\mathbb{R}^{2^n}$ if we include coordinates for X and $\emptyset$).

We can define a partial order on $\mathcal{FM}(X)$ in the following way:

Definition 2 *Let μ_1 and μ_2 be two fuzzy measures. Then we say that $\mu_1 \leq \mu_2$ if for every $A \subseteq X$, it holds $\mu_1(A) \leq \mu_2(A)$.*

A special class of fuzzy measures is the set of $\{0,1\}$-valued measures.

Definition 3 *A fuzzy measure is* **$\{0,1\}$-valued** *if it only takes values 0 and 1.*

Notice that for a $\{0,1\}$-valued measure μ, there are some subsets A satisfying the following conditions:

$$\begin{array}{ll} \mu(A) = 1, & \\ \mu(B) = 1, & \forall B \supseteq A, \\ \mu(C) = 0, & \forall C \subset A. \end{array}$$

We will call any subset satisfying these conditions a **minimal subset** for μ. Note that a $\{0,1\}$-valued measure is completely defined by its minimal subsets. Let us denote by $\mu_{\mathbf{C}}$ the fuzzy measure whose collection of minimal sets is $\mathbf{C}$.

An important instance of $\{0,1\}$-valued measures are the so-called unanimity games:

Definition 4 *A* **unanimity game** *over $A \subseteq X$, $A \neq \emptyset$ is a fuzzy measure defined by*

$$u_A(B) := \begin{cases} 1 & \text{if A} \subseteq \text{B} \\ 0 & \text{otherwise} \end{cases}$$

For $\emptyset$, we define the unanimity game by

$$u_\emptyset(B) := \begin{cases} 1 & \text{if B} \neq \emptyset \\ 0 & \text{if B} = \emptyset \end{cases}$$

Definition 5 *Let μ be a fuzzy measure over X; we define the* **dual measure of** *μ as the fuzzy measure $\overline{\mu}$ given by $\overline{\mu}(A) = 1 - \mu(A^c)$.*

In order to define the *extreme points* or *vertices* of a family of fuzzy measures we need to define the convex combination of measures.

Definition 6 *Given $\lambda \in [0,1]$ and two fuzzy measures μ_1 and μ_2 the fuzzy measure*

$$\mu = \lambda\mu_1 + (1-\lambda)\mu_2$$

is said to be a **convex combination** *of μ_1 and μ_2. A subset of fuzzy measures $\mathcal{F} \subseteq \mathcal{FM}(X)$ is* **convex** *if it contains every convex combination of its members.*

Definition 7 *Given a convex subset of fuzzy measures, $\mathcal{F} \subseteq \mathcal{FM}(X)$, we say that $\mu \in \mathcal{F}$ is a* **vertex** *or* **extreme point** *of $\mathcal{F}$ if μ can not be written as a convex combination of two measures $\mu_1, \mu_2 \in \mathcal{F}$ both different from μ.*

Obviously, the set of vertices completely determines the convex polytope $\mathcal{FM}(X)$, as any point in it can be written as a convex combination of the vertices. The vertices of $\mathcal{FM}(X)$ are given in the following result.

Proposition 1 *[22] The set of $\{0,1\}$-valued measures constitutes the set of vertices of $\mathcal{FM}(X)$.*

A similar result (see [23]) can be proved for the family of p-symmetric measures [10]. However, it does not hold [23] for the family of measures that we study in this paper: the k-additive measures, which are defined through the concept of Möbius transform.

Definition 8 *[6, 24] Let μ be a set function (not necessarily a fuzzy measure) on X. The* **Möbius transform (or inverse)** *of μ is another set function on X defined by*

$$m(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B), \, \forall A \subseteq X. \tag{1}$$

The Möbius transform given, the original set function can be recovered through the *Zeta transform* [25]:

$$\mu(A) = \sum_{B \subseteq A} m(B). \tag{2}$$

The value $m(A)$ represents the strength of the subset A in any superset which it appears. Remark that the Möbius transform can attain negative values; when it is a non-negative function, it corresponds to the *basic probability mass assignment* in Dempster-Shafer theory of evidence [26].

In order to determine a fuzzy measure, $2^n - 2$ values are necessary. The number of coefficients grows exponentially with n and so does the complexity of the problem of identifying the fuzzy measure. This drawback reduces considerably the practical use of fuzzy measures. Thus, some subfamilies of fuzzy measures have been defined in an attempt to reduce complexity, prominently k-additive measures [4].

Definition 9 *[4] A fuzzy measure μ is said to be* **k-order additive** *or* **k-additive** *if its Möbius transform vanishes for any $A \subseteq X$ such that $|A| > k$ and there exists at least one subset A with exactly k elements such that $m(A) \neq 0$.*

In this sense, a probability measure is just a 1-additive measure. Thus, k-additive measures generalize probability measures, that are very restrictive in many situations. They fill the gap between probability measures and general fuzzy measures. For a k-additive measure, the number of coefficients is reduced to

$$\sum_{i=1}^{k} \binom{n}{i}.$$

More about k-additive measures can be found e.g. in [2]. We will denote the set of all k'-additive measures on X with $k' \leq k$ by $\mathcal{FM}^k(X)$; we will use the fact that $\mathcal{FM}^k(X)$ is a convex polyhedron (the proof is straightforward considering the Möbius transform). Specially appealing is the 2-additive case, that provides a generalization of probability allowing interactions while keeping a reduced complexity.

In this paper we will be studying some properties of the extreme points of the k-additive measures. The following results are known:

ISBN: 978-989-95079-6-8

Theorem 1 *[23] There are vertices of the set $\mathcal{FM}^k(X), k > 2$, that are not $\{0,1\}$-valued measures.*

Proposition 2 *[23] The set of extreme points of $\mathcal{FM}^1(X)$ (resp. $\mathcal{FM}^2(X)$) are the $\{0,1\}$-valued measures that are in $\mathcal{FM}^1(X)$ (resp. $\mathcal{FM}^2(X)$).*

We will also need some results relating the group of isometries of $\mathcal{FM}(X)$ in next section.

Definition 10 *Let $\mathcal{F}$ be a family of fuzzy measures. A surjective function $f : \mathcal{F} \to \mathcal{F}$ is an* **isometry** *if*

$$d(\mu_1, \mu_2) = d(f(\mu_1), f(\mu_2)), \forall \mu_1, \mu_2 \in \mathcal{F}.$$

Remark that an isometry is a bijective mapping on $\mathcal{F}$ and that these isometries form a group under composition of functions. It can be seen [27] that isometries also map vertices into vertices. Let us denote by $\mathcal{G}(\mathcal{F})$ the group of isometries of $\mathcal{F}$.

In [27] we have determined the group of isometries of general fuzzy measures (a result which was later generalized on [28]). We need some previous definitions.

Definition 11 *Consider $\sigma : X \to X$ a permutation on X. We define the* **symmetry induced by** *σ, denoted S_σ, the transformation on $\mathcal{FM}(X)$ such that for any $\mu \in \mathcal{FM}(X)$, the fuzzy measure $S_\sigma(\mu)$ is defined by*

$$S_\sigma(\mu)(A) = \mu(\sigma(A)), \forall A \subseteq X.$$

Definition 12 *We define the* **dual transformation***, denoted D, the transformation on $\mathcal{FM}(X)$ given by*

$$D : \begin{array}{ccc} \mathcal{FM} & \to & \mathcal{FM} \\ \mu & \mapsto & \overline{\mu} \end{array}.$$

In most cases, the group of isometries of the general fuzzy measures is generated by symmetries and the dual transformation.

Theorem 2 *If $|X| > 2$, the group $\mathcal{G}(\mathcal{FM}(X))$ is given by symmetries and compositions of symmetries with the dual application.*

In fact, it can be seen that $\mathcal{G}(\mathcal{FM}(X))$ is the semi-direct product of the group of symmetries with the cyclic group or order 2 generated by the dual transformation, i.e. the group of isometries is given by a composition of symmetries composed with either the dual application or the identity map.

Theorem 3 *If $n = 2$, the group $\mathcal{G}(\mathcal{FM}(X))$ is isomorphic to the dihedral group D_4 (the group of isometries of the square).*

Finally, we define the concept of adjacency of extreme points of $\mathcal{FM}(X)$, which will be central in our study of the vertices of $\mathcal{FM}^k(X)$.

Definition 13 *Two vertices μ_1 and μ_2 of $\mathcal{FM}(X)$ are adjacent if the convex combination (their midpoint)*

$$\mu = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2$$

can not be written as a convex combination of extreme points in any other way.

The following results have been proved in [29] and later generalized in [30].

Lemma 1 *If μ_1 and μ_2 are adjacent vertices then $\mu_1 \leq \mu_2$ or $\mu_2 \leq \mu_1$.*

However, this condition is not sufficient. The characterization of adjacency in $\mathcal{FM}(X)$ is based on the concept of $\mathcal{C}-$decomposability, that is given in next definition.

Definition 14 *Let $\mathbf{C}$ be a collection of subsets of X, μ an extreme point and let $A_1, \ldots, A_m$ be all its minimal sets. We say that μ is* **C-decomposable** *if there exists a partition of $\{A_1, \ldots, A_m\}$ in two non-empty subsets $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{A} \not\subseteq \mathbf{C}$ and $\mathbf{B} \not\subseteq \mathbf{C}$ and if $A \in \mathbf{A}$ and $B \in \mathbf{B}$ then there exists $C \in \mathbf{C}$ such that $C \subseteq A \cup B$.*

The following theorem characterizes adjacency on $\mathcal{FM}(X)$.

Theorem 4 *Let $\mu, \mu_\mathcal{C}$ be two vertices of $\mathcal{FM}(X)$ such that $\mu \geq \mu_\mathcal{C}$. Then μ and $\mu_\mathbf{C}$ are adjacent if and only if μ is not $\mathbf{C}$-decomposable.*

3 Generating the vertices of the k-additive measures

Consider the set $\mathcal{FM}(X)$. From a geometrical point of view, $\mathcal{FM}^k(X)$ is a subpolytope of $\mathcal{FM}(X)$ with additional restrictions; namely, these restrictions are $m(A) = 0, \forall |A| > k$.

Let $|A| > k$ and consider the restriction $m(A) = 0$. Then, it can be proved that the set of vertices of the polytope of fuzzy measures with this new condition consists in (see [31]):

1. Those $\{0,1\}$-valued measures for which it is $m(A) = 0$ and
2. the cut points of the hyper-plane $m(A) = 0$ with the edges of the polytope $\mathcal{FM}(X)$.

If we repeat the procedure imposing restrictions for every set A with size bigger than k we eventually obtain all the vertices of the polytope $\mathcal{FM}^k(X)$. Thus, we obtain a characterization of the extreme points of the k-additive measures (and of all intermediate polytopes). If $\mathbf{A}$ is a collection of subsets of X we denote by $\mathcal{FM}_\mathbf{A}(X)$ the set of fuzzy measures whose Möbius transform is zero on every set belonging to $\mathbf{A}$. The following theorem, well-known in the theory of polytopes, provides the new vertices when adding a new constraint.

Theorem 5 *Let $\mathbf{A}$ be a collection of subsets of X (eventually empty), B a subset of X not in $\mathbf{A}$ and μ a fuzzy measure of $\mathcal{FM}_{\mathbf{A}\cup\{B\}}(X)$ which is not a vertex of $\mathcal{FM}_\mathbf{A}(X)$. Then μ is a vertex of $\mathcal{FM}_{\mathbf{A}\cup\{B\}}(X)$ if and only if there exist μ_1 and μ_2 two adjacent vertices of $\mathcal{FM}_\mathbf{A}(X)$ not in $\mathcal{FM}_{\mathbf{A}\cup\{B\}}(X)$ and $\lambda \in (0,1)$ such that $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$.*

As a consequence of this result, for the case of $(n-1)$-additive measures we have the following:

Theorem 6 *Let μ be a $(n-1)$-additive measure which is not $\{0,1\}$-valued. Then μ is a vertex of $\mathcal{FM}^{n-1}(X)$ if and only*

$$\mu = \lambda\mu_1 + (1-\lambda)\mu_2,$$

where $\lambda \in (0,1)$ and μ_1, μ_2 are two adjacent vertices of $\mathcal{FM}(X)$ not in $\mathcal{FM}^{n-1}(X)$.

ISBN: 978-989-95079-6-8

In this particular case we can obtain all vertices of $\mathcal{FM}^{n-1}(X)$ just from adjacency relationships in $\mathcal{FM}(X)$. However, this is not possible in general; for other values of k there exist vertices not $\{0,1\}$-valued which cannot be expressed as a convex combination of exactly two $\{0,1\}$-valued vertices, as the following example shows.

Example 1 *Consider $|X| = 5$. Let μ_1 be the $\{0,1\}$-valued measure whose minimal sets are the singletons $\{1\},\ldots,\{5\}$ and μ_2 the fuzzy measure whose minimal sets are the sets of size 4. Also, let μ_3 be the fuzzy measure with minimal sets $\{(1,2),(1,3),(2,4),(3,5),(4,5)\}$.*

From Theorem 4 it follows that μ_1 and μ_2 are adjacent vertices in $\mathcal{FM}(X)$. Analogously, μ_2 and μ_3 are also adjacent vertices.

It is easy to check that

$$\frac{4}{5}\mu_1 + \frac{1}{5}\mu_2, \frac{4}{5}\mu_3 + \frac{1}{5}\mu_2 \in \mathcal{FM}^4(X). \tag{3}$$

Thus, these measures are extreme points of $\mathcal{FM}^4(X)$ and are not $\{0,1\}$-valued. Let us denote them by μ and μ' respectively. It can be checked that these measures are adjacent in $\mathcal{FM}^4(X)$. In fact, it easy to verify that their midpoint satisfies with equality exactly $2^n - 3$ linearly independent restrictions of those defining the polytope. Moreover,

$$\frac{1}{4}\mu + \frac{3}{4}\mu' \in \mathcal{FM}^3(X). \tag{4}$$

Thus, we have obtained a vertex of the 3-additive measures, which requires three vertices of $\mathcal{FM}(X)$ in order to obtain a convex combination which generates it.

As an application of Theorem 5 we have computed the vertices of the polytopes $\mathcal{FM}^k(X)$ for some values of k and n. The following tables show the number of vertices and the number of vertices being $\{0,1\}$-valued.

Table 1: Number of vertices of the $\mathcal{FM}^k(X)$ polytopes.

$n\backslash k$	1	2	3	4	5	6
1	1					
2	2	4				
3	3	9	18			
4	4	16	303	166		
5	5	25	584740	407201	7579	
6	6	36	?	?	232871070690	7828352

Table 2: Number of $\{0,1\}$-valued vertices of the $\mathcal{FM}^k(X)$ polytopes.

$n\backslash k$	1	2	3	4	5	6
1	1					
2	2	4				
3	3	9	18			
4	4	16	68	166		
5	5	25	195	1855	7579	
6	6	36	456	10986	1322954	7828352

To compute these numbers for the cases with $n \leq 5$ we repeatedly use Theorem 5 to calculate the vertices of the intermediate polytopes. This process is very time-consuming and, in fact, took several weeks in a 2.1GHz PC to complete the calculations in the case $n = 5$. One of the main difficulties is that we lack criteria to quickly test adjacency (except for general fuzzy measures, where we can use Theorem 4). The other main problem is that the number of extreme points grows very fast, making the number of comparisons needed grow even faster. All this makes unfeasible a direct approach to calculate the number of vertices of the k-additive measures when $n = 6$.

However, it is possible to use the results on the isometries of the polytope of fuzzy measures (see Section 2) to at least obtain the number of vertices of the 5-additive measures when $n = 6$. Just observe that if σ is a permutation of X then S_σ, the symmetry induced by σ (Definition 11), is an isometry that fixes u_X and if μ is a fuzzy measure then $m_\mu(X)$ and $m_{S_\sigma(\mu)}(X)$ are both either positive, negative or zero, by Eq. (1). Also, as S_σ is an isometry, if μ_1 and μ_2 are adjacent vertices, then so are $S_\sigma(\mu_1)$ and $S_\sigma(\mu_2)$. Consequently, if $\lambda\mu_1 + (1-\lambda)\mu_2$ is a vertex of the $n-1$-additive measures then so is $\lambda S_\sigma(\mu_1) + (1-\lambda)S_\sigma(\mu_2)$.

Take μ, μ' two $\{0,1\}$-valued measures such that there exists a symmetry S_σ for which $S_\sigma(\mu) = \mu'$. If μ_1 is a $\{0,1\}$-valued measure adjacent to μ and $\alpha \in (0,1)$ such that

$$\alpha\mu + (1-\alpha)\mu_1 \in \mathcal{FM}^{n-1}(X),$$

then

$$S_\sigma(\alpha\mu + (1-\alpha)\mu_1) = \alpha S_\sigma(\mu) + (1-\alpha)S_\sigma(\mu_1)$$

also belongs to $\mathcal{FM}^{n-1}(X)$. Consequently, every vertex of $\mathcal{FM}^{n-1}(X)$ obtained from μ' is the image under S_σ of another vertex obtained from μ.

Thus, if μ and μ' are taken one into the other by some symmetry S_σ (i.e., they are in the same orbit under the action of the group of symmetries), then they will generate exactly the same number of vertices of $\mathcal{FM}^{n-1}(X)$. Therefore, in order to count the number of vertices of the $n-1$-additive measures, we can proceed as follows:

1. List all the vertices of the general fuzzy measures.
2. Classify these vertices according to the sign of their Möbius transform on X.
3. Calculate the orbits of the vertices with positive Möbius on X.
4. Pick a representative of each orbit and test its adjacency to the vertices with negative Möbius on X.
5. Multiply the number of adjacent vertices by the size of the orbit, and sum all the values obtained.
6. Add to this value the number of vertices whose Möbius transform on X is 0.

After applying these methods in the case where $n = 6$ we obtained 1322954 vertices with zero Möbius transform on X,

ISBN: 978-989-95079-6-8

3252699 with negative Möbius and also 3252699 with positive Möbius. These last vertices are organized into 6806 different orbits. After 20 hours of computation on a 2.1GHz PC, the number of non $\{0,1\}$-valued vertices of $\mathcal{FM}^5(X)$ with $|X| = 6$ was found to be 232869747736, for a total of 232871070690 vertices.

The figures in Tables 1 and 2 seem to suggest that the number of non $\{0,1\}$-valued vertices of $\mathcal{FM}^k(X)$ for $k = 3, ..., n-1$ grows much more quickly than the number of extreme points of $\mathcal{FM}(X)$. In the following Section we show that, in fact, this is the case when $k = n-1$.

4 On the number of vertices of the $n-1$-additive measures

In this section we study the asymptotic behavior of the ratio of the number of vertices of $\mathcal{FM}^{n-1}(X)$ to the number of vertices of $\mathcal{FM}(X)$. We will denote by D_n the number of vertices of $\mathcal{FM}(X)$ (with $n = |X|$), by A_n the number of vertices of $\mathcal{FM}^{n-1}(X)$, and by B_n the number of vertices of $\mathcal{FM}(X)$ which are not in $\mathcal{FM}^{n-1}(X)$ and whose minimal sets all have size at least $\lceil \frac{n-3}{2} \rceil$.

Lemma 2 *It holds that*

$$\lim_{n\to\infty} \frac{B_n}{D_n} > 0$$

Sketch of Proof:

The proof of the Theorem rests on considering, for even n, the set M_0^n of vertices μ of $\mathcal{FM}(X)$ such that:

1. every minimal set of μ has size $\frac{n}{2}-1$, $\frac{n}{2}$ or $\frac{n}{2}+1$,
2. $\mu(A) = 1$ for every A of size greater than $\frac{n}{2}+1$,
3. $\mu(A) = 0$ for every A of size less than $\frac{n}{2}-1$,
4. the number of minimal sets of μ of size $\frac{n}{2}-1$ is at most $2^{\frac{n}{2}}$.

In [32] it is proven that the ratio of the number of measures in M_0^n to the total number of measures does not tend to zero when n tends to infinity. It can also be seen that the ratio of non $(n-1)$-additive measures in M_0^n does not tend to 0 when n tends to infinity and the Lemma follows (for even n).

For odd n, the argument is similar, but considering the set M_1^n of vertices μ of $\mathcal{FM}(X)$ such that:

1. every minimal set of μ has size $\frac{n-3}{2}$, $\frac{n-1}{2}$ or $\frac{n+1}{2}$,
2. $\mu(A) = 1$ for every A of size greater than $\frac{n+1}{2}$,
3. $\mu(A) = 0$ for every A of size less than $\frac{n-3}{2}$,
4. the number of minimal sets of μ of size $\frac{n-3}{2}$ is at most $2^{\frac{n}{2}}$.

■

We can now state the main Theorem of the section.

Theorem 7 *There exist $k > 0$ and n_0 such that*

$$A_n > k\frac{2^n}{\sqrt{n}} D_n$$

for all $n \geq n_0$.

Sketch of Proof:

Denote $q = \lceil \frac{n-3}{2} \rceil$. Consider n such that $q > 4$. Consider $\mathcal{P}_1$ the collection of all partitions P of X such that

1. exactly one set in P has size $q-2$,
2. the rest of the sets in P are singletons,

and $\mathcal{P}_2$ the collection of all partitions P of X such that

1. exactly one set in P has size $q-3$,
2. the rest of the sets in P are singletons.

Consider μ a vertex of $\mathcal{FM}(X)$ not in $\mathcal{FM}^{n-1}(X)$ whose minimal sets all have size at least q. Consider P in $\mathcal{P}_1 \cup \mathcal{P}_2$ and μ' the measure whose collection of minimal sets is P. It is possible to show, with the help of Theorem 4, that μ and μ' are adjacent.

The size of $\mathcal{P}_1$ is $\binom{n}{q-2}$ and the size of $\mathcal{P}_2$ is $\binom{n}{q-3}$. Also, the sizes of the partitions in $\mathcal{P}_1$ and of the partitions in $\mathcal{P}_2$ have different parity, so their Möbius transform on X will be 1 in one case and -1 in the other (see Proposition 4 in [23]). Since μ is not $(n-1)$-additive and it is adjacent to every μ' whose minimal sets are in $\mathcal{P}_1 \cup \mathcal{P}_2$, then the convex combinations of μ with all such μ' generate at least $\binom{n}{q-3}$ vertices of $\mathcal{FM}^{n-1}(X)$. It follows that

$$A_n \geq \binom{n}{q-3} B_n$$

for all n such that $q > 4$.

From Lemma 2 we know that there exist $k_1 > 0$ and n_1 such that

$$\frac{B_n}{D_n} > k_1$$

for all $n \geq n_1$. Then, it is easy to prove that there exists $k > 0$ such that for n big enough

$$A_n \geq k\frac{2^n}{\sqrt{n}} D_n$$

■

This Theorem shows that the growth of the number of vertices of $\mathcal{FM}^{n-1}(X)$ is much faster than that of $\mathcal{FM}(X)$. Theorem 5 together with the values on Table 1 hint that this may be the case also for other polytopes $\mathcal{FM}^k(X)$ with k big enough. This suggests that, despite the simple interpretation, k-additive measures have a complex structure and the choice of the value of k should be made with care.

5 Conclusions and open problems

In this paper we have studied the vertices of the polytope of k-additive measures. We have provided a description of these vertices and a algorithm to compute them. We have also shown some examples of the results of the computation of this algorithm. Finally, we have given a lower bound on the number of vertices of the polytope of fuzzy measures which are at most $(n-1)$-additive, showing that this number grows much faster than the number of vertices of the general fuzzy measures. Similar results for other values of k seem plausible and deeper investigation of the asymptotical behavior on

ISBN: 978-989-95079-6-8

those cases might provide further insights on the polytopes $\mathcal{FM}^k(X)$.

The results presented here show that, in general, the structure of $\mathcal{FM}^k(X)$ is more complex than that of the general fuzzy measures and of the p-symmetric measures (cf.[29, 30]). For instance, knowledge of a wide number of intermediate polytopes was needed in order to compute the extreme points of the k-additive measures, making this computation very time consuming. A study of conditions for the adjacency in the polytope of k-additive measures (and in the intermediate ones) could help to decrease this computing time. Knowing the group of isometries of the polytopes may also be useful.

We also intend to investigate subfamilies of the k-additive measures which retain the modeling power of these measures. For instance, it is interesting to restrict the vertices to those which are $\{0, 1\}$-valued and study the resulting polytope.

Finally, in order to obtain a suitable initial population (for measure identification with a genetic algorithm) it would be useful to study methods for the random generation of k-additive measures.

Acknowledgements

This research has been supported in part by grant numbers MTM2007-61193 and CAM- UCM910707, and by MEC and FEDER grant TIN2007-61273.

References

[1] M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5:587–607, 1997.

[2] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, (92):167–189, 1997.

[3] L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the theory of Games*, volume II of *Annals of Mathematics Studies*, pages 307–317. Princeton University Press, 1953.

[4] M. Grabisch. k-order additive discrete fuzzy measures. In *Proceedings of 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 1345–1350, Granada (Spain), 1996.

[5] P. L. Hammer and R. Holzman. On approximations of pseudo-boolean functions. *Zeitschrift für Operations Research. Mathematical Methods of Operations Research*, (36):3–21, 1992.

[6] G. C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, (2):340–368, 1964.

[7] M. Sugeno and T. Terano. A model of learning based on fuzzy information. *Kybernetes*, (6):157–166, 1977.

[8] J.-L. Marichal. k-intolerant capacities and Choquet integrals. In *Proceedings of Tenth International Conference of Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, pages 601–608, Perugia (Italy), July 2004.

[9] J.-L. Marichal. Tolerant or intolerant character of interacting criteria in aggregation by the Choquet integral. *European Journal of Operational Research*, 155(3):771–791, 2004.

[10] P. Miranda, M. Grabisch, and P. Gil. p-symmetric fuzzy measures. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 (Suppl.):105–123, 2002.

[11] D. Dubois and H. Prade. A class of fuzzy measures based on triangular norms. *Int. J. General Systems*, 8:43–61, 1982.

[12] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, (5):131–295, 1953.

[13] P. Miranda, M. Grabisch, and P. Gil. Axiomatic structure of k-additive capacities. *Mathematical Social Sciences*, 49:153–178, 2005.

[14] C. A. Bana e Costa and J.-C. Vansnick. MACBETH- An interactive path towards the construction of cardinal value function. *International Transactions in Operational Research*, 1(4):489–500, 1994.

[15] J.-L. Marichal, P. Meyer, and M. Roubens. Sorting multiattribute alternatives : the TOMASO method. *Computer and Operations Resarch*, 32(4):861–877, 2005.

[16] M. Grabisch and J.-M. Nicolas. Classification by fuzzy integral-performance and tests. *Fuzzy Sets and Systems, Special Issue on Pattern Recognition*, (65):255–271, 1994.

[17] J. Wang Z. Wang, K. Xu and G. Klir. Using genetic algorithms to determine nonnegative monotone set functions for information fusion in environments with random perturbation. *International Journal of Intelligent Systems*, 14:949–962, 1999.

[18] E. F. Combarro and P. Miranda. Identification of fuzzy measures from sample data with genetic algorithms. *Computers and Operations Research*, 33(10):3046–3066, 2006.

[19] J. H. Holland. *Adaptation in natural and artificial systems*. Ann Arbor: The University of Michigan Press, 1975.

[20] D. Denneberg. *Non-additive measures and integral*. Kluwer Academic, 1994.

[21] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[22] D. Radojevic. The logical representation of the discrete Choquet integral. *Belgian Journal of Operations Research, Statistics and Computer Science*, 38(2–3):67–89, 1998.

[23] P. Miranda, E.F. Combarro, and P. Gil. Extreme points of some families of non-additive measures. *European Journal of Operational Research*, 33(10):3046–3066, 2006.

[24] J. H. van Lint and R. M. Wilson. *A course in Combinatorics*. Cambridge University Press, 1998.

[25] A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, (17):263–283, 1989.

[26] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, (USA), 1976.

[27] P. Miranda and E.F. Combarro. On the structure of some families of fuzzy measures. *IEEE Transactions on Fuzzy Systems*, 15(6):1068–1081, 2007.

[28] E. F. Combarro and P. Miranda. Adjacency on the order polytope with applications to the theory of fuzzy measures. *Submitted*.

[29] E. F. Combarro and P. Miranda. On the polytope of non-additve measures. *Fuzzy Sets and Systems*, 159(16):2145–2162, 2008.

[30] E. F. Combarro and P. Miranda. Characterizing the isometries on the order polytope with an application to the theory of fuzzy measures. *Submitted*.

[31] K. G. Murty. Adjacency on convex polyhedra. *SIAM Review*, 13(3):377–386, 1971.

[32] A. D. Korshunov. Monotone boolean functions. *Russian Mathematical Survey*, 58(5):929–1001, 2003.

Type-2 Fuzzy Trust Aggregation

John T. Rickard[1] Sam Hamilton[2] Wendy Hamilton[3]

1 Distributed Infinity, Inc., 4637 Shoshone Drive,
Larkspur, CO 80118, USA
2, 3 Distributed Infinity, Inc., 4033 Nabal Drive
La Mesa, CA 91941, USA
Email: {trickard, shamilton, whamilton@distributedinfinity.com}

Abstract— *Trust is a fundamental concept that is critical in human decision processes in almost all domains, but of particular relevance in the domain of computer security. In many computer systems deployed today, trust is only modeled indirectly through a serious of formal rules and regulations describing when privileges are to be granted or revoked. Recently, research has been conducted into the area of fuzzy trust modeling in order to allow a more intelligent tradeoff analysis by the computer security software. In this paper, we extend this work to support Type-2 fuzzy reasoning. The Type-2 input membership functions can be derived from multiple human experts' judgments. This allows us to incorporate the inherent imprecision in trust judgments in a mathematically principled manner.*

Keywords— Computer security, fuzzy systems, trust modeling, Type-2 fuzzy logic.

1 Introduction

In computer security, trust is described as the belief in the competence of an entity to act dependably, reliably and securely within a specific context, and distrust is described as the belief in the incompetence of an entity to act dependably, securely and reliably within a specific context [1]. In many computer systems deployed today, trust is only modeled indirectly through a serious of formal rules and regulations describing when privileges are to be granted or revoked. Manual human judgment is invoked whenever the system falls into an ambiguous state. For example, if a user tries to log on with an incorrect password too many times, many systems will revoke privileges for that user account and rely on the system administrator for the judgment of when it is appropriate to reissue those privileges.

This approach has significant weaknesses in that its strong reliance on a human in the decision process requires the Boolean security rule set to make inappropriately crisp statements, which either overly restrict users (reducing productivity) or are not restrictive enough (reducing security). Recently, research has been conducted into augmenting policy rules with Type-1 fuzzy trust modeling to help address this issue by softening the hard and fast set of computer security policy rules to allow a more intelligent tradeoff analysis [2-7].

The notion of trust relationships is traditionally modeled in a fuzzy sense, i.e., "entity A trusts entity B to a degree $x \in [0,1]$," where x is usually determined by the historical results of entity $A's$ interactions with entity B. Trust chains may also be involved in this determination, e.g., entity A can consult entity C as to its trust in entity B and factor that information into its decision. A discrete set K of contexts may also enter into this description, resulting in an ordered pair (x, y) describing trust, indicating that entity A trusts entity B to a degree $x \in [0,1]$ in a particular context $y \in K$.

The contribution of each context factor may be measured independently using two numbers, a *confidence* $c \in [-1,1]$ and a *plausibility* $p \in [0,1]$. The confidence c describes entity $A's$ belief in a particular outcome either *occurring* or *not occurring* (we use negative values to indicate the latter) in an interaction with entity B in a given context, and the plausibility p describes the degree to which entity A considers her confidence assessment to be reliable. For example, entity A might believe with high confidence (e.g., 0.8) in a particular outcome's occurrence, but if this belief is based upon little or no historical evidence, then her plausibility might be low (say, 0.1). Upon the accumulation of further direct evidence as a result of her interactions with entity B and/or consultation with other entities, she may simultaneously refine her confidence estimate and increase her plausibility in this refined confidence. Thus we treat these as independent variables.

Hamilton and Hamilton [8] developed an algebra for combining such assessments, based upon *singleton values* of input confidences and plausibilities, and producing a corresponding singleton aggregate confidence and plausibility. However, it is well known that human assessments of situations are pervasively imprecise, and thus humans are more reliable at estimating *intervals* of confidence or plausibility rather than point values. Likewise, the results of computer simulations are better described by an *interval* of outcomes than by a single number. Furthermore, it is often desirable to solicit such intervals from multiple experts, or to perform multiple simulations. In this case, the problem remains as to how to incorporate this multiplicity of interval inputs into the trust system. Multiple interval inputs can be combined into interval Type-2 input membership functions using the techniques described in [17].

In this paper, we extend the results of [8] initially to individual interval inputs of confidence and plausibility, to compute corresponding interval output membership functions for their aggregation. Then, employing α-cuts, we further extend these results to accommodate Type-1 and

ISBN: 978-989-95079-6-8

interval Type-2 fuzzy inputs. This allows us to propagate imprecision in our knowledge of the inputs to the corresponding imprecision in our knowledge of the output in a mathematically principled fashion.

2 Trust aggregation

Suppose that we have two assessments of confidence and plausibility relating to our trust in a particular context factor, which we denote by A_{c_1,p_1} and A_{c_2,p_2}, respectively, where c_1 and c_2 denote the confidences, and p_1 and p_2 the plausibilities, of the corresponding factor. Suppose further that we wish to perform an aggregation of these values, denoted by the operator "$\wedge$", to yield a sensible aggregate confidence and plausibility based upon these two inputs. Thus we wish to describe $A_{c_3,p_3} = A_{c_1,p_1} \wedge A_{c_2,p_2}$ in terms of its aggregate confidence c_3 and plausibility p_3.

Hamilton and Hamilton [8] propose the following algebraic computations for c_3 and p_3:

$$c_3 = \frac{c_1 p_1^{\alpha} + c_2 p_2^{\alpha}}{p_1^{\alpha} + p_2^{\alpha}} \tag{1}$$

$$p_3 = \frac{(p_1 + p_2 - p_1 p_2)\left(2 - |c_1 - c_2|\right)^{\beta} + (p_1 + p_2 - 2p_1 p_2)|c_1 - c_2|^{\beta}}{\left(2 - |c_1 - c_2|\right)^{\beta} + |c_1 - c_2|^{\beta}} \tag{2}$$

The parameters $\alpha > 0$ and $\beta > 0$ govern the degree of influence of the input plausibilities p_1 and p_2 upon the aggregate confidence c_3, and the degree of influence of the *differences* in the two input confidences c_1 and c_2 upon the aggregate plausibility p_3, respectively. Typical values for these parameters might be $\alpha = 2$ and $\beta = 1/4$. It is straightforward to show that when the domains of the inputs are $c_1, c_2 \in [-1,1]$ and $p_1, p_2 \in [0,1]$, the resulting ranges for the aggregate confidence and plausibility are likewise $c_3 \in [-1,1]$ and $p_3 \in [0,1]$.

From (1), we see that the aggregate confidence c_3 is a simple convex combination of the input confidences, with weights proportional to the α-power of the plausibilities. The aggregate plausibility p_3 is a more complicated convex combination of two functions of the input plausibilities. At one extreme, when the input confidences are in perfect agreement, the aggregate plausibility is given by $p_3 = p_1 + p_2 - p_1 p_2$ $(c_1 = c_2)$. At the other extreme where the confidences are diametrically opposite and of unity magnitude, the aggregate plausibility is given by $p_3 = p_1 + p_2 - 2p_1 p_2$, $(c_1 = \pm 1,\ c_2 = -c_1)$.

To illustrate further the behaviors of (1) and (2), consider the following examples, where $\alpha = 2$ and $\beta = 1/4$:

$$A_{0.8,0} \wedge A_{0.3,0.5} = A_{0.3,0.5} \tag{3}$$

$$A_{-0.8,0.9} \wedge A_{0.8,0.9} = A_{0,0.5} \tag{4}$$

$$A_{0.8,0.3} \wedge A_{0.7,0.4} = A_{0.74,0.54} \tag{5}$$

In the first case, an assessment with zero plausibility adds no new information. In the second case, strongly opposing confidences reduce aggregate plausibility. In the final case, similar confidences reinforce aggregate plausibility.

A single fuzzy scalar value of our degree of trust $t\left(A_{c,p}\right)$ can be computed as the following convex combination involving confidence and plausibility:

$$t\left(A_{c,p}\right) = \frac{0.5(1-p)^{\gamma} + \frac{(1+c)}{2} p^{\gamma}}{(1-p)^{\gamma} + p^{\gamma}} \tag{6}$$

where $\gamma > 0$ controls the degree of influence of the plausibility upon the trust. A typical value might be $\gamma = 2$. Note that if $p = 0$, then $t\left(A_{c,p}\right) = 0.5$, i.e., we have a maximally ambiguous degree of trust (i.e., the "coin toss" situation), whereas if $p = 1$, then $t\left(A_{c,p}\right) = \frac{1+c}{2}$, (i.e., trust is equivalent to a rescaling of the confidence c to lie in the interval [0,1].) In the extremes of this latter case, if $c = -1$, we have zero trust in the given outcome, while if $c = 1$, we have complete trust in the given outcome.

3 Interval inputs

Suppose now that, rather than the point values of the inputs described above, we instead have interval values. Interval inputs might naturally arise from soliciting them from a human expert, or as the results of a computer simulation that includes the uncertainty of the simulations results. Specifically, let $c_1 \in [c_{1\ell}, c_{1r}]$, $c_2 \in [c_{2\ell}, c_{2r}]$, $p_1 \in [p_{1\ell}, p_{1r}]$ and $p_2 \in [p_{2\ell}, p_{2r}]$, where the subscripts ℓ and r refer to the left- and right-hand endpoints of the intervals, respectively. Bear in mind that the c intervals are subsets of $[-1,1]$, and the p intervals are subsets of $[0,1]$.

3.1 Confidence aggregation for interval inputs

From (1), c_3 is an interval weighted average [9], where the input variables take values continuously over interval ranges. Thus c_3 takes values continuously over an interval range, which we denote by $c_3 \in [c_{3\ell}, c_{3r}]$, where

$$c_{3\ell} = \min_{c_1,c_2,p_1,p_2} \frac{c_1 p_1^{\alpha} + c_2 p_2^{\alpha}}{p_1^{\alpha} + p_2^{\alpha}} = \min_{p_1,p_2} \frac{c_{1\ell} p_1^{\alpha} + c_{2\ell} p_2^{\alpha}}{p_1^{\alpha} + p_2^{\alpha}}, \tag{7}$$

$$c_{3r} = \max_{c_1,c_2,p_1,p_2} \frac{c_1 p_1^{\alpha} + c_2 p_2^{\alpha}}{p_1^{\alpha} + p_2^{\alpha}} = \max_{p_1,p_2} \frac{c_{1r} p_1^{\alpha} + c_{2r} p_2^{\alpha}}{p_1^{\alpha} + p_2^{\alpha}}. \tag{8}$$

It is well-known [10] how to solve this simple case of a two-element interval weighted average. We first re-index the triplets $(c_{i\ell}, p_{i\ell}, p_{ir})$ in order of increasing $c_{i\ell}$, i.e., to obtain $c_{1\ell} \le c_{2\ell}$. Then using these re-indexed triplets, $c_{3\ell}$ is given by

$$c_{3\ell} = \frac{c_{1\ell} p_{1r}^{\alpha} + c_{2\ell} p_{2\ell}^{\alpha}}{p_{1r}^{\alpha} + p_{2\ell}^{\alpha}}, \tag{9}$$

which is recognized as a closed-form two-element solution of the Karnik-Mendel (KM) algorithm [10], which is used more generally to find the endpoints of the output interval corresponding to an $n-$element interval weighted average. In similar fashion, after re-indexing the triplets $(c_{ir}, p_{i\ell}, p_{ir})$ in order of increasing c_{ir}, we have

ISBN: 978-989-95079-6-8

$$c_{3r} = \frac{c_{1r} p_{1\ell}^{\alpha} + c_{2r} p_{2r}^{\alpha}}{p_{1\ell}^{\alpha} + p_{2r}^{\alpha}}. \tag{10}$$

Thus we have a closed-form analytical solution for the output interval of $c_3 \in [c_{3\ell}, c_{3r}]$ in (1) when the inputs have interval values.

3.2 Plausibility aggregation for interval inputs

From (2), p_3 is a continuous function of its input values, so it too will take values over an interval $p_3 \in [p_{3\ell}, p_{3r}]$. By inspection of (2), we may re-write this equation as

$$p_3 = \lambda(p_1 + p_2 - p_1 p_2) + (1-\lambda)(p_1 + p_2 - 2p_1 p_2) \tag{11}$$

where $\lambda \in [0,1]$ is a function of c_1 and c_2 given by

$$\lambda(c_1, c_2) = \frac{\left(2 - |c_1 - c_2|\right)^{\beta}}{\left(2 - |c_1 - c_2|\right)^{\beta} + |c_1 - c_2|^{\beta}}. \tag{12}$$

Taking partial derivatives in (11) with respect to p_1 and p_2, we obtain

$$\frac{\partial p_3}{\partial p_1} = 1 - (2-\lambda) p_2 \quad \begin{cases} > 0 \ \textit{for} \ p_2 < \dfrac{1}{2-\lambda} \\ < 0 \ \textit{for} \ p_2 > \dfrac{1}{2-\lambda} \end{cases}, \tag{13}$$

and

$$\frac{\partial p_3}{\partial p_2} = 1 - (2-\lambda) p_1 \quad \begin{cases} > 0 \ \textit{for} \ p_1 < \dfrac{1}{2-\lambda} \\ < 0 \ \textit{for} \ p_1 > \dfrac{1}{2-\lambda} \end{cases}. \tag{14}$$

Given values of c_1 and c_2, and hence a given value of λ, we see that p_3 is monotonically increasing or decreasing as a function of p_1 and p_2 away from the saddle point $\left(\frac{1}{2-\lambda}, \frac{1}{2-\lambda}\right)$, which lies in the upper right-hand quadrant of the unit square since $\lambda \in [0,1]$. Thus the minimum (maximum) of (11) as a function of p_1 and p_2 always occurs at one of the four input interval extremal points $(p_{1\ell}, p_{2\ell})$, $(p_{1\ell}, p_{2r})$, $(p_{1r}, p_{2\ell})$ or (p_{1r}, p_{2r}).

To continue this analysis, let us define four functions $f_{ij}(c_1, c_2)$, $i \in \{\ell, r\}$, $j \in \{\ell, r\}$ such that

$$f_{ij}(c_1, c_2) = \frac{(p_{1i} + p_{2j} - p_{1i} p_{2j})\left(2 - |c_1 - c_2|\right)^{\beta} + (p_{1i} + p_{2j} - 2p_{1i} p_{2j})|c_1 - c_2|^{\beta}}{\left(2 - |c_1 - c_2|\right)^{\beta} + |c_1 - c_2|^{\beta}} \tag{15}$$

It then remains to calculate

$$p_{3\ell} = \min_{i,j \in \{\ell, r\}} \left\{ \min_{c_1, c_2 \in [-1,1]} \frac{\pi_{ij}^{(1)}\left(2 - |c_1 - c_2|\right)^{\beta} + \pi_{ij}^{(2)} |c_1 - c_2|^{\beta}}{\left(2 - |c_1 - c_2|\right)^{\beta} + |c_1 - c_2|^{\beta}} \right\} \tag{16}$$

$$p_{3r} = \max_{i,j \in \{\ell, r\}} \left\{ \max_{c_1, c_2 \in [-1,1]} \frac{\pi_{ij}^{(1)}\left(2 - |c_1 - c_2|\right)^{\beta} + \pi_{ij}^{(2)} |c_1 - c_2|^{\beta}}{\left(2 - |c_1 - c_2|\right)^{\beta} + |c_1 - c_2|^{\beta}} \right\} \tag{17}$$

where

$$\pi_{ij}^{(1)} = p_{1i} + p_{2j} - p_{1i} p_{2j} \tag{18}$$

$$\pi_{ij}^{(2)} = p_{1i} + p_{2j} - 2p_{1i} p_{2j} \tag{19}$$

To this end, we employ the continuous change of variables $u = |c_1 - c_2|$, which reduces these equations to:

$$p_{3\ell} = \min_{i,j \in \{\ell, r\}} \left\{ \min_{u \in [0,2]} \frac{\pi_{ij}^{(1)}\left(2-u\right)^{\beta} + \pi_{ij}^{(2)} u^{\beta}}{\left(2-u\right)^{\beta} + u^{\beta}} \right\} \triangleq \min_{i,j \in \{\ell, r\}} \left\{ \min_{u \in [0,2]} g_{ij}^{(\min)}(u) \right\} \tag{20}$$

$$p_{3r} = \max_{i,j \in \{\ell, r\}} \left\{ \max_{u \in [0,2]} \frac{\pi_{ij}^{(1)}\left(2-u\right)^{\beta} + \pi_{ij}^{(2)} u^{\beta}}{\left(2-u\right)^{\beta} + u^{\beta}} \right\} \triangleq \max_{i,j \in \{\ell, r\}} \left\{ \max_{u \in [0,2]} g_{ij}^{(\max)}(u) \right\} \tag{21}$$

It is obvious by inspection of (18) and (19) that for any of the four combinations of $(i,j): i,j = 1,2$, $\pi_{ij}^{(1)} \geq \pi_{ij}^{(2)}$, so from (20) and (21) we observe that each of the functions $g_{ij}^{(\min)}(u)$ and $g_{ij}^{(\max)}(u)$ is a convex combination (whose coefficients are nonlinear functions of u over the interval $[0,2]$) of the values $\pi_{ij}^{(1)}$ and $\pi_{ij}^{(2)}$, having its maximum at $u = 0$ and its minimum at $u = 2$. We can show that the derivatives of these functions are uniformly non-positive over the interval $[0,2]$, and thus the functions $g_{ij}^{(\min)}(u)$ and $g_{ij}^{(\max)}(u)$ are uniformly non-increasing over any sub-interval of $[0,2]$. Therefore their minimum with respect to u in (20) over an interval $[u_\ell, u_r] \subseteq [0,2]$ will correspond to u_r, the right endpoint point of this interval, and their maximum with respect to u in (21) will correspond to u_ℓ, the left endpoint of this interval.

To determine the associated interval for u from the intervals $c_1 \in [c_{1\ell}, c_{1r}]$ and $c_2 \in [c_{2\ell}, c_{2r}]$, we must consider six cases of different interval overlap scenarios, depicted in Table 1.

Table 1: Interval overlap cases.

Case	Interval Positions	Interval for $u = \lvert c_1 - c_2 \rvert$
1	$c_{1r} < c_{2\ell}$	$u \in [c_{2\ell} - c_{1r}, c_{2r} - c_{1\ell}]$
2	$c_{1\ell} \leq c_{2\ell} \leq c_{1r} \leq c_{2r}$	$u \in [0, c_{2r} - c_{1\ell}]$
3	$c_{1\ell} \leq c_{2\ell} \leq c_{2r} \leq c_{1r}$	$u \in [0, \max(c_{1r} - c_{2\ell}, c_{2r} - c_{1\ell})]$
4	$c_{2\ell} \leq c_{1\ell} \leq c_{2r} \leq c_{1r}$	$u \in [0, c_{1r} - c_{2\ell}]$
5	$c_{2\ell} \leq c_{2r} \leq c_{1\ell} \leq c_{1r}$	$u \in [c_{1\ell} - c_{2r}, c_{1r} - c_{2\ell}]$
6	$c_{2\ell} \leq c_{1\ell} \leq c_{1r} \leq c_{2r}$	$u \in [0, \max(c_{2r} - c_{1\ell}, c_{1r} - c_{2\ell})]$

As noted above, in all cases, $[u_\ell, u_r] \subseteq [0,2]$, and due to the monotone non-increasing nature of the functions $g_{\min}(u)$ and $g_{\max}(u)$, we finally have

ISBN: 978-989-95079-6-8

$$p_{3\ell} = \min_{i,j\in\{\ell,r\}} \left\{ \min_{u\in[u_\ell,u_r]} g_{ij}^{(\min)}(u) \right\} = \min_{i,j\in\{\ell,r\}} \left\{ \frac{\pi_{ij}^{(1)}\left(2-u_r\right)^\beta + \pi_{ij}^{(2)}u_r^\beta}{\left(2-u_r\right)^\beta + u_r^\beta} \right\}, \quad (22)$$

$$p_{3r} = \max_{i,j\in\{\ell,r\}} \left\{ \max_{u\in[u_\ell,u_r]} g_{ij}^{(\max)}(u) \right\} = \max_{i,j\in\{\ell,r\}} \left\{ \frac{\pi_{ij}^{(1)}\left(2-u_\ell\right)^\beta + \pi_{ij}^{(2)}u_\ell^\beta}{\left(2-u_\ell\right)^\beta + u_\ell^\beta} \right\}. \quad (23)$$

Thus to find the interval boundaries for p_3, we simply must find the minimum (maximum) of the four function evaluations involved in (22) and (23).

3.3 Interval calculations for trust

As a final step in extending our results to interval values, we consider the trust function of equation (6) in the event that $c\in[c_\ell,c_r]$ and $p\in[p_\ell,p_r]$ are described by interval values. Since c appears only in the numerator, it is obvious by inspection that

$$\min t\left(A_{c,p}\right) = \min_{p\in[p_\ell,p_r]} \frac{0.5(1-p)^\gamma + \frac{(1+c_\ell)}{2}p^\gamma}{(1-p)^\gamma + p^\gamma} \quad (24)$$

$$\max t\left(A_{c,p}\right) = \max_{p\in[p_\ell,p_r]} \frac{0.5(1-p)^\gamma + \frac{(1+c_r)}{2}p^\gamma}{(1-p)^\gamma + p^\gamma} \quad (25)$$

Analogous to the functions $g_{ij}^{(\min)}(u)$ and $g_{ij}^{(\max)}(u)$ of the previous sections, we are here dealing with the minimum or maximum of a nonlinear convex combination of the values 0.5 and $(1+c)/2$ over the interval $p\in[p_\ell,p_r]$. We can show that the derivative of (6) with respect to p is monotonic, so that the minimum (maximum) of the above expressions occurs either at $p=p_\ell$ or $p=p_r$ depending upon the value of c (for $c<0$ the derivative is negative and for $c>0$ it is positive), with the result that $t\in[t_\ell,t_r]$, where

$$t_\ell = \min_{p\in[p_\ell,p_r]} t\left(A_{c,p}\right) = \begin{cases} \dfrac{0.5(1-p_r)^\gamma + \frac{(1+c_\ell)}{2}p_r^\gamma}{(1-p_r)^\gamma + p_r^\gamma}, & c_\ell < 0 \\ \dfrac{0.5(1-p_\ell)^\gamma + \frac{(1+c_\ell)}{2}p_\ell^\gamma}{(1-p_\ell)^\gamma + p_\ell^\gamma}, & c_\ell \geq 0 \end{cases} \quad (26)$$

$$t_r = \max_{p\in[p_\ell,p_r]} t\left(A_{c,p}\right) = \begin{cases} \dfrac{0.5(1-p_\ell)^\gamma + \frac{(1+c_r)}{2}p_\ell^\gamma}{(1-p_\ell)^\gamma + p_\ell^\gamma}, & c_r < 0 \\ \dfrac{0.5(1-p_r)^\gamma + \frac{(1+c_r)}{2}p_r^\gamma}{(1-p_r)^\gamma + p_r^\gamma}, & c_r \geq 0 \end{cases} \quad (27)$$

4 Type-1 fuzzy inputs

Suppose that instead of interval membership functions, the fuzzy membership functions for c_1, c_2, p_1 and p_2 in equations (1) and (2) are general convex Type-1 membership functions $\mu_{c_1}(x)$, $\mu_{c_2}(x)$, $\mu_{p_1}(x)$ and $\mu_{p_2}(x)$, respectively, which may not necessarily be normal (i.e., the maximum degree, or height, of the membership function is not necessarily unity). The approach we take in this case is similar to that used to analyze fuzzy weighted averages [9] via α-cuts of the input membership functions, as corrected in [10]. (Note that the term "α-cuts", as commonly used in the fuzzy systems literature, bears no relationship to the α exponent used in equation (1).)

We recall that that an α-cut of a convex fuzzy membership function $\mu(x)$ is an interval on the independent variable axis defined by

$$\left[x(\alpha)_\ell, x(\alpha)_r\right] = \left\{x : \mu(x) \geq \alpha\right\}. \quad (28)$$

We may then represent $\mu(x)$ in terms of its α-cuts by [11]

$$\mu(x) = \sup_{\alpha\in[0,1]} \alpha I_{\mu_\alpha}(x), \quad (29)$$

where $I_{\mu_\alpha}(x)$ is the indicator function for the α-cuts, i.e.,

$$I_{\mu_\alpha}(x) = \begin{cases} 1, & x\in\left[x(\alpha)_\ell, x(\alpha)_r\right] \\ 0, & otherwise \end{cases}. \quad (30)$$

Note that the α-cuts of a fuzzy membership function exist for $0\leq\alpha\leq\alpha_{\max}\leq 1$, where $\alpha_{\max}$ is the height of $\mu(x)$.

Thus, given the α-cuts

$$\left[x(\alpha)_{c_1\ell}, x(\alpha)_{c_1 r}\right] = \left\{x : \mu_{c_1}(x) \geq \alpha\right\} \quad (31)$$

$$\left[x(\alpha)_{c_2\ell}, x(\alpha)_{c_2 r}\right] = \left\{x : \mu_{c_2}(x) \geq \alpha\right\} \quad (32)$$

$$\left[x(\alpha)_{p_1\ell}, x(\alpha)_{p_1 r}\right] = \left\{x : \mu_{p_1}(x) \geq \alpha\right\} \quad (33)$$

$$\left[x(\alpha)_{p_2\ell}, x(\alpha)_{p_2 r}\right] = \left\{x : \mu_{p_2}(x) \geq \alpha\right\} \quad (34)$$

we can calculate the corresponding α-cuts $\left[x(\alpha)_{c_3\ell}, x(\alpha)_{c_3 r}\right]$ and $\left[x(\alpha)_{p_3\ell}, x(\alpha)_{p_3 r}\right]$ of the fuzzy membership functions $\mu_{c_3}(x)$ and $\mu_{p_3}(x)$ for $0\leq\alpha\leq\min_i \alpha_{\max i}$ using (9)-(10) and (22)-(23), with the input α-cut intervals specified by (31)-(34). Then the Type-1 memberships for $\mu_{c_3}(x)$ and $\mu_{p_3}(x)$ are given by:

$$\mu_{c_3}(x) = \sup_{\forall\alpha\in[0,\alpha_{\max}]} \begin{cases} \alpha, & x(\alpha)_{c_3\ell} \leq x \leq x(\alpha)_{c_3 r} \\ 0 & elsewhere \end{cases}, \quad (35)$$

$$\mu_{p_3}(x) = \sup_{\forall\alpha\in[0,\alpha_{\max}]} \begin{cases} \alpha, & x(\alpha)_{p_3\ell} \leq x \leq x(\alpha)_{p_3 r} \\ 0 & elsewhere \end{cases}, \quad (36)$$

where $\alpha_{\max} = \min_i \alpha_{\max i}$. In a similar vein, we can compute the fuzzy membership function $\mu_t(x)$ of trust from the α-cut intervals for $c\in[c_\ell,c_r]$ and $p\in[p_\ell,p_r]$ using (26)-(27) to compute the corresponding α-cut intervals $\left[x(\alpha)_{t\ell}, x(\alpha)_{tr}\right]$ for t:

$$\mu_t(x) = \sup_{\forall\alpha\in[0,\alpha_{\max}]} \begin{cases} \alpha, & x(\alpha)_{t\ell} \leq x \leq x(\alpha)_{tr} \\ 0 & elsewhere \end{cases}. \quad (37)$$

These membership functions can be de-fuzzified to a representative scalar value by computing their centroids, where the centroid ς_μ of a fuzzy membership function $\mu(x)$ is defined by

ISBN: 978-989-95079-6-8

$$\varsigma_\mu = \frac{\int x\mu(x)dx}{\int \mu(x)dx} \tag{38}$$

In practice, the integrals in (38) are approximated by summations.

5 Interval Type-2 fuzzy inputs

Now suppose that the fuzzy membership functions for c_1, c_2, p_1 and p_2 in equations (1) and (2) are interval Type-2 [13] fuzzy membership functions $\tilde{\mu}_{c_1}(x)$, $\tilde{\mu}_{c_2}(x)$, $\tilde{\mu}_{p_1}(x)$ and $\tilde{\mu}_{p_2}(x)$, respectively. Each of these functions exhibits a "footprint of uncertainty" (FOU) that represents the union of an uncountable number of Type-1 primary fuzzy membership functions lying within the FOU, each having a secondary membership equal to unity [14]. Thus we wish to propagate this imprecision in the input variables through to the corresponding imprecision in the output aggregation of two inputs.

In this case, we use an approach analogous to that of Wu and Mendel [15,16] for linguistic weighted averages, which relies upon the observation that an interval Type-2 fuzzy set is completely specified by the upper and lower convex membership functions (UMF and LMF) that bound its FOU. We denote the Type-2 membership functions for the input confidences and plausibilities by $\mu_{c_1}^{UMF}(x)$, $\mu_{c_1}^{LMF}(x)$, $\mu_{c_2}^{UMF}(x)$, $\mu_{c_2}^{LMF}(x)$, $\mu_{p_1}^{UMF}(x)$, $\mu_{p_1}^{LMF}(x)$, $\mu_{p_2}^{UMF}(x)$ and $\mu_{p_2}^{LMF}(x)$, respectively. To determine the corresponding interval Type-2 membership functions $\tilde{\mu}_{c_3}(x)$ and $\tilde{\mu}_{p_3}(x)$ for the aggregate confidence and plausibility, we employ the α-cut technique of the previous section applied separately to the UMF's and LMF's to compute the output UMF's and LMF's for confidence and plausibility, respectively. The area between these bounding functions represents the primary Type-2 fuzzy membership functions for these variables, with a secondary membership equal to unity over this FOU. In similar fashion, we can find the UMF and LMF for the trust Type-2 membership function $\tilde{\mu}_t(x)$.

Once any of the Type-2 memberships have been calculated by the above method, we have a very rich description of the imprecision of the corresponding fuzzy membership function. These Type-2 membership functions can be *type-reduced* to a Type-1 membership function by calculating the *centroid* of the Type-2 membership function via the Karnik-Mendel algorithm [12]. For interval Type-2 membership functions, the centroid membership function is itself an interval $[z_\ell, z_r]$. This is of course a reduced representation of the imprecision that is fully described by the Type-2 membership function. The interval Type-1 membership function can then be *defuzzified* by calculating the midpoint of this interval to obtain a scalar representation of the aggregate membership function.

6 Example

We now consider an example that approximates the first case depicted in equation (3), but incorporates more general interval Type-2 fuzzy membership functions to describe the distributions of confidence and plausibility. In all cases below, we use $\alpha = 2$ and $\beta = 1/4$.

Fig. 1 shows two Type-2 fuzzy membership functions for the confidence of two different observations, where the FOU of each membership is the area between the solid and dotted curves. In this example, both confidences have positive support intervals, centered upon the values 0.8 (with + shading) and 0.3 (with x shading) respectively, analogous to the point values of confidence in equation (3).

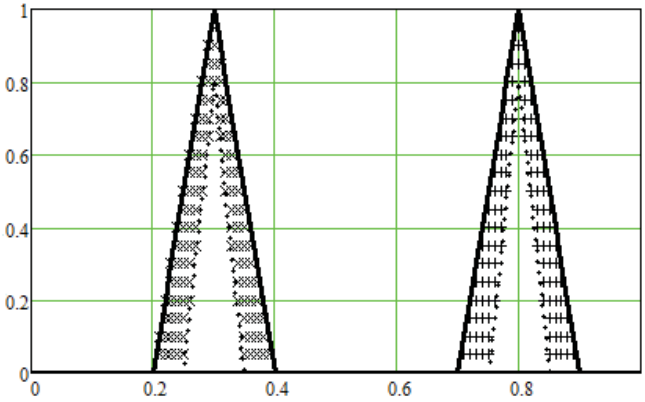


Figure 1 Interval Type-2 fuzzy membership functions for the confidence of two different observations.

In Fig. 2, we show the corresponding interval Type-2 membership functions of the plausibility of these two observations. Note the plausibility of the first observation (+ shading in Fig. 1) is concentrated near zero, while that of the second observation is centered on 0.5, but both cases exhibit a FOU of values.

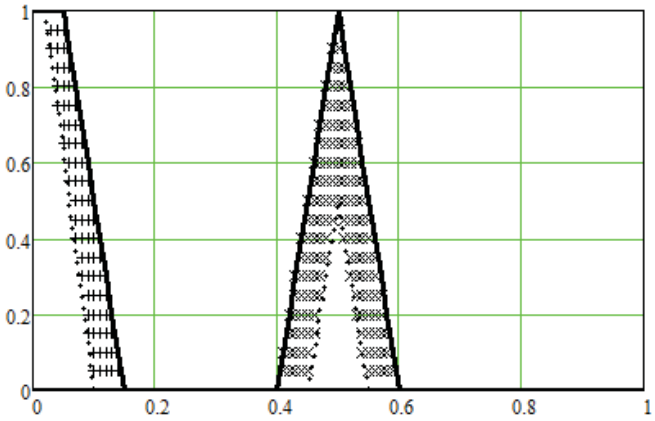


Figure 2 Interval Type-2 fuzzy membership functions for the plausibility of the observations whose confidence distributions are depicted in Fig. 1.

The mathematical apparatus described in previous sections for propagating the imprecision of these values when aggregating these two observations' confidences and plausibilities results in the aggregate confidence (+ shading) and plausibility (x shading) interval Type-2 membership functions shown in Fig. 3. The vertical dashed lines straddling the confidence FOU depict the centroid interval that results from type-reduction of the confidence membership function, while the vertical solid lines depict the midpoint (i.e., the de-fuzzified value) of the centroid intervals for both confidence and plausibility. These have

ISBN: 978-989-95079-6-8

values of 0.315 and 0.51, respectively, which roughly compares with the point wise values in equation (3), confirming that the aggregation of an observation with high confidence but very low plausibility results in essentially the same confidence and plausibility associated with the other observation. Note, however, that there is considerable additional information regarding both distributions that is lost in the type-reduction of a Type-2 membership function to a Type-1 interval membership function, and still more information is lost by representing the entire FOU by its de-fuzzified midpoint value.

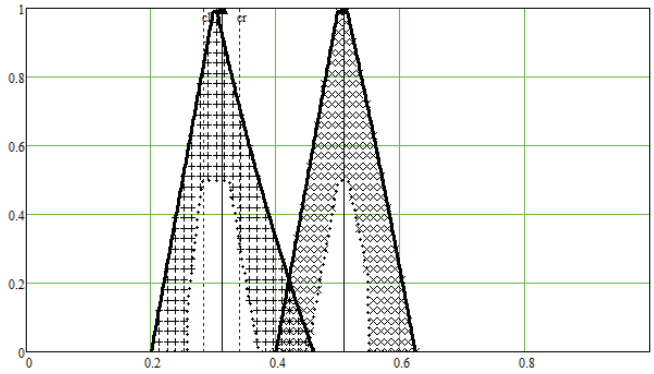


Figure 3 Interval Type-2 fuzzy membership functions for the aggregate confidence (+ shading) and plausibility (x shading) of the case depicted in Figures 4 and 5.

7 Conclusion

In this paper, we have generalized the trust aggregation algebra of Hamilton [8] for point values to admit interval, Type-1 and/or Type-2 input fuzzy membership values, and shown how these more complex inputs can be propagated in a mathematically principled manner to the corresponding output fuzzy membership functions.

References

[1] Grandison, T. and M. Sloman, "A survey of trust in internet applications," *IEEE Communications Surveys and Tutorials,* 3 pp. 2–16, 2000.

[2] Ray, I., S. Chakraborty, and I. Ray, "VTrust: a trust management system based on a vector model of trust," *Proceedings of International Conference on Information Systems Security*, Dec. 2005.

[3] Blaze, M., J. Feigenbaum and J. Lacy, "Decentralized trust management," *Proceedings of 17th IEEE Symposium on Security and Privacy*, Oakland, California, USA, pp. 164-173, IEEE Computer Society Press, 1996.

[4] Blaze, M., J. Feigenbaum and M. Strauss, "Compliance checking in the policymaker trust management system," *Proceedings of the 2nd Financial Crypto Conference*, vol. 1465 of *Lecture Notes in Computer Science*, pp. 254-274, Anguilla, Springer-Verlag, 1998.

[5] Jajodia, S., P. Samarati and V.S. Subramanian, "Logical language for expressing authorizations," *Proceedings IEEE Symposium on Security and Privacy*, pp. 31-42, May 1997.

[6] Damianou, N., N. Dulay, E. Lupu and M. Sloman. "The ponder policy specification language," *Policies for Distributed Systems and Networks*, vols. 1995, 2001.

[7] H. Hosmer, "Using fuzzy logic to represent security policies in the multipolicy paradigm," *ACM SIGSAC Review*, vol. 10, no. 4, pp. 12-21, Fall 1992.

[8] Hamilton, S. N. and Hamilton W. L. Hamilton, "Applying trust models to access control," submitted to *ACM Symposium on Access Control Models and Technologies*, June 2009.

[9] Wu, D. and J. M. Mendel, "Aggregation using the linguistic weighted average and interval Type-2 fuzzy sets," *IEEE Trans. Fuzzy Systems*, vol. 15, no. 6, pp. 1145-1161, Dec. 2007.

[10] Wu, D. and J. M. Mendel, "Corrections to "Aggregation using the linguistic weighted average and interval Type-2 fuzzy sets"," *IEEE Trans. Fuzzy Systems,* to appear.

[11] Klir, G. J. and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1995.

[12] Karnik, N. N. and J. M. Mendel, "Centroid of a type-2 fuzzy set," *Information Sciences*, vol. 132, pp. 195-220, 2001.

[13] Mendel, J. M., *Uncertain Rule-Based Fuzzy Logic Systems*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[14] Mendel, J. M., "Computing with words and its relationship with fuzzistics," *Information Sciences*, vol. 177, pp. 988-1006, 2007.

[15] Mendel, J. M. and H. Wu, "Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: Part 1: forward problem," *IEEE Trans. Fuzzy Systems*, vol. 14, pp. 781-792, 2006.

[16] Mendel, J. M. and H. Wu, "Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: Part 2: inverse problem," *IEEE Trans. Fuzzy Systems*, vol. 15, pp. 301-308, 2007.

[17] Liu, D. and J. M. Mendel, "Encoding words into interval Type-2 fuzzy sets using an *interval* approach," *IEEE Trans. Fuzzy Systems*, vol. 16, pp. 1503-1521, Dec. 2008.

ISBN: 978-989-95079-6-8

On vertices of the k-additive monotone core

Pedro Miranda[1] Michel Grabisch[2]

1. Faculty of Mathematical Sciencies, Universidad Complutense de Madrid
Madrid, Spain
2. Centre d'Économie de la Sorbonne, Université Paris I-Panthéon-Sorbonne
Paris, France
Email: pmiranda@mat.ucm.es, Michel.Grabisch@univ-paris1.fr

Abstract— Given a capacity, the set of dominating k-additive capacities is a convex polytope; thus, it is defined by its vertices. In this paper we deal with the problem of deriving a procedure to obtain such vertices in the line of the results of Shapley and Ichiishi for the additive case. We propose an algorithm to determine the vertices of the k-additive monotone core. Then, we characterize the vertices of the n-additive core and finally, we explore the possible translations for the k-additive case.

Keywords— Capacities, k-additivity, dominance, core.

1 Introduction

One of the main problems of cooperative game theory is to define a solution of a game μ, that is, supposing that all players join the grand coalition X, an *imputation* to each player represents a sharing of the total worth of the game $\mu(X)$. In the case of finite games of n players, an imputation can be written as a n-tuple $(x_1, \ldots, x_n)$ such that $\sum_{i=1}^{n} x_i = \mu(X)$. Of course, some rationality criterion should prevail when defining the sharing.

In this respect, the core is perhaps the most popular solution of a game. It is a well known fact that the core is nonempty if and only if the game is balanced [1]. However, there are games whose core is empty. It is then necessary to give an alternative solution. In this sense, many possibilities have been proposed in the literature, as the dominance core stable sets, Shapley index, the nucleolus, etc. (see e.g. [2]).

On the other hand, Grabisch has defined in [3] the concept of k-additive capacities, for a fixed value $1 \leq k \leq n$. These capacities generalize the concept of probability and they fill the gap between probabilities and general capacities. Moreover, as they are defined in terms of the Möbius transform and this transform can be applied to the characteristic function of any game (not necessarily monotone), the concept of k-additivity can be extended to games as well.

In a previous paper we have defined the so-called k-additive core. The basic idea is to remark that an imputation is nothing other than an additive game, and if the core is empty, we may allow to search for games more general than additive ones, namely k-additive games, dominating the game. We have presented a generalization of balanced games, the k-balanced games, that are those admitting a dominating k-additive game and no dominating $(k-1)$-additive game.

We have seen that for general games, any game is either balanced or 2-balanced. Moreover, the 2-additive core is not a bounded polytope but an unbounded convex polyhedron. However, when dealing with capacities, it makes sense to study the k-additive monotone core and it can be easily seen that in this particular case it is a convex polytope, whence it can be defined through describing its vertices. This paper studies these vertices. In the framework of Game Theory, it has been solved for the (1-additive) core by Shapley and Ichiishi. The vertices of the $(n-1)$-additive core has been obtained in [4].

Moreover, there are other fields in which it is interesting to find the set of probabilities dominating a capacity. For instance, Dempster [5] and Shafer [6] have proposed a representation of uncertainty based on a "lower probability" or "degree of belief", respectively, to every event. Their model needs a lower probability function, usually non-additive but having a weaker property: it is a belief function [6]. This requirement is perfectly justified in some situations (see [5]). The general form of lower probabilities has been studied by several authors (see e.g. [7, 8]). Moreover, in many decision problems, in which we have not enough information, decision makers often feel that they are only able to assign an interval value for the probability of events. In other words, they do not know the real probability distribution but there exists a set of probabilities compatible with the available information. Let us call this set of all compatible probabilities $\mathcal{P}_1$ and let us define $\mu = \inf_{P \in \mathcal{P}_1} P$; then, μ is a capacity (but not necessarily a belief function [9]); μ is called "coherent lower probability", and it is the natural "lower probability function". Of course, if P' is a probability measure dominating μ, it is clear that $E_{P'}(f) \geq \mathcal{C}_\mu(f)$, for any function f, where $\mathcal{C}_\mu$ represents Choquet integral [10]. Chateauneuf and Jaffray use this fact and that $\mu \leq P$, $\forall P \in \mathcal{P}_1$ in [11] to obtain an easy method for computing a lower bound of $\inf_{P \in \mathcal{P}_1} E_P(f)$ whenever μ is 2-monotone. Their method is based on obtaining the set of all probability distributions dominating μ. The same can be done for obtaining an upper bound. In this case, we can find a similar motivation for studying the set of all k-additive capacities dominating a capacity.

The paper is organized as follows: In next section, we give the basic concepts about k-additive capacities and about the set of dominating probabilities. Next, in Section 3 we provide an algorithm for obtaining the set of all k-additive dominating capacities. Section 4 is devoted to characterize the vertices for the n-additive case and, in Section 5, we deal with possible generalizations for the k-additive case.

2 Basic concepts

We will use the following notations: we suppose a finite universal set with n elements, $X = \{1, ..., n\}$. Subsets of X are

ISBN: 978-989-95079-6-8

denoted by capital letters A, B, and so on. The set of subsets of X is denoted by $\mathcal{P}(X)$, and the set of subsets whose cardinality is maximum k is denoted by $\mathcal{P}^k(X)$.

Definition 1 *[12] A* **game** *over X is a mapping $\mu : \mathcal{P}(X) \to \mathbb{R}$ (called* **characteristic function***) satisfying $\mu(\emptyset) = 0$.*
If, in addition,

1. *μ satisfies $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$, the game μ is said to be* **monotone***;*
2. *μ satisfies $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \subseteq X$, $A \cap B = \emptyset$, the game is said to be* **additive***;*
3. *μ satisfies $\mu(A \cup B) + \mu(A \cap B) \geq \mu(A) + \mu(B)$, for all $A, B \subseteq X$, the game is said to be* **convex***.*
4. *μ satisfies*

$$\mu(\bigcup_{i=1}^{k} A_i) \geq \sum_{\substack{K \subseteq \{1,\ldots,k\} \\ K \neq \emptyset}} (-1)^{|K|+1} \mu(\bigcap_{j \in K} A_j) \quad (1)$$

*for any family of k subsets $A_1, \ldots A_k$, $k \geq 2$, the game is said to be k-***monotone***.*

Definition 2 *A* **non-additive measure** *[13] or* **capacity** *[10] or* **fuzzy measure** *[14] μ over X is a monotone game with $\mu(X) = 1$.*

Note that any monotone game can be equivalently defined through a capacity. The set of all capacities on X is a convex polytope, that we will denote $\mathcal{FM}(X)$.

Definition 3 *[15] Let μ be a game on X. The* **Möbius transform (or inverse)** *of μ is a set function on X defined by*

$$m_\mu(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B), \, \forall A \subseteq X. \quad (2)$$

The Möbius transform given, the original characteristic function can be recovered through the *Zeta transform* [11]:

$$\mu(A) = \sum_{B \subseteq A} m(B). \quad (3)$$

Let us turn to the concept of k-additivity. In order to define a capacity, $2^n - 2$ values are necessary. The number of coefficients grows exponentially with n, and so does the complexity of the problem. This drawback reduces considerably the practical use of capacities. Then, some subfamilies of capacities have been defined in an attempt to reduce complexity. In this paper we will use k-additive capacities.

Definition 4 *[16] A game μ is said to be k-***order additive** *or k-***additive** *for some $k \in \{1, \ldots, n\}$ if its Möbius transform vanishes for any $A \subseteq X$ such that $|A| > k$, and there exists at least one subset A of exactly k elements such that $m(A) \neq 0$.*

In this sense, a probability is just a 1-additive capacity [16]. Thus, k-additive capacities generalize probabilities, that are very restrictive in many situations as they do not allow interactions between the elements of X. They fill the gap between probabilities and general non-additive capacities. We will denote by $\mathcal{FM}^k(X)$ (resp. $\mathcal{G}^k(X)$) the set of all k'-additive capacities (resp. games) with $k' \leq k$.

Let us introduce the concept of k-additive monotone core.

Definition 5 *Let $\mu\mu^*$ be two games. We say μ^** **dominates** *μ, and we denote it $\mu^* \geq \mu$, if and only if*

$$\mu^*(A) \geq \mu(A), \, \forall A \subset X, \mu^*(X) = \mu(X). \quad (4)$$

One of the main problems of cooperative game theory is to define a solution of a game ν, that is, supposing that all players join the grand coalition X, an *imputation* to each player represents a sharing of the total worth of the game $\nu(X)$. In the case of finite games of n players, an imputation can be written as a n-tuple $(x_1, \ldots, x_n)$ such that $\sum_{i=1}^n x_i = \nu(X)$. Of course, some rationality criterion should prevail when defining the sharing.

Definition 6 *Let μ be a game. We say that a vector : $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is an* **imputation** *for μ if it satisfies*

$$\sum_{i=1}^{n} x_i = \mu(X). \quad (5)$$

An imputation represents a possible pay-off for the players, i.e. supposing that all players agree to form the grand coalition, it provides a possible sharing of the value $\mu(X)$ among the players.

Remark 1 *For any $\mathbf{x} \in \mathbb{R}^n$, it is convenient to use the notation $x(A) := \sum_{i \in A} x_i$, for all $A \subseteq X$, with the convention $x(\emptyset) = 0$. Thus, x identifies with an additive game for which the values on singletons are x_i.*

The value x_i is the asset player i receives when sharing $\mu(X)$. Suppose that the imputation satisfies $x(A) \geq \mu(A)$, for all $A \subseteq X$. If this is the case, all players should agree with their pay-off, as if they try to form other coalition different of X, the corresponding value for the coalition would be worse than the value the coalition obtains with the additive game x. In other words, any such $(x_1, \ldots, x_n)$ is a possible satisfactory imputation for all players.

Definition 7 *[17] Let μ be a game. The* **core** *of μ, denoted by $\mathcal{C}(\mu)$, is defined by*

$$\mathcal{C}(\mu) := \{x \in \mathbb{R}^n \mid x(A) \geq \mu(A), \, \forall A \subseteq X, x(X) = \mu(X)\}.$$

Since by Remark 1 any $x \in \mathbb{R}^n$ induces an additive game, the core can be equivalently defined as the set of additive games dominating μ. When the core is nonempty, it is usually taken as the solution of the game. Note that for the case of the core, given a dominating additive game, the value x_i coincides with $m(i)$. However, there are games with an empty core. Then, the following definition arises:

Definition 8 *[12] A game μ is* **balanced** *if $\mathcal{C}(\mu) \neq \emptyset$.*

For the special case of μ being a capacity, if $(x_1, \ldots, x_n)$ is in the core, it follows that $(x_1, \ldots, x_n)$ determines a probability distribution on X dominating μ. Thus, in this case, $\mathcal{C}(\mu)$ coincides with the set of all probabilities dominating μ.

When non-empty, the core is a convex polytope and its vertices are known when the game is convex.

Definition 9 *A* **maximal chain** *in 2^X is a sequence of subsets $A_0 := \emptyset, A_1, \ldots, A_{n-1}, A_n := X$ such that $A_i \subset A_{i+1}$, $i = 0, \ldots, n-1$. The set of maximal chains of 2^X is denoted by $\mathcal{M}(2^X)$.*

ISBN: 978-989-95079-6-8

To each maximal chain $C := \{\emptyset, A_1, \ldots, A_n = X\}$ in $\mathcal{M}(2^X)$ corresponds a unique permutation σ on X such that $A_1 = \sigma(1)$, $A_2 \setminus A_1 = \sigma(2)$, $\ldots$, $A_n \setminus A_{n-1} = \sigma(n)$. The set of all permutations over X is denoted by $\mathfrak{S}(X)$. Let μ be a capacity. Each permutation σ (or maximal chain C) induces an additive capacity ϕ^σ (or ϕ^C) on X defined by:

$$\phi^\sigma(\{\sigma(i)\}) := \mu(\{\sigma(1), \ldots, \sigma(i)\}) - \mu(\{\sigma(1), \ldots, \sigma(i-1)\}) \quad (6)$$

or

$$\phi^C(\{\sigma(i)\}) := \mu(A_i) - \mu(A_{i-1}), \quad \forall i \in X, \quad (7)$$

with the above notation.

Theorem 1 *The following propositions are equivalent.*

1. *μ is a convex capacity.*
2. *All additive capacities ϕ^σ, $\sigma \in \mathfrak{S}(X)$, belong to the core of μ.*
3. *$\mathcal{C}(\mu) = \mathrm{co}(\{\phi^\sigma\}_{\sigma \in \mathfrak{S}(X)})$.*
4. *$\mathrm{ext}(\mathcal{C}(\mu)) = \{\phi^\sigma\}_{\sigma \in \mathfrak{S}(X)}$,*

where $\mathrm{co}(\cdot)$ and $\mathrm{ext}(\cdot)$ denote respectively the convex hull of some set, and the extreme points of some convex set.

(i) $\Rightarrow$ (ii) and (i) $\Rightarrow$ (iv) are due to Shapley [17], while (ii) $\Rightarrow$ (i) was proved by Ichiishi [18].

In a previous work [19] we have defined the so-call k-additive monotone core.

Definition 10 *For some integer $1 \leq k \leq n$, the k-**additive monotone core** of a capacity μ is defined by:*

$$\mathcal{MC}^k(\mu) := \{\phi \in \mathcal{FM}^k(X) \mid \phi(A) \geq \mu(A), \quad \forall A \subseteq X\}.$$

If non-empty, it is easy to see that $\mathcal{MC}^k(\mu)$ is a convex polytope. In next sections we will study its vertices, i.e. the capacities such that they cannot be put as a convex combination of two other capacities in the polytope.

3 An algorithm for determining vertices of the k-additive monotone core

Take $\mu \in \mathcal{FM}(X)$. The polytope $\mathcal{MC}^k(\mu)$ can be seen as a subpolytope of $\mathcal{FM}^k(X)$, given by the additional constraints

$$\mu^*(A) \geq \mu(A), \forall A \subseteq X. \quad (8)$$

I.e. we restrict the polytope to the measures $\mu^* \in \mathcal{FM}^k(X)$ dominating μ. Thus, we propose the following algorithm to determine its vertices:

- Initialization: $\mathcal{FM}^k(X)$.
- Take $A \subseteq X$ and add the constraint $\mu^*(A) \geq \mu(A)$.
- Obtain the vertices and the adjacency structure (i.e. whether two vertices are in an edge) of the new polytope.
- Repeat for any $A \subseteq X, A \neq X, \emptyset$.

Let us analyze a step. We will denote by $\mathcal{F}_1$ the polytope before introducing the new constraint $\mu^*(A) \geq \mu(A)$ and by $\mathcal{F}_2$ the resulting polytope. For $\mathcal{F}_1$, we assume that we know its vertices and its adjacency structure. The vertices of $\mathcal{F}_2$ are:

- Vertices of $\mathcal{F}_1$ satisfying the new constraint.
- New vertices, coming from the intersection of the hyperplane defined by $\mu^*(A) = \mu(A)$ and $\mathcal{F}_1$. It can be easily proved that these new vertices are in edges of $\mathcal{F}_1$.

Thus, in order to determine $\mathcal{F}_2$, it suffices to know the vertices and the edges (whether two vertices are adjacent) of $\mathcal{F}_1$.

In order to apply this procedure again, let us determine the adjacency structure of $\mathcal{F}_2$. Consider μ_1, μ_2 two vertices of $\mathcal{F}_2$. We have several cases:

- μ_1, μ_2 are vertices of $\mathcal{F}_1$.
- μ_1 is a vertex of $\mathcal{F}_1$ but μ_2 is not.
- Neither μ_1 nor μ_2 are vertices of $\mathcal{F}_1$.

We will study whether μ_1 and μ_2 are adjacent vertices in each situation.

Consider μ_1, μ_2 two vertices of $\mathcal{F}_2$ that are also vertices of $\mathcal{F}_1$. If they are adjacent vertices in $\mathcal{F}_1$, then they are adjacent in $\mathcal{F}_2$, as $\mathcal{F}_2$ is a subpolytope of $\mathcal{F}_1$ and we are done.

Lemma 1 *Assume μ_1, μ_2 are not adjacent vertices in $\mathcal{F}_1$. Then, if μ_1, μ_2 are adjacent vertices in $\mathcal{F}_2$, they satisfy $\mu_1(A) = \mu(A) = \mu_2(A)$.*

Moreover, the following holds:

Lemma 2 *Consider μ_1, μ_2 two vertices of $\mathcal{F}_1$ and assume they are adjacent vertices in $\mathcal{F}_2$. Then, if μ_1, μ_2 are not adjacent in $\mathcal{F}_1$, necessarily $\mu_1(A) = \mu(A) = \mu_2(A)$.*

Remark 2 *It can be proven that if μ_1, μ_2 are in the conditions of the previous lemma, then they are in a facet of dimension 2 of $\mathcal{F}_1$.*

Moreover, the intersection of this facet with the hyperplane $\mu^*(A) = \mu(A)$ is exactly the segment $[\mu_1, \mu_2]$. Otherwise, if we can find μ_3 outside the segment and in the facet satisfying $\mu_3(A) = \mu(A)$, we can build two linearly independent vectors in the facet and in the hyperplane, whence the facet is contained in the hyperplane and thus, the facet is contained in $\mathcal{F}_2$. But this would imply that, as μ_1, μ_2 are not adjacent in $\mathcal{F}_1$, they are not adjacent in $\mathcal{F}_2$, a contradiction.

Let us now turn to the case in which μ_1 is a vertex of $\mathcal{F}_1$ but μ_2 is not. The conditions for μ_1, μ_2 being adjacent vertices in $\mathcal{F}_2$ are given in next lemma.

Lemma 3 *Consider μ_1, μ_2 two vertices of $\mathcal{F}_2$ and suppose μ_1 is a vertex of $\mathcal{F}_1$ and μ_2 is not. If they are adjacent in $\mathcal{F}_2$, then:*

- *If $\mu_1(A) > \mu(A)$, then μ_2 is in an edge starting in μ_1.*
- *If $\mu_1(A) = \mu(A)$, then μ_1 and μ_2 are in a facet of $\mathcal{F}_1$ of dimension 2. Indeed, the intersection of the facet with the hyperplane $\mu^*(A) = \mu(A)$ is the segment $[\mu_1, \mu_2]$.*

Moreover, the following can be proved:

Lemma 4 *Consider μ_2 a vertex of $\mathcal{F}_2$ such that μ_2 is not a vertex of $\mathcal{F}_1$. Then, there exists exactly one vertex μ_1 of $\mathcal{F}_1$ such that $\mu_1(A) > \mu(A)$ that is adjacent to μ_2 in $\mathcal{F}_2$.*

ISBN: 978-989-95079-6-8

Finally, let us consider the case of μ_1, μ_2 being two vertices of $\mathcal{F}_2$ that are not vertices of $\mathcal{F}_1$.

Lemma 5 *Suppose μ_1, μ_2 are vertices of $\mathcal{F}_2$ that are not vertices of $\mathcal{F}_1$. If they are adjacent vertices of $\mathcal{F}_2$, they are in the same facet of dimension 2 of $\mathcal{F}_1$. Indeed, the intersection of the facet with the hyperplane $\mu^*(A) = \mu(A)$ is the segment $[\mu_1, \mu_2]$.*

As a final remark about this section note that, in order to apply this procedure, it is necessary to know the adjacency structure of the polytope $\mathcal{FM}^k(X)$. As proved in [20, 21], the problem of determining non-adjacency of vertices of a polytope is, in some cases, NP-complete. The vertices of $\mathcal{FM}(X)$ are $\{0,1\}$-valued measures [22]. In [23] a characterization of the adjacency in $\mathcal{FM}(X) = \mathcal{FM}^n(X)$ that allows us to check whether two vertices are adjacent in quadratic time has been obtained. The adjacency structure of $\mathcal{FM}^1(X)$ and $\mathcal{FM}^2(X)$ is also known. However, the structure of $\mathcal{FM}^k(X)$ for other values of k is more complicated [24] and the adjacency structure is not known (indeed, the vertices of the polytope have not been obtained yet). In this last case, we are forced to apply a similar algorithm with the additional constraints $m(A) = 0$ before applying the procedure.

4 The set $\mathcal{MC}^n(\mu)$.

The procedure stated in the previous section can be very time-consuming and thus unfeasible in practice for big values of $|X|$. Thus, it is interesting to look for a characterization of the vertices of the k-additive core; in this line we have the results of Shapley and Ichiishi [17, 18] for $\mathcal{C}^1(\mu)$ and the results in [4] for $\mathcal{C}^{n-1}(\mu)$ (the $(n-1)$-additive core, not restricted to capacities). In this section we provide a characterization of vertices of $\mathcal{MC}^n(\mu)$. We consider the following procedure.

- Let $\prec$ be an order on $\mathcal{P}(X) \setminus \{X, \emptyset\}$. This order allows us to rank the different subsets of X,

$$A_1 \prec A_2 \prec ... \prec A_{2^n-2}. \tag{9}$$

- Next, take a partition $\mathcal{P} = \{\mathcal{U}, \mathcal{L}\}$ on $\mathcal{P}(X) \setminus \{X, \emptyset\}$, where $\mathcal{U}$ or $\mathcal{L}$ could be empty.

- **Initializing step**: For $\prec$ and $\mathcal{P}$ fixed, let us define

$$\overline{\mu}^0(A_i) = 1, \underline{\mu}^0(A_i) = \mu(A_i), \forall A_i. \tag{10}$$

- **Iterating step**: For $i = 1$ until $i = 2^n - 2$, do:

  – If $A_i \in \mathcal{U}$, then assign

$$\mu_{\prec,\mathcal{P}}(A_i) = \overline{\mu}^{i-1}(A_i). \tag{11}$$

  Redefine:
  For $\underline{\mu}^i$, we put

$$\underline{\mu}^i(B) = \max\{\overline{\mu}^{i-1}(A_i), \underline{\mu}^{i-1}(B)\}, \text{ if } A_i \subseteq B \tag{12}$$

$$\underline{\mu}^i(B) = \underline{\mu}^{i-1}(B), \text{ otherwise.} \tag{13}$$

  For $\overline{\mu}^i$, we put

$$\overline{\mu}^i(B) = \overline{\mu}^{i-1}(B), \forall B \subset X. \tag{14}$$

  – If $A_i \in \mathcal{L}$, then assign

$$\mu_{\prec,\mathcal{P}}(A_i) = \underline{\mu}^{i-1}(A_i). \tag{15}$$

  Redefine:
  For $\overline{\mu}^i$, we put

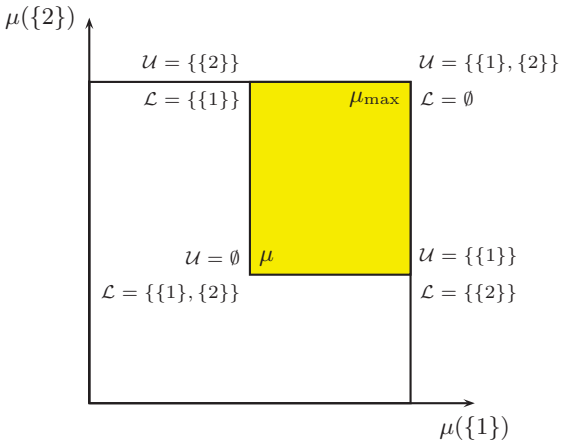
$$\overline{\mu}^i(B) = \min\{\underline{\mu}^{i-1}(A_i), \overline{\mu}^{i-1}(B)\}, \text{ if } B \subseteq A_i \tag{16}$$

$$\overline{\mu}^i(B) = \overline{\mu}^{i-1}(B), \text{ otherwise.} \tag{17}$$

  For $\underline{\mu}^i$, we put

$$\underline{\mu}^i(B) = \underline{\mu}^{i-1}(B), \forall B \subset X. \tag{18}$$

The idea of the procedure is the following: If we are in step i, the values for $A_1, ..., A_{i-1}$ are fixed. For A_i, if $A_i \in \mathcal{U}$, we assign to $\mu_{\prec,\mathcal{P}}(A_i)$ the biggest possible value keeping dominance, which is $\overline{\mu}^i(A_i)$. Similarly, if $A_i \in \mathcal{L}$, we assign to $\mu_{\prec,\mathcal{P}}(A_i)$ the smallest possible value keeping dominance, which is $\underline{\mu}^i(A_i)$. Once the value of $\mu_{\prec,\mathcal{P}}(A_i)$ is fixed, we need to actualize the lower and upper bounds for $A_j, j > i$. These lower and upper bounds are stored in $\underline{\mu}^{i+1}$ and $\overline{\mu}^{i+1}$, respectively. The Figure below explains the performance of the algorithm for the special case of $|X| = 2$. In it, we can see that $\mathcal{MC}^2(\mu)$ has four vertices. Note that μ itself is always a vertex of $\mathcal{MC}^2(\mu)$, corresponding to the case $\mathcal{L} = \mathcal{P}(X) \setminus \{X, \emptyset\}$ and $\mathcal{U} = \emptyset$. Thus, for the n-additive case, $\mu \in \mathcal{MC}^n(\mu)$ and it is the bottom element of this set. Similarly, the measure attaining value 1 for every subset is another vertex of $\mathcal{MC}^2(\mu)$, corresponding to the case $\mathcal{U} = \mathcal{P}(X) \setminus \{X, \emptyset\}$ and $\mathcal{L} = \emptyset$. Thus, for the n-additive case, it is the top element of the set. This also holds when $|X| > 2$.



Note also that this procedure generalizes for the n-additive case the Shapley-Ichiishi theorem for probabilities. In our case, the total order on $\mathcal{P}(X)\setminus\{\emptyset, X\}$ plays the role of chains; the value that is assigned to a singleton in the Shapley-Ichiishi theorem is indeed the upper and lower bound for this value in order to keep dominance. However, in our case we need two "chains" instead of one because lower and upper bounds are not the same for the general case.

Finally, remark that for the general case we do not need to impose any additional condition on μ, while for the set $\mathcal{C}(\mu) = \mathcal{MC}^1(\mu)$ convexity is required.

In next results we will prove that the function $\mu_{\prec,\mathcal{P}}$ obtained through this procedure is a vertex of $\mathcal{MC}^n(\mu)$ and that any vertex can be obtained through a suitable choice of $\prec$ and $\mathcal{P}$.

ISBN: 978-989-95079-6-8

Proposition 1 $\mu_{\prec,\mathcal{P}} \in \mathcal{MC}^n(\mu)$.

The proof of this result is based on the following lemmas, that provide us with additional properties of $\{\overline{\mu}^i(B)\}_{i=0}^{2^n-2}$ and $\{\underline{\mu}^i(B)\}_{i=0}^{2^n-2}$.

Lemma 6 *For any $B \subset X$, the sequence $\{\underline{\mu}^i(B)\}_{i=1}^{2^n-2}$ is nondecreasing. Similarly, the sequence $\{\overline{\mu}^i(B)\}_{i=0}^{2^n-2}$ is nonincreasing.*

Lemma 7 $\underline{\mu}^i, \overline{\mu}^i \in \mathcal{FM}(X), \forall i = 0, ..., 2^n - 2$.

Lemma 8 $\overline{\mu}^i \geq \underline{\mu}^i, \forall i = 0, ..., 2^n - 2$.

Moreover, the following holds.

Proposition 2 *$\mu_{\prec,\mathcal{P}}$ is a vertex of $\mathcal{MC}^n(\mu)$.*

Finally, it can be seen that all the vertices can be derived from this procedure.

Proposition 3 *If μ^* is a vertex of $\mathcal{MC}^n(\mu)$, there exists an order $\prec$ and a partition $\mathcal{P}$ of $\mathcal{P}(X) \setminus \{X, \emptyset\}$ such that $\mu^* = \mu_{\prec,\mathcal{P}}$.*

Note that these results allows us to derive an upper bound for the number of vertices of $\mathcal{MC}^n(\mu)$.

Proposition 4 *The number of vertices of $\mathcal{MC}^n(\mu)$ is bounded by 2^{2^n-2}.*

5 The case of $\mathcal{MC}^k(\mu)$

In this section we treat the general k-additive case. The basic idea is to translate the results of the previous section to this case. In order to translate the results, we will need to solve two new problems:

- For a fixed value of $\mu(A)$, the possible lower and upper bounds of $\mu(B), B \neq A$ are not trivial, as it happened for the n-additive case. Moreover, if we are dealing with $\mathcal{MC}^k(\mu)$, it could be the case that $\mu \notin \mathcal{FM}^k(X)$. Thus, in the k-additive case, it could be the case that no such bottom element for $\mathcal{MC}^k(\mu)$ exists, so we cannot define $\underline{\mu}^0$. Similarly, the measure attaining value 1 for every subset is no longer in $\mathcal{MC}^k(\mu)$, whence we cannot define $\overline{\mu}^0$.
- The structure of the polytope $\mathcal{FM}^k(X)$ is not known for $k \geq 3$. Indeed, it has been proved in [24] that there are vertices of $\mathcal{FM}^k(X), k \geq 3$ that are not $\{0,1\}$-valued measures; moreover, we do not know the vertices of the polytope. However, it could be the case that some of these vertices are in $\mathcal{MC}^k(\mu)$. How can they be characterized?

We will study in this section the particular case in which $\mu \in \mathcal{FM}^k(X)$. Notice that in the n-additive general case, this condition trivially holds. Of course, if $\mu \in \mathcal{FM}^k(X)$, then $\mathcal{MC}^k(\mu) \neq \emptyset$ and has a bottom element (μ itself). Notice again that this was the situation for the n-additive case.

- Let us consider a total order on $\mathcal{P}^k_*(X) := \mathcal{P}^k(X) \setminus \{\emptyset\}$. Notice that in the general case, we have considered an order on $\mathcal{P}(X) \setminus \{\emptyset, X\}$. This total order allows us to range the subsets in $\mathcal{P}^k_*(X)$: $A_1 \prec A_2 \prec ... \prec A_r$, where $r = \sum_{i=1}^k \binom{n}{i}$.
- Next, take a partition $\mathcal{P} = \{\mathcal{U}, \mathcal{L}\}$ on $\mathcal{P}(X) \setminus \{X, \emptyset\}$, where $\mathcal{U}$ or $\mathcal{L}$ could be empty.
- **Initializing step**: Let us define $\mu_0 := \mu$
- **Iterating step**: For $i = 1$ until $i = r$ do:
  - If $A_i \in \mathcal{L}$, then $\mu_i = \mu_{i-1}$.
  - Otherwise $A_i \in \mathcal{U}$. It is easy to see that $\mu_{i-1} \in \mathcal{FM}^k(X)$.
    Let us consider the subset given by

$$\{\mu^* \in \mathcal{FM}^k(X) \,|\, \mu^* \geq \mu_{i-1},$$
$$\mu^*(A_j) = \mu_{i-1}(A_j), j = 1, ..., i-1\}. \quad (19)$$

    As $\mathcal{FM}^k(X)$ is a polytope, so is this set. Thus, for A_i, it follows that $\mu^*(A_i)$ can vary in an interval $[\mu_{i-1}(A_i), s]$. Take the subset of measures in the set satisfying $\mu^*(A_i) = s$. Let us denote this subset by $\mathcal{A}_0$.
    As before, $\mathcal{A}_0$ is a polytope. Then, for $\mu^* \in \mathcal{A}_0$, the value $\mu^*(A_{i+1})$ can vary in an interval $[x_1^-, x_1^+]$. We define

$$\mathcal{A}_1 := \{\mu^* \in \mathcal{A}_0 \,|\, \mu^*(A_{i+1}) = x_1^-\}. \quad (20)$$

    Now, the same can be done for $\mathcal{A}_1$ and considering A_{i+2}. We reiterate until $\mathcal{A}_j$ is just a singleton (notice that necessarily $\mathcal{A}_{r-i}$ is a singleton). This capacity is μ_i.

It can be checked that this algorithm coincides with the algorithm of the previous section for the n-additive case. Let us denote by μ_r the measure obtained in the last iteration.

Note that if $k = 1$ and μ is a probability, then the set $\mathcal{MC}^1(\mu) = \mathcal{C}(\mu) = \{\mu\}$ and the problem is trivial.

Now, the following can be proved:

Lemma 9 *If $\mu \in \mathcal{FM}^k(X)$, the measure μ_r obtained in the procedure is an extreme point of $\mathcal{MC}^k(\mu)$.*

However, this method does not obtain all the vertices of $\mathcal{MC}^k(\mu)$, as next example shows:

Example 1 *Consider $|X| = 4$ and the measure $u_{\{1,4\}}$ given by $u_{\{1,4\}}(A) = 1$ if $\{1,4\} \subseteq A$ and $u_{\{1,4\}}(A) = 0$ otherwise. Then, $u_{\{1,4\}} \in \mathcal{FM}^3(X)$. Consider the measure μ^* given by*

Subset	*1*	*2*	*3*	*4*	*1,2*	*1,3*	*1,4*
μ^*	*0*	*0*	*0*	*0*	*0.5*	*0.5*	*1*

Subset	*2,3*	*2,4*	*3,4*	*1,2,3*	*1,2,4*	*1,3,4*	*2,3,4*
μ^*	*0.5*	*0*	*0*	*0.5*	*1*	*1*	*1*

It has been proved in [24] that μ^* is an extreme point of $\mathcal{FM}^3(X)$. On the other hand, $\mu^* \geq u_{\{1,4\}}$, whence μ^* is an extreme point of $\mathcal{MC}^3(u_{\{1,4\}})$. Let us check that μ^* cannot be obtained through the previous algorithm, no matter the order considered.

Assume $A_1 \in \mathcal{U}$.

ISBN: 978-989-95079-6-8

- *If A_1 is a singleton, straightforward calculus shows that we can obtain $\mu_1(A_1) = 1$, whence μ^* could not be derived. The same holds if A_1 is a pair different from $\{1,4\}$ and $\{2,3\}$.*

- *Suppose $A_1 = \{2,3\}$ and consider the measure whose Möbius transform is given by*

$$m'(\{2,3\}) = 1, m'(\{1,4\}) = 1, m'(\{2,4\}) = 1. \quad (21)$$

  In this case, we obtain a 3-additive measure and thus, it is possible to obtain $\mu_1(\{2,3\}) = 1$. Therefore, μ^* cannot be derived in this case.

- *If $A_1 = \{1,2,3\}$, then $m'(\{1,2\}) = 1$ gives $\mu_1(\{1,2,3\}) = 1$, whence μ^* cannot be derived in this case.*

- *For the other possibilities, we have $u_{\{1,4\}}(A_1) = 1$, whence the value is fixed.*

Consequently, it is not possible to obtain μ^* if $A_1 \in \mathcal{U}$. Thus, assume $A_1 \in \mathcal{L}$. This fixes $\mu_1(A_1) = u_{\{1,4\}}(A_1)$.

- *If $A_1 \in \{\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}$, then $\mu_1(A_1) = 0$, whence μ^* cannot be recovered.*

- *For other possibilities, we fix a value either 0 or 1.*

But then, we can repeat the process for A_2 with the same results. Thus, μ^* cannot be obtained through the procedure.

6 Conclusions

In this paper we have proposed an algorithm to obtain the vertices of the k-additive monotone core. This procedure can be applied to any value of k. However, it seems that it could be very time-consuming. Next, we have derived the vertices of the n-additive core; these results generalize the Shapley-Ichiishi theorem for the general case. Finally, we have treated the possible extensions for the k-additive case.

An important problem arising in the k-additive case is the number of vertices of the k-additive core. For the general n-additive case, the set of vertices of $\mathcal{FM}(X)$ coincides with the n-th Dedekind number; simulations carried on for the k-additive case seem to show that the number of vertices of $\mathcal{FM}^k(X)$ is even greater, due to vertices that are not $\{0,1\}$-valued. This problem could make unfeasible to store all the vertices of the k-additive core in some cases.

Acknowledgment

This research has been supported in part by grant numbers MTM2007-61193 and CAM-UCM910707.

References

[1] O. Bondareva. Some applications of linear programming to the theory of cooperative games. *Problemy Kibernet*, (10):119–139, 1963.

[2] T. Driessen. *Cooperative Games*. Kluwer Academic, 1988.

[3] M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5:587–607, 1997.

[4] M. Grabisch and P. Miranda. On the vertices of the k-additive core. *Discrete Mathematics*, (308):5204–5217, 2008.

[5] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, (38):325–339, 1967.

[6] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, (USA), 1976.

[7] P. Walley and T. L. Fine. Towards a frequentist theory of upper and lower probability. *Ann. of Stat.*, 10:741–761, 1982.

[8] M. Wolfenson and T. L. Fine. Bayes-like decision making with upper and lower probabilities. *J. Amer. Statis. Assoc.*, (77):80–88, 1982.

[9] P. Walley. Coherent lower (and upper) probabilities. Technical Report 22, U. of Warwick, Coventry, (UK), 1981.

[10] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, (5):131–295, 1953.

[11] A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, (17):263–283, 1989.

[12] G. Owen. *Game Theory*. Academic Press, 1995.

[13] D. Denneberg. *Non-additive measures and integral*. Kluwer Academic, 1994.

[14] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[15] G. C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, (2):340–368, 1964.

[16] M. Grabisch. k-order additive discrete fuzzy measures. In *Proceedings of 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 1345–1350, Granada (Spain), 1996.

[17] L. S. Shapley. Cores of convex games. *International Journal of Game Theory*, 1:11–26, 1971.

[18] T. Ichiishi. Super-modularity: Applications to convex games and to the Greedy algorithm for LP. *Journal of Economic Theory*, (25):283–286, 1981.

[19] P. Miranda and M. Grabisch. k-balanced games and capacities. *European Journal of Operational Research*, (Accepted).

[20] T. Matsui. NP-completeness of non-adjacency relations on some 0-1-polytopes. In *Lecture Notes in Operations Research, 1*, pages 249–258. 1995.

[21] C. Papadimitriou. The adjacency relation on the travelling salesman polytope is NP-complete. *Mathematical Programming*, 14(1), 1978.

[22] D. Radojevic. The logical representation of the discrete Choquet integral. *Belgian Journal of Operations Research, Statistics and Computer Science*, 38(2–3):67–89, 1998.

[23] E. F. Combarro and P. Miranda. On the polytope of non-additve measures. *Fuzzy Sets and Systems*, 159(16):2145–2162, 2008.

[24] P. Miranda, E.F. Combarro, and P. Gil. Extreme points of some families of non-additive measures. *European Journal of Operational Research*, 33(10):3046–3066, 2006.

ISBN: 978-989-95079-6-8

Resolution of fuzzy relation equations with sup-inf composition over complete Brouwerian lattices—a review *

Xue-ping Wang[1] Feng Sun[1]

1.Department of Mathematics, Sichuan Normal University
Chengdu, Sichuan 610066, People's Republic of China
Email: xpwang1@hotmail.com, sunfeng1005@163.com

Abstract— *This paper restates the results on fuzzy relation equations with sup-inf composition from the viewpoint of decomposition, presents a way to describe the solution set of fuzzy relation equations, and shows a necessary and sufficient condition, which partly answers the open problem for existence of minimal solutions over complete Brouwerian lattices.*

Keywords— Complete Brouwerian lattice, fuzzy relation equation, minimal solution, solution set, sup-inf composition.

1 Introduction and preliminaries

Let I, J be index sets and $A = (a_{ij})_{I\times J}$ be a coefficient matrix, $B = (b_i)_{i\in I}^T$ be a constant column vector (the sign "T" denotes the "transpose"). Then $A \odot X = B$ or

$$\bigvee_{j\in J} (a_{ij} \wedge x_j) = b_i, i \in I \tag{1}$$

is called a *fuzzy relation equation* assigned on a lattice L, where $\odot$ denotes the sup-inf composite operation, and all x_j, b_i, a_{ij}'s are in L. An X which satisfies (1) is called a *solution* of (1), the *solution set* of (1) is denoted by $\mathscr{X}_1 = \{X : A \odot X = B\}$. A special case of (1) is as follows: $A \odot X = b$ or

$$\bigvee_{j\in J} (a_j \wedge x_j) = b, \tag{2}$$

where $b \in L$, $A = (a_j)_{j\in J}$ is a row vector. Denote $\mathscr{X}_2 = \{X : A \odot X = b\}$ the solution set of (2).

The solvability of fuzzy relation equations in complete Brouwerian lattices was first proposed in connection with medical diagnosis problems in [24]. Sanchez [24] showed that every solvable fuzzy relation equation assigned on complete Brouwerian lattices has the greatest solution. Since then, the resolution of fuzzy relation equations over lattices has been a theme of continuous interest in fuzzy inference and fuzzy systems theory. A number of works in this area were published (see, e.g. [2,6,8,9,12,17,19,20,23,25,27,28,29,30,32,33,38]). Many results are obtained when L is a linear lattice (see, e.g. [3, 7, 13, 14, 15, 18, 26]). There are also many results considered fuzzy relation equations over semi-linear spaces (see [10, 16, 21, 22] for details), and many works dealt with the more general fuzzy relation equations on complete lattices, such as $sup_{j\in J}\mathscr{T}(a_j, x_j) = b$, where $\mathscr{T}$ is a t-norm (see [5]) or a pseudo-t-norm (see [11,34]) or a conjunctor (see [31,37]).

*Supported by the National Natural Science Foundation of China (No.10671138).

This paper mainly focuses on the fuzzy relation equations over complete Brouwerian lattices.

We first recall some definitions and results of lattice theory which will be used in the sequel. Let $(P, \leq)$ be a partially ordered set and $X \subseteq P$. $p \in X$ is a *minimal* element of X if there is no $x \in X$ such that $x < p$. The *greatest* element of X is an element $g \in X$ such that $x \leq g$ for all $x \in X$. A *lattice* is a poset $L = (L, \leq)$ any two of whose elements have a g.l.b. or "meet" denoted by $x \wedge y$, and a l.u.b. or "join" denoted by $x \vee y$. A lattice L is *complete* when each of its subset T has a l.u.b. and a g.l.b. in L. An element a of a lattice L is *join-irreducible* if $x \vee y = a$ implies $x = a$ or $y = a$. A lattice is *Brouwerian* if for any pair of elements $a, b \in L$, the greatest element $x \in L$, denoted by $a\alpha b$, satisfying the inequality $a \wedge x \leq b$ exists. It is easy to verify the following properties: for any a, b, c in a complete Brouwerian lattice L, $b \leq a\alpha(a \wedge b), a\alpha(b \wedge c) = (a\alpha b) \wedge (a\alpha c), a \wedge (a\alpha b) = a \wedge b$. It is well known that any Brouwerian lattice L is distributive, and a complete lattice L is Brouwerian if and only if (iff)

$$x \wedge (\bigvee_{i\in I} y_i) = \bigvee_{i\in I} (x \wedge y_i)$$

for any $x \in L$ and any family of elements $\{y_i \in L : i \in I\}$. Let p be a join-irreducible element of a distributive lattice L. Then $p \leq \bigvee_{i=1}^{k} x_i$ iff $p \leq x_i$ for some x_i. Some other definitions and results of lattice theory which we do not list here are from [1, 4].

2 The sets of solutions for (1)

Throughout the paper, L is always assumed to be a complete Brouwerian lattice with universal bounds 0 and 1, I and J are infinite index sets, unless otherwise specified. Denote $X^* = (x_j^*)_{j\in J}^T$ a greatest solution, $\underline{k} = \{1, 2, \cdots, k\}$ for any positive integer k and $A - B = \{x \in A : x \notin B\}$, where A and B are two (crisp) sets.

It is trivial that $a \wedge x = b$ (resp. $a \vee x = b$) is solvable iff $a \geq b$ (resp. $a \leq b$). Denote $[a]_b$ (resp. $\overline{[a]}_b$) its solution set. For the solvable case, if L is a distributive lattice then $[a]_b = [b) - \bigcup_{b<d\leq a} [d)$ and $\overline{[a]}_b = (b] - \bigcup_{a\leq d<a} (d]$ (see [36]); if L is a complete Brouwerian lattice then $[a]_b = [b, a\alpha b]$ and $\overline{[a]}_b = (b] - \bigcup_{a\leq d<a} (d]$, where $[a) = \{x \in L : x \geq a\}$ and $(a] = \{x \in L : x \leq a\}$ (see [12]). In what follows, for any $i \in I$, denote

$$\mathscr{R}^i = \{R^i = (q_{ij})_{j\in J}^T : b_i = \bigvee_{j\in J} q_{ij}, q_{ij} \leq a_{ij}, \forall j \in J\},$$

ISBN: 978-989-95079-6-8

$\mathcal{X}^i$ the solution set of $\bigvee_{j\in J}(a_{ij}\wedge x_j)=b_i$ and $\mathcal{Q}=\{Q=(q_{ij})_{I\times J}:\bigvee_{j\in J}q_{ij}=b_i,\forall i\in I;\bigcap_{i\in I}[a_{ij}]_{q_{ij}}\neq\emptyset,\forall j\in J\}$. Then we can verify the following criteria of solvability easily.

Theorem 2.1 (Di Nola [9]) *$\mathcal{X}_1\neq\emptyset$ iff $\mathcal{Q}\neq\emptyset$.*

Though we omit its proof, it should be pointed out that the proof of Theorem 2.1 suggests a way to build other elements, different from the greatest element $X^*=(\bigwedge_{i\in I}(a_{ij}\alpha b_i))^T_{j\in J}$, of $\mathcal{X}_1$. What is more, Theorem 2.1 holds when L is a complete lattice.

Note that for any $i\in I$, the i-th row of any $Q\in\mathcal{Q}$ forms a decomposition of the i-th component of B. Moreover, $b_i=\bigvee_{j\in J}q_{ij}$ and $q_{ij}\leq a_{ij}$, $\forall j\in J$, therefore $\bigvee_{j\in J}(a_{ij}\wedge q_{ij})=\bigvee_{j\in J}q_{ij}=b_i$, $\forall i\in I$, thus the transpose of the i-th row of Q solves the i-th equation in (1), i.e. $\mathcal{R}^i\subseteq\mathcal{X}^i,\forall i\in I$.

Let $\mathcal{Q}_j=\bigcap_{i\in I}[a_{ij}]_{q_{ij}}$ for any $j\in J$. Then the following theorem is true.

Theorem 2.2 (Di Nola [9]) *Let $\mathcal{X}_1\neq\emptyset$. Then for any $j\in J$, $\mathcal{Q}_j=[\bigvee_{i\in I}q_{ij},\bigwedge_{i\in I}(a_{ij}\alpha q_{ij})]$, where $(q_{ij})_{I\times J}\in\mathcal{Q}$. Further, put $R=(r_j)^T_{j\in J}$ with $r_j\in\mathcal{Q}_j$ for all $j\in J$, then $R\in\mathcal{X}_1$.*

Note that Theorem 2.2 gives the solution set of (1) when $|J|=1$, i.e. $[\bigvee_{i\in I}b_i,\bigwedge_{i\in I}(a_i\alpha b_i)]$ is the solution set of the system of equations $a_i\wedge x=b_i$, $i\in I$. Then the following equivalent conditions is straightforward.

Corollary 2.1 (Han [12]) *For the system of equations $a_i\wedge x=b_i$, $i\in\underline{m}$ ($m\geq 2$), the following conditions are equivalent:*
(1) The system of equations $a_i\wedge x=b_i$, $i\in\underline{m}$ is solvable;
(2) $a_i\geq b_i$ and $a_j\wedge b_i=b_j\wedge b_i$ for all $i,j\in\underline{m}$;
(3) $\bigvee_{i=1}^m b_i$ is its solution;
(4) $\bigwedge_{i=1}^m(a_i\alpha b_i)$ is its solution.
Further, the solution set is $[\bigvee_{i=1}^m b_i,\bigwedge_{i=1}^m(a_i\alpha b_i)]$.

Remark 2.1 Conditions (1), (2) and (3) in Corollary 2.1 are also equivalent when the system is assigned on distributive lattices.

By Remark 2.1, the following corollary is obvious when L is a distributive lattice.

Corollary 2.2 (Zhang [35]) *The system of equations $a_i\wedge x=b_i$, $i\in\underline{m}$ is solvable iff $\bigvee_{i=1}^m b_i$ is its solution. And the solution set is $[\bigvee_{i=1}^m a_i]_{\bigvee_{i=1}^m b_i}$.*

Combining Corollaries 2.1 and 2.2, we get a property of α operation as follows:

Proposition 2.1 *If $a_i\geq b_i$ and $a_j\wedge b_i=b_j\wedge b_i$ for all $i,j\in\underline{m}$, then $(\bigvee_{i=1}^m a_i)\alpha(\bigvee_{i=1}^m b_i)=\bigwedge_{i=1}^m(a_i\alpha b_i)$.*

Proof. If $a_i\geq b_i$ and $a_j\wedge b_i=b_j\wedge b_i$ for all $i,j\in\underline{m}$, then $a_i\wedge x=b_i$, $i\in\underline{m}$ is solvable and its solution set is $[\bigvee_{i=1}^m b_i,\bigwedge_{i=1}^m(a_i\alpha b_i)]$ by Corollary 2.1. Therefore from Corollary 2.2, we have $[\bigvee_{i=1}^m b_i,(\bigvee_{i=1}^m a_i)\alpha(\bigvee_{i=1}^m b_i)]=[\bigvee_{i=1}^m a_i]_{\bigvee_{i=1}^m b_i}=[\bigvee_{i=1}^m b_i,\bigwedge_{i=1}^m(a_i\alpha b_i)]$. Thus $(\bigvee_{i=1}^m a_i)\alpha(\bigvee_{i=1}^m b_i)=\bigwedge_{i=1}^m(a_i\alpha b_i)$. □

Notice that, in general, $(\bigvee_{i=1}^m a_i)\alpha(\bigvee_{i=1}^m b_i)=\bigwedge_{i=1}^m(a_i\alpha b_i)$ does not always hold.

Example 2.1 Let N be the set of nonnegative integers. We define $a\wedge b=$ l.c.m.$\{a,b\}$, $a\vee b=$ g.c.d.$\{a,b\}$, $a\leq b$ iff a is multiple of b, where $a,b\in N$ and l.c.m. (resp. g.c.d.) stands for the smallest (resp. greatest) common multiple (resp. divisor) between a and b. Then $L=(N,\wedge,\vee,\leq)$ is a complete Brouwerian lattice with operator "α" given by $a\alpha b=$ g.c.d.$\{x\in N:a\wedge x\leq b\}$ for any $a,b\in N$. $(2\vee 4)\alpha(3\vee 5)=2\alpha 1=1$, but $(2\alpha 3)\wedge(4\alpha 5)=3\wedge 5=15$.

When L is a distributive lattice, $I=\underline{m}$ and $J=\underline{n}$, the next two corollaries are straightforward by Theorem 2.1 and Corollary 2.2.

Corollary 2.3 (Zhang [35]) *Equation (1) is solvable iff there exists a matrix $Q=(q_{ij})_{m\times n}$ such that $\bigvee_{j=1}^n q_{ij}=b_i$, and $a_{ij}\wedge(\bigvee_{k=1}^m q_{kj})=q_{ij}$, $i\in\underline{m}$ and $j\in\underline{n}$. Denote (Q) the set of such Q. Then $\mathcal{X}_1=\bigcup_{Q\in(Q)}H_Q$, where $H_Q=\{(r_j)^T_{j\in\underline{n}}:r_j\in[\bigvee_{i=1}^m a_{ij}]_{\bigvee_{i=1}^m q_{ij}},\forall j\in\underline{n}\}$.*

Corollary 2.4 (Zhang [36]) *For given $B'=(b'_i)^T_{i\in\underline{m}}$, $A'=(a'_i)_{i\in\underline{n}}$ and B, there exists X such that $B'\odot A'\odot X=B$ iff $\bigvee_{j=1}^n a'_j\geq\bigvee_{i=1}^m b_i$ and $b'_k\wedge(\bigvee_{i=1}^m b_i)=b_k$, $\forall k\in\underline{m}$.*

Proof. If there exists $X=(x_j)^T_{j\in\underline{n}}$ such that $B'\odot A'\odot X=B$, then $\bigvee_{j=1}^n(a'_j\wedge x_j)$ is a solution of $b'_i\wedge x=b_i$, $i\in\underline{m}$, thus $\bigvee_{j=1}^n a'_j\geq\bigvee_{j=1}^n(a'_j\wedge x_j)\geq\bigvee_{i=1}^m b_i$. Further, $b'_k\wedge(\bigvee_{i=1}^m b_i)=b_k$, $\forall k\in\underline{m}$ by Corollary 2.2.

Conversely, if $\bigvee_{j=1}^n a'_j\geq\bigvee_{i=1}^m b_i$, then $\bigvee_{j=1}^n[a'_j\wedge(\bigvee_{i=1}^m b_i)]=(\bigvee_{j=1}^n a'_j)\wedge(\bigvee_{i=1}^m b_i)=\bigvee_{i=1}^m b_i$, therefore $X=(x_j)^T_{j\in\underline{n}}$ with $x_j=\bigvee_{i=1}^m b_i$ satisfies $B'\odot A'\odot X=B$ since $b'_k\wedge(\bigvee_{i=1}^m b_i)=b_k$, $\forall k\in\underline{m}$. □

Note that Corollary 2.3 was generalized in [37] to fuzzy relation equations with sup-conjunctor composition.

From Theorem 2.2 we know that $R=(r_j)^T_{j\in J}\in\mathcal{X}_1$ with $r_j\in[\bigvee_{i\in I}q_{ij},\bigwedge_{i\in I}(a_{ij}\alpha q_{ij})]$ for all $j\in J$. Particularly, take $R=(\bigvee_{i\in I}q_{ij})^T_{j\in J}$ with $(q_{ij})_{I\times J}\in\mathcal{Q}$ and $\mathcal{R}=\{R=(r_j)^T_{j\in J}:r_j=\bigvee_{i\in I}q_{ij},\forall j\in J,(q_{ij})_{I\times J}\in\mathcal{Q}\}$. It is easy

ISBN: 978-989-95079-6-8

to verify $\mathscr{R} \subseteq \mathscr{X}_1$. For any $X = (x_j)_{i\in J}^T \in \mathscr{X}_1$, we have $R^i = (q_{ij})_{j\in J}^T = (a_{ij} \wedge x_j)_{j\in J}^T \in \mathscr{R}^i, \forall i \in I$. Therefore, $\mathscr{R}$ possesses similar properties as $\mathscr{X}_1$. For instance, we have:

Proposition 2.2 *If $R^1 = (r_j^1)_{j\in J}^T$, $R^2 = (r_j^2)_{j\in J}^T \in \mathscr{R}$ and $R = (r_j)_{j\in J}^T$ such that $R^1 \leq R \leq R^2$, then $R \in \mathscr{R}$.*

Proposition 2.3 *If $R^1 = (r_j^1)_{j\in J}^T$, $R^2 = (r_j^2)_{j\in J}^T \in \mathscr{R}$, then $R^1 \vee R^2 \in \mathscr{R}$.*

Proposition 2.4 *$\mathscr{R} \neq \emptyset$ iff $R^* = ((\bigvee_{i\in I} a_{ij}) \wedge [\bigwedge_{i\in I} (a_{ij}\alpha b_i)])_{j\in J}^T \in \mathscr{R}$. Further, R^* is the greatest element of $\mathscr{R}$.*

Note that, from Proposition 2.4 we know if $|\mathscr{X}_1| = 1$ then $\bigwedge_{i\in I} (a_{ij}\alpha b_i) \leq \bigvee_{i\in I} a_{ij}$ for all $j \in J$, since $R^* \in \mathscr{R} \subseteq \mathscr{X}_1$ must be equal to the greatest element of $\mathscr{X}_1$ (otherwise, $|\mathscr{X}_1| > 1$). By the definition of $\mathscr{R}$, we can build an element of $\mathscr{X}_1$ from $\mathscr{R}^i$, $i \in I$, which may be different from the greatest element in $\mathscr{X}_1$.

Example 2.2 Let L be the lattice considered in Example 2.1. Consider the fuzzy relational equation

$$\begin{cases} (3 \wedge x_1) \vee (7 \wedge x_2) = 3, \\ (6 \wedge x_1) \vee (8 \wedge x_2) = 12. \end{cases} \tag{3}$$

We can see $(12, 21)^T \in \mathscr{R}^1$, $(12, 24)^T \in \mathscr{R}^2$, i.e. $12 \vee 21 = 3$, $12 \vee 24 = 12$. And further, $4 \in [3]_{12} \cap [6]_{12}$, $3 \in [7]_{21} \cap [8]_{24}$, then $(12 \vee 12, 21 \vee 24)^T = (12, 3)^T \in \mathscr{R} \subseteq \mathscr{X}_1$, which is different from the the greatest element $(4, 3)^T$.

Let $\mathscr{R} \neq \emptyset$. From the definition of $\mathscr{R}$, we know that R is an element of $\mathscr{R}$ if and only if there exist $R^i = (q_{ij})_{j\in J}^T \in \mathscr{R}^i$, $\forall i \in I$ such that $R = \bigvee_{i\in I} R^i$ and $\bigcap_{i\in I} [a_{ij}]_{q_{ij}} \neq \emptyset$, $\forall j \in J$. This result has a generalized version considered over semi-linear spaces (see [10, 21, 22] for details) as follows:

Theorem 2.3 (Nosková [16]) *Let (1) be solvable. Then X is a solution to (1) iff there exist $X_i \in \mathcal{X}^i$, $i \in \underline{m}$, such that $X = \bigvee_{i=1}^m X_i$ and $\bigvee_{i=1}^m X_i \leq X^*$.*

For every $i \in I$, denote $\mathcal{S}^i = \bigcup_{R^i \in \mathscr{R}^i} S_{R^i}$, where $R^i = (q_{ij})_{j\in J}^T \in \mathscr{R}^i$ and $S_{R^i} = \{S = (s_j)_{j\in J}^T : s_j \in [a_{ij}]_{q_{ij}}, \forall j \in J\}$. Then $\mathscr{R}^i \subseteq \mathcal{S}^i$ since $q_{ij} \in [a_{ij}]_{q_{ij}}$. Further,

Theorem 2.4 *$\mathcal{S}^i = \mathcal{X}^i$, $\forall i \in I$.*

Proof. For any $i \in I$, if $S = (s_j)_{j\in J}^T \in \mathcal{S}^i$, then there exists $R^i = (q_{ij})_{j\in J}^T \in \mathscr{R}^i$ such that $S \in S_{R^i}$, i.e. $s_j \in [a_{ij}]_{q_{ij}}, \forall j \in J$. Thus $\bigvee_{j\in J} (a_{ij} \wedge s_j) = \bigvee_{j\in J} q_{ij} = b_i$. Therefore $S \in \mathcal{X}^i$, i.e. $\mathcal{S}^i \subseteq \mathcal{X}^i$. Vice versa, for any $i \in I$, if $X = (x_j)_{j\in J}^T \in \mathcal{X}^i$, then $\bigvee_{j\in J} (a_{ij} \wedge x_j) = b_i$. Let $R^i = (a_{ij} \wedge x_j)_{j\in J}^T$, i.e. $R^i \in \mathscr{R}^i$. Then $X \in S_{R^i} \subseteq \mathcal{S}^i$ since $x_j \in [a_{ij}]_{a_{ij}\wedge x_j}$, $\forall j \in J$. Thus $\mathcal{X}^i \subseteq \mathcal{S}^i$, together with $\mathcal{S}^i \subseteq \mathcal{X}^i$ we conclude $\mathcal{S}^i = \mathcal{X}^i$. □

Since $\mathscr{X}_1 = \bigcap_{i\in I} \mathcal{X}^i$, then the following theorem is straightforward by Theorem 2.4.

Theorem 2.5 $\mathscr{X}_1 = \bigcap_{i\in I} \bigcup_{R^i \in \mathscr{R}^i} S_{R^i}$.

From Theorems 2.4 and 2.5, we know that it is important to find all elements of $\mathscr{R}^i$, $i \in I$, in order to determine $\mathscr{X}_1$.

In the sequel, we will consider a method to determine all elements of $\mathscr{R}^i$, $i \in I$ then describe the solution set of (1).

Theorem 2.6 *For any $i \in I$, if $\mathscr{R}^i \neq \emptyset$ and $R^* = (q_{ij}^*)_{j\in J}^T = (a_{ij} \wedge b_i)_{j\in J}^T$ is the greatest element of $\mathscr{R}^i$, then every element $R = (q_{ij})_{j\in J}^T$ of $\mathscr{R}^i$ is determined by*

$$q_{ij} \in \overline{[\bigvee_{k\in J, k\neq j} q_{ik}^j]}_{b_i} \bigcap (a_{ij}],$$

where

$$q_{ik}^j = \begin{cases} q_{ik}, & k \in D, \\ q_{ik}^*, & k \in J - D \end{cases}$$

and $D = \{j \in J : q_{ij}$ has been determined $\}$.

Proof. For any $R = (q_{ij})_{j\in J}^T \in \mathscr{R}^i$ and $j \in J$, $b_i = \bigvee_{k\in J} q_{ik}^* \geq q_{ij} \vee (\bigvee_{k\in J, k\neq j} q_{ik}^j) \geq q_{ij} \vee (\bigvee_{k\in J, k\neq j} q_{ik}) = b_i$, then $q_{ij} \in \overline{[\bigvee_{k\in J, k\neq j} q_{ik}^j]}_{b_i} \bigcap (a_{ij}]$ since $q_{ij} \leq a_{ij}, \forall j \in J$ by the definition of $\mathscr{R}^i$. Vice versa is straightforward by the definition of q_{ij} and q_{ik}^j. □

By Theorems 2.5 and 2.6 the following results is obvious.

Corollary 2.5 *For any $i \in I$, if $R^* = (q_{ij}^*)_{j\in J}^T = (a_{ij} \wedge b_i)_{j\in J}^T$ is the greatest element of $\mathscr{R}^i$, then every element $X = (x_j)_{j\in J}^T$ of $\mathscr{X}_1$ is determined by*

$$x_j \in \bigcap_{i\in I} \bigcup_{q_{ij} \in \overline{[\bigvee_{k\in J, k\neq j} q_{ik}^j]}_{b_i} \bigcap (a_{ij}]} [a_{ij}]_{q_{ij}},$$

where

$$q_{ik}^j = \begin{cases} q_{ik}, & k \in D, \\ q_{ik}^*, & k \in J - D \end{cases}$$

and $D = \{j \in J : q_{ij}$ has been determined $\}$.

Note that Corollary 2.5 generalizes Theorem 3.2 in [12].

As for (2), the corresponding $\mathscr{R}$ mentioned before is nothing but

$$\mathscr{R}_2 = \{R = (r_j)_{j\in J}^T : b = \bigvee_{j\in J} r_j, r_j \leq a_j, \forall j \in J\}.$$

When $J = \underline{n}$, let $b = s_1 = r_1 \vee s_2, s_2 = r_2 \vee s_3, \cdots, s_{n-1} = r_{n-1} \vee s_n, s_n = r_n$. Therefore, $r_1 \vee \cdots \vee r_n = b$ and the following theorem when (2) is assigned on distributive lattices is straightforward, since Theorem 2.4 also holds in distributive lattices in finite case.

Theorem 2.7 (Zhang [35]) $\mathscr{R}_2 = \bigcup_{b=s_1\geq s_2\geq\cdots\geq s_n} \{R = (r_j)_{j\in\underline{n}}^T : r_j \in \overline{[s_{j+1}]}_{s_j} \bigcap (a_j], \forall j \in \underline{n-1}, r_n = s_n\}$. *And* $\mathscr{X}_2 = \bigcup_{b=s_1\geq s_2\geq\cdots\geq s_n} \{X = (x_j)_{j\in\underline{n}}^T : x_j \in [a_j]_{r_j}, \forall j \in \underline{n}, r_j \in \overline{[s_{j+1}]}_{s_j} \bigcap (a_j], \forall j \in \underline{n-1}, r_n = s_n\}$.

ISBN: 978-989-95079-6-8

3 Conditions for existence of minimal solutions to (1)

As it was mentioned in [8] that the determination of minimal elements in $\mathscr{X}_1$, i.e. minimal solutions to (1), when (1) is assigned on complete Brouwerian lattices remains open in finite case as well as in infinite case. In this section we will consider the determination of minimal solutions.

First, we have:

Theorem 3.1 (Di Nola [9]) *If $\mathscr{R}$ has minimal elements then these elements are minimal in $\mathscr{X}_1$, and vice versa.*

As a comment to Theorem 3.1 we have to say we do not know if $\mathscr{X}_1$ (or $\mathscr{R}$) has or has not minimal elements. However, if one suppose the existence of minimal elements in $\mathscr{X}_1$, then these elements must be sought among the minimal elements of $\mathscr{R}$.

Minimal solutions do not always exist, here is an example borrowed from [29].

Example 3.1 Let L be the lattice considered in Example 2.1. If $A = (3, 7)$ and $b = 2$, then (2) has a solution $X = (2, 2)^T$ but no minimal solution since for any $X = (x_1, x_2)^T \in \mathscr{X}_2$, $X' = (2x_1, x_2)^T$ is also an element of $\mathscr{X}_2$ and $X' \leq X$ but $X' \neq X$.

Let $\mathscr{X}_1 \neq \emptyset$ and $\gamma_{j_0}(X) = \{l \in L : (a_{ij_0} \wedge l) \vee [\bigvee_{j\in J, j\neq j_0} (a_{ij} \wedge x_j)] = b_i, \forall i \in I\}$ for every $X = (x_j)^T_{j\in J} \in \mathscr{X}_1$ and $j_0 \in J$. Then:

Theorem 3.2 *For any $X = (x_j)^T_{j\in J} \in \mathscr{X}_1$, if $X' = (x'_j)^T_{j\in J} \in \mathscr{X}_1$ and $X' \geq X$ then $x_j \in \gamma_j(X')$ for all $j \in J$.*

Proof. For any $X = (x_j)^T_{j\in J} \in \mathscr{X}_1$ and $j_0 \in J$, put $\overline{X} = (\overline{x}_j)^T_{j\in J}$ with

$$\overline{x}_j = \begin{cases} x_{j_0}, & j = j_0, \\ x'_j, & j \neq j_0. \end{cases}$$

Then $X \leq \overline{X} \leq X'$, and $\overline{X} \in \mathscr{X}_1$. Therefore, $x_{j_0} \in \gamma_{j_0}(X')$. □

The following theorem gives a criterion to determine when a given solution is minimal:

Theorem 3.3 (Di Nola [9]) *$X = (x_j)^T_{j\in J}$ is a minimal element of $\mathscr{X}_1$ iff $x_{j_0} = \min \gamma_{j_0}(X)$ for every $j_0 \in J$.*

For a given element $R = (r_j)^T_{j\in J} = (\bigvee_{i\in I} q_{ij})^T_{j\in J}$ in $\mathscr{R}$ and each $j_0 \in J$, denote $\overline{\gamma}_{j_0}(R) = \{t \in L : t = \bigvee_{i\in I} t_{ij_0}, (\bigvee_{j\in J, j\neq j_0} q_{ij}) \vee t_{ij_0} = b_i, \forall i \in I\}$, then from Theorems 3.2 and 3.3 the following two theorems are obvious.

Theorem 3.4 *$R = (r_j)^T_{j\in J}$ is a minimal element of $\mathscr{R}$ iff $r_{j_0} = \min \overline{\gamma}_{j_0}(R)$ for all $j_0 \in J$.*

Theorem 3.5 *For any $R = (r_j)^T_{j\in J} \in \mathscr{R}$, if $R' = (r'_j)^T_{j\in J} \in \mathscr{R}$ and $R' \geq R$ then $r_j \in \overline{\gamma}_j(R')$ for all $j \in J$.*

As an immediate consequence of Theorem 3.3, the following theorem on the unicity of solution is obvious.

Theorem 3.6 (Di Nola [9]) *$|\mathscr{X}_1| = 1$ iff for every $j_0 \in J$, $x^*_{j_0} = \bigwedge_{i\in I} (a_{ij_0} \alpha b_i) = \min \gamma_{j_0}(X^*)$.*

In fact, if $\mathscr{X}_1 \neq \emptyset$, then from Theorem 2.2 we have that $|\mathscr{X}_1| = 1$ if and only if $\bigvee_{i\in I} q_{ij} = \bigwedge_{i\in I} (a_{ij} \alpha q_{ij})$ for any $j \in J$ and $(q_{ij})_{I\times J} \in \mathscr{Q}$ (see also Theorem 6 in [25]).

By Theorem 3.1, the following theorem is straightforward.

Theorem 3.7 (Wang [30]) *If $X_* = (x_{j_*})^T_{j\in J}$ is a minimal element of $\mathscr{X}_2$, then $a_j \wedge x_{j_*} = x_{j_*}$ for all $j \in J$, i.e. $b = \bigvee_{j\in J} x_{j_*}$ and $x_{j_*} \leq a_j, \forall j \in J$.*

Though minimal solutions do not always exist, when adding some assumptions to B, we get some interesting results. Here is a sufficient condition for existence of a minimal solution to (1) when I is finite and J is infinite.

Theorem 3.8 (Wang [28]) *If every component of B is a compact element with an irredundant finite join-decomposition, then for each $X \in \mathscr{X}_1$ there exists a minimal element $X_* \in \mathscr{X}_1$ such that $X_* \leq X$.*

Thus the following theorem follows from Theorems 3.1 and 3.8 easily.

Theorem 3.9 *Suppose that each component of B is compact and has an irredundant finite join-decomposition. Then R is minimal in $\overline{\mathscr{R}}$ iff it is minimal in $\mathscr{X}_1$, where $\overline{\mathscr{R}} = \{R : R = \bigvee_{i\in I} R^i, R^i = (q_{ij})^T_{j\in J}$ is minimal in $\mathscr{R}^i, \forall i \in I; \bigcap_{i\in I} [a_{ij}]_{q_{ij}} \neq \emptyset, \forall j \in J\}$.*

Note that Theorem 3.8 was generalized in [31] to fuzzy relation equations with sup-conjunctor composition. With Theorem 3.9 in hands, it is easy to see that if I is finite and every component of B is compact and has an irredundant finite join-decomposition, then $\mathscr{X}_1$ is completely determined by the greatest solution and all minimal solutions. Therefore, Theorem 3.9 generalizes Theorem 3.11 in [8] and the corresponding result in [13] considered over $[0, 1]$. Another generalized version of Theorem 3.11 in [8] is Theorem 4 in [16] considered fuzzy relation equations over semi-linear spaces. When J is also finite, the assumption of compact in Theorem 3.8 can be removed (see Theorem 7.1 in [29]). Further, the following theorem gives the number of minimal solutions of (2) when both I and J are finite.

Theorem 3.10 (Wang [28]) *If $\mathscr{X}_2 \neq \emptyset$ and b has an irredundant finite join-decomposition $\bigvee_{i=1}^{k} p_i$, then the number of minimal solutions of (2) is $\prod_{i\in \underline{k}} |G(p_i)|$, where $G(p_i) = \{j \in \underline{n} : a_j \geq p_i\}$. Further, all minimal solutions $X = (x_j)^T_{j\in J}$ are determined by $x_j = \bigvee_{f_X(i)=j} p_i$ for any mapping $f_X \in \prod_{i\in \underline{k}} G(p_i)$.*

Let both I and J be finite. If for any $i \in I$, b_i has an irredundant finite join-decomposition $\bigvee_{t=1}^{k_i} p_{it}$, then for any $i \in I$, $t_i \in \underline{k_i}$ we define $G(p_{it_i}) = \{j \in J : a_{ij} \geq p_{it_i}\}$ and $F_i = \{f_i : f_i \in \prod_{t_i\in \underline{k_i}} G(p_{it_i}), \bigvee_{f_i(t_i)=j} p_{it_i} \leq \bigwedge_{i\in I} (a_{ij} \alpha b_i), \forall j \in J\}$. Then:

Theorem 3.11 *Let $\mathscr{X}_1 \neq \emptyset$. Then $|\mathscr{X}_1| = 1$ iff for any $j \in J$ and $f_i \in F_i$, $\bigvee_{i\in I} \bigvee_{f_i(t_i)=j} p_{it_i} = \bigwedge_{i\in I} (a_{ij} \alpha b_i)$.*

ISBN: 978-989-95079-6-8

Proof. Note that $X^* = (\bigwedge_{i\in I}(a_{ij}\alpha b_i))_{j\in J}^T \in \mathscr{X}_1$ since $\mathscr{X}_1 \neq \emptyset$. For any $i \in I$ and $f_i \in F_i$, from Theorem 3.10 and the definition of F_i we know that $(\bigvee_{f_i(t_i)=j} p_{it_i})_{j\in J}^T \leq X^*$ and $(\bigvee_{f_i(t_i)=j} p_{it_i})_{j\in J}^T$ is minimal in $\mathcal{X}^i$. Thus $\{X = (x_j)_{j\in J}^T : x_j = \bigvee_{i\in I}\bigvee_{f_i(t_i)=j} p_{it_i}, f_i \in F_i\}$ contains all minimal elements of $\mathscr{X}_1$ from Theorem 3.9. Therefore, the thesis is straightforward. □

Note that Theorem 3 in [7] that considers the unique solvability of fuzzy relation equations over linear lattices and it is just a special case of Theorem 3.11.

When $|I| = 1$ and J is infinite, a necessary and sufficient condition for existence of a minimal solution is as follows:

Theorem 3.12 (Wang [30]) *Let $\mathscr{X}_2 \neq \emptyset$. Then for each $X \in \mathscr{X}_2$ there exists a minimal element X_* of $\mathscr{X}_2$ such that $X_* \leq X$ iff there is a subset B of L with B satisfying:*
(i) $\bigvee B = b$;
(ii) For each $p \in B$, if $p \neq 0$ then $b \neq \bigvee(B - \{p\})$;
(iii) For each $X = (x_j)_{j\in J}^T \in \mathscr{X}_2$ and each $p \in B$ there is an index $k \in J$ such that $p \leq a_k \wedge x_k$.

If I is finite, then from Theorem 3.12 we have

Theorem 3.13 (Wang [30]) *If $\mathscr{X}_1 \neq \emptyset$ and every component b_i, $i \in I$, of B is compact and for each b_i, $i \in I$, there exists a subset B_i of L such that:*
(i) $\bigvee B_i = b_i$;
(ii) For each $p_{it} \in B_i$, if $p_{it} \neq 0$ then $b_i \neq \bigvee(B_i - \{p_{it}\})$;
(iii) For each $X = (x_j)_{j\in J}^T \in \mathcal{X}^i$ and each $p_{it} \in B_i$ there is an index $k \in J$ such that $p_{it} \leq a_{ik} \wedge x_k$;
(iv) For each $p \in \bigcup_{i\in I} B_i$, if $p \neq 0$ then there is no subset Q of $\bigcup_{i\in I} B_i$ such that $p \leq \bigvee(Q - \{p\})$.
Then for each $X \in \mathscr{X}_1$, there exists a minimal element X_* of $\mathscr{X}_1$ such that $X_* \leq X$.

Notice that Theorem 3.8 is a special case of Theorem 3.13. In the following, let $G(b) = \{j \in J : a_j \geq b\}$. Then

Theorem 3.14 *Let $\mathscr{X}_2 \neq \emptyset$. Then there exists a minimal element in $\mathscr{X}_2$ iff either $G(b) \neq \emptyset$ or there is a subset B of L such that:*
(i) $b = \bigvee B$;
(ii) For every $p \in B$, there exists an index $j \in J$ such that $p \leq a_j$;
(iii) For every $p \in B, q \in L$, if $q < p$ then $b \neq [\bigvee(B - \{p\})] \bigvee q$.

Proof. Let $X_* = (x_{*j})_{j\in J}^T$ be minimal in $\mathscr{X}_2$. If $G(b) = \emptyset$ then from Theorem 3.7 $b = \bigvee_{j\in J} x_{*j}$, put $B = \{x_{*j} : x_{*j} \neq 0, j \in J\}$, then B satisfies the conditions (i), (ii) and (iii).

Conversely, if $G(b) \neq \emptyset$ then it is easy to see that there exists a minimal element in $\mathscr{X}_2$. Now suppose that there exists a subset B of L satisfying (i), (ii) and (iii), then we can construct a family of subsets $A_j, j \in J$ of B such that $\bigcup_{j\in J} A_j = B$ and $A_i \bigcap A_k = \emptyset$ when $i \neq k$ and $i, k \in J$. Define $X = (x_j)_{j\in J}^T$ with

$$x_j = \begin{cases} \bigvee_{p\in A_j} p, & A_j \neq \emptyset, \\ 0, & A_j = \emptyset. \end{cases}$$

Then $\bigvee_{j\in J}(a_j \wedge x_j) = \bigvee_{j\in J, A_j\neq\emptyset} x_j = \bigvee(\bigcup_{j\in J} A_j) = \bigvee B = b$, i.e. $X \in \mathscr{X}_2$. Let $Y = (y_j)_{j\in J}^T \in \mathscr{X}_2$ be such that $Y \leq X$. If $x_j = 0$ then $y_j = 0$. If $x_j \neq 0$ then $A_j \neq \emptyset$, therefore $y_j \leq x_j = \bigvee_{p\in A_j} p \leq a_j$ and $b = \bigvee_{j\in J}(a_j \wedge y_j) = \bigvee_{j\in J} y_j$. One can verify that for any $j \in J$ if $A_j \neq \emptyset$ then $y_j = x_j$. Indeed, if there exists an index $j_0 \in J$ such that $A_{j_0} \neq \emptyset$ but $y_{j_0} < x_{j_0}$, put $\overline{Y} = (\overline{y}_j)_{j\in J}$ with

$$\overline{y}_j = \begin{cases} y_{j_0}, & j = j_0, \\ x_j, & j \neq j_0. \end{cases}$$

Then $b = \bigvee_{j\in J} y_j \leq \bigvee_{j\in J} \overline{y}_j \leq \bigvee_{j\in J} x_j = b$ which means $(\bigvee_{j\in J, j\neq j_0} x_j) \bigvee y_{j_0} = b$, contradicts to (iii). Thus $Y = X$, i.e. X is minimal in $\mathscr{X}_2$. □

Note that Theorem 3.14 partly answers the open problem for existence of minimal solutions (see Page 46 in [8]).

It should be pointed out that most of the results in last two sections, such as Theorems 2.1, 2.4, 2.5, 3.1, 3.2 and 3.3, are also hold when L is distributive in finite case.

4 Conclusions

In this paper we gave an overview of the known results for fuzzy relation equations over complete Brouwerian lattices. In our opinions, the resolution of fuzzy relation equations over complete Brouwerian lattices can be solved from the viewpoint of decomposition.

References

[1] G. Birkhoff, Lattice theory. Vol. XXV, 3rd Ed., American Mathematical Society Colloquium Publications, Providence, RI, 1979.

[2] M. M. Bourke, D. G. Fisher, Solution algorithms for fuzzy relational equations with max-product composition, Fuzzy Sets and Systems 94 (1998) 61-69.

[3] K. Cechlarova, Unique solvability of max-min fuzzy equations and strong regularity of matrices over fuzzy algebra, Fuzzy Sets and Systems 75 (1995) 165-177.

[4] P. Crawley, R. P. Dilworth, Algebraic theory of lattices, Prentice-Hall, Englewood Cliffs, NJ, 1973.

[5] B. De Baets, Sup-T equations: state of the art, in: O. Kaynak, L. Zadeh, B. Turksen, I. Rudas (Eds.), Computational Intelligence: Soft Computing and Fuzzy-Neural Integration with Applications, NATO ASI Series F, Vol. 162, Springer, Berlin, 1998, pp. 80-93.

[6] B. De Baets, Analytical solution methods for fuzzy relational equations, in: D. Dubois, H. Prade (Eds.), Fundamentals of Fuzzy Sets, The Handbooks of Fuzzy Sets Series, Vol. 1, Kluwer Academic Publishers, Dordrecht, 2000, pp. 291-340.

[7] A. Di Nola, S. Sessa, Finite fuzzy relation equations with a unique solution in linear lattices, J. Math. Anal. Appl 132 (1988) 39-49.

ISBN: 978-989-95079-6-8

[8] A. Di Nola, S. Sessa, W. Pedrycz, E. Sanchez, Fuzzy relation equations and their applications to knowledge engineering. Kluwer Academic Publishers, Dordrecht, 1989.

[9] A. Di Nola, On solving relational equations in Brouwerian lattices, Fuzzy Sets and Systems 34 (1990) 365-376.

[10] A. Di Nola, A. Lettieri, I. Perfilieva, V. Novák, Algebraic analysis of fuzzy systems, Fuzzy Sets and Systems 158 (2007) 1-22.

[11] S. C. Han, H. X. Li, J. Y. Wang, Resolution of finite fuzzy relation equations based on strong pseudo-t-norms, Appl. Math. Lett. 19 (2006) 752-757.

[12] S. C. Han, H. X. Li, J. Y. Wang, Resolution of matrix equations over arbitrary Brouwerian lattices, Fuzzy Sets and Systems 159 (2008) 40-46.

[13] M. Higashi, G. J. Klir, Resolution of finite fuzzy relation equations, Fuzzy Sets and Systems 13 (1984) 65-82.

[14] H. Imai, M. Miyakoshi and T. Da-te, Some properties of minimal solutions for a fuzzy relation equation, Fuzzy Sets and Systems 90 (1997) 335-340.

[15] H. Imai, K. Kikuchi and M. Miyakoshi, Unattainable solutions of a fuzzy relation equation, Fuzzy Sets and Systems 99 (1998) 193-196.

[16] L. Nosková, I. Perfilieva, System of fuzzy relation equations with sup-* composition in semi-linear spaces: minimal solutions. In Proc. FUZZ-IEEE Conf. on Fuzzy Systems, July 23-36, 2007, London, pp. 1520-1525.

[17] W. Pedrycz, s-t Fuzzy relational equations, Fuzzy Sets and Systems 59 (1993) 189-195.

[18] K. Peeva, Y. Kyosev, Fuzzy relational calculus-theory, applications and software (with CD-ROM), in Series Advances in Fuzzy Systems - Applications and Theory, Vol. 22, World Scientific Publishing Company, 2004.

[19] K. Peeva, Y. Kyosev, Algorithm for solving max-product fuzzy relational equations, Soft Comput 11 (2007) 593-605.

[20] I. Perfilieva, Fuzzy function as an approximate solution to a system of fuzzy relation equations, Fuzzy Sets and Systems 147 (2004) 363-383.

[21] I. Perfilieva, Semi-linear spaces, in: H. Noguchi, H. Ishii, et al.(Eds.), Proc. of Seventh Czech-Japanese Seminar on Data Analysis and Decision Making under Uncertainty, Hyogo, Japan, 2004, pp. 127-130.

[22] I. Perfilieva, Semilinear spaces-basic structures for fuzzy systems, in: Proc. Conf. of IPMU'2006, Paris, France, 2006, pp. 1087-1094.

[23] I. Perfilieva, V. Novák, System of fuzzy relation equations as a continuous model of IF-THEN rules, Information Sciences 177 (2007) 3218-3227.

[24] E. Sanchez, Resolution of composite fuzzy relation equations, Inform. and Control 30 (1976) 38-48.

[25] S. Sessa, Finite fuzzy relation equations with unique solution in complete Brouwerian lattices, Fuzzy Sets and Systems 29 (1989) 103-113.

[26] M. Wagenknecht, K. Hartmann, On the existence of minimal solutions for fuzzy equations with tolerances, Fuzzy Sets and Systems 34 (1990) 237-244.

[27] X. P. Wang, Method of solution to fuzzy relation equations in a complete Brouwerian lattice, Fuzzy Sets and Systems 120 (2001) 409-414.

[28] X. P. Wang, Infinite fuzzy relational equations in a complete Brouwerian lattice, Indian J. Pure Appl. Math. 33 (2002) 87-95.

[29] X. P. Wang, Conditions under which a fuzzy relational equation has minimal solutions in a complete Brouwerian lattice, Advances In Mathematics 31 (2002) 220-228 (in Chinese).

[30] X. P. Wang, Infinite fuzzy relational equations on a complete Brouwerian lattice, Fuzzy Sets and Systems 138 (2003) 657-666.

[31] X. P. Wang, Q. Q. Xiong, The solution set of a fuzzy relational equation with sup-conjunctor composition in a complete lattice, Fuzzy Sets and Systems 153 (2005) 249-260.

[32] X. P. Wang, X. B. Qu, Continuous join-irreducible elements and their applications to describing the solution set of fuzzy relational equations, Acta Mathematica Sinica, Chinese Series 49 (2006) 1171-1180 (in Chinese).

[33] X. B. Qu, X. P. Wang, Some properties of infinite fuzzy relational equations on complete Brouwerian lattices, Fuzzy Sets and Systems 158 (2007) 1327-1339.

[34] Z. D. Wang, Y. D. Yu, Pseudo-t-norms and implication operators on a complete Brouwerian lattice, Fuzzy Sets and Systems 132 (2002) 113-124.

[35] K. L. Zhang, matrix equations in distributive lattices, Acta Scientiarum Naturalium Universitatis Intramongolicas 22 (1991) 311-317.

[36] K. L. Zhang, The inverse problem for distributive lattices, Fuzzy Sets and Systems 59 (1993) 109-113.

[37] K. L. Zhang, D. H. Li, L. X. Song, On finite relation equations with sup-conjunctor composition over a complete lattice, Fuzzy Sets and Systems 160 (2009) 119-128.

[38] C. K. Zhao, On matrix equations in a class of complete and completely distributive lattices, Fuzzy Sets and Systems 22 (1987) 303-320.

ISBN: 978-989-95079-6-8

Aggregation operators for conditional crispness

Doretta Vivona[1] Maria Divari[2]

1.2.Faculty of Engineering, "Sapienza" University of Rome 1
Rome , Italy
Email: vivona@dmmm.uniroma1.it, maria.divari@alice.it,

Abstract— *In axiomatic way, in fuzzy setting, we introduce the measure of crispness and the measure of conditional clearness. For these measures we propose a functional form. Then, for the last measure, we give some aggregation operators.*

Keywords— Aggregation operators, functional equations, fuzzy sets, measures of fuzziness, crispness measures, conditional crispness.

1 Introduction

De Luca and Termini gave the first definition of fuzziness measure [8, 15], based upon a condition of monotonicity, with respect to a antural relation of less or grater fuzziness. In 1975 Mukaidono presented another order, called monotonicity with respect to ambiguity [11]. Later Couso and Gil have modified the definition [7] and Roventa used the cuts of fuzzy sets in order to define this kind of entropy [12, 13, 14]. In [2] the authors, taking into account some consideration due to Wang [18] introduced small variations, leanding to a more exact order relation among fuzzy sets, which represents the idea of less or greater fuzziness.

We know that in $\mathcal{F}$ there are two different partial order: the classical inclusion relation [19] and the so called *fuzziness relation* according to the greater or smaller "proximity" to a crisp set; with respect to the last relation it is possible to define the infimum and the supremum [16]. In [2, 3, 4] we have studied the fuzziness measure and in [16] there are many examples of fuzziness measure also by using Sugeno's integral.

These researches have suggested to introduced, in axiomatic way, the concept of crispness c through a fuzziness measure d.

Then, taking into account the crispness, we give the definition of conditional crispness.

We know that an aggregation operator is a procedure by which a unique value can be associated to the result obtained through different tests or different values of data base. The unique value is kind of aritmetic mean.

In this setting, in this paper, we give the definition of measure of crispness (shortly crispness) through the fuzziness measure and the definition of conditional crispness.

Then, we propose a characterization of classes of aggregation operators for the conditional crispness.

2 Preliminaires

Given an abstract space X and a σ-algebra $\mathcal{C}$ of crisp subsets C of X, let $\mathcal{F}$ be the family of all $\mathcal{C}$-measurable fuzzy sets [18, 19]. With every fuzzy set F a membership function f is associated, $f \to [0,1]$, which is $\mathcal{C}$-measurable.

Now, we recall the order relation called *fuzziness relation* [2, 3]: F_1 is *less fuzzy* then F_2 ($F_1 \preceq F_2$) if the following condition holds $\forall\, x \in X$:

$$\begin{cases} f_1(x) \le \frac{1}{2} \implies f_1(x) \le f_2(x) \le \frac{1}{2} \\ f_1(x) \ge \frac{1}{2} \implies \frac{1}{2} \le f_2(x) \le f_1(x)\,, \end{cases}$$

which is equivalent to

$$\begin{cases} f_1(x) < \frac{1}{2} \implies f_1(x) \le f_2(x) \\ f_1(x) > \frac{1}{2} \implies f_1(x) \ge f_2(x)\,. \end{cases}$$

We remember that the fuzziness relation is different and not comparable with the inclusion relation:
$F_1 \subset F_2 \iff f_1(x) \le f_2(x)\ \forall\, x \in X$.
For crisp sets the relation $C_1 \preceq C_2$ implies $C_1 = C_2$.

We know [2] that $\left[\frac{X}{2}\right]$ and $\left[\frac{C}{2}\right]$ are those fuzzy sets, whose membership function are $f(x) = \frac{1}{2}, \forall x \in X$, and $f(x) = \frac{1}{2}, \forall x \in C$, respectively. Moreover we shall indicate with F^c the complement of F.

Definition 1 *The fuzziness measure is a map $d : \mathcal{F} \to [0,1]$ such that:*

$[A_1]\ d(C) = 0\ ,\ \forall\, C \in \mathcal{C}$;

$[A_2]\ F_1 \preceq F_2 \implies d(F_1) \le d(F_2)\ \forall F_1, F_2 \in \mathcal{F}$;

$[A_3]\ d\left(\left[\frac{X}{2}\right]\right) = 1$;

$[A_4]\ d(F) = d(F^c)\ \forall F \in \mathcal{F}$.

Moreover, we have recognized in [2] that to every fuzziness measure d we can associate a fuzzy measure μ on $\mathcal{C}$. In fact, the function $\mu : \mathcal{C} \to [0,1]$, defined by $\mu(C) = d\left(\left[\frac{C}{2}\right]\right)$, is non-decreasing with respect to the inclusion relation and $\mu(\emptyset) = 0$. This measure μ is called by us *fuzzy measure associated to d*. In [2] it is showed a form of this measure by using the Sugeno integral.

ISBN: 978-989-95079-6-8

3 Crispness and conditional crispness for fuzzy sets

In this paragraph we introduce the definition of crispness by using the fuzziness measure.

Definition 2 *Given a fuzziness measure* d, *the* crispness *of a fuzzy set* F *is a map* $c : \mathcal{F} \to [0,1]$ *defined by*

$$c(F) = 1 - d(F)\ . \tag{1}$$

Using the properties above $[A_1] - [A_4]$ the crispness (1) enjoys the following:

$[A_1^{'}]\ c(C) = 1\ ,\ \forall\ C \in \mathcal{C},\ \text{ as d(C)} = 0;$

$[A_2^{'}]\ F_1 \preceq F_2 \implies c(F_1) \geq c(F_2),$
as $d(F_1) \leq d(F_2)\ \forall F_1, F_2 \in \mathcal{F}$;

$[A_3^{'}]\ c\left(\left[\frac{X}{2}\right]\right) = 0,\ \text{ as } d\left(\left[\frac{X}{2}\right]\right) = 1;$

$[A_4^{'}]\ c(F) = c(F^c)\ \forall F \in \mathcal{F},\ \text{ as d(F)} = \text{d}(\text{F}^{\text{c}})\ .$

As an extension of crisp setting [9], we assume that two fuzzy sets F and H are *algebrically independent* if $F \cap H \neq \emptyset$.

Definition 3 *Fixed a fuzzy set* $H \in \mathcal{F}$ *and a crispness (1), the* conditional crispness *of any variable set* $F \in \mathcal{F}$ *is a map*

$$c_H(\cdot) : \mathcal{F} \to [0,1]$$

defined by the following axioms:

$(i)\ \ F_1 \preceq F_2 \implies c_H(F_1) \geq c_H(F_2), \forall F_1, F_2 \in \mathcal{F};$

$(ii)\ c_H\left(\left[\frac{X}{2}\right]\right) = 0,\ \left[\frac{X}{2}\right] \in \mathcal{C},$

$(iii)\ c_H(F) = c(F), \forall F \in \mathcal{F},$
if F is c-independent by H.

We assume the (iii) as the c-independence i.e.independence with respect to c.

As conseguence, every crisp set C is c-independent from H:
$c_H(C) = c(H),\ \forall C \in \mathcal{C}$.

4 Example

Let F be a fuzzy set composed by a group of old men. Given a fuzziness measure d, it is possible to calculate $d(F)$ and $c(F)$ according with (1). Moreover, we consider another fuzzy set H, whose elements are sick men. For us, the meaning of $c_H(F)$ is that the crispness of F influenced by the sickness of the men in H.

5 A form of measure of conditional crispness

Now, we propose the following form for the conditional crispness measure:

$$c_H(F) = \Phi\Big(c(F), c(H)\Big) \tag{2}$$

where $\Phi : [0,1] \times [0,1] \to [0,1]$. From $[(i), (ii), (iii)]$, we obtain:

$(a) F_1 \preceq F_2 \implies \Phi\Big(c(F_1), c(H)\Big) \geq \Phi\Big(c(F_2), c(H)\Big),$
$(b)\ \Phi(0, c(H)) = 0,$
$(c)\ \Phi(1, c(H)) = c(H).$

If we put $x = c(F_1), x' = c(F_2), y = c(H)$, with $, x, x', y \in [0,1]$, we get:

$$\begin{cases} (a')\ \Phi(1,y) = y \\ \\ (b')\ \Phi(0,y) = 0 \\ \\ (c')\ \Phi(x,y) \geq \Phi(x',y) \text{ if } x \geq x'\ . \end{cases}$$

For the system $[(a') - (c')]$ we have the following:

Proposition 1 *A class of solution of the system* $[(a') - (c')]$ *is*

$$\Phi = h^{-1}\Big(h(x) \cdot h(y)\Big) \tag{3}$$

where $h : [0,1] \times [0,1] \to [0,1]$ *is increasing with* $h(0) = 0$ *and* $h(1) = 1$.

Proof. The proof is immediate.

6 Example

If $h(x) = x^n, n \in (0, +\infty)$, then $\Phi(x,y) = x \cdot y$.

7 Aggregation operators for conditional crispness measures

In [17] we have studied the aggregation operators of the general conditional information. We shall use again the same procedure for the characterization of some forms of aggregation operators of the conditional crispness.

Let $\mathcal{I}$ be the family of the conditional crispness measures $c_H(\cdot)$. The aggregation operator $L : \mathcal{I} \to [0, \text{K}]$, $0 < \text{K} < [0, +\infty]$ of $n \in [0, +\infty)$ conditional crispness measures $c_H(F_1), ..., c_H(F_i), ..., c_H(F_n)$, with $F_i \in \mathcal{F}, \text{i} = 1, ..., \text{n}, \text{H} \in \mathcal{F}$ has the following properies as in [10]:

$(I)\ idempotence : c_H(F_i) = \lambda, \forall i = 1, ..., n \implies$

$$L(\underbrace{\lambda, ..., \lambda}_{n\ times}) = \lambda;$$

$(II)\ monotonicity : c_H(F_1) \leq c_H(F_1^{'}) \implies$
$L(c_H(F_1), ..., c_H(F_i), ..., c_H(F_n)) \leq$
$L(c_H(F_1^{'}), ..., c_H(F_i), ..., c_H(F_n)),$
$F_1^{'}, F_i \in \mathcal{F}, \text{i} = 1, ..., \text{n};$

$(III)\ continuity\ from\ below : c_H(F_{1,m}) \nearrow c_H(F_1) \implies$

ISBN: 978-989-95079-6-8

$c_H(F_{1,m}), ..., c_H(F_i), ..., c_H(F_n)) \quad \nearrow$
$c_H(F_1), ..., c_H(F_i), ..., c_H(F_n))$ $F_1^{'}, F_i \in \mathcal{F}, \mathrm{i} = 1, ..., \mathrm{n}.$

Putting $c_H(F_i) = x_i, i = 1, ...n, c_H(F_1^{'}) = x_1^{'}, c_H(F_{1,m}) = x_{1,m}$, with $x_i, i = 1, ...n, x_1^{'}, x_{1,m} \in [0,1]$, we obtain the following system of functional equations:

$$\begin{cases} (I')\ L(\underbrace{\lambda, ..., \lambda}_{n\ times}) = \lambda, \\ (II')\ x_1 \leq x_1^{'} \Longrightarrow L(x_1, ..., x_n) \leq L(x_1^{'}, ...x_n), \\ (III')\ x_{1,m} \nearrow x_1 \Longrightarrow L(x_{1,m}, ..., x_n) \nearrow L(x_1, ...x_n). \end{cases}$$

For the solution of the system $[(I') - (III')]$, we propose the following:

Proposition 2 *Two natural solutions of the system* $[(I') - (III')]$ *are*

$$L(x_1, ..., x_n) = \bigwedge_{i=1}^{n} x_i$$

and

$$L(x_1, ..., x_n) = \bigvee_{i=1}^{n} x_i.$$

Proof. The proof is immediate.

Proposition 3 *A class of solution of the system* $[(I') - (III')]$ *is*

$$L(x_1, ..., x_n) = h^{-1}\left(\frac{h(x_1) + ... + h(x_n)}{n}\right),$$

where $h : [0,1] \to [0,K](0 < K < +\infty)$ *is a continuous, strictly increasing function with* $h(0) = 0$ *and* $h(1) = K$.

Proof. The proof is immediate.

8 Remark

If the function h is linear, then the aggregation operator L is the aritmetic mean.

9 Conclusion

First, we have given the definitions of the crispness and the conditional crispness for a fuzzy set.

Second, we have proposed some classes of aggregation operators of the conditional crispness, solving a suitable system of functional equations.

References

[1] J. ACZEL: *Lectures on functional equations and their applications*, (1966), New Jork, Academic Press .

[2] P.BENVENUTI - D.VIVONA - M.DIVARI: On fuzziness measures via Sugeno integral, *Fuzzy logic and Soft Computing, Advances in Fuzzy Systems* **v.IV** (1995), 330-336.

[3] P.BENVENUTI - D.VIVONA - M.DIVARI: Equiordered Fuzzy Sets and Fuzziness Measure, *Proc.VII IFSA Congress,* (1997), pp.409-412.

[4] P.BENVENUTI - D.VIVONA: Order relation for fuzzy sets and entropy measures, *New Trend in Fuzzy Systems,* Ed.D.Mancini, M.Squillante and A.Ventre, World Scientific, (1998), pp.224-232.

[5] P.BENVENUTI - D.VIVONA - M.DIVARI: Aggregation operators and associated fuzzy mesures, *Int.Journ.of Uncertainty, Fuzziness and Knownledge-Based Systems,* **2**, (2001), 197-204.

[6] T.CALVO - G.MAYOR - R.MESIAR: *Aggregation operators,* New Trend and Applications, Physisc-Verlag, Heidelberg, (2002).

[7] I COUSO - P.GIL: Measures of fuzziness of type 2 fuzzy sets. *Proc.IPMU 96,* **2**, (1996), pp.581-584.

[8] A.DE LUCA - S.TERMINI: A definition of non probabilistic entropy in the setting of fuzzy sets theory. *Infor.and Contr.,* **20**, (1972), pp.301-312.

[9] B.FORTE: Measure of information. The general axiomatic theory. *R.I.R.O.* (1969), pp.63-99.

[10] E.P.KLEMENT: Binary aggregation operators which are bounded by the minimum (a mathematical overview), *Proc.AGOP05,* (2005), 13-16.

[11] M.MUKAIDONO: On some properties of fuzzy logic. *Syst.-Comput.-Control,* **6**, n.2, (1975), pp.36-43.

[12] E.ROVENTA: On the Degree of Fuzziness of a Fuzzy Sets, *FSS,* **36** (1990) pp.259-264.

[13] E.ROVENTA: On a Class of Measures of Fuzziness, *J.Math.Anal.Appl.,* **175** (1993) pp.337-341.

[14] E.ROVENTA - D.VIVONA: Representation Theorem for transom based measures of fuzziness, *Rend.Mat.Roma ,* **16** *rm (1996)* .

[15] S.TERMINI: Remarks on the measures of fuzziness. *Scuola Estiva "Logica Borosa"* El Escorial, Madrid (1992).

[16] D.VIVONA: Mathematical Aspect of the Theory of Measures of Fuzziness. *Mathware and Soft Computing,* **3** (1996), pp.211-224.

[17] D.VIVONA - M.DIVARI: Aggregation operators for conditional information without probability, *Proc.IPMU08,* (2008), pp.258-260.

ISBN: 978-989-95079-6-8

[18] Z.WANG – ZI-XIAO: Fuzzy measures and measures of fuzziness. *J.Math.An.Appl.,* **104** (1984), 581-601.

[19] L.A.ZADEH: Fuzzy sets, *Inf. and Control,* **8** (1965), pp.338-353.

ISBN: 978-989-95079-6-8

A New Conditioning Rule, Its Generalization and Evidential Reasoning

Koichi YAMADA[1] Vilany KIMALA[2] Muneyuki UNEHARA[1]

1 Faculty of Engineering, Nagaoka University of Technology
2 Graduate School of Engineering, Nagaoka University of Technology
1603-1, Kami-tomioka, Nagaoka, Niigata, Japan, 940-2188
Email: yamada@kjs.nagaokaut.ac.jp, kimala@mis.nagaokaut.ac.jp, unehara@kjs.nagaokaut.ac.jp

Abstract *- In Evidence theory, several conditioning rules for updating belief have been proposed, including Dempster's rule of conditioning. The paper views the conditioning rules proposed so far and proposes a new rule of conditioning based on three requirements. Then, it generalizes the rule to be applied to the case where condition is given by an uncertain belief. The paper also discusses a few interpretations of an equation used for evidential reasoning, one of which is interpreted as conditioning with an uncertain condition.*

Keywords— Dempster-Shafer theory of evidence, evidence theory, conditioning rule, evidential reasoning, uncertainty.

1 Introduction

The paper discusses a new conditioning rule of Evidence theory that we proposed recently [1,2] and generalizes the rule to be applied to the case where condition is given by an uncertain belief. Then, it discusses a few interpretations of an equation used for evidential reasoning, one of which is interpreted as conditioning with an uncertain condition.

The first and well-known conditioning rule in the theory is Dempster's rule of conditioning [3]. However, it has been pointed out that the rule has a few problems, and several alternatives have been proposed so far [4-13]. A well-known problem is that a small change of prior belief, which is represented as basic belief assignment (*bba*) or mass function in the paper, may produce a large change of posterior belief [4]. Another is a practical problem that conditional belief is undefined, when a given condition *totally conflicts* with the prior belief in the sense of $Pl(B) = 0$, where B is the condition represented by a subset of a frame of discernment Θ and Pl is the plausibility measure equivalent to the prior belief.

Dempster's rule of *conditioning* is derived as a special case of Dempster's rule of *combination*. Suppose there are two distinct beliefs represented by *bba*'s m, m', and the total mass of m' is assigned to a set B, *i.e.* $m'(B) = 1$. In the case, the belief combined from m and m' is the same as the conditional belief $m(\bullet \mid B)$. This means that both information of the prior belief m and the condition B are treated equivalently or symmetrically.

The paper deals with the two pieces of information asymmetrically. In the basic conditioning rule, which is one with a sure condition without any uncertainty in the paper, it is assumed that the condition is a piece of information given *a posteriori* and always certain, while the prior belief is given *a priori* and may not be correct. When they *partially conflict* with each other (*i.e.* $X \cap B = \varnothing$ for a X, where X is a focal element of the prior belief m, and $'$ is the empty set), the proposed conditioning trusts in B rather than X, and reassigns the mass $m(X)$ to B. The simple idea produces a conditioning rule different from conventional ones proposed so far. Then, the paper expands the basic conditioning to the one with an uncertain condition, and discusses the relation between the conditioning rule and evidential reasoning.

The paper is composed as follows; section 2 reviews Dempster-Shafer theory of evidence and major conditioning rules proposed so far including those with an uncertain condition represented by a belief. Section 3 discusses requirements that the basic conditioning rule should satisfy, and proposes a new basic conditioning. Section 4 expands the conditioning rule and proposes another generalized rule by removing a requirement. Section 5 discusses interpretations of an equation for evidential reasoning, one of which is a conditioning rule with an uncertain condition.

2 Conditioning Rules in Evidence Theory

2.1 Dempster's Theory of Evidence [3]

Let Θ be a frame of discernment. A function $m : 2^{\Theta} \rightarrow [0,1]$ satisfying the following equations represents a *bba*;

$$m(\varnothing) = 0, \tag{1}$$

$$\sum_{A \in 2^{\Theta}} m(A) = 1, \tag{2}$$

where 2^{Θ} is the power set of Θ. $m(A)$ is a degree of belief that a variable whose domain is Θ has the value in A, but has no information about elements and subsets of A. Subsets A satisfying $m(A) > 0$ are called focal elements.

The belief could be represented in other forms, such as Belief and Plausibility measures defined below.

$$Bel(A) = \sum_{X \subseteq A} m(X) \ . \tag{3}$$

$$Pl(A) = \sum_{X \cap A \neq \varnothing} m(X) \ . \tag{4}$$

Both measures are functions from 2^{Θ} to [0,1], and satisfy $Bel(\varnothing) = Pl(\varnothing) = 0$ and $Bel(\Theta) = Pl(\Theta) = 1$. Functions $m(\bullet)$, $Bel(\bullet)$ and $Pl(\bullet)$ are all transformable one another, and represent the same belief state.

ISBN: 978-989-95079-6-8

When two distinct beliefs m_1 and m_2 are derived from two independent information sources respectively, they could be combined by the famous rule, Dempster's rule of combination;

$$m_D(A) = \frac{\sum_{A=X\cap Y} m_1(X) \bullet m_2(Y)}{1 - \sum_{X\cap Y=\varnothing} m_1(X) \bullet m_2(Y)}. \quad (5)$$

where $A \neq \varnothing$ and $m_D(\varnothing) = 0$. If the denominator is zero (*i.e.* m_1 and m_2 totally conflict with each other) the combination is not defined.

When $m_1(\bullet) = m(\bullet)$ and $m_2(B) = 1$, the combination gives the next equation called Dempster's rule of conditioning.

$$m_D(A \mid B) = \frac{\sum_{A=X\cap B} m(X)}{1 - \sum_{X\cap B=\varnothing} m(X)}, \quad (6)$$

where $A' \in$ and $m_D(' \mid B) = 0$. When $Pl(B) = 0$ or m and B totally conflict with each other, the denominator becomes zero and $m_D(A \mid B)$ is undefined.

2.2 Alternative Conditioning Rules

As discussed in Introduction, many alternative conditioning rules have been proposed so far. Major rules are shown below ($B \neq \varnothing$ is assumed).

(i) Focusing rule

The rule is proposed in [7, 8] separately and is called focusing rule in [10].

$$Bel_F(A \mid B) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(\overline{A} \cap B)}. \quad (7a)$$

$$Pl_F(A \mid B) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(\overline{A} \cap B)}. \quad (7b)$$

Representation with *bba* is studied in [13].

(ii) Zhang's rule

Dempster's rule reassigns all mass of $m(X)$ to $X \cap B$. Zhang asserts that it is "radical" and proposes the next rule [11].

$$m_Z(A \mid B) = k \sum_{X\cap B=A} \frac{|X \cap B|}{X} m(X). \quad (8)$$

where k is a constant for normalization.

(iii) Geometric rule [5]

This is a strong rule in the sense that all mass of focal elements that are not included in B is abandoned.

$$m_G(A \mid B) = \begin{cases} m(A) \Big/ \sum_{X\in B} m(X), & \text{if } A \in B, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

(iv) Planchet's Conditional rule [6]

This is a weak rule in the sense that only mass of focal elements that have no intersection with B is abandoned.

$$m_P(A \mid B) = \begin{cases} m(A) \Big/ \sum_{X\in B'} m(X), & \text{if } A \in B', \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

(v) Ito&Inagaki's rule

Ito and Inagaki proposed a rule assuming that the condition is just an assertion of an information source [12].

a) In the case of $Bel(\overline{B}) < 1$,

$$Bel_I(A \mid B) = \begin{cases} \{1 - k_B Bel(\overline{B})\}\{Bel(A \cup \overline{B}) - Bel(\overline{B})\}, & \text{if } A \neq \Theta, \\ 1, & \text{if } A = \Theta, \end{cases} \quad (11a)$$

where k_B is a constant that satisfies $0 \le k_B \le 1/(1 - Bel(\overline{B}))$.

b) In the case of $Bel(\overline{B}) = 1$,

$$Bel_I(A \mid B) = \begin{cases} 1, & \text{if } A = ', \\ 0, & \text{otherwise.} \end{cases} \quad (11b)$$

(vi) Rules with an uncertain condition

Ichihashi and Tanaka proposed three Jeffrey-like conditioning rules with an uncertain condition [14].

a) Plausible conditioning

$$m_{pl}(A \mid m_b) = \sum_{A=X\in Y} \frac{m(X) \bullet m_b(Y)}{1 - \sum_{X\in Y=\varnothing} m(X)}, \quad (12)$$

where $m_b(\bullet)$ is a *bba* on Θ and $m_{pl}(\varnothing \mid m_b) = 0$. It is assumed $\sum_{X\cap Y=\varnothing} m(X) < 1$ for any focal element Y of m_b.

b) Credible conditioning

$$m_{cr}(A \mid m_b) = \sum_{A\subseteq Y} \frac{m(A) \bullet m_b(Y)}{1 - \sum_{X\not\subset Y, X\neq Y} m(X)}, \quad (13)$$

where $\sum_{X\in Y, X'\ Y} m(X) < 1$ for any focal element Y of m_b. $X \subset Y$ means that X is a proper subset of Y in the paper.

c) Possible conditioning

$$m_{po}(A \mid m_b) = \sum_{A\cap Y\neq\varnothing} \frac{m(A) \bullet m_b(Y)}{1 - \sum_{X\cap Y=\varnothing} m(X)}, \quad (14)$$

where $\sum_{X\cap Y=\varnothing} m(X) < 1$ for any focal element Y of m_b.

When m is Bayesian, and focal elements of m_b partition Θ, all of m_{pl}, m_{cr} and m_{po} equal to Jeffrey's rule. When $m_b(B) = 1$ for a specific B, m_{pl}, m_{cr} and m_{po} equal to

ISBN: 978-989-95079-6-8

Dempster's rule, Geometric rule and Planchet's rule, respectively.

3 Basic Conditioning Rule

In the paper, the basic conditioning rule means the one with a sure condition without any uncertainty. It assumes the following:

The given condition is exactly true information given temporally after the prior belief. It puts a strict restriction on focal elements of the posterior belief.

In short, the condition B is an established fact, and the true value actually exists in B. Thus, we trust in B rather than the prior belief m, when they partially conflict with each other, *i.e.* $B \cap X = \varnothing$ for an X, where X is a focal element of m. Based on the above idea, the requirements that the posterior belief must satisfy are discussed below.

(1) Requirement-1: $m(A \mid B) = 0$ when $A \cap \overline{B} \neq \varnothing$, where $\overline{B}$ is the complement of B. Assumed the condition B is an exact fact, the requirement might be self-evident. When the requirement holds, the flowing equations also hold.

$$m(A \mid B) > 0 \Leftarrow A' B, \tag{15}$$

$$B' A \Leftarrow Bel(A \mid B) = 1, \tag{16}$$

$$Bel(A \mid B) = Bel(A \cap B \mid B), \tag{17}$$

where $'$ is implication.

(2) Requirement-2: $m(A \mid B)$ must be defined except the case where $B = '$. This requirement comes from the assertion by Shafer [15] and Smets [9] that Evidence theory must be interpreted as a theory of belief, not as a theory of frequency. If *bba*, *Bel* and *Pl* represent degrees of a human belief, they should not be recognized as completely and precisely correct scales. Even if B is not plausible, *i.e.* $Pl(B) = 0$, a fact B may happen in the actual world. If the posterior belief conditioned by B is undefined in those cases, it might be a useless theory in real applications.

(3) Requirement 3: $m(\bullet \mid \Theta) = m(\bullet)$. The condition Θ represents the true value is in $'$. The condition is clearly unnecessary, unless we deal with an open world problem like Transferable Belief Model [16], which is an expanded evidence theory in the sense that $m(') > 0$ is allowed.

No conditioning rule in the previous section satisfies all the above requirements. All rules satisfy the 3rd one. Dempster's, Focusing, Zhang's and geometric rules satisfy the 1st one. However, nothing satisfies the 2nd requirement [1,2].

Conditioning rules could be derived from various combination rules in the same way as how Dempster's conditioning rule is derived as shown in section 2.1. However, such rules derived from combination rules proposed in [20-23] do not satisfy all the requirements, either. They do not satisfy the 1st requirement [1,2].

In this section, a new conditioning rule, which satisfies all the requirements, is proposed. In the proposed rule, the mass of focal elements of the prior belief are reassigned to another subset in the posterior belief as follows:

(*a***)** When a focal element X of the prior belief and the condition B are consistent (*i.e.* $X \cap B \neq \varnothing$), the condition B imposes restriction on the focal element, the possible area where the true value may exist. That is, the mass $m(X)$ is re-assigned to $X' B$.

(*b***)** When X and B partially conflict with each other (*i.e.* $X \cap B = \varnothing$), the grounds for the mass given to X are lost. Thus, the mass is temporarily given to $'$ representing total ignorance. Then, the restriction by B is imposed. As a result, the mass $m(X)$ is re-assigned to $\Theta \cap B = B$.

From the above (*a*) and (*b*), we get the next equations.

$$m_K(A \mid B) = \begin{cases} \Leftarrow m(X), \text{ if } A' \mid \ , A \{ B, \\ A = X \Leftarrow B \\ \Leftarrow m(X) + \Leftarrow m(X), \text{ if } A = B, \\ A = B \lfloor X \qquad X \Leftarrow B = \mid \\ 0, \text{ otherwise,} \end{cases} \tag{18a}$$

where $B \neq \varnothing$. This equation could be written as

$$m_K(A \mid B) = \sum_{A = X \cap B} m(X) + \sum_{\substack{X \cap B = \varnothing \\ A = B}} m(X), \tag{18b}$$

where $A \neq \varnothing, B \neq \varnothing$ and $m_K(\varnothing \mid B) = 0$.

The idea in (*b*) that the mass is given to $'$ in case of partial conflict is the same as one used in Yager's combination rule [21]. However, the conditioning rule derived from Yager's rule is different from the above, and does not satisfy the 1st requirement [1,2].

It is proved easily that $m_K(A \mid B)$ is a *bba* and that it satisfies the three requirements. It is also clear that $m_K(A \mid B) = m_D(A \mid B)$ when m and B are consistent (= all focal elements of m and B are consistent).

Then, the next inequality could be proved [2].

$$Pl_K(A \mid B)' Pl_D(A \mid B)' Bel_D(A \mid B)' Bel_K(A \mid B), \tag{19}$$

where $Pl_K(A \mid B)$ and $Pl_D(A \mid B)$ ($Bel_K(A \mid B)$ and $Bel_D(A \mid B)$) are plausibility (belief) functions calculated using eq.s (18) and (6), respectively.

$m_K(A \mid B)$ is not a generalization of Bayesian conditioning. When m is Bayesian, $Pl(B) = P(B)$ holds. In the case of $Pl(B) = P(B) = 0$ where $B \neq \varnothing$, $P(A \mid B)$ is undefined, while $m_K(A \mid B)$ must be defined due to Requirement-2. In addition, $m_K(A \mid B)$ is not Bayesian in general, when $|B| > 1$.

Associativity does not hold in this conditioning, namely, $m_K(A \mid B \mid C) \neq m_K(A \mid C \mid B) \neq m_K(A \mid B \cap C)$ for $B \cap C \neq \varnothing$, where $m_K(A \mid B \mid C) = m_K^B(A \mid C)$ and $m_K^B(A) =$

ISBN: 978-989-95079-6-8

$m_K(A \mid B)$. Non-associativity is caused by asymmetric nature of the conditioning between the two pieces of information, the prior belief and the condition.

Suppose the case where $X \cap B \cap C = \emptyset$, $X \cap B \neq \emptyset$, $X \cap C \neq \emptyset$ and $B \in C'\in$, where X is a focal element of the prior belief m. When calculating $m_K(\bullet \mid B)$, $m(X)$ is reassigned to $X \cap B$. Then, when calculating $m_K(\bullet \mid B \mid C)$, the mass $m(X)$ given to X' B is reassigned again to C, because $X \cap B \cap C = \emptyset$. On the other hand, the same mass $m(X)$ is reassigned to B when calculating $m_K(\bullet \mid C \mid B)$, and to $B \cap C$ when calculating $m_K(\bullet \mid B \cap C)$.

4 Generalized Rule of Conditioning

4.1 Conditioning Rule with Unreliable Condition

The basic conditioning rule is generalized by moderating the 1st requirement. First, we introduce an index λ $(0 \le \lambda \le 1)$ to represent reliability of the condition. When $\lambda < 1$, the condition is not completely correct, therefore requirement-1 is abandoned.

The mass reassignment in this case is an expansion of the one in the basic conditional rule, and is conducted as described below.

(a') When $X \in B'\in$, reassign mass of $'m(X)$ to X' B, and $(1-\lambda)m(X)$ to $X \cap \Theta = X$.

(b') When $X \cap B = \emptyset$, reassign mass of $\lambda m(X)$ to $\in'$ $B = B$, and $(1-\lambda)m(X)$ to X' $\in = X$.

In (a'), the degree that B is trusted is λ. Thus, only $\lambda m(X)$ is reassigned to $X \cap B$. The rest $(1-\lambda)m(X)$ is reassigned to $X \cap \Theta = E$, with replacing B by Θ, because B cannot be trusted at the degree of $1-\lambda$.

In the case of (b'), $\lambda m(X)$ is reassigned to $\Theta \cap B = B$, because the mass $\lambda m(X)$ is temporarily given to total ignorance Θ due to $X \cap B = \emptyset$, then restriction B is imposed. The rest $(1'\in)m(X)$ is reassigned to $X \cap \Theta = X$, because B cannot be trusted at the degree of $1-\lambda$.

From the above re-assignment, we get the next equation.

$$m_\lambda(A \mid B) = \begin{cases} \sum_{A = X \cap B} \lambda m(X) + (1-\lambda)m(A), \\ \qquad\qquad \text{if } B \supset A \neq \emptyset, \\ \sum_{A = B \subseteq X} \lambda m(X) + (1-\lambda)m(B) \\ \qquad + \sum_{X \cap B = \emptyset} \lambda m(X), \ \text{if } A = B, \\ (1-\lambda)m(A), \ \text{if } A \not\subset B, A \neq B, A \neq \emptyset, \\ 0, \ \text{otherwise}, \end{cases} \tag{19}$$

where $B'\in$. It is proved easily that $m_\lambda(A \mid B)$ is a *bba*. In addition, it is also proved that $m_\lambda(A \mid B)$ satisfies both the requirements 2 and 3, and that the next equation holds [2].

$$m_\lambda(A \mid B) = \lambda m_K(A \mid B) + (1-\lambda)m(A). \tag{20}$$

The equation could be understood as discounting of condition in conditioning, while discounting of information source in combination rule was proposed in [3].

4.2 Conditioning Rule with Uncertain Condition

The basic conditional rule is also generalized so that an uncertain condition given by *bba* m_b would be accepted in the same way as plausible, credible and possible conditioning rules in Section 2.

The mass reassignment in the case is similar to (a'') and (b''). The difference is i) reliability λ of condition B is replaced by *bba* $m_b(Y)$ of a focal element Y of m_b, and that ii) re-assignment of unreliable mass corresponding to $(1-\lambda)m(X)$ is unnecessary. The mass reassignment is done as follows:

(a'') When a focal element X of the prior belief m and a focal element Y of the uncertain condition m_b are consistent (*i.e.* $X \cap Y \neq \emptyset$ for any X and Y), the condition Y imposes restriction on the focal element X with uncertainty $m_b(Y)$. As a result, mass $m(X) \bullet m_b(Y)$ is re-assigned to $X \cap Y$.

(b'') When X and Y partially conflict with each other (*i.e.* $X \cap Y = \emptyset$), the grounds for the mass given to X are lost. Thus, the mass is temporarily given to $'$ representing total ignorance. Then, the restriction by Y is imposed with uncertainty $m_b(Y)$. As a result, the mass $m(X) \bullet m_b(Y)$ is re-assigned to $\in'$ $Y = Y$.

From (a'') and (b''), we get the following conditioning rule, if $A \neq \emptyset$.

$$\begin{aligned} m_{KYU}(A \mid m_b) &= \sum_{A = X \cap Y} m(X) \bullet m_b(Y) + \sum_{\substack{X \cap Y = \emptyset \\ A = Y}} m(X) \bullet m_b(Y) \\ &= \sum_Y m_b(Y) \left(\sum_{A = X \cap Y} m(X) + \sum_{\substack{X \cap Y = \emptyset \\ A = Y}} m(X) \right) \\ &= \sum_{Y \subseteq \Theta} m_K(A \mid Y) \bullet m_b(Y). \end{aligned} \tag{21}$$

It is proved easily that $m_{KYU}(A \mid m_b)$ is a *bba* and that it satisfies both the requirements 2 and 3.

The above equation shows that posterior belief with uncertain condition can be obtained by weighted sum of the basic conditioning rule $m_K(A \mid B)$ with different conditions B, where weight is given by the mass of the conditional set B of the belief representing the uncertain condition.

The following is obtained easily from eq. (20) and (21).

$$\begin{aligned} m_{\in}(A \mid B) &= \in n_K(A \mid B) + (1'\in)m(A) \\ &= \in n_K(A \mid B) + (1'\in)m_K(A \mid \in) \\ &= m_{KYU}(A \mid m_{(B,\in)}), \end{aligned} \tag{22}$$

ISBN: 978-989-95079-6-8

where $m_{(B,')}(B) = '$ and $m_{(B,\lambda)}(\Theta) = 1 - \lambda$.

In addition, it is proved easily that $m_{KYU}(A \mid m_b) = m_{pl}(A \mid m_b)$, when m and m_b are consistent each other, (*i.e.* $X \cap Y \neq \varnothing$ for all combination of X and Y, where X is a focal element of m and Y is of m_b). However, $m_{KYU}(A \mid m_b)$ is not an extension of Jeffrey's rule unlike $m_{pl}(A \mid m_b)$, because $m_K(A \mid B)$ is not Bayesian in general, even if the prior belief m is Bayesian.

5 Interpretations of Evidential Reasoning

The section shows that the conditioning represented by eq. (21) could be used as an interpretation of evidential reasoning. First, we review a few conventional discussions about evidential reasoning.

5.1 Representation of Rules

There are two ways to represent rules in evidential reasoning: joint form and conditional form [17]. The former represents a rule by a joint belief function on a product $\Theta_y \times \Theta_x$, where Θ_y, Θ_x are frames of discernment for variables y, x in antecedent and consequent of rules, respectively. The latter represents a rule by a conditional belief such as $m(A \mid B)$, where $B \subseteq \Theta_y$, $A \subseteq \Theta_x$. Though papers adopting joint form seem published more than those with conditional form, conditional form is more natural and easier for users to understand rules than joint form [17,18]. In addition, conditional form needs fewer values as knowledge than joint form; number of values conditional form needs is $2^{|'_x|} \bullet 2^{|'_y|} = 2^{|'_x|+|'_y|}$, while joint form needs $2^{|\Theta_x|\bullet|\Theta_y|}$, where $|\bullet|$ represents cardinality [17, 18].

5.2 Conventional Interpretations with Conditional Form

We have two kinds of knowledge. One is a set of rules described as

$$if\ y \in B\ then\ x \in A\ with\ belief\ m^{(1)}(A \mid B),$$

and the other is a prior belief $m^{(2)}(B)$ about y. We know all values of $m^{(1)}(A \mid B)$ and $m^{(2)}(B)$ for all combinations of $B \in '_y$ and $A \in '_x$ except for $m^{(1)}(A \mid \varnothing)$.

There are a few theories for evidential reasoning with conditional form of rules. Liu, *et al*. proposes the next equation for evidential reasoning [19].

$$m(A) = \sum_{B \subseteq \Theta_y} m^{(1)}(A \mid B) \bullet m^{(2)}(B). \tag{23}$$

The theoretical support of the equation is Bayesian formula of probabilistic reasoning with cylindrical extension.

$$P(A_E) = \sum_{B_E \in 2^{\Theta_y} \times 2^{\Theta_x}} P^{(1)}(A_E \mid B_E) \bullet P^{(2)}(B_E), \tag{24}$$

where P, $P^{(1)}$, $P^{(2)}$ are probability measures on $2^{\Theta_y} \times 2^{\Theta_x}$, and $B_E = \{B\}'\ 2^{\in_x}$, $A_E = 2^{\Theta_y} \times \{A\}$. It is assumed that $P(A_E) = m(A)$, $P^{(1)}(A_E \mid B_E) = m^{(1)}(A \mid B)$ and $P^{(2)}(B_E) = m^{(2)}(B)$.

Smets proposed evidential reasoning for Transferable Belief Model [16] based on conjunctive rule of combination in [18]. Conjunctive combination is equivalent to Dempster's rule combination without normalization and is defined by

$$m_\cap(A) = \sum_{A = X \cap Y} m'(X) \bullet m''(Y), \tag{25}$$

where m' and m'' are distinct *beliefs* defined on the same frame of discernment Θ.

The equation is proved to be equivalent to the next equation in [5].

$$m_\cap(A) = \sum_{Y \subseteq \Theta} m'_\cap(A \mid Y) \bullet m''(Y), \tag{26}$$

where $m_\cap(A \mid B)$ is conjunctive rule of conditioning, which is equal to $m_D(A \mid B)$ without normalization.

What should be noted is that reasoning with eq. (26) deviates from the standard evidential reasoning, if m' and m'' partially conflict with each other (*i.e.* $X \cap Y = \varnothing$ for a pair of X and Y, which are focal elements of $m'(\bullet)$ and $m''(\bullet)$, respectively), because $m_\cap(\in) > 0$ is derived.

Fortunately, the deviation from the standard evidential reasoning never happens, if m', m'' are *bba*'s on $\Theta = \Theta_y \times \Theta_x$, and *A, X and Y* are cylindrical extensions of $A' \subseteq \Theta_x$, $X \in '_x$ and $Y' \subseteq \Theta_y$, respectively (*i.e.* $A = \in_y' A \in$, $X = \Theta_y \times X'$, $Y = Y' \times \Theta_x$), because $X \cap Y = Y' \times X' \neq \varnothing$ for any non-empty X' and Y'. Therefore, $m_\cap$ obtained from eq. (26) is surely a *bba* on $\Theta_y \times \Theta_x$. Then, we can derive the following reasoning equation for the standard evidence theory.

$$m(A') = \sum_{Y' \subseteq \Theta_y} m^{(1)}(A' \mid Y') \bullet m^{(2)}(Y'), \tag{27}$$

where it is assumed that $m^{(1)}(A' \mid Y') = m'_\cap(A \mid Y)$, $m^{(2)}(Y') = m''(Y)$, $m(A') = m_\cap(A)$.

The two reasoning equations (23) and (27) are actually the same. That is, the evidential reasoning can be interpreted from two viewpoints: one is from probabilistic reasoning, and the other is from conjunctive combination. Note that conjunctive combination and conjunctive conditioning could be replaced by Dempster's combination and Dempster's conditioning, respectively in this case, because m' and m'' defined on $\Theta_y \times \Theta_x$ are consistent and normalization of Dempster's rule is unnecessary.

5.3 New Interpretation as Uncertain Conditioning

ISBN: 978-989-95079-6-8

Let us recall the conditioning rule with an uncertain condition given by eq. (21). We re-write the equation here.

$$m_{KYU}(A \mid m_b) = \sum_{Y \subseteq \Theta} m_K(A \mid Y) \bullet m_b(Y).$$

The equation is apparently similar to eq. (26). However, it is different, because $m_{\cap}(\varnothing) > 0$ may occur in eq. (26), while $m_{KYU}(\varnothing) = 0$ always holds.

However, if $m_K(A \mid Y)$ and $m_b(Y)$ are *bba*'s on $\in = \in_y{}' \in_x$, and $A = \in_y{}' A\in$, $Y = Y' \times \Theta_x$, $A\in\in{}'_x$, $Y' \subseteq \Theta_y$ similar to the case of eq. (26), $m_K(A \mid Y) = m_D(A \mid Y) = m_{\cap}(A \mid Y)$ holds as discussed in Section 3. Thus, equations (21) and (26) are actually the same in this case.

As a result, the equation (23) or (27) of evidential reasoning could be interpreted from the viewpoint of conditioning given a belief including uncertainty (eq. (21)), as well as from the viewpoint of probabilistic reasoning (eq. 24)) and conjunctive combination (eq. (25)).

6 Conclusions

The paper discussed and proposed a new basic conditioning rule of Evidence theory and generalizes the rule to be applied to the case where condition is given by an uncertain belief. Then, it discusses a few interpretations of an equation used for evidential reasoning, one of which is interpretation by conditioning with an uncertain condition.

The paper first discussed three requirements that should be satisfied by a basic conditioning rule. Then, it proposed a new basic rule that satisfies all of them, however, none of major conventional rules does.

The basic conditioning rule was generalized by abandoning one of the three requirements in order to cope with the case where condition is given as an uncertain belief. The generalized rule is obtained by weighted sum of the basic posterior beliefs with different conditions, where weight is the mass of conditional set in the belief given as an uncertain condition.

Then the paper discussed theoretical backgrounds of an equation used for evidential reasoning, which could be supported by probabilistic reasoning or by conjunctive combination of evidence. The paper showed that it could also be supported by a viewpoint of a conditioning rule with an uncertain condition.

References

[1] V. Kimala, Y. Koichi: Another conditioning rule in Evidence Theory, *11th Inter. Conf. on Humans and Computers*, Nagaoka, 323-328 (2008)

[2] V. Kimala, Y. Koichi, M. Unehara: Requirements of conditioning rules and a new conditioning in Evidence theory (submitted, in Japanese)

[3] G.A.Shafer: *Mathematical Theory of Evidence*, Princeton Univ. Press (1976)

[4] L.A.Zadeh: Review of books, a mathematical theory of evidence, *AI Magazine,* **5**-3, 81-83 (1984)

[5] D.Dubois, H.prade: On the unicity of Dempster rule of combination, *Int. J. of Intelligent Systems* **1**, 133-142 (1986)

[6] B. Planchet: Credibility and conditioning, *J. of Theoretical Probability* **2**, 289-299 (1989)

[7] L. M. De Campos, M.T. Lamata, S. Moral: The concept of conditional fuzzy measure, *Int. J. of Intelligent Systems* **5**, 237-246 (1990)

[8] J. Y. Halpern, R. Fagin : A new approach to updating beliefs, in *Uncertainty in Artificial Intelligence* **6** (eds. P. P. Bonissone, M. Henrion, L. N. Kanal, J. F. Lemmer), 347-373 (1991)

[9] P. Smets: What is Dempster-Shafer's model?, in *Advances in the Dempster-Shafer Theory of Evidence* (eds. R.R.Yager, M. Fedrizzi and J. Kacprzyk), John Wiley & Sons, 5-34 (1994)

[10] D.Dubois, H.Prade: Focusing versus updating in belief function theory, the same book as [9], 71-95.

[11] L. Zhang: Representation, independence, and combination of evidence in the Dempster-Shafer theory, the same as [9], 51-61.

[12]M.Itoh, T.Inagaki: A new conditioning rule for Belief Updating in the Dempster-Shafer theory of evidence, *J. of the society of instrument and control engineers,* **31**-12, 2011-2017 (1995) (in Japanese)

[13] E.C.Kulasekere, *et al.* : Conditioning and updating evidence, *Int.J of Approximate Reasoning* **36**, 75-108 (2004)

[14] H. Ichihashi, H. Tanaka: Jeffrey-like rules of conditioning for the Dempster-Shafer theory of evidence, *Int. J. of Approximate Reasoning* **3**, 143-156 (1989)

[15] G. Shafer: Perspective on the theory and practice of belief function, *Int. J of Approximate Reasoning* **4**, 323-362 (1990)

[16] P. Smets, R. Kennes: The transferable belief model, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **12**-5, 447-458 (1990)

[17] B. B. Yaghlane, K. Mellouli: Inference in directed evidential networks based on the transferable belief model, *Int. J. of Approximate Reasoning* **48**, 399-418 (2008).

[18] P. Smets: Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, *Int. J. of Approximate Reasoning* **9**, 1-35 (1993)

[19] W. Liu, J.G.Hughes, M.F.Mctear: Representing Heuristic knowledge and propagating beliefs in the Dempster-Shafer theory of evidence, in *the same book as [9],* 441-471.

[20] R.R.Yager: On the relationships of methods of aggregation of evidence in expert systems, *Cybernet. Systems* 16, 1/21 (1985)

[21] R.R. Yager: On the Dempster-Shafer framework and new combination rules, *Inform. Sci*. 41, 93/137 (1987)

[22] D.Dubois, H.Prade: Representation and combination of uncertainty with belief functions and possibility measures, *Comput. Intel.* 1(1),244/264(1988)

[23] H.Y. Hau, *et al*.: Belief combination and propagation in a lattice-structured inference network, *IEEE SMC* 20 (1), 45/58 (1990)

ISBN: 978-989-95079-6-8

Consistency of pair-wise comparison matrix with fuzzy elements

Jaroslav Ramík

School of Business Administration in Karviná, Silesian University in Opava
733 40 Karviná, Czech Republic

Email: ramik@opf.slu.cz

Abstract—*I*n this paper, consistency of pair-*w*ise comparison matri*x* *w*ith fuzzy elements based on the method of *g*eometric mean is investigated. A consistency inde*x* of reciprocal matri*x* *w*ith fuzzy elements is introduced based on a ne*w*ly designed method of lo*g*arithmic least s*q*uares for eliciting associated *w*eights. Some basic properties of the inde*x* are presented and t*w*o simple illustrating examples are supplied.

Key*w*ords—decision making, uncertainty, pair-*w*ise comparison, reciprocal matri*x*, consistency, triangular fuzzy numbers.

1 Introduction

The classical pair-*w*ise comparison method re*q*uires the decision-maker (DM) to e*x*press his/her preferences in the form of a precise ratio matri*x* encodi*n*g a valued preference relation. Ho*w*ever, it can often be difficult for the DM to e*x*press e*x*act estimates of the ratios of importance. Therefore many kinds of methods employing intervals or fuzzy numbers as elements of a pair-*w*ise comparison matri*x* have been proposed to cope *w*ith this problem. This allo*w*s for a more fle*x*ible specification of pair-*w*ise preference intensities accounti*n*g for the incomplete kno*w*ledge of the DM, see [4].

*I*n practice, *w*hen interval-valued matrices are employed, the DM often *g*ives ranges narro*w*er than his/her actual perception *w*ould authorize, because he/she mi*g*ht be afraid of e*x*pressi*n*g information *w*hich is too imprecise. On the other hand, a fuzzy number or fuzzy interval e*x*presses rich information because the DM can provide the core of the fuzzy interval as a rough estimate of his perceived preference and also the support set of the fuzzy interval as the ra*n*ge that the DM believes to surely contain the unkno*w*n ratio of relative importance.

*I*n this paper, consistency of pair-*w*ise comparison matri*x* *w*ith fuzzy numbers based on the method of *g*eometric mean is investigated. A consistency inde*x* of reciprocal matri*x* *w*ith triangular fuzzy elements is introduced based on a ne*w*ly desi*g*ned method of lo*g*arithmic least s*q*uares for eliciting associated *w*eights *w*ith the minimal measure of fuzziness. Here, *w*e accept the reasons discussed in [3] (and in other papers cited there) for using *g*eometric mean instead of Saaty's procedures based on the principal eigenvector of the pair-*w*ise comparison matri*x*. Our approach is ho*w*ever different to the method used by Van Laarhoven and Pedrycz, see [7]. Defini*n*g the consistency inde*x*, *w*e use the different approach to Buckley's at al. [2], the consistency inde*x* proposed there is based on the Saaty's definition. Moreover, the calculation procedure of the inde*x* is rather comple*x*.

2 Consistency of pair-wise comparison matrix

Consider an n×n pair-*w*ise comparison matri*x* **A** such that

$$\mathbf{A} = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a_{1n}} & \frac{1}{a_{2n}} & \cdots & 1 \end{bmatrix}. \qquad (1)$$

This matri*x* is reciprocal, if $a_{ij} = \frac{1}{a_{ji}}$ for each $1 \le i, j \le n$.

We say that **A** is consistent if

$$a_{ij}.a_{jk} = a_{ik}, \qquad (2)$$

for each $1 \le i, j, k \le n$.

*I*f for some integer i, *j*, k, (2) does not hold, than **A** is said to be inconsistent. *I*n the popular Analytic hierarchy process (AHP), it is assumed that $\frac{1}{9} \le a_{ij} \le 9$, for all $1 \le i, j \le n$, see [6]. The consistency of **A** is measured by the consistency inde*x* C*I* as

$$CI = \frac{\lambda_{max} - n}{n-1}, \qquad (3)$$

*w*here λ_{max} is the ma*x*imal eigenvalue of **A**. *I*t can be sho*w*n that C*I* ≥ 0, see [6].

*I*f **A** is an n×n reciprocal pair-*w*ise comparison matri*x*, then **A** is consistent iff C*I* = 0. To provide a measure independent of the order of the matri*x*, n, Saaty proposed the consistency ratio (CR). This is obtained by taki*n*g the ratio bet*w*een λ_{max} - n to its e*x*pected value over a large number of positive reciprocal matrices of order n, *w*hose entries are randomly chosen in the set of values $\left\{\frac{1}{9},...,9\right\}$. For this consistency measure, he proposed a 10% threshold for the CR to accept the estimation. *I*n practical decision situations inconsistency is "acceptable" if CR < 0,1. For the priorization procedure based on *g*eometric mean, the *g*eometric consistency ratio *G*CR *w*as proposed in [1], *w*ith an interpretation analogous to that considered for Saaty´s CR.

ISBN: 978-989-95079-6-8

3 Pair-wise comparison matrix with triangular fuzzy elements

Since using fuzzy numbers as elements of a pair-wise matrix is more expressive than using crisp values or intervals, we hope that the fuzzy approach allows a more accurate description of the decision making process. Rather than forcing the DM to provide precise representations of imprecise perceptions, we suggest using an imprecise representation instead.

We define a new consistency measure for an n×n reciprocal pair-wise comparison matrix with triangular fuzzy numbers being the elements of the matrix. The membership function of a triangular fuzzy number is piece-wise linear. There exist at least three reasons for using triangular shape of fuzzy elements:

1. The membership function of triangular fuzzy elements is piece-wise linear, i.e. relatively simple.
2. Triangular fuzzy numbers can be easily manipulated, e.g. added, multiplied etc.
3. Crisp (non-fuzzy) numbers, i.e. the most practical values, can be represented as triangular ones.

A triangular fuzzy number a can be equivalently expressed by a tripple of real numbers, i.e. $a = (a^L; a^M; a^U)$, where a^L is the Lower number, a^M is the Middle number, and a^U is the Upper number, $a^L \le a^M \le a^U$. If $a^L = a^M = a^U$, then a is said to be the crisp number (non-fuzzy number). Evidently, the set of all crisp numbers is isomorphic to the set of real numbers. In order to distinguish fuzzy and non-fuzzy numbers we shall denote fuzzy numbers, vectors and matrices by the tilde above the symbol, e.g. $\tilde{a} = (a^L; a^M; a^U)$. It is well known that the arithmetic operations +,-, * and / can be extended to fuzzy numbers by the Extension principle.

If all elements of an m×n matrix $\mathbf{A}$ are triangular fuzzy numbers we call $\mathbf{A}$ the matrix with triangular fuzzy elements and this matrix is composed of triples as follows

$$\tilde{\mathbf{A}} = \begin{bmatrix} (a_{11}^L; a_{11}^M; a_{11}^U) & \cdots & (a_{1n}^L; a_{1n}^M; a_{1n}^U) \\ \vdots & \ddots & \vdots \\ (a_{m1}^L; a_{m1}^M; a_{m1}^U) & \cdots & (a_{mn}^L; a_{mn}^M; a_{mn}^U) \end{bmatrix}.$$

Particularly, let $\tilde{\mathbf{A}}$ be an n×n matrix with triangular fuzzy elements. We say that $\tilde{\mathbf{A}}$ is reciprocal, if the following condition is satisfied: $\tilde{a}_{ij} = (a_{ij}^L; a_{ij}^M; a_{ij}^U)$ implies $\tilde{a}_{ji} = (\frac{1}{a_{ij}^U}; \frac{1}{a_{ij}^M}; \frac{1}{a_{ij}^L})$ for all i,j = 1,2,...,n, i.e.

$$\tilde{\mathbf{A}} = \begin{bmatrix} (1;1;1) & (a_{12}^L; a_{12}^M; a_{12}^U) & \cdots & (a_{1n}^L; a_{1n}^M; a_{1n}^U) \\ (\frac{1}{a_{12}^U}; \frac{1}{a_{12}^M}; \frac{1}{a_{12}^L}) & (1;1;1) & \cdots & (a_{2n}^L; a_{2n}^M; a_{2n}^U) \\ \vdots & \vdots & \ddots & \vdots \\ (\frac{1}{a_{1n}^U}; \frac{1}{a_{1n}^M}; \frac{1}{a_{1n}^L}) & (\frac{1}{a_{2n}^U}; \frac{1}{a_{2n}^M}; \frac{1}{a_{2n}^L}) & \cdots & (1;1;1) \end{bmatrix}, \quad (4)$$

where $1 \le a_{ij}^L \le a_{ij}^M \le a_{ij}^M$, i,j=1,2,...,n. Without loss of generality we assume that $1 \le a_{ij}^M \le a_{ik}^M$ whenever $i \le j \le k$. In some sense, our new inconsistency measure will become an extension of the consistency index CR. For this purpose let us consider n elements (e.g. some criteria or variants) being evaluated by the DM. The result of the pair-wise comparison evaluations of the DM is a reciprocal matrix with triangular fuzzy elements $\tilde{\mathbf{A}} = \{\tilde{a}_{ij}\}$, i.e. (4), where

$$\tilde{a}_{ji} = (\frac{1}{a_{ij}^U}; \frac{1}{a_{ij}^M}; \frac{1}{a_{ij}^L}) \quad \text{for all } i,j = 1,2,...,n.$$

Here, $\frac{1}{\sigma} \le a_{ij}^L \le a_{ij}^M \le a_{ij}^U \le \sigma$ and $S = [\frac{1}{\sigma}, \sigma]$, $\sigma > 1$, is an interval of real numbers called the scale. DMs use the scale for evaluating pairs of elements. In classical AHP, $\sigma = 9$, see [6].

Now, consider deviations

$$\frac{\tilde{w}_i}{\tilde{w}_j} \simeq \tilde{a}_{ij}, \quad (5)$$

where $\tilde{w}_k = (w_k^L; w_k^M; w_k^U)$ are triangular fuzzy weights associated with evaluated elements and extended arithmetic operations on fuzzy numbers, see [2]. The question is how these weights should be calculated so that the deviations are minimized in some (fuzzy) sense.

Here, we propose a new method for calculating w_k^L, w_k^M, w_k^U by solving the following optimization problem:

$$\sum_{i,j} \left(\log \frac{w_i^L}{w_j^L} - \log a_{ij}^L \right)^2 + \left(\log \frac{w_i^M}{w_j^M} - \log a_{ij}^M \right)^2 + \left(\log \frac{w_i^U}{w_j^U} - \log a_{ij}^U \right)^2 \longrightarrow \min, \quad (6)$$

subject to

$$w_k^U \ge w_k^M \ge w_k^L \ge 0, \; k = 1, 2,...,n. \quad (7)$$

By the well known method of making the derivatives zero (a necessary condition for optimality) it can be easily shown, that the solution of problem (3), (4) shall satisfy the following relations

$$w_k^L = C_L \cdot \left(\prod_{j=1}^n a_{kj}^L \right)^{1/n}, \; w_k^M = C_M \cdot \left(\prod_{j=1}^n a_{kj}^M \right)^{1/n}, \; w_k^U = C_U \cdot \left(\prod_{j=1}^n a_{kj}^U \right)^{1/n} \quad (8)$$

for each k = 1, 2,...,n, where coefficients C_L, C_M, C_U are arbitrary positive constants. We shall find the values of the coefficients with the "minimal measure of fuzziness" of the associated weights.

ISBN: 978-989-95079-6-8

Moreover, we require that the middle values w_k^M of the fuzzy weights $\tilde{w}_k = (w_k^L; w_k^M; w_k^U)$ satisfy the normalization condition, i.e. $\sum_{k=1}^{n} w_k^M = 1$. Hence, from normalization condition and from (7) we obtain

$$C_M = \frac{1}{\sum_{i=1}^{n} (\prod_j a_{ij}^M)^{1/n}}. \quad (9)$$

Hence, from (6), (7) we obtain

$$C_L \le C_M \min_{i=1,\dots,n} \left\{ \frac{\left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}{\left(\prod_{j=1}^{n} a_{ij}^L\right)^{1/n}} \right\},$$

$$C_U \ge C_M \max_{i=1,\dots,n} \left\{ \frac{\left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}{\left(\prod_{j=1}^{n} a_{ij}^U\right)^{1/n}} \right\}. \quad (10)$$

Moreover, we want to find weights $\tilde{w}_k = (w_k^L; w_k^M; w_k^U)$, k = 1, 2,...,n, with the minimal spread, i.e. the minimal measure of fuzziness $s_k = w_k^U - w_k^L$. Hence, by (8), (9) we obtain the weights as follows:

$$w_k^L = C_{min} \cdot \frac{\left(\prod_{j=1}^{n} a_{kj}^L\right)^{1/n}}{\sum_{i=1}^{n} \left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}, C_{min} = \min_{i=1,\dots,n} \left\{ \frac{\left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}{\left(\prod_{j=1}^{n} a_{ij}^L\right)^{1/n}} \right\}, \quad (11)$$

$$w_k^M = \frac{\left(\prod_{j=1}^{n} a_{kj}^M\right)^{1/n}}{\sum_{i=1}^{n} \left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}, \quad (12)$$

$$w_k^U = C_{max} \cdot \frac{\left(\prod_{j=1}^{n} a_{kj}^U\right)^{1/n}}{\sum_{i=1}^{n} \left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}, C_{max} = \max_{i=1,\dots,n} \left\{ \frac{\left(\prod_{j=1}^{n} a_{ij}^M\right)^{1/n}}{\left(\prod_{j=1}^{n} a_{ij}^U\right)^{1/n}} \right\}. \quad (13)$$

Particularly, if $\tilde{\mathbf{A}}$ is a crisp (i.e. nonfuzzy) matrix, i.e. $a_{ij}^L = a_{ij}^M = a_{ij}^U$ for all i, j, then $C_{min} = C_{max} = 1$, therefore, $w_k^L = w_k^M = w_k^U$ for all k. Consequently, the solution - weights are crisp, too.

4 New consistency index

Supposing the maximum of all deviations (5) for all i,j is equal to "zero", then intuitively $\tilde{\mathbf{A}}$ should be consistent. Hence, a consistency measure shall be introduced as the maximal deviation relative to the measure which could be achieved by a matrix of the same dimension.

For a given scale $S = [\frac{1}{\sigma}, \sigma]$ we define a consistency index $I_n^\sigma(\tilde{\mathbf{A}})$ of n×n reciprocal matrix $\tilde{\mathbf{A}}$ with fuzzy triangular elements as follows

$$I_n^\sigma(\tilde{\mathbf{A}}) = C_n^\sigma . \max_{i,j} \left\{ \max\{ \left| \frac{w_i^L}{w_j^U} - a_{ij}^L \right|, \left| \frac{w_i^M}{w_j^M} - a_{ij}^M \right|, \left| \frac{w_i^U}{w_j^L} - a_{ij}^U \right| \} \right\} \quad (14)$$

where w_k^L, w_k^M, w_k^U are given by (11) – (13) for all k= 1,2,…,n, and

$$C_n^\sigma = \frac{1}{\max\left\{ \sigma - \sigma^{\frac{2-2n}{n}}, \sigma^2 \left(\left(\frac{2}{n}\right)^{\frac{2}{n-2}} - \left(\frac{2}{n}\right)^{\frac{n}{n-2}} \right) \right\}}, \text{ if } \sigma < \left(\frac{n}{2}\right)^{\frac{n}{n-2}},$$

$$C_n^\sigma = \frac{1}{\max\left\{ \sigma - \sigma^{\frac{2-2n}{n}}, \sigma^{\frac{2n-2}{n}} - \sigma \right\}}, \text{ if } \sigma \ge \left(\frac{n}{2}\right)^{\frac{n}{n-2}}. \quad (15)$$

We say that $\tilde{\mathbf{A}}$ is F-consistent if $I_n^\sigma(\tilde{\mathbf{A}}) = 0$, otherwise, if $I_n^\sigma(\tilde{\mathbf{A}}) > 0$, then $\tilde{\mathbf{A}}$ is F-inconsistent.

The following theorem shows, that C_n^σ in (15) is a "normalizing" constant. The proof is easy, see [5].

Theorem 1. *If* $\tilde{\mathbf{A}}$ is an n×n reciprocal matrix with triangular fuzzy elements evaluated from the scale $[\frac{1}{\sigma}, \sigma]$, then

$$0 \le I_n^\sigma(\tilde{\mathbf{A}}) \le 1. \quad (16)$$

The following theorem states that the new concept of F-consistency is compatible with the old concept of consistency (2).

Theorem 2. Let $\tilde{\mathbf{A}} = \mathbf{A}$ be a non-fuzzy reciprocal pair-wise comparison matrix. *If* **A** is consistent then **A** is F-consistent.

5 Examples

In this section we present two examples showing that the newly defined consistency index is a convenient tool not only for measuring inconsistency of pair-wise comparison matrices with fuzzy elements, but also for measuring inconsistency of crisp pair-wise comparison matrices.

The first example gives the (crisp) pair-wise comparison matrix $\mathbf{A}^*$ with inconsistency index $I_n^\sigma(\mathbf{A}^*) = 1$.

Example 1. Consider 7×7 reciprocal matrix with crisp elements from the scale $[\frac{1}{9}, 9]$:

ISBN: 978-989-95079-6-8

$$\mathbf{A}^{*}=\begin{bmatrix} 1 & 9 & 9 & 9 & 9 & 9 & 9 \\ \frac{1}{9} & 1 & 9 & 9 & 9 & 9 & 9 \\ \frac{1}{9} & \frac{1}{9} & 1 & 9 & 9 & 9 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 & 9 & 9 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 & 9 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 & 9 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & 1 \end{bmatrix}.$$

Here, as $9>\left(\frac{7}{2}\right)^{\frac{7}{7-2}}=5.78$, we obtain

$$\frac{w_1}{w_n}-a_{1n}=\frac{\left(\sigma^{n-1}\right)^{\frac{1}{n}}}{\left(\frac{1}{\sigma^{n-1}}\right)^{\frac{1}{n}}}-\sigma=\sigma^{\frac{2n-2}{n}}-\sigma=9^{\frac{12}{7}}-9=34.24.$$

$$a_{n1}-\frac{w_n}{w_1}=\sigma-\frac{\left(\frac{1}{\sigma^{n-1}}\right)^{\frac{1}{n}}}{\left(\sigma^{n-1}\right)^{\frac{1}{n}}}-=\sigma-\sigma^{\frac{2-2n}{n}}-=9-9^{-\frac{12}{7}}=8.98.$$

Hence,

$$\max_{i,j}\left\{\left|\frac{w_i}{w_j}-a_{ij}\right|\right\}=\max\{34.24,8.98\}=34.24.$$

By (14) and (15) we get $I_7^9(\mathbf{A}^{*})=1$.

Example 2. Consider 3×3 reciprocal matrix $\widetilde{\mathbf{A}}$ with triangular fuzzy elements

$$\widetilde{\mathbf{A}}=\begin{bmatrix} (1 & 1 & 1) & (2 & 3 & 4) & (4 & 5 & 6) \\ (\frac{1}{4} & \frac{1}{3} & \frac{1}{3}) & (1 & 1 & 1) & (3 & 4 & 5) \\ (\frac{1}{6} & \frac{1}{5} & \frac{1}{4}) & (\frac{1}{5} & \frac{1}{4} & \frac{1}{3}) & (1 & 1 & 1) \end{bmatrix},$$

S = [1/9, 9] is an evaluation scale.

As $9>\left(\frac{3}{2}\right)^{\frac{3}{3-2}}=3.375$, we calculate

$$C_3^9=\frac{1}{\max\left\{9-9^{-\frac{4}{3}},9^{\frac{4}{3}}-9\right\}}=\frac{1}{\max\{8.95,9.72\}}=0.103.$$

By (11) – (13), we obtain $C_{max}=0.855$, $C_{min}=1.145$ and

$$\widetilde{w}_1=(w_1^{L};w_1^{M};w_1^{U})=(0.458\,;\,0.627\,;\,0.627),$$
$$\widetilde{w}_2=(w_2^{L};w_2^{M};w_2^{U})=(0.264\,;\,0.280\,;\,0.295),$$
$$\widetilde{w}_3=(w_3^{L};w_3^{M};w_3^{U})=(0.094\,;\,0.094\,;\,0.094).$$

Finally, we calculate the inconsistency index $I_3^9(\widetilde{\mathbf{A}})=0.219$, hence $\widetilde{\mathbf{A}}$ is F-inconsistent.

For comparison with the well known AHP method, consider the corresponding crisp matrix $\mathbf{A}^{M}$ taking into account only the middle values of $\widetilde{\mathbf{A}}$, i.e.

$$\mathbf{A}^{M}=\begin{bmatrix} 1 & 3 & 5 \\ \frac{1}{3} & 1 & 4 \\ \frac{1}{5} & \frac{1}{4} & 1 \end{bmatrix}.$$

Consistency index $I_3^9(\mathbf{A}^{M})=0.174$, hence $\mathbf{A}^{M}$ is also F-inconsistent, the consistency index is, however, smaller than the previous one, as there is no fuzziness in $\mathbf{A}^{M}$. Applying consistency index by T. Saaty to matrix $\mathbf{A}^{M}$, we obtain CR = 0.082.

6 Conclusions

In this paper, we investigated consistency of pair-wise comparison matrix with triangular fuzzy elements based on a newly designed method of logarithmic least squares for eliciting associated weights. A new consistency index of reciprocal matrix with fuzzy triangular elements was defined. The weights we derived have the minimal measure of fuzziness among all weights based on the solution of a specific optimization problem. Some properties of the consistency index were derived and two illustrating examples were presented. The newly defined inconsistency index seems to be a convenient tool not only for measuring inconsistency of pair-wise comparison matrices with fuzzy elements, but also for measuring inconsistency of crisp pair-wise comparison matrices.

Acknowledgment

Supported by GACR project No. 402090405

References

[1] Aguarón, J., Moreno-Jimenéz, J. M., The geometric consistency index: Approximated thresholds. European Journal of Operational Research 147, 2003, pp. 137–145, ISSN 0377-2217.

[2] Buckley, J. J., Feuring, T., Hayashi, Y., Fuzzy hierarchical analysis revisited. Fuzzy Sets and Systems 129, 2001, pp. 48-64, ISSN 0165-0114.

[3] Lootsma, F. A., A model for the relative importance of the criteria in the multiplicative AHP and SMART. European Journal of Operational Research 94, 1996, pp. 467-476, ISSN 0377-2217.

[4] Ohnishi, S. et al., A study on the Analytic Hierarchy Process using a fuzzy reciprocal matrix. Proceeding of the conference IPMU 2006, irit.fr, http://www.irit.fr/~Didier.Dubois/Papers0804/ODPY_IPMU06.pdf.

[5] Ramík, J., Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean. To be published in Fuzzy Sets and Systems.

[6] Saaty, T. L., Multicriteria decision making - the Analytical Hierarchy Process. Vol. I., RWS Publications, Pittsburgh, 1991.

[7] Van Laarhoven, P. J. M. and Pedrycz, W., A fuzzy extension of Saaty's priority theory. Fuzzy Sets and Systems 11, 1983, 4, pp. 229-241, ISSN 0165-0114.

ISBN: 978-989-95079-6-8

Fuzzy Continuous Evaluation in Training Systems Based on Virtual Reality

Ronei M. Moraes[1] and Liliane S. Machado[2]

1 Department of Statistics, Federal University of Paraíba
João Pessoa/PB, Brazil
2 Department of Informatics, Federal University of Paraíba
João Pessoa/PB, Brazil
Email: ronei@de.ufpb.br, liliane@di.ufpb.br

Abstract*—The approach of continuous evaluation is an important methodology in educational learning process. However, only recently it was applied in training based on virtual reality. This paper presents a methodology of evaluation that uses a fuzzy continuous evaluation approach to provide a user profile from his several training. This information can be used to improve the user performance in the real execution of a task. The methodology proposed is given by the union of fuzzy statistical measures, fuzzy statistics and fuzzy parameters (fuzzy testing of hypothesis and fuzzy regression models) as input for a fuzzy rule based expert system (FRBES). The FRBES is able to construct an individual profile for each trainee. This new approach is a diagnostic tool that enables a trainee to understand the areas in which he presents difficulties, allowing him to concentrate on improving skills related to them.*

***Keywords*—** Continuous evaluation, fuzzy rule based expert system, fuzzy statistics, fuzzy measures, fuzzy regression models.

1 Introduction

The researches in training evaluation based on virtual reality (VR) [4] are recent. The firsts evaluation systems were offline [3], in which a training based on VR was recorded in videotapes for post-analysis by experts. Recently, online methods were proposed by [13], in which the evaluation is performed during the training process and the user receives that evaluation immediately after the end of training. Since then, several papers were produced in that subject [9, 12, 15, 16, 18, 19, 20, 22, 24, 25, 26]. However, all those methodologies did not use any technique of continuous evaluation to improve trainee performance. Continuous evaluation is a educational methodology used in present and distance learning to help the construction of the knowledge and the cognitive training [1, 10].

In the present work, the goal is to construct a diagnostic to help trainees to understand their difficulties. The first methodology of this kind was proposed only in 2005 by [23] where the goal was to construct a profile to help trainees to understand their difficulties and to improve their performance. That methodology, based on a classical based rule expert system, was able to provide an Evaluation Report and a Continuous Evaluation Report, showing the performance of trainee in the last training and in all trainings performed by him/her, respectively. Morris et al. (2006) suggested the use of statistical linear regression to evaluate user's progress in a bone surgery. After that, in 2009 [17] had proposed another approach based on fuzzy based rule expert system (FRBES), using statistical measures and statistical models, statistical testing of hypothesis, as well as fuzzy measures as input for a FRBES. However, some problems had not been solved yet. The problems are related to assumptions of the statistical models and statistical testing of hypothesis, mainly the Gaussian distributions of data.

In this paper, we propose a new conception of fuzzy continuous evaluation to construct a trainee profile from his/her several trainings and to help him/her to improve his/her performance [2, 6]. In this approach we use fuzzy statistical measures, fuzzy statistics and fuzzy parameters (fuzzy testing of hypothesis and fuzzy regression models) to solve that problems with statistical assumptions. Those fuzzy parameters compose inputs for an fuzzy based rule expert system (FRBES) [30]. The FRBES combines logically all information about fuzzy statistical variables and parameters to making decisions about complex conjectures [7] and is able to construct a trainee profile.

2 Theoretical Aspects

For the reader's better understanding, we first present a short review about fuzzy sets, fuzzy statistics and fuzzy rule based expert system.

2.1 Fuzzy Sets and Fuzzy Numbers

As some variables in the training system do not present an exactly correspondence to the real world, some measures cannot be exact. Then, fuzzy sets are used to measure those variables [7]. In classical set theory a set A of a universe X can be expressed by means of a membership function $\mu_A(x)$, with $\mu_A: X \rightarrow \{0,1\}$, where for a given $a \in A$, $\mu_A(a)=1$ and $\mu_A(a)=0$ respectively express the presence and absence of a in A. Mathematically:

$$\mu_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases} \quad (1)$$

Zadeh [29] introduced the fuzzy set theory in 1965. A fuzzy set or fuzzy subset is used to model an ill-known quantity. A fuzzy set A on X is characterized by its membership function $\mu_A: X \rightarrow [0,1]$. We say that a fuzzy set A of X is "precise" when $\exists\ c^* \in X$ such that $\mu_A(c^*)=1$ and $\forall c \neq c^*$, $\mu_A(c)=0$. A fuzzy set A will be said to be "crisp", when $\forall\, c \in X$, $\mu_A(c) \in \{0,1\}$.

The intersection and union of two fuzzy sets are performed trough the use of *t-norm* and *t-conorm* operators respectively, which are commutative, associative and

ISBN: 978-989-95079-6-8

monotonic mappings from [0,1]→[0,1]. Moreover, a t-norm Γ (respec. t-conorm ⊥) has 1 (respec. 0) as neutral element (e. g.: Γ=*min*, ⊥=*max*) [8]. Thus, we can define intersection and union of two fuzzy sets as:
The intersection of two fuzzy sets *A* and *B*, with membership functions $\mu_A(x)$ e $\mu_B(x)$ is a fuzzy set *C* with membership function given by:

$$C=A \cap B \Leftrightarrow \mu_C(x)= \Gamma\{\mu_A(x),\mu_B(x)\}, \ \forall x\in X. \quad (2)$$

The union of two fuzzy sets *A* and *B*, with membership functions $\mu_A(x)$ e $\mu_B(x)$ is a fuzzy set *C* with membership function given by:

$$C=A \cup B \Leftrightarrow \mu_C(x)= \perp\{\mu_A(x),\mu_B(x)\}, \ \forall x\in X. \quad (3)$$

The complement of a fuzzy set *A* in *X*, denoted by $\neg A$ is defined by:

$$\mu_{\neg A}(x) = n(\mu_A(x)), \ \forall x\in X. \quad (4)$$

where: *n:* [0,1]→[0,1] is a negation operator which satisfies the following properties:

- $n(0)=1$ and $n(1)=0$
- $n(a) \leq n(b)$ if $a>b$
- $n(n(a))=a$, $\forall x\in$ [0,1]

and a negation is a strict negation if it is continuous and satisfies

- $n(a)<n(b)$ if $a>b$

The main negation operator which satisfies these four conditions is $n(a) = 1-a$.
The implication function between two fuzzy sets *A* and *B*, with membership functions $\mu_A(x)$ e $\mu_B(x)$ is a fuzzy set *C* with membership function given by:

$$C=A \Rightarrow B \Leftrightarrow \mu_C(x,y)= \nabla\{\mu_A(x),\mu_B(y)\}, \ \forall x\in X, \ \forall y\in Y \quad (5)$$

where ∇: $[0,1]^2\rightarrow[0,1]$ is an implication operator which obeys the following properties: $\forall a, a', b, b' \in [0,1]$:

- If $b \leq b'$ then $\nabla(a,b) \leq \nabla(a,b')$;
- $\nabla(0,b)=1$;
- $\nabla(1,b)=b$.

The pure implications obeys too:

- If $a \leq a'$ then $\nabla(a,b) \geq \nabla(a',b)$;
- $\nabla(a, \nabla(b,c))= \nabla(b, \nabla(a,c))$.

Beyond concept of fuzzy sets and their operations, it is important to the concept of fuzzy numbers [3]. In this paper we used triangular or triangular shaped fuzzy numbers, which are defined by three real numbers $d<e<f$ where the interval $[d,f]$ is the base of the triangle and *e* is a vertex.

2.2 Fuzzy Statistics

In this paper the statistical methods used were:

1. fuzzy statistical measures;
2. fuzzy statistical models (time dependent or not) and
3. fuzzy statistical testing of hypotheses.

A set of fuzzy statistical measures, commonly used for general purposes as fuzzy mean, fuzzy median, fuzzy standard deviation, etc. [3], can be used to describe user interactions during the training. Besides, fuzzy statistical models based on fuzzy regression can be used to construct fuzzy models for the way followed by user in task execution [10]. This model can be done by:

$$Y = Ax_1 + Bx_2 + Cx_3 \quad (6)$$

where *Y, A, B* and *C* are fuzzy numbers and x_1, x_2 and x_3 are real numbers.
Fuzzy statistical measures and fuzzy statistical parameters of fuzzy models (as *A, B* and *C* above) can be compared with fuzzy parameters of ideal fuzzy models using appropriate fuzzy statistical testing of hypothesis from its α-cuts [3]. As results of these comparisons, we can make fuzzy statistical decisions about equality or difference between parameters using measure of fuzzy probability of significance. The information synthesized by fuzzy statistical measures and fuzzy parameters helps to construct a profile for user, his/her Evaluation Report and his/her Continuous Evaluation Report.

2.3 Fuzzy Rule Based Expert System

Expert systems [27] use the knowledge of an expert in a given specific domain to answer non-trivial questions about that domain. For example, an expert system for image classification would use knowledge about the characteristics of the classes present in a given region to classify a pixel in an image of that region. This knowledge also includes the "how to do" methods used by the human expert. Usually, the knowledge in an expert system is represented by rules as:

```
IF <condition> THEN <conclusion>
```

Most rule-based expert systems allows the use of connectives AND or OR in the premise of a rule, and of connective AND in the conclusion. From rules and facts, new facts will be obtained through an inference process.
In several cases, we do not have precise information about conditions or conclusions. Then, the knowledge in the rules cannot be expressed in a precise manner. Thus, it could be interesting to use a fuzzy rule-based expert system [30]. An example of simple fuzzy rule could be:

```
IF <access to the help is persistent>
   AND <Global Users Performance is bad>
THEN <user is Novice>.
```

where "persistent" and "bad" can be characterized by fuzzy sets.

ISBN: 978-989-95079-6-8

3 Methodology

According to [23], a tool for continuous evaluation must be interconnected with an online evaluation system and must receive information from it about all variables of interest. The evaluation system works near a virtual reality simulator. In general, an online evaluation system should be capable to monitor user interactions while he/she operates the simulation system. In order to achieve that, it is necessary to collect the information about positions in the space, forces, torque, resistance, speeds, accelerations, temperatures, visualization and/or visualization angle, sounds, smells and etc. This information will be used to feed the evaluation system. In the Figure 1 [21], we can observe that the virtual reality simulator and the system of evaluation are independent systems, however they act simultaneously.

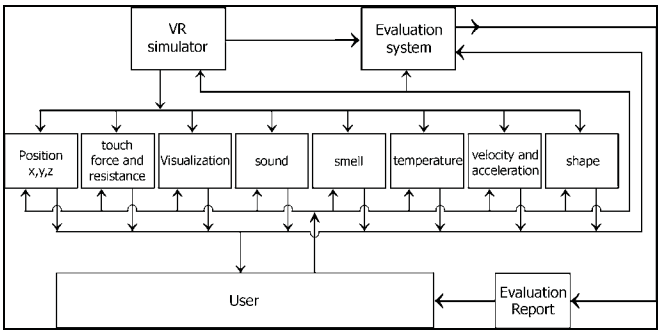


Figure 1: Diagram of A Virtual Reality Simulator With an Evaluation System.

User interactions are monitored and the information are sent to the evaluation system that analyzes the data and emits, at the end of the training, an Evaluation Report about the user performance, according *M* pre-defined classes of performance. A set of rules of the fuzzy based rule expert system (FRBES) [30] defines each one of the possible classes of performance, which are defined from specialists' knowledge. The interaction variables will be monitored according to their relevance to the training. Thus, each application will have their own set of relevant variables that will be monitored [21].

If the same user had performed other trainings, the Continuous Evaluation Tool uses the data collected from user interaction in his/her several training to create an User Profile and construct a Continuous Evaluation Report about all set of training. That information is used to evaluate the trainee and can improve his/her performance in real tasks [28]. The Figure 2 [23] shows a diagram of an Evaluation System able to perform continuous evaluation.

A Fuzzy Continuous Evaluation Tool makes a union of fuzzy statistical measures, models and testing of hypothesis, and an FRBES to construct an individual profile for trainee. Fuzzy statistical tools are programmed to make an automatic analysis of the database and construct fuzzy statistical measures, graphics fuzzy statistical models and results of fuzzy statistical hypothesis testing. FRBES uses this information to create a user profile and a continuous evaluation report. The continuous evaluation report presents the trainee profile and shows, with fuzzy statistical measures, graphics and models, the execution performance of specific tasks. They are used as input for FRBES [30]. Figure 3 shows the new methodology presented.

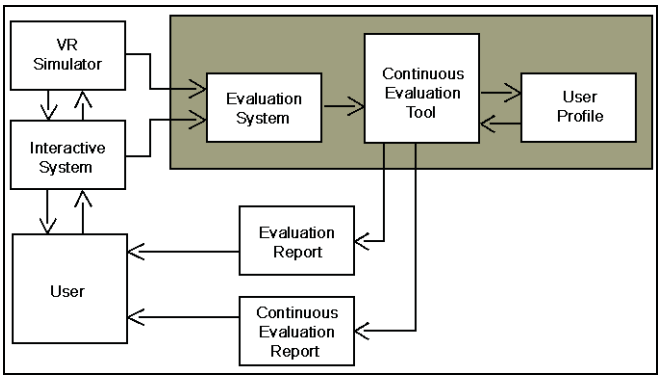


Figure 2: Diagram of an Evaluation System With Approach of Continuous Evaluation. Adapted from [23].

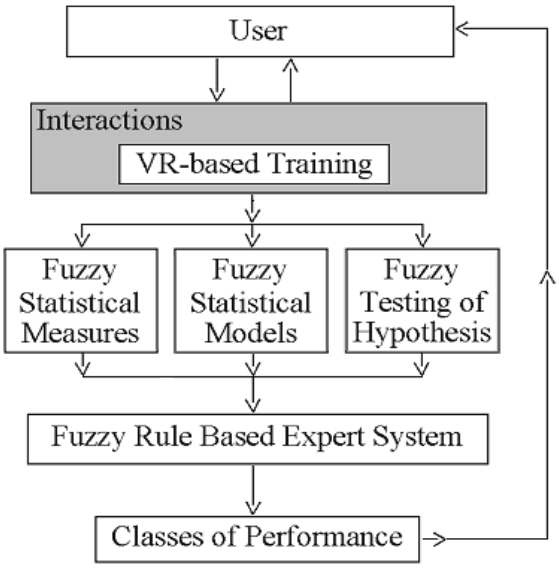


Figure 3: Diagram of The Continuous Evaluation System using FRBES.

In the first time that user executes his training, the Evaluation Report emits information about the user performance, at the end of the training, according to classes of performance previously defined. This information is stored in a User Profile for posterior evaluations with approach of continuous evaluation. In a second time when user execute his/her training, the Fuzzy Continuous Evaluation is able to construct a Continuous Evaluation Report, which presents information about user performance over specific tasks, using statistical measures, tables, graphics and models. Both reports present information from the last training. But, additionally, the Continuous Evaluation Report will show accumulated information about the sequence of trainings for this user.

ISBN: 978-989-95079-6-8

4 Application

This methodology can be applied for any activity, particularly those who offer risks to the user/trainee or to people who depends of him/her. In this context, fuzzy continuous evaluation is an interesting tool to improve knowledge construction. For example in medical area, where invasive procedures can be simulated by VR, it is necessary some kind of evaluation tool with properties of continuous evaluation. In those simulators, users perform realistic virtual procedures and can acquire dexterity and/or improve skills. If a continuous evaluation tool is attached to those applications, user interactions can be collected in real time to be used to evaluation since the tool can show to the user his/her qualities and his/her deficiencies in the execution of the medical procedure.

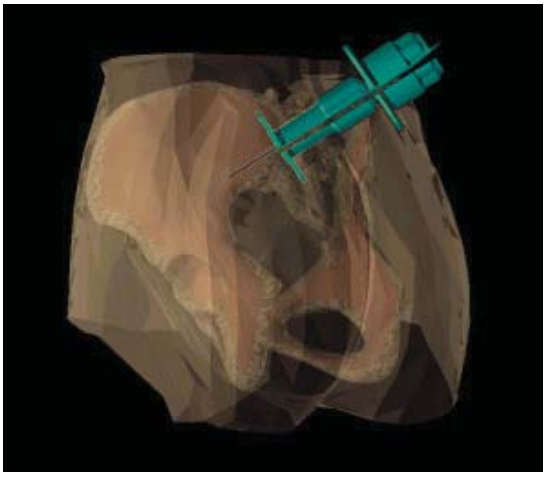


Figure 4: Screenshot of the Bone Marrow Simulator with a Semi-Transparent View of the Pelvic Region.

An example is the bone marrow simulator (Figure 4), a virtual reality simulator to training the extraction of bone marrow in children [13]. In this application the user is a novice surgeon that must acquire dexterity to insert a needle in the pelvic region of a child and find the bone marrow, located inside the iliac bone. The user manipulates a haptic device, a 6DOF interaction device, represented in the system by a needle. This device is responsible to provide tactile sensations and force feedback related to the manipulation of the needle in the system. With this device, the user can touch and feel the tissue properties of anatomy models, use his perception to identify bone location under the skin and penetrate the tissues to extract bone marrow. Figure 5 shows the haptic device used for interaction. In spite of Figure 4 presents a semi-transparent model, it is only for demonstration purposes. In the real bone marrow harvest procedure user cannot have internal visualization of the body and the touch sense is the only information perceived during needle penetration over 4 different properties tissue layers. Table 1 presents an example of description of each tissue layer present in the bone barrow harvest simulator. The hardness refers how much force must be applied to penetrate the tissue, the viscosity refers to how hard are the movements when inside the layer and surface friction indicates how slippery is the surface of the layer. The percentages are only an estimation of values and the calibration of the properties is done by a physician. The haptic device used in the simulation can read 1000 samples per second, including spatial position and applied forces. This data is an example of information used by the evaluation tool. After calibration and system setup, the same physician performs the simulation several times and in different ways to provide labels for each one of M classes of performance.

Figure 5: Haptic Device for Interaction with Tactile Sensation and Force Feedback.

Table 1: Subjective description of tissue properties for a bone marrow simulator.

Tissue	*hardness*	*viscosity*	Surface friction
Epidermis	20%	70%	60%
Subcutaneous	20%	70%	60%
Periosteum	80%	90%	20%
Hard bone	80%	90%	20%
Bone marrow	1%	5%	1%

The FRBES contains different rules to manipulate fuzzy information collected from users interaction using haptic device. The user must introduce the needle in pelvic region and go to the bone to harvest the marrow. This path can be modeled by a fuzzy regression model and the parameters of this regression can be compared with the M models previously stored by the physician. In the procedure, forces and torques applied through the needle are approximately constants between skin and bone. However, it is necessary to apply a different torque to introduce the needle into the bone. Fuzzy statistical parameters, as mean and standard deviation, can be used to compare this procedure with the parameters of procedure performed by physician. All those fuzzy statistical parameters can be compared with parameters of classes of performance using a fuzzy testing of hypothesis. The fuzzy probability of these comparisons and their correspondent hypothesis are used as input in the FRBES. If everything was performed correctly, an Evaluation Report is created to inform the user about it. In opposition, if something was performed wrong, the FRBES use their rule databases to find problems in execution and inform the user. Examples are presented in the following.

ISBN: 978-989-95079-6-8

4.1 Example 1

Let be five classes of performance done by fuzzy sets in Ω={very good, good, reasonable, bad, novice}. The fuzzy regression model for the way followed by user has parameters (for coordinates of tri-dimensional space), done by the fuzzy sets *A, B* and *C*.

Vectors of data, whose fuzzy mean and fuzzy standard deviation are given by D_{f1}, E_{f1}, G_{t1} and H_{t1}, respectively, store the force and torque applied between skin and bone. In the introduction of the needle into the bone, force and torque new vectors of data are stored, whose fuzzy mean and fuzzy standard deviation are given by D_{f2}, E_{f2}, G_{t2} and H_{t2}, respectively.

For each one class of performance, there are reference parameters for models, forces and torques. Those parameters are denoted by *A*, B** and *C** for fuzzy models and D_{fi}*, E_{fi}*, G_{ti}*, H_{ti}* , with *i=1,2* for forces and torques respectively. To compare those parameters with these reference parameters for each class, fuzzy statistical hypothesis testing are used and fuzzy probabilities for null hypothesis are obtained from them. These probabilities are treated by FRBES using rules as:

```
IF <Fuzzy Probability of Applied Forces in
    Phase 1 is Acceptable>
   AND <Fuzzy Probability of Applied Forces in
        Phase 2 is Unacceptable>
THEN <User Performance is bad>.
```

or

```
IF <Fuzzy Probability of A parameter in Fuzzy
    Model is Acceptable>
   AND <Fuzzy Probability of B parameter in
        Fuzzy Model is Acceptable>
   AND <Fuzzy Probability of C parameter in
        Fuzzy Model is Acceptable >
THEN <User Performance is very good>.
```

When the system classifies a user as good, reasonable, bad or novice, the FRBES can detect the user mistakes by the search for unacceptable fuzzy probabilities. From them, is possible to find where user made mistake. For example, it can be noticed the user made mistake in the application of forces in the Phase 2, but everything was performed correctly in the Phase 1.

```
IF <Global User Performance is bad >
   AND <Fuzzy Probability of Applied Forces in
        Phase 2 is Unacceptable>
THEN <Applied Forces in Phase 2 has
      Mistake>.
```

By the rule above, the user will know about his/her mistake. Also it is possible to know what was performed wrong. In this case, only three cases are possible:

1. Applied force was excessive and it was not possible to harvest the marrow.
2. Applied force was normal and it produced a good procedure.
3. Applied force was lower than normal, and the needle cannot be introduced into the bone.

The FRBES has rules to verify which situation really happened. For example:

```
IF <Global User Performance is bad >
   AND <User Applied Forces in Phase 2 has
        Mistake>
   AND <Applied Forces in Phase 2 is Excessive>
THEN <Applied Forces in Phase 2 is Upper
      than normal>.
```

4.2 Example 2

In the Section 4.1 it was illustrated how the Evaluation Report works. Now, we present the Continuous Evaluation Report. This report must show the evolution of user according to the sequence of training he/she performed, including the last one. The goal of this report is to help user to improve his/her performance.

Let be a training sequence for a user, denoted by $S=\{s_1,..., s_n\}$. In each s_j ($j=1,...,n$) is stored a set of vectors with: variables of training, performance by variable of training and global user performance. From this information the FRBES can produce a Continuous Evaluation Report, showing to the user his/her evolution. For example, mistakes are less frequent now than before; serious mistakes are not made there is a long time; average of the time of procedure execution is coming to ideal with small standard deviation. The rules below show how it is processed:

```
IF (<Fuzzy Correlation (serious mistakes) is
     Negative OR Approximate Zero)
    AND <Fuzzy Correlation (number of mistakes)
         is Negative OR Approximate Zero>
THEN <User's Continuous Evaluation is Good>.
```

or

```
IF (<Fuzzy Probability of Execution Time is
     Acceptable)
    AND <Fuzzy Standard Deviation (Execution
         Time) is Small>
THEN <User's Continuous Evaluation is Very
      Good>.
```

FRBES can evaluate the global users performance along the sequence of performed procedures. It is done by rules like this one:

```
IF <Fuzzy Correlation (Global User Performance)
    is Positive OR Approximate Zero>
   AND <The Last One (Global User Performance)
        is Very Good>
THEN <User's Continuous Evaluation is Very
      Good>.
```

For better user comprehension of his/her situation, some graphics are presented showing all history of evaluations. For example, a graphic illustrate the number of serious mistakes or his/her Global Users Performance along the sequence of performed procedures.

5 Conclusions

In this paper we introduced a new methodology for evaluation training using a fuzzy continuous evaluation approach. This methodology uses fuzzy statistical measures, models and results of fuzzy statistical hypothesis testing, as inputs of a FRBES. This system is able to construct an

ISBN: 978-989-95079-6-8

individual profile for trainee and emit to him/her information about his/her performance at the end of the training, according to classes of performance previously defined, as proposed in others methodologies. Moreover, this methodology can provide to user information about his performance in specific tasks in the training and show where the user made mistakes. The methodology was illustrated by examples to show its functionalities and how the Evaluation Report and Continuous Evaluation Report are made.

A system developed using the proposed methodology is a diagnostic tool, which helps a trainee to understand his/her difficulties. From information presented by a Fuzzy Continuous Evaluation system, the trainee can understand his/her difficulties and improve his performance.

Acknowledgment

This work is partially supported by Brazilian Council for Scientific and Technological Development, CNPq (Process 303444/2006-1) and Brazilian Research and Projects Financing, FINEP (Grant 01-04-1054-000).

References

[1] G.M. Aalsvoort, W.C.M. Resing and A.J.J.M. Ruijssenaars, *Learning potential assessment and cognitive training: Actual research and perspectives in theory building and methodology*. JAI/Elsevier Science, 2002.

[2] J.A. Arter and J. Mctighe, *Scoring Rubrics In The Classroom: Using Performance Criteria For Assessing And Improving Student Performance*. Corwin Press, 2000.

[3] J.J. Buckley. *Fuzzy Probability and Statistics*. Springer, 2006.

[4] G. Burdea, G. Patounakis, V. Popescu and R.E. Weiss, Virtual Reality Training for the Diagnosis of Prostate Cancer. Virtual Reality Annual International Symposium Proceedings. IEEE, pp. 190-197, 1998.

[5] G. Burdea and P. Coiffet, *Virtual Reality Technology*. 2nd ed. New Jersey, John Wiley & Sons, 2003.

[6] M.O. Carlson, G.E. Humphrey and K.S. Reinhardt, *Weaving Science Inquiry and Continuous Assessment: Using Formative Assessment to Improve Learning*. Corwin Press, 2003.

[7] D. Dubois and H. Prade, *Fuzzy sets and systems - theory and applications*. New York: Academic Press, 1980.

[8] D. Dubois and H. Prade, *Possibility theory: An approach to computerized processing of uncertainty*. Plenum Press, 1988.

[9] M. Färber et al. "Heinz Handels Training and evaluation of lumbar punctures in a VR-environment using a 6DOF haptic device", Studies in Health Tech. and Informatics, 132:112-114, 2008.

[10] Galef Institute; *Assessment: Continuous Learning*. Stenhouse Publishers, 1995.

[11] J. Kacprzyk, M. Fedrizzu (eds): Fuzzy Regression Analysis. Physica Verlag. Heidelberg, 1992.

[12] K. Kumagai, J. Yamashita, O. Morikawa, and K. Yokoyama, "A New Force-Based Objective Assessment of Technical Skills in Endoscopic Sinus Surgery". Studies in Health Technology and Informatics, 125:235-237, 2007.

[13] L.S. Machado, Virtual Reality to Modeling and Simulate Invasive Procedures for Pediatrics Oncology: A Case Study in Bone Marrow Transplant. PhD Thesis (in Portuguese). Polytechnic School - University of São Paulo, 2003.

[14] L.S. Machado, R.M. Moraes and M.K. Zuffo, A Fuzzy Rule-Based Evaluation for a Haptic and Stereo Simulator for Bone Marrow Harvest for Transplant. In: Annals of Phantom Users Group 2000. Aspen, 2000.

[15] L.S. Machado, M.C.O. Valdek and R.M. Moraes, On-Line Assessment System for a Training System with Multiple Stages Based on Virtual Reality. *Journal of Multiple-Valued Logic and Soft Computing*, 14(3-5):511-524, 2008.

[16] L.S. Machado and R.M. Moraes, Qualitative and Quantitative Assessment for a VR-Based Simulator. *Studies in Health Technology and Informatics*, 2009. In press.

[17] L.S. Machado and R.M. Moraes, Continuous Evaluation in Training Systems Based on Virtual Reality Using Fuzzy Rule Based Expert Systems. Proceedings of VI International Conference on Engineering and Computer Education (ICECE'2009). Buenos Aires, 2009. In press.

[18] P.B. McBeth et al. Quantitative Methodology of Evaluating Surgeon Performance in Laparoscopic Surgery. *Studies in Health Technology and Informatics*. 85:280-286, 2002.

[19] L. Moody et al., Objective Surgical Performance Evaluation Based on Haptic Feedback. *Studies in Health Technology and Informatics* 85:304-310, 2002.

[20] R.M. Moraes and L.S. Machado, Hidden Markov Models For Learning Evaluation In Virtual Reality Simulators. *International Journal of Computers & Applications*, 25(3):212-215, 2003.

[21] R.M. Moraes and L.S. Machado, Fuzzy Gaussian Mixture Models for Online Training Evaluation in Virtual Reality Simulators. In: Proceedings of International Conference on Fuzzy Information Processing. China, v. 2, p. 733-740, 2003.

[22] Moraes, R. M.; Machado, L. S.; Using Fuzzy Hidden Markov Models for Online Training Evaluation and Classification in Virtual Reality Simulators. *International Journal of General Systems*, 33(2-3):281-288, 2004.

[23] R.M. Moraes and L.S. Machado, Continuous Evaluation in Training Systems Based on Virtual Reality Proceedings of Global Congress on Engineering and Technology Education (GCETE'2005). Brazil, p.1048-1051, 2005.

[24] R.M. Moraes and L.S. Machado, Fuzzy Bayes Rule for On-Line Training Assessment in Virtual Reality Simulators. *Journal of Multiple-Valued Logic and Soft Computing*, 14(3-5):325-338, 2008.

[25] Morris, et al.; Visuohaptic simulation of bone surgery for training and evaluation. *IEEE Computer Graphics & Applications*, 26(6): 48-57, 2006.

[26] D.J. Pucel and L.D. Anderson, Developing Computer Simulation Performance Tests: Challenges and Criteria. Proc. 2003 IASTED International Conference on Computers and Advanced Technology in Education (CATE'2003). Greece, pp 170-174,2003.

[27] E. Rich and K. Knight, *Artificial Intelligence*. McGraw Hill, 1993.

[28] R.J. Sternberg and E. Grigorenko, *Dynamic Testing: The Nature and Measurement of Learning Potential*. Cambridge University Press, 2001.

[29] L.A. Zadeh. Fuzzy Sets. *Information Control*. (8):338-353, 1965.

[30] L.A. Zadeh. Fuzzy Logic. *Computer*, (8):83-93, 1988.

ISBN: 978-989-95079-6-8

A Comparative Analysis of Symbolic Linguistic Computational Models

R.M. Rodríguez[1] L. Martínez[1] M. Espinilla[1]
1.Computer Sciences Department, University of Jaén
Jaén, Campus Las Lagunillas s/n 23071 - Jaén. Spain.
Email: rmrodrig@ujaen.es, martin@ujaen.es, mestevez@ujaen.es

Abstract*— *There are many situations in which problems deal with vague and imprecise information. In such cases, the information could be modelled by means of numbers, however, it doesn't seem logical to model imprecise information in a precise way. Therefore, the use of linguistic modelling have been used with successful results in these problems. The use of linguistic information involves the need of carrying out processes which operate with words, so called Computing with Words (CW). In the literature exists different linguistic approaches and different computational models. We focus in this contribution on the use of the fuzzy linguistic approach (FLA) to model vague and imprecise information, but more specifically we focus on their computational models paying more attention on different symbolic computational models that have been defined to deal with linguistic information. We are going to review their main features and make a comparative analysis among them.

***Keywords*—** Computing with words (CW), fuzzy linguistic approach, linguistic 2-tuple, linguistic variable.

1 Introduction

Many problems in the real world deal with vague and imprecise information. There exist different kinds of tools to manage this type of information. The probability theory can be a powerful tool in order to treat the uncertainty and can be applied in different areas, like decision making, evaluation, planning, scheduling and so on. However, it is easy to see that many aspects of uncertainties have a non-probabilistic character since they are related to imprecision and vagueness of meanings. The use of the fuzzy linguistic approach [18] to model this kind of information provided successful results because the experts involved in such situations provide linguistic values rather than numbers. The linguistic modelling implies processes of CW, in the specialized literature can be found two classical linguistic computational models that provide linguistic operators for CW:

i) Model based on the extension principle (semantic model) [2, 4].

ii) Symbolic model [6, 8, 16].

The former provides accuracy but their results cannot be expressed by linguistic terms without an approximation process. The latter also needs an approximation process to express the results in a linguistic way, but the computational process is simpler and easier to understand by the experts involved in the problems.

Due to the previous facts, different symbolic approaches based on the fuzzy linguistic approach have been defined to improve the classical computational model. These approaches have modified the representation of the linguistic information from different points of view in order to improve the computational results. The symbolic models in which we are interested are: the 2-tuple linguistic representation model [9], the virtual linguistic model [15] and the proportional 2-tuple model [13]. These models improve the classical symbolic model by avoiding the approximation in processes of CW in order to improve the precision in the final results.

The aim of this contribution is to make a comparative analysis of the three aforementioned symbolic models to show their features and discuss their correctness regarding the fuzzy linguistic approach as their basis.

This paper is structured as follows: In Section 2, we introduce in short the fuzzy linguistic approach and its classical computational models. In Section 3, we shall review the different symbolic computational models, such as, the 2-tuple linguistic representation model, the virtual linguistic model and the proportional 2-tuple representation model. In Section 4, we shall make a comparative analysis among the different symbolic computational models, and finally we shall point out some concluding remarks.

2 Fuzzy Linguistic Approach

Many aspects of different activities in the real world cannot be assessed in a quantitative form, but rather in a qualitative one, i.e., with vague or imprecise knowledge. In that case, a better approach might be to use linguistic assessments instead of numerical values. The fuzzy linguistic approach represents qualitative aspects as linguistic values by means of linguistic variables [18]. We have to choose the appropriate linguistic descriptors for the linguistic term set and their semantics. To do so, an important aspect to be analyzed is the *granularity of uncertainty* i.e., the level of discrimination among different degrees of uncertainty. Typical values of cardinality used in the linguistic models are odd ones, such as 7 or 9, where the mid term represents an assessment of "roughly 0.5" and the rest of the terms being placed symmetrically around it [2]. Once the cardinality of the linguistic term set has been established, we must provide the linguistic terms and their semantics. There exist different possibilities to accomplish this task [1, 8, 17]. One of them consists in supplying directly the term set by considering all the terms distributed on a scale on which a total order is defined [8, 17]. For example, a set of seven terms S, could be:

$$S = \{s_0 : N, s_1 : VL, s_2 : L, s_3 : M, s_4 : H, s_5 : VH, s_6 : P\}$$

Usually, in these cases, it is required that in the linguistic term set there exists:

1. A negation operator $\text{Neg}(s_i) = s_j$ such that $j = g - i$ ($g + 1$ is the cardinality)

ISBN: 978-989-95079-6-8

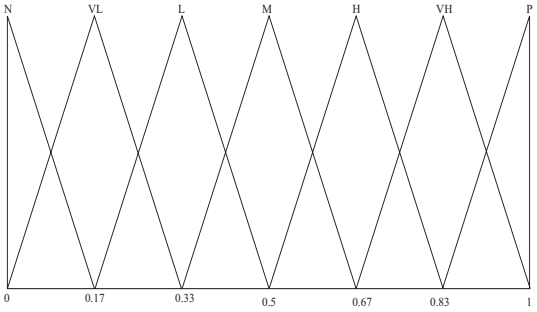


Figure 1: A Set of 7 Terms with its Semantic

2. A max operator: $\max(s_i, s_j) = s_i \; if \; s_i \geq s_j$

3. A min operator: $\min(s_i, s_j) = s_i \; if \; s_i \leq s_j$

The semantics of the terms are represented by fuzzy numbers, defined in the interval $[0, 1]$, described by membership functions. A way to characterize a fuzzy number is to use a representation based on parameters of its membership function [2]. The linguistic assessments given by the users are just approximate ones, then linear trapezoidal membership functions are good enough to capture the vagueness of those linguistic assessments [5]. This representation is achieved by the 4-tuple (a, b, c, d), where b and d indicate the interval in which the membership value is 1, with a and c indicating the left and right limits of the definition domain of the trapezoidal membership function [2]. A particular case of this kind of representation are the linguistic assessments whose membership functions are triangular, i.e., $b = d$, so we represent this kind of membership function by 3-tuples (a, b, c). An example can be:

$$\begin{array}{ll} P = (.83, 1, 1) & VH = (.67, .83, 1) \\ H = (.5, .67, .83) & M = (.33, .5, .67) \\ L = (.17, .33, .5) & VL = (0, .17, .33) \\ N = (0, 0, .17). & \end{array}$$

which is graphically shown in Figure 1.

2.1 Classical Computational Models

The use of linguistic variables implies processes of computing with words such as their fusion, aggregation, comparison, etc. To perform these computations in the fuzzy linguistic approach appeared two classical computational models:

1. **Model based on the Extension Principle (Semantic Model):** This model carries out operations with linguistic terms by means of operations associated to their membership functions based on the Extension Principle. The Extension Principle is a basic concept in the fuzzy sets theory [7] which is used to generalize crisp mathematical concepts to fuzzy sets. The use of extended arithmetic based on the Extension Principle [7] increases the vagueness of the results. Therefore, the results obtained by the fuzzy linguistic operators based on the Extension Principle are fuzzy numbers that usually do not match with any linguistic term in the initial term set. For this reason, it is necessary to carry out a linguistic approach in order to express the results in the original expression domain. In the literature, we can find different linguistic approximation operators [2, 4]. A linguistic aggregation operator based on the extension principle acts according to:

$$S^n \xrightarrow{\tilde{F}} F(\mathcal{R}) \xrightarrow{app_1(\cdot)} S \tag{1}$$

where S^n symbolizes the n Cartesian product of S. $\tilde{F}$ is an aggregation operator based on the extension principle, $F(R)$ the set of fuzzy sets over the set of real numbers R, $app_1 : F(R) \rightarrow S$ is a linguistic approximation function that returns a label from the linguistic term set S, being S the initial term set.

2. **Symbolic Model:** This model uses the ordered structure of the linguistic terms set, $S = \{s_0, s_1, \ldots, s_g\}$ where $s_i < s_j$ if $i < j$, to operate [6, 8, 16]. The intermediate results of these operations are numeric values, $\alpha \in [0, g]$, which must be approximated in each step of the process by means of an approximation function $app_2 : [0, g] \rightarrow \{0, \ldots, g\}$ that obtains a numeric value, such that, it indicates the index of the associated linguistic term, $s_{app_2(\alpha)} \in S$. Formally, it can be expressed as:

$$S^n \xrightarrow{C} [0, g] \xrightarrow{app_2(\cdot)} \{0, \ldots, g\} \longrightarrow S \tag{2}$$

where C is a symbolic linguistic aggregation operator, $app_2(\cdot)$ is an approximation function used to obtain an index $\{0, \ldots, g\}$ associated to a term in $S = \{s_0, \ldots, s_g\}$ from a value in $[0, g]$

Both models when operate with linguistic information produce loss of information due to the approximation processes and hence a lack of precision in the results. This loss of information is produced because the information representation model of the fuzzy linguistic approach is discrete in a continuous domain.

3 New Symbolic Computational Models

The necessity of dealing with linguistic information in many real world problems has driven the researchers to develop models in order to improve the processes of CW. Different models have been presented in the literature recently. In this section we focus our attention on different symbolic approaches that have developed new representation and computational models for the linguistic information in order to improve the accuracy of the results of processes of CW. Such approaches are the 2-tuple linguistic model [9], the virtual linguistic model [15] and the proportional 2-tuple linguistic model [13], which improve the limitations that present the classical symbolic computational model, regulating the loss of information and imprecision in its computations. These approaches have been widely used in problems dealing with linguistic information such as, Decision making [12, 14], Evaluation [3], Recommender Systems [11] and so on.

Due to the fact that, these symbolic approaches based on the fuzzy linguistic approach have modified the linguistic representation in order to improve the processes of CW, their review implies the study of their linguistic representation models and their computational models to accomplish the processes of CW in a symbolic and precise way.

ISBN: 978-989-95079-6-8

3.1 2-Tuple linguistic representation model

This model was presented in [9] to avoid the loss of information and to express symbolically any counting of information in the universe of the discourse.

(a) Representation model:

This representation is based on the concept of *symbolic translation* and uses it for representing the linguistic information by means of a pair of values, called *2-tuples*, (s_i, α) where s is a linguistic term and α is a numerical value representing the symbolic translation. Let $S = \{s_0, \ldots, s_g\}$ be a term set, and $\beta \in [0, g]$ a numerical value in its interval of granularity (e.g.: let β be a value obtained from a symbolic aggregation operation).

Definition 1 *The symbolic translation is a numerical value assessed in* $[-0.5, 0.5)$ *that supports the "difference of information" between a counting of information* β *assessed in the interval of granularity* $[0, g]$ *of the term set* S *and the closest value in* $\{0, \ldots, g\}$ *which indicates the index of the closest linguistic term in* S.

From this concept, is developed a linguistic representation model which represents the linguistic information by means of 2-tuples (s_i, α), $s_i \in S$ and $\alpha_i \in [-0.5, 0.5)$

- s_i represents the linguistic label of the information
- α is a numerical value expressing the value of the translation

This representation model defines a set of functions to facilitate computational processes with 2-tuples [9].

Definition 2 *Let* $S = \{s_0, \ldots, s_g\}$ *be a set of linguistic terms. The 2-tuple set associated with* S *is defined as* $\langle S \rangle = S \times [-0.5, 0.5)$. *We define the function* $\Delta : [0, g] \longrightarrow \langle S \rangle$ *given by*

$$\Delta(\beta) = (s_i, \alpha), \text{ with } \begin{cases} i = \text{round}\,(\beta), \\ \alpha = \beta - i, \end{cases} \tag{3}$$

where round *assigns to* β *the integer number* $i \in \{0, 1, \ldots, g\}$ *closest to* β.

We note that Δ is bijective [9, 10] and $\Delta^{-1} : \langle S \rangle \longrightarrow [0, g]$ is defined by $\Delta^{-1}(s_i, \alpha) = i + \alpha$. In this way, the 2-tuples of $\langle S \rangle$ will be identified with the numerical values in the interval $[0, g]$.

Remark 1 *We can consider the injective mapping* $S \longrightarrow \langle S \rangle$ *that allows us to transform a linguistic term* s_i *into a 2-tuple:* $(s_i, 0)$. *On the other hand,* $\Delta_S(i) = (s_i, 0)$ *and* $\Delta_S^{-1}(s_i, 0) = i$, *for every* $i \in \{0, 1, \ldots, g\}$.

Let's suppose a symbolic aggregation operation over labels assessed in $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$ that obtains as its result $\beta = 2.8$, then the representation of this counting of information by means of a 2-tuple will be:

$$\Delta(2.8) = (s_3, -0.2)$$

b) Computational model:

Together the representation model, a linguistic computational approach based on the functions Δ and Δ^{-1} was also defined in [9] with the following computations and operators:

1. Comparison of 2-tuples

The comparison of linguistic information represented by 2-tuples is carried out according to an ordinary lexicographic order.

Let (s_k, α_1) and (s_l, α_2) be two 2-tuples, with each one representing a counting of information:

- if $k < l$ then $(s_k, \alpha_1) < (s_l, \alpha_2)$
- if $k = l$ then
  1. if $\alpha_1 = \alpha_2$ then (s_k, α_1), (s_l, α_2) represents the same information
  2. if $\alpha_1 < \alpha_2$ then $(s_k, \alpha_1) < (s_l, \alpha_2)$
  3. if $\alpha_1 > \alpha_2$ then $(s_k, \alpha_1) > (s_l, \alpha_2)$

2. Negation operator of a 2-tuple

The negation operator over 2-tuples was defined as:

$$Neg((s_i, \alpha)) = \Delta(g - (\Delta^{-1}(s_i, \alpha))) \tag{4}$$

where $g + 1$ is the cardinality of S, $S = \{s_0, ..., s_g\}$.

3. Aggregation of 2-tuples

The aggregation of information consists of obtaining a value that summarizes a set of values, therefore, the result of the aggregation of a set of 2-tuples must be a 2-tuple. There exists several 2-tuple aggregation operators [9]. For instance, the 2-tuple arithmetic mean is defined as:

Definition 3 *Let* $x = \{(s_1, \alpha_1), \ldots, (s_n, \alpha_n)\}$ *be a set of 2-tuples, the 2-tuple arithmetic mean* $\overline{x}^e$ *is computed as,*

$$\overline{x}^e = \Delta(\sum_{i=1}^{n} \frac{1}{n} \Delta^{-1}(s_i, \alpha_i)) = \Delta(\frac{1}{n} \sum_{i=1}^{n} \beta_i) \tag{5}$$

The arithmetic mean for 2-tuples allows to compute the mean of a set of linguistic values in a precise way without any approximation process.

Example

Let's suppose an example where we have the following linguistic preference vector:

s_2	s_3	s_3	s_2

and $S = \{s_0 : N, s_1 : VL, s_2 : L, s_3 : M, s_4 : H, s_5 : VH, s_6 : P\}$ is the term set shown in the Figure 1.

We want to aggregate these values by using the arithmetic mean as aggregation operator. We follow the process below:

- The preference vector is transformed into 2-tuples as follows:

$(s_2, 0)$	$(s_3, 0)$	$(s_3, 0)$	$(s_2, 0)$

- The linguistic aggregated value obtained by the arithmetic 2-tuple is:
  $\overline{x} = \Delta(\frac{1}{4}(\Delta^{-1}(s_2, 0) + \Delta^{-1}(s_3, 0) + \Delta^{-1}(s_3, 0) + \Delta^{-1}(s_2, 0))) = \Delta(2.5) = (s_3, -0.5)$

ISBN: 978-989-95079-6-8

3.2 Virtual linguistic model

This model was presented by Xu in [15] to avoid the loss of information in processes of CW and increase the operators in processes of CW.

a) Representation model:

In this symbolic model, Xu extended the discrete term set S to a continuous linguistic term set $\bar{S} = \{s_\alpha | s_l < s_\alpha \leq s_t, \alpha \in [1,t]\}$, where, if $s_\alpha \in S$, s_α is called *the original linguistic term*, otherwise, s_α is called *virtual linguistic term* which **does not have assigned any semantics**.

In general, experts use the original linguistic terms to assess the linguistic variables, and the virtual linguistic terms appear in operations.

b) Computational model:

To accomplish processes of CW with this representation model, Xu introduced the following operational laws:

Let $s_\alpha, s_\beta \in \bar{S}$, be any two linguistic terms and $\mu, \mu_1, \mu_2 \in [0,1]$.

1. $(s_\alpha)^\mu = s_{\alpha^\mu}$
2. $(s_\alpha)^{\mu_1} \otimes (s_\alpha)^{\mu_2} = (s_\alpha)^{\mu_1+\mu_2}$
3. $(s_\alpha \otimes s_\beta)^\mu = (s_\alpha)^\mu \otimes (s_\beta)^\mu$
4. $s_\alpha \otimes s_\beta = s_\beta \otimes s_\alpha = s_{\alpha\beta}$
5. $s_\alpha \oplus s_\beta = s_{\alpha+\beta}$
6. $s_\alpha \oplus s_\beta = s_\beta \oplus s_\alpha$
7. $\mu s_\alpha = s_{\mu\alpha}$
8. $(\mu_1 + \mu_2)s_\alpha = \mu_1 s_\alpha \oplus \mu_2 s_\alpha$
9. $\mu(s_\alpha \oplus s_\beta) = \mu s_\alpha \oplus \mu s_\beta$

Example

Let's suppose the example presented previously. In order to compute the arithmetic mean with this model, we have to apply the above operational rules to the linguistic terms:

- The arithmetic mean according to Xu is defined as:

$$\overline{x}^e = \frac{\sum_{i=1}^n s_i}{n} = \frac{1}{n} s_{\sum_{i=1}^n i} \tag{6}$$

- We aggregate the preference vector and we obtain the following collective preference value

$$\overline{x} = \frac{1}{4} s_{(2+3+3+2)} = \frac{1}{4} s_{10} = s_{2.5}$$

3.3 Proportional 2-Tuples representation model

This model presented by Wang and Hao in [13] develops a new way to represent the linguistic information that is a generalization and extension of 2-tuple linguistic representation model [9]. This model deals with linguistic labels in a precise way, but it does not require that the labels are symmetrically distributed around a medium label and either having "equal distance" between them. Besides, it describes the initial linguistic information by members of a "continuous" linguistic scale domain which does not necessarily require the ordered linguistic terms of a linguistic variable being equidistant.

a) Representation model:

This model represents the linguistic information by means of proportional 2-tuples, such as $(0.2A, 0.8B)$ for the case when someone's grades in the answer scripts of a whole course are distributed as $20\%A$ and $80\%B$. The authors point out that if B were used as the approximative grade then some performance information would be lost. This approach, proportional 2-tuples, is based on the concept of *symbolic proportion* [13].

Definition 4 *Let $S = \{s_0, s_1, ..., s_g\}$ be an ordinal term set, $I = [0,1]$ and*

$$IS \equiv I \times S = \{(\alpha, s_i) : \alpha \in [0,1] \; and \; i = 0, 1, ..., g\} \tag{7}$$

where S is the ordered set of $g+1$ ordinal terms $\{s_0, ..., s_g\}$. Given a pair (s_i, s_{i+1}) of two successive ordinal terms of S, any two elements (α, s_i), (β, s_{i+1}) of IS is called a symbolic proportion pair and α, β are called a pair of symbolic proportions of the pair (s_i, s_{i+1}) if $\alpha+\beta = 1$. A symbolic proportion pair (α, s_i), $(1-\alpha, s_{i+1})$ is denoted by $(\alpha s_i, (1-\alpha)s_{i+1})$ and the set of all the symbolic proportion pairs is denoted by $\overline{S}$, i.e., $\overline{S} = \{(\alpha s_i, (1-\alpha)s_{i+1}) : \alpha \in [0,1] \; and \; i = 0, 1, \ldots, g-1\}$.

Remark 2 *Since for $i = \{2, \ldots, g-1\}$, ordinal term s_i can use either $(0s_{i-1}, 1s_i)$ or $(1s_i, 0s_{i+1})$ as its representative in $\overline{S}$, by abuse of notation.*

$\overline{S}$ is called the *ordinal proportional 2-tuple set* generated by S and the members of $\overline{S}$, *ordinal proportional 2-tuples*, which is used to represent the ordinal information for CW.

The notion of proportional 2-tuple allows experts to express their opinions using two adjacent ordinals.

In a similar way to the symbolic 2-tuple Wang and Hao introduced functions in order to facilitate the computations with this type of representation.

Definition 5 *Let $S = \{s_0, s_1, \ldots, s_g\}$ be an ordinal term set and $\overline{S}$ be the ordinal proportional 2-tuple set generated by S. The function $\pi : \overline{S} \rightarrow [0, g]$ was defined by*

$$\pi((\alpha s_i, (1-\alpha)s_{i+1})) = i + (1-\alpha), \tag{8}$$

where $i = \{0, 1, \ldots, g-1\}, \alpha \in [0,1]$ and π is called the position index function of ordinal 2-tuples.

Note that, under the identification convention which was remarked after the definition 4, the position index function π becomes a bijection from $\overline{S}$ to $[0, g]$ and its inverse $\pi^{-1} : [0, g] \rightarrow \overline{S}$ is defined by

$$\pi^{-1}(x) = ((1-\beta)s_i, \beta s_{i+1}) \tag{9}$$

where $i = E(x)$, E is the integer part function, $\beta = x - i$.

b) Computational model:

To operate with linguistic information under proportional 2-tuple contexts, Wang and Hao expanded the computational

ISBN: 978-989-95079-6-8

techniques for symbolic to proportional 2-tuples and defined the following operators:

1. Comparison of Proportional 2-tuples

The comparison of linguistic information represented by proportional 2-tuples is carried out as follows: Let $S = \{s_0, \ldots, s_g\}$ be an ordinal term set and $\bar{S}$ be the ordinal proportional 2-tuple set generated by S. For any $(\alpha s_i, (1-\alpha)s_{i+1}), (\beta s_j, (1-\beta)s_{j+1}) \in \bar{S}$, define $(\alpha s_i, (1-\alpha)s_{i+1}) < (\beta s_j, (1-\beta)s_{j+1}) \Leftrightarrow \alpha i + (1-\alpha)(i+1) < \beta j + (1-\beta)(j+1) \Leftrightarrow i + (1-\alpha) < j + (1-\beta)$.

Thus, for any two proportional 2-tuples $(\alpha s_i, (1-\alpha)s_{i+1})$ and $(\beta s_j, (1-\beta)s_{j+1})$:

- if $i < j$, then
  1. $(\alpha s_i, (1-\alpha)s_{i+1}), (\beta s_j, (1-\beta)s_{j+1})$ represents the same information when $i = j-1$ and $\alpha = 0, \beta = 1$
  2. $(\alpha s_i, (1-\alpha)s_{i+1}) < (\beta s_j, (1-\beta)s_{j+1})$ otherwise
- if $i = j$, then
  1. if $\alpha = \beta$ then $(\alpha s_i, (1-\alpha)s_{i+1}), (\beta s_j, (1-\beta)s_{j+1})$ represents the same information
  2. if $\alpha < \beta$ then $(\alpha s_i, (1-\alpha)s_{i+1}) < (\beta s_j, (1-\beta)s_{j+1})$
  3. if $\alpha > \beta$ then $(\alpha s_i, (1-\alpha)s_{i+1}) > (\beta s_j, (1-\beta)s_{j+1})$

2. Negation operator of a Proportional 2-Tuple

The negation for proportional 2-tuples is defined as:

$$Neg((\alpha s_i, (1-\alpha)s_{i+1})) = ((1-\alpha)s_{g-i-1}, \alpha s_{g-i}), \quad (10)$$

where $g+1$ is the cardinality of S, $S = \{s_0, s_1, \ldots, s_g\}$

3. Aggregation of Proportional 2-Tuple

Wang and Hao defined many aggregation operators to handle processes of CW. The definitions of these aggregation operators are based on canonical characteristic values of linguistic labels. To do so, they used the similar corresponding aggregation operators developed in [9] in order to aggregate ordinal 2-tuples through their position indexes [13].

Example

By using the example presented previously, if we apply the proportional 2-tuples we obtain the following results:

- The arithmetic mean according to Wang and Hao is defined as:

$$\begin{aligned} \overline{x} &= \pi^{-1}(\textstyle\sum_{i=1}^{n} \frac{1}{n}\pi(\alpha s_i, (1-\alpha)s_{i+1})) = \\ &= \pi^{-1}(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(i + (1-\alpha)) \end{aligned} \quad (11)$$

- The preference vector is transformed into proportional 2-tuple as follows:

$(1s_2, 0s_3)$	$(1s_3, 0s_4)$	$(1s_3, 0s_4)$	$(1s_2, 0s_3)$

- The collective preference value obtained under ordinal proportional 2-tuple contexts is

  $\overline{x} = \pi^{-1}(\frac{1}{4}(\pi(1s_2, 0s_3) + \pi(1s_3, 0s_4) + \pi(1s_3, 0s_4) + \pi(1s_2, 0s_3))) = \pi^{-1}(\frac{1}{4}(2+3+3+2)) = \pi^{-1}(2.5) = ((1-0.5)s_2, 0.5s_3) = (0.5s_2, 0.5s_3)$

4 Comparative Analysis

The aim of this contribution is to make a comparative analysis among the symbolic approaches presented in section 3. This analysis will consist of studying all the approaches from different points of view such as: representation model, computations, accuracy, comprehension and so on.

- *The Representation*

  Here, we want to point out that although the authors of the three approaches said that they are based on the fuzzy linguistic approach. It is clear that the *Virtual model* does not follow this approach because its representation does not have any semantics to interpret the linguistic information. Nevertheless, the 2-tuple and the proportional 2-tuple in spite of using additional information to the representation of the linguistic information both, keep the basis of the definition of linguistic variable provided in the fuzzy linguistic approach by using fuzzy numbers to represent the semantics of the linguistic terms.

- *The accuracy*

  By comparing the results provided by the three different approaches in the example showed we can see that the results are similar although the representation of the information would be different. The reason of similar results in this case is because we have considered a linguistic term set symmetrically distributed. However, if we go in deep through the three approaches we can see that the 2-tuple approach guarantees the accuracy when the labels are symmetrically distributed and there is only point with maximum height. The proportional 2-tuple guarantees the accuracy when the terms have the same width in their support. Finally regarding the virtual model is difficult to say anything about accuracy because it does not have semantics to represent the information.

- *The computations*

  By comparing the computational models provided by the three approaches we can observe an important difference. The Virtual model proposes many symbolic operations directly on the linguistic terms obtaining results without any meaning. However the 2-tuple and proportional 2-tuple approaches propose their symbolic computational models with symbolic operators and additionally provide transformation functions to facilitate the computations.

- *The comprehension*

  An important point that should be taken into account regarding the linguistic information is, that this type of information is not only used to represent the information but also to facilitate the comprehension of the results to the users in many different problems.

  Therefore, we can say that the virtual model is only valid for ranking issues because due to the fact the virtual terms do not have semantics are hard to understand apart of a simple order. The proportional 2-tuple provides a clear representation but comparing with the 2-tuple is a little bit more complex due to the fact of the use of four values to represent a single one.

In the table 1 we show the main differences among models.

ISBN: 978-989-95079-6-8

Table 1: Comparative analysis among models

	2-Tuple	Virtual Linguistic	Proportional 2-Tuple
Representation	symbolic and semantics	No semantics	symbolic and semantics
Accuracy	equidistant labels	always because no semantics	same width
Computation	symbolic	no symbolic	symbolic
Comprehension	easy to understand	only ranking issues	understandable

5 Concluding Remarks

In this paper we have reviewed the classical computational models and we have made a brief review of recent symbolic computational models, like the 2-tuple model, the proportional 2-tuple model and the virtual linguistic model.

We have also made a comparative analysis among them in which the most remarkable finding obtained is that the virtual model can not be considered a linguistic model in the sense of the fuzzy linguistic approach due to the fact that it does not require semantics to their linguistic terms. In fact, it could be considered a crisp approach because it does not use any real element to represent the vagueness of the qualitative information.

Acknowledgements

This work is partially supported by the Research Project TIN-2006-02121 and FEDER funds.

References

[1] P.P. Bonissone. *A fuzzy sets based linguistic approach: Theory and applications*, pages 329–339. Approximate Reasoning in Decision Analysis, North-Holland, 1982.

[2] P.P. Bonissone and K.S. Decker. *Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-Off Precision and Complexity*. In L.H. Kanal and J.F. Lemmer, Editors., Uncertainty in Artificial Intelligence. North-Holland, 1986.

[3] G. Buyukozkan, D. Ruan, and O. Feyzioglu. Evaluating e-learning web site quality in a fuzzy environment. *International Journal of Intelligent System*, 22:567–586, 2007.

[4] R. Degani and G. Bortolan. The problem of linguistic approximation in clinical decision making. *International Journal of Approximate Reasoning*, 2:143–162, 1988.

[5] M. Delgado, J.L. Verdegay, and M.A Vila. Linguistic decision making models. *International Journal of Intelligent Systems*, 7:479–492, 1992.

[6] M. Delgado, J.L. Verdegay, and M.A Vila. On aggregation operations of linguistic labels. *International Journal of Intelligent Systems*, 8:351–370, 1993.

[7] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Kluwer Academic, New York, 1980.

[8] F. Herrera, E. Herrera-Viedma, and J.L. Verdegay. A sequential selection process in group decision making with linguistic assessment. *Information Sciences*, 85:223–239, 1995.

[9] F. Herrera and L. Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems*, 8(6):746–752, 2000.

[10] F. Herrera and L. Martínez. The 2-tuple linguistic computational model. Advantages of its linguistic description, accuracy and consistency. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(Suppl.):33–49, 2001.

[11] L. Martínez, M.J. Barranco, L.G. Perez, and et al. A knowledge based recommender system with multigranular linguistic information. In *Proceedings of the international conference on intelligent systems and knowledge engineering (ISKE 2007)*, pages 312–319, Chengdu, China, 2007.

[12] J.M. Merigo and A.A. Gil-Lafuente. The linguistic generalized owa operator and its application instrategic decision making. In *Proceedings of the tenth international conference on enterprise information systems (ICEIS 2008)*, pages 219–224, Barcelona, Spain, 2008.

[13] J. Wang and J. Hao. A new version of 2-tuple fuzzy linguistic representation model for computing with words. *IEEE transactions on fuzzy systems*, 14:435, 2006.

[14] Y.J. Xu and L. Huang. An approach to group decision making problems based on 2-tuple linguistic aggregation operators. In *International Colloquium on Computing, Communication, Control, and Management*, pages 73–77, Guangzhou, China, 2008.

[15] Z.S Xu. A method based on linguistic aggregation operators for group decision making with linguistic preference relations. *Information Science*, 166:19–30, 2004.

[16] R.R. Yager. *Fuzzy Logic: State of the Art*, chapter Fuzzy screening systems, pages 251–261. Kluwer Academic Publishers, 1993.

[17] R.R. Yager. An approach to ordinal decision making. *International Journal of Approximate Reasoning*, 12:237–261, 1995.

[18] L.A. Zadeh. The concept of a linguistic variable and its applications to approximate reasoning. *Information Sciences, Part I, II, III*, 8,8,9:199–249,301–357,43–80, 1975.

ISBN: 978-989-95079-6-8

Universal Integrals Based on Level Dependent Capacities

Erich Peter Klement[1] Anna Kolesárová[2] Radko Mesiar[3,4] Andrea Stupňanová[3]

1. Department of Knowledge-Based Mathematical Systems, Johannes Kepler University
Linz, Austria
2. Institute IAM, Faculty of Chemical and Food Technology, Slovak University of Technology
Bratislava, Slovakia
3. Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology
Bratislava, Slovakia
4. Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech Republic
Email: ep.klement@jku.at, anna.kolesarova@stuba.sk, mesiar@math.sk, andy@math.sk

Abstract— *Three different types of universal integral based on level dependent capacities are introduced and discussed. Two extremal types are based on Caratheodory's idea of inner and outer measures, while the third type is introduced for copula-based universal integrals only.*

Keywords— Choquet integral, Copula, Fuzzy measure, Level dependent capacities, Sugeno integral, Universal integral.

1 Introduction

Capacities (also called fuzzy measures) express the weights of (measurable) subsets of a given universe X in a consistent way. If, for example, X represents a set of criteria then a capacity m on X assigns to each group of criteria $A \subseteq X$ a weight $m(A)$. Universal integrals aggregate the information contained in a capacity m and in a (measurable) function f to a single representative value. The range of f has no influence on the capacity we are exploiting in the aggregation process. In practical applications, however, there is often a need for a different treatment of functions with small values and functions with large values (or even in a more sophisticated way).

This idea of different weights for sets (groups of criteria) at different levels can be expressed by means of a system of capacities (level dependent capacities, see [1, 2, 3]). Note that, motivated by multi-criteria decision problems, one approach to a Choquet integral based on level dependent capacities was proposed and discussed in [1, 2]. On the other hand, one type of a Sugeno integral based on level dependent capacities was introduced in [3] as a solution of the comonotone maxitivite problem for aggregation functions.

The aim of this contribution is to introduce and to discuss universal integrals based on level dependent capacities, i.e., we are looking for extensions of the concept of universal integrals [4] which originally was defined for capacities only. Similar ideas generalizing classical measures are related to the notion of Markov kernels [5], see also [6].

2 Preliminaries

Recall that a *(binary) aggregation operator* or *aggregation function* [7, 8] is a function $A\colon [0,1]^2 \to [0,1]$ which is non-decreasing (in each component) and satisfies $A(0,0) = 0$ and $A(1,1) = 1$. If a binary aggregation function A has neutral element 1, i.e., satisfies $A(a,1) = A(1,a) = a$ for all $a \in [0,1]$, it is called a *semicopula* [9].

Universal integrals were introduced and studied recently in [10, 4]. If the set $\mathcal{H}$ is defined by

$$\mathcal{H} = \{h\colon [0,1] \to [0,1] \mid h \text{ is non-increasing with } h(0) = 1\}.$$

and if $\otimes\colon [0,1]^2 \to [0,1]$ is a semicopula then a non-decreasing functional $J\colon \mathcal{H} \to [0,1]$ is called $\otimes$-*consistent* if, for all $a, b \in [0,1]$, we have

$$J(b \cdot \mathbf{1}_{[0,a]} + (1-b) \cdot \mathbf{1}_{\{0\}}) = a \otimes b. \tag{1}$$

A *capacity space* $(X, \mathcal{A}, m)$ is a triplet consisting of a non-empty universe X, a σ-algebra $\mathcal{A}$ of subsets of X (the elements of $\mathcal{A}$ are called *measurable subsets of* X), and a capacity $m\colon \mathcal{A} \to [0,1]$, i.e., m is isotone with boundary conditions $m(\emptyset) = 0$ and $m(X) = 1$. Furthermore, denote by $\mathcal{F}$ the system of all $\mathcal{A}$-measurable functions $f\colon X \to [0,1]$. Note that, in our context, a function $f\colon X \to [0,1]$ is called $\mathcal{A}$-*measurable* if, for each Borel subset B of $[0,1]$, its preimage $f^{-1}(B) = \{x \in X \mid f(x) \in B\}$ is a measurable subset of X, i.e., belongs to the σ-algebra $\mathcal{A}$. For details about capacity spaces (also called *fuzzy measure spaces*) see [11, 12, 13, 14].

Definition 2.1 Let $J\colon \mathcal{H} \to [0,1]$ be a $\otimes$-consistent functional. A *J-universal integral* $\mathbf{I}_J$ is a mapping which can be defined for each capacity space $(X, \mathcal{A}, m)$ via $\mathbf{I}_{J,m}\colon \mathcal{F} \to [0,1]$ given by

$$\mathbf{I}_{J,m}(f) = J(h_{m,f}), \tag{2}$$

where the function $h_{m,f}\colon [0,1] \to [0,1]$ is defined by $h_{m,f}(t) = m(\{f \geq t\})$ (observe that, because of the monotonicity of m, we have $h_{m,f} \in \mathcal{H}$).

Note that two prototypical universal integrals are the *Choquet integral* [15, 16, 17], in which case we have $J = R$, the standard Riemann integral given by

$$R(h) = \int_0^1 h(t)\, dt \tag{3}$$

(here $\otimes = \Pi$, the standard product), and the *Sugeno integral* [18, 11, 12, 17] which is related to $J = S$ with

$$S(h) = \sup\{\min(t, h(t)) \mid t \in [0,1]\} \tag{4}$$

ISBN: 978-989-95079-6-8

(here $\otimes = \mathrm{Min}$, the greatest semicopula).

Next, we recall two important classes of universal integrals (for more details see [10, 4]):

Proposition 2.2 *Let $\otimes$ be a semicopula. The smallest universal integral $\mathbf{I}_\otimes$ based on $\otimes$ is given by $J_\otimes\colon \mathcal{H} \to [0,1]$,*

$$J_\otimes(h) = \sup\{t \otimes h(t) \mid t \in [0,1]\}, \tag{5}$$

i.e., $\mathbf{I}_\otimes = \mathbf{I}_{J_\otimes}$.

Evidently, because of (5), we have

$$\mathbf{I}_{\otimes,m}(f) = \sup\{t \otimes m(\{f \geq t\}) \mid t \in [0,1]\}. \tag{6}$$

Observe that for $\otimes = \mathrm{Min}$, $\mathbf{I}_{\mathrm{Min}} = \mathbf{I}_S$ is just the Sugeno integral, while $\mathbf{I}_\Pi$ is the *Shilkret integral* [19]. If $T\colon [0,1]^2 \to [0,1]$ is a strict t-norm [20], then $\mathbf{I}_T$ is the *Weber integral* [21].

Recall that a *(two-dimensional) copula* [22] is a binary aggregation function $C\colon [0,1]^2 \to [0,1]$ with annihilator 0 and neutral element 1, i.e., satisfies $C(a,0) = C(0,a) = 0$ and $C(a,1) = C(1,a) = a$ for all $a \in [0,1]$, which is also 2-*increasing*, i.e., for all $a_1, a_2, b_1, b_2 \in [0,1]$ with $a_1 \leq a_2$ and $b_1 \leq b_2$ we have

$$C(a_1,b_1) - C(a_1,b_2) + C(a_2,b_2) - C(a_2,b_1) \geq 0. \tag{7}$$

This means that each copula C induces a probability measure P_C on the Borel subsets of $[0,1]^2$ via

$$\begin{aligned} &P_C([a_1,a_2] \times [b_1,b_2]) \\ &\quad = C(a_1,b_1) - C(a_1,b_2) + C(a_2,b_2) - C(a_2,b_1). \end{aligned} \tag{8}$$

Equivalently, a copula C is a semicopula which is *supermodular*, i.e.,

$$\begin{aligned} &C(a_1,b_1) + C(a_2,b_2) \\ &\quad = C((a_1,b_1) \wedge (a_2,b_2)) + C((a_1,b_1) \vee (a_2,b_2)) \end{aligned} \tag{9}$$

for all $(a_1,b_1),(a_2,b_2) \in [0,1]^2$, where $\wedge$ and $\vee$ are the (pointwise) lattice operations on $[0,1]^2$, i.e, $\wedge = \min$ and $\vee = \max$.

Proposition 2.3 *If $C\colon [0,1]^2 \to [0,1]$ is a copula and P_C the probability measure on the Borel subsets of $[0,1]^2$ induced by C, then the functional $J_C\colon \mathcal{H} \to [0,1]$ given by*

$$J_C(h) = P_C(\{(x,y) \in [0,1]^2 \mid y \leq h(x)\}) \tag{10}$$

is C-consistent.

Given a copula $C\colon [0,1]^2 \to [0,1]$, we shall denote the universal integral $\mathbf{I}_{J_C}$ simply by $\mathbf{I}_{(C)}$. Since Π and Min are copulas, we see that $\mathbf{I}_{(\Pi)} = \mathbf{I}_R$ is the Choquet integral and $\mathbf{I}_{(\mathrm{Min})} = \mathbf{I}_S$ is the Sugeno integral. Therefore, we have two different ways to define the Sugeno integral.

3 Level dependent capacities

The notion of level dependent capacities was introduced in [1], see also [2].

Definition 3.1 Let $(X,\mathcal{A})$ be a measurable space. A *level dependent capacity* on $(X,\mathcal{A})$ is a system $M = (m_t)_{t\in[0,1]}$, where each $m_t\colon \mathcal{A} \to [0,1]$ is a capacity on $(X,\mathcal{A})$.

A special example of a level dependent capacity is a *Markov kernel* [5, 6], where each m_t is a probability measure on $(X,\mathcal{A})$ and, for each $A \in \mathcal{A}$, the function $M_A\colon [0,1] \to [0,1]$ given by $M_A(t) = m_t(A)$ is $\mathcal{A}$-measurable.

Given a level dependent capacity $M = (m_t)_{t\in[0,1]}$ on $(X,\mathcal{A})$, for each $\mathcal{A}$-measurable function $f\colon X \to [0,1]$ we define the function $h_{M,f}\colon [0,1] \to [0,1]$, which accumulates all the information contained in M and f, by

$$h_{M,f}(t) = m_t(\{f \geq t\}). \tag{11}$$

In general, the function $h_{M,f}$ is neither monotone nor even $\mathcal{A}$-measurable (compare with the function $h_{m,f} \in \mathcal{H}$ discussed in Section 2). Following the ideas of inner and outer measures in Caratheodory's approach [23], we introduce the two functions $(h_{M,f})^*, (h_{M,f})_*\colon [0,1] \to [0,1]$ by

$$(h_{M,f})^* = \inf\{h \in \mathcal{H} \mid h \geq h_{M,f}\}, \tag{12}$$

$$(h_{M,f})_* = \sup\{h \in \mathcal{H} \mid h \leq h_{M,f}\}. \tag{13}$$

Note that it is possible to show that for all $t \in [0,1]$

$$(h_{M,f})^*(t) = \sup\{h_{M,f}(u) \mid u \in [t,1]\}, \tag{14}$$

$$(h_{M,f})_*(t) = \inf\{h_{M,f}(u) \mid u \in [0,t]\}. \tag{15}$$

Evidently, both functions $(h_{M,f})^*$ and$(h_{M,f})_*$ are non-increasing and, therefore, belong to $\mathcal{H}$. If the level dependent capacity $M = (m_t)_{t\in[0,1]}$ is constant (i.e., $m_t = m$ for all $t \in [0,1]$) then the three functions considered in (11)–(13) coincide, i.e., we have

$$h_{m,f} = h_{M,f} = (h_{M,f})^* = (h_{M,f})_*. \tag{16}$$

The three functions $h_{M,f}$, $(h_{M,f})^*$ and $(h_{M,f})_*$ allow us to introduce three different extensions of universal integrals for level dependent capacities.

4 Extensions of universal integrals

If $\mathbf{I}_J$ is a J-universal integral then a mapping $\widetilde{\mathbf{I}}_J$ which can be defined on arbitrary measurable spaces $(X,\mathcal{A})$ and arbitrary level dependent capacities M on $(X,\mathcal{A})$ is called an *extension of* $\mathbf{I}_J$ if, whenever the level dependent capacity $M = (m_t)_{t\in[0,1]}$ is constant (i.e., $m_t = m$ for all $t \in [0,1]$), we have $\widetilde{\mathbf{I}}_{J,M} = \mathbf{I}_{J,m}$.

Using the functions in (12) and (13), two extremal extensions of universal integrals can be introduced.

Definition 4.1 Let $\mathbf{I}_J$ be a J-universal integral. The *upper extension* $(\mathbf{I}_J)^*$ (respectively the *lower extension* $(\mathbf{I}_J)_*$) of $\mathbf{I}_J$ for an arbitrary measurable space $(X,\mathcal{A})$, a level dependent capacity M on $(X,\mathcal{A})$, and an $\mathcal{A}$-measurable function $f \in \mathcal{F}$ are given by, respectively,

$$(\mathbf{I}_{J,M})^*(f) = J((h_{M,f})^*), \tag{17}$$

$$(\mathbf{I}_{J,M})_*(f) = J((h_{M,f})_*). \tag{18}$$

If the system $M = (m_t)_{t\in[0,1]}$ is constant (i.e., $m_t = m$ for all $t \in [0,1]$) then we clearly have

$$\mathbf{I}_{J,m} = (\mathbf{I}_{J,M})_* = (\mathbf{I}_{J,M})^*, \tag{19}$$

i.e., both $(\mathbf{I}_J)_*$ and $(\mathbf{I}_J)^*$ are indeed extensions of $\mathbf{I}_J$.

ISBN: 978-989-95079-6-8

For extensions $\widetilde{\mathbf{I}}_{J,M}$ of $\mathbf{I}_J$ which are non-decreasing in M we get

$$(\mathbf{I}_{J,M})_* \leq \widetilde{\mathbf{I}}_{J,M} \leq (\mathbf{I}_{J,M})^*. \tag{20}$$

The approach of extending a known integral to more general situations was applied, e.g., in the case of belief and plausibility measures. Indeed, in the case of a belief measure m [14, 16], the standard Lebesque integral (for probability measures) was applied to probability measures $P \geq m$, and the integral with respect to the capacity m (which, in general, is non-additive) was in this case defined as the infimum of all Lebesgue integrals (with respect to probability measures P with $P \geq m$). Similarly, for a plausibility measure m, all probability measures $P \leq m$ were taken into account, and then the supremum over all Lebesgue integrals (with respect to probability measures P with $P \leq m$) yields the integral for the (non-additive) plausibility measure m. Observe that in both cases the resulting integral is the Choquet integral with respect to m. In our case, having a universal integral $\mathbf{I}_J$ defined for any pair (m, f) of a capacity and a measurable function on the same space, we can compare such pairs based on the corresponding $h_{m,f}$: $(m_1, f_1) \leq (m_2, f_2)$ whenever $h_{m_1,f_1} \leq h_{m_2,f_2}$ (here (m_1, f_1) and (m_2, f_2) need not be defined on the same measurable space, in general). Then $(\mathbf{I}_{J,M})^*(f)$ is just the infimum of all values $(\mathbf{I}_{J,m})(g)$, the infimum being taken over all (m, g) with $h_{M,f} \leq h_{m,g}$. Similarly, $(\mathbf{I}_{J,M})_*(f)$ is just the supremum of $(\mathbf{I}_{J,m})(g)$, where the supremum is taken over all (m, g) with $h_{M,f} \geq h_{m,g}$.

For a copula-based universal integral $\mathbf{I}_C$ there is a third extension — however, it cannot necessarily be applied to any measurable function $f \in \mathcal{F}$.

Definition 4.2 Let C be a copula, $(X, \mathcal{A})$ be a measurable space and $M = (m_t)_{t\in[0,1]}$ be a level dependent capacity on $(X, \mathcal{A})$.

(i) A function $f \in \mathcal{F}$ is called *M-integrable* if $h_{M,f}$ is a measurable function.

(ii) For each M-integrable function $f \in \mathcal{F}$ the corresponding C-based universal integral $\widetilde{\mathbf{I}}_{(C)}$ is defined by $\widetilde{\mathbf{I}}_{(C),M}(f) = J_C(h_{M,f})$ (compare (10)), i.e.,

$$\widetilde{\mathbf{I}}_{(C),M}(f) = P_C(\{(u, v) \in [0, 1]^2 \mid v \leq h_{M,f}(u)\}). \tag{21}$$

Note that a similar extension is possible in the case of a universal integral $\mathbf{I}_\mu$ based on a capacity μ on the Borel subsets of $]0, 1[^2$ as proposed in [24], in which case

$$\mathbf{I}_{\mu,M}(f) = \mu(\{(u, v) \in]0, 1[^2 \mid v < h_{M,f}(u)\}). \tag{22}$$

Remark 4.3

(i) The generalization of the Choquet integral for level dependent capacities as proposed in [1] is closely related to the Riemann-integrability of the function $h_{M,f}$. For example, if X is a finite set (and $\mathcal{A} = 2^X$), and if the level dependent capacity $M = (m_t)_{t\in[0,1]}$ has the same measurability property as a Markov kernel, i.e., for each $A \subseteq X$ the function $M_A\colon [0, 1] \to [0, 1]$ given by $M_A(t) = m_t(A)$ is measurable, then also $h_{M,f}$ is measurable for each $f\colon X \to [0, 1]$. Since $h_{M,f}$ is not necessarily monotone, the Riemann integral in the original definition of the Choquet integral should be replaced by the Lebesgue integral (with respect to the standard Lebesgue measure λ on the Borel subsets of $[0, 1]$), i.e., then (21) turns into

$$\widetilde{\mathbf{I}}_{(\Pi),M}(f) = \int_{[0,1]} h_{M,f}\, d\lambda. \tag{23}$$

However, based on Definition 4.1, we have two other extensions of the Choquet integral given by

$$(\mathbf{I}_{(\Pi),M})^*(f) = \int_0^1 (h_{M,f})^*(x)\, dx, \tag{24}$$

$$(\mathbf{I}_{(\Pi),M})_*(f) = \int_0^1 (h_{M,f})_*(x)\, dx. \tag{25}$$

(ii) Similarly, we have three possible extensions of the Sugeno integral, namely,

$$\widetilde{\mathbf{I}}_{(\mathrm{Min}),M}(f) = \lambda(\{t \in [0, 1] \mid t \leq h_{M,f}(t)\}), \tag{26}$$

$$(\mathbf{I}_{(\mathrm{Min}),M})^*(f) = \sup_{t\in[0,1]} \min(t, h_{M,f}(t)), \tag{27}$$

$$(\mathbf{I}_{(\mathrm{Min}),M})_*(f) = \sup_{t\in[0,1]} \min\Big(t, \inf_{u\in[0,t]} h_{M,f}(u)\Big). \tag{28}$$

In [3] comonotone maxitivity of aggregation functions was investigated and, without any reference to integrals, $(\mathbf{I}_{(\mathrm{Min}),M})^*$ was found to be a solution, compare also [25].

Example 4.4 Let $X = [0, 1]$, $\mathcal{A}$ be the σ-algebra of Borel subsets of $[0, 1]$, and define $M = (m_t)_{t\in[0,1]}$ by

$$m_t = \begin{cases} m^* & \text{if } t \in \left[0, \frac{1}{4}\right], \\ \sqrt{\lambda} & \text{if } t \in \left[\frac{1}{2}, 1\right], \\ m_* & \text{otherwise,} \end{cases} \tag{29}$$

where m^* and m_* are the greatest and the smallest capacity on $(X, \mathcal{A})$, respectively, given by

$$m^*(A) = \begin{cases} 0 & \text{if } A = \emptyset, \\ 1 & \text{otherwise,} \end{cases} \tag{30}$$

$$m_*(A) = \begin{cases} 1 & \text{if } A = X, \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

If $f = \mathrm{id}_{[0,1]}$ then we get

$$(h_{M,f})^* = \mathbf{1}_{[0,\frac{1}{4}]} + \tfrac{1}{\sqrt{2}} \cdot \mathbf{1}_{]\frac{1}{4},\frac{1}{2}[} + \sqrt{1 - f} \cdot \mathbf{1}_{[\frac{1}{2},1]}, \tag{32}$$

$$h_{M,f} = \mathbf{1}_{[0,\frac{1}{4}]} + \sqrt{1 - f} \cdot \mathbf{1}_{[\frac{1}{2},1]}, \tag{33}$$

$$(h_{M,f})_* = \mathbf{1}_{[0,\frac{1}{4}]}. \tag{34}$$

Consequently, we have for the corresponding extensions of the Choquet integral

$$(\mathbf{I}_{(\Pi),M})^*(f) \approx 0.663, \tag{35}$$

$$\widetilde{\mathbf{I}}_{(\Pi),M}(f) \approx 0.486, \tag{36}$$

$$(\mathbf{I}_{(\Pi),M})_*(f) = 0.25. \tag{37}$$

ISBN: 978-989-95079-6-8

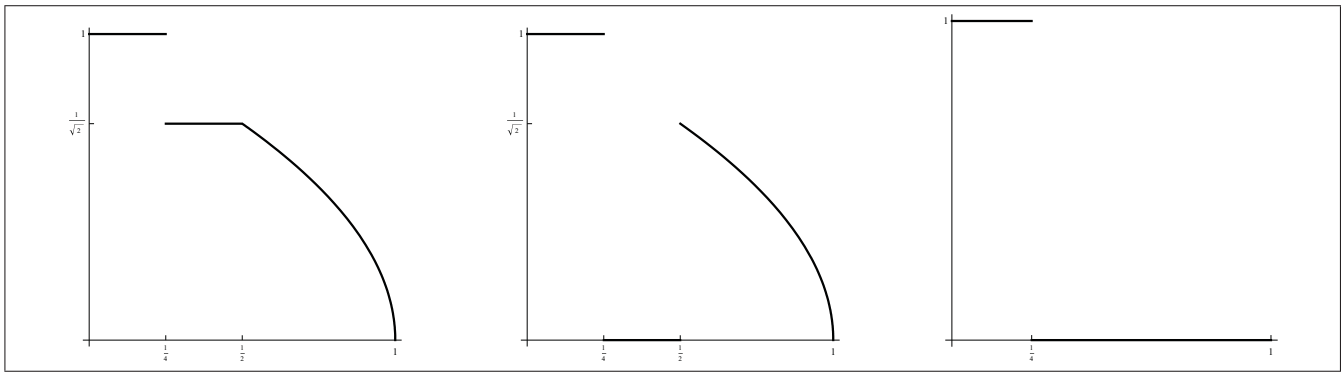


Figure 1: The functions $(h_{M,f})^*$ (left), $h_{M,f}$ (center), and $(h_{M,f})_*$ in Example 4.4

For the extensions of the Sugeno integral we get

$$(\mathbf{I}_{(\mathrm{Min}),M})^*(f) \approx 0.618, \tag{38}$$

$$\widetilde{\mathbf{I}}_{(\mathrm{Min}),M}(f) \approx 0.368, \tag{39}$$

$$(\mathbf{I}_{(\mathrm{Min}),M})_*(f) = 0.25. \tag{40}$$

The Shilkret integral [19] can be considered either as the smallest universal integral based on $\otimes = \Pi$ or as μ-based universal integral, where μ is the capacity on the Borel subsets of $]0,1[^2$ given by $\mu(E) = \sup\{x \cdot y \mid (x,y) \in E\}$. Then

$$(\mathbf{I}_{(\Pi),M})^*(f) \approx 0.385, \tag{41}$$

$$\widetilde{\mathbf{I}}_{\mu,M}(f) \approx 0.385, \tag{42}$$

$$(\mathbf{I}_{(\Pi),M})_*(f) = 0.25. \tag{43}$$

5 Conclusions

Universal integrals based on level dependent capacities can be seen as natural extensions of capacity-based universal integrals acting on different subdomains with (possibly) different capacities. Take, for example, $X = \{1,2\}$ and define the capacities $v_1, v_2 \colon 2^X \to [0,1]$ by $v_1(\{1\}) = \frac{1}{3}$, $v_1(\{2\}) = \frac{2}{3}$, and $v_2(\{1\}) = \frac{3}{4}$, $v_2(\{2\}) = \frac{1}{4}$. Both capacities are additive (i.e., discrete probability measures), and the corresponding Choquet integrals are just weighted arithmetic means, i.e., $W_1(x,y) = \frac{x+2y}{3}$ and $W_2(x,y) = \frac{3x+y}{4}$. Consider the level dependent capacity $M = (m_t)_{t\in[0,1]}$ given by

$$m_t = \begin{cases} v_1 & \text{if } t \le \frac{1}{2}, \\ v_2 & \text{otherwise.} \end{cases} \tag{44}$$

Then, for $(x,y) \in \left[0,\frac{1}{2}\right]^2$, we have

$$(\mathbf{I}_{(\Pi),M})^*(x,y) = \widetilde{\mathbf{I}}_{(\Pi),M}(x,y) = (\mathbf{I}_{(\Pi),M})_*(x,y) = W_1(x,y), \tag{45}$$

and for $(x,y) \in \left[\frac{1}{2},1\right]^2$

$$(\mathbf{I}_{(\Pi),M})^*(x,y) = \widetilde{\mathbf{I}}_{(\Pi),M}(x,y) = (\mathbf{I}_{(\Pi),M})_*(x,y) = W_2(x,y), \tag{46}$$

If $(x,y) \in \left[0,\frac{1}{2}\right] \times \left[\frac{1}{2},1\right]$ then

$$(\mathbf{I}_{(\Pi),M})^*(x,y) = \widetilde{\mathbf{I}}_{(\Pi),M}(x,y) = (\mathbf{I}_{(\Pi),M})_*(x,y) = \tfrac{8x+6y+5}{24}. \tag{47}$$

However, our three extensions of the Choquet integral lead to three different extensions of W_1 (restricted to $\left[0,\frac{1}{2}\right]^2$) and W_2 (restricted to $\left[\frac{1}{2},1\right]^2$): if $(x,y) \in \left]\frac{1}{2},1\right] \times \left[0,\frac{1}{2}\right[$ then

$$\widetilde{\mathbf{I}}_{(\Pi),M}(x,y) = \tfrac{18x+16y-5}{24}, \tag{48}$$

$$(\mathbf{I}_{(\Pi),M})^*(x,y) = W_2(x,y), \tag{49}$$

$$(\mathbf{I}_{(\Pi),M})_*(x,y) = W_1(x,y). \tag{50}$$

Note that $\widetilde{\mathbf{I}}_{(\Pi),M}$ is a continuous aggregation function while $(\mathbf{I}_{(\Pi),M})^*$ and $(\mathbf{I}_{(\Pi),M})_*$ are non-continuous. Finally, observe that $\widetilde{\mathbf{I}}_{(\Pi),M}$ is the ordinal sum extension of the aggregation functions W_1 and W_2 as proposed in [26].

We expect applications of the functionals introduced here in multi-criteria decision making, especially in situations when the weights of the criteria are related to the cardinal values of the score values.

Acknowledgment

The second, third and fourth author gratefully acknowledge the support of the grants VEGA 1/4209/07, VEGA 1/0198/09 and APVV-0012-07. The third author was also supported by the grant MSMVZ6198898701.

References

[1] S. Giove, S. Greco, and B. Matarazzo. Level dependent capacities and the generalized Choquet integral. In D. Dubois, E. P. Klement, and R. Mesiar, editors, *Proceedings 28th Linz Seminar on Fuzzy Set Theory: "Fuzzy Sets, Probability, and Statistics—Gaps and Bridges"*, pages 54–55, Linz (Austria), 2007. Johannes Kepler Universität.

[2] S. Greco, S. Giove, and B. Matarazzo. The Choquet integral with respect to a level dependent capacity. *Fuzzy Sets and Systems*. (in press).

[3] R. Mesiar, A. Mesiarová-Zemánková, and K. Ahmad. Level-dependent Sugeno integral. *IEEE Trans. Fuzzy Systems*. (in press).

[4] E. P. Klement, R. Mesiar, and E. Pap. A universal integral as common frame for Choquet and Sugeno integral. Submitted for publication.

[5] H. Bauer. *Probability Theory and Elements of Measure Theory*. Academic Press, London, 1981.

[6] E. P. Klement. Characterization of finite fuzzy measures using Markoff-kernels. *J. Math. Anal. Appl.*, 75:330–339, 1980.

ISBN: 978-989-95079-6-8

[7] T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar. Aggregation operators: properties, classes and construction methods. In Calvo et al. [8], pages 3–104.

[8] T. Calvo, G. Mayor, and R. Mesiar, editors. *Aggregation Operators. New Trends and Applications*, volume 97 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2002.

[9] F. Durante and C. Sempi. Semicopulæ. *Kybernetika (Prague)*, 41:315–328, 2005.

[10] E. P. Klement, R. Mesiar, and E. Pap. A universal integral. In *Proceedings EUSFLAT 2007, Ostrava*, volume I, pages 253–256. 2007.

[11] Z. Wang and G. J. Klir. *Fuzzy Measure Theory*. Plenum Press, New York, 1992.

[12] E. Pap. *Null-Additive Set Functions*. Kluwer Academic Publishers, Dordrecht, 1995.

[13] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy Measures and Integrals. Theory and Applications*, volume 40 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2000.

[14] E. Pap, editor. *Handbook of Measure Theory*. Elsevier Science, Amsterdam, 2002.

[15] G. Choquet. Theory of capacities. *Ann. Inst. Fourier (Grenoble)*, 5:131–292, 1953–1954.

[16] D. Denneberg. *Non-additive Measure and Integral*. Kluwer Academic Publishers, Dordrecht, 1994.

[17] P. Benvenuti, R. Mesiar, and D. Vivona. Monotone set functions-based integrals. In Pap [14], chapter 33, pages 1329–1379.

[18] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[19] N. Shilkret. Maxitive measure and integration. *Indag. Math.*, 33:109–116, 1971.

[20] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*, volume 8 of *Trends in Logic. Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 2000.

[21] S. Weber. Two integrals and some modified versions: critical remarks. *Fuzzy Sets and Systems*, 20:97–105, 1986.

[22] R. B. Nelsen. *An Introduction to Copulas*, volume 139 of *Lecture Notes in Statistics*. Springer, New York, second edition, 2006.

[23] P. R. Halmos. *Measure Theory*. Van Nostrand Reinhold, New York, 1950.

[24] E. P. Klement, R. Mesiar, and E. Pap. Measure-based aggregation operators. *Fuzzy Sets and Systems*, 142:3–14, 2004.

[25] M. Couceiro and J.-L. Marichal. Representations and characterizations of polynomial functions on chains. arXiv:0811.0309v, 2008.

[26] B. De Baets and R. Mesiar. Ordinal sums of aggregation operators. In B. Bouchon-Meunier, J. Gutiérrez-Ríos, L. Magdalena, and R. R. Yager, editors, *Technologies for Constructing Intelligent Systems*, volume 2: Tools, pages 137–148. Physica-Verlag, Heidelberg, 2002.

ISBN: 978-989-95079-6-8

A Conceptual Framework for Understanding a Fuzzy System

José M. Alonso, Luis Magdalena

European Centre for Soft Computing, Edificio Científico-Tecnológico,
Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Asturias, Spain
Email: {jose.alonso,luis.magdalena}@softcomputing.es

Abstract— *The word Interpretability is becoming more and more frequent in the fuzzy literature. It is admitted as the main advantage of fuzzy systems and it should be given a main role in fuzzy modeling. However, although researchers talk a lot about Interpretability, it is even not clear what it really means. Understanding of fuzzy systems is a subjective task which strongly depends on the person (experience, preferences, knowledge, etc.) who makes the assessment. The general context and the specific problem under consideration have a huge influence too. This paper makes a review on works related to Interpretability and presents the proposal of a conceptual framework that can help to understand fuzzy systems. Moreover, it can be considered as a starting point in order to propose a fuzzy index, easily adaptable to the context of each problem as well as to the user quality criteria, for measuring Interpretability.*

Keywords— Fuzzy systems, Interpretability definition and measurement.

1 Introduction

The concept of interpretability appears in many fields (education, medicine, computer science, etc.) under several names like understandability, comprehensibility, intelligibility, transparency, readability, etc. All these terms are usually considered as synonymous what could yield some confusion. However, some authors [1] distinguish between the term "transparency" (readability) referred as an inherent systemic property (related to the view of the model structure as a white-box) and the term "interpretability" (comprehensibility) which has more cognitive aspects because it is always related to human beings, or more specifically to *humanistic systems* (defined by Zadeh as *those systems whose behavior is strongly influenced by human judgment, perception or emotions* [2]). Notice that readability is assumed as a prerequisite for comprehensibility.

Understanding is likely to be one of the most valuable human abilities. Of course, it is related to the human intelligence and the natural language processing capabilities, because human reasoning is mainly supported by language. The most usual way of explaining something to someone is through the use of words, sentences, linguistic expressions, etc. Of course, gestures and symbols are also used as additional communication tools but they only represent other kinds of languages. Unfortunately, knowledge about these kinds of cognitive tasks is still quite reduced. However, let us underline that this work belongs to the field of artificial intelligence and it will focus on analyzing the interpretability of knowledge-based systems, and more specifically of fuzzy rule-based systems (FRBSs). The main goal of this work is to study how comprehensible are such systems from a human point of view, opening a constructive discussion.

The use of linguistic variables [2] to overcome the ineffectiveness of computers in dealing with systems whose behavior is strongly influenced by human judgment, perceptions or emotions was pointed out by Zadeh long time ago: *In order to be able to make significant assertions (...) it may be necessary to abandon the high standards of rigor and precision that we have become conditioned to expect of our mathematical analyses (...) and become more tolerant of approaches which are approximate in nature* [2]. Following the Zadeh's advice if we really want to define a useful index for system modeling, it is necessary to change our mind. Numerical indices should be forgotten and in turn fuzzy indices should be defined, i.e., the focus must be shifted *from computing with numbers to computing with words, from manipulation of measurements to manipulation of perceptions* [3]. In consequence, the right approach to assess interpretability in an effective way consists in proposing a fuzzy index instead of a numerical one. A first attempt was presented in [4] where a hierarchical fuzzy system was used to get an interpretability measure. That proposal opened this way but a lot of work remains to do.

The expressivity of linguistic rules [5] is acknowledged to be quite close to natural language what favors the interpretability because human understanding is made in terms of natural language. That is why it is useful to take into account the experience gained by natural language processing researchers. For instance, the philosopher Paul Grice established the next four conversational maxims [6] which arise from the pragmatics of natural language and they are based on the common sense:

1. *Maxim of Quality*: Do not say what you believe to be false. Do not say anything without adequate evidence.
2. *Maxim of Quantity*: Make your contribution as informative as required for the current purposes of the exchange.
3. *Maxim of Relation*: Be relevant.
4. *Maxim of Manner*: Avoid obscurity of expression. Avoid ambiguity. Be brief. Be orderly.

Keeping the Grice's maxims in mind during the fuzzy modeling process can help to make easier the understanding of FRBSs. The rule base must be coherent avoiding the use of inconsistent rules (*Maxim of Quality*), redundant rules (*Maxim of Quantity*), and ambiguity rules (*Maxim of Manner*). In addition, selecting the most relevant rules (*Maxim of Relation*) will yield more compact and robust systems.

The rest of the paper is structured as follows. Section 2 has a look on definitions of interpretability found in the literature. In addition, it makes a global review on all the aspects that should be taken into account in the interpretability

ISBN: 978-989-95079-6-8

assessment, setting a conceptual framework for characterizing interpretability. Section 3 describes how to combine the main factors included in the proposed framework for measuring interpretability of FRBSs. Finally, section 4 offers some conclusions and points out future works.

2 Understanding a fuzzy system

Authors talk a lot about interpretability but it is not easy to find a formal definition in the literature. Thus, it is necessary to think on the following question: How can interpretability be defined? The first bid to set a formal definition was made by Tarski et al. [7] a very long time ago (in 1953). He formulated a mathematical definition in the context of classical logic, setting the basis for identifying interpretable theories. In short, *assuming T and S are formal theories, T is interpretable in S if and only if there is a way to pass from T to S, assuring that every theorem of T can be translated and proved into S*.

Regarding the fuzzy literature, a similar definition is included as part of the formal framework proposed in [8]. It distinguishes between a formal language *L* (fuzzy logic) used for describing the model under consideration, and a user-oriented language *L'* (usually the natural language) used for explaining the model to the user. If the system is interpretable, the translation from *L* to *L'* should be made by the user with a small effort. In an informal way, people say that a model is interpretable if they are able to describe it easily.

A more formal definition was given by Bodenhofer and Bauer [9]: *Interpretability means possibility to estimate the system's behavior by reading and understanding the rule base only*. Since the rule base understanding strongly depends on the readability of the involved linguistic expressions, the authors focused on analyzing the interpretability at the level of fuzzy partitioning (linguistic variables) from an intuitive and mathematically exact point of view: *The obvious orderings and inclusions of linguistic terms must not be violated by the corresponding fuzzy sets*. As a result, fuzzy partitioning readability was assumed to be a prerequisite to build interpretable FRBSs.

The comprehensibility of a FRBS depends on all its components, i.e., it depends on the knowledge base (KB) transparency but also on the inference mechanism understanding. Previous works [10, 11] have thoroughly analyzed the main factors that influence the KB readability. Also, a complete study on the interpretability constraints most frequently used in fuzzy modeling has been recently published [1].

Fig. 1 describes the main factors to be considered regarding interpretability of FRBSs. It is inspired on the taxonomy of interpretability of fuzzy systems introduced by [12], which is extended adding our own notation and concepts, and also including some of the most significant constraints extracted from [1]. There are two main points of view to be considered when assessing interpretability of FRBSs (*Global description* and *Local explanation*). The global view presents the system as a whole explaining its global behavior and trend. However, the local view focuses on each individual situation, explaining specific behaviors for specific events. For instance, if we had a fuzzy controller for driving a car, the global view would give an idea on the kind of operations it can do (go straight forward, turn on the right/left, speed up, brake, etc.) and even on the driver style (aggressive, sluggish, etc.). On the contrary, the local view would explain each specific manoeuvre.

Additional information is detailed in the following subsections. Pay attention to the fact that both viewpoints could lead to contradictory goals. The first one (*Global description*) prefers rules as compact as possible, while the second one (*Local explanation*) favors the use of complete rules (*The more general rules, the larger the number of rules that can be fired at the same time*).

2.1 Global description (system structure)

In order to assess the simplicity of a FRBS the following assumption is made: *The more compact the KB, the simpler its understanding, i.e., the higher the interpretability*. This reasoning follows the principle of incompatibility formulated by Zadeh [13] in 1973: *As the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance become almost mutually exclusive characteristics. The closer one looks at a real-world problem, the fuzzier becomes its solution*.

The global description of a linguistic FRBS can be analyzed looking at different abstraction levels as illustrated on left part of Fig. 1. First, the lowest level corresponds to the level of individual fuzzy sets. It includes those constraints demanded to build interpretable fuzzy sets, regarding mathematical properties of the membership functions. At the second level, there are several constraints with respect to the combination of several fuzzy sets to form a fuzzy partition. The use of linguistic variables favors the readability, but it is not enough to ensure interpretability. Hence, some linguistic constraints must be superimposed to the fuzzy partition definition to be interpretable. Fortunately, Ruspini defined (in 1969) a special kind of partition called Strong Fuzzy Partition (SFP) [14] that satisfies most demanded semantic constraints (distinguishability, coverage, normality, convexity, etc). In practice, satisfying all constraints is almost impossible and useless because they represent a very restrictive set of conditions that usually yield systems with very small accuracy. Notice that looking for a good accuracy-interpretability trade-off is the most complex task of fuzzy modeling. Especially relevant are some recent successful biomedical applications [15].

Once a set of linguistic terms with their associated semantics has been defined, they can be used to express linguistic propositions. Then, several propositions are combined to form fuzzy rules describing the system behavior. However, in addition to the analysis of each individual rule it is needed to study the combination of several rules, achieving the highest abstraction level. Notice that defining a global semantics previous to the rule definition makes easier the rule understanding. Only if all the rules use the same linguistic terms (defined by the same fuzzy sets) it will be possible to make a rule comparison at the linguistic level. In order to get fully meaningful partitions the right linguistic terms should be selected according to the problem context. Nevertheless, matching linguistic terms and fuzzy sets is not a straightforward task, for instance finding good linguistic terms for fuzzy partitions automatically generated from data is sometimes not feasible.

To sum up, the satisfaction of all constraints enumerated on the left part of Fig. 1 guarantees the interpretability of the FRBS from the structural point of view.

ISBN: 978-989-95079-6-8

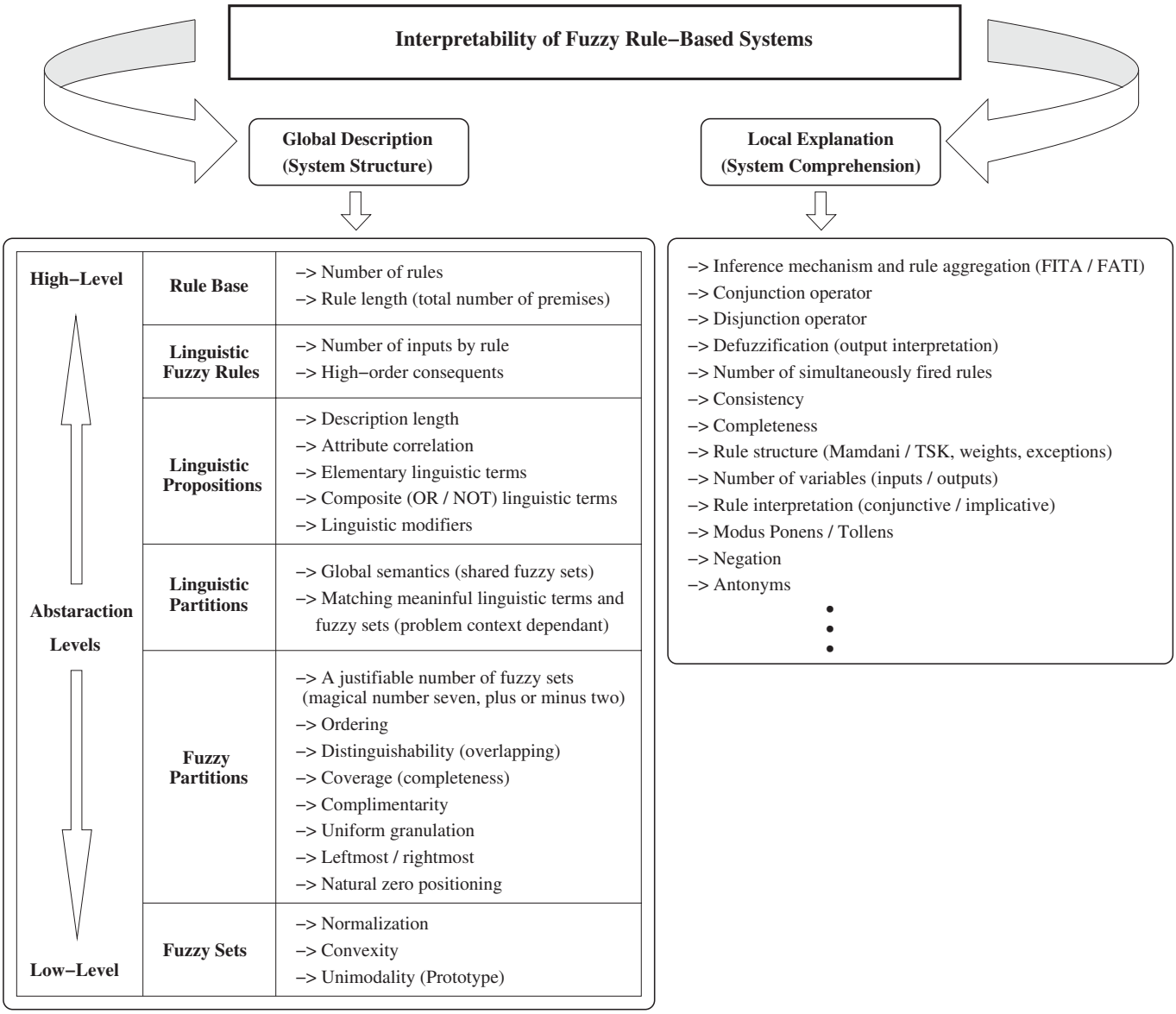


Figure 1: A conceptual framework for characterizing interpretability of FRBSs.

2.2 Local explanation (system comprehension)

Understanding the system behavior from its linguistic description is a very hard task that involves the inference level going beyond the former analysis of the system structure.

In addition, there is a need to give some comments about the inference mechanism implementation distinguishing between FITA (First Infer Then Aggregate) and FATI (First Aggregate Then Infer). It includes the fuzzy operator definitions for conjunction, disjunction, aggregation, and defuzzification. Furthermore, taking into account that as the result of a fuzzy inference several rules can be fired at the same time for a given input vector, the interpretability strongly depends on the number of rules that can be simultaneously fired. The smaller that value, the higher the interpretability. In fact, a model made up of thousand rules (where at maximum ten rules are fired together) may be seen as more interpretable than a model including only one hundred rules (where most of them are simultaneously fired). Notice that the whole rule base should be consistent (not including redundancies, contradictions, etc.) and it should cover most possible situations.

Although Mamdani rules are widely admitted as the more interpretable kind of rules, there are many other rule formats. The second most used rules are the well-known Takagi-Sugeno rules, but there are also rules with exceptions, rules with weights, and so on. The different rule formats can be compared and it is possible to discuss which one is better regarding interpretability from a structural point of view but it is a controversial issue. For instance, for many people the most interpretable rule format is the one they usually work with disregarding its complexity. This proves that many psychological aspects make influence when assessing interpretability. It is a clear example of the "Hammer principle" formulated by Zadeh [16]: *When the only tool you have is a hammer, everything begins to look like a nail*.

Finally, modus Ponens/Tollens must be carefully taken into account. Notice that the fact the all rules are fired at the same time make not easy to establish logical chains of reasoning. It is also necessary to remark that the use of negation and antonyms are quite usual in natural language but their representation using fuzzy logic is still a matter of research.

ISBN: 978-989-95079-6-8

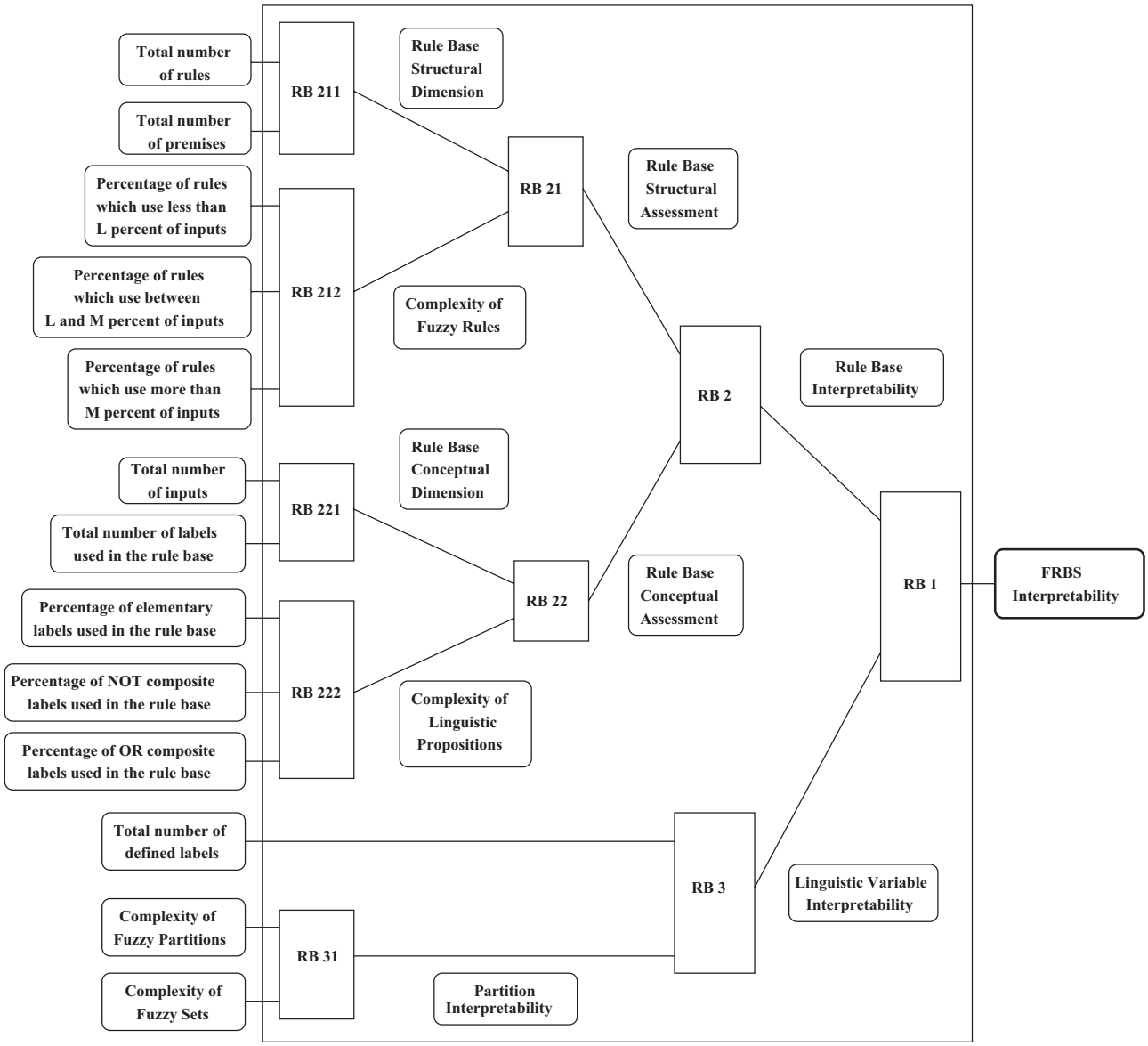


Figure 2: A conceptual framework for assessing interpretability (*Global description*) of FRBSs.

3 Measuring interpretability

Once identified all factors that should be kept in mind regarding interpretability, the definition of a universal interpretability index able to combine all of them becomes a great challenge.

The aim of this section is to introduce a conceptual framework for assessing interpretability. It is represented in the form of a hierarchical diagram in Fig. 2. Of course, this diagram only takes into account interpretability from the structural point of view (*Global description*). The local explanation will be addressed in future works.

To start with, the whole set of factors represented in left part of Fig. 1 has been summarized by a small subset that could be extended in the future. The global diagram can be seen as a flow chart with thirteen inputs (measurable factors to take into account regarding interpretability) and one output (Interpretability measure). The diagram keeps the abstraction levels shown in Fig. 1 (low-level at the bottom of the figure and high-level at the top) and selected inputs are grouped according to the information they convey.

Interpretability of a FRBS is estimated as a combination of two estimators at both low and high level. On the one hand, partition interpretability regards the complexity of each fuzzy set but also the complexity of the whole partition. An estimation of the low-level interpretability (regarding the description of all linguistic variables) is computed adding the number of labels (linguistic terms). Notice that the simple diagram depicted in Fig. 2 does not include anything with respect to the interpretability of linguistic partitions. It is assumed the use of SFPs and global semantics, not entering to the way how linguistic terms are named. On the other hand, high-level interpretability (called *rule base interpretability* in the figure above) involves analysis at the three highest sublevels (linguistic propositions, linguistic fuzzy rules, and rule base).

Drawing an analogy between a set of fuzzy rules and a set of sentences in natural language, the interpretability is assessed regarding both Syntax and Semantics. On the one hand, Syntax can be defined as the *arrangement of words in sentences, clauses, and phrases, and the study of the formation of sentences and the relationship of their component parts*.

ISBN: 978-989-95079-6-8

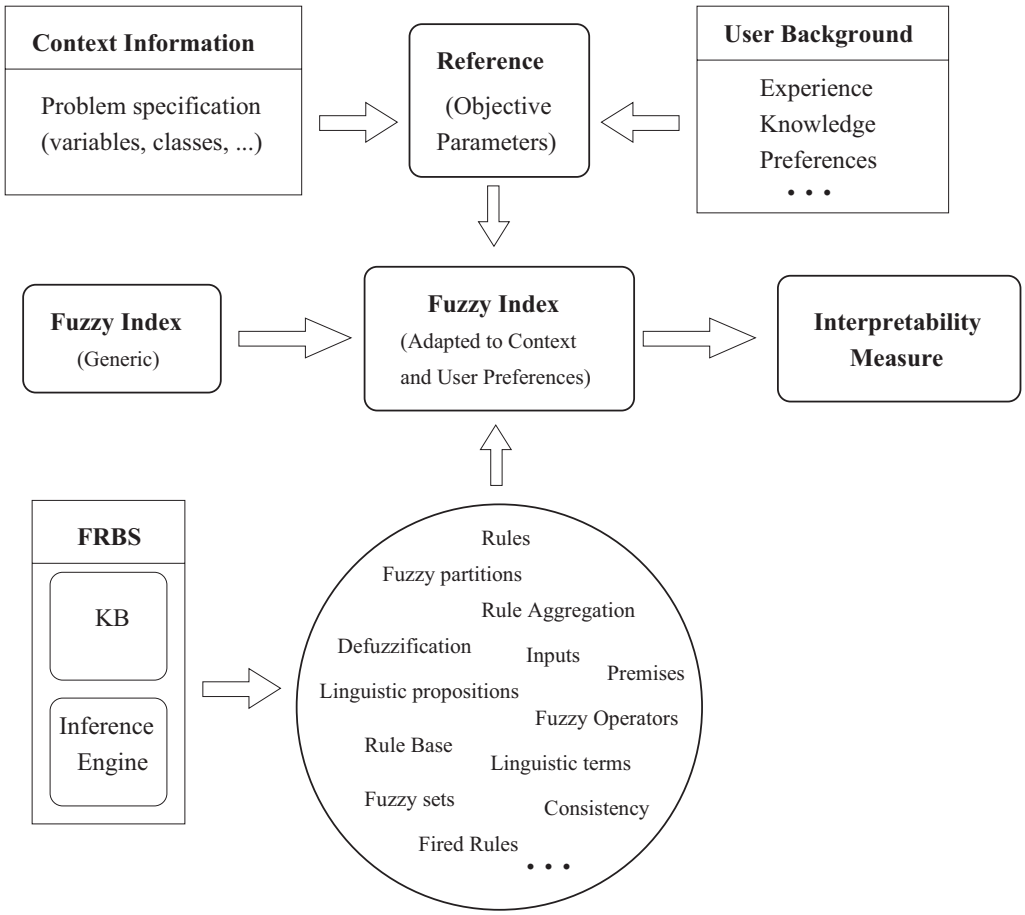


Figure 3: A fuzzy index (adaptable to context problem and user preferences) for measuring Interpretability.

On the other hand, Semantics makes reference to the *Study of meaning*[1]. In our context Syntax is related to dimension and complexity of the rule base (what we have named as *Rule Base Structural Assessment*), computed in a very simplistic way, only considering the total number of rules, premises and premise by rule. It covers both linguistic fuzzy rules and rule base abstraction levels. Semantics in turn takes into account the complexity at the level of linguistic propositions used in the rules (what we have named as *Rule Base Conceptual Assessment*).

The process of measuring something consists in comparing it with a reference (standard unit of measurement) such us a meter for measuring length. However, finding out the suitable reference is not always feasible and the task is especially difficult when measuring non-physical properties. It is widely admitted that interpretability assessment is clearly context dependant. There is not a universal reference; on the contrary the reference will change depending on the problem and depending on the person who makes the assessment. Therefore, the general proposed framework has to be adapted to the specific features of each problem under consideration as well as to the user's background and preferences. For instance, *Total number of rules* may be defined as a linguistic variable made up of five linguistic terms (*Very low*, *Low*, *Medium*, *High*, *Very high*). Nevertheless, the meaning of each linguistic term needs to be defined carefully. *What value should be taken as a prototype for small number of rules (three, ten, one hundred, etc)?* If we were analyzing a FRBS for classification among three kinds of wines, the minimum number of rules should be three, but we need to ask to the people who are going to interact with the system in order to really know how to characterize the linguistic variables. In fact, the perception of interpretability will change depending on the kind of user. The point of view of a system designer who is used to work with fuzzy systems is likely to be very different from the point of view of the domain expert who perfectly knows the problem and how should be the system behavior, but it will be even much more different from the final user who could have only a superficial knowledge of the problem, and who probably has not heard anything about fuzzy logic.

Fig. 3 describes how to build an easily adaptable fuzzy index for assessing interpretability. A generic fuzzy index like the one presented in Fig. 2 has to be tuned and adapted for each problem regarding both the problem definition and the user quality criteria. The system is flexible enough for making an easy adaptation. It consists in defining a reference with all collected information about the problem and the evaluator user (system designer, domain expert, and/or final user). Such reference yields the ranges (universes of discourse) along with the modal points of the fuzzy partitions used to define the input variables (*Total number of rules*, *Total number of premises*, *Percentage of rules which use less than L percent of inputs*, etc.) of the generic fuzzy index. The linked rule bases as well as the intermediate input-output variables could be tuned too.

[1]Both definitions were got from the Encyclopedia Britannica (http://www.britannica.com)

ISBN: 978-989-95079-6-8

4 Final remarks

Previous works has made a great effort to establish the basis for building interpretable fuzzy systems. There are many different works regarding interpretability on the fuzzy literature. Recently, some works have made a global review of the literature putting together contributions of different authors. Following that way, this work has formalized a conceptual framework for characterizing and assessing interpretability of fuzzy systems.

The use of multi-objective approaches is becoming a more and more important topic in fuzzy modeling [17] because of interpretability and accuracy are conflictive goals. In this specific field the interpretability of the model is usually only considered from the point of view of the fuzzy designer. First, it is necessary to make a qualitative and quantitative comparison of all obtained solutions. Then, the best solutions can be selected from a Pareto front regarding the accuracy-interpretability trade-off. It is possible to set a qualitative ranking of solutions based on a comparison per couples, without measuring the interpretability of each individual solution, setting some kind of pre-order is enough. Although there are several accuracy indices, interpretability is measured taking into account only basic parameters what is a strong limitation. Thus, the use of interpretability indices guiding the modeling process could help to achieve better solutions.

Setting qualitative rankings is quite common in the context of semantic web search where retrieved documents have to be ranked before presenting them as answer to a query. For instance, BUDI [18] is a meta-searcher based on fuzzy logic which uses a fuzzy similarity function for comparing documents. It regards the size of the documents, the number of series of words in the same position in both documents, but also the complexity and rarity of words and linguistic propositions. This approach could be extended to the interpretability assessment problem, considering that instead of documents what are going to be compared are the linguistic descriptions of FRBSs.

In the future, experimental analysis must be carried out in order to adapt the theoretical developments to the real worlds. In order to get a universal index adaptable to the user preferences, it is necessary to study how different kinds of user (fuzzy designer, domain expert, and final user) interact with fuzzy systems in a different way and they have different interpretability requirements.

Finally, it is necessary to advance on the paradigm of computing with words and perceptions (CWW/P) [3] which marks an evolution of fuzzy logic, an extension of current theories of fuzzy sets. It is strongly related with meaning, captured by the use of linguistic expressions, words and connectives. As a result, works regarding interpretability assessment can take profit from current research on CWW/P but also they will help to develop new ideas in relation with this new paradigm.

Acknowledgment

This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias) and Spanish government (CICYT) under grant: TIN2008-06890-C02-01.

References

[1] C. Mencar and A. M. Fanelli. Interpretability constraints for fuzzy information granulation. *Information Sciences*, 178(24):4585–4618, 2008.

[2] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Parts I, II, and III. Information Sciences*, 8, 8, 9:199–249, 301–357, 43–80, 1975.

[3] L. A. Zadeh. From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems - I: Fundamental theory and applications*, 45(1):105–119, 1999.

[4] J. M. Alonso, L. Magdalena, and S. Guillaume. HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *International Journal of Intelligent Systems*, 23(7):761–794, 2008.

[5] E. H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic systems. *IEEE Transactions on Computers*, 26(12):1182–1191, 1977.

[6] H. P. Grice. Logic and conversation. *Cole, P. and Morgan, J. (eds.) Syntax and semantics, New York: Academic Press*, 3:43–58, 1975.

[7] A. Tarski, A. Mostowski, and R. Robinson. *Undecidable Theories*. North-Holland, 1953.

[8] C. Mencar, G. Castellano, and A. M. Fanelli. Some fundamental interpretability issues in fuzzy modeling. In *Joint EUSFLAT-LFA*, pages 100–105, Barcelona, Spain, September, 7-9, 2005.

[9] U. Bodenhofer and P. Bauer. A formal model of interpretability of linguistic variables. *In [19]*, pages 524–545, 2003.

[10] J. Espinosa and J. Vandewalle. Constructing fuzzy models with linguistic integrity from numerical data-afreli algorithm. *IEEE Transactions on Fuzzy Systems*, 8(5):591–600, 2000.

[11] S. Guillaume and B. Charnomordic. Generating an interpretable family of fuzzy partitions from data. *IEEE Transactions on Fuzzy Systems*, 12 (3):324–335, 2004.

[12] S.-M. Zhou and J. Q. Gan. Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling. *Fuzzy Sets and Systems*, 159(23):3091–3131, 2008.

[13] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(1):28–44, 1973.

[14] E. H. Ruspini. A new approach to clustering. *Information and Control*, 15(1):22–32, 1969.

[15] P. Fazendeiro, J. V. de Oliveira, and W. Pedrycz. A multiobjective design of a patient and anaesthetist-friendly neuromuscular blockade controller. *IEEE Transactions on Biomedical Engineering*, 54(9):1667–1678, 2007.

[16] L. A. Zadeh. *Applied Soft Computing - Foreword*, 1(1):1–2, 2001.

[17] H. Ishibuchi and Y. Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *International Journal of Approximate Reasoning*, 44(1):4–31, 2007.

[18] J. Serrano-Guerrero et al. BUDI: Architecture for fuzzy search in documental repositories. *Mathware and Soft Computing*, 16(1):71–85, 2009.

[19] J. Casillas et al. *Interpretability issues in fuzzy modeling*, volume 128. Studies in Fuzziness and Soft Computing, Springer-Verlag, Heidelberg, 2003.

ISBN: 978-989-95079-6-8

An Experimental Study on the Interpretability of Fuzzy Systems

José M. Alonso, Luis Magdalena

European Centre for Soft Computing, Edificio Científico-Tecnológico,
Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Asturias, Spain
Email: {jose.alonso,luis.magdalena}@softcomputing.es

Abstract— *Interpretability is one of the most significant properties of Fuzzy Systems which are widely acknowledged as gray boxes against other Soft Computing techniques such as Neural Networks usually regarded as black boxes. It is essential for applications with high human interaction (decision support systems in medicine, economics, etc). The use of accuracy indices to guide the fuzzy modeling process is broadly extended. In turn, although there have been a few attempts to define Interpretability indices, we are still far away from having a universal index. With the aim of evaluating the most used indices an experimental analysis (in the form of a web poll) was carried out yielding some useful clues to keep in mind regarding Interpretability assessment. Results extracted from the poll show the inherent subjectivity of the measure because we collected a huge diversity of answers. Nevertheless, comparing carefully all the answers, it was possible to find out some interesting user profiles.*

Keywords— Fuzzy modeling, Interpretability assessment.

1 Introduction

Fuzzy modeling (FM), i.e., system modeling with fuzzy rule-based systems (FRBSs), is an important and active research line inside the fuzzy community. Fuzzy Logic (FL) was introduced by Zadeh [1] (in 1965) and its semantic expressivity, using linguistic variables [2] and linguistic rules [3], favors the interpretability of the modeled system (at least from the structural transparency viewpoint) because it is quite close to expert natural language. From 1965 to 1990, fuzzy designers focus on modeling highly interpretable systems, mainly working with expert knowledge and a few simple linguistic rules. Then, researchers realized that to deal with complex systems expert knowledge was not enough. Thus, from 1990 to 2000, the main effort was made regarding the accuracy of the final model, building complicated fuzzy rules with high accuracy (applying machine learning techniques to extract knowledge from data) but disregarding the model interpretability because automatically generated rules are rarely as readable as desired. Nowadays, a new challenge lies in looking for compact and robust systems with a good accuracy-interpretability trade-off.

Regarding the interpretability assessment of FRBSs, their comprehensibility depends on all their components, i.e., it depends on the knowledge base (KB) transparency but also on the inference mechanism understanding. There are also some crucial psychological factors; for instance, for some people the most interpretable models are the ones they are used to work with, disregarding the model complexity. Anyhow, previous works [4, 5] have thoroughly analyzed the main factors (rule base and fuzzy partitioning) that influence the KB readability. Also, a complete study on the interpretability constraints most frequently used in literature has been recently published [6]. However, once identified all such elements, the current challenge lies in how to combine them in order to obtain a good index. Unfortunately, only a few works have dealt with this issue. As explained by [7] it is possible to distinguish two main interpretability levels: Low-level or fuzzy set level, and high-level or fuzzy rule level. There are some works regarding interpretability measurement at low-level [8, 9, 10, 11, 12] which consist of mathematical formulas to evaluate the main partition properties such as distinguishability, similarity, coverage, overlapping, etc. These indices are usually used to preserve the interpretability of fuzzy systems automatically generated from data. They are also used as part of tuning processes devoted to increase the accuracy of the final model while keeping good interpretability. Furthermore, there are some simple indices, mainly applied to multi-objective fuzzy genetics-based machine learning, regarding the rule base interpretability [13]: (1) *Number of rules*; (2) *Total rule length* (addition of the number of premises defined in all the rules); (3) *Average rule length* (total rule length divided by the number of rules).

However, only a few researchers have tackled with the challenge of defining an index covering both low and high levels. The first one was the *Nauck's index* [14], a numerical index designed (in 2003) to evaluate fuzzy rule-based classification systems, and computed as the product of three terms: $I_{Nauck} = Comp \times Part \times Cov$. *Comp* represents the complexity of a classifier measured as the number of classes divided by the total number of premises. *Part* stands for the average normalized partition index overall input variables. It is computed as the inverse of the number of labels minus one (two is the minimum number of linguistic terms in a partition) for each input variable. Finally, *Cov* is the average normalized coverage degree of the fuzzy partition. It is equal to one for strong fuzzy partitions (SFPs).

A second global index was defined in 2006 [15] and improved in 2008 [16]. It consists of a *fuzzy index* which was initially inspired on the Nauck's index. Six variables (total number of rules, total number of premises, number of rules which use one input, number of rules which use two inputs, number of rules which use three or more inputs, and total number of labels defined by input) are considered as inputs of a fuzzy system and they are grouped, according to the information they convey. In consequence, the interpretability index is computed as the result of inference of a hierarchical fuzzy system made up of four linked KBs. It is specially designed for the context of classification problems solved by means of a specific kind of FRBSs generated following the HILK (Highly Interpretable Linguistic Knowledge) [16] methodology, and assuming SFPs for all system variables.

The rest of the paper is structured as follows. Section 2

ISBN: 978-989-95079-6-8

Table 1: Comparison of interpretability indices (measures).

	Method	Number of rules	Total rule length	Average rule length	Nauck's index	Fuzzy index
KB1	CL-FDT-DS-FDT	20	43	2.15	0.0174	0.452
KB2	CL-FDT-DS-FDT-S	5	9	1.8	0.1667	0.92
KB3	CL-FDT-DS-WM	53	643	12.132	0.0011	0.144
KB4	CL-FDT-DS-WM-S	8	16	2	0.1484	0.839
KB5	CL-WM-DS-FDT	21	49	2.333	0.0153	0.444
KB6	CL-WM-DS-FDT-S	5	10	2	0.2625	0.919
KB7	CL-WM-DS-WM	46	545	11.848	0.0013	0.192
KB8	CL-WM-DS-WM-S	3	6	2	0.3056	0.924
KB9	FDT-S	8	19	2.375	0.0763	0.814
KB10	WM-S	6	18	3	0.0873	0.742
KB11	FDT-P	32	94	2.937	0.0079	0.392
KB12	FDT-P-S	6	15	2.5	0.1714	0.837

Table 2: Comparison of interpretability indices (ranking).

Index	+ Interpretability -
Number of rules	KB8, KB2/KB6, KB10/KB12, KB4/KB9, KB1, KB5, KB11, KB7, KB3
Total rule length	KB8, KB2, KB6, KB12, KB4, KB10, KB9, KB1, KB5, KB11, KB7, KB3
Average rule length	KB2, KB4/KB6/KB8, KB1, KB5, KB9, KB12, KB11, KB10, KB7, KB3
Nauck's index	KB8, KB6, KB12, KB2, KB4, KB10, KB9, KB1, KB5, KB11, KB7, KB3
Fuzzy index	KB8, KB2, KB6, KB4, KB12, KB9, KB10, KB1, KB5, KB11, KB7, KB3

makes a comparison of several interpretability indices in an experimental analysis. Results extracted from a web poll show clearly the intrinsic subjectivity of the measure. As it is explained in section 3, although we got a huge diversity of answers, at first glance completely different, after looking carefully it was possible to find out some interesting user profiles. Finally, section 4 offers some conclusions and points out future works.

2 Experimental analysis

With the aim of making a fair (qualitative and quantitative) comparison of the five indices remarked in the previous section (Number of rules, Total rule length, Average rule length, Nauck's index, and fuzzy index) this experimental study deals with the well known benchmark classification problem called WINE which data set is freely available at the UCI[1] (University of California, Irvine) machine-learning repository. It contains 178 instances coming from results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

Following the HILK modeling methodology [16] twelve KBs, of several sizes, have been generated for the WINE recognition problem. Looking for maximizing the interpretability of final KBs, a global semantics (based on the use of SFPs) is defined previously to rule definition. As a result, for each KB all the rules use the same linguistic terms (defined by the same fuzzy sets), and rule comparison can be done at the linguistic level. The original data set was randomly divided into two subsets taking 50% of data for training and the remaining part for test. HILK copes with different rule induction techniques in order to get enough diversity. Second column in Table 1 contains the abbreviations of the combined methods. CL means clustering previous to rule induction, WM represents the well-known Wang and Mendel's algorithm [17], FDT stands for the popular Fuzzy Decision Tree algorithm [18], DS is data selection in training set previous to rule induction, P means pruning of the tree, and S stands for simplification procedure. All selected algorithms are implemented in KBCT [19], a free software tools for designing fuzzy systems. WM and FDT implementations differ from the original ones in the fuzzy partition design step. Interpretable fuzzy partitions are defined previous to rule induction (in our study five labels per variable were initially defined). Details about the induction algorithms are explained in the cited literature.

The five interpretability indices have been compared from both quantitative (Table 1) and qualitative (Table 2) viewpoints. Table 1 includes results of computing the five selected interpretability indices for the twelve generated KBs. The comparison of the obtained values lets us set rankings from the interpretability point of view. Notice that KBs with equivalent interpretability are set at the same level separated by symbol "/" (see Table 2). From a quantitative point of view the ranges of values are completely different among all the indices. Each individual value only makes sense in comparison with the other values obtained by the same index. From a qualitative point of view we would rather choose those indices yielding a ranking without ambiguities (Total rule length, Nauck's index, and Fuzzy index), i.e., those indices able to produce a full order distinguishing among all pairs of KBs. As expected, we have achieved five different rankings because each interpretability index follows different criteria. Nevertheless, looking carefully it is easy to appreciate that all rankings are somehow similar, the same KBs usually appear at the beginning (KB2, KB4, KB6, KB8) or at the end (KB3, KB7).

However, a last question still remains to be answered: How to know which index is the best one? Since the measure of interpretability is clearly subjective the only way to answer this question is asking people. For that reason, a web poll was addressed to FL experts (50%) as well as people who are not familiar with FL (50%). The study is made regarding the

[1] http://www.ics.uci.edu/-mlearn/MLSummary.html

ISBN: 978-989-95079-6-8

twelve KBs generated for the WINE problem. The goal is to compare the most popular interpretability criteria, including people used (and not used) to work with fuzzy systems that can be (or not) fond of wines. Since interpretability extremely depends on the kind of user, let us add a short comment. In the context of FM, there are three kinds of users:

- The *final user* of the modeled system. In most cases, he/she will interact with the system providing data and/or receiving system suggestions and advices for making decisions. The user will only trust on the system if the system output is coherent according to his/her background. Notice that the use of a comprehensible model can help the final user to understand the system output.
- The *system designer* who has to be an expert on fuzzy logic in order to produce a good model useful for the final user of the application. A transparent (gray-box) model structure is really appreciated for the future model maintenance and update.
- The *domain expert* who will explain the system behavior to the system designer during the model design stage. In addition, he/she will be in charge of validating system running. Since domain experts usually do not know anything about fuzzy logic a clearly readable model description is required to make easier the validation stage.

In our study, FL experts are mainly considered as system designers but due to the nature of the problem they also can act as domain experts and even as final users. In turn, non-FL users are only viewed as domain experts or final users. Twenty six answers were collected. They show a huge diversity what clearly illustrates how different users have very different criteria to measure interpretability. Three main questions were asked as part of the poll:

1. *How much interpretable are the twelve KBs?* Each user was asked to give an interpretability measure for each KB. Such measure was represented by an interval (min-max), i.e., the range in which it should be included, between zero and one hundred. However, only a few users were willing to answer to this question with numerical values. In fact, we realize that people find much more natural to make approximate reasoning based on the use of linguistic terms like *Highly interpretable*, *Moderately interpretable*, etc. In addition collected values show a huge variance. In consequence, it does not make sense drawing statistical conclusions from stored data. According to these results it can be argued that people get into difficulties when they have to give numerical indices as computers usually do.
2. *What is the KB interpretability ranking?* Users were asked to rank the KBs according to their preferences from the interpretability point of view: One for the most interpretable KB, and twelve for the least interpretable one. Since all users were willing to answer this question, an interesting conclusion can be drawn: *People feel much more confident setting rankings than giving numerical values*. In order to set a ranking it is necessary to compare all the KBs (by couples) but it does not imply setting individual measures. First column of Table 3 includes the user identifier, setting in brackets if the user is used to work with fuzzy systems (F) or not (NF). Second column of the table shows rankings given by users. As it can be seen at first glance there is a huge variance. Only two couples of users (1-26 and 4-11) gave exactly the same order. Nevertheless, looking carefully answers are not so different. The global order is more or less the same for all users but when two KBs are quite close regarding interpretability the final ranking choice depends in many subtle details, and as a result, there is a clearly subjective choice at the end.

   The comparison between rankings provided by the users (Table 3) and rankings derived from the computed interpretability indices (Table 2) lets us evaluate the goodness of such indices. However, only user 3 and fuzzy index yield the same ranking. In order to make a deeper analysis, we have computed Euclidean distance from each of the five interpretability indices, x (first line of the table), to all the twenty six users, y, according to equation 1 where x_i means the ranking position of KB_i regarding index x and y_i is the ranking position of KB_i regarding user y.

$$d_{x,y} = \sqrt{\sum_{i=1}^{12} |x_i - y_i|} \tag{1}$$

   Computed distances give an idea on how different (comparing positions of each KB in selected rankings) indices and user's answers are. Most user's answers are closer to the rankings obtained using *Number of rules*. There are also many answers closer to *Total rule length* and *Fuzzy index*. In fact, it is possible to identify several groups of users (look at Table 4) whose answers fit better with some indices but there is a lot of overlapping among groups. Moreover, approximately the same number of F and NF users belongs to all the groups. This is due to the fact that in general none of the computed indices fit properly in user's answers.

3. *What are the most relevant aspects to consider when assessing interpretability?* Each user was asked to give short comments explaining what he/she considers good strategies and/or key criteria to measure interpretability. Some of the most useful comments are listed below:
   - *A common heuristic reasoning is the following. First look at the total number of rules. Second, if there is ambiguity between some of the knowledge bases it is needed to check the total number of premises. Then, if there is still ambiguity it is necessary to analyze the complexity of the linguistic terms*. This suggests making the ranking in different abstraction levels, adding new criteria only when they are needed.
   - *I prefer short rules considering at most 5 features than fewer rules with a long size*. This remarks that the number of inputs by rule is a main criterion. However, different people have different views about what must be considered as a small

ISBN: 978-989-95079-6-8

Table 3: Ranking of KBs extracted from the poll results.

User	+ Interpretability -
user1 (F)	KB8, KB2, KB6, KB12, KB10, KB4, KB9, KB1, KB5, KB11, KB7, KB3
user2 (F)	KB6, KB2, KB4, KB1, KB8, KB5/KB9, KB10, KB12, KB11, KB7, KB3
user3 (F)	KB8, KB2, KB6, KB4, KB12, KB9, KB10, KB1, KB5, KB11, KB7, KB3
user4 (NF)	KB2, KB6, KB8, KB12, KB9, KB10, KB4, KB1, KB5, KB11, KB7, KB3
user5 (F)	KB2, KB6, KB8, KB9, KB10/KB12, KB4, KB1, KB5, KB11, KB7, KB3
user6 (F)	KB8/KB12, KB2/KB6, KB4/KB9/KB10, KB1/KB5, KB11, KB3, KB7
user7 (F)	KB8, KB6, KB2, KB12, KB10, KB9, KB4, KB1, KB5, KB3, KB11, KB7
user8 (F)	KB8, KB2, KB6, KB12, KB9, KB4, KB10, KB1, KB5, KB11, KB7, KB3
user9 (NF)	KB2, KB9, KB12, KB8, KB6, KB10, KB5, KB4, KB1, KB11, KB3, KB7
user10 (NF)	KB6, KB2, KB9, KB12, KB4, KB8, KB5, KB1, KB11, KB10, KB7, KB3
user11 (NF)	KB2, KB6, KB8, KB12, KB9, KB10, KB4, KB1, KB5, KB11, KB7, KB3
user12 (NF)	KB8, KB12, KB2, KB6, KB9, KB4, KB10, KB5, KB1, KB11, KB7, KB3
user13 (F)	KB2, KB6, KB8, KB9, KB12, KB4/KB10, KB1/KB5/KB11, KB3/KB7
user14 (NF)	KB8, KB2/KB6, KB12, KB9, KB4, KB10, KB1, KB5, KB11, KB7, KB3
user15 (NF)	KB8, KB2, KB12, KB6, KB10, KB1, KB5, KB9, KB4, KB11, KB7, KB3
user16 (NF)	KB8, KB6, KB2, KB12, KB10, KB4, KB9, KB5, KB1, KB11, KB7, KB3
user17 (NF)	KB8, KB6, KB2, KB12, KB4, KB9, KB5, KB1, KB11, KB10, KB3, KB7
user18 (NF)	KB2/KB12, KB4/KB6/KB8/KB10, KB5/KB9, KB11, KB1, KB7, KB3
user19 (NF)	KB2, KB4, KB6, KB11, KB5, KB1, KB9, KB12, KB10, KB8, KB7, KB3
user20 (F)	KB2, KB8, KB9, KB12, KB4, KB6, KB10, KB11, KB1, KB5, KB7, KB3
user21 (NF)	KB2, KB6, KB8, KB12, KB9, KB4, KB10, KB11, KB5, KB1, KB7, KB3
user22 (F)	KB8, KB6, KB2, KB9, KB12, KB10, KB4, KB5, KB11, KB1, KB3, KB7
user23 (NF)	KB8, KB2, KB6, KB12, KB10, KB9, KB4, KB1, KB5, KB11, KB7, KB3
user24 (F)	KB2, KB8, KB6, KB12, KB4, KB9, KB1, KB10, KB5, KB11, KB7, KB3
user25 (F)	KB8, KB2/KB4/KB6/KB9/KB10/KB12, KB1/KB5/KB11, KB3/KB7
user26 (F)	KB8, KB2, KB6, KB12, KB10, KB4, KB9, KB1, KB5, KB11, KB7, KB3

Table 4: Groups of users regarding computed interpretability indices.

Index	Users	F	NF
Number of rules	4, 5, 7, 9, 11, 13, 15, 16, 22, 23, 25	5	6
Total rule length	1, 8, 12, 14, 18, 21, 25, 26	4	4
Average rule length	2, 19	1	1
Nauck's index	6, 12, 25	2	1
Fuzzy index	3, 8, 10, 14, 17, 20, 21, 24, 25	5	4

number of inputs by rule. This problem arises from the intrinsic ambiguity of natural language: What means small? The same word has different meanings in different contexts, but even in the same context it has different meanings for different people. The use of fuzzy logic formalizes a precise meaning for each word coping with this kind of ambiguity.

- *With respect to words (linguistic variables and terms), the better choice of words within the context of the problem, the more accurate interpretation*. Understanding strongly depends on the context of the problem. For instance, it is easy to see how different the meaning of *High* is when talking about people, buildings, or mountains.

3 User profiles

With the aim of finding out some user profiles from Table 3 we have applied a hierarchical clustering analysis [20]. Two dendrograms were built using Ward's method [21] and squared Euclidean distance (see Fig. 1): The first one (on the top part of the figure) only regarding fuzzy expert users where two groups (SF1 and SF2) are clearly identified. The second one (on the bottom part of the figure) including only non-fuzzy users where only one group (SNF1) can be defined. The rest of users are progressively added by the clustering algorithm but they stay at long distance.

It is possible to extract a prototype user profile from each group. According to our experience designing and assessing interpretable fuzzy systems, and keeping in mind the conclusions derived from the web poll, ten variables were selected as tentative interpretability indicators: (1) Number of rules; (2) Total rule length; (3) Percentage of rules which use less than ten percent of inputs; (4) Percentage of rules which use between ten and thirty percent of inputs; (5) Percentage of rules which use more than thirty percent of inputs; (6) Number of inputs; (7) Number of labels used in the rule base; (8) Percentage of elementary labels used in the rule base; (9) Percentage of OR composite labels used in the rule base; (10) Percentage of NOT composite labels used in the rule base.

The task consists of discovering those indicators that can be considered as key to distinguish among groups. From the rankings provided by users it can be induced the order relation between KBs. Assuming that each ranking is based on a comparison per couples of all KBs, Table 3 is translated into a data set with the following format. Each column give the difference between the ten selected indicators (listed above) for each couple of KBs (A and B). The last column includes one

ISBN: 978-989-95079-6-8

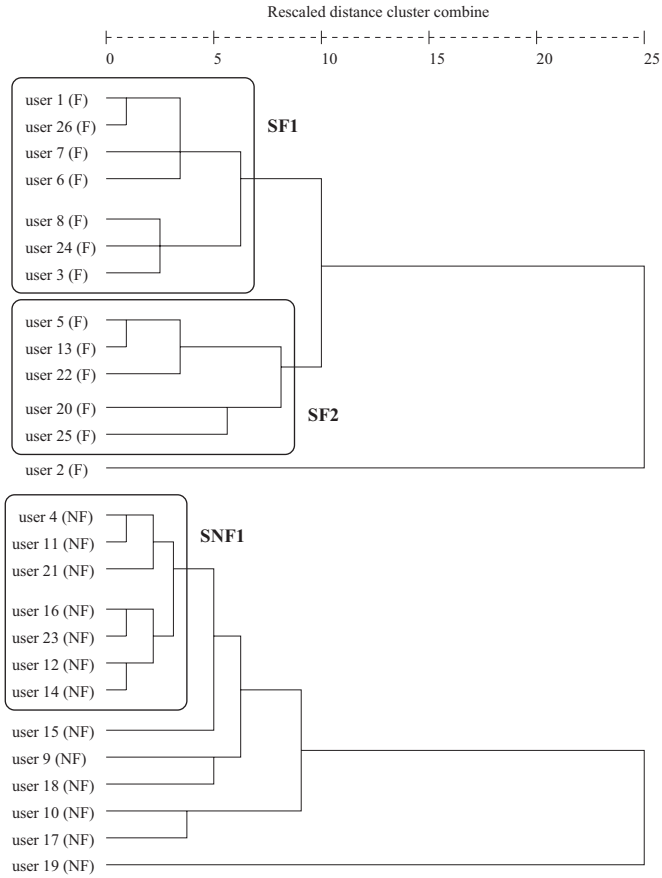


Figure 1: Groups of users by hierarchical clustering.

of three labels (1:*No*, 2:*Yes*, 3:*I don't know*) answering to the question: is A more interpretable than B? The whole data set was divided into three subsets including only the comparisons related to the users belonging to each group.

Using HILK methodology [16] (FDT-S method) and the previously generated data subsets we have built three fuzzy classifiers, one for each group of users. For each couple of KBs (A and B) the fuzzy classifiers give as output (1) *A is more interpretable than B*, (2) *B is more interpretable than A*, or (3) *I don't know*, what means that A and B are similar from an interpretability point of view. The comparison among all KBs yields a ranking for each group. Figure 2 presents obtained rankings as well as the interpretability indicators (inputs of the fuzzy classifier) that define the six prototype user profiles we were looking for. Three indicators are not selected by any of the prototypes: (3) Percentage of rules which use less than ten percent of inputs; (5) Percentage of rules which use more than thirty percent of inputs; and (9) Percentage of OR composite labels used in the rule base. It seems that all users agree that a small percentage of inputs by rule is good for interpretability and a large percentage is bad, while composite propositions including OR are seen as easily readable. Notice that each prototype is defined using at least three indicators, i.e., the use of basic indices is not enough.

In addition, for each group we have computed the distance (mean and variance) between the prototype and all users included in the group. SF1 represents a quite compact group where the prototype is defined using only three indicators, while SF2 and SNF1 are less compact and because of that they need more indicators. Although clustering of F and NF users should intuitively yield more homogeneous groups, in practice there is still a lot of diversity inside each group. In the case of F users, prototypes achieve medium mean distance with small variance. SF2 takes into account six indicators because it is made up of only five users quite heterogeneous. It results specially interesting the fact that *Total rule length* which seemed to be a very important basic indicator according to Table 4 only is taken into account by group SF2. It is not nearly the most relevant indicator in comparison with the subsets of indicators emerging from the clustering analysis. Finally, it could be argued that fuzzy users (SF1 and SF2) give more homogeneous answers but regarding more complex criteria than non-fuzzy users (SNF1). Finally, as a result of the heterogeneity of NF users only seven of them give more or less similar rankings. In consequence, four indicators were relevant for this group but yielding large mean and variance.

4 Conclusions

Assessing interpretability is a very challenging and complex task due to the inherent subjectivity of the measure. In order to evaluate existing indices we have set up a first experimental study, for simplicity limited to twelve rule bases assuring most interpretability constraints described as essential in the literature. As a result, assuming knowledge bases under study are interpretable the study focus on quantifying interpretability and comparing obtained results with assessment provided by people in a web poll. None of the evaluated indices gave good results in comparison with rankings provided by human beings. A lot of work remains still to be done so that finding a universal index. However, results derived from our experimental study offer some interesting clues.

First, it is necessary to define a new index flexible enough to be easily adaptable to the problem context and user preferences. Such index must take into account many subtle details combined in the context of computing with words and perceptions based on fuzzy logic technology. In addition, obtaining a numerical value is not needed in most applications where the important thing is to set an appropriate ranking.

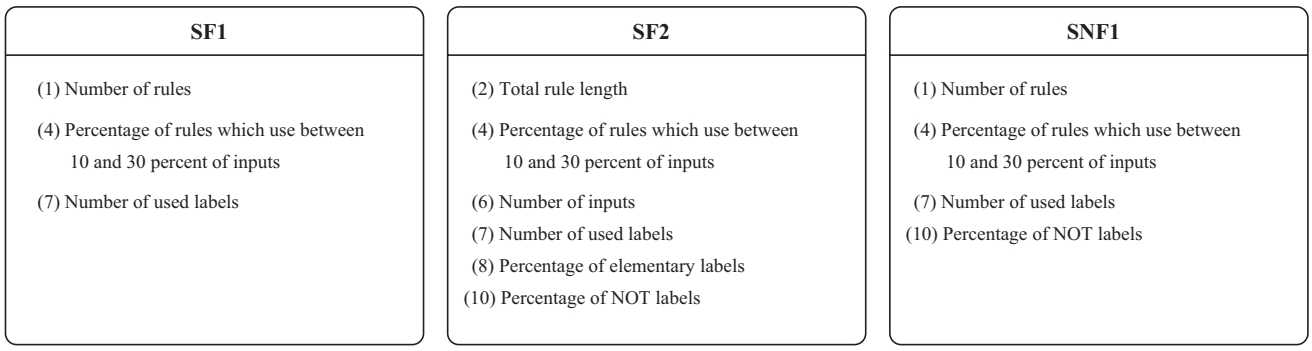
Second, a hierarchical fuzzy framework has been proved as a powerful tool to imitate the usual way of people reasoning. It mainly consists of taking a few interpretability indicators as guide to discriminate between two knowledge bases, adding more criteria only when it is necessary because the compared knowledge bases are not distinguishable at first glance.

Finally, we have focused on interpretability from a structural point of view but there are many cognitive aspects that should be analyzed. Therefore more experimental studies are needed. Obviously, as a first step our study has been limited to a very specific kind of fuzzy rule-base systems. Of course, in the future it would be interesting to make a comparison of different rule base structures.

Acknowledgment

Authors would like to acknowledge people who have filled up the web poll. Their suggestions were very interesting and they have contributed so much to this work.

This work has been partially funded by the Foundation for the Advancement of Soft Computing (Asturias) and Spanish government (CICYT) under grant: TIN2008-06890-C02-01.



	Ranking	Distance	
		Mean	Variance
SF1	KB8, KB2, KB6, KB12, KB4, KB9/KB10, KB1, KB5, KB11, KB7, KB3	2.616	0.597
SF2	KB2/KB8, KB6, KB9/KB12, KB4, KB10, KB1, KB5, KB11, KB3, KB7	3.842	0.551
SNF1	KB8, KB2, KB6, KB12, KB9, KB10, KB4, KB1, KB5, KB11, KB7, KB3	2.789	3.086

Figure 2: Prototype user profiles.

References

[1] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Parts I, II, and III. Information Sciences*, 8, 8, 9:199–249, 301–357, 43–80, 1975.

[3] E. H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic systems. *IEEE Transactions on Computers*, 26(12):1182–1191, 1977.

[4] J. Espinosa and J. Vandewalle. Constructing fuzzy models with linguistic integrity from numerical data-afreli algorithm. *IEEE Transactions on Fuzzy Systems*, 8(5):591–600, 2000.

[5] S. Guillaume and B. Charnomordic. Generating an interpretable family of fuzzy partitions. *IEEE Transactions on Fuzzy Systems*, 12(3):324–335, 2004.

[6] C. Mencar and A. M. Fanelli. Interpretability constraints for fuzzy information granulation. *Information Sciences*, 178(24):4585–4618, 2008.

[7] S.-M. Zhou and J. Q. Gan. Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling. *Fuzzy Sets and Systems*, 159(23):3091–3131, 2008.

[8] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index. *Soft Computing*, 13(5):437–449, 2009.

[9] P. Fazendeiro and J. V. de Oliveira. A working hypothesis on the semantics/accuracy synergy. In *Joint EUSFLAT-LFA*, pages 266–271, Barcelona, Spain, September, 7-9, 2005.

[10] Y. Jin, W. von Seelen, and B. Sendhoff. On generating FC^3 fuzzy rule systems from data using evolution strategies. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(6):829–845, 1999.

[11] C. Mencar, G. Castellano, and A. M. Fanelli. Distinguishability quantification of fuzzy sets. *Information Sciences*, 177(1):130–149, 2007.

[12] M. Setnes, R. Babuska, U. Kaymak, and H. R. van Nauta Lemke. Similarity measures in fuzzy rule base simplification. *IEEE Transactions on Systems, Man, and Cybernetics. Part B*, 28(3):376–386, 1998.

[13] H. Ishibuchi and Y. Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *International Journal of Approximate Reasoning*, 44(1):4–31, 2007.

[14] D. D. Nauck. Measuring interpretability in rule-based classification systems. In *FUZZ-IEEE 2003*, volume 1, pages 196–201, St. Louis, Missouri, USA, May, 25-28, 2003.

[15] J. M. Alonso, S. Guillaume, and L. Magdalena. A hierarchical fuzzy system for assessing interpretability of linguistic knowledge bases in classification problems. In *IPMU06. Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 348–355, Paris, France, July, 2-7, 2006.

[16] J. M. Alonso, L. Magdalena, and S. Guillaume. HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *International Journal of Intelligent Systems*, 23(7):761–794, 2008.

[17] L.-X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*, 22(6):1414–1427, 1992.

[18] H. Ichihashi, T. Shirai, K. Nagasaka, and T. Miyoshi. Neuro-fuzzy ID3: A method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning. *Fuzzy Sets and Systems*, 81(1):157–167, 1996.

[19] J. M. Alonso, S. Guillaume, and L. Magdalena. KBCT: A knowledge management tool for fuzzy inference systems. *Free software under GPL license, available in http://www.mat.upm.es/projects/advocate/kbct.htm*, 2003.

[20] S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

[21] J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

ISBN: 978-989-95079-6-8

Games and capacities on partitions

Michel Grabisch
Centre d'Economie de la Sorbonne, Université Paris I 106-112, Bd de l'Hôpital, 75013 Paris, France
Email: michel.grabisch@univ-paris1.fr

Abstract— We consider capacities and games whose value $v(S)$ depend on the organization of $N \setminus S$, this organization being represented by a partition. Hence we deal with capacities and games depending on two arguments: a subset (coalition) S and a partition π containing S. We call embedded coalition any such pair (S, π). We first provide a suitable structure for the set of embedded coalitions, study its properties, and find the Möbius function for this structure. Then, we define games and capacities on this structure and study their properties. We propose also a concept similar to the Shapley value, and provide an axiomatization of it. Lastly, we give some insights on the Choquet integral.

Keywords— capacity, fuzzy measure, partition, Möbius transform, Shapley value

1 Introduction

Capacities [1] or fuzzy measures [2] are usually defined on the set of subsets of the universe, assumed to be finite in this paper, or on a subcollection of it. Specifically, denoting by N the universe, let us take a capacity v defined on 2^N, and consider some subset $S \in 2^N$ together with the quantity $v(S)$. The interpretation of $v(S)$ may differ according to the context, but usually falls into two categories: either $v(S)$ represents a degree of certainty, plausibility, etc. that the true (but unknown) state of the world is contained in S, or $v(S)$ represents some power, importance, strength, monetary value of the group S of entities (agents, players, voters, criteria, etc.). Using a common word, we may say that $v(S)$ is the *worth* of S.

Let us consider more carefully the second interpretation, which is related to cooperative game theory. If N represents a society of individuals, speaking of $v(S)$ implicitly means that the society is formed of the group (or coalition) S and the coalition $N \setminus S$. Under this organization, the worth of S is $v(S)$. We may however consider less simple situations, where the set of remaining agents $N \setminus S$ is also organized in some way, say $N \setminus S = S_2 \cup S_3 \cup \cdots \cup S_k$, where $\{S_2, S_3, \ldots, S_k\}$ is a partition of $N \setminus S$. Then it is natural to think that the worth of S may depend on the organization of the remaining agents. Therefore, the expression $v(S)$ is no more enough precise, and we need to consider $v(S, \{S, S_2, S_3, \ldots, S_k\})$, for all possible organizations of the society containing S as a block of the society. The usual term for the argument of v, namely (S, π) with π the partition $\{S, S_2, S_3, \ldots, S_k\}$, is *embedded coalition*, and v is called a *game in partition function form*, or *PFF-game* for short [3]. We will call them in this paper games (or capacities) on partitions.

Although games on partitions have been studied in the game theoretic literature (essntially the Shapley value, the core), no explicit study has been done on the mathematical object (S, π) and its structure, nor on the properties of v. The aim of this paper is precisely to fill this gap. We will provide a structure for embedded coalitions, study its properties, and find the Möbius function on this structure. Then we will study properties of games, and define a Shapley value for v, and finally give some remarks on the possibility to define a Choquet-like integral. A part of this paper is based on the working paper [4].

2 The structure of embedded coalitions

Let $N := \{1, 2, \ldots, n\}$ be the universal set (set of agents, players, etc.). We denote by $S, T, \ldots$ the subsets of N (coalitions), and their cardinality by $s, t, \ldots$. We consider the set $\Pi(N)$ (denoted also by $\Pi(n)$) of all possible partitions of N (coverings of N by disjoint subsets). For a partition $\pi := \{S_1, \ldots, S_k\}$, subsets $S_1, \ldots, S_k$ are called *blocks* of π. A partition in k blocks is called a *k-partition*. A natural ordering of partitions is given by the notion of "coarsening" or "refinement". Taking π, π' partitions in $\Pi(N)$, we say that π is a *refinement* of π' (or π' is a *coarsening* of π), denoted by $\pi \leq \pi'$, if any block of π is contained in a block of π' (or every block of π' fully decomposes into blocks of π). Then $(\Pi(N), \leq)$ is a lattice, called the *partition lattice*. With this ordering, the bottom element of the lattice is the finest partition $\pi^\perp := \{\{1\}, \ldots, \{n\}\}$, while the top element is the coarsest partition $\pi^\top := \{N\}$.

An *embedded coalition* is a pair (S, π), where $S \in 2^N \setminus \{\emptyset\}$, and $\pi \in \Pi(N)$ is such that $S \in \pi$. We denote by $\mathfrak{C}(N)$ (or by $\mathfrak{C}(n)$) the set of embedded coalitions on N. For the sake of concision, we often denote by $S\pi$ the embedded coalition (S, π), and omit braces and commas for subsets (example with $n = 3$: $12\{12, 3\}$ instead of $(\{1, 2\}, \{\{1, 2\}, \{3\}\})$). Remark that $\mathfrak{C}(N)$ is a proper subset of $2^N \times \Pi(N)$.

As mentionned in the introduction, works on games on partitions do not explicitly define a structure (that is, some order) on embedded coalitions. A natural choice is to take the product order on $2^N \times \Pi(N)$:

$$(S, \pi) \sqsubseteq (S', \pi') \Leftrightarrow S \subseteq S' \text{ and } \pi \leq \pi'.$$

Evidently, the top element of this ordered set is $(N, \pi^\top)$ (denoted more simply by $N\{N\}$ according to our conventions). However, due to the fact that the empty set is not allowed in (S, π), there is no bottom element in the ordered structure $(\mathfrak{C}(N), \sqsubseteq)$. Indeed, all elements of the form $(\{i\}, \pi^\perp)$ are minimal elements. For mathematical convenience, we introduce an artificial bottom element $\perp$ to $\mathfrak{C}(N)$ (it could be considered as $(\emptyset, \pi^\perp)$), and denote $\mathfrak{C}(N)_\perp := \mathfrak{C}(N) \cup \{\perp\}$. We give as illustration the partially ordered set $(\mathfrak{C}(N)_\perp, \sqsubseteq)$ with $n = 3$ (Fig. 1).

ISBN: 978-989-95079-6-8

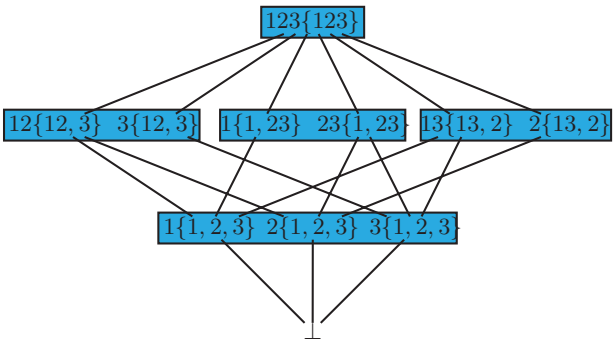


Figure 1: Hasse diagram of $(\mathfrak{C}(N)_\perp, \sqsubseteq)$ with $n = 3$. Elements with the same partition are framed in grey.

Another possibility to define an order on embedded coalitions is to basically take the Boolean lattice 2^N, and for any $S \in 2^N$, duplicate it into $S\pi_1, S\pi_2, \ldots$ for all partitions π containing S. Then $(S, \pi) \leq (S', \pi')$ if and only if $S \subseteq S'$. This seems to be the underlying structure in the work of Macho-Stadler et al. [5]. We illustrate this structure for $n = 3$ in Figure 2.

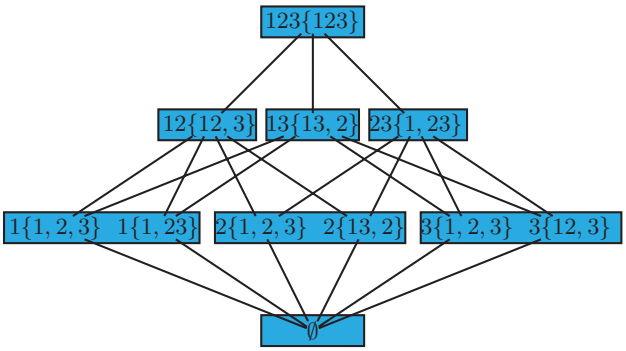


Figure 2: Hasse diagram of the structure of Macho-Stadler et al. for $n = 3$. Elements with same coalition (duplicates) are framed in grey.

In the sequel, we focus on the first structure, since it has better properties. The second one is not a lattice, since for example, $12\{12, 3\}$ and $13\{13, 2\}$ are two minimal upper bounds of $1\{1, 2, 3\}$ and $1\{1, 23\}$.

We recall that in a partially ordered set $(P, \leq)$ with a bottom element $\perp$ and a top element $\top$, a *chain from* $\perp$ *to* $\top$ is a totally ordered sequence of elements of P including $\perp, \top$. The chain is *maximal* if no other chain can contain it (equivalently, if between two consecutive elements x_i, x_{i+1} of the sequence, there is no element $x \in P$ such that $x_i < x < x_{i+1}$). If no ambiguity occurs, we say *maximal chain* instead of *maximal chain from* $\perp$ *to* $\top$. The set of maximal chains in P is denoted by $\mathcal{C}(P)$. The *length* of a maximal chain is the number of elements of the sequence minus 1.

The following is shown in [4].

Proposition 1. The following holds.

(i) $(\mathfrak{C}(N)_\perp, \sqsubseteq)$ is an upper semimodular lattice, whose top and bottom elements are $(N, \pi^\top)$ and $\perp$. All elements are complemented: for a given $S\pi$, any embedded coalition of the form $\overline{S}\pi_{\overline{S}}$ with $\overline{S}$ the complement of S and $\pi_{\overline{S}}$ any partition containing $\overline{S}$.

(ii) Every maximal chain from $\perp$ to an element (S, π) has the same length, which is $n - k + 1$, if π is a partition in k blocks. Hence, the height of the lattice is n.

(iii) The total number of elements of $\mathfrak{C}(N)_\perp$ is $\sum_{k=1}^{n} kS_{n,k} + 1$, where $S_{n,k}$ is the Stirling number of second kind.

n	1	2	3	4	5	6	7	8
$\lvert\mathfrak{C}(n)_\perp\rvert$	2	4	11	38	152	675	3264	17008

(iv) The number of maximal chains from $\perp$ to $N\{N\}$ is $\frac{(n!)^2}{2^{n-1}}$, which is also the number of maximal longest chains in $\mathfrak{C}(N)$.

n	1	2	3	4	5	6	7	8
$\lvert\mathcal{C}(\mathfrak{C}(n)_\perp)\rvert$	1	2	9	72	900	16 200	396 900	12 700 800

(v) Let (S, π) be an embedded coalition, with $\pi := \{S, S_2 \ldots, S_k\}$, and $|S| = s$. The number of maximal chains from $\perp$ to (S, π) is

$$|\mathcal{C}([\perp, (S, \pi)])| = \frac{s(n-k)!}{2^{n-k}} s!s_2! \cdots s_k!.$$

For example, there are 9 chains from $\perp$ to $123\{123, 4\}$.

(vi) Let (S, π) be an embedded coalition, with $\pi := \{S, S_2 \ldots, S_k\}$. The number of maximal chains from (S, π) to $N\{N\}$ is

$$|\mathcal{C}([(S, \pi), N\{N\}])| = \frac{1}{k}|\mathcal{C}(\mathfrak{C}(k)_\perp)| = \frac{k!(k-1)!}{2^{k-1}}.$$

The above results focus on maximal chains, since they are a fundamental concept for the sequel, especially for the Shapley value. It is important to note that this lattice is neither distributive, modular nor atomic (and hence not geometric). Atoms are all elements of the form $\{i\}\pi^\perp$, and elements representable by a join of atoms are only those of the form $S\pi^\perp_S$, where $\pi^\perp_S = \{S, i_1, \ldots, i_{n-s}\}$, with $N \setminus S = \{i_1, \ldots, i_{n-s}\}$.

3 Games and capacities on partitions

Definition 1. A *game on partitions* on N is a mapping $v : \mathfrak{C}(N)_\perp \to \mathbb{R}$, such that $v(\perp) = 0$. The set of all games on partitions on N is denoted by $\mathcal{PG}(N)$.

Although any such function could be called a game on partitions, to be meaningful, we need to assume that forming the grand coalition generates the largest total surplus, i.e., $v(N\{N\}) \geq \sum_{S \in \pi} v(S, \pi)$, for all $\pi \in \Pi(N)$.

The structure of lattice permits to define the usual notions on games and capacities.

ISBN: 978-989-95079-6-8

Definition 2. Let $v \in \mathcal{PG}(N)$.

(i) v is *monotone* if $S\pi \sqsubseteq S'\pi'$ implies $v(S\pi) \leq v(S'\pi')$. A monotone game on partitions is called a *capacity on partitions*. A capacity on partitions v is normalized if $v(N\{N\}) = 1$.

(ii) v is *supermodular* if for every $S\pi, S'\pi'$ we have

$$v(S\pi \vee S'\pi') + v(S\pi \wedge S'\pi') \geq v(S\pi) + v(S'\pi').$$

It is *submodular* if the reverse inequality holds.

(iii) A game is *additive* if it is both supermodular and submodular.

(iv) More generally, for a given $k \geq 2$, a game is *k-monotone* if for all families of k elements $S_1\pi_1, \ldots, S_k\pi_k$, we have

$$v(\bigvee_{i\in K} S_i\pi_i) \geq \sum_{J\subseteq K, J\neq\emptyset} (-1)^{|J+1|} v(\bigwedge_{i\in J} S_i\pi_i)$$

putting $K := \{1, \ldots, k\}$. A game is ∞-monotone if it is k-monotone for every $k \geq 2$.

(v) A game is a belief function if it is a normalized ∞-monotone capacity.

Above, $\vee, \wedge$ are the least upper bounds and greatest lower bounds. The greatest lower bound is simply given by:

$$(S, \pi) \wedge (S', \pi') = (S \cap S', \pi \wedge \pi') \text{ if } S \cap S' \neq \emptyset,$$

and $\perp$ otherwise. The least upper bound is more tricky to define:

$$(S, \pi) \vee (S', \pi') = (T \cup T', \rho)$$

where T, T' are blocks of $\pi \vee \pi'$ containing respectively S and S', and ρ is the partition obtained by merging T and T' in $\pi \vee \pi'$.

The following result is due to Barthélemy [6].

Proposition 2. Let L be a lattice. Then f is monotone and ∞-monotone on L if and only if it is monotone and $(|L| - 2)$-monotone.

In lattice theory, a *valuation* (or *2-valuation*) on a lattice L is a real-valued function on L being both super- and submodular (additive). More generally, for a given $k \geq 2$, a *k-valuation* satisfies

$$v(\bigvee_{i\in K} x_i) = \sum_{J\subseteq K, J\neq\emptyset} (-1)^{|J+1|} v(\bigwedge_{i\in J} x_i)$$

for every family of k elements. An ∞-valuation is a function f which is a k-valuation for every $k \geq 2$. The following results are useful (see [7, Ch. X], and also [6]).

Proposition 3. Let L be a lattice.

(i) L is modular if and only if it admits a strictly monotone valuation v (i.e., $x < y$ implies $v(x) < v(y)$).

(ii) L is distributive if and only if it admits a strictly monotone 3-valuation.

(iii) L is distributive if and only if it is modular and every strictly monotone valuation is a k-valuation for any $k \geq 2$.

(iv) Any lattice admits an ∞-valuation.

The consequence of (i) is that no strictly monotone additive game exists since $\mathfrak{C}(n)_\perp$ is not modular when $n > 2$. The question is: does it exist an additive game? The following proposition answers this question.

Proposition 4. $\mathfrak{C}(2)_\perp$ admits a strictly monotone 2-valuation (hence by Prop. 2, a strictly monotone ∞-monotone valuation). If $n > 2$, the only possible valuations on $\mathfrak{C}(n)_\perp$ are constant valuations.

Since any game v satisfies $v(\perp) = 0$, we have:

Corollary 1. The only additive game is the constant game $v = 0$.

Definition 3. Let $T\sigma \in \mathfrak{C}(N)$.

(i) The *unanimity game centered on $T\sigma$* is the game defined by

$$u_{T\sigma}(S\pi) = \begin{cases} 1, & \text{if } S\pi \sqsupseteq T\sigma \\ 0, & \text{otherwise.} \end{cases}$$

(ii) The *identity game centered on $T\sigma$* is the game defined by

$$\delta_{T\sigma}(S\pi) = \begin{cases} 1, & \text{if } S\pi = T\sigma \\ 0, & \text{otherwise.} \end{cases}$$

Unanimity games are normalized capacities while the latter ones are not. Identity games form a basis of the vector space $\mathcal{PG}(N)$ since for any $v \in \mathcal{PG}(N)$,

$$v = \sum_{T\sigma\in\mathfrak{C}(N)} v(T\sigma)\delta_{T\sigma}.$$

From Proposition 1 (iii), the dimension of the basis is $\sum_{k=1}^{n} kS_{n,k}$. A less trivial basis is given by unanimity games, through the Möbius transform, described in the next paragraph.

In capacity theory and partially ordered sets, a key point is to determine the Möbius function, enabling to compute the Möbius transform of games and capacities. We recall that for a partially ordered set $(P, \leq)$ which is locally finite, the Möbius function $\mu(x, y) : P^2 \to \mathbb{R}$ permits to solve the equation:

$$g(x) = \sum_{y\leq x} f(y)$$

where $f, g : P \to \mathbb{R}$, as follows:

$$f(x) = \sum_{y\leq x} g(y)\mu(y, x).$$

(see Rota [8], Aigner [9]). Based on properties of $\mathfrak{C}(N)_\perp$, the following fundamental result:

ISBN: 978-989-95079-6-8

Proposition 5. If P is a lattice with bottom element 0 and set of atoms $\mathcal{A}$, for every $x \in P$, the Möbius function reads

$$\mu(0,x) = \sum_{S \subseteq \mathcal{A} | \bigvee S = x} (-1)^{|S|},$$

and the fact that for the partition lattice we know that $\mu_{\Pi(n)}(\pi^{\perp}, \{N\}) = (-1)^{n-1}(n-1)!$ (see Aigner [9, p. 154]), the following can be shown.

Proposition 6. The Möbius function on $\mathfrak{C}(n)_{\perp}$ is given by:

$$\mu(\perp, S\pi) = \begin{cases} (-1)^{|S|}, & \text{if } \pi = \pi_S^{\perp} \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

$$\mu(S'\pi', S\pi) = (-1)^{k'-k}(l_1 - 1)! \cdots (l_k - 1)!, \quad (2)$$

for every $S'\pi' \leq S\pi$, where $\pi_S^{\perp}$ is the partition consisting in S and all singletons in $N \setminus S$, $S\pi = S\{S, S_2, \ldots, S_k\}$, $S'\pi' = S'\{S', S_{12}, \ldots, S_{1l_1}, S_{21}, \ldots, S_{2l_2}, \ldots, S_{k1}, \ldots, S_{kl_k}\}$, $S_i = S_{i1} \cup \cdots \cup S_{il_i}$ for $i = 2, \ldots, k$, and $k' := \sum_{i=1}^{k} l_i$.

Using this, the Möbius transform of any PFF-game v is given by

$$m(S\pi) = \sum_{S'\pi' \sqsubseteq S\pi} \mu(S'\pi', S\pi) v(S'\pi').$$

Now, $m(S\pi)$ gives the coordinates of v in the basis of unanimity games.

Example 1. Application for $n = 3$. Let us compute the Möbius transform of a game v. We have, for all distinct $i, j, k \in \{1, 2, 3\}$:

$$\begin{aligned} m(\perp) &= 0 \\ m(i\pi^{\perp}) &= v(i\pi^{\perp}) \\ m(ij\{ij,k\}) &= v(ij\{ij,k\}) - v(i\pi^{\perp}) - v(j\pi^{\perp}) \\ m(i\{i,jk\}) &= v(i\{i,jk\}) - v(i\pi^{\perp}) \\ m(123\{123\}) &= v(123\{123\}) - \sum_{i,j} v(ij\{ij,k\}) \\ &\quad - \sum_{i} v(i\{i,jk\}) + 2\sum_{i} v(i\pi^{\perp}). \end{aligned}$$

Barthélemy [6] proved the following.

Proposition 7. Let L be any lattice, and $f : L \to \mathbb{R}$. If the Möbius transform of f, denoted by m, satisfies $m(\perp) = 0$, $m(x) \geq 0$ for all $x \in L$, and $\sum_{x \in L} m(x) = 1$ (normalization), then f is a belief function.

The converse of this proposition does not hold in general (it holds for the Boolean lattice 2^N). A belief function is *invertible* if its Möbius transform is nonnegative, normalized and vanishes at $\perp$. The following counterexample shows that for $\mathfrak{C}(n)_{\perp}$, there exist belief functions which are not invertible.

Example 2. Let us take $\mathfrak{C}(3)_{\perp}$, and consider a function f whose values are given on the figure below. Monotonicity implies that $1 \geq \beta \geq \alpha \geq 0$. In order to check ∞-monotonicity, from Proposition 2 we know that it suffices to check till 7-monotonicity. We write below the most constraining inequalities only, keeping in mind that $1 \geq \beta \geq \alpha \geq 0$.

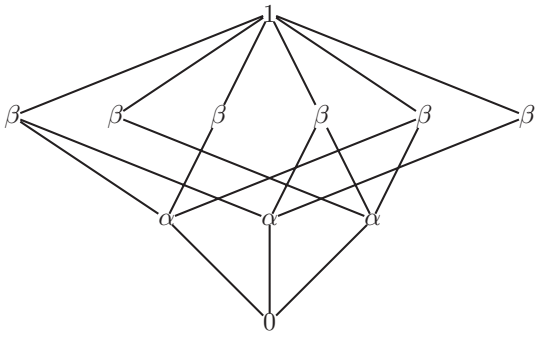


Figure 3: Hasse diagram of $(\mathfrak{C}(3)_{\perp}, \sqsubseteq)$

2-monotonicity is equivalent to $\beta \geq 2\alpha$ and $1 \geq 2\beta - \alpha$. 3-monotonicity is equivalent to $1 \geq 3\beta - \alpha$. 4-monotonicity is equivalent to $1 \geq 4\beta - 3\alpha$. 5-monotonicity is equivalent to $1 \geq 5\beta - 4\alpha$, while 6-monotonicity is equivalent to $1 \geq 6\beta - 9\alpha$. 7-monotonicity does not add further constraints.

From Example 1, nonnegativity of the Möbius transform implies $\alpha \geq 0$ (atoms), $\beta \geq 2\alpha$ (2nd level), and $6\alpha - 6\beta + 1 \geq 0$ for the top element. Then taking $\alpha = 0.1$, $\beta = 0.28$ make that f is a belief function, but $m(\top) = -0.08$.

The last point concerns necessity functions. A *necessity function* on a lattice L is a real-valued function f on L such that

$$f(x \wedge y) = \min(f(x), f(y)).$$

The following proposition by Barthélemy shows that necessity functions always exist on $\mathfrak{C}(n)_{\perp}$.

Proposition 8. Let L be a lattice and $f : L \to \mathbb{R}$ such that $f(\perp) = 0$, $f(\top) = 1$. Then f is a necessity function if and only if it is an invertible belief function whose Möbius transform is nonzero on a chain of $\mathfrak{C}(n)_{\perp}$.

In other words, taking any chain C in $\mathfrak{C}(n)_{\perp}$ and assigning nonnegative numbers (at least one of them being positive) on elements of C such that their sum is 1 generates (by Proposition 7) a belief function which is a necessity function.

4 The Shapley value

A key notion in game theory and also in applications of capacity theory is the Shapley value [10]. For classical games, it defines a rational way to share the quantity $v(N)$ among the individuals in N. Our aim is to define a similar notion for games on partitions. We give a simplified exposition of [4]. See also Macho-Stadler et al. [5] for another proposition based on their structure.

Basically, our Shapley value is defined through maximal chains in $\mathfrak{C}(N)_{\perp}$, which are called *scenarios*, because they depict a particular story of coalition formation, starting from the society of individuals and arriving at the grand coalition. The set of scenarios is denoted by $\mathfrak{S}$.

ISBN: 978-989-95079-6-8

In a scenario $\mathcal{S}$, some elements play a special role. We consider those elements $S\pi$ such that in the sequence of elements of $\mathcal{S}$ from bottom to top, $S\pi$ is the last element with S. They are called *terminal elements*. The set of terminal elements in $\mathcal{S}$ is denoted by $\mathcal{F}(\mathcal{S})$. For example, in the following scenario with $N = \{1, 2, 3, 4\}$, terminal elements are in bold:

$$\mathcal{S}_1 = \mathbf{1}\{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}\} \to 13\{13, 2, 4\} \to \mathbf{13}\{\mathbf{13}, \mathbf{24}\} \to \mathbf{N}\{\mathbf{N}\}.$$

Basically, a value represents the average contribution of players in the game. The contribution of player i is defined as the sum of his contribution in each scenario. However, this contribution is more complex to define than in classical games. Taking as example the above scenario $\mathcal{S}_1$, we see that three situations can arise:

(i) In one step, a single player enters: this happens for steps 1 and 2. This is exactly like classical games, and the contribution goes for the entering player.

(ii) In one step, several players enter together: this happens in the last step where players 2 and 4 enter together. In this case, the principle of insufficient reason tells us to divide the contribution equally among the entering players.

(iii) In one step, no new player enters: this happens in step 3 of $\mathcal{S}_1$. In this case, we wait till a new change occurs (i.e., a new player enters). Hence these steps are not taken into account.

Note that the third case implies that only steps which correspond to terminal elements are taken into account. Applying this methodology to the above scenario $\mathcal{S}_1$, we obtain:

(i) contribution of player 1: $v(1\{1, 2, 3, 4\}) - 0$

(ii) contribution of player 2: $\frac{1}{2}(v(1234\{1234\}) - v(13\{13, 24\}))$

(iii) contribution of player 3: $v(13\{13, 24\}) - v(1\{1, 2, 3, 4\})$

(iv) contribution of player 4: $\frac{1}{2}(v(1234\{1234\}) - v(13\{13, 24\}))$

Based on the above considerations, we can define the contribution of a player in a scenario.

Definition 4. The *contribution of player i in a given scenario* $\mathcal{S}$ is given by

$$\Delta_i^{\mathcal{S}}(v) := \frac{1}{|S' \setminus S|}(v(S'\pi') - v(S\pi)),$$

where $S\pi$ is the last element where i is not present (note that $S\pi \in \mathcal{F}(\mathcal{S})$), and $S'\pi'$ is the next terminal element.

Based on the idea of scenario, we define a scenario-value, from which the value will be constructed.

Definition 5. (i) A *scenario-value* is a mapping $\phi : \mathcal{PG} \to \mathbb{R}^{n\times|\mathfrak{S}|}$. Components of $\phi(v)$ are denoted by $\phi_i^{\mathcal{S}}(v)$ for scenario $\mathcal{S}$ and player i.

The *Shapley scenario-value* is defined by

$$\phi_i^{\mathcal{S}}(v) := \Delta_i^{\mathcal{S}}(v), \quad i \in N, \mathcal{S} \in \mathfrak{S}.$$

(ii) A *value* is a mapping $\phi : \mathcal{PG} \to \mathbb{R}^n$. Components of $\phi(v)$ are denoted by $\phi_i(v)$ for player i. Any scenario-value ϕ induces a value by:

$$\phi_i(v) := \frac{1}{c}\sum_{\mathcal{S}\in\mathfrak{S}} \phi_i^{\mathcal{S}}(v),$$

where c is the number of maximal chains in $\mathfrak{C}(N)_\perp$. The *Shapley value of v* is induced by the Shapley scenario-value.

It is possible to give a direct expression of the Shapley value without using the scenario-value, by expressing the game in the basis of identity games and using linearity of the Shapley value. Using the generic notation $T\sigma := T\{T, T_2, \ldots, T_k\}$, we have

$$\begin{aligned}\phi_i(v) = \frac{1}{n}v(N\{N\}) + \\ \sum_{T\sigma, T\ni i, T\neq N} \frac{2(n-k)!}{n!n!}(k-1)(k-1)!(k-2)!\times \\ t!t_2!\cdots t_k!v(T\sigma) \\ - \sum_{T\sigma, T\not\ni i} \frac{2t(n-k)!}{n!n!}(k-1)!(k-2)!\times \\ t!(t_2-1)!\cdots t_k!v(T\sigma),\end{aligned}$$

where it is assumed in the third term that $i \in T_2$, the second block of σ. An equivalent expression, although less computationally efficient but closer to the classical Shapley value, is:

$$\begin{aligned}\phi_i(v) = \frac{1}{n}v(N\{N\}) \\ + \sum_{T\sigma, T\not\ni i, T_2\supset\{i\}} \frac{2t(n-k)!}{n!n!}(k-1)!(k-2)!\times \\ t!(t_2-1)!\cdots t_k!\Big[\frac{t+1}{t}v(T\cup i\sigma_{T\cup i}) - v(T\sigma)\Big] \\ - \sum_{T\sigma, T\not\ni i, T_2=\{i\}} \frac{2t(n-k)!}{n!n!}(k-1)!(k-2)!\times \\ t!t_3!\cdots t_k!v(T\sigma),\end{aligned}$$

with $\sigma_{T\cup i}$ the partition obtained from σ by moving $i \in T_2$ to T. With $n = 3$, we obtain:

$$\begin{aligned}\phi_1(v) = \frac{1}{3}v(N\{N\}) + \frac{1}{9}v(12\{12, 3\}) + \frac{1}{9}v(13\{13, 2\}) \\ - \frac{2}{9}v(23\{23, 1\}) + \frac{1}{9}v(1\{1, 23\}) - \frac{1}{18}v(2\{2, 13\}) \\ - \frac{1}{18}v(3\{3, 12\}) + \frac{2}{9}v(1\{1, 2, 3\}) - \frac{1}{9}v(2\{1, 2, 3\}) \\ - \frac{1}{9}v(3\{1, 2, 3\}).\end{aligned}$$

We introduce axioms for the characterization of the Shapley scenario-value.

Definition 6. (i) A value is *efficient (E)* if $\sum_{i\in N}\phi_i(v) = v(N\{N\})$.

ISBN: 978-989-95079-6-8

(ii) A scenario-value is *scenario-efficient (SE)* if $\sum_{i\in N}\phi_i^{\mathcal{S}}(v) = v(N\{N\})$ for all $\mathcal{S}\in\mathfrak{S}$.

Evidently, scenario-efficiency implies efficieny of the induced value.

Proposition 9. The Shapley scenario-value is scenario-efficient. Hence, the Shapley value is efficient as well.

A scenario-value satisfies *linearity (L)* if it is linear on $\mathcal{PG}(N)$.

Definition 7. Let us consider $i\in N$, a scenario $\mathcal{S}$, and denote by $S\pi$ the last element in $\mathcal{S}$ not containing i, and $S'\pi'$ its successor in $\mathcal{F}(\mathcal{S})$. Player i is *null in scenario* $\mathcal{S}$ *for* v if $v(S\pi) = v(S'\pi')$.

> **Null axiom (N):** If i is null in scenario $\mathcal{S}$ for v, then $\phi_i^{\mathcal{S}}(v) = 0$.

The *symmetry axiom for the scenario-value (SS)* reads: For any $i\in N$, any $\mathcal{S}\in\mathfrak{S}$, and any permutation σ on N, it holds

$$\phi_i^{\mathcal{S}}(v) = \phi_{\sigma(i)}^{\sigma(\mathcal{S})}(v\circ\sigma^{-1})$$

with $\sigma(\mathcal{S}), \sigma(S,\pi)$ defined naturally as follows: $\sigma(S,\pi) = (\sigma(S),\sigma(\pi))$, where $\sigma(S) = \{\sigma(i)\mid i\in S\}$, $\sigma(\pi) = \{\sigma(S)\mid S\in\pi\}$, and $\sigma(\mathcal{S}) = \{\sigma(S,\pi)\mid (S,\pi)\in\mathcal{S}\}$.

We given the following simple characterization.

Proposition 10. The Shapley scenario-value is the unique scenario-value satisfying (L), (N), (SS), and (SE).

Another axiomatization avoiding the strong (SE) axiom is given in [4].

5 The Choquet integral

We end the paper by giving some considerations on the Choquet integral defined for capacities on partitions. The (classical) Choquet integral in its discrete version is also based on maximal chains. Specifically, taking a function $f: N\to\mathbb{R}_+$, the maximal chain induced by f is given by

$$\{\sigma(n)\}, \{\sigma(n),\sigma(n-1)\},\ldots,\{\sigma(n),\ldots,\sigma(1)\}$$

where σ is any permutation such that

$$f_{\sigma(1)}\le f_{\sigma(2)}\le\cdots\le f_{\sigma(n)}$$

where $f(i)$ is denoted by f_i for simplicity. Then the Choquet integral of f is defined by:

$$\int f\,dv := \sum_{i=1}^{n}(f_{\sigma(i)} - f_{\sigma(i-1)})v(\{\sigma(n),\ldots,\sigma(i)\}),$$

with the convention $f_{\sigma(0)} = 0$. Let us keep this idea for capacities on partitions, and consider for simplicity a function f such that $f_1 < f_2 < \cdots < f_n$. Then there is only one maximal chain in $\mathfrak{C}(N)_\perp$ induced by f, containing the subsets $\{n\}, \{n,n-1\},\ldots,N$, which is

$$\{n\}\pi^\perp, \{n,n-1\}\pi^\perp_{\{n,n-1\}},\ldots,N\{N\}.$$

Hence the Choquet integral should be defined by:

$$\int f\,dv := \sum_{i=1}^{n}(f_i - f_{i-1})v(\{n,\ldots,i\}\{\{n,\ldots,i\},1,2,\ldots,i-1\}).$$

Consider now that $f_1 = f_2 < f_3\cdots < f_n$. Since two permutations can order this function, and the result should be the same whatever the permutation used, it is not difficult to check that this can be achieved only if the above formula is used, thus excluding the use of embedded coalitions which are not of the form $S\pi^\perp_S$. But then it suffices to put $\tilde{v}(S) := v(S\pi^\perp_S)$, and we are back to the classical Choquet integral.

As a conclusion, it does not seem to make sense to consider capacities on partitions for the Choquet integral.

References

[1] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[2] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[3] R. M. Thrall and W. F. Lucas. n-person games in partition function form. *Naval Research Logistic Quarterly*, 10:281–298, 1963.

[4] M. Grabisch and Y. Funaki. A coalition formation value for games in partition function form. submitted.

[5] I. Macho-Stadler, D. Pérez-Castillo, and D. Wettstein. Sharing the surplus: An extension of the Shapley value for environments with externalities. *J. of Economic Theory*, 135:339–356, 2007.

[6] J.-P. Barthélemy. Monotone functions on finite lattices: an ordinal approach to capacities, belief and necessity functions. In J. Fodor, B. De Baets, and P. Perny, editors, *Preferences and Decisions under Incomplete Knowledge*, pages 195–208. Physica Verlag, 2000.

[7] G. Birkhoff. *Lattice Theory*. American Mathematical Society, 3d edition, 1967.

[8] G. C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:340–368, 1964.

[9] M. Aigner. *Combinatorial Theory*. Springer Verlag, 1979.

[10] L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.

ISBN: 978-989-95079-6-8

The Minimization of the Risk of Falling in Portfolios under Uncertainty

Yuji Yoshida
Faculty of Economics and Business Administration, University of Kitakyushu
4-2-1 Kitagata, Kokuraminami, Kitakyushu 802-8577, Japan
Email: yoshida@kitakyu-u.ac.jp

Abstract— *A portfolio model to minimize the risk of falling under uncertainty is discussed. The risk of falling is represented by the value-at-risk of rate of return. Introducing the perception-based extension of the value-at-risk, this paper formulates a portfolio problem to minimize the risk of falling with fuzzy random variables. In the proposed model, randomness and fuzziness are evaluated respectively by the probabilistic expectation and the mean with evaluation weights and λ-mean functions. The analytical solutions of the portfolio problem regarding the risk of falling are derived. This paper gives formulae to show the explicit relations among the following important parameters in portfolio: The expected rate of return, the risk probability of falling and bankruptcy, and the rate of falling regarding the asset prices. A numerical example is given to explain how to obtain the optimal portfolio and these parameters from the asset prices in the stock market. Several figures are shown to observe the relation among these parameters at the optimal portfolios.*

Keywords— Value-at-risk (VaR), risk-sensitive portfolio, fuzzy random variable, perception-based extension, the risk probability, the rate of falling.

1 Introduction

In financial market, the portfolio is one of the most useful risk allocation technique for stable asset management. The minimization of the financial risk as well as the maximization of the return are important themes in the asset management. In a classical portfolio theory, *Markowitz's mean-variance model* is studied by many researchers and fruitful results have been achieved, and then the variance is investigated as the risk for portfolios ([8, 9, 11]). In this paper we focus on the drastic decline of asset prices. Recently, *value-at-risk (VaR)* is used widely to estimate the risk that asset prices decline based on worst scenarios. VaR is a risk-sensitive criterion based on percentiles, and it is one of the standard criteria in asset management ([17]). VaR is a kind of risk-level values of the asset at a specified probability of decline and it is used to select portfolios after due consideration of worst scenarios in investment. Many researchers and financial traders usually use VaR by mathematical programming since it is not easy to analyze the VaR portfolio model mathematically. Because Markowitz's mean-variance criterion and the variance-minimizing criterion are represented by quadratic programming, but VaR criterion in portfolio is neither linear nor quadratic ([17]). In this paper, by use of VaR regarding the rates of return, we discuss a *portfolio selection problem not only to minimize the rates of falling and but also to maximize the expected rates of return*. In the proposed portfolio model, we can maximize the expected rate of return after due consideration of the worst scenarios. This paper derives analytical solutions for the portfolio problem to minimize the rate of falling. The main purpose in this paper is to derive the explicit relations among the important parameters γ, p and δ regarding the financial risk from the obtained analytical results, where γ is the expected rate of return, p is the risk probability of falling and bankruptcy, and δ is the rate of falling regarding the asset at the optimal portfolios.

Soft computing like fuzzy logic works well for financial models in uncertain environment. Estimation of uncertain quantities is important in decision making. To represent uncertainty in this portfolio model, we use *fuzzy random variables* which have two kinds of uncertainties, i.e. randomness and fuzziness. In this paper, randomness is used to represent the uncertainty regarding the belief degree of frequency, and fuzziness is applied to linguistic imprecision of data because of a lack of knowledge regarding the current stock market. At the financial crisis in October 2008, we have observed the serious distrust of the market that the risky information regarding banks and security companies, for example the amounts of trouble loans, risky accounts, debts and so on, may not disclose to the investors and the public, and it is surely a kind of risks occurring from the imprecision of information. The fuzziness comes from the imprecision of data because of a lack of knowledge, and such serious distrust in the stock market will be represented by the fuzziness of information in finance models. We extend the VaR for real random variables to one regarding fuzzy random variables from the viewpoint of perception-based approach in Yoshida [14]. We formulate the a portfolio problem with fuzzy random variables, and we discuss the fundamental properties of the extended VaR. Recently, Yoshida [13] introduced the mean, the variance and the measurement of fuzziness of fuzzy random variables, using *evaluation weights* and λ-mean functions. This paper estimates fuzzy numbers and fuzzy random variables by the probabilistic expectation and these criteria, which are characterized by *possibility and necessity criteria* for subjective estimation and a *pessimistic-optimistic index* for subjective decision. These parameters are decided by the investor and are based on the degree of his certainty regarding the current information in the market. In this portfolio model, we use triangle-type fuzzy numbers and fuzzy random variables for computation in actual models. Regarding falling and bankruptcy, we observe the direct relations among the rate of falling, the expected rate of return and the risk probability at the optimal portfolios by figures derived from the obtained analytical results.

2 A portfolio model and the rate of falling

In this section, we explain a portfolio model with n stocks, where n is a positive integer. Let $\mathbb{T} := \{0, 1, 2, \cdots, T\}$ be the time space with an expiration date T, and $\mathbb{R}$ denotes the set of all real numbers. Let (Ω, P) be a probability space, where P is a non-atomic probability on a sample space Ω. For an asset $i = 1, 2, \cdots, n$, a *stock price process* $\{S_t^i\}_{t=0}^T$ is given

ISBN: 978-989-95079-6-8

by *rates of return* R_t^i as follows. Let

$$S_t^i := S_{t-1}^i(1 + R_t^i) \tag{1}$$

for $t = 1, 2, \cdots, T$, where $\{R_t^i\}_{t=1}^T$ is assumed to be an integrable sequence of independent real random variables. Hence $w_t = (w_t^1, w_t^2, \cdots, w_t^n)$ is called a *portfolio weight vector* if it satisfies $w_t^1 + w_t^2 + \cdots + w_t^n = 1$, and further a portfolio $(w_t^1, w_t^2, \cdots, w_t^n)$ is said to *allow for short selling* if $w_t^i \geq 0$ for all $i = 1, 2, \cdots, n$. Then the rate of return with a portfolio $(w_t^1, w_t^2, \cdots, w_t^n)$ is given by

$$R_t := w_t^1 R_t^1 + w_t^2 R_t^2 + \cdots + w_t^n R_t^n. \tag{2}$$

Therefore, the reward at time $t = 1, 2, \cdots, T$ follows

$$S_t := S_{t-1} \sum_{i=1}^{n} w_t^i (1 + R_t^i) = S_{t-1}(1 + R_t). \tag{3}$$

In this paper, we present a portfolio model where stock price processes S_t^i take fuzzy values using fuzzy random variables. The falling of asset prices is one of the most important risks in stock markets. In this section, we discuss a portfolio model where the risk is estimated by the rate of falling. Regarding the asset (3) with the portfolio w_t, the theoretical *bankruptcy* at time t occurs on scenarios ω satisfying $S_t(\omega) \leq 0$, i.e. it follows $1 + R_t(\omega) \leq 0$ from (3). Similarly, for a constant δ satisfying $0 \leq \delta \leq 1$, a set of sample paths

$$\{\omega \in \Omega | 1 + R_t(\omega) \leq 1 - \delta\} = \{\omega \in \Omega | R_t(\omega) \leq -\delta\} \tag{4}$$

is the event of scenarios where the asset price S_t will fall from the current price S_{t-1} to a lower level than $100(1 - \delta)\,\%$ of the current price S_{t-1}, i.e. the rate of falling is $100\,\delta\,\%$. Then we say shortly that the asset is $100\,\delta\,\%$*-falling* at time t and the parameter δ is called *the rate of falling*. The probability of $100\,\delta\,\%$-falling is also given by

$$p_\delta := P(R_t \leq -\delta). \tag{5}$$

For example, p_δ denotes the probability of the falling below par value if '$\delta = 0$' and it indicates the probability of the bankruptcy if '$\delta = 1$'. In this paper, we discuss portfolios to minimize the rate of falling δ.

For a positive probability p, a value-at-risk (VaR) regarding the rate of return R_t at the probability p is given by a real number v satisfying

$$P(R_t \leq v) = p \tag{6}$$

since P is non-atomic. The value-at-risk v is the upper bound of the rate of return R_t at the worst scenarios under a given risk probability p, and then the value-at-risk v in (6) is denoted by $\mathrm{VaR}_p(R_t)$. From (5) and (6), for a risk probability $p = p_\delta$, the rate of falling is

$$\delta = -\mathrm{VaR}_p(R_t). \tag{7}$$

To minimize the rate of falling (7) under a fuzzy and random environment, in next section we discuss the fundamental properties of value-at-risks.

3 A portfolio model with fuzzy random variables

In this section, we introduce fuzzy numbers and fuzzy random variables and we give a portfolio model under uncertainty. A fuzzy number is denoted by its membership function $\tilde{a} : \mathbb{R} \mapsto [0, 1]$ which is normal, upper-semicontinuous and quasi-concave and has a compact support ([18]). $\mathcal{R}$ denotes the set of all fuzzy numbers. In this paper, we identify fuzzy numbers with their corresponding membership functions. The α-cut of a fuzzy number $\tilde{a}(\in \mathcal{R})$ is given by $\tilde{a}_\alpha := \{x \in \mathbb{R} \mid \tilde{a}(x) \geq \alpha\}$ $(\alpha \in (0, 1])$ and $\tilde{a}_0 := \mathrm{cl}\{x \in \mathbb{R} \mid \tilde{a}(x) > 0\}$, where cl denotes the closure of an interval. We write the closed intervals as $\tilde{a}_\alpha := [\tilde{a}_\alpha^-, \tilde{a}_\alpha^+]$ for $\alpha \in [0, 1]$. Hence we also introduce a partial order $\succeq$, so called the *fuzzy max order*, on fuzzy numbers $\mathcal{R}$: Let $\tilde{a}, \tilde{b} \in \mathcal{R}$ be fuzzy numbers. Then, $\tilde{a} \succeq \tilde{b}$ means that $\tilde{a}_\alpha^- \geq \tilde{b}_\alpha^-$ and $\tilde{a}_\alpha^+ \geq \tilde{b}_\alpha^+$ for all $\alpha \in [0, 1]$. An addition, a subtraction and a scalar multiplication for fuzzy numbers are defined by Zadeh's extension principle as follows: For $\tilde{a}, \tilde{b} \in \mathcal{R}$ and $\xi \in \mathbb{R}$, the addition and subtraction $\tilde{a} \pm \tilde{b}$ of $\tilde{a}$ and $\tilde{b}$ and the scalar multiplication $\xi\tilde{a}$ of ξ and $\tilde{a}$ are fuzzy numbers given by their α-cuts $(\tilde{a}+\tilde{b})_\alpha := [\tilde{a}_\alpha^- + \tilde{b}_\alpha^-, \tilde{a}_\alpha^+ + \tilde{b}_\alpha^+]$, $(\tilde{a}-\tilde{b})_\alpha := [\tilde{a}_\alpha^- - \tilde{b}_\alpha^+, \tilde{a}_\alpha^+ - \tilde{b}_\alpha^-]$ and $(\xi\tilde{a})_\alpha := [\xi\tilde{a}_\alpha^-, \xi\tilde{a}_\alpha^+]$ if $\xi \geq 0$.

A fuzzy-number-valued map $\tilde{X} : \Omega \mapsto \mathcal{R}$ is called a *fuzzy random variable* if the maps $\omega \mapsto \tilde{X}_\alpha^-(\omega)$ and $\omega \mapsto \tilde{X}_\alpha^+(\omega)$ are measurable for all $\alpha \in (0, 1]$, where $\tilde{X}_\alpha(\omega) = [\tilde{X}_\alpha^-(\omega), \tilde{X}_\alpha^+(\omega)] = \{x \in \mathbb{R} \mid \tilde{X}(\omega)(x) \geq \alpha\}$ ([7, 10]). We need to introduce expectations of fuzzy random variables in order to describe a portfolio model. A fuzzy random variable $\tilde{X}$ is said to be integrably bounded if both $\omega \mapsto \tilde{X}_\alpha^-(\omega)$ and $\omega \mapsto \tilde{X}_\alpha^+(\omega)$ are integrable for all $\alpha \in (0, 1]$. Let $\tilde{X}$ be an integrably bounded fuzzy random variable. The expectation $E(\tilde{X})$ of the fuzzy random variable $\tilde{X}$ is defined by a fuzzy number $E(\tilde{X})(x) := \sup_{\alpha \in [0,1]} \min\{\alpha, 1_{E(\tilde{X})_\alpha}(x)\}$ for $x \in \mathbb{R}$, where $E(\tilde{X})_\alpha := [\int_\Omega \tilde{X}_\alpha^-(\omega)\,\mathrm{d}P(\omega), \int_\Omega \tilde{X}_\alpha^+(\omega)\,\mathrm{d}P(\omega)]$ for $\alpha \in (0, 1]$ ([6, 10]).

Now we deal with a case where the rate of return $\{R_t^i\}_{t=1}^T$ has some imprecision. In this paper, we use triangle-type fuzzy random variables for computation, however we can apply the similar approach to general fuzzy random variables. We define a *rate of return process with imprecision* $\{\tilde{R}_t^i\}_{t=0}^T$ by a sequence of triangle-type fuzzy random variables

$$\tilde{R}_t^i(\cdot)(x) = \begin{cases} 0 & \text{if } x < R_t^i - c_t^i \\ \frac{x - R_t^i + c_t^i}{c_t^i} & \text{if } R_t^i - c_t^i \leq x < R_t^i \\ \frac{x - R_t^i - c_t^i}{-c_t^i} & \text{if } R_t^i \leq x < R_t^i + c_t^i \\ 0 & \text{if } x \geq R_t^i + c_t^i, \end{cases} \tag{8}$$

where c_t^i is a positive number. We call c_t^i a *fuzzy factor* for asset i at time t. Hence we can represent $\tilde{R}_t^i$ by the sum of the real random variable R_t^i and a fuzzy number $\tilde{a}_t^i$:

$$\tilde{R}_t^i(\omega)(\cdot) := 1_{\{R_t^i(\omega)\}}(\cdot) + \tilde{a}_t^i(\cdot) \tag{9}$$

for $\omega \in \Omega$, where $1_{\{\cdot\}}$ denotes the characteristic function of a singleton and $\tilde{a}_t^i$ is a triangle-type fuzzy number defined by

$$\tilde{a}_t^i(x) = \begin{cases} 0 & \text{if } x < -c_t^i \\ \frac{x + c_t^i}{c_t^i} & \text{if } -c_t^i \leq x < 0 \\ \frac{x - c_t^i}{-c_t^i} & \text{if } 0 \leq x < c_t^i \\ 0 & \text{if } x \geq c_t^i. \end{cases} \tag{10}$$

ISBN: 978-989-95079-6-8

For assets $i = 1, 2, \cdots, n$, we define *stock price processes* $\{\tilde{S}_t^i\}_{t=0}^T$ by the *rates of return with imprecision* $\tilde{R}_t^i$ as follows: $S_0^i := S_0^i$ is a positive number and

$$\tilde{S}_t^i = \tilde{S}_0^i \prod_{s=1}^{t} (1 + \tilde{R}_s^i) \tag{11}$$

for $t = 1, 2, \cdots, T$. For a portfolio $w = (w^1, w^2, \cdots, w^n)$, the rate of return with imprecision is given by a linear combination of fuzzy random variables

$$\tilde{R}_t := w^1 \tilde{R}_t^1 + w^2 \tilde{R}_t^2 + \cdots + w^n \tilde{R}_t^n. \tag{12}$$

In Section 4 we investigate the value-at-risk to apply the fuzzy random variable (12), and in Section 5 we discuss the portfolio problem to minimize the rate of falling regarding (12).

4 A perception-based extension of VaR

First we introduce mathematical notations of the value-at-risk for real random variables to apply it to the rates of return (12). Let $\mathcal{X}$ be the set of all integrable real random variables X on Ω with a continuous distribution function $x \mapsto F_X(x) := P(X < x)$ for which there exists a non-empty open interval I such that $F_X(\cdot) : I \mapsto (0, 1)$ is a strictly increasing and onto. Then there exists a strictly increasing and continuous inverse function $F_X^{-1} : (0, 1) \mapsto I$. We note that $F_X(\cdot) : I \mapsto (0, 1)$ and $F_X^{-1} : (0, 1) \mapsto I$ are one-to-one and onto, and we put $F_X(\inf I) := \lim_{x \downarrow \inf I} F_X(x) = 0$ and $F_X(\sup I) := \lim_{x \uparrow \sup I} F_X(x) = 1$. Then, the *value-at-risk*, shortly for *VaR*, at a risk probability p is given by the $100\,p$-percentile of the distribution function F_X: For a probability p $(0 < p < 1)$,

$$\mathrm{VaR}_p(X) := \sup\{x \in I \mid F_X(x) \le p\}, \tag{13}$$

and then we have $F_X(\mathrm{VaR}_p(X)) = p$ and $\mathrm{VaR}_p(X) = F_X^{-1}(p)$ for $0 < p < 1$. In this paper, we assume that VaR v in (6) has the following representation (14).

$$\begin{aligned}(\text{VaR } v) \quad &= (\text{the mean}) - (\text{a positive constant } \kappa) \\ &\quad \times (\text{the standard deviation}),\end{aligned} \tag{14}$$

where the positive constant κ is given corresponding to the probability p. The details are as follows: For any probability p satisfying $0 < p < 1$, there exists a positive constant κ such that a real number $v := \mu_t - \kappa\,\sigma_t$ satisfies (14) for all portfolios, where μ_t and σ_t are the expectation and the standard deviation of R_t respectively. One of the most popular sufficient condition for (14) is what the distribution of the rate of return R_t is normal ([2, 17]). In this paper, we obtain an exact estimation of κ in Theorem 1.

Let $\tilde{\mathcal{X}}$ be the set of all fuzzy random variables $\tilde{X}$ on Ω such that their α-cuts $\tilde{X}_\alpha^\pm$ are integrable and $\lambda\tilde{X}_\alpha^- + (1-\lambda)\tilde{X}_\alpha^+ \in \mathcal{X}$ for all $\lambda \in [0, 1]$ and $\alpha \in [0, 1]$. Hence, from (8) we introduce a value-at-risk for a fuzzy random variable $\tilde{X}(\in \tilde{\mathcal{X}})$ at a positive risk probability p as follows.

$$\mathrm{VaR}_p(\tilde{X})(x) := \sup_{X \in \mathcal{X} \,:\, \mathrm{VaR}_p(X) = x} \inf_{\omega \in \Omega} \tilde{X}(\omega)(X(\omega)) \tag{15}$$

for $x \in \mathbb{R}$. Yoshida [14] has studied *perception-based estimations* extending the concept of the expectations in Kruce and Meyer [6]. This definition (15) is an extension from the value-at-risk on real random variables to one on fuzzy random variables based on the perception. The value-at-risk (15) on fuzzy random variables is characterized by the following representation, which is from the continuity and the monotonicity of $\mathrm{VaR}_p(\cdot)$.

Lemma 1 ([14]). *Let $\tilde{X} \in \tilde{\mathcal{X}}$ be a fuzzy random variable and let p be a positive probability. Then the value-at-risk $\mathrm{VaR}_p(\tilde{X})$ defined by (15) is a fuzzy number whose α-cuts are*

$$\mathrm{VaR}_p(\tilde{X})_\alpha = [\mathrm{VaR}_p(\tilde{X}_\alpha^-), \mathrm{VaR}_p(\tilde{X}_\alpha^+)] \tag{16}$$

for $\alpha \in (0, 1]$.

The value-at-risk (15) on fuzzy random variables has the following properties.

Lemma 2 ([14]). *Let $\tilde{X}, \tilde{Y} \in \tilde{\mathcal{X}}$ be fuzzy random variables and let p be a positive probability. Then the value-at-risk VaR_p defined by (15) has the following properties:*

(i) *If $\tilde{X} \preceq \tilde{Y}$, then $\mathrm{VaR}_p(\tilde{X}) \preceq \mathrm{VaR}_p(\tilde{Y})$.*

(ii) *$\mathrm{VaR}_p(\zeta\tilde{X}) = \zeta\,\mathrm{VaR}_p(\tilde{X})$ for $\zeta > 0$.*

(iii) *$\mathrm{VaR}_p(\tilde{X} + \tilde{a}) = \mathrm{VaR}_p(\tilde{X}) + \tilde{a}$ for a fuzzy number $\tilde{a} \in \mathcal{R}$.*

Next since the value-at-risk $\mathrm{VaR}_p(\tilde{R}_t)$ for the rate of return (12) with a portfolio is a fuzzy number, we need to evaluate the fuzziness of fuzzy numbers and fuzzy random variables. There are many studies regarding the evaluation of fuzzy numbers. Two major approaches of them are as follows. One is to use weighting functions([1, 3, 13]) and the other is to use possibility and necessity criteria([4]). Here we adopt the former evaluation method of fuzzy numbers and fuzzy random variables. In the rest of this section we introduce the definitions from [13, 15], and in the next section we estimate the VaR regarding the rate of return (12) by the evaluation method. Yoshida [13] has studied an evaluation of fuzzy numbers by *evaluation weights* which are induced from fuzzy measures to evaluate a confidence degree that a fuzzy number takes values in an interval. With respect to fuzzy random variables, the randomness is evaluated by the probabilistic expectation and the fuzziness is estimated by the evaluation weights and the following function. Let $g^\lambda : \mathcal{I} \mapsto \mathbb{R}$ be a map such that

$$g^\lambda([x, y]) := \lambda x + (1 - \lambda)y, \quad [x, y] \in \mathcal{I}, \tag{17}$$

where λ is a constant satisfying $0 \le \lambda \le 1$ and $\mathcal{I}$ denotes the set of all bounded closed intervals. This scalarization is used for the estimation of fuzzy numbers to give a mean value of the interval $[x, y]$ with a weight λ, and g^λ is called a *λ-mean function* and λ is called a *pessimistic-optimistic index* which indicates the pessimistic degree of attitude in decision making ([3]). Let a fuzzy number $\tilde{a} \in \mathcal{R}$. A mean value of the fuzzy number $\tilde{a}$ with respect to λ-mean functions g^λ and an evaluation weight $w(\alpha)$, which depends only on $\tilde{a}$ and α, is given as follows ([15]):

$$\tilde{E}(\tilde{a}) := \frac{\int_0^1 g^\lambda(\tilde{a}_\alpha)\, w(\alpha)\, \mathrm{d}\alpha}{\int_0^1 w(\alpha)\, \mathrm{d}\alpha}, \tag{18}$$

where $\tilde{a}_\alpha = [\tilde{a}_\alpha^-, \tilde{a}_\alpha^+]$ is the α-cut of the fuzzy number $\tilde{a}$. In (18), $w(\alpha)$ indicates a *confidence degree that the fuzzy*

ISBN: 978-989-95079-6-8

number $\tilde{a}$ takes values in the interval $\tilde{a}_\alpha$ at each level α. Hence, an evaluation weight $w(\alpha)$ is called the *possibility evaluation weight* $w^P(\alpha)$ if $w^P(\alpha) := 1$ for $\alpha \in [0,1]$, and $w(\alpha)$ is called the *necessity evaluation weight* $w^N(\alpha)$ if $w^N(\alpha) := 1-\alpha$ for $\alpha \in [0,1]$. Especially, for a fuzzy number $\tilde{a} \in \mathcal{R}$, the means in the possibility and necessity cases are represented respectively by $\tilde{E}^P(\tilde{a})$ and $\tilde{E}^N(\tilde{a})$, and we consider their combination $\nu\tilde{E}^P(\tilde{a}) + (1-\nu)\tilde{E}^P(\tilde{a})$ with a parameter $\nu \in [0,1]$ ([13, 14, 16]). The mean $\tilde{E}$ has the following natural properties of the linearity and the monotonicity regarding the fuzzy max order.

Lemma 3 ([13, 15]). *Let $\lambda \in [0,1]$. For fuzzy numbers $\tilde{a}, \tilde{b} \in \mathcal{R}$ and real numbers θ, ζ, the following (i) – (iv) hold.*

(i) $\tilde{E}(\tilde{a} + 1_{\{\theta\}}) = \tilde{E}(\tilde{a}) + \theta$.

(ii) $\tilde{E}(\zeta\tilde{a}) = \zeta\tilde{E}(\tilde{a})$ *if* $\zeta \geq 0$.

(iii) $\tilde{E}(\tilde{a} + \tilde{b}) = \tilde{E}(\tilde{a}) + \tilde{E}(\tilde{b})$.

(iv) *If* $\tilde{a} \succeq \tilde{b}$, *then* $\tilde{E}(\tilde{a}) \geq \tilde{E}(\tilde{b})$ *holds.*

For a fuzzy random variable $\tilde{X}$, *the mean of the expectation* $E(\tilde{X})$ is a real number

$$E(\tilde{E}(\tilde{X})) = E\left(\frac{\int_0^1 g^\lambda(\tilde{X}_\alpha)\, w(\alpha)\, \mathrm{d}\alpha}{\int_0^1 w(\alpha)\, \mathrm{d}\alpha}\right). \tag{19}$$

Then, from Lemma 3, we obtain the following results regarding fuzzy random variables.

Lemma 4 ([13, 15]). *Let $\lambda \in [0,1]$. For a fuzzy number $\tilde{a} \in \mathcal{R}$, integrable fuzzy random variables $\tilde{X}, \tilde{Y}$, an integrable real random variable Z and a nonnegative real number ζ, the following (i) – (v) hold.*

(i) $E(\tilde{E}(\tilde{X})) = \tilde{E}(E(\tilde{X}))$.

(ii) $E(\tilde{E}(\tilde{a})) = \tilde{E}(\tilde{a})$ *and* $E(\tilde{E}(Z)) = E(Z)$.

(iii) $E(\tilde{E}(\zeta\tilde{X})) = \zeta E(\tilde{E}(\tilde{X}))$.

(iv) $E(\tilde{E}(\tilde{X} + \tilde{Y})) = E(\tilde{E}(\tilde{X})) + E(\tilde{E}(\tilde{Y}))$.

(v) *If* $\tilde{X} \succeq \tilde{Y}$, *then* $E(\tilde{E}(\tilde{X})) \geq E(\tilde{E}(\tilde{Y}))$ *holds.*

5 The Minimization of the Risk of Falling

In this section, we discuss portfolio problems under uncertainty. First we estimate the rate of return with imprecision for a portfolio. Let the mean, the variance and the covariance of the rate of return R_t^i, which are the real random variables in (2.1), by $\mu_t^i := E(R_t^i)$, $(\sigma_t^i)^2 := E((R_t^i - \mu_t^i)^2)$, and $\sigma_t^{ij} := E((R_t^i - \mu_t^i)(R_t^j - \mu_t^j))$ for $i, j = 1, 2, \cdots, n$. Hence we assume that the determinant of the variance-covariance matrix $\Sigma := [\sigma_t^{ij}]$ is not zero and there exists its inverse matrix Σ^{-1}. This assumption is natural and it can be realized easily by taking care of the combinations of assets. For a portfolio $w = (w^1, w^2, \cdots, w^n)$ satisfying $w^1 + w^2 + \cdots + w^n = 1$ and $w^i \geq 0$ $(i = 1, 2, \cdots, n)$, we calculate the expectation and the variance regarding $\tilde{R}_t = w^1\tilde{R}_t^1 + w^2\tilde{R}_t^2 + \cdots + w^n\tilde{R}_t^n$. From Lemma 4, the expectation $\tilde{\mu}_t := E(\tilde{E}(\tilde{R}_t))$ follows

$$\tilde{\mu}_t = E(\tilde{E}(\tilde{R}_t)) = \sum_{i=1}^n w^i E(\tilde{E}(\tilde{R}_t^i)) = \sum_{i=1}^n w^i \tilde{\mu}_t^i, \tag{20}$$

where $\tilde{\mu}_t^i := E(\tilde{E}(\tilde{R}_t^i))$ for $i = 1, 2, \cdots, n$. On the other hand, regarding this model, from [13] we can find that the variance $(\tilde{\sigma}_t)^2 := E((\tilde{E}(\tilde{R}_t) - \tilde{\mu}_t)^2)$ of $\tilde{R}_t$ equals to the variance $(\sigma_t)^2 := E((R_t - \mu_t)^2)$ of R_t:

$$(\tilde{\sigma}_t)^2 = (\sigma_t)^2 = \sum_{i=1}^n \sum_{j=1}^n w^i w^j \sigma_t^{ij}. \tag{21}$$

By (14), (20) and (21), the mean of $\mathrm{VaR}_p(\tilde{R}_t)$ is

$$\tilde{E}(\mathrm{VaR}_p(\tilde{R}_t)) = \sum_{i=1}^n w^i \tilde{\mu}_t^i - \kappa\sqrt{\sum_{i=1}^n \sum_{j=1}^n w^i w^j \sigma_t^{ij}} \tag{22}$$

with a positive constant κ. Now step by step we discuss a portfolio problem to minimize the rate of falling $\delta = -\tilde{E}(\mathrm{VaR}_p(\tilde{R}_t))$. First, we deal with a variance-minimizing model. For a given constant γ, which is the minimum expected rate of return to be guaranteed for the portfolio, we consider the following quadratic programming with respect to portfolios with allowance for short selling.

Variance-minimizing problem (P1): Minimize the variance

$$\sum_{i=1}^n \sum_{j=1}^n w^i w^j \sigma_t^{ij} \tag{23}$$

with respect to portfolios $w = (w^1, w^2, \cdots, w^n)$ satisfying $w^1 + w^2 + \cdots + w^n = 1$ under the following condition regarding the expected rate of return: $\sum_{i=1}^n w^i \tilde{\mu}_t^i = \gamma$.

We start from the following classical results regarding the variance-minimizing problem (P1), and we analyze the portfolio problem to minimize the rate of falling.

Lemma 5 ([15, Theorem 1]). *The solution of the variance-minimizing problem (P1) is given by*

$$w = \xi\Sigma^{-1}\mathbf{1} + \eta\Sigma^{-1}\tilde{\mu} \tag{24}$$

and then the corresponding variance is

$$\tilde{\rho} := \frac{A\gamma^2 - 2B\gamma + C}{\Delta}, \tag{25}$$

where $\tilde{\mu}^i := \mu_t^i + \tilde{E}^\lambda(\tilde{a}_t^i)$, $\tilde{\sigma}^{ij} := \sigma_t^{ij}$ $(i, j = 1, 2, \cdots, n)$, $\tilde{\Sigma} := [\tilde{\sigma}^{ij}]$, $\tilde{\mu} := [\tilde{\mu}^1\ \tilde{\mu}^2\ \cdots\ \tilde{\mu}^n]^{\mathrm{T}}$, $\mathbf{1} := [1\ 1\ \cdots\ 1]^{\mathrm{T}}$, $\xi := \frac{C - B\gamma}{\Delta}$, $\eta := \frac{A\gamma - B}{\Delta}$, $A := \mathbf{1}^{\mathrm{T}}\tilde{\Sigma}^{-1}\mathbf{1}$, $B := \mathbf{1}^{\mathrm{T}}\tilde{\Sigma}^{-1}\tilde{\mu}$, $C := \tilde{\mu}^{\mathrm{T}}\tilde{\Sigma}^{-1}\tilde{\mu}$, $\Delta := AC - B^2$ and $^{\mathrm{T}}$ denotes the transpose of a vector.

Hence, we consider a risk-sensitive model, which is not of mean-variance types but of mean-standard deviation types, in order to deal with a portfolio problem to minimize the rate of falling in the third step. For a constant γ and a positive constant κ, we discuss the following risk-sensitive portfolio problem with allowance for short selling.

Risk-sensitive problem (P2): Maximize a risk-sensitive expected rate of return

$$v(\gamma) := \sum_{i=1}^n w^i \tilde{\mu}_t^i - \kappa\sqrt{\sum_{i=1}^n \sum_{j=1}^n w^i w^j \sigma_t^{ij}} \tag{26}$$

ISBN: 978-989-95079-6-8

with respect to portfolios $w = (w^1, w^2, \cdots, w^n)$ $(w^1 + w^2 + \cdots + w^n = 1)$ under the condition $\sum_{i=1}^n w^i \tilde{\mu}_t^i = \gamma$.

Now we discuss the following VaR portfolio problem without allowance for short selling. The following form (27) comes from the value-at-risk $\tilde{E}(\mathrm{VaR}_p(\tilde{R}_t))$ given in (22).

Portfolio problem to minimize the rate of falling (P3):
Minimize the risk of falling

$$\delta = -\sum_{i=1}^n w^i \tilde{\mu}_t^i + \kappa \sqrt{\sum_{i=1}^n \sum_{j=1}^n w^i w^j \sigma_t^{ij}} \qquad (27)$$

with respect to portfolios $w = (w^1, w^2, \cdots, w^n)$ satisfying $w^1 + w^2 + \cdots + w^n = 1$ and $w^i \geq 0$ for $i = 1, 2, \cdots, n$.

Since we have

$$\inf_w (27) = \inf_\gamma \left(\inf_{w:\sum_{i=1}^n w^i \tilde{\mu}_t^i = \gamma} (27) \right) = -\sup_\gamma (26),$$

in the same way as [16] we arrive at the following analytical solutions of the portfolio problem to minimize the rate of falling (P3).

Lemma 5. *Let A and Δ be positive. Let κ satisfy $\kappa^2 > C$. The solution of the portfolio problem to minimize the rate of falling (P3) is given by $w^* = \xi\Sigma^{-1}\mathbf{1} + \eta\Sigma^{-1}\tilde{\mu}$, and then the corresponding rate of falling is $\delta(\gamma^*) = -\frac{B-\sqrt{A\kappa^2-\Delta}}{A}$ at the expected rate of return $\gamma^* := \frac{B}{A} + \frac{\Delta}{A\sqrt{A\kappa^2-\Delta}}$, where $\xi := \frac{C-B\gamma^*}{\Delta}$ and $\eta := \frac{A\gamma^*-B}{\Delta}$. Further, if $\Sigma^{-1}\mathbf{1} \geq \mathbf{0}$ and $\Sigma^{-1}\tilde{\mu} \geq \mathbf{0}$, then the portfolio (5.18) satisfies $w^* \geq \mathbf{0}$, i.e. the portfolio w^* is a portfolio without allowance for short selling. Here, $\mathbf{0}$ denotes the zero vector.*

In Lemma 5, we note that the optimal portfolio w^* not only to minimize the rate of falling $\delta(\gamma^*)$ but also to maximize the expected rate of return γ^*.

Theorem 1. *Let A and Δ be positive. Let δ satisfy $\delta > -2B/A$. Then, the assumptions in Lemma 5 are satisfied, and the following (i) and (ii) hold for the optimal portfolio in Lemma 5.*

(i) *For a rate of falling δ, the corresponding constant κ_δ and the expected rate of return γ_δ are given by*

$$\kappa_\delta := \sqrt{A\delta^2 + 2B\delta + C} \quad \text{and} \quad \gamma_\delta := \frac{B\delta + C}{A\delta + B}. \qquad (28)$$

Then the risk probability is $p_\delta = P(\tilde{E}(\tilde{R}_t) \leq -\delta)$.

(ii) *If R_t^i $(i = 1, 2, \cdots, n)$ have normal distributions, the risk probability p_δ in (i) is given by*

$$p_\delta := \Phi(-\kappa_\delta), \qquad (29)$$

where κ_δ is defined by (28) and Φ is the cumulative normal distribution function $\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}}\, dt$ for $z \in \mathbb{R}$.

Remark 1 From Theorem 1, for the optimal portfolio of (P3), the relations among the expected rate of return $\gamma = \gamma_\delta$, the rate of falling δ and the risk probability $p = p_\delta$ are illustrated as follows.

$$\left.\begin{array}{lcl} \gamma & : & \text{the expected rate of return} \\ \Updownarrow & \gamma = & \dfrac{B\delta + C}{A\delta + B} \\ \delta & : & \text{the rate of falling}(= -\mathrm{VaR}) \\ \Updownarrow & p = & \Phi(-\sqrt{A\delta^2 + 2B\delta + C}) \\ p & : & \text{the risk probability of falling} \end{array}\right\} \qquad (30)$$

These results hold for the crisp case ($c_t^i = 0$) and they will be useful not only for analytic study but also for actual management in finance. In the next section, by figures with a numerical example we show the relations (30) among the expected rate of return γ, the rate of falling δ and the risk probability p.

6 A numerical example

In this session, we give a simple example to illustrate our idea. For the numerical computation, we need to evaluate fuzzy numbers representing the rates of return (8). From [13, 14], we have evaluations of the fuzzy numbers $\tilde{a}_t^i$ by a combination $\nu\tilde{E}^P(\tilde{a}) + (1-\nu)\tilde{E}^P(\tilde{a})$ with a parameter $\nu \in [0, 1]$ with a parameter $\nu(\in [0, 1])$, which is called a *possibility-necessity weight* ([13]). Then from (20) we obtain the expected rate of return

$$\tilde{\mu}_t = \sum_{i=1}^n w^i \left(\mu_t^i + \frac{(1-2\lambda)(4-\nu)}{6} c_t^i \right) \qquad (31)$$

for the possibility-necessity weight $\nu(\in [0, 1])$ and the pessimistic-optimistic index $\lambda(\in [0, 1])$. In (31), the decision maker may choose the parameters $\lambda(\in [0, 1])$ and $\nu(\in [0, 1])$. The pessimistic-optimistic index is taken as $\lambda = 1$ if he has pessimistic personal forecast in the market and he takes careful decision, and $\lambda = 0$ if he has optimistic personal forecast and he is not nervous. The possibility-necessity weight is taken as $\nu = 1$ when he has enough confidential information about the market, and $\nu = 0$ when he does not have confidential information. In this model, $\nu = 0$ is reasonable since our objective function is VaR, which is a kind of risk, and we need to take into account of the fuzziness of information in the market. While λ depends on the decision maker's attitude in his investment. In this example, we compute the pessimistic case $\lambda = 1$.

Let $n = 4$ be the number of assets. Take the expected rate of return, a variance-covariance matrix and fuzzy factors as Table 1. We assume that the rate of return R_t^i has the normal distributions. We discuss a risk probability 1% in the normal distribution, and then the corresponding constant is $\kappa = 2.326$, which is given in (14) for VaR. Then, the conditions in Theorem 1 are satisfied and by formulae of Lemma 5 we easily obtain the optimal portfolio $w^* = (w^1, w^2, w^3, w^4) = (0.229604, 0.215551, 0.252000, 0.302845)$ for the portfolio problem to minimize the rate of falling (P3), and then the corresponding rate of falling is $\delta(\gamma^*) = 0.557042$ and the expected rate of return is $\gamma^* = 0.0498026$.

Owing to the equations (28) and (29), we can easily observe the explicit relations among the expected rate of return γ, the rate of falling δ and the risk probability p (Remark 1). Regarding this example, Fig.1 shows the relation among the rate of

ISBN: 978-989-95079-6-8

falling δ, the expected rate of return γ and the risk probability p. Figs.1 and 2 are drawn based on Remark 1. Especially Fig.2 shows the relation among the risk probability p such that the falling in the asset prices will be lower than $100(1-\delta)\,\%$, the rate of falling δ and the expected rate of return γ. For example, for a rate of falling $\delta = 0.2$, which implies that the asset price will fall lower than $80\,\%$, its risk probability is $p = 0.168993$, however we have the expected rate of return $\gamma = 0.0504023$ with $\kappa = 0.958154$, which are calculated easily from Remark 1 based on Theorem 1. We can also compute the cases of falling below par value and the complete bankruptcy, taking $\delta = 0$ and $\delta = 1$ respectively. The probability of falling below par value is $p_0 = 0.4211443$ and then the expected rate of return $\gamma_0 = 0.0545276$, and the probability of complete bankruptcy is $p_1 = 0.0000285992$ and then the expected rate of return $\gamma_1 = 0.0496258$.

Table 1: Expected rates of return, a variance-covariance matrix and fuzzy factors.

Asset	μ_t^i
w_1	0.05
w_2	0.07
w_3	0.06
w_4	0.04

Asset	c_t^i
w_1	0.006
w_2	0.008
w_3	0.007
w_4	0.005

σ_t^{ij}	w_1	w_2	w_3	w_4
w_1	0.35	0.03	0.02	-0.08
w_2	0.03	0.25	-0.06	0.08
w_3	0.02	-0.06	0.33	-0.02
w_4	-0.08	0.08	-0.02	0.24

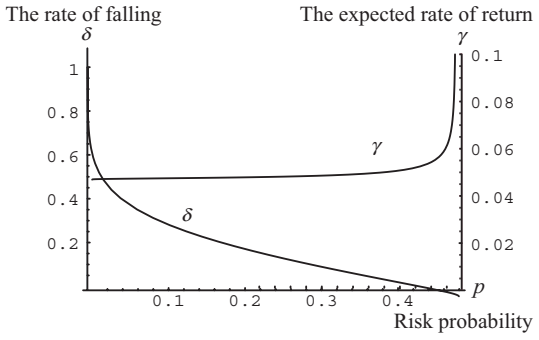


Figure 1: The rate of falling δ and the expected rate of return γ at risk probabilities p.

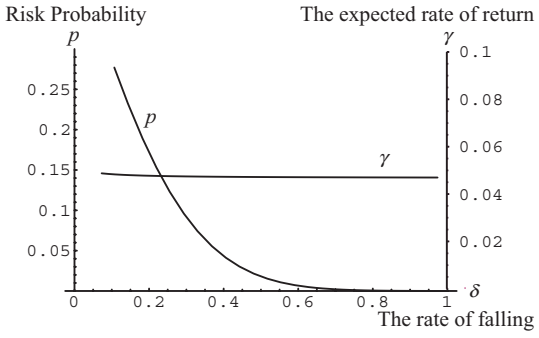


Figure 2: Risk probability p and the expected rate of return γ at the rates of falling δ.

7 Conclusion

We have derived the relation (30) among the expected rate of return γ, the rate of falling δ and the risk probability p. The relation holds for the crisp case and it will be useful not only for analytic study but also for actual management in finance.

References

[1] L.Campos and A.Munoz, A subjective approach for ranking fuzzy numbers, *Fuzzy Sets and Systems*, 29:145-153, 1989.

[2] L.El Chaoui, M.Oks and F.Oustry, Worst-case value at risk and robust portfolio optimization: A conic programming approach, *Operations Research*, 51:543-556, 2003.

[3] P. Fortemps and M. Roubens, Ranking and defuzzification methods based on area compensation, *Fuzzy Sets and Systems*, 82:319-330, 1996.

[4] M.Inuiguchi and T.Tanino, Portfolio selection under independent possibilistic information, *Fuzzy Sets and Systems*, 115:83-92, 2000.

[5] P.Jorion, *Value at Risk: Tne New Benchmark for Managing Financial Risk*, Third Edition McGraw-Hill, New York, 2007.

[6] R.Kruse and K.D.Meyer, *Statistics with Vague Data*, Riedel Publ. Co., Dortrecht, 1987.

[7] H.Kwakernaak, Fuzzy random variables – I. Definitions and theorem, *Inform. Sci.*, 15:1-29, 1978.

[8] H.Markowitz, *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Blackwell, Oxford, 1990.

[9] S.R.Pliska, *Introduction to Mathematical Finance: Discrete-Time Models*, Blackwell Publ., New York, 1997.

[10] M.L.Puri and D.A.Ralescu, Fuzzy random variables, *J. Math. Anal. Appl.*, 114:409-422, 1986.

[11] M.C.Steinbach, Markowitz revisited: Mean-variance model in financial portfolio analysis, *SIAM Review*, 43:31-85, 2001.

[12] H.Tanaka, P.Guo and L.B.Türksen, Portfolio selection based on fuzzy probabilities and possibilistic distribution, *Fuzzy Sets and Systems*, 111:397-397, 2000.

[13] Y.Yoshida, Mean values, measurement of fuzziness and variance of fuzzy random variables for fuzzy optimization, *Proceedings of SCIS & ISIS 2006*, 2277-2282, Sept. 2006.

[14] Y.Yoshida, A risk-minimizing model under uncertainty in portfolio, in: V.Torra et al. eds., *MDAI 2007*, LNAI 4529, 295-306, Springer, August, 2007.

[15] Y.Yoshida, Fuzzy extension of estimations with randomness: The perception-based approach, in: P.Melin et al. eds., *IFSA2007*, LNAI 4617, 381-391, Springer, Sept., 2007.

[16] Y.Yoshida, An estimation model of value-at-risk portfolio under uncertainty, to appear in *Fuzzy Sets and Systems*, doi:10.1016/j.fss2009.02.007.

[17] Z.Zmeškal, Value at risk methodology of international index portfolio under soft conditions (fuzzy-stochastic approach), *International Review of Financial Analysis*, 14:263-275, 2005.

[18] L.A.Zadeh, Fuzzy sets, *Inform. and Control*, 8:338-353, 1965.

ISBN: 978-989-95079-6-8

Approximations by interval, triangular and trapezoidal fuzzy numbers

Chi-Tsuen Yeh

Department of Mathematics Education, National University of Tainan
33, Sec. 2, Shu-Lin St., Tainan city 70005, Taiwan
Email: ctyeh@mail.nutn.edu.tw

Abstract— *Recently, many scholars investigated interval, triangular, and trapezoidal approximations of fuzzy numbers. These researches can be grouped into two classes: the Euclidean distance class and the non-Euclidean distance class. Most approximations in the Euclidean distance class can be calculated by formulas, but calculating approximations in the other class is more complicated. In this paper, we study interval, triangular, and trapezoidal approximations under a weighted Euclidean distance which generalize all approximations in the Euclidean distance class. First, we embed fuzzy numbers into a Hilbert space, and then introduce these weighted approximations by means of best approximations from closed convex subsets of the Hilbert space. Finally, we apply the reduction principle to simplify calculations of these approximations.*

Keywords— weighted trapezoidal approximation, triangular fuzzy number, Hilbert space

1 Introduction

In practice, fuzzy intervals are often used to represent uncertain or incomplete information. An interesting problem is to approximate general fuzzy intervals by interval, triangular, and trapezoidal fuzzy numbers, so as to simplify calculations. Recently, many scholars investigated these approximations of fuzzy numbers. According to the different aspects of distance, these researches can be grouped into two classes: the Euclidean distance class and the non-Euclidean distance class. The Euclidean distance class includes the interval approximation (proposed by Grzegorzewski in 2002 [11]), symmetric triangular approximation (proposed by Ma et al. in 2000 [18]), trapezoidal approximation (proposed by Abbasbandy and Asady in 2004 [1]), and weighted triangular approximation (proposed by Zeng and Li in 2007 [23]). The non-Euclidean distance class includes the rectangle approximation under the Hamming distance (proposed by Chanas in 2001 [6]), symmetric and non-symmetrical trapezoidal approximations under the Euclidean distance between the respective 1/2-levels (proposed by Delgado et al. in 1998 [7]), and trapezoidal approximation under the source distance (proposed by Abbasbandy and Amirfakhrian in 2006 [2]). Some other approximations are also investigated, such as the nearest parametric approximation (proposed by Nasibova and Peker in 2008 [19]), trapezoidal approximation preserving the expected interval (proposed by Grzegorzewski and Mrówka [12, 13, 14], and improved by Ban [5] and Yeh [22] in 2008, independently), approximation by π functions (proposed by Hassine et al. in 2006 [15]), and polynomial approximation (proposed by Abbasbandy and Amirfakhrian in 2006 [3]). Most approximations in the Euclidean distance class can be calculated by formulas, but calculating the approximations in the other class is more complicated. In this paper, we study interval, triangular, and trapezoidal approximations under a weighted Euclidean distance which generalize all approximations in the Euclidean distance class. In Section 2, we define a weighted L^2-distance on space of fuzzy numbers, and then embed the space into the Hilbert space $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$ by applying the weighted L^2-distance. In Section 3, we introduce weighted approximations of fuzzy numbers by means of best approximations from closed convex subsets of the Hilbert space $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$. Some preliminaries are presented. In Section 4-6, by applying the reduction principle [8, p.80] we compute straightforwardly these approximations of fuzzy numbers, and then propose several important theorems.

2 Embedding fuzzy numbers into the Hilbert space $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$

By an *inner product space* we mean that it is a (real) vector space V equipped with an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ obeying the following axioms:

1. $\langle u, u \rangle \geq 0$ for all $u \in V$, and $\langle u, u \rangle = 0$ iff (if and only if) $u = 0$,
2. $\langle u, v \rangle = \langle v, u \rangle$, for all $u, v \in V$,
3. $\langle au + bv, w \rangle = a\langle u, w \rangle + b\langle v, w \rangle$, for all $u, v, w \in V$ and all $a, b \in \mathbb{R}$.

An inner product is a metric space if the distance is defined by

$$\mathrm{d}(u, v) := \langle u - v, u - v \rangle^{\frac{1}{2}}.$$

A completely inner product space is often called a *Hilbert space*. It is well-known that the set of all L^2-integrable functions is a Hilbert space, denoted by $L^2_\lambda[0,1]$, on which the inner product is defined as

$$\langle f, g \rangle_\lambda := \int_0^1 f(t)g(t)\lambda(t)dt,$$

where $\lambda = \lambda(t)$ is a nonnegative function on $[0,1]$ with $\int_0^1 \lambda(t)dt > 0$.

Another important Hilbert space is the product space $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$, which will be discussed in this paper. Its inner product is defined by

$$\langle (f_1, f_2), (g_1, g_2) \rangle_\lambda := \int_0^1 \left[f_1(t)g_1(t) + f_2(t)g_2(t) \right] \lambda(t)dt$$

ISBN: 978-989-95079-6-8

for all (f_1, f_2) and $(g_1, g_2) \in L^2_\lambda[0,1] \times L^2_\lambda[0,1]$. We hence obtain

$$\begin{aligned}&\mathrm{d}^2_\lambda((f_1,f_2),(g_1,g_2))\\ =&\langle (f_1-g_1, f_2-g_2), (f_1-g_1, f_2-g_2)\rangle_\lambda\\ =&\int_0^1 \left(|f_1(t)-g_1(t)|^2 + |f_2(t)-g_2(t)|^2\right)\lambda(t)dt.\end{aligned}$$

Recall that, a fuzzy number $\tilde{A}$ can be represented by an ordered pair of left continuous functions $[A_L(\alpha), A_U(\alpha)]$ (called the α-cuts of $\tilde{A}$), $0 \le \alpha \le 1$, which satisfy the following conditions: (1) A_L is increasing on [0,1], (2) A_U is decreasing on [0,1], (3) $A_L(1) \le A_U(1)$. Let $\tilde{\mathbb{F}}$ denote the set of all fuzzy numbers. The weighted L^2*-distance* (*Euclidean distance*) on $\tilde{\mathbb{F}}$ is defined as

$$\begin{aligned}\mathrm{d}_\lambda(\tilde{A},\tilde{B}) :=\big[&\int_0^1 |A_L(\alpha)-B_L(\alpha)|^2\lambda(\alpha)d\alpha\\ &+\int_0^1 |A_U(\alpha)-B_U(\alpha)|^2\lambda(\alpha)d\alpha\big]^{\frac{1}{2}}.\end{aligned} \tag{1}$$

For more generality, we refer to [10] in which Grzegorzewski proposed two families of general distances on $\tilde{\mathbb{F}}$. Let $\tilde{A}$ and $\tilde{B}$ be two fuzzy numbers. The fuzzy addition and fuzzy subtraction operations on $\tilde{\mathbb{F}}$ are defined as follows:

$$\begin{aligned}\tilde{A}+\tilde{B} &:= [A_L(\alpha)+B_L(\alpha), A_U(\alpha)+B_U(\alpha)],\\ \tilde{A}-\tilde{B} &:= [A_L(\alpha)-B_U(\alpha), A_U(\alpha)-B_L(\alpha)].\end{aligned}$$

The above conditions (1)-(3) (the definition of fuzzy numbers) imply that A_L and $A_U \in L^2_\lambda[0,1]$, hence we define

$$i : \tilde{A} \mapsto (A_L, A_U) \in L^2_\lambda[0,1] \times L^2_\lambda[0,1].$$

In the following, we always use the interval notation $[A_L, A_U]$ instead of (A_L, A_U), although it may make little sense. Notice that, the fuzzy addition operation coincides with the vector addition on $L^2_\lambda[0,1]\times L^2_\lambda[0,1]$ and its inverse operation (vector subtraction) is not the fuzzy subtraction "-". Let the symbol "$\ominus$" denote the inverse operation, that is

$$\tilde{A}\ominus\tilde{B} := [A_L(\alpha)-B_L(\alpha), A_U(\alpha)-B_U(\alpha)],$$

which is often called the Hukuhara difference, see [17]. In fact, $\tilde{A}\ominus\tilde{B}$ may be not in $\tilde{\mathbb{F}}$. From Eq.(1), we find that

$$\mathrm{d}^2_\lambda(\tilde{A},\tilde{B}) = \langle \tilde{A}\ominus\tilde{B}, \tilde{A}\ominus\tilde{B}\rangle_\lambda.$$

This shows that we can embed space of fuzzy numbers into the Hilbert space $L^2_\lambda[0,1]\times L^2_\lambda[0,1]$. We hence define an inner product on $\tilde{\mathbb{F}}$ inheriting from $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$, that is

$$\langle \tilde{A},\tilde{B}\rangle_\lambda := \int_0^1 \left[A_L(\alpha)B_L(\alpha)+A_U(\alpha)B_U(\alpha)\right]\lambda(\alpha)d\alpha. \tag{2}$$

3 Approximations of fuzzy numbers

Let Ω be a subset of a Hilbert space V, then we call that:

1. Ω is a *subspace* iff $u+v \in \Omega$ and $ru \in \Omega$ for all $u, v \in \Omega$ and all $r \in \mathbb{R}$,
2. Ω is *convex* iff $ru+(1-r)v \in \Omega$ for all $u, v \in \Omega$ and all $r \in [0,1]$,
3. Ω is *chebyshev* iff for each $u \in V$ there exists a unique element $P_\Omega(u) \in \Omega$ such that

$$\mathrm{d}(u, P_\Omega(u)) \le \mathrm{d}(u,x), \quad \forall x \in \Omega,$$

and then $P_\Omega(u)$ is called the *best approximation* of u from Ω.

It is well-known that every closed convex subset (closed subspace, finite dimensional subspace) is chebyshev, see [8, p.23-24]. For any closed convex subset Ω, there is a sufficient and necessary condition of the best approximation $P_\Omega(u)$, as follows

$$\langle u - P_\Omega(u), x - P_\Omega(u)\rangle \le 0, \quad \forall x \in \Omega.$$

Furthermore, we also have

$$\mathrm{d}(P_\Omega(u), P_\Omega(v)) \le \mathrm{d}(u,v), \quad \forall u, v \in V, \tag{3}$$

refer to [21, Appendix C]. This implies P_Ω is continuous. While Ω is a closed subspace, then the above condition becomes

$$\langle u - P_\Omega(u), x\rangle = 0, \quad \forall x \in \Omega,$$

or equivalently

$$\langle u, x\rangle = \langle P_\Omega(u), x\rangle, \quad \forall x \in \Omega. \tag{4}$$

In this paper, all elements in $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$ of the form

$$[r_1 + (r_2 - r_1)\alpha, r_4 - (r_4 - r_3)\alpha]$$

are called *extended trapezoidal* fuzzy numbers. Let $\mathbb{T}$ denote the subset of all extended trapezoidal fuzzy numbers. It is easy to see that, an element $\tilde{A} = [r_1 + (r_2 - r_1)\alpha, r_4 - (r_4 - r_3)\alpha]$ is *trapezoidal* iff $\tilde{A} \in \tilde{\mathbb{F}} \cap \mathbb{T}$, that is

$$r_1 \le r_2 \le r_3 \le r_4. \tag{5}$$

Also, a trapezoidal fuzzy number $\tilde{A}$ is *triangular* (resp. *symmetric trapezoidal*, *symmetric triangular*, *interval*) iff $r_2 = r_3$ (resp. $r_2 - r_1 = r_4 - r_3$, $r_2 = r_3$ and $r_2 - r_1 = r_4 - r_3$, $r_1 = r_2$ and $r_3 = r_4$). Let $\tilde{\mathbb{T}}$, $\tilde{\mathbb{T}}_s$, $\tilde{\Delta}$, $\tilde{\Delta}_s$, and $\tilde{\mathbb{I}}$ denote the sets of all trapezoidal, symmetric trapezoidal, triangular, symmetric triangular, and interval fuzzy numbers, respectively. It is easy to verify that

1. $\mathbb{T}$ is a closed subspace of $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$, and
2. $\tilde{\mathbb{F}}$, $\tilde{\mathbb{T}}$, $\tilde{\mathbb{T}}_s$, $\tilde{\Delta}$, $\tilde{\Delta}_s$, and $\tilde{\mathbb{I}}$ are all closed convex subsets.

Hence, all of them are chebyshev. The best approximations of u from $\mathbb{T}$, $\tilde{\mathbb{T}}$, $\tilde{\mathbb{T}}_s$, $\tilde{\Delta}$, $\tilde{\Delta}_s$, and $\tilde{\mathbb{I}}$ are called the *extended trapezoidal*, *trapezoidal*, *symmetric trapezoidal*, *triangular*, *symmetric triangular*, and *interval approximations* of u, respectively. Eq.(3) implies that these approximations are continuous.

Theorem 1 (The reduction principle [8, p.80])**.** *Let K be a closed convex subset of an inner product space V and M be any chebyshev subspace of V that contains K. Then, we have that*

$$P_K(u) = P_K(P_M(u)) \quad \textit{and}$$
$$\mathrm{d}(u, P_K(u))^2 = \mathrm{d}(u, P_M(u))^2 + \mathrm{d}(P_M(u), P_K(u))^2.$$

ISBN: 978-989-95079-6-8

Now, let's define four extended trapezoidal fuzzy numbers:

$$\tilde{E}_1 := [1-\alpha, 0], \qquad \tilde{E}_2 := [\alpha, 0],$$
$$\tilde{E}_3 := [0, \alpha], \qquad \tilde{E}_4 := [0, 1-\alpha].$$

Then, each element in $\mathbb{T}$ is a linear combination of $\tilde{E}_i$, $1 \leq i \leq 4$, for instance

$$[r_1 + (r_2 - r_1)\alpha, r_4 - (r_4 - r_3)\alpha] = \sum_{i=1}^{4} r_i \tilde{E}_i.$$

This implies $\mathbb{T} = \text{Span}\,\{\tilde{E}_1, \tilde{E}_2, \tilde{E}_3, \tilde{E}_4\}$. We also define two other subspaces of $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$ as follows

$$\Delta := \text{Span}\,\{\tilde{E}_1, \tilde{E}_2 + \tilde{E}_3, \tilde{E}_4\},$$
$$\mathbb{I} := \text{Span}\,\{\tilde{E}_1 + \tilde{E}_2, \tilde{E}_3 + \tilde{E}_4\} = \text{Span}\,\{[1,0], [0,1]\}.$$

It is easy to see that $\tilde{\mathbb{I}} \subset \mathbb{I}$, $\tilde{\Delta}_s \subset \tilde{\Delta} \subset \Delta$, and $\tilde{\mathbb{T}}_s \subset \tilde{\mathbb{T}} \subset \mathbb{T}$. By applying the reduction principle, we obtain that

$$P_{\tilde{\mathbb{I}}}(u) = P_{\tilde{\mathbb{I}}}(P_{\mathbb{I}}(u)), \tag{6}$$
$$P_{\tilde{\Delta}}(u) = P_{\tilde{\Delta}}(P_{\Delta}(u)), \quad P_{\tilde{\Delta}_s}(u) = P_{\tilde{\Delta}_s}(P_{\Delta}(u)), \tag{7}$$
$$P_{\tilde{\mathbb{T}}}(u) = P_{\tilde{\mathbb{T}}}(P_{\mathbb{T}}(u)), \quad P_{\tilde{\mathbb{T}}_s}(u) = P_{\tilde{\mathbb{T}}_s}(P_{\mathbb{T}}(u)). \tag{8}$$

4 The interval approximations

In 2002, Grzegorzewski first proposed interval approximations of fuzzy numbers [11]. Let's extend his results to the case of weighted L^2-distance. We now start with computing the best approximation $P_{\mathbb{I}}(\tilde{A})$ of any fuzzy number $\tilde{A} = [A_L(\alpha), A_U(\alpha)]$ from the subspace $\mathbb{I}$. Unless otherwise stated, we fix the following real numbers:

$$\lambda_0 := \int_0^1 \lambda(\alpha) d\alpha > 0$$

and

$$L_0 := \int_0^1 A_L(\alpha)\lambda(\alpha)d\alpha, \quad U_0 := \int_0^1 A_U(\alpha)\lambda(\alpha)d\alpha.$$

From Eq.(2), we find that $\langle [1,0], [0,1] \rangle_\lambda = 0$ and

$$\langle [1,0], [1,0] \rangle_\lambda = \langle [0,1], [0,1] \rangle_\lambda = \lambda_0.$$

Since $\mathbb{I} = \text{Span}\,\{[1,0], [0,1]\}$, we may assume that

$$P_{\mathbb{I}}(\tilde{A}) = r[1,0] + s[0,1].$$

By applying Eq.(4), we can solve

$$\begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} \langle [1,0],[1,0] \rangle_\lambda & \langle [0,1],[1,0] \rangle_\lambda \\ \langle [1,0],[0,1] \rangle_\lambda & \langle [0,1],[0,1] \rangle_\lambda \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, [1,0] \rangle_\lambda \\ \langle \tilde{A}, [0,1] \rangle_\lambda \end{pmatrix}$$
$$= \lambda_0^{-1} \begin{pmatrix} L_0 \\ U_0 \end{pmatrix}.$$

Hence, we obtain

$$P_{\mathbb{I}}(\tilde{A}) = [\lambda_0^{-1} L_0, \lambda_0^{-1} U_0].$$

The fact $A_L(\alpha) \leq A_U(\alpha)$ implies $L_0 \leq U_0$, hence $P_{\mathbb{I}}(\tilde{A}) \in \tilde{\mathbb{I}}$. By applying Eq.(6), we obtain the following theorem.

Theorem 2. *Let $\tilde{A}$ be a fuzzy number. Then, its interval approximation is $P_{\tilde{\mathbb{I}}}(\tilde{A}) = [\lambda_0^{-1} L_0, \lambda_0^{-1} U_0]$.*

While $\lambda(\alpha) = 1$, we get that $\lambda_0 = 1$. Then, the above equation coincides with the Grzegorzewski's formula [11, Equations (15) and (16)]. Also, the interval $[L_0, U_0]$ is called the *expected interval* of $\tilde{A}$, which is introduced by Dubois and Prade [9] and Heilpern [16], independently.

5 The triangular approximations

Let $\lambda = \lambda(\alpha)$ be a nonnegative function on [0,1] with $\int_0^1 \lambda(\alpha)\, d\alpha > 0$. In what follows, we fix

$$a := \int_0^1 (1-\alpha)^2 \lambda(\alpha) d\alpha > 0,$$
$$b := \int_0^1 \alpha(1-\alpha)\lambda(\alpha) d\alpha > 0,$$
$$c := \int_0^1 \alpha^2 \lambda(\alpha) d\alpha > 0,$$

and

$$L_1 := \int_0^1 A_L(\alpha)\alpha\lambda(\alpha)d\alpha, \quad U_1 := \int_0^1 A_U(\alpha)\alpha\lambda(\alpha)d\alpha.$$

By applying Schwarz inequality, we get

$$ac - b^2 > 0.$$

In a similar manner, we start with computing $P_\Delta(\tilde{A})$. Recall that $\Delta := \text{Span}\,\{\tilde{E}_1, \tilde{E}_2 + \tilde{E}_3, \tilde{E}_4\}$, so we compute

$$\begin{pmatrix} \langle \tilde{E}_1, \tilde{E}_1 \rangle_\lambda & \langle \tilde{E}_2 + \tilde{E}_3, \tilde{E}_1 \rangle_\lambda & \langle \tilde{E}_4, \tilde{E}_1 \rangle_\lambda \\ \langle \tilde{E}_1, \tilde{E}_2 + \tilde{E}_3 \rangle_\lambda & \langle \tilde{E}_2 + \tilde{E}_3, \tilde{E}_2 + \tilde{E}_3 \rangle_\lambda & \langle \tilde{E}_4, \tilde{E}_2 + \tilde{E}_3 \rangle_\lambda \\ \langle \tilde{E}_1, \tilde{E}_4 \rangle_\lambda & \langle \tilde{E}_2 + \tilde{E}_3, \tilde{E}_4 \rangle_\lambda & \langle \tilde{E}_4, \tilde{E}_4 \rangle_\lambda \end{pmatrix}$$
$$= \begin{pmatrix} a & b & 0 \\ b & 2c & b \\ 0 & b & a \end{pmatrix}.$$

Let $P_\Delta(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4$. By applying Eq.(4), we can solve

$$\begin{pmatrix} r_1 \\ r_2 \\ r_4 \end{pmatrix} = \begin{pmatrix} a & b & 0 \\ b & 2c & b \\ 0 & b & a \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, \tilde{E}_1 \rangle_\lambda \\ \langle \tilde{A}, \tilde{E}_2 + \tilde{E}_3 \rangle_\lambda \\ \langle \tilde{A}, \tilde{E}_4 \rangle_\lambda \end{pmatrix}$$
$$= \frac{1}{\delta} \begin{pmatrix} 2ac - b^2 & -ab & b^2 \\ -ab & a^2 & -ab \\ b^2 & -ab & 2ac - b^2 \end{pmatrix} \begin{pmatrix} L_0 - L_1 \\ L_1 + U_1 \\ U_0 - U_1 \end{pmatrix}. \tag{9}$$

where $\delta = 2a(ac - b^2) > 0$. If we assume $\int_0^1 \lambda(\alpha)d\alpha = \frac{1}{2}$ (that is $a + 2b + c = \frac{1}{2}$) additionally, then the above $P_\Delta(\tilde{A})$ is equal to Zeng and Li's weighted triangular approximation [23]. Notice that $P_\Delta(\tilde{A})$ may be not in $\tilde{\mathbb{F}}$, refer to [21]. That shows that $P_{\tilde{\Delta}}(\tilde{A}) \neq P_\Delta(\tilde{A})$.

Lemma 3. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_\Delta(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4,$$

where r_1, r_2, and r_4 are computed by Eq.(9). Then,

1. $r_1 \leq r_4$,
2. *if $r_2 \leq r_1$, then*

$$-(a+b)L_0 + (a+3b+2c)U_0 - 2(a+2b+c)U_1 \geq 0,$$

3. *if $r_2 \geq r_4$, then*

$$-(a+3b+2c)L_0 + 2(a+2b+c)L_1 + (a+b)U_0 \geq 0.$$

ISBN: 978-989-95079-6-8

Proof. Omitted.

From Eq.(7), we find $P_{\tilde{\Delta}}(\tilde{A}) = P_{\tilde{\Delta}}(P_{\Delta}(\tilde{A}))$. Eq.(5) implies that an element

$$r\tilde{E}_1 + s(\tilde{E}_2 + \tilde{E}_3) + t\tilde{E}_4 \in \Delta$$

is triangular iff $r \le s \le t$. By Lemma 3.1, we obtain that $P_{\Delta}(\tilde{A}) \notin \tilde{\Delta}$ implies either $r_2 < r_1$ or $r_2 > r_4$, hence the best approximation (triangular approximation)

$$P_{\tilde{\Delta}}(\tilde{A}) := r_1'\tilde{E}_1 + r_2'(\tilde{E}_2 + \tilde{E}_3) + r_4'\tilde{E}_4$$

will satisfy $r_2' = r_1'$ or $r_2' = r_4'$, respectively. For instance, suppose that the fuzzy number $\tilde{A}$ has the approximation

$$P_{\Delta}(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4 \quad \text{with } r_2 < r_1.$$

Then, $P_{\tilde{\Delta}}(\tilde{A})$ will belong to Span $\{\tilde{E}_1 + \tilde{E}_2 + \tilde{E}_3, \tilde{E}_4\}$, since $r_2' = r_1'$. So, we consider the best approximation of $\tilde{A}$ from Span $\{\tilde{E}_1 + \tilde{E}_2 + \tilde{E}_3, \tilde{E}_4\}$. We hence let

$$P_{\tilde{\Delta}}(\tilde{A}) = r_1'(\tilde{E}_1+\tilde{E}_2+\tilde{E}_3)+r_4'\tilde{E}_4 = r_1'[1,\alpha]+r_4'[0,1-\alpha].$$

By applying Eq.(4), we can solve

$$\begin{aligned}\begin{pmatrix} r_1' \\ r_4' \end{pmatrix} &= \begin{pmatrix} a+2b+2c & b \\ b & a \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, [1,\alpha] \rangle_\lambda \\ \langle \tilde{A}, [0,1-\alpha] \rangle_\lambda \end{pmatrix} \\ &= \frac{1}{\delta} \begin{pmatrix} a & -b \\ -b & a+2b+2c \end{pmatrix} \begin{pmatrix} L_0 + U_1 \\ U_0 - U_1 \end{pmatrix},\end{aligned} \tag{10}$$

where $\delta = (a+b)^2 + 2(ac - b^2)$, and we have substituted by

$$\langle [1,\alpha], [1,\alpha] \rangle_\lambda = \int_0^1 (1+\alpha^2)\lambda(\alpha)d\alpha = a + 2b + 2c.$$

Notice that, the extended trapezoidal fuzzy number

$$\tilde{E}_1 + \tilde{E}_2 + \tilde{E}_3 = [1,\alpha]$$

does not belong to $\tilde{\mathbb{F}}$. But, this does not effect results of our computation. Applying Lemma 3.2, the reader can easily verify $r_1' \le r_4'$ in Eq.(10), so that

$$r_1'[1-\alpha, 0] + r_4'[\alpha, 1] \in \tilde{\mathbb{F}}.$$

Hence, we obtain

$$P_{\tilde{\Delta}}(\tilde{A}) = r_1'[1,\alpha] + r_4'[0,1-\alpha],$$

where r_1' and r_4' are computed by Eq.(10).

On the other hand, if $\tilde{A}$ has the approximation

$$P_{\Delta}(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4 \quad \text{with } r_2 > r_4$$

then its triangular approximation leads to

$$P_{\tilde{\Delta}}(\tilde{A}) = r_1'[1-\alpha, 0] + r_4'[\alpha, 1],$$

where

$$\begin{aligned}\begin{pmatrix} r_1' \\ r_4' \end{pmatrix} &= \begin{pmatrix} a & b \\ b & a+2b+2c \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, [1-\alpha, 0] \rangle_\lambda \\ \langle \tilde{A}, [\alpha, 1] \rangle_\lambda \end{pmatrix} \\ &= \frac{1}{\delta} \begin{pmatrix} a+2b+2c & -b \\ -b & a \end{pmatrix} \begin{pmatrix} L_0 - L_1 \\ L_1 + U_0 \end{pmatrix}.\end{aligned} \tag{11}$$

where $\delta = (a+b)^2 + 2(ac-b^2)$. Again, applying Lemma 3.3 the reader can easily verify $r_1' \le r_4'$ in Eq.(11).

Theorem 4. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_{\Delta}(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4,$$

where r_1, r_2 and r_4 are computed by Eq.(9). Then, the triangular approximation $P_{\tilde{\Delta}}(\tilde{A})$ can be determined in the following cases:

1. *If $r_1 \le r_2 \le r_4$, then*
$$P_{\tilde{\Delta}}(\tilde{A}) = [r_1 + (r_2 - r_1)\alpha, r_4 - (r_4 - r_2)\alpha].$$
2. *If $r_2 < r_1$, then $P_{\tilde{\Delta}}(\tilde{A}) = [r_1', r_4' - (r_4' - r_1')\alpha]$, where r_1' and r_4' are computed by Eq.(10).*
3. *If $r_2 > r_4$, then $P_{\tilde{\Delta}}(\tilde{A}) = [r_1' + (r_4' - r_1')\alpha, r_4']$, where r_1' and r_4' are computed by Eq.(11).*

Next, we compute the symmetric triangular approximation $P_{\tilde{\Delta}_s}(\tilde{A})$ which was first proposed by Ma et al. [18]. Recall that, an extended trapezoidal element

$$r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4 \in \Delta$$

is symmetric triangular iff

$$r_2 - r_1 = r_4 - r_2 \ge 0.$$

Hence, by substituting $r_2 = \frac{1}{2}(r_1 + r_4)$ we get

$$r_1\tilde{E}_1+r_2(\tilde{E}_2+\tilde{E}_3)+r_4\tilde{E}_4 = r_1[1-\frac{1}{2}\alpha, \frac{1}{2}\alpha]+r_4[\frac{1}{2}\alpha, 1-\frac{1}{2}\alpha].$$

Let's consider the best approximation $P_{\Delta_s}(\tilde{A})$ of $\tilde{A}$ from the subspace Δ_s, where Δ_s is defined by

$$\Delta_s := \text{Span } \{[1-\frac{1}{2}\alpha, \frac{1}{2}\alpha], [\frac{1}{2}\alpha, 1-\frac{1}{2}\alpha]\}.$$

Suppose that

$$P_{\Delta_s}(\tilde{A}) = r_1[1-\frac{1}{2}\alpha, \frac{1}{2}\alpha] + r_4[\frac{1}{2}\alpha, 1-\frac{1}{2}\alpha].$$

By applying Eq.(4), we can solve

$$\begin{aligned}&\begin{pmatrix} r_1 \\ r_4 \end{pmatrix} \\ &= \begin{pmatrix} a+b+\frac{c}{2} & b+\frac{c}{2} \\ b+\frac{c}{2} & a+b+\frac{c}{2} \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, [1-\frac{1}{2}\alpha, \frac{1}{2}\alpha] \rangle_\lambda \\ \langle \tilde{A}, [\frac{1}{2}\alpha, 1-\frac{1}{2}\alpha] \rangle_\lambda \end{pmatrix} \\ &= \frac{1}{\delta} \begin{pmatrix} a+b+\frac{c}{2} & -(b+\frac{c}{2}) \\ -(b+\frac{c}{2}) & a+b+\frac{c}{2} \end{pmatrix} \begin{pmatrix} L_0 - \frac{1}{2}L_1 + \frac{1}{2}U_1 \\ \frac{1}{2}L_1 + U_0 - \frac{1}{2}U_1 \end{pmatrix},\end{aligned} \tag{12}$$

where $\delta = a^2 + 2ab + ac$. The reader can verify that $r_1 \le r_4$ for any $\tilde{A} \in \tilde{\mathbb{F}}$. This implies $P_{\Delta_s}(\tilde{A}) \in \tilde{\mathbb{F}}$, so that

$$P_{\tilde{\Delta}_s}(\tilde{A}) = P_{\Delta_s}(\tilde{A}).$$

Theorem 5. *Let $\tilde{A}$ be a fuzzy number. Then, its symmetric approximation is*

$$P_{\tilde{\Delta}_s}(\tilde{A}) = [r_1 + \frac{r_4 - r_1}{2}\alpha, r_4 - \frac{r_4 - r_1}{2}\alpha],$$

where r_1 and r_4 are computed by Eq.(12).

While $\lambda(\alpha) = 1$, we get that, $a = \frac{1}{3}$, $b = \frac{1}{6}$, and $c = \frac{1}{3}$. Substituting into the above equation, we obtain

$$P_{\tilde{\Delta}_s}(\tilde{A}) = [x_0 - \sigma(1-\alpha), x_0 + \sigma(1-\alpha)],$$

where $x_0 = \frac{L_0+U_0}{2}$ and $\sigma = \frac{3}{2}(-L_0 + L_1 + U_0 - U_1)$. This formula coincides with [18, Equations (8) and (9)].

ISBN: 978-989-95079-6-8

6 The trapezoidal approximations

The (extended) trapezoidal approximation was first proposed by Abbasbandy and Asady in 2004 [1]. Afterwards, Grzegorzewski and Mrówka proposed the (extended) trapezoidal approximation preserving the expected interval [12]. Since the expected interval of any fuzzy number $\tilde{A}$ is equal to $P_{\mathbb{I}}(\tilde{A})$ and $\mathbb{I} \supseteq \mathbb{T} \supseteq \tilde{\mathbb{T}}$, by the reduction principle we get that these two (extended) trapezoidal approximations are equal. Now, we start with computing the extended trapezoidal approximation $P_{\mathbb{T}}(\tilde{A})$.

Because that $\mathbb{T} = \text{Span}\,\{\tilde{E}_1, \tilde{E}_2, \tilde{E}_3, \tilde{E}_4\}$, we let

$$P_{\mathbb{T}}(\tilde{A}) = t_1\tilde{E}_1 + t_2\tilde{E}_2 + t_3\tilde{E}_3 + t_4\tilde{E}_4.$$

In the same vein, by applying Eq.(4) we can solve

$$\begin{aligned}
\begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{pmatrix} &= \begin{pmatrix} a & b & 0 & 0 \\ b & c & 0 & 0 \\ 0 & 0 & c & b \\ 0 & 0 & b & a \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, \tilde{E}_1 \rangle_\lambda \\ \langle \tilde{A}, \tilde{E}_2 \rangle_\lambda \\ \langle \tilde{A}, \tilde{E}_3 \rangle_\lambda \\ \langle \tilde{A}, \tilde{E}_4 \rangle_\lambda \end{pmatrix} \\
&= \frac{1}{ac - b^2} \begin{pmatrix} c & -b & 0 & 0 \\ -b & a & 0 & 0 \\ 0 & 0 & a & -b \\ 0 & 0 & -b & c \end{pmatrix} \begin{pmatrix} L_0 - L_1 \\ L_1 \\ U_1 \\ U_0 - U_1 \end{pmatrix} \qquad (13) \\
&= \frac{1}{ac - b^2} \begin{pmatrix} cL_0 - (b+c)L_1 \\ -bL_0 + (a+b)L_1 \\ -bU_0 + (a+b)U_1 \\ cU_0 - (b+c)U_1 \end{pmatrix}.
\end{aligned}$$

While $\lambda(\alpha) = 1$, the above equation coincides with Grzegorzewski's formula [12, Equations (29)-(32)]. Note that, the extended trapezoidal approximation $P_{\mathbb{T}}(\tilde{A})$ may be not in $\tilde{\mathbb{F}}$, refer to [4, 5, 20]. This is because that $P_{\mathbb{T}}(\tilde{A})$ may happen $t_2 > t_3$.

Lemma 6. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_{\mathbb{T}}(\tilde{A}) = t_1\tilde{E}_1 + t_2\tilde{E}_2 + t_3\tilde{E}_3 + t_4\tilde{E}_4,$$

where t_i, $1 \le i \le 4$, are computed by Eq.(13). Then,

$$t_1 \le t_2 \quad \textit{and} \quad t_3 \le t_4.$$

Proof. Omitted.

From Eq.(8), we find

$$P_{\tilde{\mathbb{T}}}(\tilde{A}) = P_{\tilde{\mathbb{T}}}(P_{\mathbb{T}}(\tilde{A})). \qquad (14)$$

Hence, if $P_{\mathbb{T}}(\tilde{A})$ is in $\tilde{\mathbb{F}}$ (by applying Lemma 6, it is equivalent to $t_2 \le t_3$), we obtain

$$P_{\tilde{\mathbb{T}}}(\tilde{A}) = P_{\mathbb{T}}(\tilde{A}).$$

Otherwise, we have $t_2 > t_3$. Consequently, Eq.(14) implies that the trapezoidal approximation $P_{\tilde{\mathbb{T}}}(\tilde{A})$ will be restricted to triangular fuzzy numbers $\tilde{\Delta}$. This leads to

$$P_{\tilde{\mathbb{T}}}(\tilde{A}) = P_{\tilde{\Delta}}(\tilde{A}).$$

By applying Theorem 4, we prove the following theorem which is a generalization of [21, Theorem 4.4].

Theorem 7. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_{\mathbb{T}}(\tilde{A}) = t_1\tilde{E}_1 + t_2\tilde{E}_2 + t_3\tilde{E}_3 + t_4\tilde{E}_4,$$

where t_i, $1 \le i \le 4$, are computed by Eq.(13). If $t_2 \le t_3$, then the trapezoidal approximation of $\tilde{A}$ is

$$P_{\tilde{\mathbb{T}}}(\tilde{A}) = [t_1 + (t_2 - t_1)\alpha, t_4 - (t_4 - t_3)\alpha].$$

Otherwise, by Eq.(9) compute

$$P_{\Delta}(\tilde{A}) = r_1\tilde{E}_1 + r_2(\tilde{E}_2 + \tilde{E}_3) + r_4\tilde{E}_4.$$

Then, the trapezoidal approximation $P_{\tilde{\mathbb{T}}}(\tilde{A})$ can be determined in the following cases:

1. *If $r_1 \le r_2 \le r_4$, then*
$$P_{\tilde{\mathbb{T}}}(\tilde{A}) = [r_1 + (r_2 - r_1)\alpha, r_4 - (r_4 - r_2)\alpha].$$
2. *If $r_2 < r_1$, then $P_{\tilde{\mathbb{T}}}(\tilde{A}) = [r'_1, r'_4 - (r'_4 - r'_1)\alpha]$, where r'_1 and r'_4 are computed by Eq.(10).*
3. *If $r_2 > r_4$, then $P_{\tilde{\mathbb{T}}}(\tilde{A}) = [r'_1 + (r'_4 - r'_1)\alpha, r'_4]$, where r'_1 and r'_4 are computed by Eq.(11).*

Proof. Omitted.

In 1998, Delgado et al. [7] proposed a symmetric trapezoidal approximation of $\tilde{A}$ under the Euclidean distance between the respective 1/2-levels. In the following, we turn to study the symmetric trapezoidal approximation of $\tilde{A}$ under a weighted L^2-distance. Recall that, an extended trapezoidal fuzzy number $t_1\tilde{E}_1 + t_2\tilde{E}_2 + t_3\tilde{E}_3 + t_4\tilde{E}_4$ is symmetric trapezoidal iff

$$t_2 - t_1 = t_4 - t_3 \ge 0 \quad \text{and} \quad t_2 \le t_3.$$

Let $d = t_2 - t_1$. By substituting $t_2 = t_1 + d$ and $t_3 = t_4 - d$, we obtain

$$t_1\tilde{E}_1 + t_2\tilde{E}_2 + t_3\tilde{E}_3 + t_4\tilde{E}_4 = t_1[1,0] + d[\alpha, -\alpha] + t_4[0,1].$$

Let $\mathbb{T}_s := \text{Span}\,\{[1,0], [\alpha, -\alpha], [0,1]\}$. Suppose that the best approximation of $\tilde{A}$ from $\mathbb{T}_s$ is

$$P_{\mathbb{T}_s}(\tilde{A}) := t_1[1,0] + d[\alpha, -\alpha] + t_4[0,1].$$

By applying Eq.(4), we can solve

$$\begin{aligned}
\begin{pmatrix} t_1 \\ d \\ t_4 \end{pmatrix} &= \begin{pmatrix} \lambda_0 & \lambda_1 & 0 \\ \lambda_1 & 2\lambda_2 & -\lambda_1 \\ 0 & -\lambda_1 & \lambda_0 \end{pmatrix}^{-1} \begin{pmatrix} \langle \tilde{A}, [1,0] \rangle_\lambda \\ \langle \tilde{A}, [\alpha, -\alpha] \rangle_\lambda \\ \langle \tilde{A}, [0,1] \rangle_\lambda \end{pmatrix} \\
&= \frac{1}{\delta} \begin{pmatrix} 2\lambda_0\lambda_2 - \lambda_1^2 & -\lambda_0\lambda_1 & -\lambda_1^2 \\ -\lambda_0\lambda_1 & \lambda_0^2 & \lambda_0\lambda_1 \\ -\lambda_1^2 & \lambda_0\lambda_1 & 2\lambda_0\lambda_2 - \lambda_1^2 \end{pmatrix} \begin{pmatrix} L_0 \\ L_1 - U_1 \\ U_0 \end{pmatrix} \qquad (15)
\end{aligned}$$

where $\lambda_i := \int_0^1 \alpha^i \lambda(\alpha) d\alpha$, $i = 0, 1, 2$, and

$$\delta := 2\lambda_0(\lambda_0\lambda_2 - \lambda_1^2) > 0.$$

Lemma 8. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_{\mathbb{T}_s}(\tilde{A}) = t_1[1,0] + d[\alpha, -\alpha] + t_4[0,1],$$

where t_1, d, and t_4 are computed by Eq.(15). Then, $d \ge 0$.

ISBN: 978-989-95079-6-8

Proof. Omitted.

If $P_{\mathbb{T}_s}(\tilde{A}) \in \tilde{\mathbb{F}}$ (by applying Lemma 8, it is equivalent to $t_1 + d \leq t_4 - d$), then it equals the symmetric trapezoidal approximation $P_{\tilde{\mathbb{T}}_s}$. Otherwise, we will have $d > \frac{1}{2}(t_4 - t_1)$. This shows that, in the case the symmetric trapezoidal approximation

$$P_{\tilde{\mathbb{T}}_s}(\tilde{A}) = t_1'[1,0] + d'[\alpha, -\alpha] + t_4'[0,1]. \qquad (16)$$

will satisfy $d' = \frac{1}{2}(t_4' - t_1')$. Substituting into Eq.(16) by $d' = \frac{1}{2}(t_4' - t_1')$, we get

$$P_{\tilde{\mathbb{T}}_s}(\tilde{A}) := t_1'[1 - \frac{1}{2}\alpha, \frac{1}{2}\alpha] + t_4'[\frac{1}{2}\alpha, 1 - \frac{1}{2}\alpha].$$

Since $\Delta_s = \text{Span}\,\{[1 - \frac{1}{2}\alpha, \frac{1}{2}\alpha], [\frac{1}{2}\alpha, 1 - \frac{1}{2}\alpha]\}$, we get

$$P_{\tilde{\mathbb{T}}_s}(\tilde{A}) \in \Delta_s \cap \tilde{\mathbb{F}} = \tilde{\Delta}_s.$$

This implies $P_{\tilde{\mathbb{T}}_s}(\tilde{A})$ is equal to the symmetric triangular approximation $P_{\tilde{\Delta}_s}(\tilde{A})$. Consequently, by applying Theorem 5 we obtain the following theorem.

Theorem 9. *Let $\tilde{A}$ be a fuzzy number, and let*

$$P_{\mathbb{T}_s}(\tilde{A}) = t_1[1,0] + d[\alpha, -\alpha] + t_4[0,1],$$

where t_1, d, and t_4 are computed by Eq.(15). If $d \leq \frac{1}{2}(t_4 - t_1)$, then the symmetric trapezoidal approximation of $\tilde{A}$ is

$$P_{\tilde{\mathbb{T}}_s}(\tilde{A}) = [t_1 + d\alpha, t_4 - d\alpha].$$

Otherwise, it is

$$P_{\tilde{\mathbb{T}}_s}(\tilde{A}) = [r_1 + \frac{r_4 - r_1}{2}\alpha, r_4 - \frac{r_4 - r_1}{2}\alpha],$$

where r_1 and r_4 are computed by Eq.(12).

7 Conclusions

Recently, many scholars studied on computation of interval, triangular, and trapezoidal approximations approximations of fuzzy numbers by applying Langrange multiplier method or Karush-Kuhn-Tucker theorem. In the present paper, we propose a new method for computing these approximations under a weighted distance by applying function approximation theory on Hilbert space.We embed fuzzy numbers into the Hilbert space $L^2_\lambda[0,1] \times L^2_\lambda[0,1]$. Then, by introducing extended trapezoidal fuzzy numbers and applying the reduction principle (Theorem 1), it suffices to compute the best approximations of an extended trapezoidal fuzzy number. Hence, we can easily determine these approximations by choosing a suitable basis. In fact, the weighted distance can be generalized to more general form, as follows

$$\mathrm{d}_\lambda(\tilde{A}, \tilde{B}) = \int_0^1 |A_L - B_L|^2 d\mu_1 + \int_0^1 |A_U - B_U|^2 d\mu_2,$$

where μ_1 and μ_2 are arbitrary positive measures on [0,1].

Acknowledgment

The author is very grateful to Prof. P. Grzegorzewski and Prof. L. Stefanini for their invitation and suggestions, which have been very helpful in improving the presentation of the paper. The author also wishes to thank my advisor Prof. Chen-Lian Chuang for his assistance. This work was supported by National Science Council of Taiwan.

References

[1] S. Abbasbandy and B. Asady, The nearest trapezoidal fuzzy number to a fuzzy quantity, Applied Mathematics and Computation 156(2004) 381-386.

[2] S. Abbasbandy and M. Amirfakhrian, The nearest trapezoidal form of a generalized left right fuzzy number, International Journal of Approximate Reasoning 43(2006) 166-178.

[3] S. Abbasbandy and M. Amirfakhrian, The nearest approximation of a fuzzy quantity in parametric form, Applied Mathematics and Computation 172(2006) 624-632.

[4] T. Allahviranloo and M.A. Firozja, Note on "Trapezoidal approximation of fuzzy numbers", Fuzzy Sets and Systems 158(2007) 747-754.

[5] A. Ban, Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the expected interval, Fuzzy Sets and Systems, 159(2008) 1327-1344.

[6] S. Chanas, On the interval approximation of a fuzzy number, Fuzzy Sets and Systems 122(2001) 353-356.

[7] M. Delgado, M.A. Vila and W. Voxman, On a canonical representation of fuzzy number, Fuzzy Sets and Systems 93(1998) 125-135.

[8] F. Deutsch, Best approximation in inner product space, Springer, New York, 2001.

[9] D. Dubois and H. Prade, The mean value of a fuzzy number, Fuzzy Sets and Systems 24(1987) 279-300.

[10] P. Grzegorzewski, Metrics and orders in space of fuzzy numbers, Fuzzy Sets and Systems 97(1998) 83-94.

[11] P. Grzegorzewski, Nearest interval approximation of a fuzzy number, Fuzzy Sets and Systems 130(2002) 321-330.

[12] P. Grzegorzewski and E. Mrówka, Trapezoidal approximations of fuzzy numbers, Fuzzy Sets and Systems 153(2005) 115-135.

[13] P. Grzegorzewski and E. Mrówka, Trapezoidal approximations of fuzzy numbers-revisited, Fuzzy Sets and Systems 158(2007) 757-768.

[14] P. Grzegorzewski and E. Mrówka, Trapezoidal approximations of fuzzy numbers preserving the expected interval - algorithms and properties, Fuzzy Sets and Systems, 159(2008) 1354-1364.

[15] R. Hassine, F. Karray, A.M. Alimi and M. Selmi, Approximation properties of piece-wise parabolic functions fuzzy logic systems, Fuzzy Sets and Systems 157(2006) 501-515.

[16] S. Heilpern, The expected value of a fuzzy number, Fuzzy Sets and Systems 47(1992) 81-86.

[17] M. Hukuhara, Integration des applications measurables dont la valeur est un compact convexe, Funkcialaj Ekvacioj 10(1967) 205-223.

[18] M. Ma, A. Kandel and M. Friedman, A new approach for defuzzification, Fuzzy Sets and Systems 111(2000) 351-356.

[19] E. N. Nasibov and S. Peker, On the nearest parametric approximation of a fuzzy number, Fuzzy Sets and Systems, 159(2008) 1365-1375.

[20] C.T. Yeh, A note on trapezoidal approximations of fuzzy numbers, Fuzzy Sets and Systems 158(2007) 747-754.

[21] C.T. Yeh, On improving trapezoidal and triangular approximations of fuzzy numbers, International Journal of Approximate Reasoning, 48/1 (2008) 297-313.

[22] C.T. Yeh, Trapezoidal and triangular approximations preserving the expected interval, Fuzzy Sets and Systems, 159(2008) 1345-1353.

[23] W. Zeng and H. Li, Weighted triangular approximation of fuzzy numbers, International Journal of Approximate Reasoning, 46/1 (2007) 137-150.

ISBN: 978-989-95079-6-8

Fuzzy c–Means Herding

Thomas A. Runkler

Siemens Corporate Technology, Munich, Germany
Email: Thomas.Runkler@siemens.com

Abstract— *Herding is the process of bringing individuals (e.g. animals) together into a group. More specifically, we consider self–organized herding as the process of moving a set of individuals to a given number of locations (cluster centers) without any external control. We formally describe the relation between herding and clustering and show that any clustering model can be used to control herding processes. For the specific case of the fuzzy c–means model we derive the equations of the fuzzy c–means herding algorithm using a gradient descent approach with limited step size. Several experiments related to an autonomous mobile robot scenario show that fuzzy c–means herding yields smooth trajectories, well–balanced clusters, and fast convergence.*

Keywords— Fuzzy clustering, herding, robot swarms, swarm intelligence.

1 Introduction

This paper deals with the problem of *online herding* of n individuals specified by the state vectors $x_1, \ldots, x_n \in \mathbb{R}^p$, into c clusters. Application examples for herding are illustrated by the following four scenarios:

1. A swarm of autonomous micro–robots explores the Mars surface. Just before their individual energy supplies are exhausted, the robots move to a couple of individual locations from where they are collected and returned to the spaceship.
2. Museum visitors individually walk through an exhibition. At the beginning of the guided tours, each visitor receives instructions where to go to join one of the tour groups.
3. Mountain hikers have been caught by a severe thunderstorm. They receive instructions on their mobile phones where to go, so that they can be picked off from several evacuation points by helicopter.
4. Wireless sensor devices are attached to individual components of an industrial plant for condition monitoring purposes. At power on, each sensor automatically adapts its communication protocol so that several clusters of networked sensors emerge.

Herding mobile robots (as in our first scenario) is a very active field in the robotics community. A neuro–fuzzy control approach for herding (also called convoying) of multiple robots in presented in [1]. A similar approach to herding very heterogeneous types of robots is presented in [2]. Also herding in evacuation processes (as in our third scenario) is currently actively discussed in the scientific community, for example in [3]. Herding has also been applied to swarm based learning processes in order to balance exploration and exploitation behavior [4]. Some work has also been done concerning herding using outside agents [5], for example the so–called *dog and sheep* herding [6]. In this paper, however, we only consider *self–organizing* herding, where all individuals equally contribute to the herding process, without any external control.

In each of the four scenarios presented above, the individual state vectors $x_1, \ldots, x_n$ (for example the locations of the micro–robots) are successively updated, until a clustering is achieved, i.e. each of the individuals x_k is (fuzzily) assigned to cluster i with a membership $u_{ik} \in [0, 1]$. When the herding process is converged, each state vector is equal to the prototypical state $v_i \in \mathbb{R}^p$, $i = 1, \ldots, c$, of its corresponding cluster, and is crisply assigned to this cluster, $u_{ik} = 1$ and $u_{jk} = 0$ for all $j \neq i$. More formally, herding can be described by a sequence

$$X \rightarrow (U, V, X) \rightarrow (U, V, X) \rightarrow (U, V, X) \rightarrow \ldots \quad (1)$$

which converges to $X = V$ and $U \in M_{hcn}$,

$$M_{hcn} = \left\{ \; U \in \{0,1\}^{c \times n} \;\middle|\; \sum_{i=1}^{c} u_{ik} = 1, \quad k = 1, \ldots, n, \quad \sum_{k=1}^{n} u_{ik} > 0, \quad i = 1, \ldots, c \; \right\} \quad (2)$$

In practical scenarios we will not be able to run this algorithm for an infinite time until convergence, but will terminate after a previously specified number of t steps, or when an appropriate termination criterion holds, for example when the maximum change in the state vectors is lower than a previously specified threshold. Even though the limit of the final partition is crisp and somewhat trivial (because $X = V$), we pursue a fuzzy approach here to *smoothly* control the herding process.

Apparently, herding is closely related to clustering. A discussion of the relation between herding and clustering from a theoretical physics perspective can be found in [7]. Just as herding, clustering also assigns each object to a cluster, and each cluster can be described by a prototype as well, but clustering does not change the data vectors X. More formally, in contrast to (1) clustering is described by a sequence

$$X \rightarrow (U, V) \rightarrow (U, V) \rightarrow (U, V) \rightarrow \ldots \quad (3)$$

where the initial X remains constant. Despite this difference between clustering and herding, the herding process can be performed by optimizing *any* conventional clustering model. The main difference is that in clustering we consider (U, V)

ISBN: 978-989-95079-6-8

as the set of free variables for optimization, and in herding we consider (U, V, X). In this paper we illustrate this for the case of the *fuzzy c–means* [8] clustering model. Other suitable clustering models are, for example, hard c–means [9], possibilistic c–means [10], noise clustering [11], or high–dimensional fuzzy c–means [12].

This paper is structured as follows: Sections 2 and 3 briefly repeat the fuzzy c–means objective function and how this model is applied to clustering. Section 4 shows how the fuzzy c–means model can be used to control a herding process, leading to the fuzzy c–means herding algorithm. Section 5 presents experiments with the fuzzy c–means herding algorithm referring to the Mars robot scenario illustrated above. Section 6 finally summarizes the conclusions and gives an outline for future work in this field.

2 The Fuzzy c–Means Objective Function

The *fuzzy c–means* objective function [8] is a dissimilarity–based measure for the clustering of a data set $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^p$ into $c \in \{2, \ldots, n-1\}$ clusters specified by a partition matrix $U \in M_{fcn}$ where

$$M_{fcn} = \left\{ \; U \in [0,1]^{c \times n} \;\middle|\; \sum_{i=1}^{c} u_{ik} = 1, \quad k = 1, \ldots, n, \quad \sum_{k=1}^{n} u_{ik} > 0, \quad i = 1, \ldots, c \; \right\} \tag{4}$$

and a set of prototypes (cluster centers) $V = \{v_1, \ldots, v_c\} \subset \mathbb{R}^p$. The FCM objective function is the sum of the dissimilarities between data and prototypes, weighted by the (fuzzy) memberships. Using a fuzziness index $m > 1$, it is formally defined as

$$J(U, V, X, c, m) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m \, |x_k - v_i|^2. \tag{5}$$

For $m \to 1$ this objective function converges to the *hard c–means objective function* [9]. For many applications good results were reported for $m = 2$. In this paper we always use the Euclidean distance $|.| = \|.\|$. Lower values of J indicate a better clustering of (U, V, X). The theoretical minimum of J is zero. This minimum is achieved, if for each $k = 1, \ldots, n$ we have an $i \in \{1, \ldots, c\}$ with $x_k = v_i$, $u_{ik} = 1$ and $u_{jk} = 0$ for all $j \neq i$, i.e. $X = V$ and $U \in M_{hcn}$. This means that herding converges to $J = 0$.

3 Fuzzy c–Means Clustering

As pointed out above we assume X to be given and fixed in clustering and only find U and V. In fuzzy c–means clustering we find U and V by minimizing J. In clustering we are usually not able to reduce J to zero. We may use different optimization methods such as genetic algorithms [13, 14], particle swarm optimization [15, 16, 17], ant colony optimization [18], or wasp swarm optimization [19]. In this paper we apply direct minimization using the necessary conditions for local extrema or saddle points of J. An equation to compute the optimal U for a given V under the constraints in (4) is obtained by defining the Langrangian

$$L = J - \sum_{k=1}^{n} \lambda_k \left(1 - \sum_{i=1}^{c} u_{ik} \right) \tag{6}$$

and setting $\partial L / \partial u_{ik} = 0$, from which follows

$$u_{ik} = 1 \Big/ \sum_{j=1}^{c} \left(\frac{|x_k - v_i|}{|x_k - v_j|} \right)^{\frac{2}{m-1}}, \tag{7}$$

$i = 1, \ldots, c$, $k = 1, \ldots, n$. An equation to compute the optimal V for a given U is obtained by setting $\partial J / \partial v_i = 0$, from which follows

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m}, \tag{8}$$

$i = 1, \ldots, c$. The equations for V depend on U and the equations for U depend on V, so optimization can be done in an alternating scheme, the so–called alternating optimization. The clustering algorithm terminates when an appropriate termination criterion holds, e.g. when successive changes in V are lower than a threshold ε. The fuzzy c–means clustering algorithm can then be formally defined as follows:

Algorithm 1: Fuzzy c–Means Clustering.

```
input X, c, m
randomly initialize V ∈ R^p
repeat
   update U(V, X, m) by (7)
   update V(U, X, m) by (8)
until max max ΔV < ε
output U, V
```

4 Fuzzy c–Means Herding

As pointed out in the introduction, herding does not only update U and V at each optimization step but also X. If we do the optimization of $J(U, V, X)$ in an alternating scheme, then we also have to determine an appropriate update equation for X. In the scenarios described above we do not arbitrarily change the values of X (for example the position of the robots) in each step but we can only *gradually* change X (for example because the robots can only move with a limited speed). Therefore, in contrast to the *direct* optimization of U and V in clustering, we only apply a *gradient descent* of X with a limited step size $s \in \mathbb{R}^+$ in herding. The step size s divided by the time for one step represents the maximum speed of the individuals, e.g. robots. The direction vectors for the gradient descent on X are obtained as

$$\partial J / \partial x_k = \Delta x_k = -\sum_{i=1}^{c} 2 u_{ik}^m \, (x_k - v_i), \tag{9}$$

$k = 1, \ldots, n$. So, the update equations for X are

$$x_k = x_k + \frac{\min\{s, \|\Delta x_k\|\}}{\|\Delta x_k\|} \cdot \Delta x_k, \tag{10}$$

ISBN: 978-989-95079-6-8

$k = 1, \ldots, n$. This makes sure that the individuals move in the direction of the gradient but that the step length never exceeds the threshold s. The herding algorithm terminates when all elements of X have approached the corresponding elements of V, i.e. when $\max \Delta X < s$. The fuzzy c–means herding algorithm can then be formally defined as follows:

Algorithm 2: Fuzzy c–Means Herding.

```
input X, c, m, s
randomly initialize V ∈ ℝ^p
repeat
   update U(V,X,m) by (7)
   update V(U,X,m) by (8)
   update X(U,V,m) by (10) and (9)
until max‖ΔX‖ < s
output U, V, X
```

5 Experiments

We illustrate the operation of the fuzzy c–means herding algorithm using an artificial data set that is inspired by the autonomous robot scenario presented at the beginning of section 1. We randomly generate a set X of $n = 100$ data vectors of the dimension $p = 2$ using a Gaussian distribution with mean zero and variance one. This data set is illustrated in Fig. 1. The idea is that these data vectors represent the current positions of the exploring robots in the two–dimensional plane just before the herding process begins. Then, we start the fuzzy c–means herding algorithm with a fuzziness index of $m = 2$, a step size $s = 0.05$ (representing the speed of the robots), and different numbers of clusters $c \in \{2, 3, 4, 5\}$. Each experiment is performed for the same initial vectors X and V. Additional experiments with other random initializations of X and V (not presented in detail here) produced very similar results. The trajectories of the vectors V and X are shown in Fig. 2. The solid curves in these figures show the trajectories of X. Each of these trajectories starts at one of the initial points of X (as shown in Fig. 1), smoothly moves through the two–dimensional plane, and terminates in one of the final cluster centers. The data vectors do not directly move towards one of the cluster centers, but are apparently initially fuzzily assigned to several clusters, have a tendency to move towards the mean of the data set and then gradually focus on one of the cluster centers. We counted the number of steps t until all of the data points eventually reached one of the cluster centers and obtained the following table:

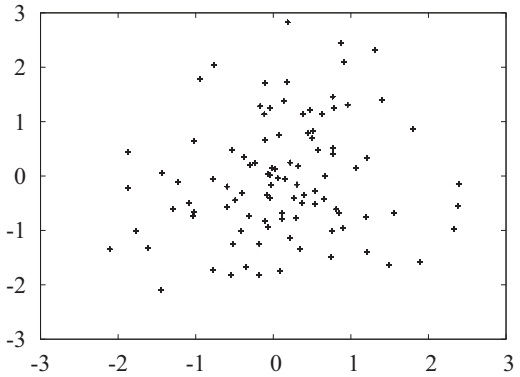


Figure 1: Initial data set (robot positions).

Table 1: Convergence steps for fuzzy c–means herding, $m = 2$.

c	2	3	4	5
t	34	29	25	23

A detailed study of the step size showed that lower values of s produce smoother trajectories and lead to higher numbers of convergence steps t. Correspondingly, higher values of s produce rather coarse trajectories and lead to lower numbers of convergence steps t. Despite the influence of s on the absolute values of t, the relation between the convergence steps for the different experiments are almost independent of s.

The ticks $\times$ in the figures show the trajectories of V. The cluster centers in V are not fixed but also move through the two–dimensional plane with an approximately decreasing step size as the herding process proceeds. The data points that are assigned to each cluster originally covered roughly the same area in the two–dimensional plane. Notice that a similar effect can be observed with the fuzzy c–means clustering algorithm.

The reader might wonder if it wasn't more efficient to arbitrarily pick fixed cluster centers, say c of the initial data vectors in X, assign each data point to the closest data center and let each data point directly move towards its pre–assigned data point. The trajectories obtained by this experiment are shown in Fig. 3. The trajectories of X (solid curves) form stars whose centers are the initially randomly chosen cluster centers. Unlike the clusters obtained by fuzzy c–means herding, the data points in each of these clusters do not represent similar areas, but some clusters obtain only very few, and others obtain many data points, depending on the cluster centers chosen. Apparently, the linear trajectories of some of the data points are considerably shorter than those obtained by fuzzy c–means herding. However, some of the trajectories are considerably longer than their herding equivalents. Also for this experiment we counted the number of steps t until all of the data points reached their closest cluster centers:

Table 2: Convergence steps for pre–assigned cluster centers (one specific run).

c	2	3	4	5
t	35	29	27	26

Notice that fuzzy c–means herding yields the same or even lower values of t for all numbers of clusters. However, this result depends on the particular choices of the cluster centers. Therefore, we repeated each of these experiments with 1000 different random choices of cluster centers and computed the numbers of steps until termination. The histograms of these step numbers are shown in Fig. 4 and the mean numbers of steps $\bar{t}$ are given in the following table:

Table 3: Convergence steps for pre–assigned cluster centers (average of 1000 runs).

c	2	3	4	5
$\bar{t}$	33.9	30.3	28.0	26.4

The average number of steps is almost equal or larger than the number of steps needed for fuzzy c–means herding with

ISBN: 978-989-95079-6-8

$m = 2$. In order to reduce the number of convergence steps in fuzzy c–means herding we reduce the value of the fuzziness parameter m, so the individual data points are less fuzzily assigned to the clusters in the beginning and more directly orientate towards their corresponding cluster centers. To avoid numerical problems that sometimes occur when m is very close to one, we pick $m = 1.1$. For this case, the trajectories of V and X are shown in Fig. 5. Just as in Fig. 2 and in contrast to Fig. 3 each cluster corresponds to approximately the same area in the two–dimensional plane. The trajectories are a little less smooth than in Fig. 2 and rather resemble the star–like trajectories in Fig. 3. This observation is corroborated by the fact that the number of convergence steps is much lower than in Fig. 2, as shown in the following table:

Table 4: Convergence steps for fuzzy c–means herding, $m = 1.1$.

c	2	3	4	5
t	28	22	20	18

A comparison of Table 4 with Table 3 reveals the fact that the number of convergence steps for fuzzy c–means herding is even lower than (on average) for pre–assigned cluster centers. We marked the numbers of convergence steps of fuzzy c–means herding for $m = 2$ and for $m = 1.1$ as vertical lines in Fig. 4. Apparently, for $m = 1.1$ the number of steps is very close to the minimum of the histograms, i.e. it compares with the number of steps when the clusters by chance have been approximately optimally assigned. In other words, fuzzy c–means herding implicitly finds an (almost) optimal cluster assignment, leads to a considerably low number of convergence steps, and yields smoothly herding trajectories at the same time.

6 Conclusions

In this paper we have shown that minimization of clustering models can be used to control herding processes. In particular, we have focussed on herding by minimization of the fuzzy c–means clustering model. Th experiments with a gradient descent of the data vectors with limited step size can be illustrated by a scenario with automous mobile robots that have to move to a certain number of locations (cluster centers) in a self–organizing manner. The experiments with fuzzy c–means herding yield very smooth trajectories. The individuals that are collected at each cluster center approximately represent the same area (or, more generally, hypervolume) of the initial state space. This means that the final clusters are nicely balanced. Compared with herding with pre–defined cluster centers, fuzzy c–means herding on average needs less steps until termination. For low values of the fuzziness index the trajectories become less smoother, but the number of steps is further reduced. For $m = 1.1$ the number of steps is even close to the global minimum. Thus, fuzzy c–means herding is an attractive herding algorithm that yields smooth trajectories, well–balanced clusters, and fast convergence.

References

[1] K. C. Ng and M. M. Trivedi. Multirobot convoying using neuro-fuzzy control. In *International Conference on Pattern Recognition*, pages 417–421, Vienna, Austria, August 1996.

[2] L. E. Parker, B. Kannan, X. Fu, and Y. Tang. Heterogeneous mobile sensor net deployment using robot herding and line-of-sight formations. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2488–2493, Las Vegas, October 2003.

[3] A. Kirchner and A. Schadschneider. Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics. *Physica A*, 312(1–2):260–276, 2002.

[4] M. R. Akbarzadeh, H. Rezaei, and M. B. Naghibi. A fuzzy adaptive algorithm for expertness based cooperative learning, application to herding problem. In *International Conference of the North American Fuzzy Information Processing Society*, pages 317–322, Chicago, July 2003.

[5] J. M. Lien, O. B. Bayazit, R. T. Sowell, S. Rodriguez, and N. M. Amato. Shepherding behaviors. In *IEEE International Conference on Robotics and Automation*, pages 4159–4164, New Orleans, April 2004.

[6] P. Kachroo, S. A. Shedied, J. S. Bay, and H. Vanlandingham. Dynamic programming solution for a class of pursuit evasion problems: the herding problem. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 31(1):35–41, 2001.

[7] D. Costantini, S. Donadio, U. Garibaldi, and P. Viarengo. Herding and clustering: Ewens vs. Simon-Yule models. *Physica A*, 355(1):224–231, 2005.

[8] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[9] G. B. Ball and D. J. Hall. Isodata, an iterative method of multivariate analysis and pattern classification. In *IFIPS Congress*, 1965.

[10] R. Krishnapuram and J. M. Keller. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110, May 1993.

[11] R. N. Davé. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12:657–664, 1991.

[12] T. A. Runkler. A single objective fuzzy clustering model for very high–dimensional data. In *GMA/GI Workshop Fuzzy Systems and Computational Intelligence, Dortmund, Germany*, pages 234–243, December 2008.

[13] J. C. Bezdek and R. J. Hathaway. Optimization of fuzzy clustering criteria using genetic algorithms. In *IEEE Conference on Evolutionary Computation, Orlando*, volume 2, pages 589–594, June 1994.

[14] F. Klawonn and A. Keller. Fuzzy clustering with evolutionary algorithms. *International Journal of Intelligent Systems*, 13(10–11):975–991, 1998.

[15] C. Y. Chen and F. Ye. Particle swarm optimization and its application to clustering analysis. In *International Conference on Networking, Sensing and Control*, pages 789–794, Taipeh, Taiwan, March 2004.

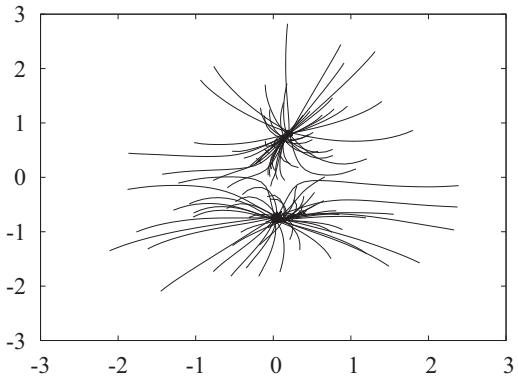
[16] X. Cui, T. E. Potok, and P. Palathingal. Document clustering using particle swarm optimization. In *IEEE Swarm Intelligence Symposium, Pasadena*, pages 185–191, June 2005.

[17] T. A. Runkler and C. Katz. Fuzzy clustering by particle swarm optimization. In *IEEE International Conference on Fuzzy Systems*, pages 3065–3072, Vancouver, July 2006.

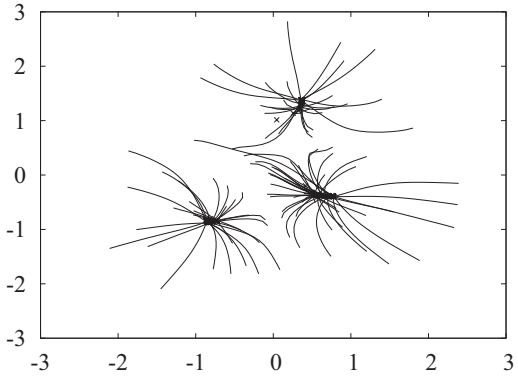
[18] T. A. Runkler. Ant colony optimization of clustering models. *International Journal of Intelligent Systems*, 20(12):1233–1261, December 2005.

[19] T. A. Runkler. Wasp swarm optimization of the c-means clustering model. *International Journal of Intelligent Systems*, 23(3):269–285, February 2008.

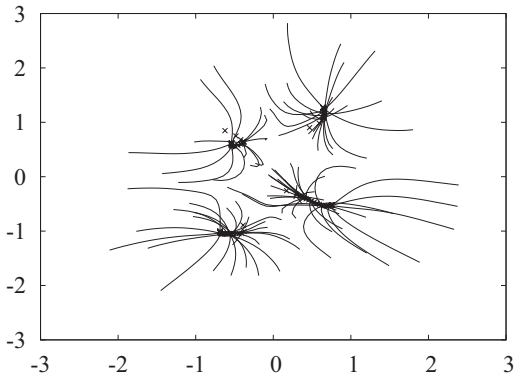
ISBN: 978-989-95079-6-8



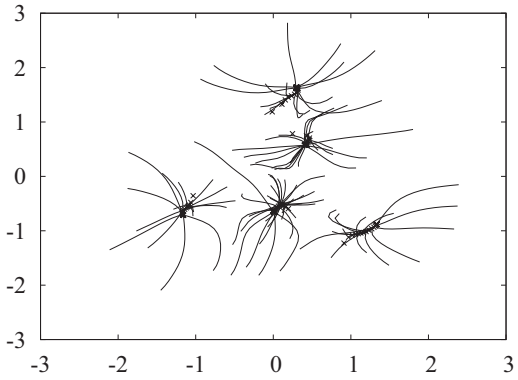
$c = 2$



$c = 3$

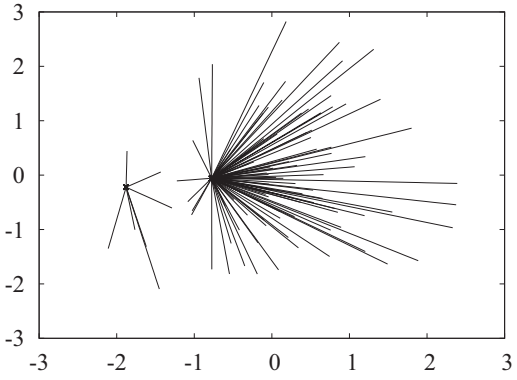


$c = 4$

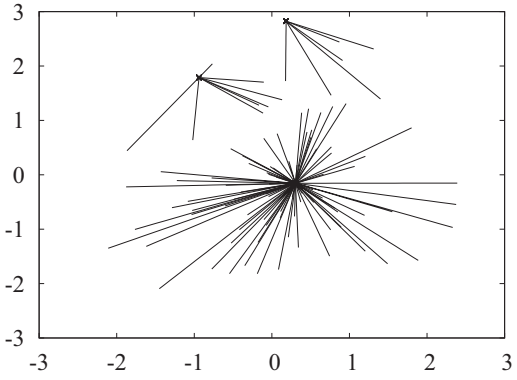


$c = 5$

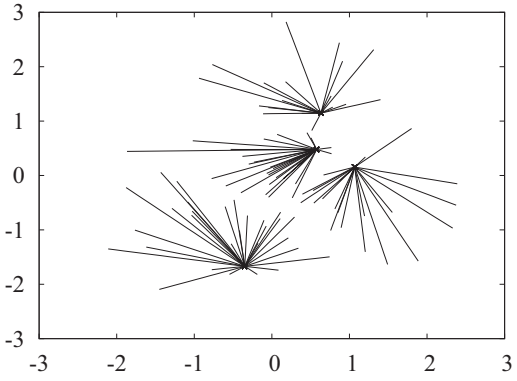
Figure 2: Herding trajectories for $m = 2$.



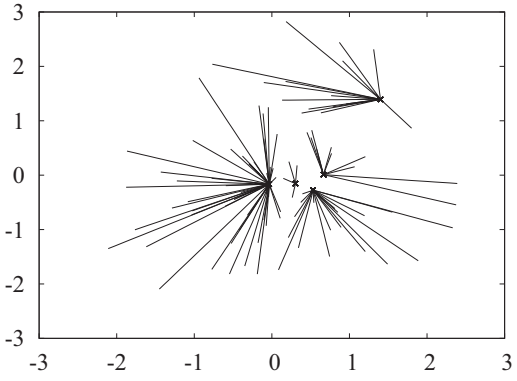
$c = 2$



$c = 3$



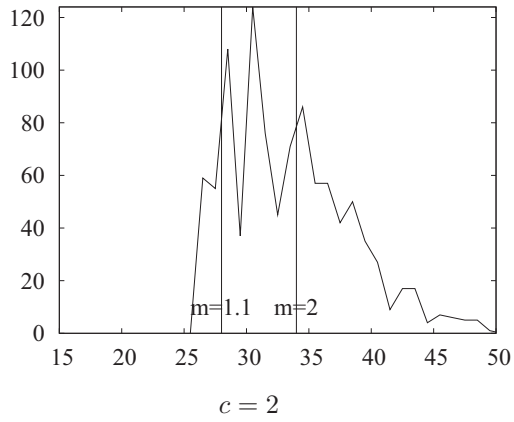
$c = 4$



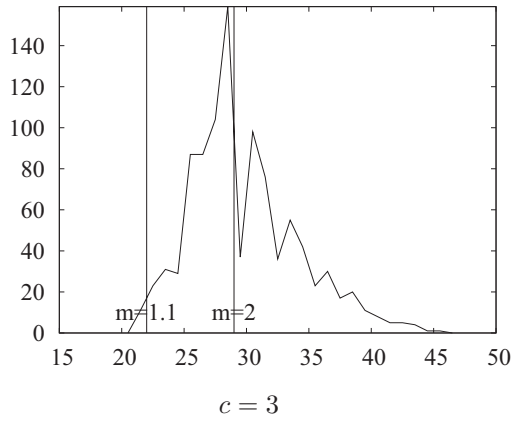
$c = 5$

Figure 3: Trajectories with pre-assigned arbitrary cluster centers.

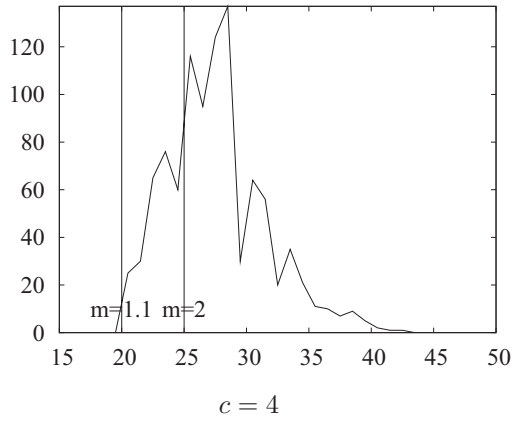
ISBN: 978-989-95079-6-8



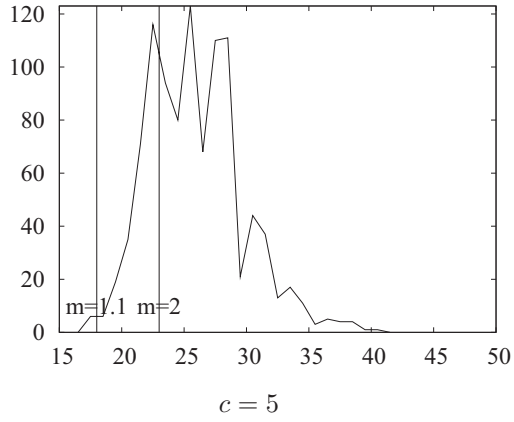
$c = 2$



$c = 3$

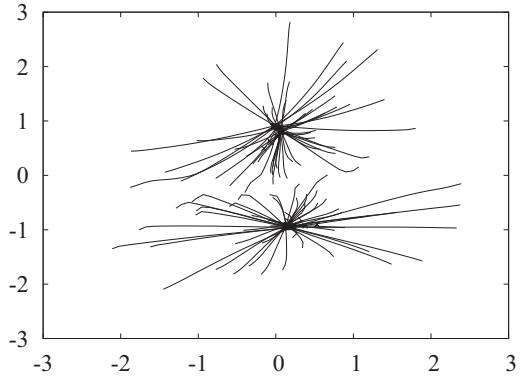


$c = 4$

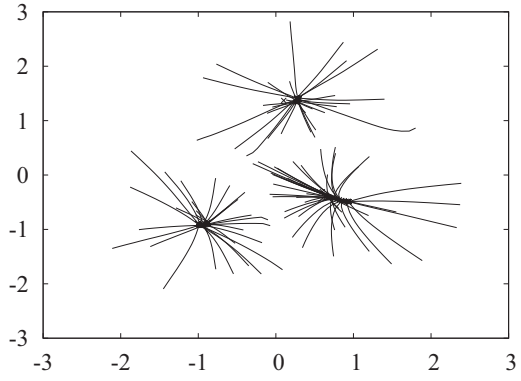


$c = 5$

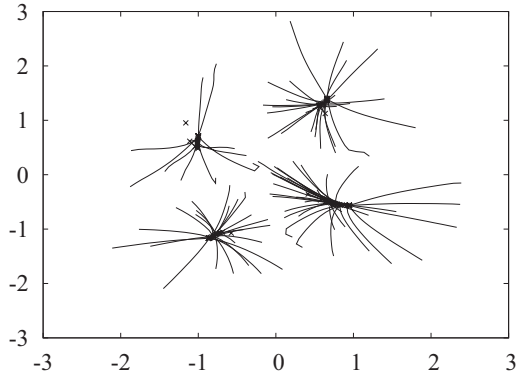
Figure 4: Histograms of the number of steps for preassigned cluster centers.



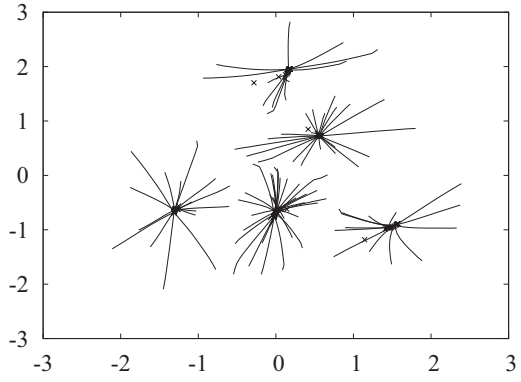
$c = 2$



$c = 3$



$c = 4$



$c = 5$

Figure 5: Herding trajectories for $m = 1.1$.

ISBN: 978-989-95079-6-8

Interval AHP for a group of decision makers

Tomoe Entani[1] 1.Department of Economics and Social Sciences, Kochi University
2-5-1 Akebono Kochi, Japan
Email: entani@cc.kochi-u.ac.jp

Abstract— *This paper considers Interval Analytic Hierarchy Process (Interval AHP) in group decision making for encouraging communication. Interval AHP is suitable method to handle subjective judgments since the induced results are intervals which can include uncertainty of given information. The decision makers' opinions can be aggregated at some stages of decision making process to reach consensus. One of the approaches is to aggregate the given judgments considering outliers and from them the group preferences are obtained. By the other approach, first the individual preferences are obtained from the respective decision maker's judgments and then they are aggregated. In the sense of reducing communication barriers, obtaining individual priority weights of alternatives beforehand helps decision makers realize their own and the others' opinions. The judgments and preferences can be aggregated based on the possibility view or by introducing importance weights of decision makings.*

Keywords— group decision making, importance weight, interval AHP, intuitive judgments, possibility, uncertainty

1 Introduction

In most organizations decision making is made by two or more people, regardless of whether the organization is public or private and the decision problem is local or global. Although all members do not need to be located in the same physical location, they are aware of one another and act as a member of the group which makes a decision. The apparent purpose of the group is to reach a decision, that is, to choose one alternative which seems to be acceptable and agreeable for all members. However, it is sometimes difficult to reach a consensus among group members [1]. Especially when there are various members who are not face-to-face, there exist some barriers to understand one another. These communication barriers make decision making tough. The technological advancements, such as computers and networks, accelerate such situations. In addition, the decision makers have some difficulties to represent and recognize their own opinions which fit their intuitive judgments and experiences, because the problems are often very complicated. It may happen that some members may exaggerate their preferences to influence the group decision. In this sense, it is also important to support the interpersonal information exchange, as well as to find the optimal alternative, in the group decision making. As a preparation for the consensus, it becomes necessary to remove communication barriers by representing individual opinions simply and clearly [1, 2]. This paper concerns on these design phase more than the choice phase in group decision support system by focusing on the aggregation of decision makers' opinions.

The group decision support system is discussed from the scope of AHP (Analytic Hierarchy Process). AHP is a useful method in multi-criteria decision making problems [3]. It is structured hierarchically as criteria and alternatives. The priority or weight for each element of the hierarchy is obtained by eigenvector method [3]. Then, the priority weight for each alternative is calculated with them. The advantages of AHP are the following two points. It helps decision maker structure complex problems hierarchically. The decision makers can give their subjective judgments directly by comparing a pair of attributes.

The group decision making with AHP is discussed in [4, 5, 6, 7]. On this assumption, more than two comparison matrices are given. The decision makers' individual opinions are aggregated in several stages of decision making process by several approaches. It can be done at the beginning stage by searching consensus on the given judgments. It is also possible to take the average of individually induced preferences with importance weights of the group members at the last stage. Between these stages, the decomposed intermediate judgments may be aggregated by voting and so on [2]. This paper concerns on the aggregations at the beginning and last stages based on two concepts such as possibility and nominal average. Furthermore, in order to reflect uncertainty of the given comparisons, Interval AHP [8, 9], where the interval weights, instead of crisp weights in the conventional AHP, are obtained is applied.

This paper is based on possibility concept in dealing with the given data and obtained results and consists as follows. As a preliminary, the definition and properties of the interval probability which are used for normalization of intervals are explained in Section 2. At first, Interval AHP as a tool to represent each decision maker's preference is shown briefly in Section 3. Then, in Section 4, the approaches to aggregate various opinions from several viewpoints are proposed. Finally, the proposed models are tested with a numerical example in case of a group of four decision makers in Section 5.

2 Interval probability as preliminary

The interval probabilities are defined by a set of intervals as follows. This definition is originally proposed in [10] and also is used in [11, 12]. The conventional crisp probabilities are extended into interval ones.

Definition 1 (Interval probability) *The set of intervals denoted as* $\{W_1, ..., W_n\}$ *where* $W_i = [\underline{w}_i, \overline{w}_i]$ *are called interval probabilities if and only if*

$$\begin{array}{ll} 1) & 0 \le \underline{w}_i \le \overline{w}_i \ \forall i \\ 2) & \sum_{i \ne j} \overline{w}_i + \underline{w}_j \ge 1 \forall j \\ 3) & \sum_{i \ne j} \underline{w}_i + \overline{w}_j \le 1 \ \forall j. \end{array} \quad (1)$$

From Definition 1, following two inequalities hold

$$\textstyle\sum_i \underline{w}_i \le 1 \ \text{ and } \ \sum_i \overline{w}_i \ge 1.$$

ISBN: 978-989-95079-6-8

Definition 1 is regarded as a normality condition of intervals corresponding to the conventional one $\sum_i w_i = 1$. It is noted that in interval probabilities there are many combinations of crisp values whose sum is one.

The combination of a pair of interval probability sets is denoted as follows.

Property 1 Combination rule *Assuming a pair of interval probabilities on n elements as $\{W_1^A, ..., W_n^A\}$ and $\{W_1^B, ..., W_n^B\}$ which satisfy Definition 1, their combination is denoted as $\{W_1^{AB}, ..., W_n^{AB}\}$. Each of elements is an interval $W_i^{AB} = [\underline{w}_i^{AB}, \overline{w}_i^{AB}]$ denoted as follows.*

$$\begin{array}{l} \underline{w}_i^{AB} = \min\{\underline{w}_i^A, \underline{w}_i^B\} \\ \overline{w}_i^{AB} = \max\{\overline{w}_i^A, \overline{w}_i^B\} \end{array} \tag{2}$$

The set of combined intervals $\{W_1^{AB}, ..., W_n^{AB}\}$ also satisfies Definition 1 so that it is interval probability.

Property 2 Weighted sum of interval probabilities *Assuming m sets of interval probabilities on n elements as $\{W_1^{A_k}, ..., W_n^{A_k}\}$ $k = 1, ..., m$, and the weights for the sets as $\{p_k \forall k\}$ where $p_k \geq 0$ and $\sum_k p_k = 1$, the weighted sum of interval probabilities is $\{W_1, ..., W_n\}$. Each of elements is an interval $W_i = [\underline{w}_i, \overline{w}_i]$ denoted as follows.*

$$\begin{array}{l} \underline{w}_i = \sum_k p_k \underline{w}_i^{A_k} \\ \overline{w}_i = \sum_k p_k \overline{w}_i^{A_k} \end{array} \tag{3}$$

As far as the sum of weights is 1, the weighted sum of interval probabilities becomes also interval probability. It is verified as follows. They apparently satisfy the condition 1) in Definition 1. The condition 2) holds as follows.

$$\begin{array}{l} \sum_{i \neq j} \sum_k p_k \overline{w}_i^{A_k} + \sum_k p_k \underline{w}_j^{A_k} \\ = \sum_k p_k (\sum_{i \neq j} \overline{w}_i^{A_k} + \underline{w}_j^{A_k}) \text{ [Definition 1-2)]} \\ \geq \sum_k p_k = 1 \end{array} \tag{4}$$

The other condition 3) is satisfied by the similar way.

$$\begin{array}{l} \sum_{i \neq j} \sum_k p_k \underline{w}_i^{A_k} + \sum_k p_k \overline{w}_j^{A_k} \\ = \sum_k p_k (\sum_{i \neq j} \underline{w}_i^{A_k} + \overline{w}_j^{A_k}) \\ \leq \sum_k p_k = 1 \end{array} \tag{5}$$

Then, the weighted sum of interval probabilities satisfy the conditions in Definition 1.

3 Interval AHP

AHP is an approach to multi-criteria decision making problems. The problem is decomposed into hierarchy by criteria and alternatives. The choice or preferences of alternatives are induced as a final decision from the decision maker's judgments given as pariwise comparison matrix. The decision maker compares all pairs of alternatives and gives the pairwise comparison matrix for n alternatives as follows [3].

$$A = \begin{pmatrix} 1 & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{n1} & \cdots & 1 \end{pmatrix} \tag{6}$$

where a_{ij} shows the importance ratio of alternative i comparing to alternative j. The comparison matrix satisfies the following relations so that the number of given comparisons is $n(n-1)/2$.

$$\begin{array}{ll} a_{ii} = 1 & \text{identical} \\ a_{ij} = 1/a_{ji} & \text{reciprocal} \end{array} \tag{7}$$

The decision maker can give his/her judgment intuitively without caring about the relative relations of comparisons. Therefore, the given comparisons are not always consistent each other. The consistent comparisons satisfy the following transitivity relations.

$$a_{ij} = a_{ik} a_{kj} \ \forall (i, j) \tag{8}$$

In the following, inconsistency means that (8) is not satisfied. The proposed models in this paper deals with such inconsistency from the possibility view [13].

In the conventional AHP, crisp priority weights are obtained from the given comparison matrix. They are extended to intervals in Interval AHP [8, 9]. The given comparisons are inconsistent each other, that is, they do not always satisfy (8). In order to reflect such inconsistency, the priority weight of alternative is denoted as the following interval.

$$W_i = [\underline{w}_i, \overline{w}_i] \ \forall i \tag{9}$$

For their normalization, they are represented as interval probabilities so that they satisfy (1) in Definition 1.

The pairwise comparison is an intuitive ratio of two alternatives so that they are approximated by the following interval.

$$\frac{W_i}{W_j} = \left[\frac{\underline{w}_i}{\overline{w}_j}, \frac{\overline{w}_i}{\underline{w}_j} \right] \tag{10}$$

where $0 < \underline{w}_i \forall i$ and the upper and lower bounds of the approximated comparison are defined as the maximum range with respect to the two intervals.

In the approximation model the probabilities are determined so as to include the given pairwise comparisons. Thus, from the possibility view, the obtained interval probabilities satisfy the following inclusion relations.

$$a_{ij} \in \frac{W_i}{W_j} \ \forall (i, j) \tag{11}$$

which leads to the following inequalities.

$$\frac{\underline{w}_i}{\overline{w}_j} \leq a_{ij} \leq \frac{\overline{w}_i}{\underline{w}_j} \Leftrightarrow \left\{ \begin{array}{l} \underline{w}_i \leq a_{ij} \overline{w}_j \ \forall (i, j) \\ \overline{w}_i \geq a_{ij} \underline{w}_j \ \forall (i, j) \end{array} \right. \tag{12}$$

The approximations by the obtaining interval priority weights include the given inconsistent comparisons.

For any inconsistent comparisons, assuming $[\underline{w}_i, \overline{w}_i] = [0, 1] \ \forall i$, the above inclusion relation (12) is apparently satisfied. A decision maker does not need to revise their intuitive judgments so as to be consistent. When a decision maker gives completely inconsistent judgments, the obtained priority weights of all alternatives are equally $[0, 1]$. It represents complete ignorance and fits our natural sense. Inconsistency among the given comparisons is reflected in the uncertainty of interval probabilities.

ISBN: 978-989-95079-6-8

The constraint conditions for determining the interval probabilities are (1) and (12). In order to obtain the least uncertain probabilities, the uncertainty of interval probabilities should be minimized. The uncertainty of interval probabilities can be measured by several indices, such as widths of intervals and entropy [14]. For simplicity, the sum of widths of intervals is used in this paper. The problem to determine the interval priority weights is formulated as follows.

$$\begin{array}{ll} I = & \min \quad \sum_i (\overline{w}_i - \underline{w}_i) \\ \text{s.t.} & \sum_{i \neq j} \overline{w}_i + \underline{w}_j \geq 1 \; \forall j \\ & \sum_{i \neq j} \underline{w}_i + \overline{w}_j \leq 1 \; \forall j \\ & \underline{w}_i \leq a_{ij} \overline{w}_j \; \forall (i,j) \\ & \overline{w}_i \geq a_{ij} \underline{w}_j \; \forall (i,j) \\ & \underline{w}_i \geq \epsilon \; \forall i \end{array} \tag{13}$$

The greater optimal objective function value is, the more uncertain the given interval priority weight becomes.

4 Group of Decision Makers

Interval AHP is introduced to the group decision making by aggregating individual opinions. Each decision maker gives pairwise comparisons for alternatives based on his/her subjective judgments. The comparison matrix given by the decision maker k, where $k = 1, ..., m$, is denoted as follows.

$$A_k = \begin{pmatrix} 1 & \cdots & a_{1nk} \\ \vdots & a_{ijk} & \vdots \\ a_{n1k} & \cdots & 1 \end{pmatrix} \quad \forall k \tag{14}$$

In the following sections, the given judgments and the obtained priority weights are aggregated at the beginning and last stages of the decision making process, respectively. The procedures of these approaches are illustrated in Figure 1.

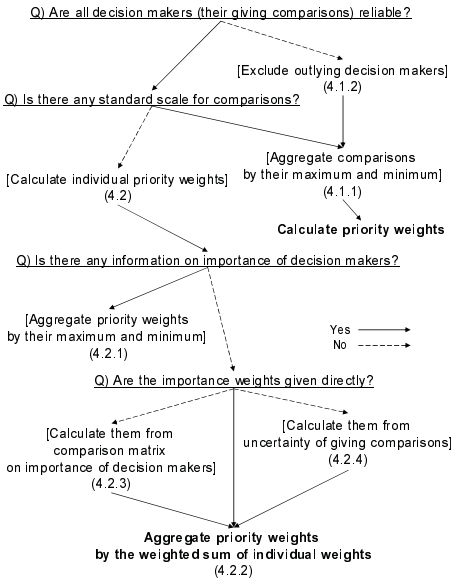


Figure 1: Steps for group decision making

4.1 Aggregation of given judgments

When the judgments which are the comparisons on alternatives in this paper are given, they can be aggregated at the beginning stage of group decision making process. It is usually said that the smaller group is more efficient since all the given judgments are reliable and tend not to be too much diverse. As for the well experienced group, for instance, the standard of judgments might be previously required or satisfied and then it is useful to aggregate the given judgments at the beginning stage. From the view of possibility concept, this is a natural and simple aggregation method. The decision makers can see their differences on their giving comparisons.

4.1.1 Directly aggregation of comparisons

The approach where the geometric mean of comparisons is taken has been proposed in [4, 5, 15, 16]. Since the aggregated comparison matrix satisfies (7), the eigenvector method can be also used to obtain the priority weights. In this section, the comparisons given by m decision makers are aggregated from the possibility view by taking their minimum and maximum.

$$A_{ij} = [\underline{a}_{ij}, \overline{a}_{ij}] = [\min_k a_{ijk}, \max_k a_{ijk}] \;\; \forall (i,j) \tag{15}$$

Since the aggregated comparisons are intervals, the inclusion constraints (12) can be written as follows.

$$A_{ij} \in \frac{W_i}{W_j} \;\Leftrightarrow\; \frac{\underline{w}_i}{\overline{w}_j} \leq \underline{a}_{ij}, \overline{a}_{ij} \leq \frac{\overline{w}_i}{\underline{w}_j} \tag{16}$$

where the aggregated comparisons are interaval and included in the approximated ones by the interval priority weights. The inclusion relation (16) with interval comparisons is an extension of (12) with crisp ones.

By solving the revised problem (13) from constraint (12) into (16), the bounds of the interval priority weights are obtained. The approximated comparison with them surely includes individually given comparisons.

4.1.2 Finding and excluding outliers

The decision making with a big group is also required. Instead of including opinions of all the decision makers in Section 4.1.1, the way to find outliers is proposed in this section. The decision makers whose opinions are apparently different from the others is called outliers. Such outliers are removed one by one till the number of removed decision makers reaches the pre-fixed number. It is noted that the number of appropriate decision makers should be determined beforehand. Then, the following function is used to measure his/her outlierness.

$$J_o = \sum_{(i,j)} \frac{\max_k a_{ijk}}{\max_{k \neq o} a_{ijk}} + \frac{\min_{k \neq o} a_{ijk}}{\min_k a_{ijk}} \geq n(n-1) \tag{17}$$

The evaluation function J_o represents the amount of extension by adding the decision maker o into the group. The increase of their upper bounds is measured as a ratio since the comparisons are ratio measure. As for the lower bounds, the decrease ratio is measured by the same way. The decision maker $\{o^* | J_o^* = \max_o J_o\}$, who gives the maximum of J is removed. If there are more than two decision makers who give the maximum, all of them are excluded. When the number of excluded decision makers reach the pre-fixed number, the comparisons given by the others are aggregated by (15) into the interval comparison matrix. The following process is the same as in Section 4.1.1.

ISBN: 978-989-95079-6-8

4.2 Aggregation of obtained individual preferences

The pariwise comparisons are relative values and each decision maker gives comparisons based on his/her own standard. In order to deal with these variations of comparisons, the decision makers' opinions are aggregated after obtaining individual preferences from respective given comparison matrix. First, the priority weights are obtained from the individually given pairwise comparison matrix. The interval priority weights based on the comparison matrix given by kth decision maker are denoted as $[\underline{w}_{ik}, \overline{w}_{ik}]$. Each decision maker can realize his/her priority weights on the alternatives, as well as others' ones. Then, in order to reach a consensus of the group, the obtained individual priority weights are aggregated.

4.2.1 Directly aggregation of priority weights

When there is not enough information on the importance of the decision makers, they are aggregated simply by taking their maximum and minimum.

$$W_i^1 = [\underline{w}_i, \overline{w}_i] = [\min_k w_{ik}, \max_k w_{ik}] \; \forall i \tag{18}$$

The obtained set of intervals are also interval probabilities by Property 1. The aggregated interval can be said to give an agreeable result, since the priority weights by all members are included in it. This is based on possibility concept and the assumption that all the decision makers give reasonable information. When the individual preferences are very different one another such as in a big group, the widths of the aggregated priority weights become large so that they are uncertain. The diversity of opinions is reflected in uncertainty of the aggregated weights. Based on the individual preferences, it is also possible to exclude outliers similarly in Section 4.1.2.

4.2.2 Aggregation with directly given importance weights

When the importance of decision makers and/or their confidence degrees of the given judgments are different, it is better to reflect such information. Therefore, the weighted sum of priority weights is proposed in the following sections.

When the importance of each decision maker is given by a supervisor, it is introduced as his/her weight directly. Denoting the given importance weight as $p_k \geq 0 \forall k$, where $\sum_k p_k = 1$, the aggregated priority weights are obtained as follows.

$$W_i^2 = [\sum_k p_k \underline{w}_{ik}, \sum_k p_k \overline{w}_{ik}] \; \forall i \tag{19}$$

As far as the weights are normalized so as to make the sum one, the set of the aggregated priorities is also interval probability by Property 2. The situation that there is not enough information is also represented by giving all the decision makers the same importance weights $1/m$. In the next section, it is represented differently from the view of interval probabilities.

4.2.3 Aggregation with importance weights by AHP

It is also possible to use conventional AHP and Interval AHP in order to induce the importance weights of decision makers. It is easier for a supervisor to give pariwise comparisons on importance of decision makers than to give their importance weights directly. The similar idea has been proposed in [6, 16], where a level of decision makers is added to the existing criteria levels. The importance weights of decision makers are determined at the level. Then, they are used as weights in the way of summing the local weights up to the overall priority weights. Using Interval AHP in obtaining weights, the overall weights are obtained as intervals in [17].

The supervisor gives comparisons on the importance of decision makers.

$$B = \begin{pmatrix} 1 & \cdots & b_{1m} \\ \vdots & b_{kl} & \vdots \\ b_{m1} & \cdots & 1 \end{pmatrix} \tag{20}$$

where each comparison represents how important is the decision maker comparing to the other.

Applying the conventional AHP in [3], the importance weights are obtained so as to make the sum one. Then, the following process is the same as in Section 4.2.2.

The Interval AHP formulated by (13) can be used to determine importance weights. The given comparisons on alternatives a_{ij} and the obtained priority weights $W_i = [\underline{w}_i, \overline{w}_i]$ in (13) are replaced into those on decision makers b_{kl} and their interval importance weights $P_k = [\underline{p}_k, \overline{p}_k]$, respectively. By solving the revised problem (13), the interval importance weights of decision makers are obtained. The problems to obtain the upper and lower bounds of the aggregated priority weight of alternative i are formulated as follows, respectively.

$$\begin{aligned} \underline{w}_i = \quad & \min \textstyle\sum_k p_{ik} \underline{w}_{ik} \\ & \text{s.t.} \; \textstyle\sum_k p_{ik} = 1 \\ & \underline{p}_k \leq p_{ik} \leq \overline{p}_k \; \forall k \\ \overline{w}_i = \quad & \max \textstyle\sum_k p_{ik} \overline{w}_{ik} \\ & \text{s.t.} \; \textstyle\sum_k p_{ik} = 1 \\ & \underline{p}_k \leq p_{ik} \leq \overline{p}_k \; \forall k \end{aligned} \tag{21}$$

In both problems, the variables are the importance weights p_{ik} included in the obtained interval importance weights. For each alternative i and its upper and lower bounds, the optimal solutions might be different. Denoting their optimal solutions for the alternative i as p_{ik*} and p_{ik}^*, respectively, the aggregated priority weight is obtained as the following intervals.

$$W_i^3 = [\sum_k p_{ik*} \underline{w}_{ik}, \sum_k p_{ik}^* \overline{w}_{ik}] \; \forall i \tag{22}$$

They are also interval probabilities by Properties 1 and 2.

From the view of interval probability, complete ignorance is represented as giving all decision makers the same interval weights $P_k = [0, 1]$. With them, the priority weights (22) by solving (21) is the same as the directly aggregated ones (18). Apparently, the aggregated priority weights with importance weights are included in the directly aggregated priority weights, that is, $W_i^2, W_i^3 \in W_i^1$. The results with importance weights are less uncertain. If any information on importance of decision makers is available, it is reasonable to introduce them as their importance weights.

4.2.4 Aggregation with importance weights based on uncertainty of information

As shown in Section 3, the sum of widths represents the uncertainty of the interval weights. The uncertainty of the given information may depend on the confidence or expert levels of decision makers on the problem. It is possible to assume such uncertainty degree of judgments as the decision maker's

ISBN: 978-989-95079-6-8

importance among the group. The importance weights of the decision makers are determined by reflecting uncertainty of their information. The decision maker who gives less uncertain comparisons will be evaluated as more important. The importance weight of decision maker l is as follows.

$$p_l = \frac{\sum_{k \neq l} I_k}{(m-1)\sum_k I_k} \tag{23}$$

where I_k is the optimal objective value of (13) with the given information by the kth decision maker and represents his/her uncertainty degree. It is noted that their sum is one, that is, $\sum_l p_l = (\sum_l \sum_{k \neq l} I_k)/((m-1)\sum_k I_k) = 1$. The importance weight of the decision maker who gives interval weights with great widths becomes small.

The aggregated priority weights are denoted as follows and also interval probabilities by Property 2.

$$W_i^4 = [\sum_k p_l \underline{w}_{ik}, \sum_k p_l \overline{w}_{ik}] \; \forall i \tag{24}$$

5 Numerical example

Assuming four decision makers, $k = 1, 2, 3, 4$, each of them gives the pairwise comparison matrix on four alternatives, $i = 1, 2, 3, 4$, shown in Table 1. The interval priority weights W_k and their uncertainty indices I_k are obtained by (13) and its optimal objective function value, respectively.

Table 1: Comparison matrices by four decision makers

A_1				W_1	
1	2	3	4	0.500	J_1=12
	1	2	3	0.250	I^1=0.083
		1	2	[0.125,0.167]	p_1=0.297
			1	[0.083,0.125]	
A_2				W_2	
1	2	4	6	0.545	J_2=13.2
	1	3	4	0.273	I^1=0.109
		1	4	[0.091,0.145]	p_2=0.285
			1	[0.036,0.091]	
A_3				W_3	
1	3	3	4	0.571	J_3=13.5
	1	2	4	0.190	I^1=0.190
		1	1	[0.095,0.190]	p_3=0.250
			1	[0.048,0.143]	
A_4				W_4	
1	1	2	2	0.375	J_4=16.5
	1	3	1	[0.219,0.375]	I^1=0.375
		1	3	[0.125,0.188]	p_4=0.168
			1	[0.063,0.219]	

5.1 Aggregation of comparisons

By (15), the comparisons given by four decision makers are aggregated and shown in Table 2. The interval priority weights are obtained by solving (13) with constraints (16). Each decision maker can accept the results more or less, since they are based on his/her giving comparisons by aggregating them into interval comparisons.

Table 2: Aggregated comparison matrix

A (aggregated)				W
1	[1,3]	[2,4]	[2,6]	[0.333,0.500]
	1	[2,3]	[1,4]	[0.167,0.333]
		1	[1,4]	[0.111,0.167]
			1	[0.042,0.167]

In this example, there are only four decision makers so that there is no need to exclude any of them. However, in order to make sure how the outliers are found, the appropriate number of decision makers is assumed as three so that one of four decision makers will be excluded. The evaluation function values J_k by (17) are shown in the right column of Table 1. They represent the changes of aggregated comparisons' bounds by excluding each decision maker and are measured as the sum of increase and decrease ratios. The fact, $J_1 = 12 = n(n-1)$, tells that A_1 does not effect the bounds of aggregated comparisons. Since J_4 is the maximum of the four, the others, A_1, A_2 and A_3, are aggregated and the comparison matrix is transformed from Table 2 into Table 3. The obtained interval priority weights are also shown in Table 3. Since outlier A_4 is excluded, the interval priority weights in Table 3 is less uncertain than those in Table 2.

Table 3: Aggregated comparison matrix without outlier A_4

A (aggregated)				W
1	[2,3]	[3,4]	[4,6]	0.522
	1	[2,3]	[3,4]	[0.174,0.261]
		1	[1,4]	[0.087,0.174]
			1	[0.043,0.130]

5.2 Aggregation of priority weights

Instead of aggregating comparisons at the beginning stage, it is possible to aggregate the calculated individual preferences at the last stage. The obtained individual priority weights shown in Table 1 help the decision makers realize their own preferences, as well as understand one another.

The aggregated priority weights with and without the importance weights of decision makers are shown in Table 5. Its left column represents the case with no information on the importance of decision makers so that they are aggregated directly by (18). The results equal to those obtained by (21) with $P_k = [0, 1] \forall k$. All decision makers' individual priority weights are included in the aggregated ones, $W_k \in W^1$ $k = 1, 2, 3, 4$. In this sense, the results are easily acceptable for all decision makers. On the other hand, the right two columns represent the cases where the importance weights of decision makers are determined based on Interval AHP and the uncertainty of their giving judgments, respectively. The comparison matrix for importance of decision makers, B, is given by a supervisor and the obtained interval importance weights, P, are shown in Table 4. With the importance weights, the priority weight of each alternative W^3 is obtained as an interval by (21). The uncertainty of the given judgments, I_k, which is the

ISBN: 978-989-95079-6-8

optimal function value of (13) and the respective importance weight, p_k, calculated by (23) are shown in Table 1. The A_1 and A_4 who give the least and most uncertain information are considered as the most and least important, respectively. With the importance weights, the priority weights of alternatives W^4 are obtained as intervals by (24). The uncertainty of W^3 and W^4 is less than that of W^1 which represents complete ignorance on decision makers' importance, since their widths of intervals are smaller. The individually obtained priory weights are reflected depending on their importance among the group.

Table 4: Comparison matrix on importance of decision makers

B				P
1	2	2	4	0.462
	1	3	4	[0.231,0.317]
		1	3	[0.106,0.231]
			1	[0.077,0.115]

Table 5: Aggregated interval priority weights with and without importance weights of decision makers

W^1	Interval AHP W^3	uncertainty W^4
[0.375,0.571]	[0.508,0.517]	0.510
[0.190,0.375]	[0.239,0.265]	[0.236,0.263]
[0.091,0.190]	[0.110,0.169]	[0.108,0.170]
[0.036,0.219]	[0.062,0.131]	[0.058,0.136]

6 Conclusion

The group decision support system based on Interval AHP has been discussed. Interval AHP is particularly suitable for dealing with uncertainty of human intuitive judgments. By Interval AHP the priority weights of alternatives and importance weights of decision makers are obtained as interval from the given pariwise comparison matrix. The obtained interval priority weights reflect all the possibilities in the given information. The approaches of aggregating group members' opinions at two different stages have been proposed. At the beginning stage, the given judgments are aggregated from the possibility view and the outliers are excluded if there are some. From the aggregated comparison matrix whose elements are intervals, the priority weights of alternatives are obtained by Interval AHP as intervals. The other approach is to aggregate individual preferences at the last stage. At first, from each given comparison matrix the interval priority weights are obtained by Interval AHP. They represent the decision makers' individual preferences. Then, they are aggregated based on the possibility view or the average with the importance weights of the group members. The importance weights of decision makers can be given by a supervisor directly, from a pariwise comparison matrix for their importance or based on uncertainty of their giving information. The aggregated preferences by the proposed approaches help the group members to understand one another and reach consensus.

References

[1] G. Desanctis and R. B. Gallupe. A foundation for the study of group decision support systems. *Management Science*, 33(5):589–609, 1987.

[2] A. A. Salo. Interactive decision aiding for group decision support. *European Journal of Operational Research*, 84:134–149, 1994.

[3] T. L. Saaty. *The Analytic Hierarchy Process*. McGraw-Hill, New York, 1980.

[4] M. T. Escobarm, J. Aguaron, and J. M. Moreno-Jimenez. A note on AHP group consistency for the row geometric mean priorization procedure. *European Journal of Operational Research*, 153(2):318–322, 2004.

[5] T. L. Saaty. Group decision making and the AHP. In *B.L. Golden et.al., Editors, The Analytic Hierarchy Process: Applications and Studies*, pages 59–67. McGraw-Hill, New York, 1989.

[6] R. F. Dyer and E. H. Forman. Group decision support with the Analytic Hierarchy Process. *Decision Support Systems*, 8:94–124, 1992.

[7] E. Forman and K. Peniwati. Aggregating individual judgments and priorities with the Analytic Hierarchy Process. *European Journal of Operational Research*, 108(1):165–169, 1998.

[8] K. Sugihara and H. Tanaka. Interval evaluations in the Analytic Hierarchy Process by possibilistic analysis. *Computational Intelligence*, 17(3):567–579, 2001.

[9] K. Sugihara, H. Ishii, and H. Tanaka. Interval priorities in AHP by interval regression analysis. *European Journal of Operational Research*, 158(3):745–754, 2004.

[10] L. M. de Campos, J. F. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty*, 2(2):167–196, 1994.

[11] H. Tanaka, K. Sugihara, and Y. Maeda. Non-additive measures by interval probability functions. *Information Sciences*, 164:209–227, 2004.

[12] H. Tanaka and T. Entani. Properties of evidences based on interval probabilities obtained by pairwise comparisons in AHP. In *Proceedings of Taiwan-Japan Symposium on Fuzzy Systems and Innovational Computing*, pages 35–40, 2006.

[13] H. Tanaka and P. Guo. *Possibilistic Data Analysis for Operations Research*. Physica-Verlag, A Springer-Verlag Company, Heidelberg, New York, 1999.

[14] T. Entani and H. Tanaka. Management of ignorance by interval probabilities. In *Proceedings of 2007 IEEE International Conference on Fuzzy Systems*, pages 841–846, 2007.

[15] J. Brazilia and B. Golany. AHP rank reversal, normalization and aggregation rules. *Information Systems and Operational Research*, 32(2):14–20, 1994.

[16] R. Ramanathan and L. S. Ganesh. Group reference aggregation methods in AHP: An evaluation and an intrinsic process for deriving members' weightages. *European Journal of Operational Research*, 79(2):249–268, 1994.

[17] T. Entani and H. Tanaka. Modified interval global weight in AHP. In *Proceedings of 9th Fuzzy Days Dortmund*, pages 415–424, 2006.

ISBN: 978-989-95079-6-8

A Monge algorithm for computing the Choquet integral on set systems

Ulrich Faigle[1] Michel Grabisch[2]

1. Zentrum für Angewandte Informatik Köln, Universität zu Köln
Weyertal 80, D-50931 Köln, Germany
2. Centre d'Economie de la Sorbonne, Université Paris I 106-112, Bd de l'Hôpital, 75013 Paris, France
Email: faigle@zpr.uni-koeln.de, michel.grabisch@univ-paris1.fr

Abstract— In works dealing with capacities (fuzzy measures) and the Choquet integral on finite spaces, it is usually considered that all subsets of the universe are measureable. Hence, all functions are integrable in the sense of Choquet. We consider the situation where some subsets are not measurable (not feasible), so that there are non-integrable functions. Since this is a severe limitation in applications, we study how to extend the Choquet integral to any function. Our results mainly deal with the case where the set of feasible subsets is a distributive lattice.

Keywords— Choquet integral, fuzzy measure, capacity, Monge algorithm, set system, distributive lattice

1 Introduction

The Choquet integral [1] has become a widely used tool in decision making and its properties are well known. With general spaces, the issue of measurability is closely related to integration, that is, a function is integrable if all subsets involved in the integration are measurable. In the case of the Choquet integral, these sets are the level sets of the integrand. Although these facts are standard when considering infinite spaces, most of the time it is considered that for finite spaces, all subsets are measurable. However, there are many situations where this assumption cannot be always true. For example, if the space N is a set of players in a game, some coalitions may not be feasible. If N is a set of states of nature, some event may be impossible to realize, if it is a set of criteria, some combinations of criteria may not correspond to conceivable objects, etc.

In cooperative game theory, already a great deal of research has been done concerning games with nonfeasible coalitions, making various assumptions on the structure of feasible coalitions (distributive lattices, as in Faigle and Kern [2], and in Grabisch and Lange [3], convex geometries as in Bilbao [4], antimatroids [5], regular set systems [6], etc.).

Computing the Choquet integral over these sets systems can be done simply by taking the usual definition. However, there will be many nonintegrable functions, for which the Choquet integral remains undefined. This could be a serious limitation in applications related to the above fields. To the best of our knowledge, the question of how to define the Choquet integral for any function on these sets systems has not been addressed. This is precisely what we aim for in this paper, supposing most of the time that the set system is a distributive lattice. We introduce a Monge algorithm, whose output coincides with the Choquet integral for every integrable function. We will show that the output of this algorithm, called the Monge-Choquet integral, is the smallest extension of the Choquet integral. Dually, we introduce the greatest extension and its associated dual Monge algorithm.

2 Preliminaries

In the whole paper, $N := \{1, 2, \ldots, n\}$.

2.1 Capacities and the Choquet integral

A *set system* $\mathcal{F}$ is any collection of sets in 2^N containing the empty set and N. For any set $A \subseteq N$, we define $\mathcal{F}(A) := \{F \in \mathcal{F} \mid F \subseteq A\}$. A *game* on $\mathcal{F}$ is any function $v : \mathcal{F} \to \mathbb{R}$ such that $v(\emptyset) = 0$. A game is *monotone* if for any $A, B \in \mathcal{F}$ such that $A \subseteq B$, we have $v(A) \leq v(B)$. A monotone game is usually called a *capacity* [1] or *fuzzy measure* [7].

Consider a function $f : N \to \mathbb{R}_+$, denoting $f(i)$ by f_i. We say that f is *measurable* if the family of sets $\{j \in N \mid f(j) \geq f(i)\}, i = 1, \ldots, n$ belongs to $\mathcal{F}$.

Definition 1. Let v be a game on $\mathcal{F}$, and $f : N \to \mathbb{R}_+$ be an measurable function. The *Choquet integral* of f w.r.t. v is defined by:

$$\int f \, dv := \sum_{i=1}^{n} (f_{\sigma(i)} - f_{\sigma(i-1)}) v(\{\sigma(i), \ldots, \sigma(n)\}) \quad (1)$$

with $f_{\sigma(0)} := 0$, $f_{\sigma(1)} \leq \cdots \leq f_{\sigma(n)}$.

For any $A \subseteq N$, let 1_A denotes the characteristic vector of A. A fundamental property of the Choquet integral is that

$$\int 1_A \, dv = v(A), \quad (2)$$

for any A in $\mathcal{F}$. We give some fundamental results on the Choquet integral, useful for the sequel.

Proposition 1. The Choquet integral is monotonic vs. the game, i.e., if $v \leq v'$ pointwise, then $\int f \, dv \leq \int f \, dv'$ for any measurable f.

Two functions $f, f' \in \mathbb{R}^n_+$ are *comonotonic* if there is no $i, j \in N$ such that $f_i > f_j$ and $f'_i < f'_j$. A functional $I : \mathbb{R}^n_+ \to \mathbb{R}$ is *comonotonic additive* if for any two comonotonic functions $f, f' \in \mathbb{R}^n_+$, we have $I(f + f') = I(f) + I(f')$.

ISBN: 978-989-95079-6-8

Proposition 2. (Characterization theorem of Schmeidler, established for continuous spaces [8]) Let $I : \mathbb{R}^n_+ \to \mathbb{R}$ be a functional. Then I is the Choquet integral w.r.t. a capacity v on 2^N if and only if I is nondecreasing, comonotonic additive, and $I(0) = 0$. Then v is uniquely determined by (2).

Consider the hypercube $[0,1]^n$. The *canonical simplices* of $[0,1]^n$ are those induced by any permutation on N as follows:

$$[0,1]^n_\sigma := \{x \in [0,1]^n \mid x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}\}.$$

Proposition 3. (Interpolation theorem, see Singer [9], and [10, Ch. 5]) The unique functional $I : [0,1]^n \to [0,1]$, whose value is fixed on vertices of $[0,1]^n$, being linear in each canonical simplex, continuous, and vanishing at 0 is the Choquet integral w.r.t. a game v on 2^N, with $v(A) = I(1_A)$, for all $A \subseteq N$.

2.2 Posets, lattices, and set systems

(see, e.g., [11]) A *poset* is any set P endowed with a binary relation $\leq$, being reflexive, antisymmetric and transitive (partial order). A *lattice* $(L, \leq)$ is a poset such that for any $x, y \in L$ their least upper bound, denoted by $x \vee y$, and greatest lower bound $x \wedge y$ always exist. When L is finite, there always exist a greatest element and a least element in L, called the *top* and the *bottom* element.

Let $(P, \leq)$ be any poset. $Q \subseteq P$ is a *downset* of P if $x \in Q$ and $y \leq x$ imply $y \in Q$. The set of all downsets of P is denoted by $\mathcal{O}(P)$. Observe that if Q, Q' are downsets of P, then so are $Q \cup Q'$ and $Q \cap Q'$. For any $x \in P$, we denote by $\downarrow x$ the downset generated by x (principal ideal of x), i.e.:

$$\downarrow x := \{y \in P \mid y \leq x\}.$$

More generally, for any subset Q of P, one can compute $\downarrow Q$, the *downset generated by* Q, defined by $\downarrow Q := \bigcup_{x \in Q} \downarrow x$. It is indeed a downset since downsets are closed under union.

For any two elements $x, y \in P$, x *is covered by* y or y *covers* x (denoted by $x \prec y$ or $y \succ x$) if $x < y$ and there is no z such that $x < z < y$. A sequence of elements such that $x \leq y_1 \leq y_2 \leq \cdots \leq z$ is called a *chain* from x to z. If in addition $x \prec y_1 \prec y_2 \cdots \prec z$, the chain is *maximal*.

A lattice is *distributive* if $\vee, \wedge$ obey distributivity. Any poset $(P, \leq)$ generates a distributive lattice of subsets of P ordered by inclusion, which is $\mathcal{O}(P)$ (and reciprocally). Its bottom and top elements are $\emptyset$ and P, respectively. In the distributive lattice $\mathcal{O}(P)$, any maximal chain C from bottom to top has length $|P|$, and corresponds to a permutation σ on P, such that $C = \{\emptyset, \{\sigma(n)\}, \{\sigma(n), \sigma(n-1)\}, \ldots, P\}$.

Let us consider as in Section 2.1 the set N of n elements. By defining a partial order $\leq$ on N, $(N, \leq)$ is a poset and $\mathcal{O}(N)$ is a distributive lattice with $\emptyset, N$ as bottom and top elements, hence it is a set system, having all its maximal chains from $\emptyset$ to N of length n. Conversely, a set system whose all maximal chains from $\emptyset$ to N are of length n is not necessarily a distributive lattice. It is called a *regular set system* [6]. An *antimatroid* $\mathcal{F}$ is a regular set system which is closed under union. A *convex geometry* is a regular set system closed under intersection.

2.3 The Monge algorithm

The original idea of the Monge algorithm goes back to [12]. Monge studied a geometric transportation problem in which a set of locations $s_1, \ldots, s_n$ of mass points has to be matched optimally (in the sense of minimizing the total cost) with another set of locations $t_1, \ldots, t_n$, and proved that optimality was reached if the transportation lines do not cross. This geometric fact can be expressed as follows: if the costs c_{ij} of matching objects s_i with t_j have the "uncrossing" property:

$$c_{ij} + c_{k\ell} \geq c_{\max(i,k),\max(j,\ell)} + c_{\min(i,k),\min(j,\ell)}$$

then the optimal matching is $(s_1, t_1), \ldots, (s_n, t_n)$. This is also called the "north-west corner rule". Translated into the language of set functions, the uncrossing property is in fact submodularity:

$$v(A) + v(B) \geq v(A \cup B) + v(A \cap B).$$

The following algorithm, which we call "Monge algorithm", is based on the previous ideas (see also the greedy algorithm of Lovász for maximizing a linear function over the core of a submodular game [13]). Let $A = [a_{ij}] \in \{0,1\}^{m \times n}$ be a $(0,1)$-matrix with m rows $A_1, \ldots, A_m$ and n columns $A^1, \ldots, A^n$. We assume throughout the paper that A does not contain the null row. Let us put $N = \{1, \ldots, n\}$ the set of column indices. Hence, any row of the matrix can be interpreted as the characteristic vector of a nonempty subset of N.

Consider an (input) vector $f \in \mathbb{R}^n_+$. The Monge algorithm produces an output line-vector $y \in \mathbb{R}^m_+$ as follows.

MONGE ALGORITHM

(M0) Initialization. Put $f' = f$, $A' = [a'_{ij}] = A$, $N' = N$, $y = 0 \in \mathbb{R}^m$.

(M1) If the current matrix A' is empty, STOP.

(M2) Let s be the smallest row index in A'.

(M3) Find the smallest value α of f'_j for all $j \in N'$ such that $a'_{sj} = 1$. Put t the smallest such j.

(M4) Updating: $y_s \leftarrow \alpha$, $f'_j \leftarrow f'_j - \alpha$ for all $j \in N'$ such that $a'_{sj} = 1$, $N' \leftarrow N' \setminus t$, suppress all rows A'_i in A' such that $a'_{it} = 1$. Go to (M1).

3 The Monge algorithm and the Choquet integral

Consider a set system $\mathcal{F}$ on N. It can be encoded by a $(0,1)$-matrix A as given in Section 2.3. The rows of A order the subsets of $\mathcal{F}$ in a particular way. An interesting case arises when this order is a linear extension of the decreasing inclusion order, that is, for any subsets S, T such that $S \subseteq T$, the corresponding rows A_{i_S}, A_{i_T} are such that $i_S \geq i_T$. We say that in this case the row order *is compatible with decreasing inclusion*. A row order compatible with increasing inclusion can be defined as well.

Consider next a game v on $\mathcal{F}$. It can be encoded as a vector in $\mathbb{R}^m$, with $m = |\mathcal{F}| - 1$. The next result shows that the Choquet integral can be computed by yv, where y is the output of the Monge algorithm, provided rows of A are ordered according to the (decreasing) inclusion order.

ISBN: 978-989-95079-6-8

Proposition 4. Let $\mathcal{F}$ be a set system, and an measurable function $f \in [0,1]^N$. Let v be a game on $\mathcal{F}$.

On the other hand, consider a (0,1)-matrix A encoding the set system $\mathcal{F}$ (except the empty set), such that the row order is compatible with decreasing inclusion. Then the output of the Monge algorithm for f is a vector y, such that yv is the Choquet integral defined in (1).

A question is whether the condition on row ordering is necessary to get this result. The answer is given in the next proposition.

Proposition 5. Assume A is a (0-1)-matrix with no two identical rows. For each nonempty subset S corresponding to a row of A, say A_{i_S}, the output of the algorithm is $y = 1_{i_S}$ (i.e., $y_i = 0$ for all i except for $i = i_S$) when the input is $f = 1_S$, if and only if the row order is compatible with decreasing inclusion.

In other words, for any set system $\mathcal{F}$ and any game v on $\mathcal{F}$, for $f = 1_S$, $S \in \mathcal{F}$, we have $yv = v(S)$, which is by (2) the output of the Choquet integral. This proves the necessity of the condition on the row order in Proposition 4.

4 Extended Choquet integrals for distributive lattices

For any set system $\mathcal{F}$, we define the *Monge-Choquet integral* as the output of the Monge algorithm for any nonnegative input vector. In this section, we restrict to the case of distributive lattices induced by a poset on N, denoted by $(N, \leq)$. We know from Section 3 that the Monge-Choquet integral coincides with the usual Choquet integral for measurable input vectors, provided the rows of A are arranged in an order extending the decreasing inclusion order (otherwise stated, this assumption holds from now on). Hence, the Monge-Choquet integral is an extension of the Choquet integral.

4.1 Computation of the Monge-Choquet integral

We show that the Monge-Choquet integral can be computed in a much simpler way, independent of the order of rows of A, provided this order is compatible with decreasing inclusion. Take f any input vector in $\mathbb{R}^n_+$. We can assume w.l.o.g. (this will be shown later on) that there is a unique permutation σ on N such that $f_{\sigma(1)} < \cdots < f_{\sigma(n)}$.

In the first step of the algorithm, we have $s = 1$, $\alpha = f_{\sigma(1)}$, and $t = \sigma(1)$. In the (initially null) vector y, we put $f_{\sigma(1)}$ at position 1, $\sigma(1)$ is deleted from N, any row in A containing $\sigma(1)$ is deleted, and $f' = f - f_{\sigma(1)}$.

In the second step, s corresponds to the largest subset F in $\mathcal{F}$ not containing $\sigma(1)$, which is uniquely determined by

$$F = \{i \in N \mid i \not\geq \sigma(1)\}$$

where $\geq$ is understood for the poset $(N, \leq)$. $F \in \mathcal{F}$ because F is a downset of $(N, \leq)$, and it is the unique largest such set in $\mathcal{F}$ because $\mathcal{F}$ is closed under union. Since it is the largest one not containing $\sigma(1)$, it is ranked first in A', hence s corresponds to this subset. Next, we look for $\sigma(i) \in F$ such that f' is minimum on F, and we put $t = \sigma(i)$, $\alpha = f_{\sigma(i)} - f_{\sigma(1)}$, and α is put in y at position s. Then $\sigma(i)$ is deleted, and all rows of A' containing $\sigma(i)$. The process continues till A' is empty.

The above shows that the following algorithm computes the Monge-Choquet integral in a simpler way. We recall that for any $A \subseteq N$, $\mathcal{F}(A) := \{F \in \mathcal{F} \mid F \subseteq A\}$.

COMPUTATION OF THE MONGE-CHOQUET INTEGRAL FOR DISTRIBUTIVE LATTICES

(MC0) Initialization: $N' \leftarrow N$.

(MC1) Take F to be the largest subset in $\mathcal{F}(N')$. It is given by $F = \{i \in N \mid i \not\geq j, \forall j \in N \setminus N'\}$. If $F = \emptyset$, goto (MC3).

(MC2) Find the minimum of f over F, call i the smallest index in F for which f is minimum. Put $N' \leftarrow N' \setminus i$. Go to (MC1).

(MC3) Denote by $F_1 = N, F_2, \ldots, F_k$ the sequence of subsets obtained in (MC1), and by $i_1, i_2, \ldots, i_k$ the sequence of indices obtained in (MC2). Then

$$\int f\, dv = \sum_{j=1}^{k} (f_{i_j} - f_{i_{j-1}}) v(F_i), \quad (3)$$

with $f_{i_0} := 0$.

Example 1. Take $n = 5$, and the poset represented on Fig. 1 (left). The corresponding set system $\mathcal{F}$ is on the right. Many

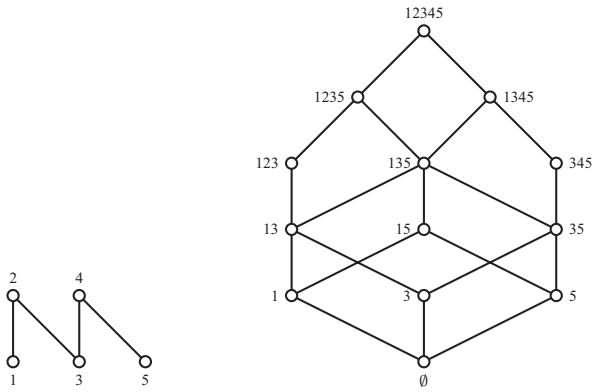


Figure 1: Example of poset with $N = \{1, 2, 3, 4, 5\}$ (left) and the corresponding set system $\mathcal{F}$.

permutations are not compatible with $\mathcal{F}$ (i.e., the induced maximal chain contains sets which are not in $\mathcal{F}$), for example 35241, 12345, 43125, etc. We can anyway compute the corresponding Monge-Choquet integrals:

$$35241 : \int f\, dv = f_3 v(N) + (f_5 - f_3) v(15) + (f_1 - f_5) v(1)$$

$$12345 : \int f\, dv = f_1 v(N) + (f_3 - f_1) v(345) + (f_5 - f_3) v(5)$$

$$43125 : \int f\, dv = f_4 v(N) + (f_3 - f_4) v(1235) + (f_1 - f_3) v(15) + (f_5 - f_1) v(5).$$

ISBN: 978-989-95079-6-8

4.2 Properties of the Monge-Choquet integral

The most important property to check is whether the functional is continuous. Obviously it is for a given permutation, but when passing from a permutation to another one, we have to check continuity. If continuity does not hold, we could even say that the integral is not well defined, because a given f may be rearranged in increasing order by several permutations.

Let us take f with permutation σ, such that

$$f_{\sigma(1)} < f_{\sigma(2)} < \cdots < f_{\sigma(i)} = f_{\sigma(i+1)} < \cdots < f_{\sigma(n)}.$$

Hence the permutation σ' obtained by a switch between $\sigma(i)$ and $\sigma(i+1)$ is also rearranging f in nondecreasing order. Observe that in any case, all "compatible" permutations can be obtained by a sufficient number of switches, hence it is enough to study what happens for one switch.

To avoid a heavy notation, and without loss of generality, we may assume that the two following orders are compatible with f:

$$1, 2, \ldots, i, i+1, \ldots, n$$
$$1, 2, \ldots, i+1, i, \ldots, n.$$

The first and second orders will be used by the algorithm when computing the integral of $f' := f + \epsilon 1_{i+1}$ and $f'' := f + \epsilon 1_i$ respectively. Let us check whether the two expressions coincide when ϵ tends towards 0. Clearly, the sequences $F'_1, \ldots, F'_{k'}$ and $F''_1, \ldots, F''_{k''}$ for f' and f'' coincide as long as both $i, i+1$ belong to the sets, and also from the point where both $i, i+1$ disappear. Let us denote respectively by F_0 and F^0 these sets (note that $F^0 = \emptyset$ is possible). Let i_0 be the index of (MC2) when $F = F_0$. Then several cases can occur.

(i) $i_0 < i$ and $i_0 < i+1$ (in the poset). Then F in the next step (MC1) will contain neither i nor $i+1$ for both f', f'' (hence $F = F^0$). Consequently, there will be no difference between the two expressions.

(ii) $i_0 < i$, $i_0 \not< i+1$. Then, the next terms obtained while processing $i, i+1$ will be:

$$\text{for } f'\text{: } (f'_{i+1} - f'_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\})$$
$$\text{for } f''\text{: } (f''_{i+1} - f''_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\}).$$

(iii) $i_0 < i+1$, $i_0 \not< i$. Similarly,

$$\text{for } f'\text{: } (f'_i - f'_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\})$$
$$\text{for } f''\text{: } (f''_i - f''_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\}).$$

(iv) $i_0 \not< i$, $i_0 \not< i+1$, and $i < i+1$. Then, the next terms obtained while processing $i, i+1$ will be:

$$\text{for } f'\text{: } (f'_i - f'_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\})$$
$$\text{for } f''\text{: } (f''_{i+1} - f''_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\}) +$$
$$(f''_i - f''_{i+1})v(F_0 \cap \{j \in N \mid j \not\geq i_0, j \not\geq i+1\}).$$

(v) $i_0 \not< i$, $i_0 \not< i+1$, and $i+1 < i$. Same as above, permuting the roles of f' and f''.

(vi) $i_0 \not< i$, $i_0 \not< i+1$, and $i || i+1$. Then we obtain

$$\text{for } f'\text{: } (f'_i - f'_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\}) +$$
$$(f'_{i+1} - f'_i)v(F_0 \cap \{j \in N \mid j \not\geq i_0\}, j \not\geq i)$$
$$\text{for } f''\text{: } (f''_{i+1} - f''_{i_0})v(F_0 \cap \{j \in N \mid j \not\geq i_0\}) +$$
$$(f''_i - f''_{i+1})v(F_0 \cap \{j \in N \mid j \not\geq i_0, j \not\geq i+1\}).$$

In all cases, the expressions become identical when ϵ tends towards 0.

In addition, this proves that in the algorithm, when taking the smallest i when several such i's exist, this choice is unimportant.

From continuity and (3), we easily deduce the following:

Proposition 6. The Monge-Choquet integral is nondecreasing w.r.t. the integrand, positively homogeneous, and is comonotonic additive.

4.3 Smallest and largest extended Choquet integrals

From (3) and the algorithm, it is easy to compute the Monge-Choquet integral for any binary function 1_A, $A \subseteq N$. From Proposition 5, we already know that for any $F \in \mathcal{F}$, $\int 1_F \, dv = v(F)$. Let us define $\hat{v}(A) := \int 1_A \, dv$, for any $A \subseteq N$, which could be seen as an extension of v on 2^N.

Proposition 7. For any $A \subseteq N$, $\hat{v}(A)$ is obtained by

$$\hat{v}(A) = v(F), \quad \text{with } F \text{ the largest subset of } \mathcal{F}(A).$$

Moreover, if v is a capacity, then so is $\hat{v}$, and it is the smallest extension of v over 2^N.

From continuity of the Monge-Choquet integral and Proposition 3, we deduce the following fundamental result.

Proposition 8. For any $f \in \mathbb{R}^n_+$,

$$\int f \, dv = \int f \, d\hat{v}$$

where the left integral is the Monge-Choquet integral, and the right one, the classical Choquet integral.

Hence, the Monge-Choquet integral inherits all properties from the Choquet integral w.r.t. the smallest extension of v over 2^N. In particular, by monotonicity of the integral w.r.t. the capacity (see Proposition 1), we get by Proposition 2:

Corollary 1. Let v be a capacity on $\mathcal{F}$. The Monge-Choquet integral w.r.t. v is the smallest functional $I : \mathbb{R}^n_+ \to \mathbb{R}$ being nondecreasing, comonotonic additive, and such that $I(1_F) = v(F)$ for each F in $\mathcal{F}$.

It is easy to define the largest extension of v over 2^N, denoted by $\check{v}$. It is given by

$$\check{v}(A) := v(\downarrow A)$$

where $\downarrow A$ is the downset generated by A in the poset $(N, \leq)$. Indeed, this is the smallest set in $\mathcal{F}$ containing A. It is possible to obtain $\check{v}$ by an algorithm which is dual to the Monge algorithm.

ISBN: 978-989-95079-6-8

DUAL MONGE ALGORITHM. We assume that the row order is compatible with *increasing* inclusion.

(DM0) Init: $A' = [a'_{ij}] = A$, $y = 0 \in \mathbb{R}^m$, $\alpha =$largest value of f, $N' = N$.

(DM1) Choose the smallest index $t \in N'$ such that $f_t = \alpha$. Put $N' \leftarrow N' \setminus t$.

(DM2) Suppress all rows in A' not containing t (i.e., $a'_{s,t} = 0$). Take the smallest row index s in A'.

(DM3) If $N' = \emptyset$ or s is the last row of A' (i.e., it corresponds to N), put $y_s = f_t$ and STOP. Otherwise, look for the highest value α of f_j for all j such that $a'_{s,j} = 0$, put $y_s = f_t - \alpha$, and go to (DM1).

Again, due to the structure of F, a simpler version can be given.

COMPUTATION OF THE DUAL MONGE-CHOQUET INTEGRAL FOR DISTRIBUTIVE LATTICES

(DMC0) Put $N' = \{i_0\}$, where i_0 is the smallest index maximising f.

(DMC1) Take F the smallest set in $\mathcal{F}$ containing N', which is given by $F =\downarrow N'$. If $F = N$, go to (DMC3).

(DMC2) Find the smallest index i maximising f over $N \setminus F$. Put $N' \leftarrow N' \cup \{i\}$. Go to (DMC1).

(DMC3) Denote by $F_1, \ldots, F_k$ and $i_0, i_1, \ldots, i_{k-1}$ the sequence of sets obtained in (DMC1) and the sequence of indices obtained in (DMC0) and (DMC2) respectively. Then

$$\int f\, dv = \sum_{j=1}^{k} (f_{i_{j-1}} - f_{i_j}) v(F_j) \qquad (4)$$

with $f_{i_k} := 0$.

Example 2. Taking the same poset as in Example 1, we obtain:

$$35241 : \int f\, dv = (f_1 - f_4)v(1) + (f_4 - f_2)v(1345) + f_2 v(N)$$

$$12345 : \int f\, dv = (f_5 - f_4)v(5) + (f_4 - f_2)v(345) + f_2 v(N)$$

$$43125 : \int f\, dv = (f_5 - f_2)v(5) + (f_2 - f_4)v(1235) + f_4 v(N).$$

Proposition 9. The dual Monge-Choquet integral is continuous.

The consequence is that, applying the same reasoning as above, we can deduce the following.

Proposition 10. (i) For any $f \in \mathbb{R}^n_+$,

$$\int^* f\, dv = \int f\, d\check{v}$$

where the left integral is the dual Monge-Choquet integral, and the right one, the classical Choquet integral.

(ii) If v is a capacity on $\mathcal{F}$, then the dual Monge-Choquet integral is the greatest functional $I : \mathbb{R}^n_+ \to \mathbb{R}_+$ being nondecreasing, comonotonic additive, and such that $I(1_F) = v(F)$ for each $F \in \mathcal{F}$.

(iii) For any measurable f, the Monge-Choquet integral and the dual Monge-Choquet integral coincide.

5 The case of regular set systems, antimatroids and convex geometries

5.1 Regular set systems

We know from Section 2.1 that a function is measurable if it exists a permutation σ on N ordering f such that the induced sets $\{\sigma(i), \ldots, \sigma(n)\}$, $i = 1, \ldots, n$, form a maximal chain from $\emptyset$ to N in $\mathcal{F}$, of length n. Hence if $\mathcal{F}$ does not contain maximal chains of length n, there will be no measurable function. This has motivated the definition of regular set systems, whose maximal chains from $\emptyset$ to N are all of length n. In this respect, it would be interesting to know what happens for nonmeasurable functions over regular set systems. It is not difficult to see that, unfortunately, the output of the Monge algorithm depends on the row order, even if it is compatible with inclusion (see example below). The reason is that a regular set system need not be closed under union (nor under intersection), and this property is fundamental for proving results in Section 4. Therefore, the Monge-Choquet integral is not well-defined in this case.

Example 3. Consider $n = 5$ and the regular set system of Fig. 3. Remark that it is not closed under union nor intersection. Hence it is neither an antimatroid nor a convex geometry. Take any function f such that $f_4 < f_3 < f_2 < f_1 < f_5$. Then,

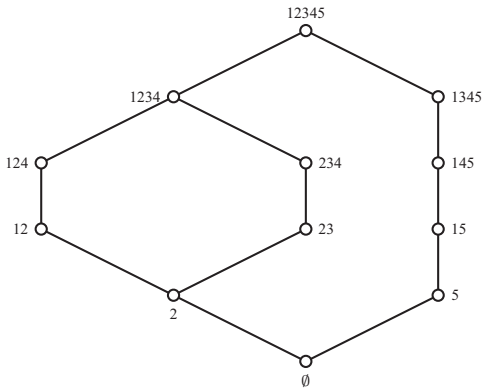


Figure 2: Example of regular set system

in the 1st iteration, $s = 1$, $\alpha = f_4$, $t = 4$, and $f' = f - f_4$. At second iteration, 4 has been discarded, so we have 3 candidates for the next subset: 12, 23, and 15. If 12 is ranked first in the matrix A', we get $\alpha = f_2$, $t = 2$, and 2 is discarded, so only 1 remains. Finally, we get:

$$\int d\, dv = f_4 v(N) + (f_2 - f_4)v(12) + (f_1 - f_2)v(1).$$

ISBN: 978-989-95079-6-8

Now, if 23 is ranked first, $\alpha = f_3$, $t = 3$, and we will get:

$$\int f\, dv = f_4 v(N) + (f_3 - f_4)v(23) + (f_2 - f_3)v(2).$$

Hence, the result depends on the ordering.

5.2 Antimatroids and convex geometries

Antimatroids are regular set systems closed under union. This crucial property (as exemplified above) makes most of previous results to hold. More precisely, there will be always a unique largest subset in $\mathcal{F}(N')$ for all $N' \subseteq N$. Consequently, the algorithm of computation of the Monge-Choquet integral and especially Equation (3) given in Section 4.1 remain valid, except for Step (MC1), where F is still the largest subset in $\mathcal{F}(N')$, but now it is no more possible to give it explicitly.

An important question is whether continuity still holds. Using a similar argument, one can prove that continuity still holds in the case of antimatroids. A sketch of the proof goes as follows. Keeping the same notation as in Section 4.2, consider functions f', f'', whose orderings differ only on i and $i+1$. There is a common chain from N to F_0 for both functions f' and f''. From F_0 which contains both i and $i+1$, functions f' and f'' may have different chains, but these chains will necessarily rejoin in F^0 (because there is a unique largest subset contained in $\{i+2, \ldots, n\}$). At F_0, the new term in the integral is $(f'_i - f'_k)v(F_0)$ for f' and $(f''_{i+1} - f''_k)v(F_0)$ for f'' for some k among $1, \ldots, i-1$. At F^0, the new term is $(f'_k - f'_{i+1})v(F^0)$ for f' and $(f''_k - f''_i)v(F^0)$ for f'' for some k in F^0, assuming $i, i+1$ are present in the preceding step (other cases work similarly). Between F_0 and F^0, there is at most one set F on each chain, where the new term will be $(f'_{i+1} - f'_i)v(F)$, and similarly for f''. When ϵ tends towards 0, both integrals coincide.

Consequently, Propositions 6, 7, 8, and Corollary 1 still hold.

On the contrary, it is no more possible to define the largest extension $\check{v}$ and consequently the dual Monge-Choquet integral, because since an antimatroid is not closed under intersection, it may happen that there is no smallest set of $\mathcal{F}$ containing in a given subset of N.

In a dual way, since convex geometries are closed under intersection but not under union, the dual Monge-Choquet integral exists and possesses all properties given in Section 4.3, while the Monge-Choquet integral is not well-defined.

References

[1] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[2] U. Faigle and W. Kern. The Shapley value for cooperative games under precedence constraints. *Int. J. of Game Theory*, 21:249–266, 1992.

[3] M. Grabisch and F. Lange. Games on lattices, multichoice games and the Shapley value: a new approach. *Mathematical Methods of Operations Research*, 65:153–167, 2007.

[4] J.M. Bilbao and P.H. Edelman. The Shapley value on convex geometries. *Discrete Applied Mathematics*, 103:33–40, 2000.

[5] E. Algaba, J. M. Bilbao, R. van den Brink, and A. Jiménez-Losada. Cooperative games on antimatroids. *Discrete Mathematics*, 282:1–15, 2004.

[6] A. Honda and M. Grabisch. An axiomatization of entropy of capacities on set systems. *Eur. J. of Operational Research*, 190:526–538, 2008.

[7] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[8] D. Schmeidler. Integral representation without additivity. *Proc. of the Amer. Math. Soc.*, 97(2):255–261, 1986.

[9] I. Singer. Extensions of functions of 0-1 variables and applications to combinatorial optimization. *Numerical Functional Analysis and Optimization*, 7(1):23–62, 1984.

[10] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation functions*. Number 127 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2009.

[11] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Orders*. Cambridge University Press, 1990.

[12] G. Monge. Déblai et remblai. Mém. de l'Académie des Sciences, Paris, 1781.

[13] L. Lovász. Submodular functions and convexity. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical programming. The state of the art*, pages 235–257. Springer Verlag, 1983.

ISBN: 978-989-95079-6-8

Decomposition of Interval-Valued Fuzzy Morphological Operations by Weak $[\alpha_1, \alpha_2]$-cuts

Tom Mélange, Mike Nachtegael, Etienne Kerre[1] Peter Sussner[2]

1.Ghent University, Dept. of Applied Math. and Computer Science
Krijgslaan 281-S9, 9000 Ghent, Belgium
2.University of Campinas, Dept. of Applied Mathematics
Campinas, SP, 13083-859, Brazil
Email: {tom.melange,mike.nachtegael,etienne.kerre}@ugent.be, sussner@ime.unicamp.br

Abstract— *One of the extensions of binary morphology to greyscale images is given by the classical fuzzy mathematical morphology. Interval-valued fuzzy mathematical morphology further extends the latter theory by now also allowing uncertainty in the grey values of the image. In this paper, the decomposition of the interval-valued fuzzy morphological operations into their binary counterparts is studied both in a general continuous framework and a discrete framework. It will be shown that some properties that do not hold in the continuous framework, do hold in the discrete framework, which is the framework that is used in practice.*

Keywords— Decomposition, interval-valued fuzzy sets, mathematical morphology.

1 Introduction

The aim of many image processing theories is to extract specific information out of an image. Mathematical morphology is one of those theories. The basic morphological operations dilation, erosion, opening and closing transform a given image into another image by the help of a structuring element. In the original binary morphology [1], both the image and the structuring element needed to be binary images (i.e., black-white). The threshold approach [1] extended the binary morphology by allowing greyscale images. Another extension, the umbra approach [2], even allowed both greyscale images and structuring elements. Some time later, one more greyscale approach was introduced, based on fuzzy set theory [3, 4]. Greyscale images and fuzzy sets can namely both be modelled as a mapping from a universe $\mathcal{U}$ into the unit interval $[0, 1]$. Remark however that fuzzy set theory is only used as a tool here and not to deal with uncertainty. Recently, also extensions of the fuzzy mathematical morphology have arised [5, 6]. The interval-valued fuzzy extension that we consider in this paper, now allows uncertainty regarding the grey values in the image, since a coordinate in the image domain is no longer mapped onto a specific grey value belonging to the unit interval $[0, 1]$, but onto an interval of grey values ($\subseteq [0, 1]$) to which the uncertain grey value is expected to belong. The uncertainty in grey values can have several causes. (i) In practice, we have to deal with the fact that any device rounds off a captured grey value to an element of a finite set of allowed values; (ii) Taking several shots of the same image may result in different grey values for some of the image pixels. This can sometimes happen under identical recording circumstances and is very likely to be the case under changing circumstances (e.g. illumination changes due to clouds passing the sun). The difference in grey value can also be the result of a slight shift in position of the camera or an object in the image between takes. This will especially cause (mostly large) uncertainty at pixels belonging to the edge of an object. (iii) There might also exist uncertainty regarding the grey values in the structuring element that is used. This structuring element can be chosen by the user and sometimes the user might doubt which weight (importance) to give to a certain pixel in the structuring element. In all of the above cases, it might be advisible to work with intervals of likely values instead of one specific value. The interval-valued fuzzy set theory now thus also serves as a model and not only as a tool [6].

In this paper, the relationships between the weak $[\alpha_1, \alpha_2]$-cuts of the interval-valued fuzzy dilation, erosion, opening and closing and the corresponding binary operations are studied. Such decomposition properties are useful (i) to compute interval-valued fuzzy morphological operations by applying several binary operations and combining the obtained results and (ii) to approximate interval-valued fuzzy morphological operations by only considering a finite number of $[\alpha_1, \alpha_2]$-cuts.

The remainder of the paper is organized as follows: the basic notions on interval-valued fuzzy mathematical morphology are introduced in section 2; section 3 studies the decomposition of the interval-valued fuzzy morphological operations by weak $[\alpha_1, \alpha_2]$-cuts in a continuous framework and investigates whether a discrete framework affects those results. Section 4 finally, concludes the paper.

2 Interval-Valued Fuzzy Mathematical Morphology

2.1 Interval-valued Fuzzy Set Theory

A fuzzy set [7] F defined over a universe $\mathcal{U}$, maps every element $u \in \mathcal{U}$ onto its membership degree $F(u) \in [0, 1]$ in that set F and allows in this way a gradual transition between belonging to (membership degree 1) and not belonging to (membership degree 0). Interval-valued fuzzy sets [8] extend the classical fuzzy sets by allowing uncertainty concerning the membership degree. They map an element of the universe onto an interval of values instead of one specific membership degree. In other words, they are modelled by mappings from the universe $\mathcal{U}$ into the class of closed intervals $L^I = \{[x_1, x_2] | [x_1, x_2] \subseteq [0, 1]\}$. This means that for an interval-valued fuzzy set G in a universe $\mathcal{U}$, $G(u) = [G_1(u), G_2(u)] \subseteq [0, 1]$, $\forall u \in \mathcal{U}$. In this paper, the class of

ISBN: 978-989-95079-6-8

interval-valued sets over the universe $\mathcal{U}$ will be denoted by $\mathcal{IVFS}(\mathcal{U})$. The lower and upper bound of an element x of L^I is denoted by respectively x_1 and x_2, i.e., $x = [x_1, x_2]$. Further, we will restrict the universe $\mathcal{U}$ in the sequel to $\mathbb{R}^n$, corresponding to the coordinates of an n-dimensional image.

Consider now the following partial ordering $\leq_{L^I}$ on L^I:

$$x \leq_{L^I} y \Leftrightarrow x_1 \leq y_1 \text{ and } x_2 \leq y_2,\ \forall x, y \in L^I.$$

It can be shown that $(L^I, \leq_{L^I})$ forms a complete lattice [9]. The infimum and supremum of a subset S of L^I are then given by $\inf S = [\inf_{x \in S} x_1, \inf_{x \in S} x_2]$ and $\sup S = [\sup_{x \in S} x_1, \sup_{x \in S} x_2]$ respectively. In the remainder, we will use the notations 0_{L^I} for $\inf L^I = [0, 0]$ and 1_{L^I} for $\sup L^I = [1, 1]$ and we will also use the following orderings on L^I ($\forall x, y \in L^I$):

$$\begin{aligned} x \ll_{L^I} y &\Leftrightarrow x_1 < y_1 \text{ and } x_2 < y_2, \\ x \geq_{L^I} y &\Leftrightarrow y \leq_{L^I} x, \\ x \gg_{L^I} y &\Leftrightarrow y \ll_{L^I} x. \end{aligned}$$

2.2 Binary and Interval-Valued Fuzzy Mathematical Morphological Operations

In this paper, the decomposition of the interval-valued fuzzy morphological operations into their binary counterparts is studied. Therefore, we first refresh the binary morphological operations.

Definition 1. [*I*] *Let $A, B \subseteq \mathbb{R}^n$. The binary dilation $D(A, B)$, erosion $E(A, B)$, closing $C(A, B)$ and opening $O(A, B)$ are the sets given by:*

$$\begin{aligned} D(A, B) &= \{y | T_y(B) \cap A \neq \emptyset\}, \\ E(A, B) &= \{y | T_y(B) \subseteq A\}, \\ C(A, B) &= E(D(A, B), -B), \\ O(A, B) &= D(E(A, B), -B), \end{aligned}$$

with $T_y(B) = \{x \in \mathbb{R}^n | x - y \in B\}$ and $-B = \{-b | b \in B\}$.

It is clear that the notions of intersection and inclusion are quite important in the above definitions. Hence, to extend the binary morphological operations to interval-valued fuzzy operators, the underlying Boolean conjunction and implication are extended by conjunctors and implicators on L^I. First, however, we extend the Boolean negation by a negator on L^I.

A *negator* $\mathcal{N}$ on L^I is a decreasing $L^I - L^I$ mapping that coincides with the Boolean negation on $\{0, 1\}$ ($\mathcal{N}(0_{L^I}) = 1_{L^I}$ and $\mathcal{N}(1_{L^I}) = 0_{L^I}$). If $(\forall x \in L^I)(\mathcal{N}(\mathcal{N}(x)) = x)$, then the negator $\mathcal{N}$ is called *involutive*. The standard negator $\mathcal{N}_s$, given by $\mathcal{N}_s([x_1, x_2]) = [1-x_2, 1-x_1]$, for all $x = [x_1, x_2] \in L^I$, is an example of an involutive negator on L^I.

A *conjunctor* $\mathcal{C}$ on L^I is an increasing $(L^I)^2 - L^I$ mapping that coincides with the Boolean conjunction on $\{0, 1\}^2$ ($\mathcal{C}(0_{L^I}, 0_{L^I}) = \mathcal{C}(0_{L^I}, 1_{L^I}) = \mathcal{C}(1_{L^I}, 0_{L^I}) = 0_{L^I}$ and $\mathcal{C}(1_{L^I}, 1_{L^I}) = 1_{L^I}$). If a conjunctor $\mathcal{C}$ satisfies $(\forall x \in L^I)(\mathcal{C}(1_{L^I}, x) = \mathcal{C}(x, 1_{L^I}) = x)$, then it is called a *semi-norm* on L^I. Furthermore, if a semi-norm $\mathcal{C}$ is commutative and associative, then we call it a *t-norm* on L^I. The conjunctor $\mathcal{C}_{\min}$, defined by $\mathcal{C}_{\min}(x, y) = [\min(x_1, y_1), \min(x_2, y_2)]$, for all $(x, y) \in (L^I)^2$, is an example of a t-norm on L^I.

Finally, an *implicator* $\mathcal{I}$ on L^I is a hybrid monotonic $(L^I)^2 - L^I$ mapping (i.e., decreasing in the first argument and increasing in the second argument) that coincides with the Boolean implication on $\{0, 1\}^2$ ($\mathcal{I}(0_{L^I}, 0_{L^I}) = \mathcal{I}(0_{L^I}, 1_{L^I}) = \mathcal{I}(1_{L^I}, 1_{L^I}) = 1_{L^I}$ and $\mathcal{I}(1_{L^I}, 0_{L^I}) = 0_{L^I}$). It can be checked that for every implicator $\mathcal{I}$, the operation $\mathcal{N}_\mathcal{I}$ defined by $\mathcal{N}_\mathcal{I}(x) = \mathcal{I}(x, 0_{L^I}), \forall x \in L^I$, is a negator on L^I. If an implicator $\mathcal{I}$ satisfies $(\forall x \in L^I)(\mathcal{I}(1_{L^I}, x) = x)$, then it is called a *border implicator* on L^I. Furthermore, if a border implicator $\mathcal{I}$ is contrapositive w.r.t. its induced negator, i.e., $(\forall (x, y) \in (L^I)^2)(\mathcal{I}(x, y) = \mathcal{I}(\mathcal{N}_\mathcal{I}(y), \mathcal{N}_\mathcal{I}(x)))$, and if it fulfills the exchange principle, i.e., $(\forall (x, y, z) \in (L^I)^3)(\mathcal{I}(x, \mathcal{I}(y, z)) = \mathcal{I}(y, \mathcal{I}(x, z)))$, then we call it a *model implicator* on L^I. The implicator $\mathcal{I}_{\max, \mathcal{N}_s}$ given by $\mathcal{I}_{\max, \mathcal{N}_s}(x, y) = [\max(1 - x_2, y_1), \max(1 - x_1, y_2)]$, for all $(x, y) \in (L^I)^2$, is an example of a model implicator on L^I.

Using the above concepts, we can extend the binary morphological operations to the interval-valued fuzzy case.

Definition 2. *Let $\mathcal{C}$ be a conjunctor on L^I, let $\mathcal{I}$ be an implicator on L^I, and let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$. The interval-valued fuzzy dilation $D^I_\mathcal{C}(A, B)$, erosion $E^I_\mathcal{I}(A, B)$, closing $C^I_{\mathcal{C},\mathcal{I}}(A, B)$ and opening $O^I_{\mathcal{C},\mathcal{I}}(A, B)$ are the interval-valued fuzzy sets in $\mathbb{R}^n$ defined for all $y \in \mathbb{R}^n$ by:*

$$\begin{aligned} D^I_\mathcal{C}(A, B)(y) &= \sup_{x \in T_y(d_B) \cap d_A} \mathcal{C}(B(x - y), A(x)), \\ E^I_\mathcal{I}(A, B)(y) &= \inf_{x \in T_y(d_B)} \mathcal{I}(B(x - y), A(x)), \\ C^I_{\mathcal{C},\mathcal{I}}(A, B)(y) &= E^I_\mathcal{I}(D^I_\mathcal{C}(A, B), -B)(y), \\ O^I_{\mathcal{C},\mathcal{I}}(A, B)(y) &= D^I_\mathcal{C}(E^I_\mathcal{I}(A, B), -B)(y), \end{aligned}$$

with $d_A = \{x | x \in \mathbb{R}^n$ and $A(x) \neq 0_{L^I}\}$, $d_B = \{x | x \in \mathbb{R}^n$ and $B(x) \neq 0_{L^I}\}$ and $(-B)(x) = B(-x), \forall x \in \mathbb{R}^n$.

Remark that if $y \notin D(d_A, d_B)$, then $D^I_\mathcal{C}(A, B)(y) = 0_{L^I}$.

In [11] it is shown that fuzzy mathematical morphology is compatible with binary morphology and if we restrict ourselves to semi-norms and border implicators it is also compatible with greyscale morphology based on the threshold approach. Since the interval-valued fuzzy morphology is compatible with the fuzzy morphology and because we want to preserve also the compatibility with the threshold approach, we will restrict ourselves in the remainder to semi-norms and border implicators on L^I.

2.3 The Discrete Framework

For the practical processing of an image, one has to deal with the technical limitations of a computer. To store an image, the domain is downsampled from the continuous space $\mathbb{R}^n$ to the discrete space $\mathbb{Z}^n$, and the image is represented by a matrix with a given number of rows and columns. Also the possible grey values are sampled. Grey values do not longer belong to the complete continuous unit interval $[0, 1]$, but are downsampled to a finite subchain of it. In the interval-valued fuzzy mathematical morphology, the used intervals of possible grey values are now thus subsets of the finite subchain $L^I_{r,s}$ of L^I, with $L^I_{r,s} = \{[\frac{r-k}{r-1}, \frac{s-l}{s-1}] | k, l \in \mathbb{Z}$ and $1 \leq k \leq r$ and $1 \leq l \leq s\}$ for given integers r and s. We denote the class of all interval-valued fuzzy sets in $\mathbb{Z}^n$ with membership intervals in $L^I_{r,s}$ as $\mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$. Further, remark that for an interval-valued fuzzy set $A \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n), \forall x \in \mathbb{Z}^n, A_1(x) \in I_r = \{\frac{r-k}{r-1} | k \in \mathbb{Z}$ and $1 \leq k \leq r\}$ and $A_2(x) \in I_s\{\frac{s-l}{s-1} | l \in \mathbb{Z}$ and $1 \leq l \leq s\}$.

ISBN: 978-989-95079-6-8

The concepts of negators, conjunctors and implicators on the chain $L^I_{r,s}$ can be adopted from the previous subsection by replacing L^I by $L^I_{r,s}$.

The definitions of the discrete interval-valued fuzzy dilation and erosion can now be written as follows:

Definition 3. *Let $\mathcal{C}$ be a conjunctor on $L^I_{r,s}$, let $\mathcal{I}$ be an implicator on $L^I_{r,s}$, and let $A, B \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$. The discrete interval-valued fuzzy dilation $D^I_{\mathcal{C}}(A,B) \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$ and erosion $E_{\mathcal{I}}(A,B) \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$ are defined by:*

$$\begin{aligned} D^I_{\mathcal{C}}(A,B)(y) &= [\max_{x \in T_y(d_B) \cap d_A} \mathcal{C}(B(x-y), A(x))_1, \\ &\quad \max_{x \in T_y(d_B) \cap d_A} \mathcal{C}(B(x-y), A(x))_2], \\ E^I_{\mathcal{I}}(A,B)(y) &= [\min_{x \in T_y(d_B)} \mathcal{I}(B(x-y), A(x))_1, \\ &\quad \min_{x \in T_y(d_B)} \mathcal{I}(B(x-y), A(x))_2]. \end{aligned}$$

3 Decomposition of Interval-valued Fuzzy Morphological Operations

3.1 Weak $[\alpha_1, \alpha_2]$-cuts

We first introduce the different weak $[\alpha_1, \alpha_2]$-cuts of an interval-valued fuzzy set [10].

Definition 4. *Let $A \in \mathcal{IVFS}(\mathbb{R}^n)$. For $[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}$, the weak $[\alpha_1, \alpha_2]$-cut $A^{\alpha_2}_{\alpha_1}$ of A is given by:*

$$\begin{aligned} A^{\alpha_2}_{\alpha_1} &= \{x | x \in \mathbb{R}^n,\ A_1(x) \geq \alpha_1 \text{ and } A_2(x) \geq \alpha_2\} \\ &= \{x | x \in \mathbb{R}^n \text{ and } A(x) \geq_{L^I} [\alpha_1, \alpha_2]\}. \end{aligned}$$

For $\alpha_1 \in]0,1]$, the weak α_1-subcut A_{α_1} of A is given by:

$$A_{\alpha_1} = \{x | x \in \mathbb{R}^n \text{ and } A_1(x) \geq \alpha_1\}.$$

For $\alpha_2 \in]0,1]$, the weak α_2-supercut A^{α_2} of A is given by:

$$A^{\alpha_2} = \{x | x \in \mathbb{R}^n \text{ and } A_2(x) \geq \alpha_2\}.$$

The cases $[\alpha_1, \alpha_2] = 0_{L^I}$, $\alpha_1 = 0$ and $\alpha_2 = 0$ are excluded for respectively the weak $[\alpha_1, \alpha_2]$-cut, α_1-subcut and α_2-supercut. Since $\{x | x \in \mathbb{R}^n,\ A_1(x) \geq 0 \text{ and } A_2(x) \geq 0\} = \{x | x \in \mathbb{R}^n \text{ and } A_1(x) \geq 0\} = \{x | x \in \mathbb{R}^n \text{ and } A_2(x) \geq 0\} = \mathbb{R}^n$, these cases don't yield new information.

3.2 Decomposition of the Interval-valued Fuzzy Dilation

Lemma 1. *If $\mathcal{C}$ is a semi-norm on L^I, then it holds that $\mathcal{C} \leq \mathcal{C}_{\min}$, i.e.: $(\forall (x,y) \in (L^I)^2)(\mathcal{C}(x,y) \leq_{L^I} \mathcal{C}_{\min}(x,y))$.*

Proof. Let $\mathcal{C}$ be a semi-norm on L^I, then it holds for all $(x,y) \in (L^I)^2$ that on the one hand $\mathcal{C}(x,y) \leq_{L^I} \mathcal{C}(x, 1_{L^I})$ and $\mathcal{C}(x, 1_{L^I}) = x$, and on the other hand $\mathcal{C}(x,y) \leq_{L^I} \mathcal{C}(1_{L^I}, y)$ and $\mathcal{C}(1_{L^I}, y) = y$, from which it can be concluded that $\mathcal{C}(x,y) \leq_{L^I} \mathcal{C}_{\min}(x,y)$. □

3.2.1 Decomposition by weak $[\alpha_1, \alpha_2]$-cuts

There is no relationship between the weak $[\alpha_1, \alpha_2]$-cut $D^I_{\mathcal{C}}(A,B)^{\alpha_2}_{\alpha_1}$ and the binary dilation $D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1})$ that holds in general for an arbitrary semi-norm $\mathcal{C}$.

Example 1. Let $[\alpha_1, \alpha_2] = [1/4, 1]$, $A(x) = [x/2, x]$ for all $x \in [0,1[$, $A(x) = 0_{L^I}$ for all $x \in \mathbb{R} \backslash [0,1[$, $B(x) = 1_{L^I}$ for all $x \in [0,1]$ and $B(x) = 0_{L^I}$ for all $x \in \mathbb{R} \backslash [0,1]$. So $d_A =]0,1[$, $d_B = [0,1]$ and $D(d_A, d_B) =]-1, 1[$ and consequently $0 \in D(d_A, d_B)$. Let $\mathcal{C}$ be an arbitrary semi-norm.

It then holds that

$$D^I_{\mathcal{C}}(A,B)(0) = \sup_{x \in T_0(d_B) \cap d_A} \mathcal{C}(B(x), A(x)) = \sup_{x \in]0,1[} \mathcal{C}(1_{L^I}, [x/2, x]) = \sup_{x \in]0,1[} [x/2, x] = [1/2, 1],$$

which means that $0 \in D^I_{\mathcal{C}}(A,B)^1_{0.25}$.

On the other hand, however, since $A^1_{0.25} = \emptyset$ also $D(A^1_{0.25}, B^1_{0.25}) = \emptyset$ and thus $0 \notin D(A^1_{0.25}, B^1_{0.25})$. As a consequence $D^I_{\mathcal{C}}(A,B)^1_{0.25} \not\subseteq D(A^1_{0.25}, B^1_{0.25})$.

Neither does it hold in general that $D^I_{\mathcal{C}}(A,B)^{\alpha_2}_{\alpha_1} \supseteq D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1})$. Take for example $[\alpha_1, \alpha_2] = [1/4, 1/2]$, $\mathcal{C}(r,s) = [r_1 \cdot s_1, r_2 \cdot s_2]$ for all $r, s \in L^I$, $A(x) = [0.3, 0.6]$ for all $x \in [0,1]$, $A(x) = 0_{L^I}$ for all $x \in \mathbb{R} \backslash [0,1]$, $B(x) = [0.4, 0.7]$ for all $x \in [0,1]$ and $B(x) = 0_{L^I}$ for all $x \in \mathbb{R} \backslash [0,1])$.

Then on the one hand $0 \in D(A^{0.5}_{0.25}, B^{0.5}_{0.25}) = D(d_A, d_B) = [-1, 1]$.

On the other hand however:

$$D^I_{\mathcal{C}}(A,B)(0) = \sup_{x \in T_0(d_B) \cap d_A} \mathcal{C}(B(x), A(x)) = \sup_{x \in [0,1]} [0.3 \cdot 0.4, 0.6 \cdot 0.7] = [0.12, 0.42] \not\geq_{L^I} [0.25, 0.5],$$

or thus $0 \notin D^I_{\mathcal{C}}(A,B)^{0.5}_{0.25}$. As a consequence $D^I_{\mathcal{C}}(A,B)^{0.5}_{0.25} \not\supseteq D(A^{0.5}_{0.25}, B^{0.5}_{0.25})$.

♢

For the semi-norm $\mathcal{C} = \mathcal{C}_{\min}$, we have the following partial result.

Proposition 1. *Let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, then it holds for all $[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}$ that:*

$$D^I_{\mathcal{C}_{\min}}(A,B)^{\alpha_2}_{\alpha_1} \supseteq D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1}).$$

Proof. Let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, and let $[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}$. Then we have:

$$\begin{aligned} & y \in D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1}) \\ \Leftrightarrow\ & T_y(B^{\alpha_2}_{\alpha_1}) \cap A^{\alpha_2}_{\alpha_1} \neq \emptyset \\ \Leftrightarrow\ & (\exists x \in T_y(d_B) \cap d_A)(x \in T_y(B^{\alpha_2}_{\alpha_1}) \text{ and } x \in A^{\alpha_2}_{\alpha_1}) \\ \Leftrightarrow\ & (\exists x \in T_y(d_B) \cap d_A) \\ & (B(x-y) \geq_{L^I} [\alpha_1, \alpha_2] \text{ and } A(x) \geq_{L^I} [\alpha_1, \alpha_2]) \\ \Leftrightarrow\ & (\exists x \in T_y(d_B) \cap d_A)([\min(B_1(x-y), A_1(x)), \\ & \min(B_2(x-y), A_2(x))] \geq_{L^I} [\alpha_1, \alpha_2]) \\ \Leftrightarrow\ & (\exists x \in T_y(d_B) \cap d_A) \\ & (\mathcal{C}_{\min}(B(x-y), A(x)) \geq_{L^I} [\alpha_1, \alpha_2]) \\ \Rightarrow\ & \sup_{x \in T_y(d_B) \cap d_A} \mathcal{C}_{\min}(B(x-y), A(x)) \geq_{L^I} [\alpha_1, \alpha_2] \\ \Leftrightarrow\ & D^I_{\mathcal{C}_{\min}}(A,B)(y) \geq_{L^I} [\alpha_1, \alpha_2] \\ \Leftrightarrow\ & y \in D^I_{\mathcal{C}_{\min}}(A,B)^{\alpha_2}_{\alpha_1}. \end{aligned}$$

□

The reverse inclusion $D^I_{\mathcal{C}_{\min}}(A,B)^{\alpha_2}_{\alpha_1} \subseteq D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1})$ does not hold in general as already illustrated in Example 1.

Remark that the above decomposition property for weak $[\alpha_1, \alpha_2]$-cuts remains valid in the discrete framework.

ISBN: 978-989-95079-6-8

3.2.2 Decomposition by weak sub- and supercuts

There is also no relationship between the weak sub- and supercut $D^I_{\mathcal{C}}(A,B)_{\alpha_1}$ and $D^I_{\mathcal{C}}(A,B)^{\alpha_2}$ and the binary dilations $D(A_{\alpha_1},B_{\alpha_1})$ and $D(A^{\alpha_2},B^{\alpha_2})$ that holds in general for an arbitrary semi-norm $\mathcal{C}$. To illustrate this, we can use Example 1 again, where the weak $[\alpha_1,\alpha_2]$-cuts and the weak super- and subcuts of A and B coincide and the results thus remain valid for weak sub- and supercuts.

For the semi-norm $\mathcal{C}=\mathcal{C}_{\min}$, we have the following partial result.

Proposition 2. *Let $A,B\in\mathcal{IVFS}(\mathbb{R}^n)$, then it holds that:*

(i) $(\forall\alpha_1\in]0,1])(D^I_{\mathcal{C}_{\min}}(A,B)_{\alpha_1}\supseteq D(A_{\alpha_1},B_{\alpha_1}))$,
(ii) $(\forall\alpha_2\in]0,1])(D^I_{\mathcal{C}_{\min}}(A,B)^{\alpha_2}\supseteq D(A^{\alpha_2},B^{\alpha_2}))$.

Proof. Analogous to the proof of Proposition 1. □

The reverse inclusion does not hold, as already illustrated in Example 1, where replacing the weak $[\alpha_1,\alpha_2]$-cuts of A and B by the coinciding weak sub- and supercuts doesn't affect the results.

In the discrete framework, not only does Proposition 2 still remain valid, but the result now also holds for arbitrary semi-norms. Further, for $\mathcal{C}_{\min}$ also the reverse inclusion now holds.

Proposition 3. *Let $A,B\in\mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$, then it holds that:*

(i) $(\forall\alpha_1\in]0,1]\cap I_r)(D^I_{\mathcal{C}_{\min}}(A,B)_{\alpha_1}=D(A_{\alpha_1},B_{\alpha_1}))$,
(ii) $(\forall\alpha_2\in]0,1]\cap I_s)(D^I_{\mathcal{C}_{\min}}(A,B)^{\alpha_2}=D(A^{\alpha_2},B^{\alpha_2}))$.

Proof. Analogous to the proof of Proposition 2, where now in the discrete case also

$$(\exists x\in T_y(d_B)\cap d_A)(\mathcal{C}_{\min}(B(x-y),A(x))_1\geq\alpha_1)$$
$$\Updownarrow$$
$$\sup_{x\in T_y(d_B)\cap d_A}\mathcal{C}_{\min}(B(x-y),A(x))_1\geq\alpha_1.$$

□

Proposition 4. *Let $A,B\in\mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$, then it holds that:*

(i) $(\forall\alpha_1\in]0,1]\cap I_r)(D^I_{\mathcal{C}}(A,B)_{\alpha_1}\subseteq D(A_{\alpha_1},B_{\alpha_1}))$,
(ii) $(\forall\alpha_2\in]0,1]\cap I_s)(D^I_{\mathcal{C}}(A,B)^{\alpha_2}\subseteq D(A^{\alpha_2},B^{\alpha_2}))$.

Proof. Analogous to the proof of Proposition 3, but for an arbitrary semi-norm $\mathcal{C}$, so that

$$D^I_{\mathcal{C}_{\min}}(A,B)(y)_1\geq\alpha_1$$
$$\Uparrow$$
$$D^I_{\mathcal{C}}(A,B)(y)_1\geq\alpha_1$$
$$\Updownarrow$$
$$y\in D^I_{\mathcal{C}}(A,B)_{\alpha_1}.$$

□

3.3 Decomposition of the Interval-valued Fuzzy Erosion

Based on its induced negator $\mathcal{N}_{\mathcal{I}}$, a border implicator $\mathcal{I}$ can be classified as an upper or a lower border implicator as follows. A border implicator on L^I is called an upper border implicator (respectively lower border implicator) if $\mathcal{N}_{\mathcal{I}}\geq\mathcal{N}_s$ (respectively $\mathcal{N}_{\mathcal{I}}\leq\mathcal{N}_s$).

Lemma 2. *If $\mathcal{I}$ is an upper border implicator on L^I, then it holds that $\mathcal{I}\geq\mathcal{I}_{\max,\mathcal{N}_s}$, i.e.: $(\forall(x,y)\in(L^I)^2)(\mathcal{I}(x,y)\geq_{L^I}\mathcal{I}_{\max,\mathcal{N}_s}(x,y))$.*

Proof. Let $\mathcal{I}$ be an upper border implicator on L^I. For all $(x,y)\in(L^I)^2$ it holds that on the one hand $\mathcal{I}(x,y)\geq_{L^I}\mathcal{I}(1_{L^I},y)$ and $\mathcal{I}(1_{L^I},y)=y$, and on the other hand $\mathcal{I}(x,y)\geq_{L^I}\mathcal{I}(x,0_{L^I})$ and $\mathcal{I}(x,0_{L^I})\geq_{L^I}\mathcal{N}_s(x)$, from which it follows that $\mathcal{I}(x,y)\geq_{L^I}\mathcal{I}_{\max,\mathcal{N}_s}(x,y)$. □

3.3.1 Decomposition by weak $[\alpha_1,\alpha_2]$-cuts

As illustrated below, there is no relationship between the weak $[\alpha_1,\alpha_2]$-cut $E^I_{\mathcal{I}}(A,B)^{\alpha_2}_{\alpha_1}$ and the binary erosion $E(A^{\alpha_2}_{\alpha_1},B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}})$ that holds in general for an arbitrary upper border implicator $\mathcal{I}$.

Example 2. Let $[\alpha_1,\alpha_2]=[0.4,0.6]$, $A(x)=[0.3,0.5]$ for all $x\in[0,1]$, $B(x)=[0.5,0.7]$ for all $x\in[0,1]$ and $A(x)=B(x)=0_{L^I}$ for all $x\in\mathbb{R}\backslash[0,1]$. Let $\mathcal{I}$ be the upper border implicator given by $\mathcal{I}_L(x,y)=[\min(1,1-x_2+y_1),\min(1,1-x_1+y_2)]$, $\forall(x,y)\in(L^I)^2$, which is a generalisation of the Łukasiewicz implicator on $[0,1]$.

On the one hand, we have that

$$E^I_{\mathcal{I}_L}(A,B)(0)=\inf_{x\in T_0(d_B)}\mathcal{I}_L(B(x),A(x))=$$
$$\inf_{x\in[0,1]}\mathcal{I}_L([0.5,0.7],[0.3,0.5])=[0.6,1],$$

or thus $0\in E^I_{\mathcal{I}_L}(A,B)^{0.6}_{0.4}$.

On the other hand, $E(A^{0.6}_{0.4},B^{\overline{0.6}}_{\overline{0.4}})=E(\emptyset,[0,1])=\emptyset$ and thus $0\notin E(A^{0.6}_{0.4},B^{\overline{0.6}}_{\overline{0.4}})$, which implies that $E^I_{\mathcal{I}_L}(A,B)^{0.6}_{0.4}\not\subseteq E(A^{0.6}_{0.4},B^{\overline{0.6}}_{\overline{0.4}})$.

Neither does it hold in general that $E^I_{\mathcal{I}}(A,B)^{\alpha_2}_{\alpha_1}\not\supseteq E(A^{\alpha_2}_{\alpha_1},B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}})$. Consider for example the upper border implicator $\mathcal{I}=\mathcal{I}_{\max,\mathcal{N}_s}$, $[\alpha_1,\alpha_2]=[0.3,0.4]$, $A(x)=[0.4,0.5]$ for all $x\in[0,0.5]$ and $A(x)=[0.2,0.3]$ for all $x\in]0.5,1]$, $B(x)=[0.7,0.8]$ for all $x\in[0,0.5]$ and $B(x)=[0.4,0.8]$ for all $x\in]0.5,1]$ and $A(x)=B(x)=0_{L^I}$ for all $x\in\mathbb{R}\backslash[0,1]$.

The binary erosion $E(A^{\alpha_2}_{\alpha_1},B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}})$ is then equal to the set $E([0,0.5],[0,0.5])=\{0\}$.

Further, it also holds that:

$$E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)(0)=\inf_{x\in T_0(d_B)}\mathcal{I}_{\max,\mathcal{N}_s}(B(x),A(x))=$$
$$\inf(\inf_{x\in[0,0.5]}\mathcal{I}_{\max,\mathcal{N}_s}(B(x),A(x)),$$
$$\inf_{x\in]0.5,1]}\mathcal{I}_{\max,\mathcal{N}_s}(B(x),A(x)))=\inf([0.4,0.5],[0.2,0.6])=$$
$$[0.2,0.5]\not\geq_{L^I}[\alpha_1,\alpha_2].$$

We conclude that $E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2}_{\alpha_1}\not\supseteq E(A^{\alpha_2}_{\alpha_1},B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}})$. ♢

For the upper border implicator $\mathcal{I}=\mathcal{I}_{\max,\mathcal{N}_s}$, we have the following partial result.

Proposition 5. *Let $A,B\in\mathcal{IVFS}(\mathbb{R}^n)$, then it holds for all $[\alpha_1,\alpha_2]\in L^I\backslash\{0_{L^I}\}$ that:*

$$E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2}_{\alpha_1}\subseteq E(A^{\alpha_2}_{\alpha_1},B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}}).$$

ISBN: 978-989-95079-6-8

Proof. Let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, and let $[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}$. It holds that:

$$
\begin{aligned}
& y \in E(A^{\alpha_2}_{\alpha_1}, B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}}) \\
\Leftrightarrow \quad & T_y(B^{\overline{1-\alpha_1}}_{\overline{1-\alpha_2}}) \subseteq A^{\alpha_2}_{\alpha_1} \\
\Leftrightarrow \quad & (\forall x \in T_y(d_B)) \\
& ((B_1(x-y) > 1-\alpha_2 \text{ and } B_2(x-y) > 1-\alpha_1) \\
& \Rightarrow (A_1(x) \geq \alpha_1 \text{ and } A_2(x) \geq \alpha_2)) \\
\Leftrightarrow \quad & (\forall x \in T_y(d_B))((B_1(x-y) \leq 1-\alpha_2 \text{ or } B_2(x-y) \\
& \leq 1-\alpha_1) \text{ or } (A_1(x) \geq \alpha_1 \text{ and } A_2(x) \geq \alpha_2)) \\
\Leftrightarrow \quad & (\forall x \in T_y(d_B)) \\
& ((1-B_1(x-y) \geq \alpha_2 \text{ or } 1-B_2(x-y) \geq \alpha_1) \text{ or} \\
& (A_1(x) \geq \alpha_1 \text{ and } A_2(x) \geq \alpha_2)) \\
\Leftarrow \quad & (\forall x \in T_y(d_B))(\max(1-B_2(x-y), A_1(x)) \geq \alpha_1 \\
& \text{and } \max(1-B_1(x-y), A_2(x)) \geq \alpha_2) \\
\Leftrightarrow \quad & (\forall x \in T_y(d_B)) \\
& (\mathcal{I}_{\max,\mathcal{N}_s}(B(x-y), A(x)) \geq_{L^I} [\alpha_1, \alpha_2]) \\
\Leftrightarrow \quad & \inf_{x \in T_y(d_B)} \mathcal{I}_{\max,\mathcal{N}_s}(B(x-y), A(x)) \geq_{L^I} [\alpha_1, \alpha_2] \\
\Leftrightarrow \quad & E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A, B)(y) \geq_{L^I} [\alpha_1, \alpha_2] \\
\Leftrightarrow \quad & y \in E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A, B)^{\alpha_2}_{\alpha_1}
\end{aligned}
$$

□

As already illustrated in Example 2, the reverse inclusion does not hold.

Remark that Proposition 5 remains valid in the discrete framework.

3.3.2 Decomposition by weak sub- and supercuts

Proposition 6. *Let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, then it holds that*

(i) $(\forall \alpha_1 \in]0,1])(E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A, B)_{\alpha_1} = E(A_{\alpha_1}, B^{\overline{1-\alpha_1}}))$,

(ii) $(\forall \alpha_2 \in]0,1])(E^I_{\mathcal{I}_{\max,\mathcal{N}_s}}(A, B)^{\alpha_2} = E(A^{\alpha_2}, B_{\overline{1-\alpha_2}}))$.

Proof. Analogous to the proof of Proposition 5. Only now it also holds that

$$
\begin{gathered}
(\forall x \in T_y(d_B))(1-B_2(x-y) \geq \alpha_1 \text{ or } A_1(x) \geq \alpha_1) \\
\Updownarrow \\
(\forall x \in T_y(d_B))(\max(1-B_2(x-y), A_1(x)) \geq \alpha_1)
\end{gathered}
$$

Analogously for the decomposition by weak supercuts. □

Proposition 7. *Let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$ and let $\mathcal{I}$ be an upper border implicator on L^I, then it holds that*

(i) $(\forall \alpha_1 \in]0,1])(E^I_{\mathcal{I}}(A, B)_{\alpha_1} \supseteq E(A_{\alpha_1}, B_{\overline{1-\alpha_1}}))$,

(ii) $(\forall \alpha_2 \in]0,1])(E^I_{\mathcal{I}}(A, B)^{\alpha_2} \supseteq E(A^{\alpha_2}, B_{\overline{1-\alpha_2}}))$.

Proof. The proof is completely analogous to the one from proposition 6. We only have that due to lemma 2

$$
\begin{gathered}
\inf_{x \in T_y(d_B)} \mathcal{I}_{\max,\mathcal{N}_s}(B_2(x-y), A_1(x))_1 \geq \alpha_1 \\
\Downarrow \\
\inf_{x \in T_y(d_B)} \mathcal{I}(B_2(x-y), A_1(x))_1 \geq \alpha_1
\end{gathered}
$$

only holds in one direction for an arbitrary upper border implicator $\mathcal{I}$ on L^I. Analogously for the decomposition by weak supercuts. □

The reverse inclusion does not hold in general, as illustrated in the first part of Example 2, where replacing the weak $[\alpha_1, \alpha_2]$-cuts by the coinciding weak sub- or supercuts does not affect the results.

Further, the two above properties remain valid in the discrete framework.

3.4 Decomposition of the Interval-valued Fuzzy Closing and Opening

In what follows, we will need the following lemma:

Lemma 3. *Let $A \in \mathcal{IVFS}(\mathbb{R}^n)$ and let $[\alpha_1, \alpha_2] \in L^I$, then it holds that:*

(i) $\alpha_2 \in]0, 0.5] \quad \Rightarrow \quad A^{\alpha_2} \supseteq A^{\overline{\alpha_2}} \supseteq A_{\overline{1-\alpha_2}}$,

(ii) $\alpha_1 \in]0.5, 1] \quad \Rightarrow \quad A_{\alpha_1} \subseteq A^{\overline{1-\alpha_1}}$.

Proof. As an example we prove (i). Suppose that $x \in A_{\overline{1-\alpha_2}}$ and $\alpha_2 \in]0, 0.5]$. The latter implies that $1-\alpha_2 \geq \alpha_2$. Further, since $x \in A_{\overline{1-\alpha_2}}$, we also have that $A_1(x) > 1-\alpha_2$. If we combine the above with the fact that $A_2(x) \geq A_1(x)$, then we find that $A_2(x) > \alpha_2$ and consequently also $A_2(x) \geq \alpha_2$.

(ii) follows in an analogous way. □

3.4.1 Decomposition by weak sub- and supercuts

Proposition 8. *Let $\mathcal{I}$ be an upper border implicator on L^I and let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, then it holds for all $\alpha_1 \in]0, 1]$ that*

(i) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)_{\alpha_1} \supseteq E(D(A_{\alpha_1}, B_{\alpha_1}), -B^{\overline{1-\alpha_1}})$,

(ii) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)_{\alpha_1} \supseteq D(E(A_{\alpha_1}, B^{\overline{1-\alpha_1}}), -B_{\alpha_1})$,

and for all $\alpha_2 \in]0, 1]$ that

(iii) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq E(D(A^{\alpha_2}, B^{\alpha_2}), -B_{\overline{1-\alpha_2}})$,

(iv) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq D(E(A^{\alpha_2}, B_{\overline{1-\alpha_2}}), -B^{\alpha_2})$.

Proof. As an example we prove (i). Let $\mathcal{I}$ be an upper border implicator on L^I, let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$ and let $\alpha_1, \alpha_2 \in]0, 1]$. From respectively Proposition 6, Proposition 2, and because the binary erosion is increasing in its first argument, we have that:

$$
\begin{aligned}
C^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)_{\alpha_1} &= E^I_{\mathcal{I}}(D^I_{\mathcal{C}_{\min}}(A, B), -B)_{\alpha_1} \\
&\supseteq E(D^I_{\mathcal{C}_{\min}}(A, B)_{\alpha_1}, -B^{\overline{1-\alpha_1}}) \\
&\supseteq E(D(A_{\alpha_1}, B_{\alpha_1}), -B^{\overline{1-\alpha_1}}).
\end{aligned}
$$

(ii), (iii) and (iv) follow in an analogous way. □

Under the restriction of $\alpha_2 \in]0, 0.5]$, the above result leads to the following relationships between the weak α_2-supercut of the interval-valued fuzzy closing and opening and the binary counterparts.

Proposition 9. *Let $\mathcal{I}$ be an upper border implicator on L^I and let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$, then it holds for $\alpha_2 \in]0, 0.5]$ that*

(i) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq C(A^{\alpha_2}, B^{\alpha_2})$,

(ii) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq C(A^{\alpha_2}, B_{\overline{1-\alpha_2}})$,

and

(iii) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq O(A^{\alpha_2}, B^{\alpha_2})$,

(iv) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A, B)^{\alpha_2} \supseteq O(A^{\alpha_2}, B_{\overline{1-\alpha_2}})$.

Proof. As an example, we prove (i). Let $\mathcal{I}$ be an upper border implicator on L^I, let $A, B \in \mathcal{IVFS}(\mathbb{R}^n)$ and let $\alpha_2 \in]0, 0.5]$.

ISBN: 978-989-95079-6-8

From proposition 8, lemma 3 and the fact that the binary erosion is decreasing in its second argument, it follows that:

$$\begin{aligned} C^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} &\supseteq E(D(A^{\alpha_2},B^{\alpha_2}),-B_{\overline{1-\alpha_2}}) \\ &\supseteq E(D(A^{\alpha_2},B^{\alpha_2}),-B^{\alpha_2}) \\ &= C(A^{\alpha_2},B^{\alpha_2}). \end{aligned}$$

(ii), (iii) and (iv) follow in an analogous way. □

The above results for weak sub- and supercuts remain valid in the discrete framework. Because of the new decomposition properties of the interval-valued fuzzy dilation in the discrete framework compared to the continuous framework (Proposition 3 and 4), also a new property holds for the discrete interval-valued fuzzy closing and opening.

Proposition 10. *Let $\mathcal{C}$ be a semi-norm on $L^I_{r,s}$ and $\mathcal{I}$ an upper border implicator on $L^I_{r,s}$ and let $A, B \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$, then it holds for all $\alpha_1 \in]0,1] \cap I_r$ that*

(i) $C^I_{\mathcal{C}_{\min},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} = E(D(A_{\alpha_1},B_{\alpha_1}),-B^{\overline{1-\alpha_1}})$,
(ii) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)_{\alpha_1} \supseteq E(D(A_{\alpha_1},B_{\alpha_1}),-B^{\overline{1-\alpha_1}})$,
(iii) $C^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq E(D(A_{\alpha_1},B_{\alpha_1}),-B^{\overline{1-\alpha_1}})$,
(iv) $O^I_{\mathcal{C}_{\min},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} = D(E(A_{\alpha_1},B^{\overline{1-\alpha_1}}),-B_{\alpha_1})$,
(v) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)_{\alpha_1} \supseteq D(E(A_{\alpha_1},B^{\overline{1-\alpha_1}}),-B_{\alpha_1})$,
(vi) $O^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq D(E(A_{\alpha_1},B^{\overline{1-\alpha_1}}),-B_{\alpha_1})$,

and for all $\alpha_2 \in]0,1] \cap I_s$ that

(vii) $C^I_{\mathcal{C}_{\min},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2} = E(D(A^{\alpha_2},B^{\alpha_2}),-B_{\overline{1-\alpha_2}})$,
(viii) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq E(D(A^{\alpha_2},B^{\alpha_2}),-B_{\overline{1-\alpha_2}})$,
(ix) $C^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2} \subseteq E(D(A^{\alpha_2},B^{\alpha_2}),-B_{\overline{1-\alpha_2}})$,
(x) $O^I_{\mathcal{C}_{\min},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2} = D(E(A^{\alpha_2},B_{\overline{1-\alpha_2}}),-B^{\alpha_2})$,
(xi) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq D(E(A^{\alpha_2},B_{\overline{1-\alpha_2}}),-B^{\alpha_2})$,
(xii) $O^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)^{\alpha_2} \subseteq D(E(A^{\alpha_2},B_{\overline{1-\alpha_2}}),-B^{\alpha_2})$.

Proof. Follows in an analogous way as in the proof of Proposition 8. □

Under the restriction of $\alpha_1 \in]0.5,1] \cap I_r$, the above result leads to the following relationships between the weak α_1-subcut of the interval-valued fuzzy closing and opening and the binary counterparts.

Proposition 11. *Let $\mathcal{C}$ be a semi-norm on $L^I_{r,s}$ and let $A, B \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$, then it holds for all $\alpha_1 \in]0.5,1] \cap I_r$ that*

(i) $C^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq C(A_{\alpha_1},B_{\alpha_1})$,
(ii) $C^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq C(A_{\alpha_1},B^{\overline{1-\alpha_1}})$,
(iii) $O^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq O(A_{\alpha_1},B_{\alpha_1})$,
(iv) $O^I_{\mathcal{C},\mathcal{I}_{\max,\mathcal{N}_s}}(A,B)_{\alpha_1} \subseteq O(A_{\alpha_1},B^{\overline{1-\alpha_1}})$,

Proof. Follows in an analogous way as in the proof of Proposition 9. □

Under the restriction of $\alpha_2 \in]0,0.5] \cap I_s$, we come to the same results as in the continuous case (Proposition 9).

Proposition 12. *Let $\mathcal{I}$ be an upper border implicator on $L^I_{r,s}$ and let $A, B \in \mathcal{IVFS}_{r,s}(\mathbb{Z}^n)$, then it holds for all $\alpha_2 \in]0,0.5] \cap I_s$ that*

(i) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq C(A^{\alpha_2},B^{\alpha_2})$,
(ii) $C^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq C(A^{\alpha_2},B_{\overline{1-\alpha_2}})$,
(iii) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq O(A^{\alpha_2},B^{\alpha_2})$,
(iv) $O^I_{\mathcal{C}_{\min},\mathcal{I}}(A,B)^{\alpha_2} \supseteq O(A^{\alpha_2},B_{\overline{1-\alpha_2}})$.

Proof. Follows in an analogous way as in the proof of Proposition 9. □

4 Conclusion

In this paper we have introduced the basic interval-valued fuzzy morphological operations and we have studied their decomposition by weak $[\alpha_1,\alpha_2]$-cuts both in the general continuous case and the practical discrete case. We found out that a few properties that do not hold in the continuous case do hold in the practical discrete case. Finally, we would like to mention that similar, but not completely analogous results hold for the decomposition by strict $[\alpha_1,\alpha_2]$-cuts. Those results will be published elsewhere.

Acknowledgment

This work was financially supported by the GOA-project B/04138/01 IV 1 of Ghent University.

References

[1] J. Serra. *Image Analysis and Mathematical Morphology*, Academic Press, London, 1982.

[2] R.M. Haralick, R.S. Sternberg and X. Zhuang. Image analysis using mathematical morphology, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):532-550, 1987.

[3] B. De Baets. Fuzzy morphology: a logical approach, *Uncertainty Analysis in Engineering and Sciences: Fuzzy Logic, Statistics, and Neural Network Approach* (Ayyub, B.M. & Gupta, M.M., editors), Kluwer Academic Publishers, Boston, pp. 53-67, 1997.

[4] P. Sussner and M.E. Valle: Classification of fuzzy mathematical morphologies based on concepts of inclusion measure and duality, *Journal of Mathematical Imaging and Vision* , 32(2):139-159, October 2008.

[5] I. Bloch. Mathematical Morphology on Bipolar Fuzzy Sets, *Proceedings of ISMM 2007 (International Symposium on Mathematical Morphology)*, pp. 3-4, 2007.

[6] M. Nachtegael, P. Sussner, T. Mélange and E.E. Kerre. Modelling numerical and spatial uncertainty in grayscale image capture using fuzzy set theory, *Proceedings of NASTEC 2008*, 2008.

[7] L. Zadeh. Fuzzy Sets, *Information and Control*, 8:338-353, 1965.

[8] R. Sambuc. *Fonctions Φ-floues. Application à l'aide au diagnostic en pathologie thyroidienne*, Ph.D. thesis, Univ. Marseille, France, 1975.

[9] G. Deschrijver and E.E. Kerre. On the relationship between some extensions of fuzzy set theory, *Fuzzy Sets and Systems*, 133:227-235, 2003.

[10] G. Deschrijver and E.E. Kerre. On the cuts of intuitionistic fuzzy compositions, *Kuwait Journal of Science & Engineering*, 32(1):17-38, 2005.

[11] M. Nachtegael and E.E. Kerre. Connections between binary, gray-scale and fuzzy mathematical morphology, *Fuzzy Sets and Systems*, 124:73-86, 2001.

An alternative operationalization of fuzzy consideration set. Application to tourism

Luisa L. Lazzari Patricia I. Mouliá Mariano Eriz

Faculty of Economics Science, University of Buenos Aires
Buenos Aires, Argentina
Email: ilazzari@econ.uba.ar, pimoulia@hotmail.com, erizmariano@hotmail.com

Abstract—*Consideration sets provide a standard tool for the exploration of consumer choice. It is appropriate to analyze it as a fuzzy set because not all the brands of a certain category of products or services that belong to it are considered of the same importance.*

In this work we propose an alternative operationalization of the aggregate fuzzy consideration set. It is based on a new approach for linguistic framework and linguistic aggregation operators which compute with words directly, defined by Xu. The model of consideration presented assigns to each brand or service the degree of membership to the evoked set through weighed aggregate interest.

The aggregate fuzzy consideration set of tourist destinations of a given of individuals segment is obtained as illustrative example of the proposed approach.

Keywords— aggregation operators, computing with words, fuzzy consideration set, linguistic model, tourist destinations.

1 Introduction

Several variables take place in the study of the consumer's behaviour, which is a complex activity. To face it in a more suitable manner, flexible models are appropriate; these will consider the uncertainty implied by its treatment, particularly those using the theory of the fuzzy sets. This theory provides an important background for the representation of preferences and the uncertainty contained in decision making. This approach can be applied to the analysis of the consideration set or evoked set.

Though the individual has a certain number of possibilities to choose among, it is possible that he does not take into account all the available alternatives. These alternatives considered during the decision making form what is known as the *consideration set* or *evoked set*.

The consideration set is a subset of all alternatives available for the consumer in a category of products or services. It is made up by his preferred brands, with which he is acquainted, those he remembers and those he positively evaluates for purchasing and consumption [1, 2].

Howard [3] was the first to introduce the term evoked set in the literature of marketing to talk collectively about the set of brands that the consumer considers in his process of purchases decision. Several authors elaborated this concept later [4, 5, 6].

Consideration sets have been the subject of much research since 1986. The level of interest in this study can be attributed to several factors [7]. Among others, a consideration set represents a task-simplifying heuristic that consumers use to cope with complex choice problems [8]; models that ignore consideration set information may have biased parameter estimates [9] and when a brand enters the consideration set of a consumer, the chances that the consumer will choose that brand increase even if it is not the most preferred. Its exclusion prevents the selection of the brand even if it is likely to be preferred [10].

Lilien *et al.* [9] states that studying the composition of the consideration set is important for two reasons: the first one because it will help the company that does not belong to the set to know its own lacks and the characteristics of those companies or brands that belong to it; the second because it is an important part of the general study of the consumption processes. Some research have shown that to include the consideration set in the stages of decision model allows to obtain better prognoses and a more suitable handling of diagnoses [11].

Fotheringham [12] suggests that the consideration set is not binary, but it might be treated as a fuzzy set, maybe because the consideration is not a discreet process of the consumers or because the investigators are not capable of measuring it. He expresses that not all the brands taken into account in the phases of evaluation and election are given the same importance.

In previous works [13, 14] we have developed a linguistic model to obtain the membership function of the aggregate fuzzy consideration set, for a category of products or services, by means of the usual linguistic framework and linguistic aggregation operators based on the Extension Principle [15, 16, 17, 18].

In this paper an alternative linguistic model is proposed, to obtain the aggregate fuzzy consideration set which is based on a new approach for linguistic framework and on the linguistic weighted arithmetic averaging (LWAA) operator, which enables to compute with words directly, defined by Xu [18, 19, 20, 21].

Then, this methodology is applied to obtain the aggregate fuzzy consideration set of tourist destinations (TDs) of the young, with the data obtained in a survey carried out on students of the University of Buenos Aires.

The cardinality of the evoked set and its support is obtained, in order to find its magnitudes. Besides, a measure of fuzziness is calculated to know the entropy that affects the fuzzy consideration set.

The paper is organized as follows. In Section 2 we introduce a linguistic framework presented by Xu [18, 20] to define some linguistic aggregation operators and the LWAA operator definition. In Section 3 we give an alternative operationalization of the aggregate fuzzy consideration set.

ISBN: 978-989-95079-6-8

In Section 4 the aggregate fuzzy consideration set of TDs is obtained and in Section 5 some concluding remarks are pointed out.

2 Fuzzy linguistic approach

The fuzzy linguistic approach is an appropriate technique to deal with qualitative aspects of problems [22].

Following Xu [18, 19, 20, 21] we consider a finite and totally ordered label set $S=\{s_\alpha/\alpha=-t,...,-1,0,1,...,t\}$, which cardinality value is odd, and t is a positive integer. Each term s_α represents a possible value for a linguistic variable [18, 19, 20, 21] and it must have the following characteristics:

i) $s_\alpha > s_\beta$ iff $\alpha > \beta$.

ii) There is the negation operator: $\text{neg}(s_\alpha)=s_{-\alpha}$; mainly $\text{neg}(s_0)=s_0$.

The mid linguistic label s_0 represents an assessment of "indifference" and the rest of labels are defined around it symmetrically.

To preserve all the given information, Xu [18, 19] extended the discrete linguistic label set S to a continuous linguistic label set $\overline{S}=\{s_\alpha/\alpha\in[-q,q]\}$, where q $(q>t)$ is a sufficiently large positive integer.

If $s_\alpha \in S$, then s_α is called an original linguistic label, otherwise, s_α is called a virtual linguistic label. Generally, the decision maker uses the original linguistic terms to evaluate attributes and alternatives, and the virtual linguistic labels can only appear in calculations [20, 21].

Considering any two linguistic terms $s_\alpha, s_\beta \in \overline{S}$, and $\lambda\in[0,1]$, Xu [19] introduces two operational laws of linguistic variables as follows:

- $s_\alpha \oplus s_\beta = s_\beta \oplus s_\alpha = s_{\alpha+\beta}$. (1)
- $\lambda . s_\alpha = s_{\lambda\alpha}$. (2)

Based on (1) and (2) Xu [18, 19, 21] developed various linguistic aggregation operators, which compute with words directly.

In this section we present the linguistic weighted arithmetic averaging (LWAA) operator due to the fact that it will be the linguistic operator used in our aggregate fuzzy consideration sets model.

Definition [18]: Let $LWAA: \overline{S}^n \to \overline{S}$. If

$$LWAA_w(s_{\alpha_1}, s_{\alpha_2}, ..., s_{\alpha_n}) = w_1 s_{\alpha_1} \oplus w_2 s_{\alpha_2} \oplus ... \oplus w_n s_{\alpha_n} = s_{\overline{\alpha}} \quad (3)$$

Where $\overline{\alpha}=\sum_{j=1}^{n} w_j.\alpha_j$, $w=(w_1, w_2, ..., w_n)$ is the weight vector of the linguistic label s_{α_j}, $j=1,...,n$, $w_j\in[0,1]$ and $\sum_{j=1}^{n} w_j = 1$, then *LWAA* is called the linguistic weighted arithmetic averaging operator.

The fundamental aspect of the *LWAA* operator is that it computes the linguistic labels taking into account the importance of the information [18].

3 The fuzzy consideration set

3.1 Preliminaries

When forming the evoked set of a certain category of products or services, the consumers do not show the same interest in their purchase, due to the variety of attributes these products or services have and to other factors that in most cases are subjective, thus, it is appropriate to analyze it as a fuzzy set [7, 9, 12, 13, 14].

When developing its model of consideration, Fotheringham [12] states that the evoked set can be considered fuzzy because uncertainty about its composition exists. Although the author affirms that the consideration sets are fuzzy, his formulation does not derive from the theory of the fuzzy sets. Besides, it does not explain which the fuzziness source is or how it obtains the membership function.

Wu and Rangaswamy [7] developed a model of consideration and election, which they call fuzzy. His work suggests that a fuzzy approach is useful to represent the fuzziness of the consideration set, their model is derived from probability theory and they do not explain how to obtain the membership function.

The model of consideration presented in this paper assigns to each brand or service the degree of membership to the evoked set through weighed aggregate interest, calculated by means of the *LWAA* operator. To obtain the value of the membership function of a brand or service, the aggregate degree of interest is weighed in agreement with the quantity of individuals who chose the above mentioned brand or service.

3.2 Building aggregate fuzzy consideration sets

Surveys to potential clients interested in the purchase of a type of goods or services are made to obtain the aggregate fuzzy consideration set of a category of products or services of a given of individuals segment. In the questionnaire we ask about:

- The names of the well-known brands.
- The names of the brands that they would consider buying.
- The degree of interest with which these are considered for their purchase.

For the design of questionnaires for the surveys a set of linguistic labels is chosen which fulfils the characteristics given in 2., so that the consulted individuals express their degree of interest in the purchase of the goods or the services.

First of all, the referential set of the evoked set (E) is obtained with the brands that appear in the survey. It is named set of well-known brands or awareness set.

To obtain the value of the membership function of each element the aggregate fuzzy consideration set, the following steps are involved:

Step 1. *The aggregate degree of interest.* If m is the quantity of proposed brands in the surveys and n is the cardinality of the label set S, the aggregate degree of interest

ISBN: 978-989-95079-6-8

(q_i) of each brand or service (c_i) considered is obtained by means of the application of (1).

$$q_i = LWAA_{c_i}\left(s_{\alpha_1}, s_{\alpha_2}, \ldots, s_{\alpha_n}\right) = s_{\overline{\alpha}_i} \quad i = 1,\ldots, m \qquad (4)$$

Where $\overline{\alpha}_i = \sum_{j=1}^{n} w_j \alpha_j$ and the weighting vector of the linguistic level s_{α_j} is $w = (w_1, w_2, \ldots, w_n)$. If r_{ij} is the quantity of individuals that selected the brand or service c_i with the level s_{α_j} and v_i is the quantity of all individuals that selected this brand, then:

$$w_j = \frac{r_{ij}}{v_i}, j = 1,\ldots, n;\ w_j \in [0,1],\ \sum_{j=1}^{n} w_j = 1 \qquad (5)$$

Step 2. *The value of the membership function, without weighing:* We define it on the basis of the aggregate degree of interest, displacing the scale of order $\frac{n-1}{2}$, later standardized.

$$\kappa_i = \frac{2\overline{\alpha}_i + n - 1}{2(n-1)}, \quad i = 1,\ldots, m,\ 0 \le \kappa_i \le 1 \qquad (6)$$

Step 3. *The weighting vector*: $p = (p_1, \ldots, p_m)$ is obtained to reflect the number of times each brand is chosen in the survey. Specifically, weights p_i are calculated as follows:

$$p_i = \frac{F_i}{\max F_i};\ 0 < p_i \le 1; \forall i = 1,\ldots, m \qquad (7)$$

where F_i is the frequency of occurrence for each brand.

Step 4. *The aggregate fuzzy consideration set*: The value of the membership function for each brand or service (c_i) is calculated as follows:

$$\mu_C(c_i) = p_i . \kappa_i \qquad (8)$$

which represent the intentions of purchase of the individuals selecting the brand or service c_i.

Finally, the aggregate fuzzy consideration set is:

$$C: E \to [0,1] / \mu_C(c_i) = p_i . \kappa_i \qquad (9)$$

4 Application

In this section, the aggregate fuzzy consideration set of TDs of a given of individuals segment is obtained as an illustrative example of the proposed approach.

The consideration set of TDs is a subset of the set of all consumers' available alternatives and it is constituted by those TDs with which the consumer is acquainted, which he remembers or which he considers would be interesting to visit, and which were evaluated positively to spend his holidays [23, 24].

As not all the TDs taken into account in the phases of evaluation and election are considered to have the same importance, it is possible to state that they belong to the evoked set with a different degree [13, 14]. For this reason we propose to treat the consideration set of TDs as a fuzzy set on the basis of the set of the places known by the consumers.

Since mid 1900s, tourism has turned into one of the economic activities of major growth in the world. The tourist activity mobilises public and private funds, affecting the economic and social conditions of the population, and it can be an essential tool to achieve a sustainable development as it is a generator of employment and currency. This justifies the systematical study of the tourist activity in a region or country.

Young tourism is an increasing world trend. The characteristic of this segment is the performing of trips, different from those of the traditional market. In Argentina, the young tourism represents almost 20 % of the tourist entire market.

Then, the methodology presented in Section 3.2 is applied to obtain the aggregate fuzzy consideration set of TDs of young students, with the data obtained in a survey.

A random closed survey to 940 students of the University of Buenos Aires (UBA) between 19 and 29 years of age was done in September 2008. We obtained information about the tourist places that students would like to visit in the next summer vacation (January, February, 2009) and the degree of interest to visit particular locations.

A linguistic term set of five values was used, and every student expressed his degree of interest in each TD considered by one of this label of the set S.

$S = \{ s_{-2} = \text{very low (VL)},\ s_{-1} = \text{low (L)},\ s_0 = \text{middle (M)},\ s_1 = \text{high (H)},\ s_2 = \text{very high (VH)} \}$

The data obtained in the survey were processed. The TDs of the support of aggregate fuzzy consideration set are:

$c_1 = \text{Bariloche}$
$c_2 = \text{Cataratas del Iguazú}$
$c_3 = \text{Córdoba}$
$c_4 = \text{El Calafate}$
$c_5 = \text{Jujuy}$
$c_6 = \text{Mar del Plata}$
$c_7 = \text{Mendoza}$
$c_8 = \text{Pinamar}$
$c_9 = \text{Puerto Madryn}$
$c_{10} = \text{Salta}$
$c_{11} = \text{San Martín de los Andes}$
$c_{12} = \text{Ushuaia}$

As an example of the calculation of the value of the membership function, we will obtain it for $c_1 = \text{Bariloche}$.

Step 1: We assume $s_{\alpha_1} = s_{-2}$; $s_{\alpha_2} = s_{-1}$; $s_{\alpha_3} = s_0$; $s_{\alpha_4} = s_1$; $s_{\alpha_5} = s_2$.

By (4) and (5), we have $w = (\frac{2}{273}, \frac{9}{273}, \frac{59}{273}, \frac{103}{273}, \frac{100}{273})$, thus,

$$\overline{\alpha}_1 = \frac{2}{273} \times (-2) \oplus \frac{9}{273} \times (-1) \oplus \frac{59}{273} \times 0 \oplus \frac{103}{273} \times 1 \oplus \frac{100}{273} \times 2$$

therefore, $\overline{\alpha}_1 = 1.06$ and $q_1 = s_{1.06}$.

Step 2: By means of (6) $\kappa_1 = \frac{2(1.06) + 4}{8} = 0.77$.

Step 3: Using (7) $p_1 = \frac{273}{306} = 0.89$.

Step 4: Finally, $\mu(c_1) = 0.69$.

ISBN: 978-989-95079-6-8

The following results are contained in Table 1.

In column 1: The TDs of the support of aggregate fuzzy consideration set.

In column 2: The frequency of each TD (F_i).

In column 3: The aggregate degree of interest of each TD (q_i).

In column 4: The value of membership function of each TD, without weighing (κ_i).

In column 5: The weight of each TD (p_i).

In column 6: The value of membership function of each TD ($\mu(c_i)$).

Table 1: Results of 1st survey - UBA 2008

TDs	F_i	q_i	κ_i	p_i	$\mu(c_i)$
c_1	273	$s_{1.06}$	0.77	0.89	0.69
c_2	306	$s_{0.97}$	0.74	1	0.74
c_3	231	$s_{0.78}$	0.70	0.75	0.53
c_4	152	$s_{1.43}$	0.86	0.50	0.43
c_5	161	$s_{1.07}$	0.77	0.53	0.41
c_6	182	$s_{0.45}$	0.61	0.59	0.36
c_7	247	$s_{1.02}$	0.75	0.81	0.61
c_8	100	$s_{0.67}$	0.67	0.33	0.22
c_9	91	$s_{1.09}$	0.77	0.30	0.23
c_{10}	269	$s_{1.12}$	0.78	0.88	0.69
c_{11}	91	$s_{1.24}$	0.81	0.30	0.24
c_{12}	162	$s_{1.29}$	0.82	0.53	0.43

It is possible to express the linguistic aggregate degree of interest (LADI) of each TD using a label of set S. Approximating the subindex of the virtual label to an integer value by means of the usual round operation, i.e., round $(\overline{\alpha}_i)$. The results can be observed in Table 2. This information is useful to the decision maker.

Table 2: Linguistic aggregate degree of interest

TD	q_i	$s_\alpha \in S$	LADI
c_1	$\alpha_{1.06}$	α_1	High
c_2	$\alpha_{0.97}$	α_1	High
c_3	$\alpha_{0.78}$	α_1	High
c_4	$\alpha_{1.43}$	α_1	High
c_5	$\alpha_{1.07}$	α_1	High
c_6	$\alpha_{0.45}$	α_0	Middle
c_7	$\alpha_{1.02}$	α_1	High
c_8	$\alpha_{0.67}$	α_1	High
c_9	$\alpha_{1.09}$	α_1	High
c_{10}	$\alpha_{1.12}$	α_1	High
c_{11}	$\alpha_{1.24}$	α_1	High
c_{12}	$\alpha_{1.29}$	α_1	High

The destinations that belong to the aggregate fuzzy consideration set have obtained high levels of aggregate interest, except for Mar del Plata which obtained middle level. The predominant destinations are "Cataratas del Iguazú", "Salta" and "Bariloche", with the biggest values of membership (Table 1).

To have an idea of the magnitude of the obtained sets, their cardinality has been calculated.

Awareness set: $|E| = 62$.

Support of aggregate fuzzy consideration set: $|S(C)| = 12$.

Aggregate fuzzy consideration set: $\|C\| = 5.22$.

To determine the degree of uncertainty contained in the information provided by the aggregate fuzzy consideration set, we obtained the Yager's measure of fuzziness [25].

$$e_Y(C) = 1 - \frac{\left[\sum_{i=1}^{n} \left|\mu_C(x_i) - \mu_{\overline{C}}(x_i)\right|^p\right]^{\frac{1}{p}}}{|\mathrm{supp}(C)|}, \; p \in N - \{0\} \qquad (10)$$

$0 \le e_Y(C) \le 1$.

If $p = 1$, then $e_Y(C) = 0.68$.

We can states that the fuzziness of the obtained set is moderate. This value indicates a measure of the information quality about the uncertainty of the group's consideration.

The utility of the entropy has its roots in the possibility of comparison as well as in the study of fuzziness' evolution existing in a system.

Empirical available evidence of similar studies shows that the fuzziness of the aggregated fuzzy consideration set is moderate in the first instance of its evaluation and diminishes its value as you approach the moment to make the decision about your holidays.

Reference [14] shows that if $p = 1$, then (10) is equal to Kaufmann's linear index of fuzziness.

5 Conclusions

This paper offers a new approach to build aggregate fuzzy consideration set.

We verified that it enables to find more easily results similar to those obtained with the method that uses linguistic operators based on the Extension Principle.

The operator used in the model presented compute with words directly and weighs the linguistic argument itself, instead of the ordered position. It allows for a continuous representation of the linguistic information on its domain, and thus, they can represent any counting of information obtained in an aggregation process without any loss of information.

In this work we have presented preliminary results of a research on young tourism begun in 2008. In this project we will analyze young people behaviour in choosing a TD in Argentina. National University population will be the aim of this study, since attending students have heterogeneous characteristics.

With the information obtained through longitudinal exploratory studies, we will determine and analyze the aggregate fuzzy consideration sets of tourist destinations in Argentina at local and regional level, in winter and summer. We will study the dynamics and the fuzziness of the aggregate sets and the attributes the young have in mind when choosing possible destinations to spend their vacations. Obtained results will be compared considering tourist season and geographic areas.

ISBN: 978-989-95079-6-8

Finally, after each vacation period, the same group of students will be asked about which TD was visited, to check the forecast of the model.

This study will allow us to obtain relevant information, to make a diagnosis of the situation of the young tourism in the considered regions and to plan actions tending to increase this activity.

At this stage of the research we obtained and analyzed the aggregated awareness sets and consideration set of tourist destinations of Argentina for the young people in 2009 summer vacation.

The results allow to infer that the hypothesis referred to consideration set's fuzziness is valid as the consulted young people stated that not all the TDs considered to spend their next vacation have aroused the same interest of visit.

We observed that the destinations considered are traditional and well-known. They are distributed all around the whole country and present a variety of sceneries and attractions, such as the sea, the mountains, lakes, glaciers, rivers, waterfalls and cities with great cultural activity.

The favorites TDs are Bariloche, Mar del Plata, Iguazú Waterfalls and Córdoba, because they present the largest degree of consideration.

Acknowledgement

This work has been supported by the Argentinian Agencia Nacional de Promoción Científica, Tecnológica y de Innovación (PICT 383) and by the University of Buenos Aires (UBACyT E018).

References

[1] D.L. Loudon and A.J. Della Bitta, *Consumer Behavior: Concepts and Applications*. New Jersey: McGraw-Hill, Inc., 2001.

[2] L.G. Schiffman and L.L. Kanuk, *Consumer Behavior*. New Jersey: Prentice Hall Inc., 2000.

[3] J.A. Howard, *Marketing Management: Analysis and Planning*. Illinois: Richard D. Irwin Co., 1963.

[4] J.E. Brisoux and M. La Roche, Evoked set formation and composition: An empirical investigation under a routinized response behavior situation. *Advances in Consumer Research*, 8: 357-361, 1980.

[5] J.A. Howard and J.N. Sheth, *The Theory of Buyer Behavior*. New York: Wiley, 1969.

[6] C.L. Narayana and R.J. Markin, Consumer behavior and product performance: An alternative conceptualization. *Journal of Marketing*, 39: 1-6, 1975.

[7] J. Wu and A. Rangaswamy, A fuzzy set model search and consideration with an application to an online market. *Marketing Science*, 22: 411-434, 2003.

[8] B.J. Bronnenberg and W.R. Vanhonacker, Limited choice sets, local price response and implied measures of price competition. *Journal of Marketing Research*, 33: 163-173, 1996.

[9] G.L. Lilien *et. al.*, *Marketing Models*. New Jersey: Prentice Hall International, Inc., 1992.

[10] R.L. Andrews and T.C. Srinivasan, Studying consideration effects in empirical choice models using scanner panel data *Journal of Marketing Research*, 32: 30-41, 1995.

[11] D.H. Gensch and J.A. Svestka, A maximum likelihood hierarchical disaggregate model for predicting choice of individuals. *Journal of Mathematical Psychology*, 28: 160-178, 1984.

[12] A.S. Fotheringham, Consumer store choice and choice set definition. *Marketing Science*, 7(3): 299-310, 1988.

[13] M. Eriz *et al.* Consideration set in tourist destinations as a fuzzy set. *Proceedings of International Conference on Modelling & Simulation* (ICMS'04), Valladolid, 2004.

[14] L.L. Lazzari, *El comportamiento del consumidor desde una perspectiva fuzzy*. Tesis doctoral. Universidad de Valladolid, Valladolid, 2006.

[15] M. Delgado *et al.* Aggregation of linguistic information based on a symbolic approach. In: L.A. Zadeh and J. Kacprzyk (edits.). *Computing with Words in Information/Intelligent Systems* 1. Heidelberg: Physica-Verlag, 1999.

[16] F. Herrera and E. Herrera-Viedma, Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, 115: 67-82, 2000.

[17] E. Herrera-Viedma, *et al.*, Evaluation the information quality of web sites: A methodology based on fuzzy computing with words. *Journal of American Society for Information Science and Technology*, 57(4): 538-549, 2006.

[18] Z. Xu, Linguistic aggregation operators: An overview, in: H. Bustince *et al.* (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*. Berlin: Springer-Verlag, 2008.

[19] Z. Xu, EOWA and EOWG operators for aggregating linguistic labels based on linguistic preference relations. *International Journal of Uncertainty Fuzziness and Knowledge-Based- Systems*, 12: 791-810, 2004.

[20] Z. Xu, Deviation measures of linguistic preference relations in group decision making. *Omega*, 33: 249-254, 2005.

[21] Z. Xu, A note on linguistic hybrid arithmetic averaging operator in multiple attribute group decision making with linguistic information. *Group Decision and Negotiation*, 15: 593-604, 2006.

[22] L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning. Part I, *Information Sciences*, 8 : 199-249. Part II, *Information Sciences*, 8: 301-357. Part III, *Information Sciences*, 9: 43-80, 1975.

[23] J.L. Crompton and P.K. Ankomah, Choice set propositions in destination decisions. *Annals of Tourism Research*, 20: 461-476, 1993.

[24] H.R. Seddighi and A.L. Theocharous, A model of tourism destination choice: a theoretical and empirical analysis. *Tourism Management*, 23: 475-487, 2002.

[25] R.R. Yager, On the measure of fuzziness and negation. Part I: Membership in the unit interval. *International Journal of General Systems*, 5: 221-229, 1979.

ISBN: 978-989-95079-6-8

Ant Colony Optimization for Distributed Routing Path Optimization in Optical Burst-Switched Networks

João Pedro[1, 2], João Pires[2], and Joao Paulo Carvalho[3]

1 Nokia Siemens Networks Portugal S.A., R. Irmãos Siemens 1, 2720-093 Amadora, Portugal
2 Instituto de Telecomunicações, Instituto Superior Técnico, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
3 INESC-id, Instituto Superior Técnico, R. Alves Redol 9, 1000-029 Lisboa, Portugal
Email: joao.pedro@nsn.com, jpires@lx.it.pt, joao.carvalho@inesc-id.pt

Abstract—*This work describes a distributed framework for routing path optimization in Optical Burst-Switched (OBS) networks that loosely mimics the foraging behaviour of ants observed in nature, which in the past has originated the Ant Colony Optimization (ACO) metaheuristic. The framework consists of additional data structures stored at the nodes and of special control packets that traverse the network, estimate the goodness of their paths and update accordingly the data structures of the nodes. The performance of the framework, which has been implemented on an event-driven OBS network simulator, is evaluated on several network topologies and compared with that obtained with centralized routing path optimization. The simulation results show that the distributed routing path selection framework significantly improves the performance of OBS networks by reducing data loss and that it attains a performance comparable to that of the centralized algorithm. The results also suggest that the framework is robust to changes in its parameters.*

Keywords—Ant colony optimization, optical burst switching, routing path optimization, swarm intelligence.

1 Introduction

The foraging behaviour of some ant species has inspired the ant colony optimization metaheuristic [1], which has become one of the most successful swarm intelligence methods for problem solving [2]. Ant colonies are distributed systems that take advantage of their structured social organization to accomplish complex tasks while relying only on very simple individuals. This property, where a colony is able to perform tasks whose complexity far exceeds that of an individual, has made ant colonies a source of inspiration to develop artificial distributed systems and algorithms [3].

It has been shown that ACO algorithms can efficiently solve routing problems in some types of telecommunication networks, namely in circuit-switched [4] and packet-switched [5] networks. Recently, ACO algorithms were also devised for the routing and wavelength assignment problem that arises in Wavelength Division Multiplexing (WDM) optical networks [6], [7].

Circuit-switched WDM optical networks are worldwide used as backbone and metropolitan networks, due to their high capacity and low transmission losses. However, their coarse granularity limits their effectiveness to support bursty Internet Protocol (IP) packet traffic. Research efforts have been directed to finer-grained optical networking paradigms, such as optical burst switching [8]. At the edge nodes of OBS networks, multiple IP packets are assembled into bursts of data, which are then transmitted towards their egress nodes. A one-way resource reservation mechanism without acknowledgement of successful reservation in the burst path is used to allocate wavelength channels to the upcoming data bursts. Hence, there is the possibility that two or more bursts contend for the same wavelength channel at the core nodes, which is further augmented by the lack of efficient optical buffers (limiting the capability to delay data bursts). Thus, contention is a major issue in OBS networks and minimizing it is of paramount importance. One approach to mitigate contention consists of optimizing the routing paths of data bursts as to balance the traffic load across the network links.

This paper describes a distributed framework for routing path optimization in OBS networks inspired on ACO and assesses its performance using network simulation. The simulation results show that this framework can efficiently reduce contention, thus improving network performance, without requiring a fine tuning of its main parameters.

The remainder of the paper is organized as follows. Section 2 describes the ant foraging behaviour observed in nature and its utilization for network routing optimization, whereas section 3 introduces the OBS network architecture and the relevance of optimizing the paths used to route data bursts. The proposed distributed routing path optimization framework for OBS networks, embedding collective intelligence and problem solving capabilities observed in colonies of ants is described in section 4. The performance of the proposed framework is evaluated through network simulation and compared to that obtained with shortest path routing and a centralized routing path optimization algorithm in section 5. Finally, section 6 presents the conclusions.

2 Ant colony optimization and its applications

2.1 Ants in Nature

Individual ants are very simple insects whose behaviour appears to have a large random component. Despite their individual simplicity, when acting collectively they are able to execute complicated tasks with significant consistency and reliability. For instance, some ant species can collectively cooperate in regulating the nest temperature, raiding particular areas for food, building and protecting the nest, carrying large items, preferentially exploiting the richest available food source and, more related to this work, finding the shortest routes from the nest to a food source [4]. These

ISBN: 978-989-95079-6-8

relatively complex tasks seem to emerge from interactions between large numbers of ants and their environment. In some cases, individual ants coordinate their activities using a principle known as *stigmergy* [1], whose main characteristics are: (i) indirect and non-symbolic form of communication mediated by the environment, that is, ants exchange information by modifying their environment; (ii) stigmergic information is local, which means that it can only be accessed by ants that visit the immediate vicinity of the place where it was released.

The communication between individual ants is based on the use of specific chemicals, called pheromones. Particularly important for the social life of some ant species is the *trail pheromone*, which is a specific type of pheromone they use for marking paths on the ground. While walking between food sources and the nest, ants deposit pheromones on the ground, forming a pheromone trail. Subsequent ants sense the pheromone and tend to choose, probabilistically, paths with a stronger pheromone concentration, further reinforcing it. After some time, ants converge to use a single shortest path. A compendium of seminal works on the pheromone trail-laying and trail-following behaviour of some ant species can be found in [3].

2.2 Ant Colony Optimization for Network Routing

The basic idea underlying ant colony-based algorithms and systems is to use a form of *artificial stigmergy* to coordinate the activities of *artificial agents*. The majority of ACO applications have been for solving $\mathcal{NP}$-hard problems, for which the best known algorithms that guarantee an optimal solution have an exponential time worst case complexity. In some of these problems, ACO algorithms have been shown to determine high-quality, although not guaranteed to be optimal, solutions in relatively short time.

The routing problems in telecommunication networks are particularly interesting for ACO, because the intrinsically distributed nature of ACO can fit well with them. However, these problems pose additional challenges, because the routing algorithm should also be able to react rapidly to changing network and traffic conditions and to maintain exploration capabilities so that it can effectively evaluate alternative paths which, in view of the problem dynamics, can become the best ones at some point in time [1].

The different ACO algorithms for network routing have in common the fact that pheromone tables are stored at the network nodes and artificial agents, typically in the form of control packets, are employed to collect node/link state information and update the pheromone tables. ACO was first applied to network routing in circuit-switched networks, like the Plain Old Telephone Service (POTS), in [3]. The main purpose of Ant-Based Control (ABC) is to relieve congestion on the network by preferably routing calls (circuits) via parts of the network with larger spare capacity, consequently reducing call blocking probability. Later, ACO was applied to routing in packet-switched networks, such as IP networks, in [5]. These networks use store-and-forward nodes, which contain buffers to store the incoming packets and delay them until there is free capacity on the preferred output link. With AntNet [5], an artificial ant is launched at regular intervals of time from a network node towards another node to discover a low-cost path between the nodes and to investigate the load status of the network. The selection of the next node by an ant is made by assigning to the neighbouring nodes a probability of being selected that depends not only on the correspondent pheromone value but also on the occupancy of the queue associated to the link between the current node and the neighbouring node. Hence, the chances of selecting a neighbouring node increase with its pheromone value but decrease with the queue waiting time associated to it.

ACO has also found application on routing in optical WDM networks, both in Optical Circuit-Switched (OCS) [6], [7] and Optical Packet-Switched (OPS) networks [9]. AntNet was adapted in [9] for load balancing in OPS networks. The nodes of these networks use Fibre Delay Line (FDL) buffers, which have a very limited contention resolution capability, when compared to electronic buffers. The proposal modifies the solution construction phase of the original AntNet implementation [5] using a suitable expression for computing the buffer occupancy, since FDL buffers behave differently from electronic buffers.

Despite the recent research activity aiming to use ACO for routing in optical WDM networks, both in OCS and OPS, to the best of the author's knowledge ACO has not yet been adapted to the case of OBS networks.

3 Optical burst-switched networks

In current optical transport networks, circuits are established between nodes during long periods of time (weeks, months) using wavelength channels with a capacity of several Gbit/s each. However, the dominant IP packet traffic carried over these networks significantly fluctuates over relatively small time scales. Hence, current circuit-switched optical networks are inefficient to support this type of traffic due to both their coarse granularity and slow circuit setup time. OPS networks would provide the best adaptation between the IP traffic and the optical network resources used, but they cannot be implemented with state-of-the-art technology, namely due to the already referred lack of efficient optical buffers. OBS networks have attracted considerable interest because their switching granularity and optical technology complexity are in between those of OCS and OPS [10].

3.1 Network Architecture

At the ingress edge nodes of OBS networks, IP packets are assembled into large chunks of data, called bursts. Each data burst is transmitted using a wavelength channel and must be switched at the core nodes until it reaches the egress edge node, where it is disassembled and its packets delivered to the local IP network. Resource reservation for each burst is made using a control message, known as Burst Header Packet (BHP), sent ahead of the data burst on a separate wavelength channel. Each node along the routing path of the burst processes the BHP and configures its switching matrix to direct the upcoming data burst to an available wavelength channel on the appropriate output fibre link. The interested reader can find more information on OBS networks in [10].

ISBN: 978-989-95079-6-8

3.2 Routing Path Optimization

In OBS networks, multiple data bursts, overlapping in time, may contend at a node for the same wavelengths of an output fibre link. Unresolved contention leads to dropping some of the data bursts, degrading network performance. Therefore, it is important to minimize the probability that the number of overlapping data bursts directed to a fibre link exceeds the number of wavelength channels. This can be achieved by optimizing the routing paths in order to balance the traffic load offered to the network links, reducing the traffic load on the most congested links at the expense of increasing it on the less congested ones.

Routing path optimization can either be centralized, when all the paths are computed at a central point of the network using global information of the offered traffic load [11], or distributed [12]. Although centralized strategies can reduce contention more efficiently, by exploiting global knowledge of the network state, distributed strategies are often preferred because they are more scalable and robust and faster to adapt the routing paths to changes in the traffic pattern.

4 Distributed Ant-inspired Burst Routing

The framework for distributed routing path optimization in OBS networks proposed in this work is based on principles of ACO. Although it inherits some concepts from previous ACO algorithms for network routing, such as AntNet and ABC, the proposed Distributed Ant-inspired Burst Routing (DABR) framework was specifically designed for bufferless OBS networks and incorporates functions that account for the particular characteristics of these networks.

Common to all applications of ACO for network routing, the DABR framework uses artificial ants, in the form of control packets, and data structures at the network nodes to support stigmergic communication between the artificial ants. The key features of DABR are:

- Forward ants carry a fictitious data burst reservation and, at each node, the instantaneous congestion of the output fibre links is evaluated based on the number of wavelengths available to transmit the fictitious burst.
- Backward ants update the pheromone tables using the congestion information collected by the forward ants as input to a loss performance model that estimates the burst blocking probability on the path traversed by the forward ant.
- In order to mitigate path stagnation, three mechanisms are used: the routing of forward ants employs both a minimum pheromone concentration per adjacent node (noise) and pheromone-heuristic control, whereas the backward ants use privileged pheromone laying [13].
- The nodes keep separate routing and pheromone tables because an additional path integrity mechanism is used to completely avoid circular path formation.

Consider an OBS network modelled as a directed graph $G = (V, E)$, where $N = |V|$ is the number of nodes and $L = |E|$ is the number of unidirectional fibre links. Assume each fibre link supports a set of W wavelengths, $\{\lambda_1, \lambda_2, \ldots, \lambda_{W-1}, \lambda_W\}$. The following sub-sections describe the DABR framework.

4.1 Control Packets and Local Data Structures

In DABR, artificial ants are transmitted on the control channel as modified BHPs. Two types of artificial ants are used: Explorer Ant Packet (EAP) and Referee Ant Packet (RAP). The former ant is used to collect information on the resource availability along a given routing path and to update the pheromone tables, loosely mimicking the behaviour of real ants, whereas the latter one is used to avoid circular path formation. Both EAP and RAP have two modes of operation: forward (F), from the ingress to the egress node, and backward (B), in the reverse path. Two data structures are stored at each node: one pheromone table and one routing table. The pheromone table is used for routing the F-EAPs and it is updated in their return trips as B-EAPs. The routing table can only be updated by RAPs.

Unlike previous ACO routing algorithms, routing of artificial ants and data bursts in DABR is based not only on the egress node but also on the ingress node, providing an additional degree of freedom for load balancing at the expense of increasing the size of the tables. Let A_i denote the set of adjacent nodes of $i \in V$ (nodes connected to i by a fibre link) and τ_{ij}^{sd} denote the pheromone value associated to using node $j \in A_i$ to route at node i a burst generated at s and directed to d. The sum of the pheromone values related with the same ingress and egress node pair must equal one,

$$\sum_{j \in A_i} \tau_{ij}^{sd} = 1, \quad i, s, d \in V, s \neq d . \tag{1}$$

Let h_{ij} denote the number of links of the shortest path between i and j. The pheromone tables are initialized as to be slightly biased towards the shortest paths,

$$\tau_{ij}^{sd} = \begin{cases} \dfrac{1/(1+h_{jd})}{\sum_{k \in A_i} 1/(1+h_{kd})}, & j \in A_i, i \neq d, s \neq d \\ 0, \text{ otherwise} \end{cases} . \tag{2}$$

4.2 Routing of Forward Explorer Ants

Each time node s assembles a data burst directed to d, it generates with probability p_{ant}, where $0 < p_{\text{ant}} \leq 1$, an F-EAP that is routed from s to d to collect information on the resource availability of a particular path. The F-EAP carries a fictitious burst reservation request whose duration equals that of the burst that triggered the generation of the artificial ant and a list V_{EAP} with the nodes it has already visited.

At each node $i \in V$ visited by an F-EAP, the next node of its routing path is selected among the adjacent nodes not visited yet. In case $A_i \subseteq V_{\text{EAP}}$, the F-EAP is dropped to avoid looping. Otherwise, the adjacent node j is selected with a probability p_{ij}^{sd} that is the normalized sum of the pheromone value and a heuristic value η_{ij} that accounts for the resource availability on the fibre link between nodes i and j [5],

$$p_{ij}^{sd} = \begin{cases} \dfrac{\tau_{ij}^{sd}}{(1+\alpha) \sum_{\substack{k \in A_i \\ k \notin V_{\text{EAP}}}} \tau_{ik}^{sd}} + \alpha \dfrac{\eta_{ij}}{(1+\alpha) \sum_{\substack{k \in A_i \\ k \notin V_{\text{EAP}}}} \eta_{ik}}, & j \in A_i, j \notin V_{\text{EAP}} \\ 0, \quad \text{otherwise} \end{cases} . \tag{3}$$

ISBN: 978-989-95079-6-8

The pheromone value is the result of pheromone laying by previous artificial ants and represents the long-term goodness of selecting the fibre link from *i* to *j* to route the F-EAP, whereas the heuristic value corresponds to the instantaneous goodness of selecting this fibre link. In order to compute η_{ij}, the wavelength availability of the fibre link *ij* is assessed by determining if the fictitious data burst could be scheduled on each of the *W* wavelengths of the fibre link. Let $\hat{W}_{ij}$ denote the number of available wavelengths to schedule the data burst on this fibre link. The heuristic value η_{ij} is given by

$$\eta_{ij} = \hat{W}_{ij} / W . \quad (4)$$

The link state weight α determines the importance of the instantaneous resource availability of the fibre link over its long-term pheromone concentration.

The node selected *j** is added to the F-EAP, as well as the number of available wavelengths $\hat{W}_{ij^*}$. When it reaches the egress node, the artificial ant is switched to B-EAP.

4.3 Pheromone Table Update by Backward Explorer Ants

The B-EAP follows the reverse path of the F-EAP, updating the pheromone tables in each node. In an OBS network, the average burst blocking probability on a fibre link can be estimated with the Erlang-B formula, assuming an average offered traffic load of *x* and a Poisson burst arrival process,

$$\mathrm{Erl}(x, W) = \left(x^W / W! \right) \cdot \left(\sum_{k=0}^{W} x^k / k! \right)^{-1} . \quad (5)$$

The wavelength availability $\hat{W}_{ij}$, collected by the F-EAP, is used to estimate the traffic load offered to the fibre link between *i* and *j* as $x = W - \hat{W}_{ij}$. Under the link independence assumption [14], the average burst blocking probability in the routing path between nodes *s* and *d* can be estimated by

$$B^{sd} = 1 - \prod_{k=1}^{|V_{\mathrm{EAP}}|-1} (1 - \hat{B}_{v_k v_{k+1}}) , \quad (6)$$

where $\{v_1, v_2, ..., v_k, ..., v_{|V_{\mathrm{EAP}}|}\}$ denotes the ordered set of the nodes visited by the F-EAP and $\hat{B}_{v_k v_{k+1}}$ the average burst blocking probability in the fibre link between v_k and v_{k+1}. The goodness of a routing path is quantified in DABR as

$$q^{sd} = 1 / B^{sd} . \quad (7)$$

Each node *i* maintains the goodness of the best solution q^{sd}_{best} estimated by the previous *Q* artificial ants between *s* and *d* and the pheromone reinforcement value *r* is given by

$$r = \begin{cases} q^{sd} / q^{sd}_{\mathrm{best}}, & q^{sd} < q^{sd}_{\mathrm{best}} \\ 1, & \text{otherwise} \end{cases} . \quad (8)$$

In order to enhance the rewarding of good solutions (high values of *r*), while saturating the rewards for low values of *r*, the following squash function *s*(*x*) is used [5],

$$s(x) = \left(1 + \exp\left(\frac{1}{r \mid V_{\mathrm{EAP}} \mid} \right) \right)^{-1} , \quad (9)$$

$$r \leftarrow \tau_{\max} \cdot s(r) / s(1) , \quad (10)$$

where $\tau_{\max}$ is the maximum pheromone value that a single artificial ant can deposit. This parameter is used to regulate the impact of a single artificial ant on the pheromone tables.

The pheromone update has to ensure a minimum amount of pheromone $\tau_{\min}$ uniformly divided by the adjacent nodes,

$$\tau^{sd}_{ik} \geq \tau_{\min} / | A_i | , \quad k \in A_i , \quad (11)$$

and it also has to guarantee that equation (1) holds. Hence, the following expressions have been devised for updating the pheromone value associated to the adjacent node *j* the F-EAP has used and that associated to the remaining adjacent nodes,

$$\tau^{sd}_{ij} \leftarrow \tau^{sd}_{ij} + r \cdot \left(1 - \tau^{sd}_{ij} - \tau_{\min} \cdot (| A_i | - 1) / | A_i | \right) , \quad (12)$$

$$\tau^{sd}_{ik} \leftarrow \tau^{sd}_{ik} - r \cdot \left(\tau^{sd}_{ik} - \tau_{\min} / | A_i | \right) , \quad k \in A_i , k \neq j . \quad (13)$$

In case the update has modified the adjacent node that has the highest pheromone value, designated as the dominant adjacent node, a specific flag of the B-EAP is activated.

4.4 Routing Table Update and Circular Path Avoidance

In the initial implementation of DABR, data bursts were routed through the path obtained from the adjacent nodes with the largest pheromone concentration. However, it was found that circular paths were occasionally formed, which resulted in data bursts getting trapped on a loop. Although these paths are only temporary, the resulting communication disruption is not tolerable in a high capacity optical network. Hence, it is imperative to include in DABR a mechanism to avoid the formation of circular paths. The novel mechanism devised for this purpose separates the pheromone and routing tables, guaranteeing that the update of the routing paths never allows the formation of loops without introducing additional restrictions to the update of the pheromone tables.

Let v^{sd}_i denote the adjacent node of node *i* used for routing data bursts from node *s* to node *d*. The update of the routing path between *s* and *d* is triggered by the arrival at node *s* of a B-EAP with the flag that signals a new dominant adjacent node activated. As a result, node *s* generates a F-RAP directed to node *d*. This control packet includes a list to record the nodes it visits. At each node *i* the F-RAP visits, the dominant adjacent node *j** is determined, that is,

$$j^* = \arg\max_{j \in A_i} \left\{ \tau^{sd}_{ij} \right\} , \quad (14)$$

and compared with the nodes previously visited. In case node *j** has already been visited, a loop is detected and the F-RAP is discarded, keeping the routing path unchanged. Otherwise, the F-RAP adds the dominant adjacent node *j** to the list of nodes visited and moves to it. In case *j** = *d*, the new routing path is non-circular. Thus, the F-RAP is switched to B-RAP, follows the inverse path and, at each node *i*, the adjacent node used to route data bursts, v^{sd}_i, is updated to be the same as the node visited by the F-RAP immediately after *i*.

The routing of data bursts in DABR is determined by the routing tables of the nodes. In particular, when a BHP, preceding the correspondent data burst, arrives at node *i*, the node determines the next node of the data burst from v^{sd}_i.

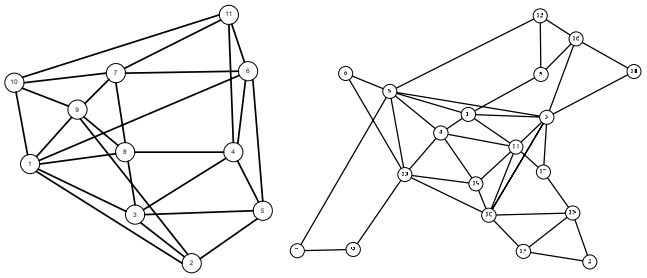
ISBN: 978-989-95079-6-8

5 Simulation Results

In OBS networks, the most relevant performance metric is the average burst blocking probability. The loss performance of OBS networks is evaluated using the network simulator of [15] modified to support DABR. All the simulations use the JET resource reservation mechanism [8], each fibre link has W=32 wavelength channels, each with a capacity μ=10 Gb/s, the time to process one BHP or artificial packet is 100 μs and the switch fabric configuration time is 160 μs. The burst size and interarrival time are exponentially distributed and the average burst size is 10 MB. The average offered traffic load normalized to the network capacity is given by

$$\Gamma = \frac{\sum_{s,d \in V} \gamma^{sd} \cdot h_{sd}}{L \cdot W \cdot \mu}, \qquad (15)$$

where γ^{sd} is the average offered traffic load between nodes s and d. The performance evaluation is made with the network topologies of Fig. 1, assuming a uniform traffic pattern for EON and a non-uniform traffic pattern for COST 239. The average burst blocking probability and a 95% confidence interval on this value are obtained from 10 independent simulations with 5×10^6 data bursts each.

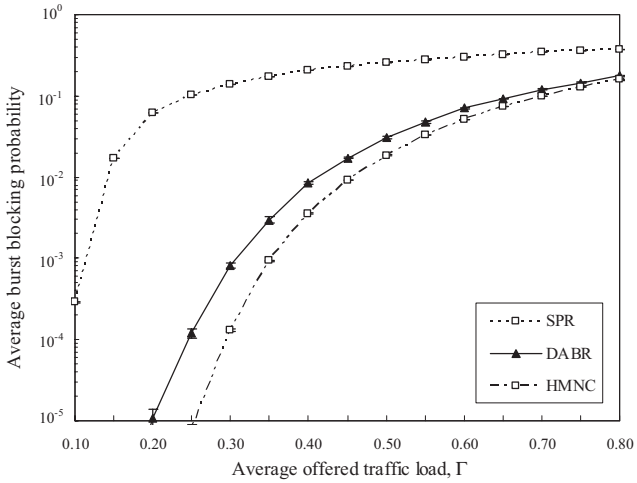


(a) COST 239 (b) EON

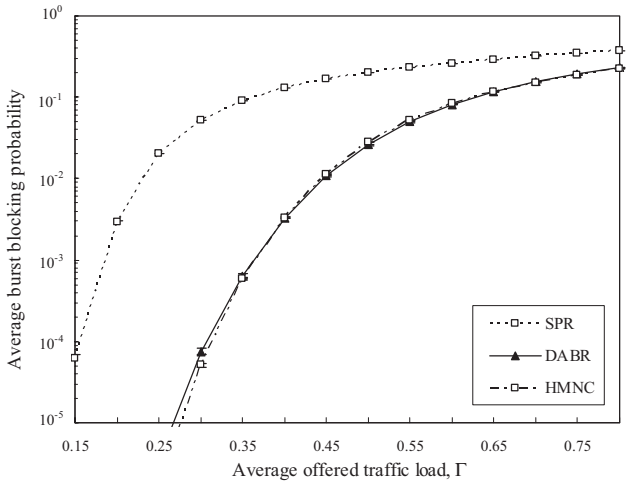
Figure 1: Network topologies.

Figure 2 plots the average burst blocking probability as a function of the average offered traffic load for the COST 239 and EON networks, respectively. The parameters used for DABR are p_{ant}=0.05, α=0.25, τ_{min}=0.2, τ_{max}=0.1, and Q=50, aiming a compromise between performance and adaptability to changing conditions. For comparison purposes the plots also include the performance obtained with Shortest Path Routing (SPR) and the centralized HMNC algorithm of [11].

The simulation results show that the performance obtained with DABR lies between that of SPR and HMNC, although always closer to the latter. This is an expected outcome. On one hand, the simpler SPR does not optimize the traffic load distribution over the fiber links, which can significantly increase contention in some of the links. On the other hand, the HMNC algorithm is more effective because it exploits the global knowledge of accurate traffic statistics. However, in more dynamic traffic scenarios, where traffic statistics are less reliable, HMNC may become less effective than the proposed DABR. The simulation results also show that larger confidence intervals are obtained with DABR than that with SPR and HMNC. This is a consequence of the online modification of the routing paths observed when using DABR, while with SPR and HMNC the same set of routing paths is used during the entire simulation.



(a) 11-node COST 239 network



(b) 19-node EON network

Figure 2: Network performance of DABR, SPR and HMNC.

The performance impact of varying some of the parameters of DABR is assessed in the COST 239 network. In all cases, the parameter values given above are used with the exception of the parameter being varied. The average burst blocking probability as a function of the minimum pheromone value, maximum pheromone increment per EAP and size of sliding window used to record the best EAP solution are plotted in Figures 3, 4 and 5, respectively.

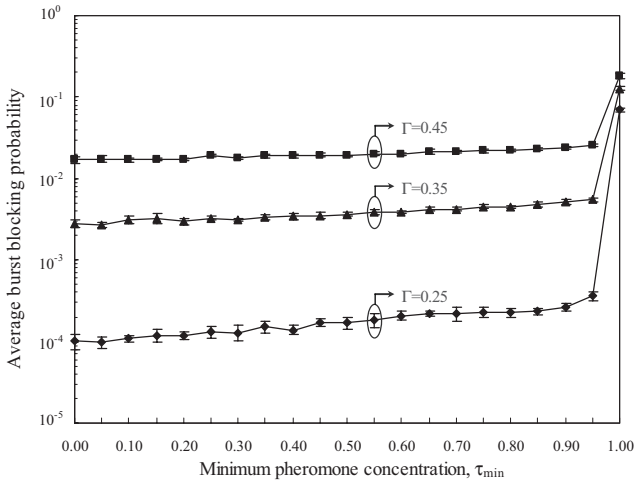


Figure 3: Performance impact of τ_{min}.

ISBN: 978-989-95079-6-8

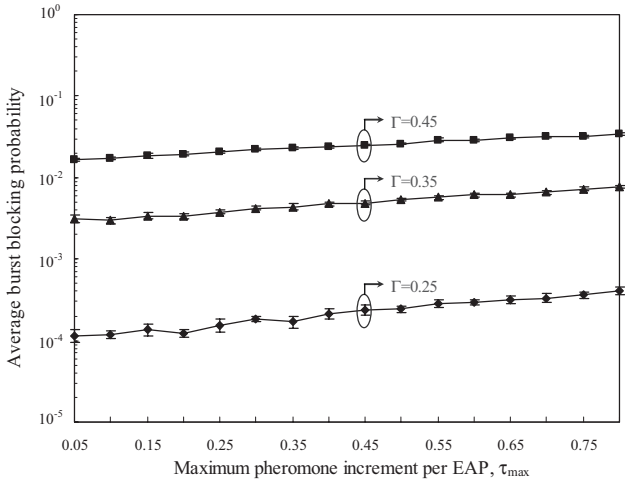


Figure 4: Performance impact of τ_{max}.

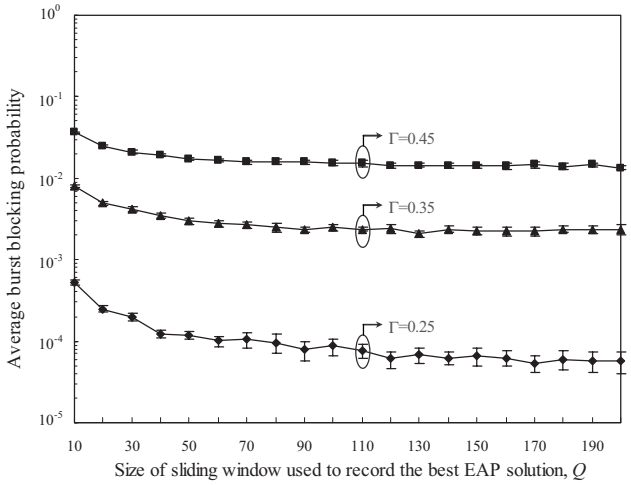


Figure 5: Performance impact of Q.

The results above consider three values of average offered traffic load in order to confirm that the behaviour observed is independent of the traffic load. According to the simulation results of Fig. 3, the loss performance slightly deteriorates as the minimum pheromone concentration on the adjacent nodes is increased. This is because the random component of routing the EAPs becomes larger, increasing the chances of modifications in the routing paths. In the limit case $\tau_{min}=1$, the network cannot optimize the set of routing paths.

From Fig. 4, it can be seen that the average burst blocking probability is slightly increased as the EAPs are allowed to deposit more pheromone in a single trip. This behaviour is attributed to more frequent modifications of the routing paths as a consequence of increased fluctuations of the pheromone values. A smaller pheromone increment makes the routing paths less sensitive to the goodness of the solution found by a single artificial ant, which means they are changed only when a significant number of EAPs have followed a given routing path and confirmed its superior quality.

As shown in Fig. 5, the use of a large Q can improve the network performance because it enhances the accuracy of the pheromone update process.

Overall, given the relatively small range of variation of the average burst blocking probability, the results suggest that DABR is quite robust to changes in its main parameters. The results also suggest that a larger Q and a smaller τ_{min} and τ_{max} slightly improve loss performance. However, it should be stressed that the simulations consider a single traffic pattern (in terms of average offered traffic load), whereas in more dynamic traffic scenarios a very large Q and very small τ_{min} and τ_{max} should be avoided because they can slow down the convergence to a new set of optimized routing paths whenever significant changes in the traffic pattern take place.

6 Conclusions

This paper has proposed a novel distributed routing path optimization framework for OBS networks based on ACO concepts. Simulation results have shown that the proposed framework can significantly improve network performance without needing to fine tune its parameters.

References

[1] M. Dorigo, M. Birattari and T. Stützle, Ant colony optimization: Artificial ants as a computational intelligence technique, *IEEE Computational Intelligence Magazine*, 1 (4): 28-39, Nov 2006.

[2] E. Bonabeau, M. Dorigo and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, Oxford University Press, 1999.

[3] M. Dorigo and T. Stützle, *Ant Colony Optimization*, Cambridge: MIT Press, 2004.

[4] R. Schoonderwoerd, O. Holland, J. Bruten and L. Rothkrantz, Ant-based load balancing in telecommunications networks, *Adaptive Behavior*, 5 (2): 169-207, 1996.

[5] G. Di Caro, and M. Dorigo, AntNet: Distributed stigmergetic control for communications networks, *Journal of Artificial Intelligence Research*, 9: 317-365, 1998.

[6] R. Garlick and R. Barr, Dynamic wavelength routing in WDM networks via ant colony optimization, in *Proc. of International Workshop on Ant Algorithms* (*ANTS*) *2002*, 250-255, Brussels, Belgium.

[7] S. Ngo, X. Jiang and S. Horiguchi, An ant-based approach to dynamic RWA in optical WDM networks, *Photonic Network Communications*, 11 (1): 39-48, January 2006.

[8] C. Qiao, and M. Yoo, Optical Burst Switching (OBS) – A new paradigm for an optical Internet, *Journal of High Speed Networks*, 8 (1): 69-84, January 1999.

[9] G. Pavani and H. Waldman, Traffic engineering and restoration in optical packet switching networks by means of ant colony optimization, in *Proc. of BROADNETS 2006*, San Jose, USA.

[10] Y. Chen, C. Qiao and X. Yu, Optical burst switching: A new area in optical network research, *IEEE Network*, 18 (3): 16-23, May 2004.

[11] J. Castro, J. Pedro and P. Monteiro, Burst loss reduction in OBS networks by minimizing network congestion, in *Proc. of ConfTele 2005*, Tomar, Portugal.

[12] G. Thodine, V. Vokkarane and J. Jue, Dynamic congestion-based load balanced routing in optical burst-switched networks, in *Proc. of IEEE GLOBECOM 2003*, 5: 2628-2632, San Francisco, USA.

[13] K. Sim and W. Sun, Ant colony optimization for routing and load-balancing: Survey and new directions, *IEEE Transactions on Systems, Man, and Cybernetics*, 33 (5): 560-572, September 2003.

[14] F. Kelly, Loss networks, *The Annals of Applied Probability*, 1 (3): 319-378, 1991.

[15] J. Pedro, J. Castro, P. Monteiro and J. Pires, On the modelling and performance evaluation of optical burst-switched networks, in *Proc. of IEEE CAMAD 2006*, 30-37, Trento, Italy.

ISBN: 978-989-95079-6-8

Non-conjunctive and non-disjunctive uninorms in Atanassov's intuitionistic fuzzy set theory

Glad Deschrijver

Fuzziness and Uncertainty Modelling Research Unit,
Department of Applied Mathematics and Computer Science, Ghent University,
Krijgslaan 281 (S9), B–9000 Gent, Belgium
Email: Glad.Deschrijver@UGent.be

***Abstract*— *Uninorms are a generalization of t-norms and t-conorms for which the neutral element is an element of* $[0,1]$ *which is not necessarily equal to* 0 *(as for t-norms) or* 1 *(as for t-conorms). Uninorms on the unit interval are either conjunctive or disjunctive, i.e. they aggregate the pair* $(0,1)$ *to either* 0 *or* 1. *In real-life applications, this kind of aggregation may be counter-intuitive. Atanassov's intuitionistic fuzzy set theory is an extension of fuzzy set theory which allows to model uncertainty about the membership degrees. In Atanassov's intuitionistic fuzzy set theory there exist uninorms which are neither conjunctive nor disjunctive. In this paper we study such uninorms more deeply and we investigate the structure of these uninorms. We also give several examples of uninorms which are neither conjunctive nor disjunctive.*

***Keywords*—** Conjunctive, disjunctive, interval-valued fuzzy set, intuitionistic fuzzy set, uninorm.

1 Introduction

Interval-valued fuzzy set theory [1, 2] is an extension of fuzzy set theory in which to each element of the universe a closed subinterval of the unit interval is assigned which approximates the unknown membership degree (using interval-valued fuzzy sets is not always the best approach to deal with uncertainty, see [3] for more information). Another extension of fuzzy set theory is intuitionistic fuzzy set theory introduced by Atanassov [4]. In [5] it is shown that intuitionistic fuzzy set theory is equivalent to interval-valued fuzzy set theory and that both are equivalent to L-fuzzy set theory in the sense of Goguen [6] w.r.t. a special lattice $\mathcal{L}^I$.

Uninorms are an important generalization of t-norms and t-conorms introduced by Yager and Rybalov [7]. Uninorms allow for a neutral element lying anywhere in the unit interval rather than at one or zero as is the case for t-norms and t-conorms. Uninorms on the unit interval are either conjunctive or disjunctive, i.e. they aggregate the pair $(0,1)$ to either 0 or 1. In real-life applications, this kind of aggregation may be counter-intuitive, e.g. in customer satisfaction modelling, if an aspect of the product receives a negative evaluation and another aspect a positive evaluation, then in general the global evaluation will neither be very negative or very positive, but rather be quite uncertain. This situation can be modelled by using uninorms in Atanassov's intuitionistic fuzzy set theory, which can be neither conjunctive nor disjunctive (see [8]). In this paper we therefore investigate such uninorms more deeply.

Definition 1.1 *We define* $\mathcal{L}^I = (L^I, \leq_{L^I})$, *where*

$$L^I = \{[x_1,x_2] \mid (x_1,x_2) \in [0,1]^2 \text{ and } x_1 \leq x_2\}, \tag{1}$$

$$[x_1,x_2] \leq_{L^I} [y_1,y_2] \iff (x_1 \leq y_1 \text{ and } x_2 \leq y_2), \tag{2}$$

for all $[x_1,x_2],[y_1,y_2]$ *in* L^I.

Similarly as Lemma 2.1 in [5] it can be shown that $\mathcal{L}^I$ is a complete lattice.

Definition 1.2 [1, 2] *An interval-valued fuzzy set on* U *is a mapping* $A : U \to L^I$.

Definition 1.3 [4] *An intuitionistic fuzzy set on* U *is a set*

$$A = \{(u, \mu_A(u), \nu_A(u)) \mid u \in U\}, \tag{3}$$

where $\mu_A(u) \in [0,1]$ *denotes the membership degree and* $\nu_A(u) \in [0,1]$ *the non-membership degree of* u *in* A *and where for all* $u \in U$, $\mu_A(u) + \nu_A(u) \leq 1$.

An intuitionistic fuzzy set A on U can be represented by the $\mathcal{L}^I$-fuzzy set A given by

$$\begin{array}{rcl} A : U & \to & L^I : \\ u & \mapsto & [\mu_A(u), 1-\nu_A(u)], \end{array} \tag{4}$$

In Figure 1 the set L^I is shown. Note that each $x = [x_1,x_2] \in L^I$ is represented by the point $(x_1,x_2) \in \mathbb{R}^2$.

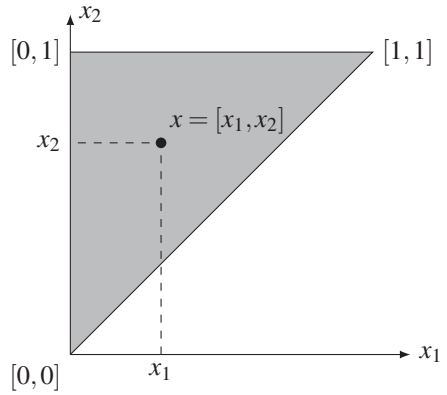


Figure 1: The grey area is L^I.

In the sequel, if $x \in L^I$, then we denote its bounds by x_1 and x_2, i.e. $x = [x_1,x_2]$. The smallest and the largest element of $\mathcal{L}^I$

ISBN: 978-989-95079-6-8

are given by $0_{\mathcal{L}^I} = [0,0]$ and $1_{\mathcal{L}^I} = [1,1]$. Note that, for x, y in L^I, $x <_{L^I} y$ is equivalent to $x \leq_{L^I} y$ and $x \neq y$, i.e. either $x_1 < y_1$ and $x_2 \leq y_2$, or $x_1 \leq y_1$ and $x_2 < y_2$. If for x, y in L^I it holds that either $x_1 < y_1$ and $x_2 > y_2$, or $x_1 > y_1$ and $x_2 < y_2$, then x and y are incomparable w.r.t. $\leq_{L^I}$, denoted as $x \|_{L^I} y$. We define for further usage the set $D = \{[x_1, x_1] \mid x_1 \in [0,1]\}$.

Definition 1.4 *A t-norm on $\mathcal{L}^I$ is a commutative, associative, increasing mapping $\mathcal{T} : (L^I)^2 \to L^I$ which satisfies $\mathcal{T}(1_{\mathcal{L}^I}, x) = x$, for all $x \in L^I$.*

A t-conorm on $\mathcal{L}^I$ is a commutative, associative, increasing mapping $\mathcal{S} : (L^I)^2 \to L^I$ which satisfies $\mathcal{S}(0_{\mathcal{L}^I}, x) = x$, for all $x \in L^I$.

Definition 1.5 *A negation on $\mathcal{L}^I$ is a decreasing mapping $\mathcal{N} : L^I \to L^I$ for which $\mathcal{N}(0_{\mathcal{L}^I}) = 1_{\mathcal{L}^I}$ and $\mathcal{N}(1_{\mathcal{L}^I}) = 0_{\mathcal{L}^I}$. If $\mathcal{N}(\mathcal{N}(x)) = x$, for all $x \in L^I$, then $\mathcal{N}$ is called involutive.*

Let N be a negation on $([0,1], \leq)$. Then the mapping $\mathcal{N}_N : L^I \to L^I$ defined by, for all $x \in L^I$,

$$\mathcal{N}_N(x) = [N(x_2), N(x_1)], \tag{5}$$

is a negation on $\mathcal{L}^I$.

We will also need the following result and definition (see [9, 10, 11, 12, 13]).

Theorem 1.1 *Let $(T_\alpha)_{\alpha \in A}$ be a family of t-norms on $([0,1], \leq)$ and $(]a_\alpha, e_\alpha[)_{\alpha \in A}$ be a family of non-empty, pairwise disjoint open subintervals of $[0,1]$. Then the function $T : [0,1]^2 \to [0,1]$ defined by, for all x, y in $[0,1]$,*

$$T(x,y) = \begin{cases} a_\alpha + (e_\alpha - a_\alpha) \cdot T_\alpha\left(\dfrac{x - a_\alpha}{e_\alpha - a_\alpha}, \dfrac{y - a_\alpha}{e_\alpha - a_\alpha}\right), \\ \quad \text{if } (x,y) \in [a_\alpha, e_\alpha]^2, \\ \min(x,y), \quad \text{otherwise}, \end{cases} \tag{6}$$

is a t-norm on $([0,1], \leq)$.

Definition 1.6 *Let $(T_\alpha)_{\alpha \in A}$ be a family of t-norms on $([0,1], \leq)$ and $(]a_\alpha, e_\alpha[)_{\alpha \in A}$ be a family of non-empty, pairwise disjoint open subintervals of $[0,1]$. The t-norm T defined by (6) is called the ordinal sum of the summands $\langle a_\alpha, e_\alpha, T_\alpha \rangle$, $\alpha \in A$, and we will write*

$$T = (\langle a_\alpha, e_\alpha, T_\alpha \rangle)_{\alpha \in A}. \tag{7}$$

2 Uninorms on $\mathcal{L}^I$

The following definition of a uninorm on $\mathcal{L}^I$ is a straightforward generalization of the definition of a uninorm on the unit interval introduced by Yager and Rybalov [7, 14].

Definition 2.1 [8] *A uninorm on $\mathcal{L}^I$ is a commutative, associative, increasing mapping $\mathcal{U} : (L^I)^2 \to L^I$ for which there exists an $e \in L^I$ such that $\mathcal{U}(e,x) = x$, for all $x \in L^I$. The element e is called the neutral element of $\mathcal{U}$.*

For any uninorm U on the unit interval, there exist increasing bijections $\phi_e : [0,e] \to [0,1]$ and $\psi_e : [e,1] \to [0,1]$ with increasing inverse, a t-norm T_U and a t-conorm S_U on $([0,1], \leq)$ such that [14]

(i) $(\forall (x,y) \in [0,e]^2)(U(x,y) = \phi_e^{-1}(T_U(\phi_e(x), \phi_e(y))))$;

(ii) $(\forall (x,y) \in [e,1]^2)(U(x,y) = \psi_e^{-1}(S_U(\psi_e(x), \psi_e(y))))$.

Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in L^I$. We define $E = \{x \mid x \in L^I \text{ and } x \leq_{L^I} e\}$ and $E' = \{x \mid x \in L^I \text{ and } x \geq_{L^I} e\}$. In [8] it is shown that if $e \notin D$, then there does not exist increasing bijections $\Phi_e : E \to L^I$ and $\Psi_e : E' \to L^I$ such that Φ_e^{-1} and Ψ_e^{-1} are increasing. On the other hand, if $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$, then the mappings $\Phi_e : E \to L^I$ and $\Psi_e : E' \to L^I$ defined by, for all $x \in L^I$,

$$\Phi_e(x) = \left[\frac{x_1}{e_1}, \frac{x_2}{e_1}\right], \tag{8}$$

$$\Psi_e(x) = \left[\frac{x_1 - e_1}{1 - e_1}, \frac{x_2 - e_1}{1 - e_1}\right]. \tag{9}$$

are increasing bijections for which the inverse is also increasing. As a consequence, the above result can only be extended if $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$.

From now on, we denote for any t-norm T and t-conorm S on $([0,1], \leq)$, $T_{\phi_e} = \phi_e^{-1} \circ T \circ (\phi_e \times \phi_e)$ and $S_{\psi_e} = \psi_e^{-1} \circ S \circ (\psi_e \times \psi_e)$, where $\times$ denotes the product operation [15]. A similar notation will be used for t-(co)norms and bijections on $\mathcal{L}^I$.

Theorem 2.1 [8] *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Then:*

(i) the mapping $\mathcal{T}_\mathcal{U} : (L^I)^2 \to L^I$ defined by, for all $x, y \in L^I$,

$$\mathcal{T}_\mathcal{U}(x,y) = \Phi_e(\mathcal{U}(\Phi_e^{-1}(x), \Phi_e^{-1}(y))) \tag{10}$$

is a t-norm on $\mathcal{L}^I$;

(ii) the mapping $\mathcal{S}_\mathcal{U} : (L^I)^2 \to L^I$ defined by, for all $x, y \in L^I$,

$$\mathcal{S}_\mathcal{U}(x,y) = \Psi_e(\mathcal{U}(\Psi_e^{-1}(x), \Psi_e^{-1}(y))) \tag{11}$$

is a t-conorm on $\mathcal{L}^I$.

Theorem 2.2 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Then for all x, y in L^I,*

$$x \leq_{L^I} e \leq_{L^I} y \implies \inf(x,y) \leq_{L^I} \mathcal{U}(x,y) \leq_{L^I} \sup(x,y). \tag{12}$$

These properties show that uninorms are well suited to model human evaluations (e.g. customer satisfaction). Customers which evaluate the performance of all aspects of a certain product high, have a tendency to give the global satisfaction degree an even higher value; on the other hand customers which globally consider the performance of the various aspects as insufficient, will give a low global evaluation. So we observe "reinforcement": a collection of high (low) rates "reinforce" each other and yield a global evaluation rate that is even higher (resp. lower) than each individual rate. If, however, a customer gives high scores only to some aspects and low scores for other aspects, then the global score will in general be located between the lowest and the highest value. This is "compensation". From Theorem 2.1 it follows that $\mathcal{U}\big|_{E^2}$ behaves like a t-norm, in particular $\mathcal{U}(x,y) \leq_{L^I} \inf(x,y)$, for all x, y in E. On the other hand, $\mathcal{U}\big|_{E'^2}$ behaves like a t-conorm, so $\mathcal{U}(x,y) \geq_{L^I} \sup(x,y)$, for all x, y in E'. Finally, if $x \leq_{L^I} e$ and $y \geq_{L^I} e$ (or conversely), then $\mathcal{U}(x,y)$ is a number between

ISBN: 978-989-95079-6-8

$\inf(x,y)$ and $\sup(x,y)$. So, clearly, uninorms show a reinforcing behaviour on E^2 and E'^2, and a compensating behaviour on $E \times E'$ and $E' \times E$ (see [16, 17, 18, 19] for more details).

For uninorms on the unit interval, however, $U(0,1)$ can only have two values: 0 or 1 (see [14]). In the first case the uninorm is called "conjunctive" and in the second case "disjunctive". However, in both cases the compensatory behaviour of the uninorm is violated. For uninorms on $\mathcal{L}^I$ we have the following property.

Theorem 2.3 [8] *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Then either $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = 0_{\mathcal{L}^I}$ or $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = 1_{\mathcal{L}^I}$ or $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \parallel_{L^I} e$.*

Hence uninorms on $\mathcal{L}^I$ are not necessarily conjunctive or disjunctive. It is possible that a uninorm on $\mathcal{L}^I$ shows compensatory behaviour between $0_{\mathcal{L}^I}$ and $1_{\mathcal{L}^I}$. If one aspect of a product has a very negative evaluation ($0_{\mathcal{L}^I}$) and another aspect is very positively evaluated ($1_{\mathcal{L}^I}$), then in general it will be very difficult to give a global evaluation of the product, in fact the global evaluation will contain a lot of uncertainty. Therefore it makes more sense to use a uninorm $\mathcal{U}$ for which $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \parallel_{L^I} e$.

3 Uninorms on $\mathcal{L}^I$ which are neither conjunctive nor disjunctive

In this section we try to obtain more information about the structure of uninorms which are neither conjunctive nor disjunctive by investigating the possible values of $\mathcal{U}(x,y)$ with x, y in L^I. First we give an example of a uninorm on $\mathcal{L}^I$ that is neither conjunctive nor disjunctive, in order to show that such uninorms do exists.

Example 3.1 *Let for all $e_1 \in]0,1[$, U_{e_1} be the uninorm on $([0,1], \leq)$ defined by, for all x_1, y_1 in $[0,1]$,*

$$U_{e_1}(x_1,y_1) = \begin{cases} \max(x_1,y_1), & \text{if } x_1 \geq e_1 \text{ and } y_1 \geq e_1; \\ \min(x_1,y_1), & \text{else.} \end{cases} \quad (13)$$

Let now, for all x, y in L^I,

$$\mathcal{U}(x,y) = [U_{e_1}(x_1,y_1), 1 - U_{1-e_1}(1-x_2, 1-y_2)]. \quad (14)$$

Then $\mathcal{U}$ is a uninorm on $\mathcal{L}^I$ with neutral element $e = [e_1, e_1]$. Since $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = [0,1]$, $\mathcal{U}$ is neither conjunctive nor disjunctive.

In general, if U_1 is an arbitrary conjunctive uninorm and U_2 an arbitrary disjunctive uninorm on $([0,1], \leq)$, then the mapping $\mathcal{U} : (L^I)^2 \to L^I : (x,y) \mapsto [U_1(x_1,y_1), U_2(x_2,y_2)]$, for all x, y in L^I, is a uninorm on $\mathcal{L}^I$ for which $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = [0,1]$.

Lemma 3.1 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Then, for all $x \in L^I$,*

(i) either $\mathcal{U}(0_{\mathcal{L}^I}, x) = 0_{\mathcal{L}^I}$ or $\mathcal{U}(0_{\mathcal{L}^I}, x) \notin E$,

(ii) either $\mathcal{U}(1_{\mathcal{L}^I}, x) = 1_{\mathcal{L}^I}$ or $\mathcal{U}(1_{\mathcal{L}^I}, x) \notin E'$.

Theorem 3.2 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. If $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \parallel_{L^I} e$, then, for all $x \in L^I$,*

(i) $\mathcal{U}(0_{\mathcal{L}^I}, x) \parallel_{L^I} e$ or $\mathcal{U}(0_{\mathcal{L}^I}, x) = 0_{\mathcal{L}^I}$,

(ii) $\mathcal{U}(1_{\mathcal{L}^I}, x) \parallel_{L^I} e$ or $\mathcal{U}(1_{\mathcal{L}^I}, x) = 1_{\mathcal{L}^I}$.

If one aspect of a product has a negative evaluation $x \in L^I$ with $x \leq_{L^I} e$ and another aspect has a positive evaluation $y \in L^I$ with $y \geq_{L^I} e$, then the global evaluation will be rather neutral and contain some uncertainty. Therefore it is natural to expect that $\mathcal{U}(x,y) \parallel_{L^I} e$. We investigate for which x and y in L^I this is the case.

Lemma 3.3 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Assume that $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \parallel_{L^I} e$.*

(i) Let arbitrarily $x \in E$. If $\mathcal{U}(1_{\mathcal{L}^I}, x) = 1_{\mathcal{L}^I}$, then $\mathcal{U}(1_{\mathcal{L}^I}, [x_1, y_2]) = 1_{\mathcal{L}^I}$, for all $y_2 \in [x_1, e_1]$.

(ii) Let arbitrarily $x \in E'$. If $\mathcal{U}(0_{\mathcal{L}^I}, x) = 0_{\mathcal{L}^I}$, then $\mathcal{U}(0_{\mathcal{L}^I}, [y_1, x_2]) = 0_{\mathcal{L}^I}$, for all $y_1 \in [e_1, x_2]$.

Theorem 3.4 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. If $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \parallel_{L^I} e$, then*

(i) there exists an $\alpha \in D \cap E$ such that (see Figure 2)

- *$\mathcal{U}(1_{\mathcal{L}^I}, x) \parallel_{L^I} e$ for all $x \in L^I$ satisfying $x_1 < \alpha_1$ and $x_2 \leq e_1$, and*
- *$\mathcal{U}(1_{\mathcal{L}^I}, x) = 1_{\mathcal{L}^I}$, for all $x \in L^I$ satisfying $x_1 > \alpha_1$,*

(ii) there exists a $\beta \in D \cap E'$ such that

- *$\mathcal{U}(0_{\mathcal{L}^I}, x) \parallel_{L^I} e$ for all $x \in L^I$ satisfying $x_1 \geq e_1$ and $x_2 > \beta_1$, and*
- *$\mathcal{U}(0_{\mathcal{L}^I}, x) = 0_{\mathcal{L}^I}$, for all $x \in L^I$ satisfying $x_2 < \beta_1$.*

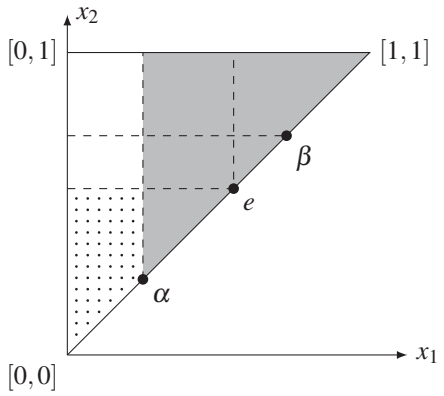


Figure 2: The grey area is the set of elements x for which $\mathcal{U}(x, 1_{\mathcal{L}^I}) = 1_{\mathcal{L}^I}$, the dotted area is the set of elements x for which $\mathcal{U}(x, 1_{\mathcal{L}^I}) \parallel_{L^I} e$.

Example 3.2 We give an example of a uninorm which satisfies the results in Theorem 3.4 for a non-trivial α and β. Let arbitrarily $\alpha \in D \cap E \setminus \{0_{\mathcal{L}^I}, e\}$ and $\beta \in D' \cap E' \setminus \{e, 1_{\mathcal{L}^I}\}$. Let T_{1a} and T_{1b} be arbitrary t-norms, S_{1a} and S_{1b} arbitrary t-conorms on $([0,1], \leq)$, and define

$$T_1 = (\langle 0, \phi_e(\alpha_1), T_{1a} \rangle, \langle \phi_e(\alpha_1), 1, T_{1b} \rangle), \quad (15)$$

$$S_2 = (\langle 0, \psi_e(\beta_1), S_{2a} \rangle, \langle \psi_e(\beta_1), 1, S_{2b} \rangle). \quad (16)$$

ISBN: 978-989-95079-6-8

Define the mappings $U_1 : [0,1]^2 \to [0,1]$ and $U_2 : [0,1]^2 \to [0,1]$ by, for all x_1, y_1, x_2, y_2 in $[0,1]$,

$$U_1(x_1,y_1) = \begin{cases} (T_1)_{\phi_e}(x_1,y_1), \text{ if } \max(x_1,y_1) \le e_1, \\ (S_1)_{\psi_e}(x_1,y_1), \text{ if } \min(x_1,y_1) \ge e_1, \\ 1, \text{ if } (x_1 > \alpha_1 \text{ and } y_1 = 1) \\ \quad \text{or } (y_1 > \alpha_1 \text{ and } x_1 = 1), \\ \min(x_1,y_1), \text{ else}, \end{cases} \tag{17}$$

$$U_2(x_2,y_2) = \begin{cases} (T_2)_{\phi_e}(x_2,y_2), \text{ if } \max(x_2,y_2) \le e_1, \\ (S_2)_{\psi_e}(x_2,y_2), \text{ if } \min(x_2,y_2) \ge e_1, \\ 0, \text{ if } (x_2 < \beta_1 \text{ and } y_2 = 0) \\ \quad \text{or } (y_2 < \beta_1 \text{ and } x_2 = 0), \\ \max(x_2,y_2), \text{ else}, \end{cases} \tag{18}$$

Then U_1 is a conjunctive uninorm and U_2 is a disjunctive uninorm on $([0,1],\le)$. The mapping $\mathcal{U} : (L^I)^2 \to L^I$ defined by, for all x, y in L^I,

$$\mathcal{U}(x,y) = [U_1(x_1,y_1), U_2(x_2,y_2)], \tag{19}$$

is a uninorm on $\mathcal{L}^I$ for which $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = [0,1]$ and for which the results in Theorem 3.4 hold for the given α and β.

From now on α and β will be the elements of $\mathcal{L}^I$ introduced in Theorem 3.4.

Lemma 3.5 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. If $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \|_{L^I} e$, then for all $x \in E$ and $y \in E'$ satisfying $x_1 < \alpha_1$ and $y_2 > \beta_1$ it holds that $\mathcal{U}(x,y) \|_{L^I} e$.*

Theorem 3.6 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. If $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \|_{L^I} e$, then for all $x \in E$ and $y \in E'$ satisfying $x_1 < \alpha_1$ and $y_2 > \beta_1$ it holds that $(\mathcal{U}(x,y))_1 \le \alpha_1$ and $(\mathcal{U}(x,y))_2 \ge \beta_1$.*

Corollary 3.7 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Assume that $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \|_{L^I} e$.*

(i) Let arbitrarily $a = [\alpha_1, a_2] \in E$ and $y \in E'$ such that $y_2 > \beta_1$. Then

$$\lim_{\substack{x \to a \\ x_1 < \alpha_1}} (\mathcal{U}(x,y))_1 = \alpha_1. \tag{20}$$

(ii) Let arbitrarily $b = [b_1, \beta_2] \in E'$ and $x \in E$ such that $x_1 < \alpha_1$. Then

$$\lim_{\substack{y \to b \\ y_2 > \beta_2}} (\mathcal{U}(x,y))_2 = \beta_2. \tag{21}$$

In the above, the limits are calculated using on L^I the Euclidean metric function $d^E(x,y) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}$, for all x, y in L^I.

Theorem 3.8 *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Assume that $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) \|_{L^I} e$.*

(i) For all $x \in E$ and $y \in E'$ satisfying $x_1 > \alpha_1$ and $y_2 > \beta_1$ it holds that $\mathcal{U}(x,y) \ge_{L^I} [\alpha_1, \beta_1]$.

(ii) For all $x \in E$ and $y \in E'$ satisfying $x_1 < \alpha_1$ and $y_2 < \beta_1$ it holds that $\mathcal{U}(x,y) \le_{L^I} [\alpha_1, \beta_1]$.

4 The value of $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I})$

In this section we check which are the possible values for $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I})$ in the case that $\mathcal{U}$ is neither conjunctive nor disjunctive.

Lemma 4.1 *For any $\alpha \in L^I$ and $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$ such that $\alpha \|_{L^I} e$, $\alpha_1 > 0$ and $\alpha_2 < 1$, there exists an involutive negation N on $([0,1],\le)$ such that $N(\alpha_1) = \alpha_2$ and $N(e_1) = e_1$.*

Theorem 4.2 *Let $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$, $\alpha \in L^I$, T_1 and T_2 be t-norms, S_1 and S_2 be t-conorms on $([0,1],\le)$ such that*

(i) $\alpha \|_{L^I} e$,

(ii) there exist t-norms T_{1a} and T_{1b} on $([0,1],\le)$ such that $T_1 = (\langle 0, \phi_e(\alpha_1), T_{1a}\rangle, \langle \phi_e(\alpha_1), 1, T_{1b}\rangle)$,

(iii) there exist t-conorms S_{2a} and S_{2b} on $([0,1],\le)$ such that $S_2 = (\langle 0, \psi_e(\alpha_2), S_{2a}\rangle, \langle \psi_e(\alpha_2), 1, S_{2b}\rangle)$,

(iv) $T_1(x_1,y_1) \le T_2(x_1,y_1)$ and $S_1(x_1,y_1) \le S_2(x_1,y_1)$, for all x_1, y_1 in $[0,1]$.

Define the mapping $\mathcal{U} : (L^I)^2 \to L^I$ by, for all x, y in L^I,

$$(\mathcal{U}(x,y))_1 = \begin{cases} \alpha_1, \text{ if } (x_1 < \alpha_1 \text{ and } y_1 \ge \alpha_1 \text{ and } y_2 > e_1) \\ \quad \text{or } (y_1 < \alpha_1 \text{ and } x_1 \ge \alpha_1 \text{ and } x_2 > e_1), \\ U_1(x_1,y_1), \text{ else}, \end{cases} \tag{22}$$

$$(\mathcal{U}(x,y))_2 = \begin{cases} \alpha_2, \text{ if } (x_2 > \alpha_2 \text{ and } y_2 \le \alpha_2 \text{ and } y_1 < e_1) \\ \quad \text{or } (y_2 > \alpha_2 \text{ and } x_2 \le \alpha_2 \text{ and } x_1 < e_1), \\ U_2(x_2,y_2), \text{ else}. \end{cases} \tag{23}$$

where, for all x_1, y_1, x_2, y_2 in $[0,1]$,

$$U_1(x_1,y_1) = \begin{cases} (T_1)_{\phi_e}(x_1,y_1), & \text{if } \max(x_1,y_1) \le e_1, \\ (S_1)_{\psi_e}(x_1,y_1), & \text{if } \min(x_1,y_1) \ge e_1, \\ \min(x_1,y_1), & \text{else}, \end{cases} \tag{24}$$

$$U_2(x_2,y_2) = \begin{cases} (T_2)_{\phi_e}(x_2,y_2), & \text{if } \max(x_2,y_2) \le e_1, \\ (S_2)_{\psi_e}(x_2,y_2), & \text{if } \min(x_2,y_2) \ge e_1, \\ \max(x_2,y_2), & \text{else}. \end{cases} \tag{25}$$

Then $\mathcal{U}$ is a uninorm on $\mathcal{L}^I$ with neutral element e for which $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = \alpha$.

Theorem 4.2 shows that for any $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$ and any $\alpha \in L^I$ such that $\alpha \|_{L^I} e$, there exists a uninorm $\mathcal{U}$ on $\mathcal{L}^I$ with neutral element e such that $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = \alpha$.

In the following theorem we show that for most values of $\alpha \in L^I$ such that $\alpha \|_{L^I} e$, it is even possible to find uninorms satisfying $\mathcal{U}(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}) = \alpha$, which are self-dual.

Theorem 4.3 *Let $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$, $\alpha \in L^I$, T be a t-norm, S a t-conorm and N a negation on $([0,1],\le)$ such that*

(i) $\alpha \|_{L^I} e$ and either $\alpha \|_{L^I} [0,1]$ or $\alpha = [0,1]$,

(ii) N is involutive, $N(\alpha_1) = \alpha_2$ and $N(e_1) = e_1$,

(iii) there exist t-norms T_a and T_b on $([0,1],\le)$ such that $T = (\langle 0, \phi_e(\alpha_1), T_a\rangle, \langle \phi_e(\alpha_1), 1, T_b\rangle)$,

ISBN: 978-989-95079-6-8

(iv) $T_{\phi_e}(x_1,y_1) \leq N(S_{\psi_e}(N(x_1),N(y_1)))$, *for all* x_1 *and* y_1 *in* $[0,e_1]$.

Define the mapping $\mathcal{U} : (L^I)^2 \to L^I$ *by, for all* x,y *in* L^I,

$$(\mathcal{U}(x,y))_1 = \begin{cases} \alpha_1, \text{ if } (x_1 < \alpha_1 \text{ and } y_1 \geq \alpha_1 \text{ and } y_2 > e_1) \\ \quad \text{or } (y_1 < \alpha_1 \text{ and } x_1 \geq \alpha_1 \text{ and } x_2 > e_1), \\ U(x_1,y_1), \text{ else}, \end{cases} \tag{26}$$

$$(\mathcal{U}(x,y))_2 = N((\mathcal{U}(\mathcal{N}_N(x),\mathcal{N}_N(y)))_1). \tag{27}$$

where, for all x_1,y_1 *in* $[0,1]$,

$$U(x_1,y_1) = \begin{cases} T_{\phi_e}(x_1,y_1), & \text{if } \max(x_1,y_1) \leq e_1, \\ S_{\psi_e}(x_1,y_1), & \text{if } \min(x_1,y_1) \geq e_1, \\ \min(x_1,y_1), & \text{else}. \end{cases} \tag{28}$$

Then $\mathcal{U}$ *is a uninorm on* $\mathcal{L}^I$ *with neutral element* e *for which* $\mathcal{U}(0_{\mathcal{L}^I},1_{\mathcal{L}^I}) = \alpha$ *and, for all* x,y *in* L^I,

$$\mathcal{U}(x,y) = \mathcal{N}_N(\mathcal{U}(\mathcal{N}_N(x),\mathcal{N}_N(y))). \tag{29}$$

Example 4.1 *Let arbitrarily* $e \in D$ *and* $\alpha \in L^I$ *with* $\alpha \|_{L^I} e$ *and* $\alpha \|_{L^I} [0,1]$. *Define for all* $x_1 \in [0,1]$,

$$N(x_1) = \begin{cases} 1 - \frac{1-\alpha_2}{\alpha_1}x_1, & \text{if } x_1 \in [0,\alpha_1], \\ e_1 + \frac{\alpha_2 - e_1}{\alpha_1 - e_1}(x_1 - e_1), & \text{if } x_1 \in [\alpha_1,e_1], \\ e_1 + \frac{\alpha_1 - e_1}{\alpha_2 - e_1}(x_1 - e_1), & \text{if } x_1 \in [e_1,\alpha_2], \\ -\frac{\alpha_1}{1-\alpha_2}(x_1 - e_1), & \text{if } x_1 \in [\alpha_2,1]. \end{cases} \tag{30}$$

Then N *is an involutive negation with* $N(\alpha_1) = \alpha_2$ *and* $N(e_1) = e_1$. *Define* $T = (\langle 0, \phi_e(\alpha_1), P\rangle, \langle \phi_e(\alpha_1), 1, \min\rangle)$, *where* P *is the product t-norm on the unit interval. Then for all* $(x_1,y_1) \in [0,e_1]^2$,

$$T_{\phi_e}(x_1,y_1) = \begin{cases} \frac{1}{\alpha_1}x_1y_1, & \text{if } (x_1,y_1) \in [0,\alpha_1]^2, \\ \min(x_1,y_1), & \text{else}. \end{cases} \tag{31}$$

Let now for all $(x_1,y_1) \in [e_1,1]^2$, $S_{\psi_e} = N \circ T_{\phi_e} \circ (N \times N)$. *Define* U, $(\mathcal{U}(x,y))_1$ *and* $(\mathcal{U}(x,y))_2$ *in a similar way as in Theorem 4.3. Then* $\mathcal{U}$ *is a uninorm on* $\mathcal{L}^I$ *with neutral element* e *for which* $\mathcal{U}(0_{\mathcal{L}^I},1_{\mathcal{L}^I}) = \alpha$ *and which is self-dual w.r.t.* $\mathcal{N}_N$.

The question remains whether for any $e \in D \setminus \{0_{\mathcal{L}^I},1_{\mathcal{L}^I}\}$, any $\alpha \in L^I$ such that $\alpha \|_{L^I} e$, and also any t-norm $\mathcal{T}$ and any t-conorm $\mathcal{S}$ on $\mathcal{L}^I$, there exists a uninorm $\mathcal{U}$ on $\mathcal{L}^I$ with neutral element e such that $\mathcal{U}(0_{\mathcal{L}^I},1_{\mathcal{L}^I}) = \alpha$, $\mathcal{T}_{\mathcal{U}} = \mathcal{T}$ and $\mathcal{S}_{\mathcal{U}} = \mathcal{S}$.

5 Conclusion

In this paper we have studied uninorms on the lattice $\mathcal{L}^I$, which is the underlying lattice of both Atanassov's intuitionistic fuzzy set theory and interval-valued fuzzy set theory. Such uninorms $\mathcal{U}$ can be neither conjunctive nor disjunctive, in which case $\mathcal{U}(0_{\mathcal{L}^I},1_{\mathcal{L}^I})$ is an element of L^I which is incomparable to the neutral element of $\mathcal{U}$. We have investigated the value $\mathcal{U}(x,y)$ in the case that x and y are located in certain areas of L^I and we have found several restrictions. For any value of $\alpha \in L^I$ which is incomparable to an arbitrary element e, we have constructed a uninorm $\mathcal{U}$ with neutral element e and for which $\mathcal{U}(0_{\mathcal{L}^I},1_{\mathcal{L}^I}) = \alpha$.

References

[1] M. B. Gorzałczany. A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, 21(1):1–17, Jan 1987.

[2] R. Sambuc. *Fonctions Φ-floues. Application à l'aide au diagnostic en pathologie thyroidienne*. PhD thesis, Université de Marseille, France, 1975.

[3] D. Dubois. On ignorance and contradiction considered as truth-values. *Logic Journal of the IGPL*, 16(2):195–216, 2008.

[4] K. T. Atanassov. *Intuitionistic fuzzy sets*. Physica-Verlag, Heidelberg, New York, 1999.

[5] G. Deschrijver and E. E. Kerre. On the relationship between some extensions of fuzzy set theory. *Fuzzy Sets and Systems*, 133(2):227–235, 2003.

[6] J. A. Goguen. L-fuzzy sets. *Journal of Mathematical Analysis and Applications*, 18(1):145–174, Apr 1967.

[7] R. R. Yager and A. Rybalov. Uninorm aggregation operators. *Fuzzy Sets and Systems*, 80(1):111–120, May 1996.

[8] G. Deschrijver and E. E. Kerre. Uninorms in L^*-fuzzy set theory. *Fuzzy Sets and Systems*, 148(2):243–262, 2004.

[9] A. H. Clifford. Naturally totally ordered commutative semigroups. *Amer. J. Math.*, 76:631–646, 1954.

[10] S. Jenei. A note on the ordinal sum theorem and its consequence for the construction of triangular norms. *Fuzzy Sets and Systems*, 126(2):199–205, Mar 2002.

[11] C.-H. Ling. Representation of associative functions. *Publ. Math. Debrecen*, 12:189–212, 1965.

[12] S. Saminger. On ordinal sums of triangular norms on bounded lattices. *Fuzzy Sets and Systems*, 157(10):1403–1416, May 2006.

[13] B. Schweizer and A. Sklar. Associative functions and abstract semigroups. *Publ. Math. Debrecen*, 10:69–81, 1963.

[14] J. C. Fodor, R. R. Yager, and A. Rybalov. Structure of uninorms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5(4):411–427, Aug 1997.

[15] M. Eisenberg. *Axiomatic theory of sets and classes*. Holt, Rinehart and Winston, Inc., New York, 1971.

[16] B. Depaire, K. Vanhoof, and G. Wets. The application of uninorms in importance-performance analysis. In *FS'06: Proceedings of the 7th WSEAS International Conference on Fuzzy Systems*, pages 1–7, Stevens Point, Wisconsin, USA, 2006. World Scientific and Engineering Academy and Society (WSEAS).

[17] K. Vanhoof, P. Pauwels, J. Dombi, T. Brijs, and G. Wets. Penalty-reward analysis with uninorms: A study of customer (dis)satisfaction. In D. Ruan, C. Chen, E. E. Kerre, and G. Wets, editors, *Intelligent Data Mining: Techniques and Applications*, volume 5, pages 237–252. 2005.

[18] B. Depaire, K. Vanhoof, and G. Wets. Managerial opportunities of uninorm-based importance-performance analysis. *WSEAS Transactions on Business and Economics*, 3(3):101–108, 2007.

[19] B. Depaire, K. Vanhoof, and G. Wets. Expectation-performance compatibility in a customer satisfaction context modelled by means of aggregation operators. In B. De Baets, K. Maes, and R. Mesiar, editors, *Proceedings of the 4th International Summer School on Aggregation Operators*, pages 45–50. Academia Press, 2007.

ISBN: 978-989-95079-6-8

Fuzzy Descriptions Logics with Fuzzy Truth Values

Fernando Bobillo[1] Umberto Straccia[2]

1. Dpt. of Computer Science & Systems Engineering University of Zaragoza, Spain
2. ISTI - CNR, Pisa, Italy
Email: fbobillo@unizar.es, straccia@isti.cnr.it

***Abstract*— *Fuzzy Description Logics are a family of logics which allow to deal with structured knowledge affected by vagueness. Although a relatively important amount of work has been carried out in the last years, current fuzzy DLs are open to be extend with several features worked out in the fuzzy logic literature. In this work, we extend fuzzy DLs with fuzzy truth values, allowing to state sentences such as "Tina is young is* almost *true".**

***Keywords*—** Fuzzy Description Logics, Fuzzy Truth Values

1 Introduction

Description Logics (DLs) [1], have gained popularity due to their application in the context of the *Semantic Web* [2]. *Ontologies* play a key role in the Semantic Web. An ontology consists of a hierarchical description of important concepts in a particular domain, along with the description of the properties (of the instances) of each concept. DLs play a particular role in this context as they are essentially the theoretical counterpart of the *Web Ontology Language OWL DL*, a state of the art language to specify ontologies.

It is well-known that "classical" ontology languages are not appropriate to deal with *fuzzy/vague/imprecise knowledge*, which is inherent to several real world domains [3, 4]. Since fuzzy set theory and fuzzy logic are suitable formalisms to handle these types of knowledge, fuzzy ontologies emerge as useful in several applications, such as (multimedia) information retrieval, image interpretation, ontology mapping, matchmaking and the Semantic Web [5]. So far, several fuzzy extensions of DLs can be found in the literature (see the survey in [5]) and some fuzzy DL reasoners exist, such as FUZZYDL [6], DELOREAN [7], FIRE [8] or DLMEDIA [9].

In this paper we allow fuzzy DL sentences to be qualified with *fuzzy truth values* [10], and, thus, allow expressions such as "Tina is young is *very true*" and "Tina is young is *almost true*". We show the syntax, semantics and reasoning algorithms for the extension provided in this paper.

We proceed as follows. To make the paper self-contained, the next section recalls salient notions of mathematical fuzzy logics [11]. Section 3 introduces fuzzy $\mathcal{ALC}$ [3], which is the fuzzy variant of the Description Logic $\mathcal{ALC}$. $\mathcal{ALC}$ is usually considered as a reference language, whenever new features are introduced into DLs. Section 4 extends fuzzy DLs with fuzzy truth values and Section 5 provides reasoning algorithms. Section 6 concludes and describes future work.

2 Preliminaries: Mathematical Fuzzy Logic

In the setting of fuzzy logics, the convention prescribing that a statement is either true or false is changed and is a matter of degree measured on an ordered scale $\mathcal{S}$ that is no longer $\{0, 1\}$, but, e.g., the unit interval $[0, 1]$. This degree of fit is called *degree of truth* of the statement ϕ in the interpretation $\mathcal{I}$. Fuzzy logics logics provide compositional calculi of degrees of truth, including degrees between "true" and "false". A statement is now not true or false only, but may have a truth degree taken from a *truth space* $\mathcal{S}$, usually $[0, 1]$ (in that case we speak about *Mathematical Fuzzy Logic* [11]).

In the illustrative fuzzy logic that we consider in this section, *fuzzy statements* have the form $\phi \geqslant l$ or $\phi \leqslant u$, where $l, u \in [0, 1]$ [13, 11] and ϕ is a statement, which encode that the degree of truth of ϕ is *at least* l resp. *at most* u. For example, $ripe\_tomato \geqslant 0.9$ says that we have a rather ripe tomato (the degree of truth of $ripe\_tomato$ is at least 0.9).

A *fuzzy interpretation* $\mathcal{I}$ maps each basic statement p_i into $[0, 1]$ and is then extended inductively to all statements:

$$\begin{array}{lcl} \mathcal{I}(\phi \wedge \psi) & = & \mathcal{I}(\phi) \otimes \mathcal{I}(\psi)\,; \\ \mathcal{I}(\phi \vee \psi) & = & \mathcal{I}(\phi) \oplus \mathcal{I}(\psi)\,; \\ \mathcal{I}(\phi \rightarrow \psi) & = & \mathcal{I}(\phi) \Rightarrow \mathcal{I}(\psi)\,; \\ \mathcal{I}(\neg\phi) & = & \ominus \mathcal{I}(\phi)\,, \end{array}$$

where $\otimes$, $\oplus$, $\Rightarrow$, and $\ominus$ are so-called *truth combination functions*, namely, *triangular norms* (or *t-norms*), *triangular conorms* (*t-conorms*), *implication functions*, and *negation functions*, respectively, which extend the classical Boolean conjunction, disjunction, implication, and negation, respectively, to the fuzzy case (see [11] for a formal definition of these functions and their properties). Several t-norms, t-conorms, implication functions, and negation functions have been given in the literature, giving raise to different fuzzy logics with different logical properties. In fuzzy logic, one usually distinguishes three different logics (see Fig. 1), namely Łukasiewicz, Gödel, and Product logic [11]. Zadeh "logic" (fuzzy operators originally considered by Zadeh [12]) is entailed by Łukasiewicz logic.

A *fuzzy set* R over a countable crisp set X is a function $R\colon X \rightarrow [0, 1]$. A (binary) *fuzzy relation* R over two countable crisp sets X and Y is a function $R\colon X \times Y \rightarrow [0, 1]$. The *degree of subsumption* between two fuzzy sets A and B is defined as $\inf_{x \in X} A(x) \Rightarrow B(x)$.

The notions of satisfiability and logical consequence are defined in the standard way. A fuzzy interpretation $\mathcal{I}$ *satisfies* a fuzzy statement $\phi \geqslant l$ (resp., $\phi \leqslant u$) or $\mathcal{I}$ is a *model* of $\phi \geqslant l$ (resp., $\phi \leqslant u$), denoted $\mathcal{I} \models \phi \geqslant l$ (resp., $\mathcal{I} \models \phi \leqslant u$), iff $\mathcal{I}(\phi) \geqslant l$ (resp., $\mathcal{I}(\phi) \leqslant u$). $\phi \geqslant l$ is a *tight logical consequence* of a set of fuzzy statements $\mathcal{K}$ iff l is the infimum of $\mathcal{I}(\phi)$ subject to all models $\mathcal{I}$ of $\mathcal{K}$. The latter value is equivalent to $l = \sup\{r \mid \mathcal{K} \models \phi \geqslant r\}$, it is called *Best Entailment Degree* (BED), and is denoted $bed(\mathcal{K}, \phi)$, while the *Best Satisfiability Degree*, denoted as $bsd(\mathcal{K}, \phi)$ is defined as $\sup_{\mathcal{I} \models \mathcal{K}} \mathcal{I}(\phi)$. We refer the reader to [11, 13, 14] for reasoning algorithms for fuzzy propositional and First-Order Logics (FOLs).

ISBN: 978-989-95079-6-8

	Łukasiewicz Logic	Gödel Logic	Product Logic	Zadeh Logic
$a \otimes b$	$\max(a+b-1,0)$	$\min(a,b)$	$a \cdot b$	$\min(a,b)$
$a \oplus b$	$\min(a+b,1)$	$\max(a,b)$	$a+b-a \cdot b$	$\max(a,b)$
$a \Rightarrow b$	$\min(1-a+b,1)$	$\begin{cases} 1 & \text{if } a \leq b \\ b & \text{otherwise} \end{cases}$	$\min(1,b/a)$	$\max(1-a,b)$
$\ominus a$	$1-a$	$\begin{cases} 1 & \text{if } a=0 \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 1 & \text{if } a=0 \\ 0 & \text{otherwise} \end{cases}$	$1-a$

Figure 1: Combination functions of various fuzzy logics.

For illustrative purposes, we provide a simple and effective way to solve the entailment problem, in the case of Zadeh logic and Lukasiewicz logic, in terms of Mixed Integer Linear Programming (MILP) –see also [14]. The calculus depends on the t-norm, t-conorm and negation functions considered. Suppose we are looking for $bed(\mathcal{K},\phi)$, then,

$$bed(\mathcal{K},\phi) = \min\{x \mid \mathcal{K} \cup \{\phi \leq x\} \text{ is satisfiable.}\} \quad (1)$$

Indeed, suppose the minimal value is $\bar{n}$. We will know then that for any interpretation $\mathcal{I}$ satisfying $\mathcal{K}$, it cannot be $\mathcal{I}(\phi) < \bar{n}$, and, thus, $\mathcal{I}(\phi) \geq \bar{n}$ has to hold.

The above problem can be solved by means of MILP (we consider Zadeh logic only, for Lukasiewicz logic the procedure is similar) [1]. For a formula ϕ consider a variable x_ϕ. The intuition is that the degree of truth of ϕ is greater or equal to x_ϕ. The MILP problem determining $bed(\mathcal{K},\phi)$ is as follows:

$$\begin{array}{l} \min x. \text{ such that } x \in [0,1] \\ \quad x_{\neg\phi} \geq 1-x, \sigma(\neg\phi), \\ \quad \text{for all } (\phi' \geq n) \in \mathcal{K}, x_{\phi'} \geq n, \sigma(\phi'), \\ \quad \text{for all } (\phi' \leq n) \in \mathcal{K}, x_{\neg\phi'} \geq 1-n, \sigma(\neg\phi') \end{array} \quad (2)$$

where the function σ, transforming a many-valued proposition into a set of inequations, is inductively defined as follows:

$$\sigma(\phi) = \begin{cases} x_p \in [0,1] & \text{if } \phi = p \\ x_{\phi'} = 1 - x_\phi, x_\phi \in [0,1] & \text{if } \phi = \neg\phi' \\ x_{\phi_1} \geq x_\phi, x_{\phi_2} \geq x_\phi, \\ \sigma(\phi_1), \sigma(\phi_2), x_\phi \in [0,1] & \text{if } \phi = \phi_1 \wedge \phi_2 \\ x_{\phi_1} + x_{\phi_2} = x_\phi, x_{\phi_1} \leq y, \\ x_{\phi_2} \leq 1-y, \sigma(\phi_1), \sigma(\phi_2), \\ y \in \{0,1\}, x_\phi \in [0,1], \\ \text{where } y \text{ is a new binary variable} & \text{if } \phi = \phi_1 \vee \phi_2 \end{cases}.$$

In a similar way, we may determine $bsd(\mathcal{K},\phi)$ as

$$\begin{array}{l} \min -x. \text{ such that } x \in [0,1] \\ \quad x_\phi \geq x, \sigma(\phi), \\ \quad \text{for all } (\phi' \geq n) \in \mathcal{K}, x_{\phi'} \geq n, \sigma(\phi'), \\ \quad \text{for all } (\phi' \leq n) \in \mathcal{K}, x_{\neg\phi'} \geq 1-n, \sigma(\neg\phi') \end{array} \quad (3)$$

Note that we may verify whether $\mathcal{K}$ is satisfiable by checking if $bed(\mathcal{K},p) = 1$, where p is a new propositional letter not occurring in $\mathcal{K}$, and that under Łukasiewicz logic and Zadeh semantics we end up with a *bounded Mixed Integer Linear Program* (bMILP) optimization problem [15].

3 Fuzzy $\mathcal{ALC}$

Syntax. Consider an alphabet of *concepts names* (denoted A), *abstract roles names* (denoted R), *abstract individual names* (denoted a). From a First-Order Logic point of view, concepts may be seen as a formulae with one free variable (and, thus, may be seen as class descriptors), while roles as binary predicates (and, thus, may be used to describe properties of a class). *Concepts* (denoted C or D) of the language can be built inductively from atomic concepts (A), top concept $\top$, bottom concept $\bot$, abstract roles (R) as:

$C, D \rightarrow$	A	$\mid$	(atomic concept)
	$\top$	$\mid$	(top concept)
	$\bot$	$\mid$	(bottom concept)
	$C \sqcap D$	$\mid$	(concept conjunction)
	$C \sqcup D$	$\mid$	(concept disjunction)
	$\neg C$	$\mid$	(concept negation)
	$\forall R.C$	$\mid$	(universal quantification)
	$\exists R.C$		(existential quantification)

As illustrative purpose, Fig. 2 provides a First-Order Logic translation of $\mathcal{ALC}$ concepts and examples.

Syntax			*FOL*	*Example*
C, D	$\rightarrow$	$\top$	$\top(x)$	
		$\bot$	$\bot(x)$	
		A	$A(x)$	$Human$
		$C \sqcap D$	$C(x) \wedge D(x)$	$Human \sqcap Male$
		$C \sqcup D$	$C(x) \vee D(x)$	$Nice \sqcup Rich$
		$\neg C$	$\neg C(x)$	$\neg Meat$
		$\exists R.C$	$\exists y.R(x,y) \wedge C(y)$	$\exists has\_child.Blond$
		$\forall R.C$	$\forall y.R(x,y) \rightarrow C(y)$	$\forall has\_child.Human$

Figure 2: $\mathcal{ALC}$ concepts and First-Order Logic reading.

A *Fuzzy Knowledge Base* $\mathcal{K}$ comprises a fuzzy ABox $\mathcal{A}$ and a fuzzy TBox $\mathcal{T}$. A *fuzzy ABox* consists of a finite set of *fuzzy assertions* of one of the following types: a fuzzy *concept assertion* of the form $\langle a{:}C, n\rangle$ (with informal meaning, individual a is an instance of concept C with degree at least n) or a fuzzy *role assertion* of the form $\langle (a,b){:}R, n\rangle$ (the pair of individuals (a,b) is an instance of role R with degree at least n). In FOL, $\langle a{:}C, n\rangle$ may be seen as a fuzzy statement of the form $C(a) \geqslant n$, while $\langle (a,b){:}R, n\rangle$ may be seen as a fuzzy statement of the form $R(a,b) \geqslant n$.

In general, a *fuzzy TBox* $\mathcal{T}$ is a finite set of *fuzzy concept inclusion axioms* $\langle C \sqsubseteq D, n\rangle$, where C, D are concepts and $n \in (0,1]$. Informally, $\langle C \sqsubseteq D, n\rangle$ states that all instances of concept C are instances of concept D to degree n.$C = D$ is a shorthand for the two axioms $\langle C \sqsubseteq D, 1\rangle$ and $\langle D \sqsubseteq C, 1\rangle$.

However, for computational reasons, we will restrict TBoxes to be *acyclic*. That is, $\mathcal{T}$ is a finite set of *fuzzy concept inclusion axioms* $\langle A \sqsubseteq C, n\rangle$, and *concept definitions* $A = C$, where A is an atomic concept. Furthermore, we assume that $\mathcal{T}$ verifies two additional constraints: *(i)* there is no concept A such that it appears more than once on the left hand side of some axiom in $\mathcal{T}$. *(ii)* no cyclic definitions are present in $\mathcal{T}$. We will say that A *directly uses* primitive concept B in $\mathcal{T}$, if there is some axiom $\tau \in \mathcal{T}$ such that A is on the left hand side of τ and B occurs in the right hand side of τ. Let *uses* be the transitive closure of the relation *directly uses* in $\mathcal{T}$. $\mathcal{T}$ is cyclic iff there is A such that A uses A in $\mathcal{T}$.

[1] Note that this is an optimized version w.r.t. [13, 14].

ISBN: 978-989-95079-6-8

It is well known that such TBoxes can be eliminated through a finite (although it can create an exponential growth of the KB), *expansion process*, both in the crisp and in the fuzzy case [3]. Instead, we will use a extension to the fuzzy case of the *lazy expansion* technique [16], which has proved to be more useful in practice.

Semantics. From a semantics point of view, a fuzzy interpretation $\mathcal{I}$ is a pair $(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consisting of a non empty set $\Delta^{\mathcal{I}}$ (the interpretation domain) and a fuzzy interpretation function $\cdot^{\mathcal{I}}$ mapping: (i) an *abstract individual* a onto an element $a^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ such that if $a \neq b$ then $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ (Unique Name Assumption); (ii) a *concept* C onto a function $C^{\mathcal{I}} : \Delta^{\mathcal{I}} \to [0,1]$; (iii) an *abstract role* R onto a function $R^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \to [0,1]$.

Given arbitraries t-norm $\otimes$, t-conorm $\oplus$, negation function $\ominus$ and implication function $\Rightarrow$, the fuzzy interpretation function is extended to *complex concepts and roles* as shown in Fig. 3, where $C^{\mathcal{I}}$ denotes the membership function of the fuzzy concept C with respect to the fuzzy interpretation $\mathcal{I}$. $C^{\mathcal{I}}(x)$ gives us the degree of being the individual x an element of the fuzzy concept C under $\mathcal{I}$. Similarly, $R^{\mathcal{I}}$ denotes the membership function of the fuzzy role R with respect to $\mathcal{I}$. $R^{\mathcal{I}}(x,y)$ gives us the degree of being (x,y) an element of the fuzzy role R under $\mathcal{I}$. The fuzzy interpretation function is extended to *fuzzy axioms* in Fig. 3.

Concept		Semantics
$(\top)^{\mathcal{I}}(x)$	=	1
$(\bot)^{\mathcal{I}}(x)$	=	0
$(A)^{\mathcal{I}}(x)$	=	$A^{\mathcal{I}}(x)$
$(C \sqcap D)^{\mathcal{I}}(x)$	=	$C^{\mathcal{I}}(x) \otimes D^{\mathcal{I}}(x)$
$(C \sqcup D)^{\mathcal{I}}(x)$	=	$C^{\mathcal{I}}(x) \oplus D^{\mathcal{I}}(x)$
$(\neg C)^{\mathcal{I}}(x)$	=	$\ominus C^{\mathcal{I}}(x)$
$(\forall R.C)^{\mathcal{I}}(x)$	=	$\inf_{y \in \Delta^{\mathcal{I}}} \{R^{\mathcal{I}}(x,y) \Rightarrow C^{\mathcal{I}}(y)\}$
$(\exists R.C)^{\mathcal{I}}(x)$	=	$\sup_{y \in \Delta^{\mathcal{I}}} \{R^{\mathcal{I}}(x,y) \otimes C^{\mathcal{I}}(y)\}$
Axiom		**Semantics**
$(a:C)^{\mathcal{I}}$	=	$C^{\mathcal{I}}(a^{\mathcal{I}})$
$((a,b):R)^{\mathcal{I}}$	=	$R^{\mathcal{I}}(a^{\mathcal{I}}, b^{\mathcal{I}})$
$(C \sqsubseteq D)^{\mathcal{I}}$	=	$\inf_{x \in \Delta^{\mathcal{I}}} \{C^{\mathcal{I}}(x) \Rightarrow D^{\mathcal{I}}(x)\}$

Figure 3: Semantics of fuzzy concepts and axioms.

A fuzzy interpretation $\mathcal{I}$ *satisfies* (is a *model* of) a fuzzy statement $\langle \alpha, n \rangle$ iff $\alpha^{\mathcal{I}} \geq n$. The notions of logical consequence, best entailment degree and best satisfiability degree of α are as for Section 2. We additionally define the *Best Satisfiability Degree* [6] of a concept C w.r.t. a fuzzy KB $\mathcal{K}$ as $bsd(\mathcal{K}, C) = \sup_{\mathcal{I} \models \mathcal{K}} \sup_{x \in \Delta^{\mathcal{I}}} C^{\mathcal{I}}(x)$.

4 Fuzzy DLs with fuzzy truth values

So far, we have seen that fuzzy statements are of the form $\langle \alpha, n \rangle$, where $n \in [0,1]$. As next, we extend such statements to allow fuzzy truth values, such as "very true, almost true, almost false", in place of a "crisp" value $n \in [0,1]$. Essentially, we have fuzzy truth-qualified statements in which the truth is now a fuzzy set and, thus, we allow statements such as "Tina is young is very true".

As pointed out in [17] (see also [4, 6, 18]), there are many functions to specify fuzzy set membership degrees in fuzzy set theory and practice. The most frequently used are the trapezoidal (Fig. 4 (a)), the triangular (Fig. 4 (b)), the L-function (left-shoulder function, Fig. 4 (c)), the R-function (right-shoulder function, Fig. 4 (d)) and linear hedges (Fig. 4 (e), where $a = c/(c+1), b = 1/(c+1)$). We will call these functions (defined over $[0,1]$) *truth qualifiers* and allow them to be used to modify the degree of truth of a sentence. We will use

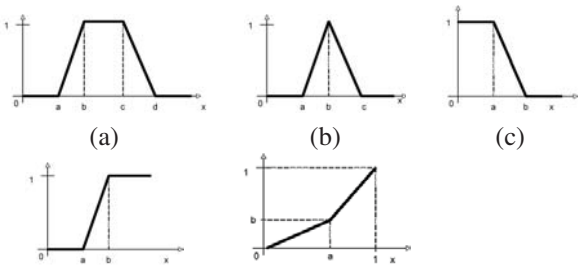


Figure 4: (a) Trapezoidal function; (b) Triangular function; (c) L-function; (d) R-function; (e) Linear function.

the abbreviations $\mathsf{trpz(a,b,c,d)}, \mathsf{tri(a,b,c)}, \mathsf{ls(a,b)}, \mathsf{rs(a,b)}$ and $\mathsf{ln(c)}$ to denote them, respectively.

Syntax. Let α be a concept assertion $a{:}C$, a role assertion $(a,b){:}R$ or a GCI $C \sqsubseteq D$, and $n \in [0,1]$. Let q be a truth qualifier, as in Fig. 4. Then we extend $\mathcal{ALC}$ axioms $\langle \alpha, n \rangle$ also to be of the form $\langle \alpha, q \rangle$, where a truth qualifier q may occur in place of a value $n \in [0,1]$. For instance, "Tina is young is very true" may be represented with $\langle \mathsf{tina{:}Young}, \mathsf{ln(4)} \rangle$. Similarly, we may represent an imprecision about the actual degree of truth, such as, "Tina is young is true to degree around 0.7", as an axioms of the form $\langle \mathsf{tina{:}Young}, \mathsf{tri(0.6, 0.7, 0.8)} \rangle$.

In the following, a knowledge base will be split into two parts, $\mathcal{K} = \mathcal{K}_U \cup \mathcal{K}_Q$ in which $\mathcal{K}_U$ contains only unqualified fuzzy axioms of type $\langle \alpha, n \rangle$, where $n \in [0,1]$, while $\mathcal{K}_Q$ contains only qualified fuzzy axioms of type $\langle \alpha, q \rangle$, where q is a truth qualifier. In FOL, we may see $\mathcal{K}$ as the formula

$$\Gamma_{\mathcal{K}} = \bigwedge_{\tau_i \in \mathcal{K}_U} \tau_i \wedge \bigwedge_{\tau_j \in \mathcal{K}_Q} \tau_j \,. \tag{4}$$

We will also use Γ_U for the left conjunct, while use Γ_Q for the right conjunct and, thus, $\Gamma_{\mathcal{K}} = \Gamma_U \wedge \Gamma_Q$.

Semantics. So far, given an interpretation $\mathcal{I}$, an axiom of the form $\langle \alpha, n \rangle$ is crisp under $\mathcal{I}$, in the sense that either $\mathcal{I}$ satisfies $\langle \alpha, n \rangle$ (i.e., $\alpha^{\mathcal{I}} \geq n$) or $\mathcal{I}$ does not satisfy $\langle \alpha, n \rangle$ (i.e., $\alpha^{\mathcal{I}} < n$). Once we move to truth qualified axioms, axioms are no longer true or false, but have a degree of truth depending on the qualifier. More precisely, the *degree of truth* of a fuzzy axiom τ of the form $\langle \alpha, q \rangle$ under $\mathcal{I}$, denoted $\mathcal{I}(\tau)$ is defined as the value $q(\alpha^{\mathcal{I}})$ (the application of the qualifier q to the truth value $\alpha^{\mathcal{I}}$). So, if "Tina is Young" is true to degree 0.9 under $\mathcal{I}$, then "Tina is young is very true" to degree 0.6 under $\mathcal{I}$. We may also extend this notion the axioms of the form $\langle \alpha, n \rangle$ by defining the *degree of truth* of a fuzzy axiom τ of the form $\langle \alpha, n \rangle$ to be $\mathcal{I}(\tau) = 1$ if $\alpha^{\mathcal{I}} \geq n$, otherwise $\mathcal{I}(\tau) = 0$, which is compatible with what we have defined so far.

Now, consider a knowledge base $\mathcal{K} = \mathcal{K}_U \cup \mathcal{K}_Q$, a fuzzy axiom τ and an interpretation $\mathcal{I}$. The *degree of truth* of $\mathcal{K}_x$ ($x \in \{U, Q\}$) under $\mathcal{I}$, denoted $\mathcal{I}(\mathcal{K}_x)$ is defined as

$$\mathcal{I}(\mathcal{K}_x) = \bigotimes_{\tau_i \in \mathcal{K}_x} \mathcal{I}(\tau_i) \,.$$

In FOL, this is the same as $\mathcal{I}(\Gamma_x)$. Note that if $\mathcal{I}$ is a model of $\mathcal{K}_U$ then $\mathcal{I}(\mathcal{K}_U) = 1$, else $\mathcal{I}(\mathcal{K}_U) = 0$. Furthermore, the *degree of entailment* of a τ w.r.t. $\mathcal{K}$ under $\mathcal{I}$, denoted $\mathcal{I}(\mathcal{K}, \tau)$, is defined as

$$\mathcal{I}(\mathcal{K}, \tau) = (\mathcal{I}(\mathcal{K}_U) \otimes \mathcal{I}(\mathcal{K}_Q)) \Rightarrow \mathcal{I}(\tau) \,. \tag{5}$$

ISBN: 978-989-95079-6-8

Essentially, the degree of entailment under $\mathcal{I}$ is the evaluation under $\mathcal{I}$ of the FOL implication $(\Gamma_U \wedge \Gamma_Q) \rightarrow \tau$, which is quite natural. Please note that if $\mathcal{I} \not\models \mathcal{K}_U$ then $\mathcal{I}(\mathcal{K}_U) = 0$ and, thus, $\mathcal{I}(\Gamma_\mathcal{K} \rightarrow \tau) = 1$, while if $\mathcal{I} \models \mathcal{K}_U$ then $\mathcal{I}(\Gamma_\mathcal{K} \rightarrow \tau) = \mathcal{I}(\Gamma_Q \rightarrow \tau)$. The *Best Entailment Degree* of a τ w.r.t. $\mathcal{K}$, denoted $bed(\mathcal{K}, \tau)$ is defined as $bed(\mathcal{K}, \tau) = \inf_\mathcal{I} \mathcal{I}(\mathcal{K}, \tau)$. By the previous observations, it easily follows that

$$bed(\mathcal{K}, \tau) = \inf_{\mathcal{I} \models \mathcal{K}_U} \mathcal{I}(\Gamma_Q \rightarrow \tau) \,, \tag{6}$$

where $\inf \emptyset = 1$. Finally, for an axiom α of the form $a{:}C$, $(a,b){:}R$ or a $C \sqsubseteq D$, the *Best Entailment Degree* of α w.r.t. $\mathcal{K}$, denoted $bed(\mathcal{K}, \alpha)$ is defined as

$$bed(\mathcal{K}, \alpha) = \sup\{n \mid 1 = bed(\mathcal{K}, \langle \alpha, n \rangle)\} \,, \tag{7}$$

and the *Best Satisfiability Degree* of a C w.r.t. $\mathcal{K}$, is

$$bsd(\mathcal{K}, C) = \sup_{\mathcal{I} \models \mathcal{K}_U} \sup_{x \in \Delta^\mathcal{I}} \mathcal{I}(\Gamma_Q) \Rightarrow C^\mathcal{I}(x) \,.$$

5 Reasoning

We next provide a reasoning algorithm for fuzzy $\mathcal{ALC}$ with truth qualifiers. To start with, we require a calculus for fuzzy $\mathcal{ALC}$ without truth qualifiers. Our algorithm is inspired by the one implemented within the FUZZYDL system [6] (which follows from the one presented in [17]). However, the presence of truth qualifiers requires some modifications. Furthermore, the version we present here for $\mathcal{ALC}$, introduces some optimizations that will require less "variables" and, thus, is expected to be more efficient. For the sake of this paper, *we provide a calculus under Łukasiewicz logic*, as in [17, 19]. A similar algorithm under product logic is expected to be more involved (see, e.g., [20, 21]).

W.l.o.g., we may assume that concepts are in *Negation Normal Form* (NNF), which is obtained by pushing in the usual manner negation on front of concept names only, by applying recursively the equivalences below.

$\neg\top \equiv \bot$	$\neg(C_1 \sqcup C_2) \equiv \neg C_1 \sqcap \neg C_2$	$\neg\exists R.C \equiv \forall R.\neg C$
$C \sqcap \bot \equiv \bot$	$C \sqcup \bot \equiv C$	$\neg\neg C \equiv C$
$\neg\bot \equiv \top$	$\neg(C_1 \sqcap C_2) \equiv \neg C_1 \sqcup \neg C_2$	$\neg\forall R.C \equiv \exists R.\neg C$
$C \sqcap \top \equiv C$	$C \sqcup \top \equiv \top$	

Now, our goal is to provide a terminating procedure determining $bed(\mathcal{K}, \alpha)$, where α is of the form $a{:}C$, $(a,b){:}R$ or a $C \sqsubseteq D$, and $bed(\mathcal{K}, \tau,)$, for truth qualified fuzzy axioms τ of the form $\langle \alpha, q \rangle$ (q is a truth qualifier).

5.1 Reasoning in $\mathcal{ALC}$ without truth qualifiers

Let's focus first on $bed(\mathcal{K}, \alpha)$, where $\mathcal{K}$ *does not* contain any truth qualifier, i.e. $\mathcal{K}_Q = \emptyset$. If α is a fuzzy role assertion of the form $(a,b){:}R$ then, in order to determine $bed(\mathcal{K}, (a,b){:}R)$, we may reduce it to the BED for concept assertions, as

$$bed(\mathcal{K}, (a,b){:}R) = bed(\mathcal{K} \cup \{\langle b{:}B, 1\rangle\}, a{:}\exists R.B) \,,$$

where B is a new concept (i.e., it does not occur in $\mathcal{K}$).

Consider $\mathcal{K} = \langle \mathcal{A}, \mathcal{T} \rangle$, where $\mathcal{T}$ is acyclic. The basic idea behind our reasoning algorithm is based on the observations and algorithm of Section 2. So, not surprisingly, in order to determine $bed(\mathcal{K}, \alpha)$, we will compute it as

$$bed(\mathcal{K}, \alpha) = \min x. \text{ such that } \mathcal{K} \cup \{(\alpha \leq x)\} \text{ satisfiable.} \tag{8}$$

Then by applying satisfiability preserving rules, we generate new *inequations* over $[0,1]$-valued variables. These inequations have to hold in order to respect the semantics of the DL constructors. Finally, in order to determine the BED, we *minimize* the original variable x such that all constraints are satisfied. More specifically,

$$\begin{array}{lcl} bed(\mathcal{K}, a{:}C) & = & \min x \text{ such that } \mathcal{K} \cup \{\langle a{:}\neg C, 1-x\rangle\} \text{ satisfiable} \\ bed(\mathcal{K}, C \sqsubseteq D) & = & \min x \text{ such that } \mathcal{K} \cup \{\langle b{:}C \sqcap \neg D, 1-x\rangle\} \text{ satisfiable} \\ bsd(\mathcal{K}, C) & = & \min -x \text{ such that } \mathcal{K} \cup \{\langle b{:}C, x\rangle\} \text{ satisfiable} \,, \end{array}$$

where b is a new individual. This means that determining the minimum degree of satisfiability [2] of a KB is the main reasoning issue to be addressed. Note also that we may determine if $\mathcal{K}$ has a model by e.g. checking whether $bed(\mathcal{K}, b{:}A) = 1$, where b and A do not occur in $\mathcal{K}$.

Like most of the tableaux algorithms, our algorithm works on *completion-forests* since an ABox might contain several individuals with arbitrary roles connecting them. A completion-forest $\mathcal{F}$ is a collection of trees whose distinguished roots are arbitrarily connected by edges. The forest has associated a set $\mathcal{C}_\mathcal{F}$ of constraints of the form $l \leq l'$, $l = l'$, $x_i \in [0,1], y_i \in \{0,1\}$, where l, l' are arithmetic expressions, on the variables occurring the node labels and edge labels.

Each node v is labelled with a set $\mathcal{L}(v)$ of concepts C. If $C \in \mathcal{L}(v)$ then we consider a variable $x_{v:C}$. The intuition here is that v is an instance of C to degree equal or greater than of the value of the variable $x_{v:C}$ in a minimal solution. Essentially, $x_{v:C}$ will hold the degree of truth of $v{:}C$. Each edge $\langle v, w\rangle$ is labelled with a set $\mathcal{L}(\langle v, w\rangle)$ of roles R. If $R \in \mathcal{L}(\langle v, w\rangle)$ then we consider a variable $x_{(v,w):R}$ (the intuition here is that $\langle v, w\rangle$ is an instance of R to degree equal or greater than the value of the variable $x_{(v,w):R}$ in a minimal solution (as for concept assertions, $x_{(v,w):R}$ will hold the degree of truth of $(v,w){:}R$). We will assume that there is a bijection between assertions α and variables x_α.

We are ready now to present our calculus. We first start with an initialization step, which builds the starting completion-forest. Then we apply to it a set of completion-forest transforming rules until no more rule can be applied. Finally, we solve the MILP problem associated to the set of constraints.

The algorithm initializes a forest $\mathcal{F}$ as follows. Consider $\mathcal{K}$: (i) $\mathcal{F}$ contains a root node a_i for each individual a_i occurring in $\mathcal{A}$; (ii) $\mathcal{F}$ contains an edge $\langle a, b\rangle$ for each fuzzy assertion $\langle (a,b){:}R, n\rangle \in \mathcal{A}$; (iii) for each fuzzy assertion $\langle a{:}C, n\rangle \in \mathcal{A}$, add both C to $\mathcal{L}(a)$ and $x_{a:C} \geq n$ to the set of constraints $\mathcal{C}_\mathcal{F}$; (iv) for each fuzzy assertion $\langle (a,b){:}R, n\rangle \in \mathcal{A}$, add both R to $\mathcal{L}(\langle a, b\rangle)$ and $x_{(a,b):R} \geq n$ to the set of constraints $\mathcal{C}_\mathcal{F}$; (v) for all introduced variables x_α, add $x_\alpha \in [0,1]$ to $\mathcal{C}_\mathcal{F}$.

$\mathcal{F}$ is then expanded by repeatedly applying the completion rules described below. The completion-forest is complete when none of the completion rules are applicable. Then, the bMILP problem on the set of constraints $\mathcal{C}_\mathcal{F}$ is solved.

As anticipated, we will use a extension to the fuzzy case of the *lazy expansion* technique in order to remove the axioms in $\mathcal{T}$. The basic idea is as follows (recall that there are only two types of fuzzy concept inclusions): given $\langle A \sqsubseteq C, n\rangle$, add C only to nodes with a label containing A, and given $\langle C \sqsubseteq A, 1\rangle$, add $\neg C$ only to nodes with a label containing $\neg A$.

[2] In the last equation, we use the fact that in our setting $\max x \; s.t. \; x \in S = \min -x \; s.t. \; x \in S$.

ISBN: 978-989-95079-6-8

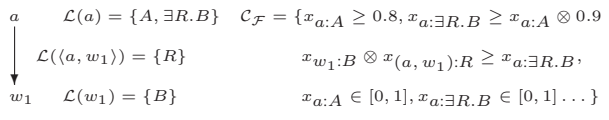


Figure 5: Complete forest for Example 1.

We assume a fixed rule application strategy as e.g., the order of rules below, such that the rule for $(\exists)$ is applied as last. Also, all expressions in node labels are processed according to the order they are introduced into $\mathcal{F}$. Note that we do not need a notion of *blocking* as $\mathcal{T}$ is acyclic.

Now we are ready to present the inference rules:

$(\bar{A})$. If $\neg A \in \mathcal{L}(v)$ then $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:A} = 1 - x_{v:\neg A}\} \cup \{x_{v:A} \in [0,1]\}$.

$(\bot)$. If $\bot \in \mathcal{L}(v)$ then $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:\bot} = 0\}$.

$(\top)$. If $\top \in \mathcal{L}(v)$ then $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:\top} = 1\}$.

$(\sqcap)$. If *(i)* $C \sqcap D \in \mathcal{L}(v)$, and *(ii)* not both $C \in \mathcal{L}(v)$ and $D \in \mathcal{L}(v)$ then add C and D to $\mathcal{L}(v)$, and $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:C} \otimes x_{v:D} = x_{a:C \sqcap D}\} \cup \{x_{v:C}, x_{v:D} \in [0,1]\}$.

$(\sqcup)$. If *(i)* $C \sqcup D \in \mathcal{L}(v)$, and *(ii)* not both $C \in \mathcal{L}(v)$ and $D \in \mathcal{L}(v)$ then add C and D to $\mathcal{L}(v)$, and $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:C} \oplus x_{v:D} = x_{a:C \sqcup D}\} \cup \{x_{v:C}, x_{v:D} \in [0,1]\}$.

$(\forall)$. If *(i)* $\forall R.C \in \mathcal{L}(v)$, $R \in \mathcal{L}(\langle v, w\rangle)$, and *(ii)* the rule has not been already applied to this pair then add C to $\mathcal{L}(w)$, and $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{w:C} \geq x_{a:\forall R.C} \otimes x_{(v,w):R}\} \cup \{x_{w:C} \in [0,1]\}$.

$(\sqsubseteq)$. If *(i)* $\langle A \sqsubseteq C, n\rangle \in \mathcal{T}$, *(ii)* $A \in \mathcal{L}(v)$, and *(iii)* v is a node to which this rule has not yet been applied then *(i)* append C to $\mathcal{L}(v)$, and *(ii)* $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:C} \geq x_{v:A} \otimes n\} \cup \{x_{v:C} \in [0,1]\}$.

$(\bar{\sqsubseteq})$. If *(i)* $\langle C \sqsubseteq A, 1\rangle \in \mathcal{T}$, *(ii)* $\neg A$, and *(iii)* v is a node to which this rule has not yet been applied then *(i)* append $\neg C$ to $\mathcal{L}(v)$, and *(ii)* $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{v:\neg A} \geq x_{v:\neg C}\} \cup \{x_{v:\neg C} \in [0,1]\}$.

$(\exists)$. If $\exists R.C \in \mathcal{L}(v)$ then create a new node w, add R to $\mathcal{L}(\langle v, w\rangle)$, add C to $\mathcal{L}(w)$, and $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{w:C} \otimes x_{(v,w):R} = x_{v:\exists R.C}\} \cup \{x_{(v,w):R}, x_{w:C} \in [0,1]\}$.

Note that in order to write the fuzzy operators, we may need to create some new control variables. For example, under Łukasiewicz t-norm, $x_1 \otimes x_2 \geq l$ can be written as $\{l \leq y, x_1 + x_2 \leq 1 + l, x_1 + x_2 - l \geq y, y_i \in \{0,1\}\}$. If $y = 0$, then $l = 0$ (it simulates the case where $x_1 + x_2 \leq 1$, and hence $x_1 \otimes x_2 = 0$), and if $y = 1$, then $l = x_1 + x_2 - 1$ [17, 19].

Example 1 illustrates the behaviour of the algorithm.

Example 1 *Consider* $\mathcal{K} = \{\langle A \sqsubseteq \exists R.B, 0.9\rangle, \langle a{:}A, 0.8\rangle\}$. *Let's show that* $\mathcal{K}$ *is satisfiable. The forest* $\mathcal{F}$ *is initialized with a root node* a*, labelled with* $\mathcal{L}(a) = \{A\}$*, and set of constraints* $\mathcal{C}_\mathcal{F} = \{x_{a:A} \geq 0.8\}$*. The we apply the* $(\sqsubseteq)$ *rule and add* A *and* $\exists R.B$ *to* $\mathcal{L}(a)$*, and add to the constraint set* $\{x_{a:\exists R.B} \geq x_{a:A} \otimes 0.9\} \cup \{x_{a:A}, x_{a:\exists R.B} \in [0,1]\}$*. As next, we apply the* $(\exists)$ *rule to node* a *and, thus, we create a new node* w_1*, labelled with* $\mathcal{L}(w_1) = \{B\}$*, and an edge* $\langle a, w_1\rangle$ *labelled with* $\mathcal{L}(\langle a, w_1\rangle) = \{R\}$*, and we update the constraint set with* $\mathcal{C}_\mathcal{F} = \mathcal{C}_\mathcal{F} \cup \{x_{w_1:B} \otimes x_{(a,w_1):R} = x_{a:\exists R.B}\} \cup \{x_{(a,w_1):R}, x_{w_1:B} \in [0,1]\}$*. Rule* $(\sqsubseteq)$ *is not applicable to node* w_1 *because* $A \notin \mathcal{L}(b)$*.*

The complete forest $\mathcal{F}$ *in Figure 5 shows the computation so far. It only remains to find a solution to the inequalities.* □

Note that there is a significative difference with other similar algorithms for fuzzy DLs combining tableau algorithms with optimization problems [17, 20, 9]. In those algorithms, every time a concept C appears in the list of expressions of a node v, a new variable x is created. Instead, we introduce a variable $x_{v:C}$ once, and reuse it the following times. This reduction in the number of generated variables is important because it makes the bMILP problem easier to solve. We have:

Proposition 1 *For each KB* $\mathcal{K}$*, the tableau algorithm terminates and computes* $bed(\mathcal{K}, \alpha)$ *and* $bsd(\mathcal{K}, C)$*.*

5.2 Reasoning in $\mathcal{ALC}$ with truth qualifiers

Let us now consider the case $\mathcal{K}_Q \neq \emptyset$. We next address the problem of determining $bed(\mathcal{K}, \tau)$ and $bed(\mathcal{K}, \alpha)$. From Eq. (6), we now that if $\mathcal{K}_U$ has no model then immediately $bed(\mathcal{K}, \tau) = 1$ and, thus, this case is not of particular interest. So, let us assume that $\mathcal{K}_U$ has a model. The case $\tau = \langle \alpha, n\rangle$, with $n \in [0,1]$, is also not of particular interest as it suffices to compute $m = bed(\mathcal{K}, \alpha)$ and check whether $m \geq n$. So, it remains to determine both $bed(\mathcal{K}, \tau)$, for a fuzzy truth qualified statement τ, and $bed(\mathcal{K}, \alpha)$, where α is of the form $a{:}C$, $(a,b){:}R$ or a $C \sqsubseteq D$. From what we have seen so far, it is immediate that (where a is new individual [3])

$$bed(\mathcal{K}, \tau) = \min x. \text{ such that } \mathcal{K}_U \cup \{(\Gamma_Q \rightarrow \tau) \leq x\} \text{ satisfiable.} \quad (9)$$

$$bsd(\mathcal{K}, C) = \min -x. \text{ such that } \mathcal{K}_U \cup \{(\Gamma_Q \rightarrow a{:}C) \geq x\} \text{ satisfiable.} \quad (10)$$

$$bed(\mathcal{K}, \alpha) = \min x. \text{ such that } \mathcal{K}_U \cup \{(\Gamma_Q \rightarrow (\alpha \leq x)) = 1\} \text{ satisfiable.} \quad (11)$$

Now, let us address first the problem (9). The calculus is of the same style as in the previous section. We have that

$$bed(\mathcal{K}, \tau) = \min x. \text{ such that } \mathcal{K}_U \cup \{(\Gamma_Q \wedge \neg\tau) \geq 1 - x\} \text{ satisfiable .}$$

Therefore, it suffices to provide rules for encoding $(\Gamma_Q \wedge \neg\tau) \geq 1 - x$ as a set of MILP constraints. This is obtained as follows. Consider $\Gamma_Q = \bigwedge_{\tau_j \in \mathcal{K}_Q} \tau_j$ and assume $\tau = \langle \alpha, q\rangle$. Consider variables $x_{QC}, x_{\Gamma_Q}, x_{\neg\tau}$, and x_{τ_j}, where x_{Γ_Q} will hold the degree of truth of Γ_Q, x_{τ_j} ($x_{\neg\tau}$) will hold the degree of truth of τ_j ($\neg\tau$) and x_{QC} will hold the degree of truth of $(\Gamma_Q \wedge \neg\tau)$. Then, to encode $(\Gamma_Q \wedge \neg\tau)) \geq 1 - x$ we need

$$\begin{array}{lcl} x_{QC} & \geq & 1 - x \\ x_{\Gamma_Q} \otimes x_{\neg\tau} & = & x_{QC} \\ x_{\Gamma_Q} & = & \bigotimes_{\tau_j \in \mathcal{K}_Q} x_{\tau_j} \\ x_{QC}, x_{\Gamma_Q}, x_{\tau_j}, x_{\neg\tau} & \in & [0,1]\,. \end{array} \quad (12)$$

and we add $\langle \alpha, x_\alpha\rangle$ to $\mathcal{K}_U$. We now have to connect correctly the value of x_{τ_j} ($x_{\neg\tau}$) to the degree of truth of $\tau_j = \langle \alpha_j, q_j\rangle$ ($\neg\tau$). To this end, we need the constraints

$$x_{\tau_j} = q_j(x_{\alpha_j}) \quad (13)$$

$$x_{\neg\tau} = 1 - q(x_\alpha) \quad (14)$$

and we add $\langle \alpha, x_\alpha\rangle$ to $\mathcal{K}_U$. Note that, e.g. in Eq. 13, the equation $x_{\tau_j} = q(x_\alpha)$ is MILP expressible as q is MILP expressible function (see, e.g. [17]). For instance, if q is $\mathsf{ls}(\mathsf{a}, \mathsf{b})$ then $x_{\tau_j} = q(x_\alpha)$ may be expressed as

$$\begin{array}{l} x_\alpha + (1-a)y_1 \leq 1, x_\alpha - ay_2 \geq 0, x_\alpha + (1-b)y_2 \leq 1, \\ (b-a)x_{\tau_j} \geq x_\alpha - a - (1-a)y_2, \\ (b-a)x_{\tau_j} \leq x_\alpha - a - (1+a)y_1 + (b-a)y_2, \\ x_\alpha - by_3 \geq 0, x_{\tau_j} + y_3 \leq 1, y_1 + y_2 + y_3 = 1, \\ x_\alpha \in [0,1], y_i \in \{0,1\}\}, \end{array}$$

[3] Concerning Equation (10), suppose the minimal value is $\bar{n}$. We will know then that for any interpretation $\mathcal{I}$ satisfying $\mathcal{K}_U$, it cannot be $\mathcal{I}(\Gamma_Q \rightarrow (\alpha < \bar{n})) = 1$, and, thus, $\mathcal{I}(\Gamma_Q \rightarrow (\alpha \geq \bar{n})) = 1)$ has to hold and, thus, $bed(\mathcal{K}, \langle \alpha, \bar{n}\rangle) = 1$ and $\bar{n}$ is tight.

ISBN: 978-989-95079-6-8

where y_i are new variables (Eq. 14 is managed similarly). Now, we may proceed as for Section 5.1, where we consider these additional constraints and in which expressions of the form $\langle \alpha, x_\alpha \rangle$ are handled as for the $\langle \alpha, n \rangle$ case, except that the value n is replaced with the variable x_α instead [4]. The following can be shown:

Proposition 2 *For each KB $\mathcal{K}$, the tableau algorithm terminates and computes $bed(\mathcal{K}, \tau)$.*

As next, let us address the problem of determining $bsd(\mathcal{K}, C)$. The way we proceed is pretty similar as for $bed(\mathcal{K}, \tau)$. Now, x_{QC} will hold the degree of truth of $(\Gamma_Q \rightarrow a{:}C)$. To encode $(\Gamma_Q \rightarrow a{:}C) \geq x$, we replace the constraints (12) with

$$\begin{array}{lcl} x_{QC} & \geq & x \\ x_{\Gamma_Q} \Rightarrow x_{a:C} & = & x_{QC} \\ x_{\Gamma_Q} & = & \bigotimes_{\tau_j \in \mathcal{K}_Q} x_{\tau_j} \\ x_{QC}, x_{\Gamma_Q}, x_{\tau_j} & \in & [0,1]\,, \end{array} \tag{15}$$

we add $\langle a{:}C, x_{a:C} \rangle$ to $\mathcal{K}_U$, and then we proceed as for determining $bed(\mathcal{K}, \tau)$. The following can be shown.

Proposition 3 *For each KB $\mathcal{K}$, the tableau algorithm terminates and computes $bsd(\mathcal{K}, C)$.*

Finally, let us address the problem of determining $bed(\mathcal{K}, \alpha)$. Since $bed(\mathcal{K}, a{:}C) = \min x.$ such that $\mathcal{K}_U \cup \{(\Gamma_Q \rightarrow (a{:}\neg C \geq 1 - x)) = 1\}$ satisfiable, $bed(\mathcal{K}, (a,b){:}R) = bed(\mathcal{K} \cup \{\langle b{:}B, 1\rangle\}, a{:}\exists R.B)$, and $bed(\mathcal{K}, C \sqsubseteq D) = \min x.$ such that $\mathcal{K}_U \cup \{(\Gamma_Q \rightarrow (a{:}C \sqcap \neg D \geq 1 - x)) = 1\}$ satisfiable, we may restrict our attention to show how to encode $(\Gamma_Q \rightarrow (a{:}E \geq 1 - x)) = 1$ in MILP. To this end, let $\tau = \langle a{:}E, 1 - x \rangle$. Then we replace the constraints (12) with (recall that τ is either true or false)

$$\begin{array}{lcl} x_{\Gamma_Q} & \leq & 1 - y \\ x_{a:E} & \geq & 1 - x - y \\ x_{\Gamma_Q} & = & \bigotimes_{\tau_j \in \mathcal{K}_Q} x_{\tau_j} \\ x_{\Gamma_Q}, x_{\tau_j}, x_{a:E} & \in & [0,1] \\ y & \in & \{0,1\}\,, \end{array} \tag{16}$$

we add $\langle a{:}E, x_{a:E} \rangle$ to $\mathcal{K}_U$, and then we proceed as for determining $bed(\mathcal{K}, \tau)$. The following can be shown.

Proposition 4 *For each KB $\mathcal{K}$, the tableau algorithm terminates and computes $bed(\mathcal{K}, \alpha)$.*

Note that we may safe some obvious variables and equations in the constraints derived so far, such as in Eq 12–16.

6 Conclusions & Outlook

We have addressed the problem of allowing to deal with fuzzy truth qualified statements with fuzzy DLs and, thus, allowing statements such as "Tina is young is *almost true*". We have provided syntax, semantics of a fuzzy DL with truth qualified axioms, and a calculus addressing the various reasoning problems presented under Łukasiewicz logic.

We have also provided a novel simplified calculus for fuzzy DLs without truth qualified axioms, which is closer to the usual tableau algorithms for DLs, and which reduces the number of generated variables, and hence the size of the optimization problem to be solved.

It would be possibly to unify both approaches by using truth qualifiers of the form $\mathtt{trpz}(n, n, 1, 1)$ for fuzzy DLs without truth qualified axioms. However, it is more efficient to deal with these two cases using different strategies.

Along the line of qualified statements, we think it is useful to further extend the language by allowing so-called *probability qualified* statements (cf. [10], page 222) and, thus, allowing statements such as "The probability of the temperature t at given place and time is around 35°C is likely", where both "around" and "likely" are fuzzy sets (with, e.g. a triangular membership function to model "around", and a right shoulder function to model "likely").

[4] Note that in Step 3 and 4 of the forest initialization, in case we consider a fuzzy assertion of the form $\langle \alpha, x_\alpha \rangle$, we may omit to add $x_\alpha \geq x_\alpha$ to the constraint set $\mathcal{C}_\mathcal{F}$.

References

[1] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[2] I. Horrocks, P. F. Patel-Schneider and F. van Harmelen. From $\mathcal{SHIQ}$ and RDF to OWL: The making of a web ontology language. *Journal of Web Semantics*, 1(1):7–26, 2003.

[3] U. Straccia. Reasoning within fuzzy Description Logics. *Journal of Artificial Intelligence Research*, 14:137–166, 2001.

[4] U. Straccia. A fuzzy Description Logic for the Semantic Web. In Elie Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, chapter 4, pages 73–90. Elsevier, 2006.

[5] U. Straccia. Managing uncertainty and vagueness in Description Logics, logic programs and Description Logic programs. In *Reasoning Web, 4th International Summer School, Tutorial Lectures*, Lecture Notes in Computer Science 5224, pages 54–103. Springer Verlag, 2008.

[6] F. Bobillo and U. Straccia. fuzzyDL: An expressive fuzzy Description Logic reasoner. In *Proceedings of the 2008 International Conference on Fuzzy Systems (FUZZ-IEEE 2008)*, pages 923–930. IEEE Computer Society, 2008.

[7] F. Bobillo, M. Delgado and J. Gómez-Romero. DeLorean: A reasoner for fuzzy OWL 1.1. In *Proc. of the 4th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW 2008)*, CEUR Workshop Proceedings 423, 2008.

[8] G. Stoilos, N. Simou, G. Stamou and S. Kollias. Uncertainty and the Semantic Web. *IEEE Intelligent Systems*, 21(5):84–87, 2006.

[9] U. Straccia and G. Visco. DL-Media: an ontology mediated multimedia information retrieval system. In *Proc. of the 4th Int. Workshop on Uncertainty Reasoning for the Semantic Web (URSW 2008)*, CEUR Workshop Proceedings 423, 2008.

[10] G. J. Klir and B. Yuan. *Fuzzy sets and fuzzy logic: theory and applications*. Prentice-Hall, Inc., 1995.

[11] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer, 1998.

[12] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[13] R. Hähnle. Advanced many-valued logics. In Dov M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd Edition*, volume 2. Kluwer, 2001.

[14] R. Hähnle. Many-valued logics and mixed integer programming. *Annals of Mathematics and Artificial Intelligence*, 3,4(12):231–264, 1994.

[15] H. Salkin and M. Kamlesh. *Foundations of Integer Programming*. North-Holland, 1988.

[16] F. Baader, E. Franconi, B. Hollunder, B. Nebel, and H. J. Profitlich. An empirical analysis of optimization techniques for terminological representation systems, or: Making KRIS get a move on. *Applied Artificial Intelligence* 4:109–132, 1994.

[17] U. Straccia. Description logics with fuzzy concrete domains. In Fahiem Bachus and Tommi Jaakkola, editors, *21st Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pages 559–567, 2005. AUAI Press.

[18] T. Lukasiewicz and U. Straccia. Managing uncertainty and vagueness in Description Logics for the Semantic Web. *Journal of Web Semantics*, 6:291–308, 2008.

[19] U. Straccia and F. Bobillo. Mixed integer programming, general concept inclusions and fuzzy Description Logics. *Mathware & Soft Computing*, 14(3):247–259, 2007.

[20] F. Bobillo and U. Straccia. A fuzzy Description Logic with product t-norm. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE-07)*, pages 652–657. IEEE Computer Society, 2007.

[21] F. Bobillo and U. Straccia. Fuzzy Description Logics with general t-norms and datatypes. In *Fuzzy Sets and Systems*. To appear.

ISBN: 978-989-95079-6-8

A new approach to fuzzy location of cephalometric landmarks in craniofacial superimposition

Oscar Ibáñez[1] Oscar Cordón[1] Sergio Damas[1] Sergio Guadarrama[1] José Santamaría[2]

[1]European Centre for Soft Computing. Mieres, Spain
[2]Computer Science Department, University of Jaén. Jaén, Spain
Email: {oscar.ibanez,oscar.cordon,sergio.damas,sergio.guadarrama}@softcomputing.es, jslopez@ujaen.es

Abstract— *Craniofacial superimposition is the second stage of a complex forensic technique that aims to identify a missing person from a photograph (or video shot) and the skull found. This specific task is devoted to find the most appropriate pose of the skull model to be projected onto the photograph. The process is guided by a number of landmarks identified both in the skull (craniometric landmarks) and in the face (cephalometric landmarks). In this contribution we extend our previous genetic algorithm-based approach to the problem by considering the uncertainty involved in the location of the cephalometric landmarks and its influence in the matching between these landmarks and the craniometric ones. The new proposal is tested over two instances of a real case solved by the Physical Anthropology lab at the University of Granada (Spain).*

Keywords— Craniofacial superimposition, fuzzy landmarks, genetic algorithms, image registration.

1 Introduction

Photographic supra-projection [1] is a forensic process where photographs or video shots of a missing person are compared with the skull that is found. By projecting both photographs on top of each other (or, even better, matching a scanned three-dimensional skull model against the face photo/video shot), the forensic anthropologist can try to establish whether that is the same person. To do so, an accurate 3D model of the skull is first demanded. Next, the matching of two sets of radiometric points (facial anthropometric landmarks in the subject photograph, and cranial anthropometric landmarks in the obtained skull model) is considered to guide the superimposition of the skull 3D model and the face photograph [1]. Then, a decision making stage starts by analyzing the different kinds of achieved matches between landmarks. Due to physiognomic characteristics, some of them will perfectly match, some will partially do so, and finally some others will not. After the whole process, the forensic expert must declare whether the analyzed skull corresponds to the missing person or not.

This procedure is very time consuming and there is no systematic methodology but every expert often applies a particular process. Hence, there is a strong interest in designing automatic methods to support the forensic anthropologist to put it into effect. In a previous proposal [2], we tackled the second stage of the process, i.e. the craniofacial superimposition, by means of evolutionary algorithms assuming no uncertainty was involved in the process. Next, in [3] we identified two sources of uncertainty in the problem. On the one hand, the *landmark matching uncertainty* (not yet modeled in any of our works) will refer to the imprecision involved in the matching of the landmarks that correspond to the two different objects: the face and the skull (as said, there is a clear partial matching situation). On the other hand, the *landmark location uncertainty* is related to the extremely difficult task of locating the landmarks [4] in an invariable place, with the accuracy required for this application. Indeed, every forensic anthropologist is prone to locate the landmarks in a slightly different position. The ambiguity may also arise from reasons like variation in shade distribution depending on light condition during photographing, unsuitable camera focusing, poor image quality, etc.

In our previous contribution [3], we tackled the landmark location uncertainty considering the cephalometric landmarks as rectangular zones of different sizes, instead of using crisp locations, taking inspiration from [5]. However, this is a too simple way to represent the underlying uncertainty since all the posible crisp points in the rectangle are equally likely to be the actual location, which is not so realistic.

In this contribution we will improve the previous approach by using fuzzy sets to model the uncertainty related to each cephalometric landmark location. In addition, we will also consider fuzzy distances to model the distance between each pair of craniometric and cephalometric landmarks. The resulting genetic fuzzy system [6] is tested on two superimposition problems derived from a real-world identification case solved by the Physical Anthropology lab at the University of Granada.

The structure of the contribution is as follows. Section 2 is devoted to review our previous proposal on evolutionary craniofacial superimposition. Our new proposal is described in Section 3. Then, Section 4 presents the experimental study. Finally, Section 5 collects some concluding remarks and future works.

2 Real-coded genetic algorithm for craniofacial superimposition

In [2] we first formulated this complex task in forensic identification as a numerical optimization problem related to the well known field of image registration (IR) [7]. Then we adapted three different evolutionary algorithms to solve it: two variants of a real-coded genetic algorithm (GA) and an evolution strategy.

A sensible way to design an automatic craniofacial superimposition procedure is through the use of a IR technique to properly align the 3D skull model and the 2D face photograph in a common coordinate frame. Taking a previous proposal [8] as a base, we modeled the required perspective transformation as a set of geometric operations involving translation, rota-

ISBN: 978-989-95079-6-8

tion, scaling and projection. That registration transformation is defined by 12 parameters which are encoded in a real-coded chromosome. The search for the best transformation parameters is guided by a number of pairings between craniometric and cephalometric landmarks located in the skull and the face, respectively.

Different definitions of the objective function were studied and the one achieving the best results was the mean error[1]:

$$ME = \frac{\sum_{i=1}^{N} ||f(cl^i) - fl^i||}{N} \quad (1)$$

where $||\cdot||$ is the 2D Euclidean distance, N is the number of considered landmarks (only four in Nickerson et al.'s proposal), cl^i corresponds to every 3D craniometric landmark, fl^i refers to every 2D facial landmark, f is the function which defines the geometric 3D-2D transformation, and $f(cl^i)$ represents the position of the transformed skull 3D landmark cl^i in the projection plane.

Among the three different evolutionary designs developed to solve the problem, a real-coded generational GA using tournament selection [9], Simulated Binary crossover (SBX) [10] and random mutation operators [11] was proposed and is the one considered in this contribution.

Later on, we introduced the uncertainty treatment in the location of cephalometric landmarks [3]. In that first approach, based on the work by Sinha [5], we considered the cephalometric landmarks as rectangular zones, instead of using crisp locations. The larger the rectangle that defines each landmark, the lower the contribution of the corresponding landmark to the fitness function. Finally, we calculated the Euclidean distances between craniometric and cephalometric landmarks, using the centroid of the rectangle related to the uncertain ones. Thus, once the centroid of the uncertain landmarks was considered, the problem of computing distances between a set of uncertain landmarks and a set of crisp ones became the problem of measuring a set of Euclidean distances between different pairs of crisp landmarks. That was just a first approach to model the location uncertainty, which did not take into account the inherent uncertainty involved when we are measuring distances between fuzzy and crisp points.

3 New proposal for the fuzzy location of cephalometric landmarks

The use of fuzzy landmarks in this contribution aims to face the landmark location uncertainty in the photograph of the missing person. Furthermore, the present approach also deals with the uncertainty involved when distances between crisp points and fuzzy ones are measured, as a consequence of the fact that craniometric landmarks are considered as crisp values[2]. In order to cope with the fact that each forensic expert could place each facial landmark in different positions in the 2D image, the higher the uncertainty related to a landmark, the broader the region where the forensic experts would locate that landmark in the photograph.

[1]Notice that, mean square error is not used because of its negative effect when image ranges are normalized in [0,1]

[2]As said, we have not considered the location uncertainty in craniometric landmarks due to the big resolution (one mm per pixel) and high quality of the 3D skull model.

In addition, by using fuzzy landmarks, those experts are able to locate a larger set of landmarks with the proper level of confidence (using fuzzy regions of different sizes). In contrast, following the classical crisp approach, they would only be able to mark the landmarks whose position they can determine precisely. Due to different reasons as the pose of the missing person, the quality of the image, or partial occlusions of landmark regions, it can be difficult to do so for many of the existing cephalometric landmarks. Hence, this approach presents an important advantage since the bigger the set of landmarks, the more complete the information will be to guide the genetic search of the best transformation to properly superimpose the skull on the photograph. In this way, it will also allow us to improve the performance of our previous proposal in some problematic instances with several specific poses (see Figures 4 and 8).

To ease the comprehension of our formulation to the reader, we review some required basic concepts from Fuzzy Sets Theory [12] as follows. For each $\alpha \in (0,1]$ the α-level set $\widetilde{A}_\alpha$ of a fuzzy set $\widetilde{A}, \mu_{\widetilde{A}} :\rightarrow [0,1]$, is $\widetilde{A}_\alpha = \{x \in X; \mu_{\widetilde{A}(x)} \geq \alpha\}$. Hence, the core $\widetilde{A}_1 = \{x \in X; \mu_{\widetilde{A}(x)} = 1\}$ of a fuzzy set is the subset of X whose elements have membership equal to 1. The support $\widetilde{A}_0$ is defined as the closure of the union of all its level sets, that is

$$\widetilde{A}_0 = \overline{\bigcup_{\alpha \in (0,1]} \widetilde{A}_\alpha}$$

3.1 Distance between a point and a set of points

Given a point x of $\mathbb{R}^n$ and a nonempty subset A of $\mathbb{R}^n$ we can define a distance $d : \mathbb{R}^n \times \mathbb{P}(\mathbb{R}^n) \rightarrow \mathbb{R}^+$ by:

$$d(x, A) = inf\{||x - a||; a \in A\}$$

for a certain norm $||.||$ on $\mathbb{R}^n$. Thus, $d(x, A) \geq 0$ and $d(x, A) = 0 \Rightarrow x \in A$.

3.2 Distance between a point and a fuzzy set of points

Now we can define the distance between a point x of $\mathbb{R}^n$ an a fuzzy set of points $A : \mathbb{R}^n \rightarrow [0,1]$ by:

$$d^*(x, \widetilde{A}) = \int_0^1 d(x, \widetilde{A}_\alpha) d\alpha$$

Lemma 3.1. *The distance from the point x to the fuzzy set $\widetilde{A}$ is lesser or equal than the distance to the core of $\widetilde{A}$ and greater or equal than the distance to the support of $\widetilde{A}_0$. That is,*

$$d(x, \widetilde{A}_1) \leq d^*(x, \widetilde{A}) \leq d(x, \widetilde{A}_0)$$

The proof is straight forward.

In case that we have discrete fuzzy set of points $\widetilde{A} = x_1/\alpha_1 + \cdots + x_m/\alpha_m$, the distance can be expressed by:

$$d^*(x, \widetilde{A}) = \frac{\sum_{i=1}^{m} d(x, \widetilde{A}_{\alpha_i}) * \alpha_i}{\sum_{i=1}^{m} \alpha_i}$$

Following the idea of metric spaces in [13] we will define a fuzzy landmark as a fuzzy convex set of points having a

ISBN: 978-989-95079-6-8

nonempty core and a bounded support. That is, all its α-levels are nonempty, bounded and convex sets.

In our case, since we are dealing with 2D photographs with an $x \times y$ resolution, we can define the fuzzy landmarks as 2D masks represented as a matrix m with $m_x \times m_y$ points. Each fuzzy landmark will have a different size depending on the imprecision on its localization but at least one pixel (i.e. crisp point related to a matrix cell) will have membership with degree one. An example of a 5×5 mask is:

	1	2	3	4	5
1	0.1	0.3	0.5	0.3	0.1
2	0.3	0.5	0.7	0.5	0.3
3	0.5	0.7	1	0.7	0.5
4	0.3	0.5	0.7	0.5	0.3
5	0.1	0.3	0.5	0.3	0.1

These masks are easily built starting from two triangular fuzzy sets $\widetilde{V}$ and $\widetilde{H}$ representing the approximate vertical and horizontal position of the landmark, thus becoming two-dimensional fuzzy sets. Each triangular fuzzy set $\widetilde{A}$, is defined by its center c and its offsets l, r as follows:

$$\widetilde{A}(x) = \begin{cases} 1 - \frac{|x-c|}{c-l}, & \text{if } l \leq x \leq c \\ 1 - \frac{|x-c|}{r-c}, & \text{if } c < x \leq r \\ 0, & \text{otherwise} \end{cases}$$

and the membership functions of the fuzzy landmarks are calculated using the product t-norm by:

$$\mu_{\widetilde{F}}(i,j) = \mu_{\widetilde{V}}(i) \cdot \mu_{\widetilde{H}}(j)$$

Now we can calculate the distance between a point (which will be the projection of the 3D craniometric landmark on the 2D face photo) and the fuzzy landmark (representing the position of the cephalometric landmark), as depicted in Figure 1. Note that the implemented distance between a point and a fuzzy set of points is quite similar to the one defined in [14].

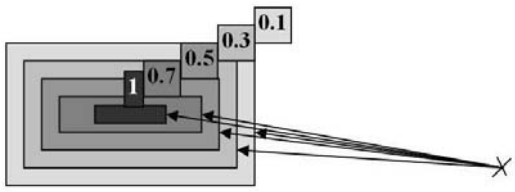


Figure 1: Distance between a point and a fuzzy point

If we denote as $d_i = d(x, \widetilde{F}_{\alpha_i})$ the distance from point x to the α-level set $\widetilde{F}_{\alpha_i})$, then the distance from the point to the fuzzy landmark $\widetilde{F}$, can be expressed by:

$$d^*(x, \widetilde{F}) = \frac{\sum_{i=1}^{m} d_i \cdot \alpha_i}{\sum_{i=1}^{m} \alpha_i}$$

In the example of Figure 1, taking $\{\alpha_1 = 0.1, \alpha_2 = 0.3, \alpha_3 = 0.5, \alpha_4 = 0.7, \alpha_5 = 1\}$ and assuming $\{d_1 = 4.5, d_2 = 5.4, d_3 = 6.3, d_4 = 7.3, d_5 = 9\}$, we calculate the distance as:

$$d^*(x, \widetilde{F}) = \frac{d_1 \cdot \alpha_1 + \cdots + d_5 \cdot \alpha_5}{\alpha_1 + \cdots + \alpha_5} = \frac{19.33}{2.6} = 7.43$$

Therefore, we have modified the previous definition of our genetic craniofacial superimposition techniques's fitness function (see section 2) as follows:

$$\textit{fuzzy ME} = \frac{\sum_{i=1}^{N} d^*(f(cl^i), \widetilde{F}^i)}{N} \quad (2)$$

where N is the number of considered landmarks, cl^i corresponds to every 3D craniometric landmark, f is the function which defines the geometric 3D-2D transformation, $f(cl^i)$ represents the position of the transformed skull 3D landmark cl^i in the projection plane, that is to say, a crisp point. $\widetilde{F}^i$ represents the fuzzy set points of each cephalometric landmark and finally, $d^*(f(C_i), \widetilde{F}^i)$ is the distance between a point and a fuzzy set points.

4 Experiments

After explaining the sources of uncertainty and our proposal to deal with the landmark location uncertainty, we will study its performance as follows. Section 4.1 presents the considered experimental design. Sections 4.2 and 4.3 are devoted to the analysis of one real-world case of study with two different photographs (two different poses).

4.1 Experimental design

Two different types of landmark sets for our case of study were provided by the forensic experts. The first type is the one classically used in the manual superimposition process. It is composed of crisp landmarks, those the forensic anthropologists can place in a single pixel. The second one is a set of fuzzy landmarks, that is to say, a region for each landmark. As said, in this second set, the forensic expert could place more landmarks than in the first one, due to the possibility of drawing bigger (in size) fuzzy sets of points. Notice that, in the former case, the forensic anthropologist would only identify those landmarks which are clearly located in a specific pixel without any doubt. According to different criteria concerning the features of both the considered landmarks and the photograph characteristics (pose, quality, etc.), forensic experts used areas of different size to determine the position of each landmark in that second set.

We compare the results of the genetic craniofacial superimposition based on crisp landmarks with those reached by using fuzzy location of cephalometric landmarks. In order to perform a significant and fair comparison between crisp and fuzzy approaches concerning the number of landmarks, we considered the following experimental design: two different sets of fuzzy landmarks are used, one with the same size (and, of course, the same specific landmarks) as the crisp set and another also including the additional landmarks identified thanks to the use of the fuzzy location approach.

The case study is a real-world one happened in Cádiz, Spain. The skull 3D model (327,641 points stored as x, y, z coordinates), see Figure 2, was acquired using Konica-Minolta 3D Lasserscanner VI-910. Two photographs[3] were provided by the family (see Figures 2 and 6). They were acquired at different moments and in different poses and conditions. Hence this case consists of two distinct superimposition

[3]Notice that we have processed the photographs to hide the subject identity following legal issues.

ISBN: 978-989-95079-6-8

problems. It was initially solved following a manual approach. The forensic experts tried to use both photographs but finally they were restricted to only one of them (pose 1 photograph), because they did not achieve a proper superimposition for the second due to the face pose.

Experiments consider the best performing GA parameter values in [2]: number of generations = 600, population size = 1,000, SBX η parameter = 1, mutation probability = 0.2, crossover probability = 0.9, and tournament size = 2. For all the tests, 30 runs of the GA are considered. Tables 1 and 2 show the minimum (m), maximum (M), mean (μ) and standard deviation (σ) of ME relative to the Euclidean distances between cephalometric landmarks and the transformed craniometric landmarks achieved by the GA. Notice that, in the case of crisp landmarks, the fitness function is directly defined as the "crisp" ME. However, when we use fuzzy landmarks the fitness function is not the ME, but a fuzzy distance between the transformed skull landmarks and the fuzzy regions (the average of the fuzzy distances). Owing to that, numerical results are not fully comparable. Nevertheless, we present the ME values in both cases as the only way to provide an error measurement.

4.2 Case study 1, pose 1

Figure 2 depicts this data set. In this first pose, the anthropologists identified nine crisp cephalometric landmarks. In addition, they marked these nine landmarks following a fuzzy approach, which also allowed them to identified a new set with five more (see Figure 3). Specially recall the large amount of fuzziness of the *vertex* landmark, located in the top part of the head. There is a strong uncertainty on its location due to the woman's hair and the forensic expert would have never trusted in that landmark in case a crisp location have to be defined. The corresponding craniometric landmarks were manually extracted from the skull 3D model.

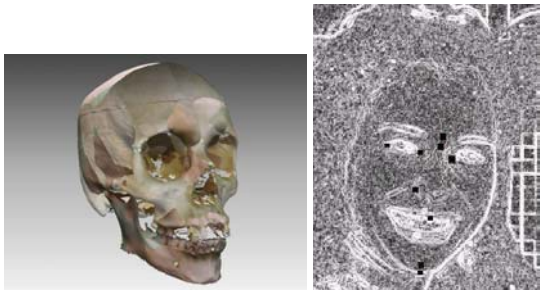


Figure 2: Case study 1, pose 1. 3D model of the skull (left) and photograph of the missing person with a set of nine crisp landmarks, represented by squares (right).

Table 1 presents the ME values for the obtained craniofacial superimpositions, distinguishing between crisp and fuzzy locations. It is important to remind that results are not fully comparable since the superimposition process using fuzzy landmarks does not minimize the ME but a different function (see Equations 1 and 2). Figures 4 and 5 present the best superimposition results of the crisp and the fuzzy approaches to allow a visual comparison.

On the one hand, referring to the numerical results, a very similar behavior is observed. Both approaches demonstrate a robust conduct. Notice that, in the case of the larger set of landmarks (fourteen fuzzy landmarks), Table 1 shows the ME corresponding just to the nine landmarks it shares with the smaller sets of landmarks. This higher ME value demonstrates that a larger error (i.e., distance between landmarks) value not necessarily means a worst superimposition (as we will see below). On the other hand, visual superimposition results show again a very similar behavior in both approaches for a smaller set of landmarks (Figure 4). These results are unacceptable for identification purposes. Finally, we can clearly identify the improvement in the superimposition when a larger set of fuzzy landmarks is used (Figure 5). Our forensic anthropologists confirmed the final superimposition result is good enough to be used in the final decision making stage of the photographic supra-projection process.

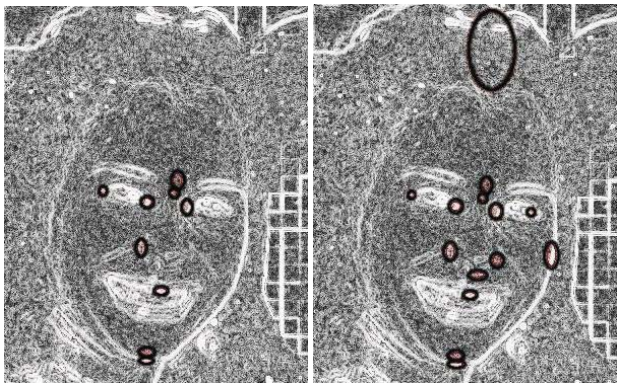


Figure 3: Case study 1, pose 1. Photograph of the missing person with two different sets of fuzzy landmarks: one of nine (left) and the other with fourteen (right) fuzzy landmarks. Landmarks are represented by ellipses.

Table 1: Case study 1, pose 1. Superimposition results.

Landmark set	ME			
	m	M	μ	σ
Nine crisp l.	0.0083	0.0086	0.0084	0.0000
Nine fuzzy l.	0.0084	0.0334	0.0095	0.0045
Fourteen fuzzy l. (ME over nine)	0.0122	0.0133	0.0124	0.0002

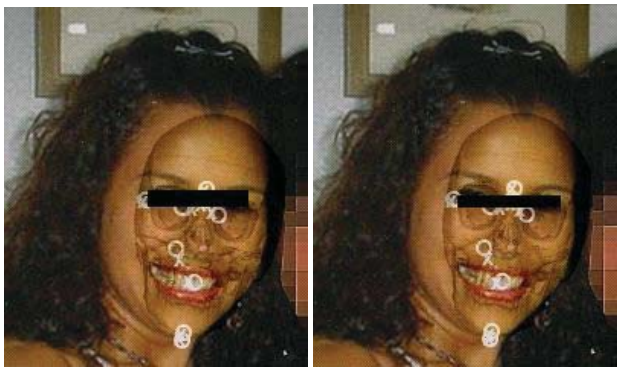


Figure 4: Case study 1, pose 1. Best superimposition results achieved using nine crisp (left) and nine fuzzy (right) landmarks, respectively.

ISBN: 978-989-95079-6-8

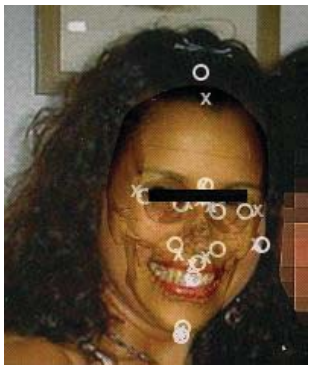


Figure 5: Case study 1, pose 1. Best superimposition result achieved using fourteen fuzzy landmarks.

4.3 Case study 1, pose 2

In this second pose, the anthropologists originally identified twelve crisp cephalometric landmarks following a crisp approach (see Figure 6). They also marked three more landmarks following a fuzzy approach. Figure 7 depicts both sets of fuzzy landmarks.

Figure 6: Case study 1, pose 2. Photograph of the missing person with the corresponding set of twelve crisp landmarks.

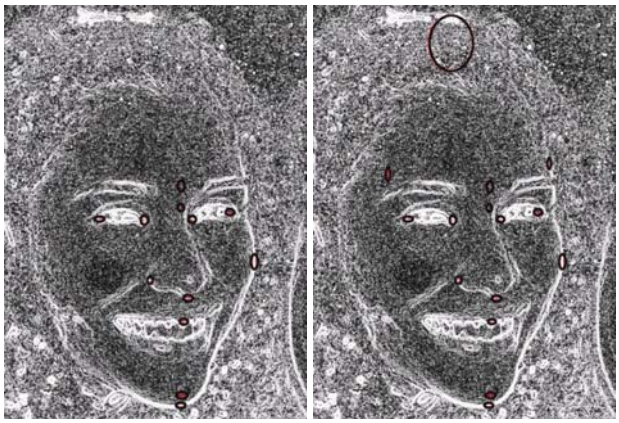


Figure 7: Case study 1, pose 2. Photograph of the missing person with two different sets of fuzzy landmarks, twelve (left) and fifteen (right).

Table 2 presents the ME figures for the obtained craniofacial superimpositions. In spite of the different fitness functions minimized, a very similar behavior is observed. As in the previous pose, the ME value of the larger set of landmarks was calculated just using the twelve landmarks that compose the smaller set. Both approaches demonstrate a robust conduct. Besides, it can be recognized how the fuzzy approach becomes more robust when considering a larger set of landmarks. On the other hand, Figure 8 presents the visual representations of the superimposition results which again show a very similar bad behavior in both approaches for a reduced set of landmarks. Finally, we can clearly see the improvement in the superimposition when a larger set of fuzzy landmarks is used (Figure 9, left). The final superimposition result is good enough to be used in the following decision making stage. In fact, this automatic superimposition is much better than that manually achieved by the forensic anthropologists (Figure 9, right), which was restricted by the computer vision tools they considered and by the lateral pose of the face photograph. The latter issues forced them to discard this photograph in the identification decision.

Table 2: Case study 1, pose 2. Superimposition results obtained using crisp and fuzzy locations

Landmark set	ME			
	m	M	μ	σ
Twelve crisp l.	0.0347	0.0354	0.0350	0.0002
Twelve fuzzy l.	0.0343	0.0613	0.0357	0.0048
Fifteen fuzzy l. (ME over twelve)	0.0382	0.0412	0.0391	0.0001

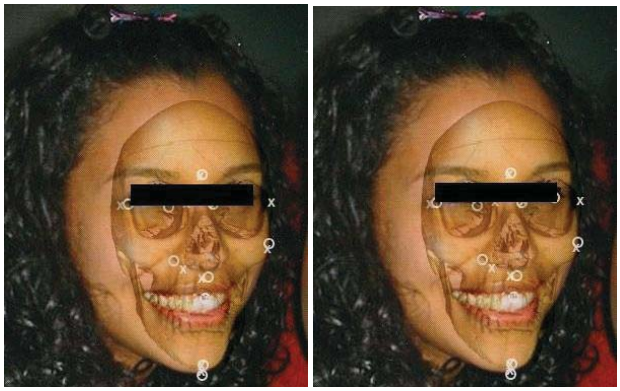


Figure 8: Case study 1, pose 2. Best superimposition results achieved using twelve crisp (left) and fuzzy (right) landmarks.

5 Concluding remarks and future works

We have proposed the use of fuzzy landmarks to tackle the uncertainty related to landmark location for a complex forensic identification task called craniofacial superimposition. We have used two different sets of landmarks provided by the forensic anthropologists for two photographs of the same identification case. We have compared the automatic superimpositions obtained by a GA considering crisp landmarks and those achieved using fuzzy landmarks. Promising results have been achieved following a fuzzy approach, showing that it is

ISBN: 978-989-95079-6-8

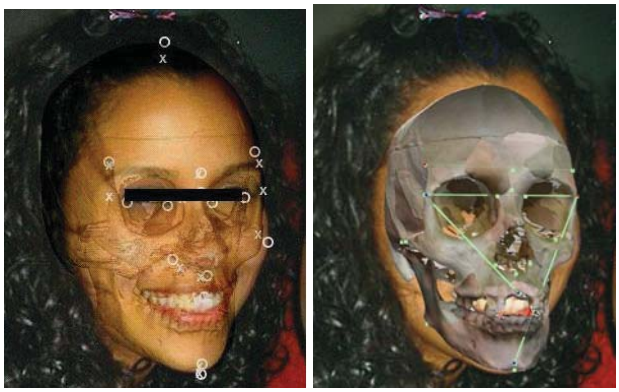


Figure 9: Case study 1, pose 2. Best superimposition result achieved using fifteen fuzzy landmarks (left) and best manual superimposition obtained by the forensic experts (right).

able to model the inherent uncertainty involved in the location of the landmarks. None of the achieved results following a crisp approach provided a better superimposition than the fuzzy approach. Thus, our proposal becomes more robust when tackling different locations of the landmarks.

By using fuzzy landmarks, forensic experts are able to place a larger landmark set with the proper level of confidence (using different sizes for the landmarks representing different levels of uncertainty in their actual location). In contrast, following the classical crisp approach, they are only able to mark those landmarks whose position they are able to determine precisely. As we expected, results show a much better performance of the GA when the number of landmarks is larger and that was only possible by using fuzzy location of landmarks.

As a drawback, the fuzzy approach implies more computational operations with the consequent increment in the run time required for the GA: from the 20-40 seconds per run using crisp landmarks to the 2-4 minutes using fuzzy landmarks. However, it is still a significantly short time if we compare it with the usual time needed by the forensic experts to perform a manual superimposition, around 24 hours.

Besides, these results need to be confirmed in a more extensive study, with a larger number of cases and using other different fuzzy distances [15]. As additional future work, we aim to determine the specific transformation parameters of the manual superimposition achieved by the forensic expert. We will use those parameters as the ground truth solution to measure the accuracy of the automatic superimposition processes. We will also make a poll between different forensic anthropologists in order to define the most appropriate shapes and sizes for the fuzzy landmarks in several face photographs corresponding to real-world previously solved cases. That poll will be also helpful in order to have different sets of crisp landmarks for comparison purposes. In addition, we are planning to introduce the landmark matching uncertainty (see Section 1). Finally, we aim to tackle the identification stage, i.e. the final decision making process, by using fuzzy logic, in order to assist the forensic expert in the final identification decision.

Acknowledgments

This work was partially supported by the Spain's Ministerio de Educación y Ciencia (ref. TIN2006-00829) and by the Andalusian Dpto. de Innovación, Ciencia y Empresa (ref. TIC1619), both including EDRF funding. We want to acknowledge all the team of the Physical Anthropology lab at the University of Granada (headed by Dr. Botella and Dr. Alemán) for their support during the data acquisition and validation processes. Dr. Guadarrama's work has been supported by the Spanish Department of Science and Innovation (MICINN) under program Juan de la Cierva JCI-2008-3531, and the European Social Fund.

References

[1] M. Y. Iscan. Introduction to techniques for photographic comparison. In M. Y. Iscan and R. Helmer, editors, *Forensic Analysis of the Skull*, pages 57–90. Wiley, 1993.

[2] O. Ibáñez, L. Ballerini, O. Cordón, S. Damas, and J. Santamaría. An experimental study on the applicability of evolutionary algorithms to craniofacial superimposition in forensic identification. *Information Sciences,* in press*, http://dx.doi.org/10.1016/j.ins.2008.12.029*, 2009.

[3] O. Ibáñez, O. Cordón, S. Damas, and J. Santamaría. Craniofacial superimposition based on genetic algorithms and fuzzy location of cephalometric landmarks. In *Hybrid artificial intelligence systems*, number 5271 in LNAI, pages 599–607, 2008.

[4] J. Richtsmeier, C. Paik, P. Elfert, T.M. Cole, and F. Dahlman. Precision, repeatability and validation of the localization of cranial landmarks using computed tomography scans. *The Cleft Palate-Craniofacial Journal*, 32(3):217–227, 1995.

[5] P. Sinha. A symmetry perceiving adaptive neural network and facial image recognition. *Forensic Science International*, 98(1-2):67–89, November 1998.

[6] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena. Ten years of genetic fuzzy systems: Current framework and new trends. *Fuzzy Sets and Systems*, 141(1):5–31, 2004.

[7] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.

[8] B. A. Nickerson, P. A. Fitzhorn, S. K. Koch, and M. Charney. A methodology for near-optimal computational superimposition of two-dimensional digital facial photographs and three-dimensional cranial surface meshes. *Journal of Forensic Sciences*, 36(2):480–500, March 1991.

[9] T. Blickle. Tournament selection. In T. Bäck, D. B. Fogel, and Z. Michalewicz, editors, *Handbook of Evolutionary Computation*, pages C2.3:1–C2.3:4. IOP Publishing Ltd and Oxford University Press, 1997.

[10] K. Deb and R. B. Agrawal. Simulated binary crossover for continuous search space. *Complex Systems*, 9:115–148, 1995.

[11] T. Bäck, D. B. Fogel, and Z. Michalewicz, editors. *Handbook of Evolutionary Computation*. IOP Publishing Ltd and Oxford University Press, 1997.

[12] G. J. Klir and B. Yuan. *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A. Zadeh*. World Scientific Publishing Co., Inc., 1996.

[13] P. Diamond and P. Kloeden. Metric topology of fuzzy numbers and fuzzy analysis. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy Sets*, The Handbooks of Fuzzy Sets, chapter 11, pages 583–637. Kluwer Academic, 2000.

[14] D. Dubois and H. Prade. On distance between fuzzy points and their use for pausible reasoning. In *International Conference on Systems, Man and Cybernetics*, pages 300–303, 1983.

[15] I. Bloch. On fuzzy distances and their use in image processing under imprecision. *Pattern Recognition*, 32:1873–1895, 1999.

ISBN: 978-989-95079-6-8

Incorporating Fuzziness in Spatial Susceptible-Infected Epidemic Models

Jan M. Baetens Bernard De Baets

KERMIT, Department of Applied Mathematics, Biometrics and Process Control
Ghent University, Coupure Links 653, 9000 Gent, Belgium
Email: jan.baetens@ugent.be, bernard.debaets@ugent.be

Abstract— *In this paper we propose a coupled-map lattice for modelling epidemic spread in a fuzzy setting. The presented model complies with the need for adequate modelling tools to describe and/or predict spatio-temporal phenomena, following the growing availability of spatio-temporal data. Furthermore, our approach does not rely on partial differential equations making it particularly suited to model epidemics in a fuzzy setting. It will be shown that the presented model allows to describe epidemic spread when the magnitude of the initial outbreak and/or the epidemic properties are only imprecisely known.*

Keywords— epidemic, discrete model, fuzzy initial condition, spatio-temporal dynamics

1 Introduction

Ordinary differential equations (ODEs) are widely used and well established to model various biological phenomena, as illustrated extensively in [1]. Partial differential equations (PDEs) are often resorted to if one is not merely interested in the process' temporal dynamics but also in the spatial patterns it generates [2]. Recently, several researchers have addressed the (numerical) solution of fuzzy ODEs (FODEs) [3, 4], endorsed by their potential importance in various scientific fields, for including imprecise information into well-established mathematical models (see [3, 5]). At present, the study of FODEs is still growing [6, 7], while thorough research on fuzzy PDEs is not yet carried out. Two distinct approaches have been developed within the theory of FODEs, differing only in whether they rely upon the notion of fuzzy derivatives or not [8]. This dichotomy and the immaturity of the theory of FODEs and FPDEs might hamper the widespread consideration of fuzziness within mathematical biology, despite often being faced with imprecise information. To overcome these barriers, we propose a coupled-map lattice (CML), developed by [9], to work with fuzzy initial conditions and/or parameters easily in a spatially explicit context. A short overview of fuzziness in discrete dynamical systems will be presented in the first section of this paper. In the second section we will introduce the CML which will be used to model spatio-temporal epidemic dynamics. In the third section we will deal with fuzzy initial conditions, while fuzzy parameters will be treated in the last section.

2 Fuzziness in discrete dynamical systems

A cellular automaton (CA), first introduced by von Neumann and Ulam as 'cellular spaces', and explored in detail by Wolfram [10, 11], is a mathematical model, discrete in all its senses, e.g. space is represented by an infinite lattice of cells, updates occur only at discrete time steps and the states can only take a finite number of values [10]. CA and CML are closely related, but in the latter states can take arbitrary values [12]. Although the first notion of a fuzzy CA dates back to the late 60s [13], literature on fuzzy CA is considered insufficient [14]. Fuzzy automata were first defined by Wee and Fu [13], but several alternative definitions have been proposed [14, 15]. Essentially, every state is attributed a membership value in a fuzzy CA [15], as such relaxing the condition of merely allowing discrete states. Recently, a review on fuzzy automata has been published [14], following the increasing number of articles published on the topic (see [16, 17, 18]). Despite the definition given in [13], fuzzy CA are often simply understood as spatial extensions of fuzzy rule-based models such as Mamdami-Assilian [19, 20] or Takagi-Sugeno models [21], e.g. CA in which the local transition function is a fuzzy rule-based model (see [22, 23, 24, 25, 26, 27]). To our knowledge, the notion of fuzziness in CML models has not been addressed yet.

3 Spatially explicit modelling of epidemics

3.1 The model

Several authors, including Kaneko [12] and Wolfram [11], have argued that CML and CA provide a suitable framework to deal with spatio-temporal dynamics. This is illustrated by the rich variety of such models that has been developed during the last decade for describing various spatial biological phenomena such as epidemics [28, 29, 30], population dynamics [31, 32], tumor growth [33, 34, 35], biofilm development [36, 37] and much other phenomena [38, 39].

Recently, Baetens and De Baets [9] proposed a generalized CML for modelling various biological phenomena, traditionally described by means of PDEs. Their model is general in two senses. Firstly, exploiting graph notations makes it independent from the spatial subdivisions used to discretize the space domain and secondly, it can serve as a basis for modelling various biological phenomena. With regard to an epidemic sweeping through a region which is subdivided into irregular polygons, and involving only non-reproducing susceptible and infected individuals, it can be written in a simplified form as

$$\begin{cases} S_j^{t+1} = S_j^t - S_j^t \displaystyle\sum_{P_k \in N(P_j)} w_{jk}\, G\left(\mathbf{I}_j, d_{jk}\right)\, I_k^t \\ I_j^{t+1} = I_j^t + S_j^t \displaystyle\sum_{P_k \in N(P_j)} w_{jk}\, G\left(\mathbf{I}_j, d_{jk}\right)\, I_k^t \end{cases} \tag{1}$$

where S_j^t, resp. I_j^t, represent the fraction of susceptible, resp. infected individuals within polygon P_j at the t-th time step

ISBN: 978-989-95079-6-8

and G is a function describing the effect of landscape and connectivity characteristics on the epidemic spread, d_{jk} is the distance measured on a graph between the vertices v_j and v_k, representing polygons P_j and P_k, $N(P_j)$ is the set of polygons of which the infected inhabitants can affect I_j^{t+1}, e.g. the neighbourhood of P_j defined as

$$N(P_j) = \{P_k \in P \mid d_{ij} \leq \epsilon\} \,, \tag{2}$$

where ϵ is the neighbourhood radius. Further, w_{jk} is a weighting factor, weighing the influence of every $P_k \in N(P_j)$ in the determination of S_j^{t+1}, and $\mathbf{I}_j$ contains information about P_j. This model may be regarded as a discrete analogue of the PDE-based SI-model (see [2]). Assuming $S_j^t + I_j^t = 1$, at all t, and for all P_j, we only have to keep track of one of the system's equations in order to follow its dynamics. From (1) it follows clearly that the system has a single fixed point $\left(S_j^*, I_j^*\right) = (0, 1)$.

Within the framework of this paper, only a rectangular lattice consisting of 101×101 equally-sized cells will be considered, since we rather want to focus on the fuzziness in (1) than on its general applicability as discussed in [9]. Moreover, we will assume that the region is spatially homogeneous, e.g. G does not depend on $\mathbf{I}_j$, reducing (1) to

$$I_j^{t+1} = I_j^t + S_j^t \sum_{P_k \in N(P_j)} w_{jk}\, H\left(d_{jk}\right)\, I_k^t \,, \tag{3}$$

where we introduced the function H, only dependent on d_{jk}, to distinguish from G, potentially dependent on both d_{jk} and $\mathbf{I}_j$. Further, we will define

$$H\left(d_{jk}\right) = \begin{cases} \nu_0, \text{ if } d_{jk} = 0, \\ \nu_1, \text{ if } d_{jk} = 1, \\ \quad \vdots \\ \nu_\epsilon, \text{ if } d_{jk} = \epsilon, \end{cases} \tag{4}$$

where ϵ is the neighbourhood radius and ν_ϵ is a measure for the epidemic's virulence, in such a way that

$$\sum_{P_k \in N(P_j)} w_{jk}\, H\left(d_{jk}\right) \leq 1, \quad \forall\, t,\, P_j \,, \tag{5}$$

assuring that $0 \leq I_j^t \leq 1$, at all t, and for all P_j. For this paper, let $\epsilon = 1$, $w_{jk} = \frac{1}{8}$ for all j, k and $j \neq k$ and $w_{jk} = 1$ if $j = k$. Consequently, P_j's eight nearest neighbours influence I_j^{t+1} to the same degree.

3.2 Fuzziness in the proposed model

In the remainder of this paper S_j^t and I_j^t are considered fuzzy intervals in $[0, 1]$, while the notations $\left[S_j^t\right]^\alpha$ and $\left[I_j^t\right]^\alpha$ will refer to their respective α-cuts defined by $\left[S_j^t\right]^\alpha = \left\{s \in [0,1] \mid S_j^t(s) \geq \alpha\right\}$ if $\alpha > 0$ and $\left[S_j^t\right]^0 = \text{cl}\left\{s \in [0,1] \mid S_j^t(s) > 0\right\}$ (the closure of the support) if $\alpha = 0$, and analogously for $\left[I_j^t\right]^\alpha$. In consequence, we can write $\left[S_j^t\right]^\alpha = \left[s_{j,1}^t(\alpha), s_{j,2}^t(\alpha)\right]$ and $\left[I_j^t\right]^\alpha = \left[i_{j,1}^t(\alpha), i_{j,2}^t(\alpha)\right]$. A triangular fuzzy interval A will be denoted as $A \equiv (l, u, r)$ where $[A]^0 = [l, r]$ and $[A]^1 = \{u\}$. Two sources of fuzziness in (3) will be examined more closely. A first concerns the initial conditions I_k^0 necessary to iteratively solve (3). As indicated by Beale [40], spatial epidemiological data are becoming increasingly available, providing the means to deduce I_k^0. Nonetheless, one has to be aware of the imprecision in the outbreak data which can be modeled with a fuzzy set approach. Analogously, the parameters in (4) might only be known imprecisely, and will be regarded as the second source of fuzziness. Since the mathematical functions in (3) are restricted to the basic arithmetic operations and involve no derivatives, it becomes relatively straightforward to obtain the fuzzy system's spatio-temporal dynamics. Nonetheless, one has to bear in mind the coupling between S_j^t and I_j^t through the condition $S_j^t + I_j^t = 1$, at all t, and for all P_j. As such, S_j^t and I_j^t cannot take values independently of each other and may be called interactive fuzzy variables [41]. Their joint possibility distribution C_j^t is not given by $S_j^t \times I_j^t$ as illustrated in Fig. 1(a), but can be defined as [42]

$$\left[C_j^t\right]^\alpha = \left\{l\left(s_{j,1}^t(\alpha), i_{j,2}^t(\alpha)\right) + (1-l)\left(s_{j,2}^t(\alpha), i_{j,1}^t(\alpha)\right) \mid l \in [0,1]\right\} \,, \tag{6}$$

for all $\alpha \in [0, 1]$, and is depicted in Fig. 1(b). The extension $f(S_j^t, I_j^t)$ is defined by [43]

$$f(S_j^t, I_j^t)\,(z) = \sup_{f(x,y)=z} C_j^t\,(x, y) \,. \tag{7}$$

Following the theory outlined in [44], S_j^t and I_j^t may be called completely negatively 'correlated'. In view of the existing interactivity we can write

$$\left[S_j^t + I_j^t\right]^\alpha = \left[s_{j,1}^t(\alpha) + i_{j,2}^t(\alpha), s_{j,2}^t(\alpha) + i_{j,1}^t(\alpha)\right] \,, \tag{8}$$

and

$$\left[S_j^t \cdot I_j^t\right]^\alpha = \left[\min\left(s_{j,2}^t(\alpha) \cdot i_{j,1}^t(\alpha), s_{j,1}^t(\alpha) \cdot i_{j,2}^t(\alpha)\right), \max\left(s_{j,2}^t(\alpha) \cdot i_{j,1}^t(\alpha), s_{j,1}^t(\alpha) \cdot i_{j,2}^t(\alpha)\right)\right] \,. \tag{9}$$

All simulations were performed in Mathematica 6.0 (Wolfram Research, Inc.) on a desktop PC with an Intel Dual Core 1.86 GHz processor. Although a quadratic lattice was used in this paper, the described simulations can easily be performed when irregular spatial subdivisions would be utilized.

4 Fuzzy initial conditions

In this section, we assume $\nu_0 = 0.5$ and $\nu_1 = 0.5$, meaning that the spread of an infection in a polygon P_j can be equally attributed to infected individuals living in P_j as to infected individuals residing in P_j's neighbourhood. First, we will consider the situation in which a disease outbreak was recorded in the center polygon P_c of the grid, but only imprecise information was available on the proportion of initially infected individuals. As such, we adopted the following initial condition

$$\begin{cases} I_j^0 = (0.2, 0.3, 0.4) \,, \text{ if } j = c \\ I_j^0 = 0 \qquad\qquad\quad \,, \text{ else.} \end{cases} \tag{10}$$

Figure 2 illustrates the CML evolution during the first time interval: starting from the initial condition (Fig. 2(a), 2(d)),

ISBN: 978-989-95079-6-8

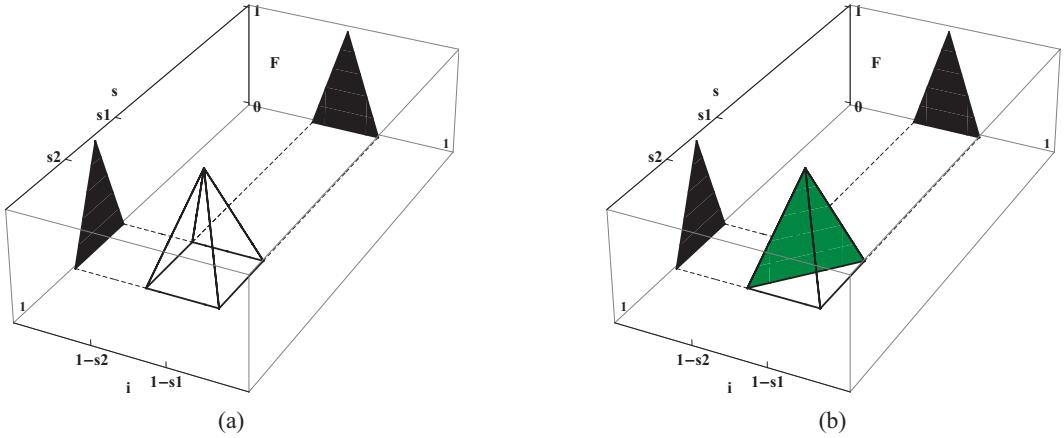


Figure 1: Non-interactive (a) and interactive (b) joint possibility distributions.

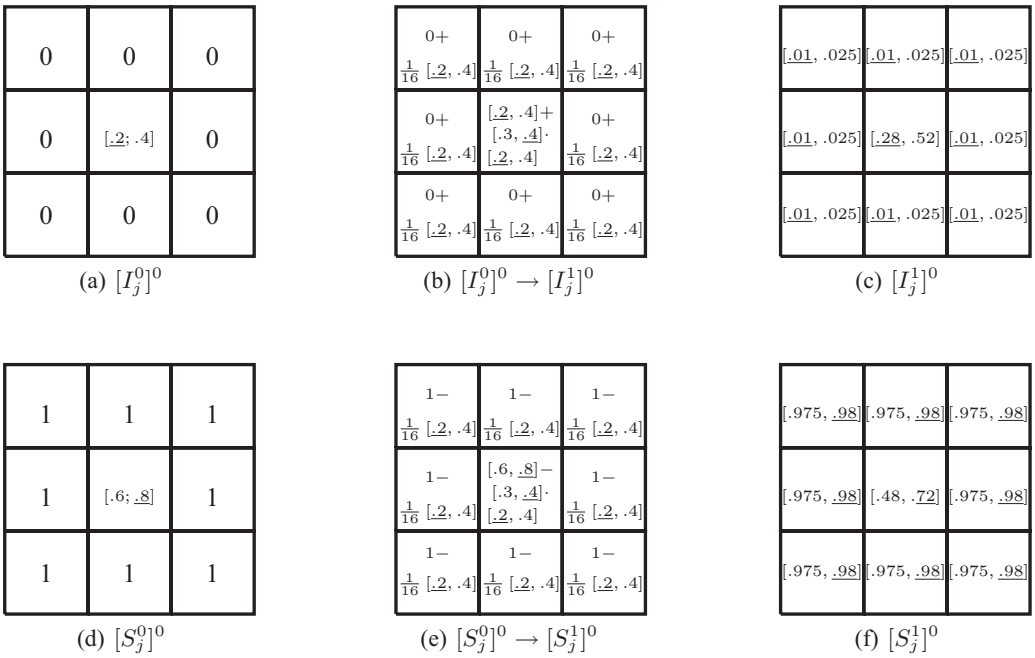


Figure 2: Evolution from $\left[I_j^0\right]^0$ to $\left[I_j^1\right]^0$ (a, b, c) and from $\left[S_j^0\right]^0$ to $\left[S_j^1\right]^0$ (e, d, f).

the lattice sites are updated according to (3), yielding the updated fuzzy states (Fig. 2(c), 2(f)). Since the neighbourhood was restricted to the nearest neighbours, it suffices to plot only the center polygon and its nearest neighbours to view all state changes during the first time interval. Underlining was used in order to visualize how the interaction between S_j^t and I_j^t was taken into account. For instance, $\left[I_j^1\right]^0$ was calculated as

$$\begin{aligned}\left[I_j^1\right]^0 &= \left[I_j^0\right]^0 + \left[s_{j,2}^0(0)\cdot\tfrac{1}{2}\,i_{j,1}^0(0),\; s_{j,1}^0(0)\cdot\tfrac{1}{2}\,i_{j,2}^0(0)\right]\\ &= [0.2, 0.4] + [0.8\cdot 0.1, 0.6\cdot 0.2]\\ &= [0.28, 0.52]\end{aligned}$$

Figure 3 shows the proportion of infected individuals after two, five, ten and fifteen time steps following the epidemic outbreak. For reasons of clarity, we limited the depicted spatial extent. The fill color of the polygons corresponds to the value of $\left[I_j^t\right]^1$, while the vertical axis indicates the membership value of all $i \in [0, 1]$. For each polygon, the fuzzy interval $I_j^t(i)$ is depicted at the center of the Y-direction, with the i-values marked along the X-direction, as indicated by the additional axis drawn at the back of the plots depicted in Fig. 3(a). This figure shows clearly that, although the epidemic was caused by an imprecise number of infected individuals residing in the center polygon, I_c^5's support was narrower than that of I_c^0, meaning that the imprecision of the model prediction decreased in this polygon during the considered time interval. This suggests that I_j^t tended to a crisp number as $t \rightarrow \infty$ which was verified experimentally by running the model for 150 time steps. This tendency could be expected since $(S_c^*, I_c^*) = (0, 1)$ is an asymptotic fixed point of the CML. Further, it is shown in this figure that the imprecision, initially only present in the outbreak data from the center polygon, propagated radially as a traveling wavefront giving rise to fuzziness in the model predictions at other locations. Though,

ISBN: 978-989-95079-6-8

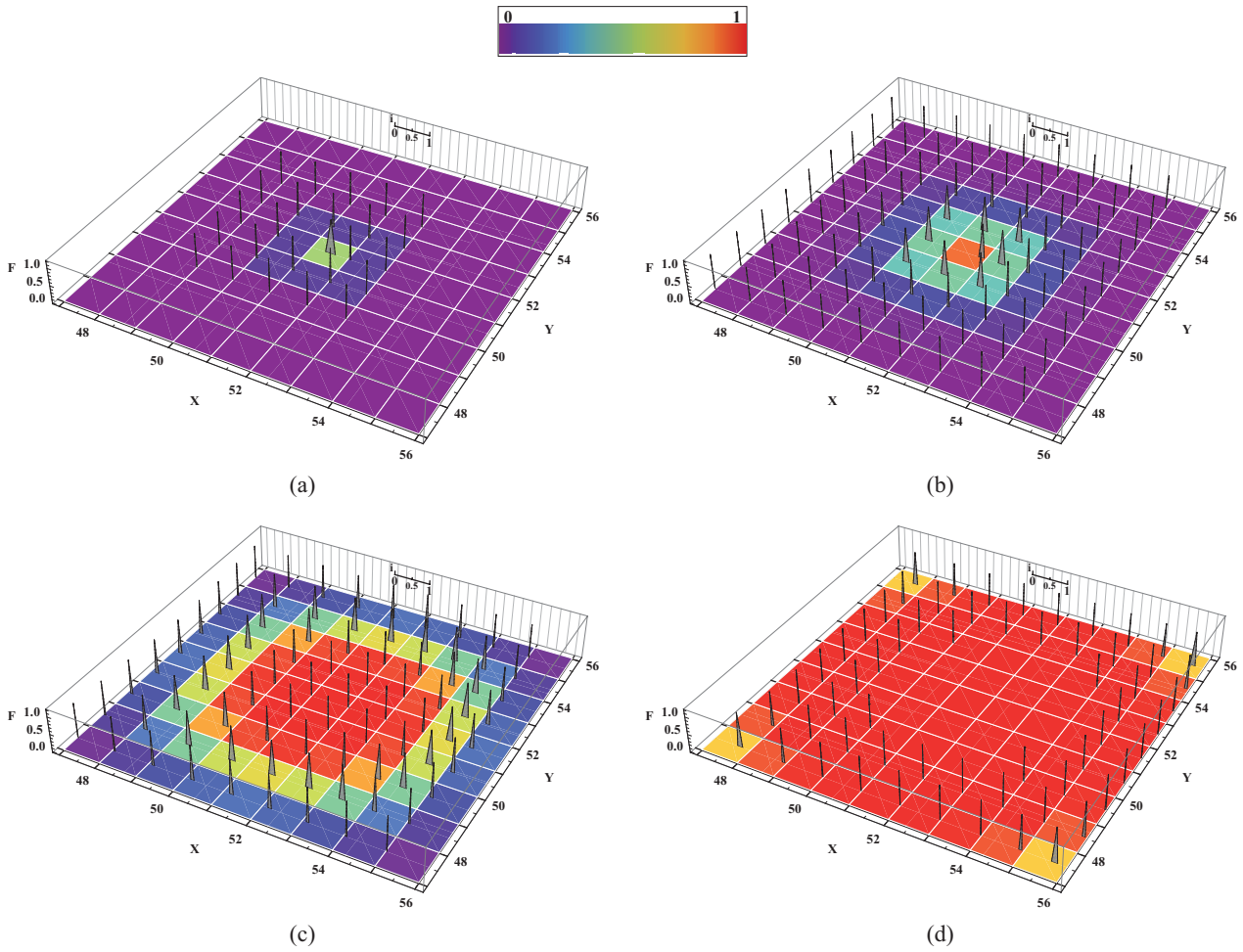


Figure 3: CML state two (a), five (b), ten (c) and fifteen (d) time steps after an outbreak was recorded in the center polygon.

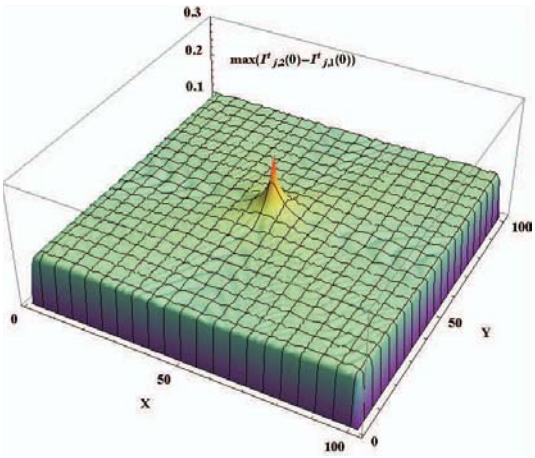


Figure 4: Maximum $[I_c^t]^0$ appeared during the first 125 time steps after an outbreak was recorded in the center polygon of the grid.

the degree of imprecision, expressed as the maximum width of I_j^t's support for all t, decreased asymptotically to 0.1 as d_{cj} increased (Fig. 4).

Also the situation of simultaneous epidemic outbreaks at different locations was studied more closely. For this setup, it can be expected that the imprecision linked to the outbreak data in each of the initially infected polygons might 'interact' when the epidemic waves meet. This was confirmed by the small roughnesses between the two peaks in Fig. 5 showing the maximum width of I_j^t's support during the first 50 time steps after an epidemic was initiated in two polygons which are at distance 14 from one another. Similar to the above-described situation, I_j^t tended to a crisp number as $t \to \infty$ and the CML evolved towards its fixed point.

5 Uncertain epidemic properties

In this section we again consider the model given by (3) with fuzzy initial conditions (10), but in addition we assume that ν_0 was not known accurately and could be described as a triangular fuzzy interval $\nu_0 = (0.2, 0.35, 0.5)$. Comparing Figs. 6 and 3 one clearly sees that I_j^t's support was wider when both the initial condition and ν_0 were only imprecisely known. As the wavefront propagated, the maximum width of the support spanned the entire unit interval in polygons further from P_c due to the successive non-interactive multiplication of ν_0 and $I_{P_i}^t$. Yet, I_j^t tended to a crisp number as $t \to \infty$ since the CML evolved to its fixed point.

6 Conclusions

In this paper it was shown that imprecise information, described using a fuzzy set approach, can be used easily within a spatially explicit model, based upon the coupled-map lattice

ISBN: 978-989-95079-6-8

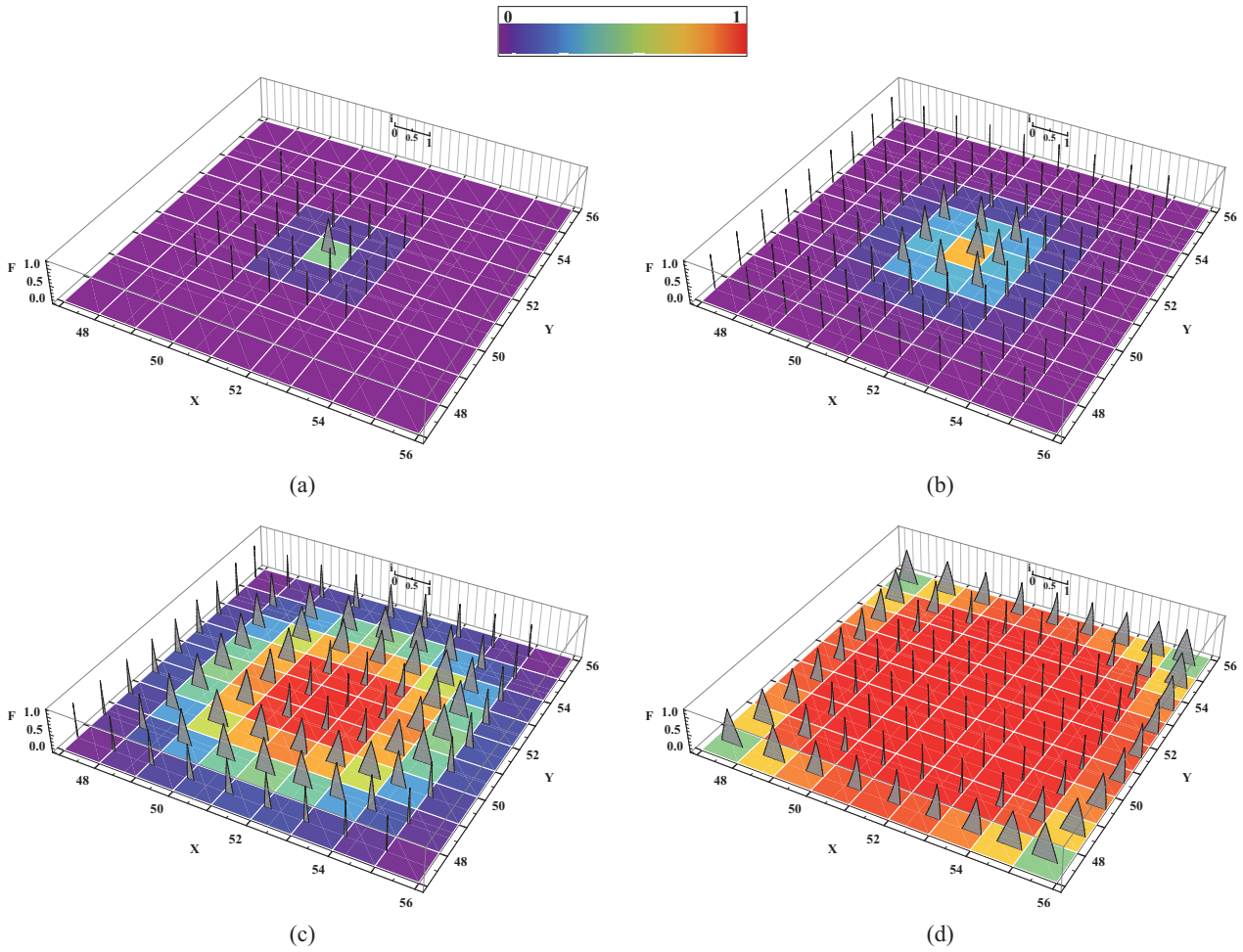


Figure 6: CML state two (a), five (b), ten (c) and fifteen (d) time steps after an outbreak was recorded in the center polygon.

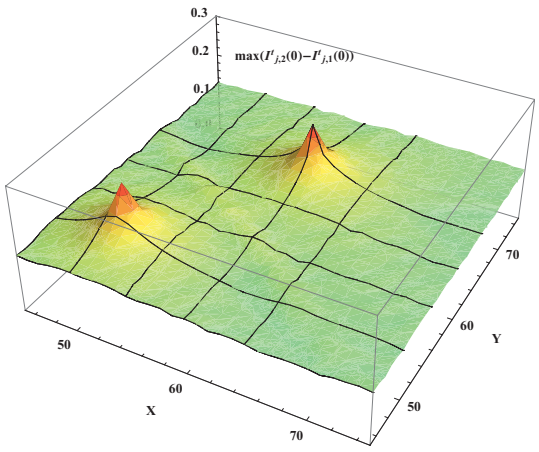


Figure 5: Maximum $[I_c^t]^0$ appeared during the first 125 time steps after an outbreak was recorded in two polygons, which are at distance 14 from one another.

paradigm. The presented modelling framework is perfectly suited to cope with the growing importance and availability of spatio-temporal data. Further work on this topic will consist of extending the presented model to cover also recovered individuals.

References

[1] J.D. Murray. *Mathematical Biology: I. An Introduction*. Springer, Berlin, Germany, second edition, 1993.

[2] J.D. Murray. *Mathematical Biology: II. Spatial Models and Biomedical Applications*. Springer, Berlin, Germany, third edition, 2007.

[3] M. Oberguggenberger and S. Pittschmann. Differential equations with fuzzy parameters. *Mathematical and Computer Modelling of Dynamical Systems*, 5:181–202, 1999.

[4] V. Lakshmikantham and R.N. Mohapatra. Theory of fuzzy differential equations and inclusions. In R.P. Agarwal and D. O'Regan, editors, *Series in Mathematical Analysis and Applications*, volume 6. Taylor & Francis, New York, United States, 2003.

[5] M. Guo and R. Li. Impulsive functional differential inclusions and fuzzy populations models. *Fuzzy Sets and Systems*, 138:601–615, 2003.

[6] J.J. Buckley and T. Feuring. Fuzzy differential equations. *Fuzzy Sets and Systems*, 110:43–54, 2000.

ISBN: 978-989-95079-6-8

[7] Y. Chalco-Cano and H. Román-Flores. On new solutions of fuzzy differential equations. *Chaos, Solitons and Fractals*, 38:112–119, 2008.

[8] D. Vorobiev and S. Seikkala. Towards the theory of fuzzy differential equations. *Fuzzy Sets and Systems*, 125:231–237, 2002.

[9] J.M. Baetens and B. De Baets. A generalized coupled-map lattice to model biological phenomena. *Mathematical Biology*, 2009. Submitted.

[10] S. Wolfram. Statistical mechanics of cellular automata. *Reviews of Modern Physics*, 55:601–644, 1983.

[11] S. Wolfram, editor. *A New Kind of Science*. Wolfram Media Inc., Champaign, United States, 2002.

[12] K. Kaneko, editor. *Theory and Applications of Coupled Map Lattices*. John Wiley & Sons Ltd., Chichester, United Kingdom, 1193.

[13] W.G. Wee and K.S. Fu. A formulation of fuzzy automata and its application as a model of learning systems. *IEEE Transactions on Systems Science and Cybernetics*, 5:215–223, 1969.

[14] M. Doostfatemeh and S.C. Kremer. New directions in fuzzy automata. *International Journal of Approximate Reasoning*, 38:175–214, 2005.

[15] J.N. Mordeson and D.S. Malik. *Fuzzy automata and languages, theory and applications*. Chapman and Hall/CRC, Boca Raton, United States, 2002.

[16] G. Cattaneo, P. Flocchini, C. Quaranta Vogliotti, and N. Santoro. Cellular automata in fuzzy backgrounds. *Physica D*, 105:105–120, 1997.

[17] P. Flocchini, F. Geurts, and N. Santoro. CA-like error propagation in fuzzy CA. *Parallel Computing*, 23:1673–1682, 1997.

[18] D. Dunne and A. Mingarelli. On the dynamics of some exceptional fuzzy cellular automata. In *7th International Conference on Cellular Automata for Research and Industry, ACRI 2006*, pages 78–87, Perpignan, France, September 2006.

[19] S. Assilian. *Artificial Intelligence in the Control of Real Dynamical Systems*. PhD thesis, London University.

[20] E.H. Mamdani. Application of fuzzy algorithms for control of simple dynamic plant. *Proceedings of the IEEE*, 121:1585–1588, 1974.

[21] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15:116–131, 1985.

[22] C. Bone, S. Dragicevic, and A. Roberts. A fuzzy-constrained cellular automata model of forest insect manifestations. *Ecological Modelling*, 192:107–125, 2006.

[23] F. Wu. Simulating urban encroachment on rural land with fuzzy-logic-controlled cellular automata in a geographical information system. *Journal of Environmental Management*, 5:293–308, 1998.

[24] M. da Silva Peixoto, L. Carvalho de Barros, and R. Carlos Bassanezi. A model of cellular automata for the spatial and temporal analysis of citris sudden death with the fuzzy parameter. *Ecological Modelling*, 214:45–52, 2008.

[25] M. Mraz, N. Zimic, and J. Virant. Intelligent bush fire spread prediction using fuzzy cellular automata. *Journal of Intelligent and Fuzzy Systems*, 7:203–207, 1999.

[26] Q. Chen and A.E. Mynett. Modelling algal bloom in the Dutch coastal waters by integrated numerical and fuzzy cellular automata approaches. *Ecological Modelling*, 199:73–81, 2006.

[27] L.A. Mentelas, P. Prastacos, and T. Hatzichristos. Modeling urban growth using fuzzy cellular automata. In *Proceedings of the 11th AIGLE Conference (CD-ROM)*, Girona, Spain, 2008.

[28] S. Hoya White, A. Martín del Rey, and G.Rodríguez Sanchez. Modeling epidemics using cellular automata. *Applied Mathematics and Computation*, 186:193–202, 2007.

[29] J. Milne and S C. Fu. Epidemic modelling using cellular automata. In *Proceedings of the 1st Australian Conference on Artificial Life (ACAL'03)*, pages 43–57, Canberra, December 2003.

[30] R.J. Doran and S.W. Laffan. Simulating the spatial dynamics of foot and mouth disease outbreaks in feral pigs and livestock in Queensland, Australia, using a susceptible-infected-recovered cellular automata model. *Preventive Veterinary Medicine*, 70:133–152, 2005.

[31] A.K. Dewdney. Sharks and fish wage an ecological war on the toroidal planet. *Scientific American*, pages 14–22, December 1984.

[32] H. Baltzer, P.W. Braun, and W. Köhler. Cellular automata models for vegetation dynamics. *Ecological Modelling*, 107:113–125, 1998.

[33] A. Qi, X. Zheng, C. Du, and B. An. A cellular automaton model of cancerous growth. *Journal of Theoretical Biology*, 161:1–12, 1993.

[34] D.G. Mallet and L.G. De Pillis. A cellular automata model of tumor-immune system interactions. *Journal of Theoretical Biology*, 239:334–350, 2006.

[35] L. Preziosi. *Cancer Modelling and Simulation*. Chapman & Hall, Boca Raton, United States, 2003.

[36] C. Picioreanu, M.C.M. van Loosdrecht, and J.J. Heijnen. Mathematical modeling of biofilm structure with a hybrid differential-discrete cellular automaton approach. *Biotechnology and Bioengineering*, 58:101–116, 1998.

[37] G. Pizarro, D. Griffeath, and D.R. Noguera. Quantitative cellular automaton model for biofilms. *Journal of Environmental Engineering*, 127:782–789, 2001.

[38] L.B. Kier, P.G. Seybold, and C. Cheng. *Modelling Chemical Systems using Cellular Automata*. Springer, Dordrecht, The Netherlands, 2005.

[39] J.L. Schiff. *Cellular Automata: A Discrete View of the World*. John Wiley & Sons Ltd., Chichester, United Kingdom, 2008.

[40] L. Beale, J.J. Abellan, S. Hodgson, and L. Jarup. Methodologic issues and approaches to spatial epidemiology. *Environmental Health Perspectives*, 116:1105–1110, 2008.

[41] D. Dubois and H. Prade. *Possibility Theory: an Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, United States, 1988.

[42] R. Fullér and P. Majlender. On interactive fuzzy numbers. *Fuzzy Sets and Systems*, 143:355–369, 2004.

[43] P. Majlender. *A Normative Approach to Possibility Theory and Soft Decision Support*. PhD thesis, Åbo Akademi University, Finland, 2004.

[44] C. Carlsson, R. Fullér, and P. Majlender. On possibilistic correlation. *Fuzzy Sets and Systems*, 155:425–445, 2005.

ISBN: 978-989-95079-6-8

Recurrent Neural Kalman Filter Identification and Indirect Adaptive Control of a Continuous Stirred Tank Bioprocess

Ieroham S. Baruch[1] Carlos R. Mariaca-Gaspar[1]

1 Department of Automatic Control, CINVESTAV-IPN,
Ave. IPN No 2508, Col. Zacatenco, A.P. 14-740,
07360 Mexico City, MEXICO
Email: {baruch,cmariaca}@ctrl.cinvestav.mx

Abstract— *The aim of this paper is to propose a new Kalman Filter Recurrent Neural Network (KFRNN) topology and a recursive Levenberg-Marquardt (L-M) algorithm of its learning capable to estimate states and parameters of a highly nonlinear Continuous Stirred Tank Bioreactor (CSTR) in noisy environment. The estimated parameters and states obtained by the proposed KFRNN identifier are used to design an indirect adaptive sliding mode control scheme. The obtained simulation results of the real-time neural identification and control of a CSTR model, taken from the literature, exhibited fast convergence, noise filtering, and low mean squared error of reference tracking. A 20 runs comparative validation experiment in noisy environment is also done. It gives some priority of the L-M learning over the BP one.*

Keywords— Backpropagation learning, continuous stirred tank bioreactor, indirect adaptive sliding mode control, Kalman filter recurrent neural network, Levenberg-Marquardt learning.

1 Introduction

The Recent advances in understanding of the working principles of artificial neural networks has given a tremendous boost to identification, prediction and control tools of nonlinear systems, [1], [2]. The main network property namely the ability to approximate complex non-linear relationships without prior knowledge of the model structure makes them a very attractive alternative to the classical modeling and control techniques, [3]. It is particularly useful in applications where the complexity of the data or tasks makes the design of such a functions by hand impractical. Among several possible network architectures the ones most widely used are the Feedforward NN (FFNN) and Recurrent NN (RNN). In a FFNN the signals are transmitted only in one direction, starting from the input layer, subsequently through the hidden layers to the output layer, which requires applying a tap delayed global feedbacks and a tap delayed inputs to achieve a Nonlinear Autoregressive Moving Average (NARMAX) neural dynamic plant model. A RNN has local feedback connections to some of the previous layers. Such a structure is suitable alternative to the FFNN when the task is to model dynamical systems. Its main advantage is the reduced complexity of the network structure. However, the analysis of state of the art in the area of classical RNN-based modelling and control has also shown some of their inherent limitations as follows: 1. The RNN input vector consists of a number of past system inputs and outputs and there is not a systematic way to define the optimal number of past values, and usually, the method of trials and errors is performed, [2]. 2. The RNN model is naturally formulated as a discrete model with fixed sampling period, therefore, if the sampling period is changed, the neural network has to be trained again; 3. It is assumed that the plant order is known, which represents a quite strong modelling assumption in general, [1]; 4. The managing of noisy input/output plant data required to augment the filtering capabilities of the identification RNNs, [4]. Driven by these limitations, a new Kalman Filter Recurrent Neural Network (KFRNN) topology and the recursive Backpropagation (BP) type learning algorithm in vector-matrix form was derived and its convergence was studied, [5]. But the recursive BP algorithm, applied for KFRNN learning, is a gradient descent first order learning algorithm which not permits to augment the precision and to accelerate the learning. So, the aim of the paper is to use a second order learning algorithm for the KFRNN like the Levenberg-Marquardt (L-M) algorithm is, [4, 6]. The KFRNN with L-M learning will be applied for Continuous Stirred Tank Reactor (CSTR) model identification, [7]. In [8, 9] a comparative study of linear, nonlinear and neural-network-based adaptive controllers for a CSTR is done. The papers proposed to use the neuro-fuzzy and adaptive nonlinear control systems design, applying FFNNs (multilayer perceptron and radial basis functions NN). The proposed control gives a good adaptation to the nonlinear plants dynamics, better with respect to the other methods of control, but the used FFNNs have a great complexity, and the plant order has to be known. The application of KFRNNs, [5], together with the recursive L-M, [4], could avoid these problems improving the learning and the precision of the plant states and parameters estimation. Here the obtained from the KFRNN identifier parameters and states will be used to design an indirect adaptive sliding mode control scheme, [5].

2 Topology and learning of the KFRNN

2.1 KFRNN Topology

A Kalman Filter Recurrent Neural Network model and its learning algorithm of dynamic Backpropagation-type, together with the explanatory figures and stability proofs, are described in [5]. The KFRNN topology (see Fig. 1) is described by the following vector-matrix equations:

ISBN: 978-989-95079-6-8

$$X(k+1) = A_1 X(k) + BU(K) - Dy(k) \quad (1)$$

$$Z(k) = \mathrm{G}[X(k)] \quad (2)$$

$$V_1(k) = CZ(k) \quad (3)$$

$$V(k+1) = A_2 V(k) + V_1(k) \quad (4)$$

$$Y(k) = \mathrm{F}[V(k)] \quad (5)$$

$$A_1 = block\text{-}diag\,(A_{1,i})\,;\ |A_{1,i}| < 1$$
$$A_2 = block\text{-}diag\,(A_{2,i})\,;\ |A_{2,i}| < 1 \quad (6)$$

Where: Y, X, and U are, respectively, output, state and input vectors with dimensions l, n, m; A_1 and A_2 are (nxn) and (lxl) block-diagonal local feedback weight matrices; $A_{1,i}$ and $A_{2,i}$ are i-th diagonal block of A_1 and A_2 with (nxn) and (lxl) dimensions, respectively. Equation (6) represented the local stability conditions, imposed on all blocks of A_1 and A_2; B and C are (nxm) and (lxn) input and output weight matrices; D is a (nxl) global output feedback weight matrix; G(.), F(.) are vector-valued sigmoid or hyperbolic tangent-activation functions, Z, V_1, V are vector variables with corresponding dimensions; the integer k is a discrete-time variable.

2.2 Backpropagation learning of the KFRNN

The general BP learning algorithm is given by the following equation:

$$W(k+1) = W(k) + \eta \Delta W(k) + \alpha \Delta W(k-1) \quad (7)$$

Where: W is the weight matrix, being modified (A_1, A_2, B, C, D); ΔW is the weight matrix correction (ΔA_1, ΔA_2, ΔB, ΔC, ΔD); η and α are learning rate parameters. Applying the diagrammatic method, [10], and using the block- diagram of the KFRNN topology (see Fig. 1), we could design an error predictive adjoint KFRNN (see Fig. 2). Following this adjoint KFRNN block diagram we could obtain the next matrix KFRNN weight updates:

$$\Delta C(k) = E_1(k) Z^T(k) \quad (8)$$

$$\Delta A_2(k) = E_1(k) V^T(k) \quad (9)$$

$$E_1(k) = F'[Y(k)]E(k);\ E(k) = Y_p(k) - Y(k) \quad (10)$$

$$\Delta B(k) = E_3(k) U^T(k) \quad (11)$$

$$\Delta A_1(k) = E_3(k) X^T(k) \quad (12)$$

$$\Delta D(k) = E_3(k) Y^T(k) \quad (13)$$

$$E_3(k) = G'[Z(k)]E_2(k);\ E_2(k) = C^T E_1(k) \quad (14)$$

$$\Delta vA_1(k) = E_3(k) \oplus X(k);\ \Delta vA_2(k) = E_1(k) \oplus V(k) \quad (15)$$

Where: ΔA_1, ΔA_2, ΔB, ΔC, ΔD are weight corrections of the of the learned matrices A_1, A_2, B, C, D, respectively; E is an error vector of the output KFRNN layer, where Yp is a desired target vector and Y is a KFRNN output vector, both with dimensions l; X is a state vector, and E1, E2, E3 are error vectors, shown on Fig. 2; F'(.), G'(.) are diagonal

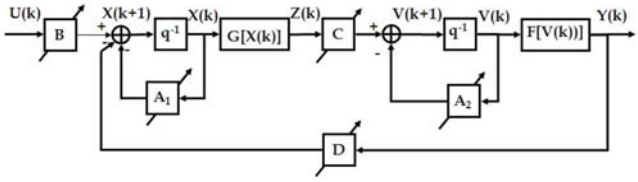


Figure 1: Block-diagram of the KFRNN topology.

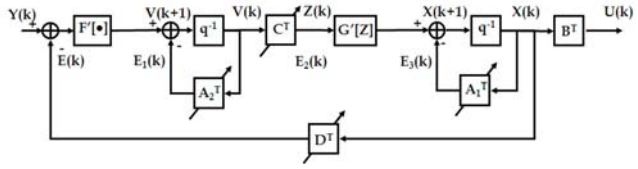


Figure 2: Block-diagram of the adjoint KFRNN topology.

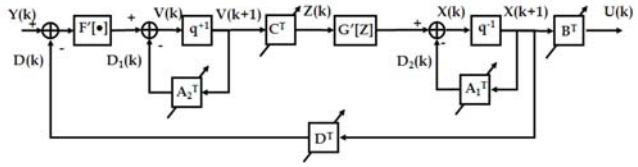


Figure 3: Block-diagram of the adjoint KFRNN used for the L-M algorithm.

Jacobean matrices with appropriate dimensions. The diagonal elements of these matrices are derivatives of the activation functions. The equations (12), (9) represented the weight update of the feedback weight matrices A_1, A_2, as full (nxn), (lxl) matrices. The equation (15) gives the learning solution as an element by element vector product when these matrices are diagonal vA_1, vA_2 which is our case.

2.3 Recursive Levenberg-Marquardt KFRNN learning

The general recursive L-M algorithm of learning, [4, 6], is given by the following equations:

$$W(k+1) = W(k) + P(k)\nabla Y[W(k)]E[W(k)] \quad (16)$$

$$Y[W(k)] = g[W(k), U(k)] \quad (17)$$

$$E^2[W(k)] = \{Y_p(k) - g[W(k), U(k)]\}^2 \quad (18)$$

$$DY[W(k)] = \frac{\partial}{\partial W} g[W, U(k)]\Big|_{W=W(k)} \quad (19)$$

Where: W is a general weight matrix (A_1, A_2, B, C, D), under modification; P is the covariance matrix of the weights estimates, being updated; DY is a nw-dimensional gradient vector; Y is the KFRNN output vector which depended on the updated weights and the input; E is an error vector; Yp is the plant output vector which is in fact the target vector. Using the same KFRNN adjoint block diagram (see Fig.2), we could obtain the values of the gradients DY for each updated weight, propagating the value D(k) = I through it and following the block diagram of Fig. 3, we could apply equation (19) for each element of the weight matrices (A_1, A_2, B, C, D) to be updated. The corresponding gradient components are as follows:

$$DY[C_{ij}(k)] = D_{1,i}(k) Z_j(k) \quad (20)$$

$$DY[A_{2ij}(k)] = D_{1,i}(k) V_j(k) \quad (21)$$

ISBN: 978-989-95079-6-8

$$D_{1,i}(k) = F_i'[Y_i(k)] \tag{22}$$

$$DY[A_{1ij}(k)] = D_{2,i}(k) X_j(k) \tag{23}$$

$$DY[B_{ij}(k)] = D_{2,i}(k) U_j(k) \tag{24}$$

$$DY[D_{ij}(k)] = D_{2,i}(k) Y_j(k) \tag{25}$$

$$D_{2,i}(k) = G_i'[Z_i(k)] C_i D_{1,i}(k) \tag{26}$$

So the Jacobean matrix could be formed as:

$$DY[W(k)] = [DY(C_{ij}(k)), DY(A_{2ij}(k)), DY(B_{ij}(k)) \\ DY(A_{1ij}(k)), DY(D_{ij}(k))] \tag{27}$$

The P(k) matrix is computed recursively by the equation:

$$P(k) = \alpha^{-1}(k)\{P(k-1) - \\ -P(k-1)\Omega[W(k)] S^{-1}[W(k)] \Omega^T[W(k)] P(k-1)\} \tag{28}$$

Where: the S(.), and Ω(.) matrices are given as follows:

$$S[W(k)] = \alpha(k)\Lambda(k) + \Omega^T[W(k)] P(k-1)\Omega[W(k)] \tag{29}$$

$$\Omega^T[W(k)] = \begin{bmatrix} & \nabla Y^T[W(k)] & \\ 0 \quad \cdots & 1 & \cdots \quad 0 \end{bmatrix}; \\ \Lambda(k)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix}; 10^{-4} \le \rho \le 10^{-6}; \\ 0.97 \le \alpha(k) \le 1; 10^3 \le P(0) \le 10^6 \tag{30}$$

The matrix Ω(.) has dimension (nwx2), where the second row has only one unity element (the others are zero). The position of that element is computed by:

$$i = k \bmod (nw) + 1; k > nw \tag{31}$$

Next the given up topology and learning are applied for CSTR system identification and control.

3 Indirect Adaptive Control Scheme (Sliding Mode Control)

The indirect adaptive control using the RTNN as plant identifier has been described in, [11]. Later the proposed indirect control has been derived as a Sliding Mode Control (SMC) and applied for control of unknown hydrocarbon biodegradation processes, [12]. The block diagram of the indirect adaptive control scheme is shown on Fig. 4. It contains identification and state estimation KF RNN and a sliding mode controller. The stable nonlinear plant is identified by a KFRNN model with topology, given by equations (1)-(6) learned by the stable BP-learning algorithm, given by equations (7)-(15), or using the second order LM-learning algorithm, given by equations (16)-(31). The simplification and linearization of the neural identifier equations (1)-(2), omitting the DY(.) term, leads to the next

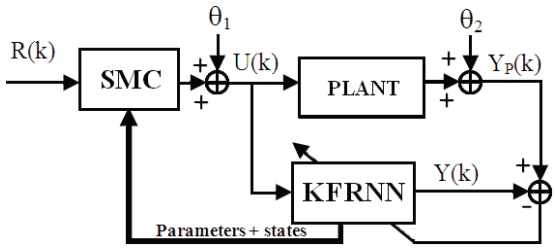


Figure 4: Block diagram of the closed-loop system containing KF RNN identifier and a SMC.

local linear plant model, extracted from the complete KF RNN model:

$$X(k+1) = A_1 X(k) + BU(k) \tag{32}$$

$$Z(k) = HX(k); \; H = CG'(Z) \tag{33}$$

Where G'(.) is the derivative of the activation function and L = M, is supposed.

In [13], the sliding surface is defined with respect to the state variables, and the SMC objective is to move the states form an arbitrary space position to the sliding surface in finite time. In [14], the sliding surface is also defined with respect to the states but the states of the SISO systems are obtained from the plant outputs by differentiation. In [15], the sliding surface definition and the control objectives are the same. The equivalent control systems design is done with respect to the plant output, but the reachability of the stable output control depended on the plant structure. In [12], the sliding surface is derived directly with respect to the plant outputs which facilitated the equivalent SMC systems design. Let us define the following SS as an output tracking error function:

$$S(k+1) = E(k+1) + \sum_{i=1}^{P} \gamma_i E(k-i+1); \; |\gamma| < 1 \tag{34}$$

Where: S(.) is the Sliding Surface Error Function (SSEF) defined with respect to the plant output; E(.) is the systems output tracking error; γ_i are parameters of the desired stable SSEF; p is the order of the SSEF. The tracking error and their iterate are defined as:

$$E(k) = R(k) - Z(k); E(k+1) = R(k+1) - Z(k+1) \tag{35}$$

Where R(k), Z(k) are L-dimensional reference and output vectors of the local linear plant model. The objective of the sliding mode control systems design is to find a control action which maintains the systems error on the sliding surface which assure that the output tracking error reaches zero in P steps, where P < N. So, the control objective is fulfilled if:

$$S(k+1) = 0 \tag{36}$$

ISBN: 978-989-95079-6-8

Now, let us to iterate (33) and to substitute (32) in it so to obtain the input/output local plant model, which yields:

$$Z\left(k+1\right) = FX\left(k+1\right) = F\left[AX\left(k\right) + BU\left(k\right)\right] \quad (37)$$

From (34), (35), and (36) it is easy to obtain:

$$R\left(k+1\right) - Z\left(k+1\right) + \sum_{i=1}^{P} \gamma_i E\left(k-i+1\right) = 0 \quad (38)$$

The substitution of (37) in (38) gives:

$$R\left(k+1\right) - FAX\left(k\right) + FBU\left(k\right) - \sum_{i=1}^{P} \gamma_i E\left(k-i+1\right) \quad (39)$$

As the local approximation plant model (32), (33), is controllable, observable and stable (see [5]), the matrix A_1 is diagonal, and L = M, then the matrix product (HB), representing the plant model static gain, is nonsingular, and the plant states X(k) are smooth non-increasing functions. Now, from (39) it is easy to obtain the equivalent control capable to lead the system to the sliding surface which yields:

$$U_{eq}\left(k\right) = \left(FB\right)^{-1}\left[-FAX\left(k\right) + R\left(k+1\right) + \sum_{i=1}^{P} \gamma_i E\left(k-i+1\right)\right] \quad (40)$$

Following [13], the SMC avoiding chattering is taken using a saturation function instead of sign one. So the SMC takes the form:

$$U^*(k) = \begin{cases} U_{eq}\left(k\right) & \text{if } \left\|U_{eq}\left(k\right)\right\| < U_0 \\ -U_0 U_{eq}\left(k\right) / \left\|U_{eq}\left(k\right)\right\| & \text{if } \left\|U_{eq}\left(k\right)\right\| \geq U_0 \end{cases} \quad (41)$$

The SMC substituted the multi-input multi-output coupled high order dynamics of the linearized plant with desired decoupled low order one.

4 Description of the CSTR bioprocess

The CSTR model given in [16, 17] was chosen as a realistic example for application of the KFRNN and the SMC for solution of system identification and control problems. The CSTR is described by the following continuous time, nonlinear, system of ordinary differential equations:

$$\frac{dC_A(t)}{dt} = \frac{Q}{V}\left(C_{Af} - C_A(t)\right) - k_0 C_A(t) \exp\left(-\frac{E}{RT(t)}\right) \quad (42)$$

$$\frac{dT(t)}{dt} = \frac{Q}{V}(T_f - T(t)) + \frac{(-\Delta H)C_A(t)}{\rho C_p}\exp\left(\frac{E}{RT(t)}\right) + \frac{\rho_c C_{pc}}{\rho C_p V} Q_c(t)\left[1 - \exp\left(\frac{-hA}{Q_c(t)\rho_c C_{pc}}\right)\right](t_{ef} - T(t)) \quad (43)$$

It suffices to know that within the CSTR two chemicals are mixed and react to produce a product compound A at a concentration CA(t), with the temperature of the mixture being T(t). The reaction is exothermic and producing heat which slows down the reaction. By introducing a coolant flow-rate Qc(t), the temperature can be varied and hence the product concentration controlled. CAf is the inlet feed concentration, Q is the process flow-rate, Tf and Tef are the inlet feed and coolant temperatures, respectively, all of which are assumed constant at nominal values. Likewise k0, E/R, V, ΔH, ρ, Cpc, Cp and ρc are thermodynamic and chemical constant relating to this particular problem. Numerical values for the parameters and nominal operating conditions of this model are given in Table 1.

The quantities Qc0, T0, and CA0 shown in Table 1 are steady values for steady operating point in the CSTR. The objective is to control the product compound A, manipulating the variable $Q_c(t)$. The operating values of the variables are obtained from the papers [16], [17]. The authors of these papers examined the performances of neural networks and of fuzzy-neural networks based control systems, respectively.

5 Simulation Results

The proposed indirect adaptive SMC is applied for the given up CSTR plant. The Fig. 5 and Fig. 7 showed the comparison between the reference and the plant output signals applying the BP and L-M algorithms of learning for the KFRNN, respectively. Detailed comparative graphical simulation results of CSTR KFRNN plant identification by means of the BP and the L-M learning are given in Fig.6 and Fig.8. A 10% white noise is added to the plant inputs and outputs and the behavior of the plant identification and control has been studied accumulating some statistics of the final MSE% (ξ_{av}) for BP and L-M learning, which results are given in Table 3 and Table 4 for 20 runs. The mean average cost for all runs (ε) of control and the standard deviation (σ) with respect to the mean value and the deviation (Δ) are presented in Table 2 for BP and L-M algorithms of learning.

Table 1: CSTR parameters and operating conditions

$Q = 100 \quad (L/\min)$	$E/R = 9.95\times10^3 \quad (K)$
$C_{Af} = 1.0 \quad (mol/L)$	$-\Delta H = 2\times10^5 \quad (cal/mol)$
$T_f = 350 \quad (K)$	$\rho\cdot\rho_c = 1000 \quad (g/L)$
$T_{cf} = 350 \quad (K)$	$C_p C_{pc} = 1 \quad (cal/gK)$
$V = 100 \quad (L)$	$Q_{c0} = 103.41 \quad (L/\min)$
$hA = 7\times10^5 \quad (cal/\min K)$	$T_0 = 440.2 \quad (K)$
$k_0 = 7.2\times10^{10} \quad (1/\min)$	$C_{A0} = 0.0836 \quad (mol/L)$

Table 2: Standard deviation and mean average values using BP and L-M algorithms of learning.

BP Algorithm	LM Algorithm
ε = 1.0997	ε =1.0370
σ = 0.0295	σ =0.0232

ISBN: 978-989-95079-6-8

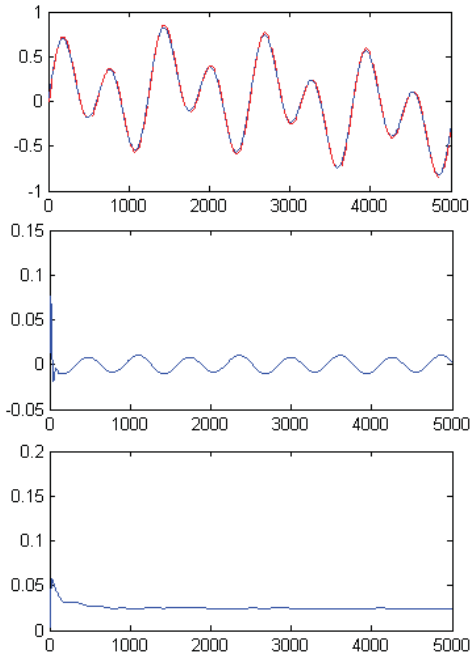


Figure 5: Detailed graphical simulation results of CSTR plant Sliding Mode Indirect Control using BP KFRTNN learning a) comparison between the plant output and the reference signal; b) control signal; c) MSE% of control.

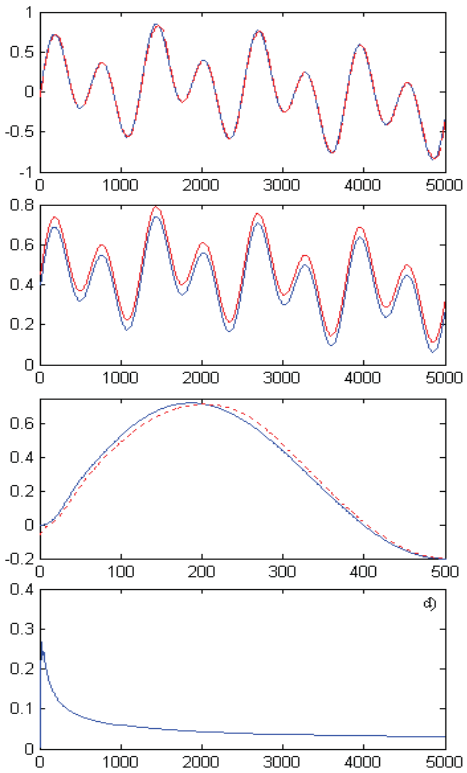


Figure 6: Graphical results of identification using BP KFRNN learning. a) Comparison of the plant output (continuous line) and KFRNN output (pointed line); b) state variables; c) comparison of the plant output (continuous line) and KFRNN output (pointed line) in the first instants; d) MSE% of identification

The values are computed using the following formulas:

$$\varepsilon = \frac{1}{n}\sum_{k=1}^{n}\xi_{av_k}, \quad \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\Delta_i^2}, \quad \Delta = \xi_{av} - \varepsilon \tag{44}$$

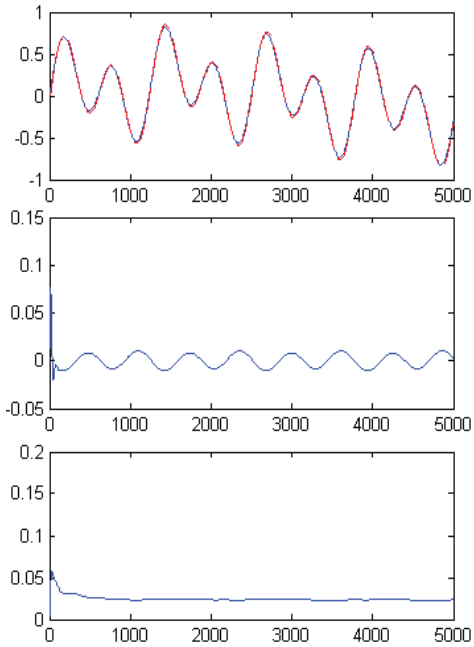


Figure 7: Detailed graphical simulation results of CSTR plant Sliding Mode Indirect Control using L-M KFRTNN learning a) comparison between the plant output and the reference signal; b) control signal; c) MSE% of control.

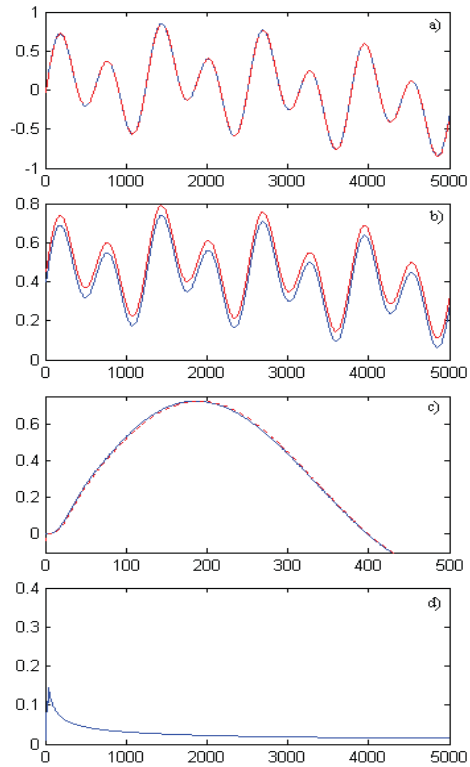


Figure 8: Graphical results of identification using L-M KFRNN learning. a) Comparison of the plant output (continuous line) and KFRNN output (pointed line); b) state variables; c) comparison of the plant output (continuous line) and KFRNN output (pointed line) in the first instants; d) MSE% of identification

ISBN: 978-989-95079-6-8

Table 3: Final MSE% of control for 20 runs of control program using the Backpropagation algorithm.

No.	**1**	**2**	**3**	**4**	**5**
MSE%	1.1096	1.1530	1.1331	1.0660	1.1375
No.	**6**	**7**	**8**	**9**	**10**
MSE%	1.0701	1.1171	1.1202	1.1311	1.0899
No.	**11**	**12**	**13**	**14**	**15**
MSE%	1.0964	1.1007	1.1139	1.0685	1.0502
No.	**16**	**17**	**18**	**19**	**20**
MSE%	1.1330	1.1039	1.0724	1.0609	1.0665

Table 4: Final MSE% of control for 20 runs of control program using the Levenberg-Marquardt algorithm.

No.	**1**	**2**	**3**	**4**	**5**
MSE%	1.0296	1.0462	1.0472	1.0670	1.0134
No.	**6**	**7**	**8**	**9**	**10**
MSE%	1.0078	1.0396	1.0678	1.0017	1.0609
No.	**11**	**12**	**13**	**14**	**15**
MSE%	1.0019	1.0364	1.0135	1.0501	1.0175
No.	**16**	**17**	**18**	**19**	**20**
MSE%	1.0654	1.0096	1.03652	1.0627	1.0660

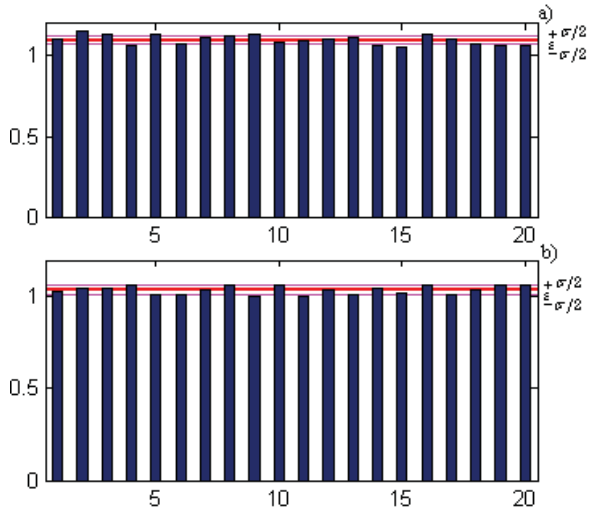


Figure 9: Comparison between the final MSE% for 20 runs of control program: a) using BP algorithm, b) using L-M algorithm.

6 Conclusions

The paper proposed a new KFRNN model for systems identification and states estimation of nonlinear plants. The KFRNN is learned by the second order recursive learning algorithm of Levenberg-Marquardt. The estimated parameters and states of the RNN model are used for indirect (SM) adaptive trajectory tracking control systems design. The applicability of the proposed neural control system is confirmed by simulation results with a CSTR plant. The results showed good convergence of both L-M and BP learning algorithms. The L-M algorithm of learning is more precise (see Table 2) but more complex then the BP one.

Acknowledgment

The Ph.D. student Carlos-Roman Mariaca-Gaspar is thankful to CONACYT, Mexico for the fellowship received during his studies at CINVESTAV-IPN, Mexico.

References

[1] K.S. Narendra, and K. Parthasarathy, Identification and Control of Dynamic Systems Using Neural Networks, *IEEE Transactions on Neural Networks,* 1(1): 4 -27, 1990.

[2] K.J. Hunt, D. Sbarbaro, R. Zbikowski, and P.J. Gawthrop, Neural Network for Control Systems - A Survey, *Automatica*, 28(6): 1083-1112, 1992.

[3] S.A. Pao, S.M. Phillips, and D.J. Sobajic, Neural Net Computing and Intelligent Control Systems, *International Journal of Control*, 56: 263-289, 1992.

[4] I. S. Baruch, S. Escalante, and C. R. Mariaca-Gaspar, Identification, Filtering and Control of Nonlinear Plants by Recurrent Neural Networks Using First and Second Order Algorithms of Learning, *Dynamics of Continuous, Discrete and Impulsive Systems, International Journal, Series A: Mathematical Analysis, Special Issue on Advances in Neural Networks-Theory and Applications*, Watam Press, Canada, ISSN 1201-3390, 14: 512-521, 2007.

[5] I.S. Baruch, C.R. Mariaca-Gaspar, and J. Barrera-Cortes, Recurrent Neural Network Identification and Adaptive Neural Control of Hydrocarbon Biodegradation Processes, *In: Recurrent Neural Networks, Xiaolin Hu, and P. Balasubramaniam (Eds.)*, ISBN 978-3-902613-28-8, I-Tech Education and Publishing KG, Vienna, Austria, Chapter 4: 61-88, 2008.

[6] L.S. Ngia, and J. Sjöberg, Efficient Training of Neural Nets for Nonlinear Adaptive Filtering Using a Recursive Levenberg-Marquardt Algorithm, *IEEE Transactions on Signal Processing*, 48: 1915-1927, 2000.

[7] L. D. Schmidt, *The Engineering of Chemical Reactions,* New York: Oxford University Press, ISBN 0-19-510588-5, 1998.

[8] Shi-Rong Liu, and Jin-Shou Yu, Robust Control Based on Neuro Fuzzy Systems for a Continuous Stirred Tank Reactor, *In: Proc. of the First International Conference on Machine Learning and Cybernetics,* Beijing, 4-5 November 2002, 1483-1488, 2002.

[9] T. Zhang, and M. Guay. Adaptive Nonlinear Control of Continuously Stirred Tank Reactor Systems, *In: Proc. of the American Control Conference, Arlington*, 25-27 June 2001, 1274-1279, 2001.

[10] E. Wan, and F. Beaufays, Diagrammatic Method for Deriving and Relating Temporal Neural Networks Algorithms, *Neural Computations*, 8: 182-201, 1996.

[11] I.S. Baruch, J.M.Flores, F. Thomas, and R. Garrido, Adaptive Neural Control of Nonlinear Systems, *In: Artificial Neural Networks-ICANN 2001, LNCS 2130, G. Dorffner, H. Bischof, K. Hornik (Eds.)*, ISBN 3-540-42486-5, Springer, Berlin, 930-936, 2001.

[12] I.S. Baruch, C.R. Mariaca-Gaspar, I. Cruz-Vega, and J. Barrera-Cortes, Sliding Mode Control of a Hydrocarbon Degradation in Biopile System Using Recurrent Neural Network Model, *In: MICAI 2007: Advances in Artificial Intelligence, LNAI 4827, A. Gelbukh, A.F. Kuri-Morales (Eds.)*, ISBN-10 3-540-76630-8, Springer, Berlin Heidelberg New York, 1184-1194, 2001.

[13] K. D. Young, V.I. Utkin, and U. Ozguner, A Control Engineer's Guide to Sliding Mode Control, *IEEE Transactions on Control Systems Technology*, 7 (3): 328-342, 1999.

[14] A. Levent, Higher Order Sliding Modes, Differentiation and Output Feedback Control, *International Journal of Control, Special Issue Dedicated to Vadim Utkin on the Occasion of his 65th Birthday (Guest editor: Leonid M. Fridman),* ISSN 0020-7179, 76(9/10): 924-941, 15 June-10 July 2003.

[15] C. Eduards, S.K. Spurgeon, and R.G. Hebden, On the Design of Sliding Mode Output Feedback Controllers, *International Journal of Control, Special Issue Dedicated to Vadim Utkin on the Occasion of his 65th Birthday (Guest editor: Leonid M. Fridman),* ISSN 0020-7179, 76(9/10): 893-905, 15 June-10 July 2003.

[16] G. Lightbody, and G. W. Irwin, Nonlinear Control Structures Based on Embedded Neural Systems Models, *IEEE Transactions on Neural Networks*, 8: 553-557, 1997.

[17] A. Aoyama, F. J. Doyle, and V. Venkatasubramanian, Control-Affine Fuzzy Neural Networks Approach for Nonlinear Process Control, *Journal of Process Control*, 8: 375-386, 1995.

A Multi-granular Linguistic Promethee Model

Nesrin Halouani[1], Luis Martínez[2], Habib Chabchoub[1], Jean-Marc Martel[3], Jun Liu[4]

[1] University of Economic Sciences and Management, Sfax, Tunisia,
[2] University of Jaén, Spain,
[3] University of Laval, Canada,
[4] University of Ulster, Nothern Ireland, UK
{halouani.nesrin@gmail.com, martin@ujaen.es}

Abstract— *In Multi-criteria Decision Making (MCDM) problems dealing with qualitative criteria and uncertain information the use of linguistic values is suitable for the experts in order to express their judgments. It is common that the group of experts involved in such problems have different degrees of knowledge about the criteria, so we propose a multi-granular linguistic framework such that each expert can provide his/her evaluations in different linguistic term sets according to his/her knowledge. MCDM problems have been solved in the literature by using different methods, in this contribution we focus on PROMETHEE method and our proposal consists of developing tools and operators for the PROMETHEE method to deal with multi-granular linguistic information.*

Keywords— Multi-criteria decision making, linguistic hierarchies, Promethee.

1 Introduction

Decision-making is a common human activity and its multidimensional nature of real world decision problems is well addressed by multi-criteria decision aid (MCDA). The focal point of interest within the methodological framework of MCDA is the analysis and the modelling of the multiple decision makers' preferences. This special characteristic of MCDA implies that a comprehensive model of a decision situation cannot be developed, but instead the model should be developed to meet the expert's requirements. However, sometimes tradeoffs between some criteria may also be too difficult to define for the experts, and they can be then reluctant to express any measurable opinions.

Different methods belonging to MCDA are the following [1]: the *Electre* family developed by Roy and his co-workers, *Promethee* (Brans, Mareshal, Vincke), *Oreste* (Pastijin and Leyson), *Melchior* (Leclerq), *Qualifex* (Paelink), *Regime* (Hinloopen, Nijkamp, Rietvald), *Macbetch* (Bana e Costa, Vansnick), *Ahp* (Saaty), *Topsis* (Hwang and Yoon). Often these methods require a group of experts to express their preferences over the criteria involved in the decision process.

In the real-world, many decision problems are characterized by two overarching concerns: to consider conflict between the criteria of the problem, and to take into account the uncertainty inherent in decision making that depends on the outcome of unknown future events.

To deal with these concerns, the Promethee method is used. Promethee [2] is a popular decision method that has been successfully applied in the selection of the final solution of a problem. It generates a ranking of available alternatives, according to the expert's preferences, and the best ranked one is considered the favourite final solution.

In this paper, we focus on decision under uncertainty because is one of the most frequent situations in practical decision making, namely in planning activities in many fields. Traditional studies of such issues are conducted by using probabilistic tools and techniques. However, it is not difficult to see in many problems that aspects related to imprecision or vagueness clearly have a non probabilistic character since they are related to imprecision of meanings. Usually, when we deal with certain knowledge in a quantitative setting the information provided by the experts is expressed by means of numerical precise values. However, when we work in a qualitative setting, that is, with vague or imprecise knowledge, it could not be estimated with an exact numerical value. Then, a more realistic approach may be to use linguistic assessments instead of numerical values [11].

The use of linguistic variables makes experts' evaluations more flexible and reliable, but implies processes of computing with words (CW). The main problem that presents the traditional linguistic approaches to carry out the CW processes is the loss of information and hence a lack of precision in the final results. Different linguistic computational models have been developed Semantic model [4], Symbolic [5] or the 2-tuple one [3] that provides a model to deal with CW processes in a precise way.

Our aim in this contribution is focused on MCDM problems where different experts can have different degree of knowledge about the criteria so they can use different linguistic term sets to provide their information defining a multi-granular linguistic context. Again, the main problem is to carry out the CW processes in such a context, in order to overcome this drawback, Herrera and Martinez [6] have developed a model based on linguistic hierarchies. Thus, the CW processes in such contexts can be carried out without loss of information.

Accordingly, a flexible and realistic multi-granular hierarchical linguistic approach based on Promethee method is presented in this paper. The main advantage of this approach is to tackle the uncertainty of both performance of criteria and experts' knowledge without loss of information.

The structure of the paper is the following one. Basic concepts about Promethee method are introduced in section 2. A brief linguistic background is presented in section 3. An aggregation process for multi-granular linguistic information in PROMETHEE is proposed in section 4. In section 5, we

ISBN: 978-989-95079-6-8

apply this approach to an example. The conclusions are pointed out in section 6.

2 The Promethee Method

The PROMETHEE method (Preference Ranking Organization METHod for Enrichment Evaluation) is a multicriteria decision-making method, belonging to the family of outranking methods [2]. It is a ranking method quite simple in conception and application compared to other methods for multicriteria analysis. It is well adapted to problems where a finite number of alternatives are ranked considering several conflicting criteria. The evaluation table is the starting point of this method. In this table, the alternatives are evaluated according to different criteria. The implementation of Promethee requires two additional types on information, namely: information about the relative importance, w_j, (i.e. the weights) of the criteria considered, and information on the expert's preference modeling, which it uses when comparing the contribution of the alternatives in terms of each separate criterion.

The Promethee method encompasses two phases: (i) the aggregation of information about the alternatives and the criteria, (ii) the exploitation of the outranking relation for decision aid.

The aggregation phase requires that each point of view would be associated with a generalized criterion to assess the preference for an alternative a_i with regards to a_k as a function of $P_j(a_i,a_k) = H_j\left(f_j(a_i) - f_j(a_k)\right)$. A generalized criterion is thus a function $H_j\left(f_j(a_i) - f_j(a_k)\right)$ which is null when $\left(f_j(a_i) - f_j(a_k)\right)$ is negative, non-decreasing with $\left(f_j(a_i) - f_j(a_k)\right)$ varying between 0 and 1. Six different types of generalized criteria (for a further description see [2]) are proposed to experts, in each case at most two parameters from these thresholds q, p and s have to be fixed. Indifference threshold, q, is the largest deviation to consider as negligible on that criterion. It is a small value with regards to the scale of measurement. Preference threshold, p, is the smallest deviation to consider decisive in the preference of one alternative over another. It is a large value with respect to the scale of measurement. Gaussian threshold, s, is only used with the Gaussian preference function. It is usually fixed as intermediate value between an indifference and a preference threshold.

The outranking relation can be then represented by an oriented graph. The value of each arc is the multi-criteria preference index $\pi(.,.)$, which is defined for all ordered pairs of alternatives. These indices that may take any value in the interval $[0,1]$ define a fuzzy outranking relation. For each $(a_i,a_k) \in A \times A$, Promethee permits the computation of the following quantities for alternatives a_i and a_k:

$$\pi(a_i,a_k) = \frac{\sum_{j=1}^{n} w_j P_j(a_i,a_k)}{\sum_{j=1}^{n} w_j} \quad (1)$$

$$\phi^+(a_i) = \sum_{a_k \in A} \pi(a_i,a_k), \quad (2)$$

$$\phi^-(a_i) = \sum_{a_k \in A} \pi(a_k,a_i),$$

$$\phi(a_i) = \phi^+(a_i) - \phi^-(a_i)$$

For each alternative a_i, belonging to the set A of alternatives, $\pi(a_i,a_k)$ is an overall preference index of a_i over a_k. The leaving flow $\phi^+(a_i)$ defines the strength of the alternative a_i, how much a_i dominates all the other alternatives of A. Symmetrically, the entering flow $\phi^-(a_i)$ defines the weakness of the alternative, how much a_i is dominated by all the other alternatives of A. $\phi(a_i)$ represents a value function, whereby a higher value reflects a higher attractiveness of alternative a_i. We call $\phi(a_i)$ the net flow of alternative a_i.

According to Promethee I, alternative a_i is better than a_k if the leaving flow of a_i ($\phi^+(a_i)$) is greater than the leaving flow of a_k ($\phi^+(a_k)$) and the entering flow of a_i ($\phi^-(a_i)$) is smaller than the entering flow of a_k ($\phi^-(a_k)$).

Equality in $\phi^+(a_i)$ and $\phi^-(a_i)$ indicates indifference between the two alternatives. In the case where the leaving flows indicate that a_i is better than a_k, while the entering flows indicate the reverse, a_i and a_k are considered incomparable. Therefore, the Promethee I provide a partial ranking of the alternatives.

In Promethee II, the net flow $\phi(a_i)$ is used in order to obtain a complete ranking of all alternatives. The alternative with the higher net flow is better.

3 Linguistic Background

Due to the fact that, our proposal consists in dealing with MCDM problems defined in multi-granular linguistic contexts that implies processes of CW, here we review briefly the 2-tuple linguistic representation model and the linguistic hierarchies structure that are necessary concepts to achieve our aim.

3.1 The 2-tuple Fuzzy Representation Model

This model was presented in [3] for overcoming the drawback of the loss of information presented by the classical linguistic computational models [12], i.e., (i) the semantic model [4], and (ii) the symbolic one [5].

The 2-tuple fuzzy linguistic representation model is based

ISBN: 978-989-95079-6-8

on the symbolic method and takes as the base of its representation the concept of Symbolic Translation.

Definition 1. *The Symbolic Translation of a linguistic term* $s_i \in S = \{s_0,...,s_g\}$ *is a numerical value assessed in [-0.5, 0.5) that supports the "difference of information" between an amount of information* $\beta \in [0, g]$ *and the closest value in {0,..., g} that indicates the index of the closest linguistic term in S (*s_i*), being [0,g] the interval of granularity of S.*

From this concept a new linguistic representation model is developed, which represents the linguistic information by means of 2-tuples (s_i,α_i), $s_i \in S$ and $\alpha_i \in$ [-0.5, 0.5).

Definition 2. *Let* $S = \{s_0,...,s_g\}$ *be a linguistic term set and* $\beta \in [0,g]$ *a value supporting the result of a symbolic aggregation operation. Then the 2-tuple that expresses the equivalent information to* β *is obtained with the following function:*

$$\Delta : [0, g] \to S \times [-0.5,.0.5) \qquad (3)$$

$$\Delta(\beta) = (s_i, \alpha), \begin{cases} s_i & i = round(\beta) \\ \alpha = \beta - i & \alpha \in [-0.5, 0.5) \end{cases}$$

where round(·) is the usual round operation, s_i has the closest index label to "β" and "α" is the value of the symbolic translation.

Proposition 1. Let $S = \{s_0,...,s_g\}$ be a linguistic term set and (s_i, α_i) be a linguistic 2-tuple. There is always a Δ^{-1} function, such that, from a 2-tuple it returns its equivalent numerical value $\beta \in$ [0, g] in the interval of granularity of *S*.

Proof. It is trivial, we consider the following function:

$$\Delta^{-1} : S \times [-0,5, 0.5) \to [0, g] \qquad (4)$$

$$\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta$$

Remark 1. From Definitions 1 and 2 and Proposition 1, it is obvious that the conversion of a linguistic term into a linguistic 2-tuple consists of adding a value 0 as symbolic translation: $s_i \in S \Rightarrow (s_i, 0)$.

This model has a computational technique based on the 2-tuples [3]:

- Aggregation of 2-tuples
  The aggregation of linguistic 2-tuples consist of obtaining a value that summarizes a set of values, therefore, the result of the aggregation of a set of 2-tuples must be a linguistic 2-tuple. We can find several 2-tuple aggregation operators in [5] based on classical aggregation operators as the arithmetic mean and weighted mean operators.
- Comparison of 2-tuples
  The comparison of information represented by 2-tuples is carried out according to an ordinary lexicographic order.
  Let (s_k, α_1) and (s_l, α_2) be two 2-tuples represented two assessments:
  - If $k < l$ then (s_k, α_1) is smaller than (s_l, α_2);
  - If $k=l$ then
    1) If $\alpha_1 = \alpha_2$ then (s_k, α_1) and (s_l, α_2) represent the same value;
    2) If $\alpha_1 < \alpha_2$ then (s_k, α_1) is smaller than (s_l, α_2);
    3) If $\alpha_1 > \alpha_2$ then (s_k, α_1) is bigger than (s_l, α_2).
- Negation Operator of a 2-tuple
  The negation operator over 2-tuples is defined as:

$$Neg(s_i, \alpha) = \Delta\ (g - \Delta^{-1}(s_i, \alpha)) \qquad (5)$$

where g+1 is the cardinality of *S*.

3.2 Linguistic Hierarchies

The Linguistic Hierarchies were introduced in [6] in order to accomplish processes of CW with multi-granular linguistic information in a precise way. A Linguistic Hierarchy is a set of levels, where each level represents a linguistic term set with different granularity to the remaining levels. Each level is denoted as *l*(*t, n*(*t*)):

- *t* a number that indicates the level of the hierarchy
- *n*(*t*) the granularity of the term set of the level *t*

We assume that levels containing linguistic terms are triangular shaped, symmetrical and uniformly distributed. In addition, the linguistic term sets have an odd number of linguistic terms being the middle one the value of indifference.

The levels belonging to a linguistic hierarchy are ordered according to their granularity, i.e., for two consecutive levels *t* and *t*+1, *n*(*t*+1) > *n*(*t*). Therefore, the level *t*+1 is a refinement of the previous level *t*.

From the above concepts, we define a linguistic hierarchy, LH, as the union of all levels *t*:

$$LH = \bigcup_t l(t, n(t)) \qquad (6)$$

Given an LH, we denote as $S^{n(t)}$ the linguistic term set of LH corresponding to the level *t* of LH characterized by a granularity of uncertainty *n*(*t*):

$$S^{n(t)} = \{s_0^{n(t)},...,s_{n(t)-1}^{n(t)}\} \qquad (7)$$

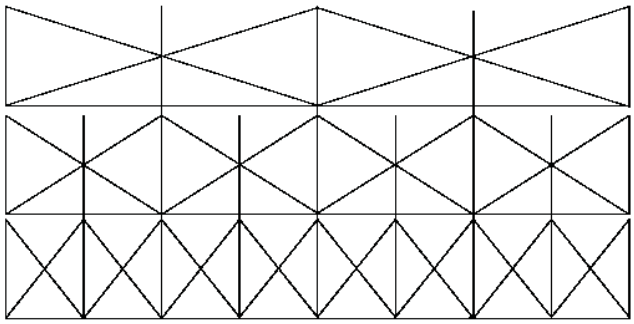
Generically, we can say that the linguistic term set of level *t* + 1 is obtained from its predecessor as:

$$l(t, n(t)) \to l(t+1, 2 \bullet n(t) - 1) \qquad (8)$$

A graphical example of a linguistic hierarchy can be seen in Figure 1.

ISBN: 978-989-95079-6-8

Fig. 1. Linguistic Hierarchy with term sets of 3,5 and 9 terms.



In [6] different transformation functions between labels of different levels were developed without loss of information. To understand how these functions are working, there were defined transformation functions between two consecutive levels and afterwards between any levels of the hierarchy, those transformation functions use the linguistic 2-tuple computational model. Here, we present the transformation function between any levels.

Definition 3. *Let* $LH = \bigcup_t l(t, n(t))$ *be a linguistic hierarchy whose linguistic term sets are denoted as* $S^{n(t)} = \{s_0^{n(t)}, ..., s_{n(t)-1}^{n(t)}\}$, *and let us consider the 2-tuple linguistic representation. The transformation function from a linguistic label in level t to a label in level t' is defined as*:

$$TF_{t'}^{t} : l(t, n(t)) \to l(t', n(t')) \quad \textbf{(9)}$$

$$TF_{t'}^{t}(s_i^{n(t)}, \alpha^{n(t)}) = \Delta_{n(t')}\left(\frac{\Delta_{n(t)}^{-1}(s_i^{n(t)}, \alpha^{n(t)}) \bullet (n(t')-1)}{n(t)-1}\right)$$

Proposition 2. *The transformation function between linguistic terms in different levels of the linguistic hierarchy is bijective:*

$$TF_{t}^{t'}(TF_{t'}^{t}(s_i^{n(t)}, \alpha^{n(t)})) = (s_i^{n(t)}, \alpha^{n(t)}) . \quad \textbf{(10)}$$

4 A Linguistic Multi-Granular Promethee model

The model developed in this paper is the result of the integration between the aggregation operators of the PROMETHEE method and the multi-granular linguistic model to combine multiple experts' assessments defining a multi-granular linguistic framework. So, each expert can express his/her evaluations in the suitable scale according to his/her knowledge in the table 1.

Table 1. Multiple experts' assessments scheme.

				Experts					
		e_1			e_l			e_E	
Alt.	c_1	c_i	c_n	c_1	c_i	c_n	c_1	c_i	c_n
a_1	C_{11}^1	C_{1j}^1	C_{1n}^1	C_{11}^2	C_{1j}^2	C_{1n}^2	C_{11}^E	C_{1j}^E	C_{1n}^E
a_j	C_{j1}^1	C_{ij}^1	C_{jn}^1	C_{j1}^2	C_{ij}^2	C_{jn}^2	C_{j1}^3	C_{ij}^3	C_{jn}^3
a_m	C_{m1}^1	C_{mj}^1	C_{mn}^1	C_{m1}^2	C_{mj}^2	C_{mn}^2	C_{m1}^3	C_{mj}^3	C_{mn}^3

Being $A = \{a_1, ..., a_j, ..., a_m\}$ a set of alternatives, $C = \{c_1, ..., c_i, ..., c_n\}$ a set of criteria, $E = \{e_1, ..., e_l, ..., e_E\}$ a set of experts. The assessments C_{ij}^e provided by the experts, e_e, can be assessed in a linguistic term sets of the linguistic hierarchy that can have different granularity of uncertainty. Therefore $C_{ij}^e \in S^e$ and $S^e \in LH$.

The proposed model to deal with MCDM problems defined in multi-granular linguistic contexts consists of three steps. Due to the use of multi-granular linguistic information, the aggregation step is divided in two steps:

- *Normalization step*. The multi-granular linguistic information is expressed in a unique linguistic expression domain.
- *Aggregation step*. The unified information expressed in a unique linguistic term set is aggregated.
- *Exploitation step*. The collective preference values are ordered in a decreasing way and the solution set is composed of the best alternative/s.

4.1 Normalization step

At the beginning, a linguistic term set to unify the multi-granular linguistic information must be selected called S_T. Any linguistic term set to do it can be chosen because the transformations between levels in a LH are bijective (see Proposition 2). In order to reduce the number of computations, the linguistic term set that the most of experts express their preferences in it shall be chosen.

Let us suppose that $S^e = S^{n(t')}$ and $S_T = S^{n(t'')}$, so a transformation function between linguistic terms in different levels of the linguistic hierarchy is obtained as follows:

$$TF_{t''}^{t'}(C_{ij}^e) = (s_{ij}^e, \alpha) . \quad \textbf{(11)}$$

with, $s_{ij}^e \in S_T$.

4.2 Aggregation step

In this step, two types of preference indexes, individual ($\pi^e(a_i, a_k)$) and collective ($\pi(a_i, a_k)$), will be computed. The individual preference index which translates the intensity of the preference of the alternative a_i compared to the alternative a_k according to the point of view of each expert is expressed as:

$$\pi^e : A \times A \to S^{n(t)} \times [-0.5, 0.5) \quad \textbf{(12)}$$

$$\pi^e(a_i, a_k) = \Delta\left[\frac{\sum_{j=1}^{n} \Delta^{-1}\left(w_j^e . P_j^e(a_i, a_k)\right)}{\Delta^{-1}(W^e)} . n(t)\right] = \left(s_i^{e,n(t)}, \alpha_i^e\right)$$

w_j^e is the weight assigned by each expert to each criterion, $W^e = \sum_{j=1}^{n} w_j^e$, $P_j^e(a_i, a_k)$ is a preference function among the six types of functions proposed in [2], a_i and a_k are two alternatives belonging to A, $n(t)$ is the cardinality of the chosen linguistic term set, $\left(s_i^{e,n(t)}, \alpha_i^e\right)$ is a linguistic

ISBN: 978-989-95079-6-8

2-tuple, $i,k \in \{1,...,m\}$ alternatives, $j \in \{1,...,n\}$ criteria, $e \in \{1,...,E\}$ experts.

Thus, the collective preference index corresponds to the aggregation of the linguistic values computed. The aggregation operator could be different in each problem. Different aggregation operators were defined in [3] to deal with linguistic 2-tuples, for the sake of simplicity in this case we have chosen the arithmetic mean. The expression of the collective preference index is:

$$\pi : A \times A \to S^{n(t)} \times [-0.5, 0.5) \quad \textbf{(13)}$$

$$\pi(a_i, a_k) = \Delta\left[\sum_{e=1}^{E} \frac{1}{E} \Delta^{-1}\left(\pi^e(a_i, a_k)\right)\right] = \left(s_i^{n(t)}, \alpha_i\right).$$

With, $E = \{e_1,...,e_E\}$ is the set of experts and $\left(s_i^{n(t)}, \alpha_i\right)$ is a 2-tuple.

4.3 Exploitation step

The exploitation step generates a solution set of alternatives (the best ones) for the decision problem. To do so, this step uses a total ranking of the alternatives in a decreasing way according to a choice function. Different choices functions have been proposed in the literature [8-10]. In this paper, a choice function that computes the dominance degree for each alternative, a_i, over the other alternatives is used as follows:

$$\Lambda(a_i) = \frac{1}{m-1} \sum_{k,i=1,k\neq i}^{m} \pi(a_i, a_k). \quad \textbf{(14)}$$

Then, the best alternative(s) are in the head of ranking should be chosen as solution set of alternatives.

5 Numerical Example

In this section, we present an investment example to show the integration between the aggregation operators of PROMETHEE method and the linguistic hierarchies.

5.1 Input Data

An investment company wants to invest a sum of money in the best option. There is a panel with four possible alternatives $A = \{a_1,...,a_4\}$ of investment possibilities. a_1 is a car industry, a_2 is a food company, a_3 is a computer company, and a_4 is an arms industry. The investment company chooses four experts $E = \{e_1,...,e_4\}$ from four consultancy departments: risk analysis, growth analysis, social-political analysis, and environmental impact analysis departments respectively, to construct a decision group throughout a set of three criteria $C = \{c_1, c_2, c_3\}$ being, c_1 profit, c_2 pollution, and c_3 employment.

These experts use different linguistic term sets from the LH (showed in Fig. 1) to provide their preferences over the alternative set as following: e_1 provides his preferences in $l(3,9)$, e_2 provides his preferences in $l(2,5)$, e_3 provides his preferences in $l(1,3)$, and e_4 provides his preferences in $l(3,9)$.

After a deep study, each expert provides the following preference values:

Table 2. Input data of each expert.

		e1			e2			e3			e4	
	c1	c2	c3	c1	c2	c3	c1	c2	c3	c1	c2	c3
a1	S_6^9	S_3^9	S_2^9	S_4^5	S_2^5	S_1^5	S_2^3	S_1^3	S_1^3	S_8^9	S_1^9	S_3^9
a2	S_7^9	S_2^9	S_4^9	S_3^5	S_3^5	S_3^5	S_2^3	S_1^3	S_1^3	S_5^9	S_2^9	S_1^9
a3	S_8^9	S_5^9	S_5^9	S_3^5	S_1^5	S_2^5	S_2^3	S_1^3	S_2^3	S_7^9	S_3^9	S_5^9
a4	S_8^9	S_6^9	S_1^9	S_4^5	S_3^5	S_2^5	S_2^3	S_2^3	S_2^3	S_8^9	S_5^9	S_2^9
wj	S_8^9	S_6^9	S_4^9	S_3^5	S_4^5	S_1^5	S_1^3	S_2^3	S_1^3	S_7^9	S_1^9	S_5^9
Type	II	III	IV	II	III	IV	II	III	IV	II	III	IV
pj	S_4^9	S_2^9	S_5^9	S_3^5	S_3^5	S_2^5	S_2^3	S_1^3	S_2^3	S_5^9	S_2^9	S_4^9
qj	S_2^9	S_1^9	S_3^9	S_1^5	S_2^5	S_1^5	S_1^3	S_0^3	S_1^3	S_3^9	S_1^9	S_3^9

5.2 Normalization Step

In this example, the linguistic term set $l(3,9)$ shall be chosen to unify the multi-granular linguistic information, since the most of experts have expressed their preferences in it.

5.3 Aggregation Step

This step is carried out by the computation of both individual and collective preference indexes. The obtained results are shown in the tables below:

Table 3. Individual preference index of each expert.

π^1	a1	a2	a3	a4
a1	-	$(S_3^9, 0)$	$(S_0^9, 0)$	$(S_0^9, 0)$
a2	$(S_1^9, 0)$	-	$(S_0^9, 0)$	$(S_0^9, 0)$
a3	$(S_8^9, 0)$	$(S_3^9, 0)$	-	$(S_1^9, 0)$
a4	$(S_7^9, 0)$	$(S_3^9, 0)$	$(S_3^9, 0)$	-

Table 4. Collective preference index.

π	a_1	a_2	a_3	a_4
a_1	-	$(S_3^9, -0.24)$	$(S_1^9, 0.4)$	$(S_0^9, 0)$
a_2	$(S_3^9, 0.27)$	-	$(S_3^9, 0.26)$	$(S_0^9, 0.25)$
a_3	$(S_2^9, 0.14)$	$(S_0^9, 0.75)$	-	$(S_3^9, -0.27)$
a_4	$(S_4^9, 0.12)$	$(S_4^9, 0.5)$	$(S_2^9, 0.33)$	-

ISBN: 978-989-95079-6-8

5.4 Exploitation Step

This step provides a total ranking of the alternatives in a decreasing way (Table 6) according to a choice function (Table 5).

Table 5. Dominance degree.

$\Lambda(a_1)=(S_1^9,0.04)$	$\Lambda(a_2)=(S_2^9,-0.3)$
$\Lambda(a_3)=(S_1^9,0.4)$	$\Lambda(a_4)=(S_3^9,-0.3)$

Where the degrees are computed as:

$$\Lambda(a_1)=\frac{1}{3}\sum_{k,i=1,k\neq 1}^{3}\pi(a_1,a_k)=\frac{1}{3}\left(\pi(a_1,a_2)+\pi(a_1,a_3)+\pi(a_1,a_4)\right)$$
$$=\frac{1}{3}\left(\left(S_3^9,-0.24\right)+\left(S_1^9,0.4\right)+\left(S_0^9,0\right)\right)$$
$$=\left(S_1^9,0.4\right)$$

Table 6. Alternatives ranking.

a_4	a_2	a_3	a_1

According to the ranking of the alternatives, the company should choose the alternative, a_4, for its investment.

6 Conclusions

In order to manage multi-granular linguistic information in MCDM problems, we extended aggregation operators of PROMETHEE method for combining the linguistic values by the direct computation on labels. In this paper, a multi-criteria, multi-expert method has been presented to obtain the overall linguistic value without loss of information, taking into account the particular nature of the criteria and the specific differences among the experts through the aggregation process. The proposed model is computationally simple and quick.

Acknowledgments

This paper has been partially supported by the R&D project TIN2006-02121, P08-TIC-3548 and FEDER funds.

References

[1] Maystere, L.Y., Pictet, J., Simos, J.: Méthodes Multicritères ELECTRE Lausanne. Presses Polytechniques et Universitaires, Romandes (1994)

[2] Brans, J.P., Mareschal, B., Vincke, P.: PROMETHEE: A New Family of Outranking Methods in Multicriteria Analysis. In: Brans (eds.): Operational Research'84, Elsevier Science Publishers, North Holland, (1984) 408-421

[3] Herrera, F., Martinez, L.: A 2-Tuple Fuzzy Linguistic Representation Model For Computing with Words. IEEE Transactions on Fuzzy Systems, Vol. 8. (2000) 746-752

[4] Degani, R., Bortolan, G.: The Problem of Linguistic Approximation in Clinical Decision Making. International Journal of Approximate Reasoning, Vol. 2. (1988) 143-162

[5] Delgado, M., Verdegay, J.L., Vila, M.A.: On Aggregation Operations of Linguistic Labels. International Journal of Intelligent Systems, Vol. 2. (1993) 351-370

[6] Herrera, F., Martinez, L.: A Model Based on Linguistic 2-Tuples for Dealing with Multigranularity Hierarchical Linguistic Contexts in Multiexpert Decision-Making. IEEE Transactions on Systems, Man and Cybernetics, Vol. 31. (2001) 33-49

[7] Herrera, F., Martinez, L., Sanchez, P.J. : Managing Non-Homogeneous Information in Group Decision Making. European Journal of Operational Research, Vol. 166. (2005) 115-132

[8] Arrow, K.J.: Social Choice and Individual Values. Yale University Press, New Haven, CT (1963)

[9] Orlovsky, S.A.: Decision-Making with a Fuzzy Preference Relation. Fuzzy Sets and Systems, Vol. 1. (1978) 155-167

[10] Roubens, M.: Some Properties of Choice Functions Based on Valued Binary Relations. European Journal of Operational Research, Vol. 40. (1989) 309-321.

[11] L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning. Part I, Information Sciences 8 (1975) 199-249, Part II, Information Sciences 8 (1975) 301-357, Part III, Information Sciences 9 (1975) 43-80.

[12] F. Herrera, L. Martínez, The 2-tuple Linguistic Computational Model. Advantages of its linguistic description, accuracy and consistency. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2001, Vol 9 pp. 33-48.

ISBN: 978-989-95079-6-8

Generating Minimum Dispersion Densities from an Interval-Valued Fuzzy Set

Carol L. Walker[1] Elbert A. Walker[1] Ronald R. Yager[2]

1. Department of Mathematical Sciences
Las Cruces, New Mexico, USA
2. Machine Intelligence Institute, Iona College
New Rochelle, New York, USA
Email: hardy@nmsu.edu, elbert@nmsu.edu, yager@panix.com

***Abstract*—Let f be an interval-valued fuzzy subset of a finite set U of size n. This yields n closed intervals inside the unit interval. Picking a point in each interval and dividing by the sum of the points gives rise to a probability density on the set of intervals. For a given measure of dispersion, such as entropy, the problem is to pick points that maximize (or minimize) this dispersion. The particular case considered here is to pick points that minimize the sum of the squares of probabilities in the density. We provide an algorithm for picking such points.**

***Keywords*—interval-valued fuzzy set; minimum dispersion density; probability density; sum of squares**

I. Introduction

A problem that arises in several situations is this. Given a family of probability densities, choose one from that family that has the largest value for some measure of dispersion. Some instances involving entropy are discussed in [1], [2], [3], [4], [5], [6], [7]. But there are other measures of dispersion, for which a general formulation is provided in [8].

The problem we consider is, given n subintervals $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_n, b_n]$ of $[0, 1]$, choose $x_i \in [a_i, b_i]$ such that for an n-tuple $\mathbf{x} = (x_1, ..., x_n)$ the value

$$H(\mathbf{x}) = \sum_j \left(\frac{x_j}{\sum_i x_i} \right)^2 \tag{1}$$

is minimum. This means that among all the possible choices $\mathbf{x}$ for the n-tuple of x_i's, this quantity is minimum.

One motivation for this problem is the following. Let U be a finite set and f a function from U into the set of closed intervals of $[0, 1]$. That is, $f : U \to \{[a, b] : 0 \leq a \leq b \leq 1\}$. This yields a finite set of closed intervals $\{[a_i, b_i] : i = 1, 2, \ldots, n\}$. If not all the $a_i = 0$, than choosing $x_i \in [a_i, b_i]$ yields a probability density $P([a_i, b_i]) = x_i / \Sigma x_j$ on $\{[a_i, b_i] : i = 1, 2, \ldots, n\}$. Maximizing, (or minimizing) a dispersion measure gives a canonical way to pick out a point x_i in each interval $[a_i, b_i]$, a possible first step toward defuzzifying the interval-valued fuzzy set f.

Several remarks are in order.

1) It should be noted that the intervals do not have to be distinct, and that an interval can be a single point. Intervals such as $[0, 0]$ and $[0, a]$ are allowed, as are $[1, 1]$ and $[b, 1]$. And of course, an interval may be contained in another.
2) If all the intervals are $[0, 0]$, then each $x_i = 0$ and each probability is $\frac{0}{0}$, which is undefined so *we assume this is not the case*. Thus at least one of the intervals is not equal to $[0, 0]$, and there is at least one $S = \sum_{i=1}^n x_i$ that is positive. In this case, the interval $[0, 0]$ has associated probability 0, and its contribution to the sum of the squares is 0. Thus, in developing an algorithm for finding the x_i's that give minimum dispersion, *we can assume that no interval is* $[0, 0]$, that is, $b_i > 0$ for all i.
3) If the intersection of the intervals is non-empty, then choosing all the x_i to be any point $x_i = x > 0$ in that intersection yields minimum dispersion, namely
$$\sum_{j=1}^n \left(\frac{x}{\sum_{i=1}^n x} \right)^2 = \frac{1}{n} \tag{2}$$
In particular, the solution may not be unique. In fact, in this case, if the intersection is not a single point, then there are uncountably many solutions, namely any positive point in the intersection. That $\frac{1}{n}$ is the minimum possible dispersion is a special case of the Cauchy-Schwarz inequality, which says that
$$\left(\sum_{i=1}^n x_i y_i \right)^2 \leq \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i^2 \right) \tag{3}$$
Letting each $y_i = 1$, we get
$$\left(\sum_{i=1}^n x_i \right)^2 \leq \left(\sum_{i=1}^n x_i^2 \right) n \tag{4}$$
and
$$\frac{1}{n} \leq \sum_{j=1}^n \left(\frac{x_j}{(\sum_{i=1}^n x_i)} \right)^2 \tag{5}$$
4) If $a_i = 0$ for all i, then since we are assuming $b_i > 0$ for all i, the intersection of the intervals contains a point $x > 0$, and we are in case 3 above. In particular, we can assume that $a_i > 0$ for some i. In this situation, $S = \sum x_i$ is always positive, and *there are x_i that minimize*

ISBN: 978-989-95079-6-8

$H(\mathbf{x}) = \sum_{j=1}^{n}\left(\frac{x_j}{\sum_{i=1}^{n} x_i}\right)^2$ since H is a continuous function on the compact space $\prod_{i=1}^{n}[a_i, b_i]$.

5) An x_i does not have to be an end-point of its interval. For example, if $n = 3$, and the three intervals are disjoint, then there will be a unique solution, namely the right endpoint of the left most interval, the left endpoint of the right most interval, and generally some interior point of the middle interval. We will see an explicit example of this below.

II. A Reformulation

Based on the discussion above, we are in the following situation. We have a finite set $[a_1, b_1], [a_2, b_2], ..., [a_n, b_n]$ of closed subintervals of the unit interval $[0, 1]$, all the b_i are positive, and at least one a_i is positive. We wish to choose $\mathbf{x} = (x_1, x_2, ..., x_n)$ with each $x_i \in [a_i, b_i]$ such that $H(\mathbf{x}) = \sum_{j=1}^{n}\left(\frac{x_j}{\sum_{i=1}^{n} x_i}\right)^2$ is minimum. In fact, we will provide an algorithm that will give an exact solution in terms of the end points of the intervals $[a_i, b_i]$. First we prove a lemma that provides a reformulation into a somewhat simpler problem. We term $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ a **solution** if $H(\mathbf{x})$ is minimum.

Lemma 1: Let $x_i \in [a_i, b_i]$ with $x_1 \leq x_2 \leq \cdots \leq x_k < x_{k+1} \leq \cdots \leq x_n$, and let $S = \sum_{i=1}^{n} x_i$. Then there exists $h > 0$ such that

$$\left(\frac{x_1 + h}{S + h}\right)^2 + \sum_{j=2}^{n}\left(\frac{x_j}{S + h}\right)^2 < \sum_{j=1}^{n}\left(\frac{x_j}{S}\right)^2 \tag{6}$$

Proof: Letting $f(h) = \left(\frac{x_1+h}{S+h}\right)^2 + \sum_{j=2}^{n}\left(\frac{x_j}{S+h}\right)^2$, we have

$$\begin{aligned} f'(h) &= 2\left(\frac{x_1 + h}{S + h}\right)\left(\frac{S + h - (x_1 + h}{(S + h)^2}\right) \\ &\quad + \sum_{j=2}^{n} 2\left(\frac{x_j}{S + h}\right)\left(\frac{-x_j}{(S + h)^2}\right) \end{aligned} \tag{7}$$

Now $f'(h)$ is negative if and only if

$$\begin{aligned} &(x_1 + h)(S - x_1) - \sum_{j=2}^{n} x_j^2 \\ &= (x_1 + h)\left(\sum_{j=2}^{n} x_j\right) - \sum_{j=2}^{n} x_j^2 \\ &= \left(x_1 \sum_{j=2}^{n} x_j\right) - \sum_{j=2}^{n} x_j^2 + h\sum_{j=2}^{n} x_j \end{aligned} \tag{8}$$

is negative. Since $x_1 < x_{k+1}$, $\left(x_1 \sum_{j=2}^{n} x_j\right) - \sum_{j=2}^{n} x_j^2$ is negative, and hence h can be chosen positive and small enough so that $f'(h)$ is negative. The lemma follows. ■

If $x_1 < b_1$, then h can be chosen so that also $x_1 + h \in [a_1, b_1]$. Thus we have the following.

Corollary 2: If $x_i \in [a_i, b_i]$ such that $x_1 \leq x_2 \leq ... \leq x_k < x_{k+1} \leq ... \leq x_n$ is a solution, then $x_1 = b_1$.

Note that the proof uses the fact that x_1 is the smallest x_i. If $x_2 = x_1$, then of course $x_2 = b_2$.

For intervals $[a_i, b_i]$, let $R = \min\{b_i\}_{i=1}^{n}$ and $L = \max\{a_i\}_{i=1}^{n}$. Note that if $L \leq R$, then $[L, R]$ is the intersection of the intervals $[a_i, b_i]$, and picking for each x_i the same element x in $[L, R]$ yields a solution $\mathbf{x}$ with $H(\mathbf{x}) = \frac{1}{n}$, the minimum possible value. Also, from the lemma above, we see that for $\mathbf{x}$ with at least two distinct entries, $H(\mathbf{x}) > \frac{1}{n}$. Thus *we may assume that* $R < L$. Otherwise the problem is trivial. If $R < L$, then any $\mathbf{x}$ has at least two distinct entries, and by our corollary, all $x_i \geq b_1 = R$. Similarly, we may show that all x_i are $\leq L$. This means that the intervals $[a_i, b_i]$ may be replaced by the intervals $[a_i \vee R, b_i \wedge L]$. This turns out to be a significant reduction of the original problem. For example, it already provides us with two elements of a solution $\mathbf{x}$. The smallest and the largest entry in any solution will be R and L, respectively, since two intervals are $[R, R]$ and $[L, L]$. Here is our reformulation.

Reformulation: Let $0 < R < L \leq 1$, and let $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_n, b_n]$ be subintervals of $[R, L]$ with $a_1 = b_1 = R$ and $a_n = b_n = L$. Find $x_i \in [a_i, b_i]$ which minimizes $\sum_{i=1}^{n}\left(\frac{x_i}{\sum_{j=1}^{n} x_j}\right)^2$.

The reformulated problem has the same solution (or perhaps solutions) as the original problem. In the subintervals of $[R, L]$, one is $[R, R]$, and one is $[L, L]$. Call these $[a_1, b_1]$ and $[a_n, b_n]$ respectively. So $x_1 = a_1 = b_1$, and $x_n = a_n = b_n$, and we have two of the required x_i immediately. We assume from now on that we are in this reformulated situation. There are some additional lemmas we need.

Lemma 3: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution and if $a_k < b_k = L$, then $x_k < b_k$. Similarly, if $R = a_k < b_k$, then $a_k < x_k$.

Proof: Of course, $k < n$. Let $x_i \in [a_i, b_i]$, $i = 1, 2, \ldots, n$, and suppose that $a_k < b_k = b_n = x_k = L$. Let $S = \sum_{j=1}^{n} x_j$ and consider

$$f(h) = \sum_{i=1, i\neq k}^{n}\left(\frac{x_i}{S - h}\right)^2 + \left(\frac{x_k - h}{S - h}\right)^2 \tag{9}$$

Then a calculation shows that

$$f'(h) = \frac{2}{(S - h)^3}\left(\sum_{\substack{i=1\\ i\neq k}}^{n} x_i^2 - x_k \sum_{\substack{i=1\\ i\neq k}}^{n} x_i + h\sum_{\substack{i=1\\ i\neq k}}^{n} x_i\right) \tag{10}$$

If $h = 0$, then $f'(h) < 0$ since $x_i \leq x_k$ for all i, and $x_1 = R < x_k$. For h sufficiently small and positive, $f'(h) < 0$. A similar proof works for $R = a_k < b_k$. ■

Lemma 4: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, then for no two distinct entries $x_i < x_j$ in $\mathbf{x}$ can it be that $a_i < b_i$, $a_j < b_j$, and $x_i \in [a_i, b_i)$ and $x_j \in (a_j, b_j]$.

Proof: Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, suppose that $x_1 < x_2$, and suppose that $x_1 \in [a_1, b_1)$ and $x_2 \in (a_2, b_2]$. Consider the density given by $(x_1 + h, x_2 - h, x_3, \ldots, x_n)$. Since $x_1 + h +$

ISBN: 978-989-95079-6-8

$x_2 - h + \sum_{i=3}^{n} x_i = \sum_{i=1}^{n} x_i = S$, we have

$$f(h) = \left(\frac{x_1+h}{S}\right)^2 + \left(\frac{x_2-h}{S}\right)^2 + \sum_{j=3}^{n}\left(\frac{x_j}{S}\right)^2 \quad (11)$$

and its derivative with respect to h is

$$f'(h) = \frac{2}{S}\left(\frac{x_1+h}{S}\right) - \frac{2}{S}\left(\frac{x_2+h}{S}\right) = \frac{2}{S^2}(x_1 - x_2) \quad (12)$$

This latter quantity is negative whenever $x_2 - h > x_1 + h$. Since $x_1 \in [a_i, b_i)$ and $x_2 \in (a_2, b_2]$ we can so choose h, keeping x_1 in $[a_1, b_1]$ and x_2 in $[a_2, b_2]$, and decreasing H. Thus $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ cannot be a solution. ■

This lemma just says that if $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, and $x_i < x_j$, then x_i cannot be moved to the right and x_j to the left, keeping them in their original intervals, and keeping $x_i < x_j$. That is, if $x_i < x_j$ is part of a solution, then either $x_i = b_i$ or $x_j = a_j$.

Corollary 5: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, then no two distinct entries in $\mathbf{x}$ can be in the interior of their intervals.

Corollary 6: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, and x_i is in the interior of its interval, then for every $x_j < x_i$, $x_j = b_j$ and for every $x_j > x_i$, $x_j = a_j$.

As mentioned earlier, the function

$$H(\mathbf{x}) = \sum_{j=1}^{n}\left(\frac{x_j}{\sum_{i=1}^{n} x_i}\right)^2 \quad (13)$$

assumes a minimum since it is continuous on a compact space. Whatever that minimum, it consists of some endpoints of some of the intervals $[a_i, b_i]$ and interior points of the rest. But this corollary says that all those interior points must be equal. *So any solution consists of endpoints of some of the intervals, and a common interior point of the remainder of the intervals.*

We need one more lemma.

Lemma 7: Let $0 < k < n$. Suppose that $[a, b]$ is the intersection of the intervals $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_k, b_k]$, $a < b$, and $x_i \in [a_i, b_i]$, $i = k+1, \ldots, n$. Letting $S = kx + \sum_{i=k+1}^{n} x_i$, an $x \in [a, b]$ that minimizes

$$f(x) = \left(k\left(\frac{x}{S}\right)^2 + \sum_{i=k+1}^{n}\left(\frac{x_j}{S}\right)^2\right) \quad (14)$$

is either an endpoint a or b, or an interior point x such that $f'(x) = 0$. The solution to $f'(x) = 0$ is the point

$$x_0 = \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i} \quad (15)$$

which may or may not be in the interval $[a, b]$. If $x_0 < a$, then f assumes its minimum on $[a, b]$ at a. If $x_0 > b$, then f assumes its minimum on $[a, b]$ at b. Otherwise, f assumes its minimum on $[a, b]$ at x_0.

Proof: The function f has a minimum at a, b, or at a point in the interior of the interval $[a, b]$ where the derivative of f is zero. Since the derivative of S is k, the derivative of $f(x)$ with respect to x is

$$\begin{aligned} f'(x) &= 2k\left(\frac{x}{S}\right)\left(\frac{S-kx}{S^2}\right) - 2k\sum_{i=k+1}^{n}\left(\frac{x_i}{S}\right)\left(\frac{x_i}{S^2}\right) \quad (16)\\ &= 2k\left(\frac{x}{S}\right)\left(\frac{S-kx}{S^2}\right) - 2k\sum_{i=k+1}^{n}\left(\frac{x_i}{S}\right)\left(\frac{x_i}{S^2}\right)\\ &= 2k\left(\frac{x}{S}\right)\left(\frac{\sum_{i=k+1}^{n} x_i}{S^2}\right) - 2k\sum_{i=k+1}^{n}\left(\frac{x_i}{S}\right)\left(\frac{x_i}{S^2}\right)\\ &= \left(\frac{2k}{S^3}\right)\left(x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2\right) \end{aligned}$$

This implies that $f'(x) = 0$ only when

$$x = \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i} \quad (17)$$

The sign of the derivative $f'(x)$ is the same as the sign of $x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2$. If $x < \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$, then

$$\begin{aligned} x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2 &< \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2\\ &= \sum_{i=k+1}^{n} x_i^2 - \sum_{i=k+1}^{n} x_i^2 = 0 \quad (18) \end{aligned}$$

Similarly, if $x > \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$, then $x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2 > 0$. Thus, $f(x)$ is decreasing when $x < \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$, increasing when $x > \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$,. and reaches its minimum at $x = \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$. The lemma follows. ■

Note that this lemma only minimizes H *given the* x_i, $i = k+1, k+2, \ldots, n$, and note that we have assumed that $0 < k < n$.

III. An Algorithm

Corollary 5 says that any solution is an endpoint of some set of intervals and a common interior point of the intersection of the rest of the intervals. So we search out all such situations and choose the one with minimum H. The goal is to search out all such situations efficiently. Here is an algorithm that yields a solution $\mathbf{x}$ with $H(\mathbf{x})$ minimum. In the next section, we give some examples illustrating this algorithm.

An Algorithm:

1) If $\cap_{i=1}^{n}[a_i, b_i] = [a, b] \neq \varnothing$, then let each x_i be any element in $[a, b]$.
2) If $\cap_{i=1}^{n}[a_i, b_i] = \varnothing$, replace $[a_i, b_i]$ by $[a_i \vee R, b_i \wedge L]$, and call these new intervals again $[a_i, b_i]$. (This is the reformulation mentioned earlier, which says we may assume $0 < R < L \leq 1$ with $[a_1, b_1]$, $[a_2, b_2]$, ...,

ISBN: 978-989-95079-6-8

$[a_n, b_n]$ subintervals of $[R, L]$, with $a_1 = b_1 = R$, and $a_n = b_n = L$.)

3) Find the intersection of all distinct pairs of intervals from $[a_i, b_i]$, $i = 2, 3, ..., n-1$, such that these intersections are proper closed intervals; that is, are of the form $[a, b]$ with $a < b$. Let $\mathcal{I}$ be the set of these intersections together with all the proper intervals $[a_i, b_i]$. (The intersection of any nonempty family $\mathcal{S}$ of the intervals $[a_i, b_i]$ is either a member of $\mathcal{S}$ or can be obtained as the intersection of two members of $\mathcal{S}$.)
4) For each interval $I \in \mathcal{I}$, form the family $\mathcal{S}_I$ of all intervals $[a_i, b_i]$ containing I, so for each intersection $[a, b]$, $\mathcal{S}_{[a,b]}$ is the family of all the intervals in $\{[a_i, b_i] : i = 2, 3, ..., n-1\}$ containing $[a, b]$. Note that $\mathcal{S}_{[a,b]}$ has a multiplicity, namely the number of (not necessarily distinct) intervals containing $[a, b]$.
5) For those intervals $[a_i, b_i]$ not in $\mathcal{S}_I$ and such that $a_i < b_i$, choose an endpoint in the interior of $[R, L]$. If $a_i < a_j < b_i < b_j$, do not choose a_j and b_i, as such endpoints cannot be part of a solution minimizing H by Lemma 4. If no possible choices of endpoints are possible, proceed to another $\mathcal{S}_I$. Choose all endpoints that are of the form $[a, a]$, which include $[a_1, b_1]$ and $[a_n, b_n]$. This gets for each $[a_i, b_i] \notin \mathcal{S}_I$, an endpoint x_i, except for those I for which no proper selection of endpoints is possible.
6) For each $I \in \mathcal{I}$, with each possible set of endpoints gotten from step 5, use Lemma 7 to calculate the resulting H. There may be several sets of possible endpoints for a given $I \in \mathcal{I}$.
7) Repeat for all $I \in \mathcal{I}$. Now choose from the resulting candidates the one that gives minimum H.

Note that the resulting solution given by this algorithm is an exact expression in the endpoints of the given intervals. For all we know, there may be more than one solution, even when $R < L$, but we suspect not. And, of course, there may be a more efficient algorithm.

IV. Examples

We give three examples illustrating the algorithm. The first illustrates the simplest nontrivial case.

Example 8: Consider the three intervals $I_1 = [0.25, 0.33]$, $I_2 = [0.5, 0.67]$, $I_3 = [0.75, 1]$.

1) The intersection of the intervals is empty.
2) $R = 0.33$ and $L = 0.75$. Replacing the intervals by $[a_i \vee R, b_i \wedge L]$ yields the intervals

$$\begin{aligned} &[0.33, 0.33] \\ &[0.5, 0.67] \\ &[0.75, 0.75] \end{aligned} \tag{19}$$

3) $\mathcal{I}$ has only one element, namely $[0.5, 0.67]$.
4) The only interval containing $[0.5, 0.67]$ is $[0.5, 0.67]$ itself.
5) The only endpoints to be chosen are 0.33 and 0.75.
6) Lemma 7 says that the minimum H is the minimum value of

$$\left(\frac{0.33}{1.08 + x}\right)^2 + \left(\frac{x}{1.08 + x}\right)^2 + \left(\frac{0.75}{1.08 + x}\right)^2 \tag{20}$$

for $x \in [0.5, 0.67]$. Setting the derivative equal to 0 gives

$$x = \frac{0.33^2 + 0.75^2}{0.33 + 0.75} = 0.621\,67 \tag{21}$$

which is in $[0.5, 0.67]$. So by Lemma 7 the solution is $\mathbf{x} = (0.33, 0.621\,67, 0.75)$. Calculating, we get $H(0.33, 0.621\,67, 0.75) = 0.365\,33$. Notice that $0.36533 > \frac{1}{n} = \frac{1}{3}$.
7) There was only one I to deal with.

Example 9: Consider the four intervals $I_1 = [0.25, 0.33]$, $I_2 = [0.5, 0.67]$, $I_3 = [0.69, 0.72]$, $I_4 = [0.75, 1]$.

1) The intersection of the intervals is empty.
2) $R = 0.33$ and $L = 0.75$. Replacing the intervals by $[a_i \vee R, b_i \wedge L]$ yields the intervals

$$\begin{aligned} &[0.33, 0.33] \\ &[0.5, 0.67] \\ &[0.69, 0.72] \\ &[0.75, 0.75] \end{aligned} \tag{22}$$

3) $\mathcal{I}$ has two elements, namely $[0.5, 0.67]$ and $[0.69, 0.72]$.
4) The only interval containing $[0.5, 0.67]$ is $[0.5, 0.67]$ itself and the only endpoints to choose with this interval are 0.33, 0.69, and 0.75.
5) Lemma 7 says that the minimum of H for $[0.5, 0.67]$ occurs for

$$x_0 = \frac{0.33^2 + 0.69^2 + 0.75^2}{0.33 + 0.69 + 0.75} = 0.648\,305 \tag{23}$$

since $0.648\,305$ is in $[0.5, 0.67]$. So by Lemma 7 the minimum solution for $[0.5, 0.67]$ occurs for $\mathbf{x} = (0.33, 0.648\,305, 0.69, 0.75)$. Calculating, we get $H(0.33, 0.648\,305, 0.69, 0.75) = 0.268\,082$.
6) The only interval containing $[0.69, 0.72]$ is $[0.69, 0.72]$ itself and the only endpoints to choose with this interval are 0.33, 0.67, and 0.75.
7) Lemma 7 says that the minimum H for this interval occurs for

$$x_0 = \frac{0.33^2 + 0.67^2 + 0.75^2}{0.33 + 0.67 + 0.75} = 0.640\,171 \tag{24}$$

if this is in the interval, but it is to the left of the interval, and thus by Lemma 7 the minimum occurs for $x = 0.69$. So the solution is $\mathbf{x} = (0.33, 0.67, 0.69, 0.75)$. Calculating, we get $H(0.33, 0.67, 0.69, 0.75) = 0.268\,14$.
8) Notice that $0.268\,14 > 0.268\,082 > \frac{1}{n} = 0.25$. There are no more intervals to check, so the minimum is $H(\mathbf{x}) = 0.268\,082$, which occurs for $\mathbf{x} = (0.33, 0.648\,305, 0.69, 0.75)$.

From the two preceding examples, it is easy to see what happens for any family of *mutually disjoint* intervals: for each

ISBN: 978-989-95079-6-8

proper interval $[a_i, b_i]$, choose all the right endpoints to the left of the interval and all the left endpoints to the right of the interval. Use the point in $[a_i, b_i]$ determined by Lemma 7 and compute the resulting $H(\mathbf{x})$. Then compare these and find the minimum.

In the case the intervals are not all mutually disjoint, the situation is a bit more complicated and the more subtle points in the algorithm come into play. The next example is more illustrative of the algorithm.

Example 10: Consider the seven intervals

$$\begin{aligned} J_1 &= [0.2, 0.3] \\ J_2 &= [0.35, 0.5] \\ J_3 &= [0.15, 0.375] \\ J_4 &= [0.1, 0.3] \\ J_5 &= [0.35, 0.7] \\ J_6 &= [0.25, 0.5] \\ J_7 &= [0.4, 0.8] \end{aligned} \tag{25}$$

1) The intersection of the intervals is empty.
2) $R = 0.3$ and $L = 0.4$. Replacing $[a_i, b_i]$ by $[a_i \vee R, b_i \wedge L]$ yields the intervals

$$\begin{aligned} I_1 &= [0.3, 0.3] \\ I_2 &= [0.35, 0.4] \\ I_3 &= [0.3, 0.375] \\ I_4 &= [0.3, 0.3] \\ I_5 &= [0.35, 0.4] \\ I_6 &= [0.3, 0.4] \\ I_7 &= [0.4, 0.4] \end{aligned} \tag{26}$$

3) We get the new interval $I_8 = [0.35, 0.375]$, and $\mathcal{I} = \{I_2, I_3, I_6, I_8\}$. ($I_2 = I_5$, so we need not list I_5.)
4) For each interval $I \in \mathcal{I}$, forming the family $\mathcal{S}_I$ of all intervals containing I yields
   a) $\mathcal{S}_{I_2} = \{I_2, I_5, I_6\}$
   b) $\mathcal{S}_{I_3} = \{I_3, I_6\}$
   c) $\mathcal{S}_{I_6} = \{I_6\}$
   d) $\mathcal{S}_{I_8} = \{I_2, I_3, I_5, I_6\}$
5) For each of these four families there happens to be only one choice for endpoints for each of the intervals not in $\mathcal{S}_I$.
   a) $\mathcal{S}_{I_2} = \{I_2, I_5, I_6\}$ pairs with $\mathcal{P}_{I_2} = \{P_1 = 0.3,\ P_3 = 0.375,\ P_4 = 0.3,\ P_7 = 0.4\}$
   b) $\mathcal{S}_{I_3} = \{I_3, I_6\}$ pairs with $\mathcal{P}_{I_3} = \{P_1 = 0.3,\ P_2 = 0.35,\ P_4 = 0.3,\ P_5 = 0.35,\ P_7 = 0.4\}$
   c) $\mathcal{S}_{I_6} = \{I_6\}$ pairs with $\mathcal{P}_{I_6} = \{P_1 = 0.3,\ P_2 = 0.35,\ P_3 = 0.375,\ P_4 = 0.3,\ P_5 = 0.35,\ P_7 = 0.4\}$. But by Lemma 4, the endpoints P_2 and P_3 cannot both be part of a solution, so we discard this option.
   d) $\mathcal{S}_{I_8} = \{I_2, I_3, I_5, I_6\}$ pairs with $\mathcal{P}_{I_8} = \{P_1 = 0.3,\ P_4 = 0.3,\ P_7 = 0.4\}$
6) For each of these three remaining options $\mathcal{S}_{I_2}$, $\mathcal{S}_{I_3}$, and $\mathcal{S}_{I_8}$ we compute the values

$$f_I(x) = k_I \left(\frac{x}{S_I}\right)^2 + \sum_{x_j \in \mathcal{P}_I} \left(\frac{x_j}{S_I}\right)^2 \tag{27}$$

where k_I is the number of intervals in $\mathcal{S}_I$, $S_I = k_I x + \sum_{x_j \in \mathcal{P}_I} x_j$, and x is either an endpoint of the intersection of $\mathcal{S}_I$ that lies in the interior of $[R, L]$, or

$$x = \frac{\sum_{x_i \in \mathcal{P}_I} x_i^2}{\sum_{x_i \in \mathcal{P}_I} x_i}. \tag{28}$$

   a) For $\mathcal{S}_{I_2} = \{I_2, I_5, I_6\}$ we have $I_2 = [0.35, 0.4]$ and $\mathcal{P}_{I_2} = \{0.3, 0.375, 0.3, 0.4\}$. Then

$$\begin{aligned} x &= \frac{0.3^2 + 0.375^2 + 0.3^2 + 0.4^2}{0.3 + 0.375 + 0.3 + 0.4} \\ &= 0.349\,55 \notin [0.35, 0.4] \end{aligned} \tag{29}$$

   By Lemma 7, since $0.349\,55 < 0.35$, on $I_2 = [0.35, 0.4]$, the minimum is assumed at 0.35. We have

$$3\,(0.35) + 0.3 + 0.375 + 0.4 = 2.\,125 \tag{30}$$

   and so the value of H for this set of points is

$$\begin{aligned} f_{I_2}(0.35) &= 3\left(\frac{0.35}{2.\,125}\right)^2 + \left(\frac{0.3}{2.\,125}\right)^2 \\ &+ \left(\frac{0.375}{2.\,125}\right)^2 + \left(\frac{0.4}{2.\,125}\right)^2 \\ &= 0.167\,89 \end{aligned} \tag{31}$$

   b) For $\mathcal{S}_{I_3} = \{I_3, I_6\}$, we have $I_3 \cap I_6 = [0.3, 0.375]$ and $\mathcal{P}_{I_3} = \{0.3, 0.35, 0.3, 0.35, 0.4\}$, and

$$\begin{aligned} x &= \frac{0.3^2 + 0.35^2 + 0.3^2 + 0.35^2 + 0.4^2}{0.3 + 0.35 + 0.3 + 0.35 + 0.4} \\ &= 0.344\,12 \in [0.3, 0.375] \end{aligned} \tag{32}$$

   By Lemma 7, since $0.344\,12 \in [0.3, 0.375]$, on $[0.3, 0.375]$, the minimum is assumed at $0.344\,12$.which has multiplicity 2. We have

$$2\,(0.344\,12) + 0.3 + 0.35 + 0.3 + 0.35 + 0.4 = 2.\,388\,2 \tag{33}$$

   and so the value of H for this set of points is

$$\begin{aligned} f_{I_3}(0.344\,12) &= 2\left(\frac{0.344\,12}{2.\,388\,2}\right)^2 + \left(\frac{0.3}{2.\,388\,2}\right)^2 \\ &+ \left(\frac{0.35}{2.\,388\,2}\right)^2 + \left(\frac{0.3}{2.\,388\,2}\right)^2 \\ &+ \left(\frac{0.35}{2.\,388\,2}\right)^2 + \left(\frac{0.4}{2.\,388\,2}\right)^2 \\ &= 0.144\,09 \end{aligned} \tag{34}$$

ISBN: 978-989-95079-6-8

d. For $\mathcal{S}_{I_8} = \{I_2, I_3, I_5, I_6\}$ we have $I_2 \cap I_3 \cap I_4 \cap I_6 = [0.35, 0.375]$ and $\mathcal{P}_{I_8} = \{0.3, 0.3, 0.4\}$, and

$$x = \frac{0.3^2 + 0.3^2 + 0.4^2}{0.3 + 0.3 + 0.4} = 0.34 \notin [0.35, 0.375] \tag{35}$$

By Lemma 7, since $0.34 < 0.35$, the minimum is assumed at 0.35,which have multiplicity 4. We have

$$4\,(0.35) + 0.3 + 0.3 + 0.4 = 2.4 \tag{36}$$

and so the value of H for this set of points is

$$\begin{aligned} f_{I_8}(0.35) &= 4\left(\frac{0.35}{2.4}\right)^2 + \left(\frac{0.3}{2.4}\right)^2 \\ &+ \left(\frac{0.3}{2.4}\right)^2 + \left(\frac{0.4}{2.4}\right)^2 \\ &= 0.144\,10 \end{aligned} \tag{37}$$

Note that the values computed are all $> \frac{1}{n} = \frac{1}{7}$, as they should be. Comparing the values 1.6789, 1.4409, and 1.4410, we see that H assumes its minimum as indicated below.

$$\begin{aligned} 0.3 &\in [0.3, 0.3] \\ 0.35 &\in [0.35, 0.4] \\ 0.344\,12 &\in [0.3, 0.375] \\ 0.3 &\in [0.3, 0.3] \\ 0.35 &\in [0.35, 0.4] \\ 0.344\,12 &\in [0.3, 0.4] \\ 0.4 &\in [0.4, 0.4] \end{aligned} \tag{38}$$

In terms of the original intervals, we have

$$\begin{aligned} 0.3 &\in [0.2, 0.3] \\ 0.35 &\in [0.35, 0.5] \\ 0.344\,12 &\in [0.15, 0.375] \\ 0.3 &\in [0.1, 0.3] \\ 0.35 &\in [0.35, 0.7] \\ 0.344\,12 &\in [0.25, 0.5] \\ 0.4 &\in [0.4, 0.8] \end{aligned} \tag{39}$$

V. Comments

We have not shown for $R < L$ that there is a unique solution. This seems to be an interesting technical problem but may yield to more sophisticated analytical techniques. We also have not shown that our algorithm is the fastest possible. This was done in [7] for the case of computing entropy under interval uncertainty, and it may be possible to do something similar for minimizing the sum of squares.

The algorithm provided to find a solution is tedious, but could easily be programmed. It is emphasized that each member of a solution is an exact expression in terms of the endpoints of the given intervals.

There are many measures of dispersion, and in [8], several are discussed. Also, a general definition of dispersion is formulated.

References

[1] J. Abellan and S. Moral. Maximum of entropy for credal sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11:587-597, 2003.

[2] E. Jaynes. *Probability Theory: The Logic of Science*. Cambridge Univ. Press, 2003.

[3] A. Meyerowitz, F. Richman, E. Walker. Calculating maximum-entropy probability densities for belief functions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:377-389, 1994.

[4] H. Nguyen and E. Walker. On decision-making using belief functions. In *Advances in the Dempster-Shafer Theory of Evidence*, R. Yager, M. Fedrizzi, and J. Kacprzyk, Eds., John Wiley & Sons: New York, 311-330, 1994.

[5] H. Nguyen and E. Walker. *A First Course in Fuzzy Logic*, 3rd ed. Chapman & Hall/CRC: Boca Raton, Florida, 2006.

[6] C. Walker, E. Walker, and R. Yager. Generating a Maximum Entropy Probability Density from an Interval-Valued Fuzzy Set. *2008 Proceedings of the NAFIPS International Conference*, New York, 2224-27, May 2008.

[7] G. Xiang, M. Ceberio, and V. Kreinovich. Population Variance and Entropy under Interval Uncertainty: Linear-Time Algorithms. *Reliable Computing*, 13(6):467-488, 2007.

[8] R. Yager. On the dispersion measure of OWA operators. *Technical Report #MII-2901*, Machine Intelligence Institute, Iona College.

ISBN: 978-989-95079-6-8

Approximate Reasoning in Surgical Decisions

Elisabeth Rakus-Andersson

School of Engineering, Department of Mathematics and Science, Blekinge University of Technology
Karlskrona, Sweden
Email: Elisabeth.Andersson@bth.se

Abstract—*Approximate reasoning is one of the most effective fuzzy systems. The compositional rule of inference founded on the logical law modus ponens is furnished with a true conclusion, provided that the premises of the rule are true as well. Even though there exist different approaches to an implication, being the crucial part of the rule, we modify the early implication proposed by Zadeh* [1] *in our practical model concerning a medical application. The approximate reasoning system presented in this work considers evaluation of a risk in the situation when physicians weigh necessity of the operation on a patient. The patient's clinical symptom levels, pathologically heightened, indicate the presence of a disease possible to recover by surgery. We wish to evaluate the extension of the operation danger by involving particularly designed fuzzy sets in the algorithm of approximate reasoning.*

Keywords—Approximate reasoning, compositional rule of inference, Zadeh's implication, operation risk, symptom levels, parametric membership functions.

1 Introduction

The technique of approximate reasoning, earliest evolved by Zadeh [1, 2] quickly found many adherents who differentiated the foundations of the theory. Especially, the changes concerned the implication IF…THEN…ELSE…, which constitutes an important factor of the reasoning system. In [3, 4, 5] we can trace the discussion revealing definitions of the implication generated by Kleene and Denies, Willmot, Mamdani and Assilian, Larsen, Gödel et al. The trials of inserting individually created operations on fuzzy sets discern the approaches mentioned above. Even the item of compositional rule of inference was debated from separate points of views [6, 7, 8, 9, 10]. We can mention the Yager conception [10] and the Sugeno design [3] as the most original modifications of the initial version of the rule.

For a practitioner an applicable meaning of approximate reasoning is essential, especially in technique and natural sciences where vagueness of input and output is often expected. Although some technical trials of applications are remarkable, it can happen coincidentally to counterpart the approximate reasoning in medicine. The only contribution in the topic, found by the author in [11], is a discussion of the model employing a pharmacological example.

Since members of surgical staff make decisions about operations on severely-ill patients with the highest care then we wish to support these verdicts by results coming from reasoning systems. We adopt Zadeh's approach to the rule [1, 2, 12], which is slightly modified by us and based on Lukasiewicz's definition of the fuzzy implication [1, 6, 12]. We still find this rule to be the most appealing for the reason of simply performed operations and clearly interpretable results. Then we build an own original apparatus accommodated to medical assumptions. Particular fuzzy sets that contain input data and output effects are designed in compliance with the physician's hint. The discussion about how to find the objective of reasoning, i.e. operation risk, is accomplished in Section 2. Fuzzy sets, taking place in the model, are furnished with appropriate membership degrees in Section 3. Section 4, added as a presentation of efficiency of the algorithm, reveals some risks in cancer surgery.

2 Adoption of approximate reasoning to operation decisions

For patients, who suffer from e.g. cancer, decisions concerning their operations are made with the highest thoughtfulness. In the later or the last stage of the disease the possibility to cure the patient totally of cancer by operating him/her for tumors is rather little. As a physician does not want the patient to run the risk to suffer even more after an unnecessary operation, he ought to judge thoroughly the consequences of the surgery.

We intend to involve approximate reasoning to support mathematically the extraction of a proper decision when discerning the operation danger. The most decisive clinical symptoms found in an individual patient will be taken into consideration to evaluate the risk.

Let us ponder a logical compound statement

$$\begin{aligned}&\text{IF}\,(p\,\text{AND}\,((\text{IF}\,p\,\text{THEN}\;q)\,\text{ELSE}\,(\text{IF}\,(\text{NOT}\,p)\\&\text{THEN}\,(\text{NOT}\,q)))\,)\,\text{THEN}\,q\end{aligned} \tag{1}$$

whose primitive statements p and q are included in the equivalent form of (1) derived as

$$p\wedge((p\rightarrow q)\wedge(\neg p\rightarrow\neg q))\rightarrow q. \tag{2}$$

The logical joint ELSE is interpreted in (2) as the conjunction $\wedge$ in compliance with the suggestions made by Lukasiewicz and Zadeh [1, 6].

The logical statement (2) is a tautology, which can be easily confirmed by the method of truth tables. We also prove that thesis q in (2) will become true if the premises p and $(p\rightarrow q)\wedge(\neg p\rightarrow\neg q)$ constitute true statements as well. In order to accomplish the last proof we utilize the method of denying the truth of the thesis q. Let $v(p)$ and $v(q)$ denote the truth values of p and q according to the convention of binary logic. If, on behalf of the proof, we

ISBN: 978-989-95079-6-8

assume that the thesis q is not true then $v(q) = 0$. From the previous assumption $v((p \to q) \wedge (\neg p \to \neg q)) = 1$ if $v(p \to q) = 1$ and $v(\neg p \to \neg q) = 1$. But $v(q) = 0$, which suggests that $v(p) = 0$ as well to warrant $v(p \to q) = 1$. On the other hand we have already assumed that premise p is true. As the suggestion $v(q) = 0$ leads to the contradiction "p is false" against "p is true" then we will accept $v(q) = 1$.

In accordance with the extended law *modus ponens* proposed by Zadeh [1, 2] we interpret (2) as a sentence

IF
$p`$ (premise)
AND (3)
(IF p THEN q) ELSE (IF (NOT p) THEN (NOT q)) (premise)
THEN
$q`$ (thesis)

provided that the semantic meaning of p and $p`$ ($q$ and $q`$ respectively) is very close.

Let p be visualized by a fuzzy set P in the universe X and let q be expressed by another fuzzy set Q in the universe of discourse Y. Analogously, the fuzzy set $P` \subset X$ constitutes a mathematical formalization of the primitive statement $p`$ whereas $Q` \subset Y$ replaces formally the sense of $q`$. The modus ponens rule thus becomes

IF
$p` = P`$ (premise)
AND (4)
(IF $p = P$ THEN $q = Q$) ELSE (IF (NOT $p = CP$) THEN (NOT $q = CQ$)) (premise)
THEN
$q` = Q`$ (thesis)

The sets CP and CQ are complements of P and Q.

When making a feedback to the medical task previously outlined, we wish to use a technique of accommodating actual theoretical assertions to concrete formulations letting us evaluate the operation decision in some grades of risk.

Let S denote a symptom possessing the most decisive power in the evaluation of the operation risk. We regard S as either the complex qualitative symptom or the symptom whose intensity is assimilated with level codes. These codes, determined for both descriptions of S's complexion, form the universe X = "*symptom levels*" = $\{1,\ldots,k,\ldots,n\}$. Let us assume that level 1 is associated with the slightly heightened symptom values whereas level n indicates their critical status.

The statement $p`$

$p`$ = "symptom S is found in patient on level k"

is now addressed to a fuzzy set $P`$ introduced by

$$P` = \frac{\mu_{P`}(1)}{1} + \cdots + \frac{\mu_{P`}(k)}{k} + \cdots + \frac{\mu_{P`}(n)}{n}. \quad (5)$$

The sentence p built by

"p = "rising levels of S are essential for operation risk"

is dedicated to a fuzzy set P given by

$$P = \frac{\mu_P(1)}{1} + \cdots + \frac{\mu_P(k)}{k} + \cdots + \frac{\mu_P(n)}{n}. \quad (6)$$

Another category of elements, constituting a content of the universe Y, is determined in the model as risk grades. We set risk grades in Y = "*operation risk grades*" = $\{L_0$ = "*none*", L_1 = "*little*", L_2 = "*moderate*", L_3 = "*great*", L_4 = "*total*"$\}$, on condition that Y is experimentally restricted to five risk grades only.

For sentence q

q = "operation risk exists for patient"

a creation of a fuzzy set Q is supported by

$$Q = \frac{\mu_Q(L_0)}{L_0} + \frac{\mu_Q(L_1)}{L_1} + \frac{\mu_Q(L_2)}{L_2} + \frac{\mu_Q(L_3)}{L_3} + \frac{\mu_Q(L_4)}{L_4}. \quad (7)$$

At last, we define $q`$ containing the final risk judgment as a statement

$q`$ = "patient runs estimated risk of being operated",

where *risk* is graded by membership degrees of the corresponding fuzzy set Q' proposed as

$$Q` = \frac{\mu_{Q`}(L_0)}{L_0} + \frac{\mu_{Q`}(L_1)}{L_1} + \frac{\mu_{Q`}(L_2)}{L_2} + \frac{\mu_{Q`}(L_3)}{L_3} + \frac{\mu_{Q`}(L_4)}{L_4}. \quad (8)$$

In the next paragraph we accomplish the discussion about an apparatus providing us with membership degrees of sets (5)–(8).

Due to *modus ponens* rule (4) we set all decision data in the scheme

IF
"symptom S is found in patient on level k" = $P`$ (premise)
AND
(IF "rising levels of S are essential for operation risk" = P THEN "operation risk exists for patient" = Q) ELSE (IF ("rising levels of S are not essential for operation risk" = CP THEN operation risk does not exists for patient = CQ) (premise)
THEN
"patient runs estimated risk of being operated" = $Q`$ (thesis)

In conformity with [1, 6, 12] we first prognosticate a mathematical expression of the implication

ISBN: 978-989-95079-6-8

(IF "rising levels of *S* are essential for operation risk" = *P* THEN "operation risk exists for patient" = *Q*) ELSE (IF ("rising levels of *S* are not essential for operation risk" = *CP* THEN operation risk does not exists for patient = *CQ*)

performed as matrix *R*. Even though several approaches to membership functions of implications were made [1, 2, 3, 5, 6, 8, 12] we still feel attracted by the Lukasiewicz [6, 12] conception of fuzzy implication *R* with a membership function derived as

$$\mu_R(k,L_l) = 1 \wedge ((1-\mu_P(k)) + \mu_Q(L_i)) \\ \wedge (\mu_P(k) + (1-\mu_Q(L_i))), \tag{9}$$

$k = 1,\ldots,n$, $l = 0,\ldots,4$, for all $x \in X$ and all $y \in Y$.

The membership degrees of set $Q`$ will be visualized after composing set $P`$ with relation R due to Zadeh's compositional rule [1]

$$Q` = P' \circ R \tag{10}$$

designated by the membership function

$$\mu_{Q`}(L_l) = \max_{k \in X}(\min(\mu_{P`}(k), \mu_R(k,L_l))) . \tag{11}$$

The comparisons of magnitudes of membership degrees in set $Q`$ yield indications referring to judgments of the risk grades after consideration of symptom level *k* verified in the patient.

As the operations of maximum and minimum have a tendency to filter the input data, which sometimes does not result in a clear-cut decision, then we will propose another set of composition operations in (10). In accordance with [13] we propose

$$Q` = P' \underset{\cdot +}{\circ} R \tag{12}$$

assisted by membership degrees

$$\mu_{Q`}(L_l) = \frac{\sum_{k=1}^{n} \mu_{P`}(k) \cdot \mu_R(k,L_l)}{\sum_{k=1}^{n} \mu_R(L_l)} . \tag{13}$$

To be able to apply (13) we ought to prove that the value of the quotient $\mu_{Q`}(L_l)$ is a number belonging to the interval [0, 1]. To verify this we first notice that $\mu_{P`}(k) \cdot \mu_R(k,L_l) \le \mu_R(k,L_l)$ since both $\mu_{P`}(k)$ and $\mu_R(k,L_l)$ are less than one for all *k* and *l*, $k = 1,\ldots,n$, $l = 0,\ldots,4$. This causes the value of a product to be lesser than the values of both factors. We thus conclude that the numerator is less than or equal to the denominator, which guarantees that the entire value of the quotient is a member of [0, 1]; therefore it can be approved as a membership degree of L_l coming from the support of $Q`$.

We also notice that the sum placed in the denominator of the quotient never becomes equal to zero, since almost all risk grades will be designed as positive quantities. This assumption prohibits membership degrees of the risk grades from being undefined structures.

Values $\mu_R(k,L_l)$ are adaptable to be treated as weights of level importance assigned to a distinct risk. These, as the entries of matrix *R* are invariants in the system promoting the same diagnostic model, contrary to information concerning different patients that is changeable. And, additionally, we can prove that operation (13) satisfies the criteria of OWA operators [13].

3 Mathematical design of data sets

The decision model designed in Section 2 includes operations on fuzzy sets furnished with symbolically established membership degrees. In the current paragraph we put some life into theoretical symbols by assigning to them mathematical structures. The set $P`$ a.k.a. (5) now gets assigned

$$P` = \frac{\mu_{P`}(1)}{1} + \cdots + \frac{\mu_{P`}(k)}{k} + \cdots + \frac{\mu_{P`}(n)}{n} \\ = \cdots + \frac{\frac{n-2}{n}}{k-2} + \frac{\frac{n-1}{n}}{k-1} + \frac{1}{k} + \frac{\frac{n-1}{n}}{k+1} + \frac{\frac{n-2}{n}}{k+2} + \cdots \tag{14}$$

for the k^{th} symptom level certified in the patient examined.

Another set *P*, concerning the same symptom levels in the support, is found by (6) and modified as

$$P = \frac{\mu_P(1)}{1} + \cdots + \frac{\mu_P(k)}{k} + \cdots + \frac{\mu_P(n)}{n} \\ = \frac{\frac{1}{n}}{1} + \cdots + \frac{\frac{k}{n}}{k} + \cdots + \frac{\frac{n}{n}}{n}, \tag{15}$$

due to the previously made assumptions, which suggest the tendency to ascending values of the membership degrees in *P*.

The set *Q* is more sophisticated to design as a fuzzy set whose support consists of other fuzzy sets L_l, $l = 0,\ldots,4$, commonly defined in a symbolic risk reference set $Z = [0, 1]$. We also intend to determine the membership degrees of *Q* as some characteristic quantities from [0, 1]. Evaluation of these numbers is founded on a procedure involving a linguistic variable

"*operation risk grades*" = {L_0 = "*none*", L_1 = "*little*", L_2 = "*moderate*", L_3 = "*great*", L_4 = "*total*"},

experimentally restricted to five risk grades only.

We first fuzzify the expressions concerning the items of the list to continue further with their defuzzification in order to attach numerical equivalents to the words from the list. Each word assists now a fuzzy set L_l, $l = 0, 1, 2, 3, 4$, whose constraint is grounded on an *s*-class mapping defined for *z* in $Z = [0, 1]$ as [14]

ISBN: 978-989-95079-6-8

$$\mu_{L_l}(z)=\mu_{L_0(l)}(z)=\begin{cases}\text{left}(\mu_{L_0(l)}(z))= \\ \text{right}(\mu_{L_0(l)}(z))= \\ s(z,\alpha_{L_0},\beta_{L_0},\gamma_{L_0},l\cdot h), & \text{for} \quad z\le\gamma_{L_0}, \\ 1-s(z,\alpha_{L_0}+h,\beta_{L_0}+h,\gamma_{L_0}+h,l\cdot h) & \text{for} \quad z>\gamma_{L_0}. \end{cases} \tag{16}$$

We clarify the fact that formulas of all membership functions are derived from only one predetermined subject defining $\mu_{L_0}(z)$. The equality $\mu_{L_l}(z)=\mu_{L_0(l)}(z)$ reveals that $\mu_{L_l}(z)$ is dependent on a parameter l equal to level number l, l = 0,…,4. The h unit determines a distance between α_{L_l} and $\alpha_{L_{l+1}}$ (respectively β_{L_l} and $\beta_{L_{l+1}}$ or γ_{L_l} and $\gamma_{L_{l+1}}$) for symmetric functions s.

We prepare constraints for L_0, which are affected by $\alpha_{L_0}=-0.25$, $\beta_{L_0}=-0.125$ and $\gamma_{L_0}=0$ as

$$\text{left}(\mu_{L_0}(z))=\begin{cases}2\left(\frac{z-(-0.25)}{0-(-0.25)}\right)^2 \\ \text{for} \quad -0.25\le z<-0.125, \\ 1-2\left(\frac{z-0}{0-(-0.25)}\right)^2 \\ \text{for} \quad -0.125\le z<0, \end{cases} \tag{17}$$

and

$$\text{right}(\mu_{L_0}(z))=\begin{cases}1-2\left(\frac{z-0}{0.25-0}\right)^2 \\ \text{for} \quad 0\le z<0.125, \\ 2\left(\frac{z-0.25}{0.25-0}\right)^2 \\ \text{for} \quad 0.125\le z<0.25. \end{cases} \tag{18}$$

By inserting in (17) and (18) the current value l, l = 0,…,4, and the distance h, casually determined as h = 0.25, we obtain a formula of the left branch of L_l

$$\text{left}(\mu_{L_l}(z))= \begin{cases}2\left(\frac{z-(-0.25+l\cdot 0.25)}{0-(-0.25)}\right)^2 \\ \text{for} \quad -0.25+l\cdot 0.25\le z<-0.125+l\cdot 0.25, \\ 1-2\left(\frac{z-(0+l\cdot 0.25)}{0-(-0.25)}\right)^2 \\ \text{for} \quad -0.125+l\cdot 0.25\le z<0+l\cdot 0.25, \end{cases} \tag{19}$$

and a function shaping its right branch

$$\text{right}(\mu_{L_l}(z))= \begin{cases}1-2\left(\frac{z-(0+l\cdot 0.25)}{0.25-0}\right)^2 \\ \text{for} \quad 0+l\cdot 0.25\le z<0.125+l\cdot 0.25, \\ 2\left(\frac{z-(0.25+l\cdot 0.25)}{0.25-0}\right)^2 \\ \text{for} \quad 0.125+l\cdot 0.25\le z<0.25+l\cdot 0.25. \end{cases} \tag{20}$$

Figure 1 collects plots of L_0–L_4 in conformity with different values of l included in (19) and (20).

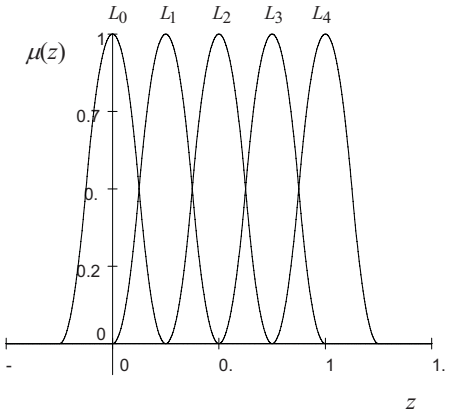


Figure 1: The terms of "*operation risk grades*" as fuzzy sets L_0–L_4

Actually, we have an intension to emphasize the meaning of parametric nature of the L_l membership functions, which deprives the model of many distinct formulas. Apart from this advantage we focus on generating the functions that represent elegant structures mathematically expressed.

In the process of defuzzification we consider only z-values for which the sets L_0–L_4 get the status of normal sets, i.e., z = 0, z = 0.25, z = 0.5, z = 0.75 and z = 1. For these, another fuzzy set "*numerical operation risk*" is projected by developing its membership function in the form of

$$\mu_{\text{"numerical operation risk"}}(z)=\begin{cases}2\left(\frac{z-0}{1-0}\right)^2 \\ \text{for} \quad 0\le z<0.5, \\ 1-2\left(\frac{z-1}{1-0}\right)^2 \\ \text{for} \quad 0.5\le z<1. \end{cases} \tag{21}$$

Via the selected z-quantities above, we tie their membership degrees calculated by means of (21) to expressions from the list in order to establish relations between words and their numerical replacements. Therefore, the set Q finally obtains a shape of

$$Q=\frac{0}{L_0}+\frac{0.125}{L_1}+\frac{0.5}{L_2}+\frac{0.875}{L_3}+\frac{1}{L_4}. \tag{22}$$

We now wish to demonstrate the action of approximate reasoning accustomed to the judgment of surgical risk.

4 Risks grades in cancer surgery

In patients, who suffer from cancer as the recognized diagnosis, one of the symptoms, namely, *CRP* (*C*-reactive proteins) is carefully measured and discussed with a view to make a decision about accomplishing a successful operation. The heightened values of *CRP* (measured in milligrams per liter) are theoretically discerned in four levels stated as

1 = "*almost normal*" for $CRP < 10$,
2 = "*heightened*" if $10 \le CRP \le 20$,
3 = "*very heightened*" if $20 \le CRP \le 25$,

ISBN: 978-989-95079-6-8

4 = "*dangerously heightened*" for $CRP > 25$.

Due to (15) set P is expressed as

$$P = \frac{0.25}{1} + \frac{0.5}{2} + \frac{0.75}{3} + \frac{1}{4} \tag{23}$$

in $X = \{1,\ldots,4\}$.

Suppose that an individual patient examined reveals the *CRP*-value to be 23. *CRP* is thus classified in level 3 and set P` characteristic of the patient is stated in the form of

$$P` = \frac{0.5}{1} + \frac{0.75}{2} + \frac{1}{3} + \frac{0.75}{4} \ . \tag{24}$$

according to (14).

The sets (23) and (22) together with

$$\mathrm{C}P = \frac{0.75}{1} + \frac{0.5}{2} + \frac{0.25}{3} + \frac{0}{4} \tag{25}$$

and

$$\mathrm{C}Q = \frac{1}{L_0} + \frac{0.875}{L_1} + \frac{0.5}{L_2} + \frac{0.125}{L_3} + \frac{0}{L_4} \tag{26}$$

generate matrix R with the entries computed in compliance with (9). R is expanded as a two-dimensional table

$$R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} L_0 & L_1 & L_2 & L_3 & L_4 \end{array} \\ \begin{bmatrix} 0.75 & 0.875 & 0.75 & 0.375 & 0.25 \\ 0.5 & 0.625 & 1 & 0.625 & 0.5 \\ 0.25 & 0.375 & 0.75 & 0.875 & 0.75 \\ 0 & 0.125 & 0.5 & 0.875 & 1 \end{bmatrix} \end{array} \tag{27}$$

which, inserted in (10) for P` determined by (24), provides us with

$$Q` = \frac{0.5}{L_0} + \frac{0.625}{L_1} + \frac{0.75}{L_2} + \frac{0.875}{L_3} + \frac{0.75}{L_4} . \tag{28}$$

By interpreting the meaning of (28) we understand that there exists a risk when considering an operation in patient whose *CRP*-index is evaluated on the third level. The most possible risk is evaluated as "*great*" according to the highest quantity of the membership degree. The total danger of accomplishing the surgical operation is evaluated as essential with the membership degree 0.75.

Even the results of implementing (13) given as

$$Q` = \frac{0.66}{L_0} + \frac{0.69}{L_1} + \frac{0.75}{L_2} + \frac{0.795}{L_3} + \frac{0.725}{L_4} . \tag{29}$$

fully confirm the risk extension judged by (28).

We hope that the classical model of approximate reasoning, modified by us and adapted to the problem of operation decision can constitute its complementary solution, especially when a decision of saving somebody's life via surgery is crucial.

5 Conclusions

Via the way of forming the text of this paper we have already come to substantial conclusions. We only summarize that we have used approximated reasoning to introduce the own initial interpretation of the system to approximate the operation risk concerning patients with rising values of a biological index. The formulas of membership degrees and membership functions have been expanded by applying a formal mathematical design. We expect that the study makes a contribution in the domain of mathematical models projected for medical applications.

In future works we wish to examine a model consisted of several symptoms that are divided in different numbers of levels. The symptoms should be included in the pattern simultaneously, which may expose some internal interactions among them. In other words, the operation risk will be a criterion that can employ many data factors. We count on finding some helpful remarks in [15] to implement an algorithm supporting the method newly planned.

References

[1] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision process. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. MSC-3(1): 28-44, 1973.

[2] L. A. Zadeh. A theory of approximate reasoning. *Machine Intelligence*, 9: 149-174, 1979.

[3] R. Fuller. On fuzzy reasoning schemes. http:// www.abo. fi/~ rfuller/ pgs97.pdf.

[4] E. Hüllermeier**.** *Case based approximate reasoning*, Berlin Heidelberg: Springer, 2007.

[5] P. M. van den Broek. Efficient algorithms for approximate reasoning. *Proceedings of the 6th International Conference on Neural Information Processing, ICONIP 1999*, Perth, Western Australia, 1999.

[6] J. F. Baldwin and B. W. Pilsworth. A model of fuzzy reasoning through multi-valued logic and set theory. *Int. J. of Man-Machine Studies*, 11: 351-380, 1979.

[7] C. Cornelis, M. de Cock and E. Kerre. Efficient approximate reasoning with positive and negative information. *Lecture Notes in Artificial Intelligence*, 3214: 779-785, 2004.

[8] G. Gerla. *Fuzzy logic: mathematical tools for fuzzy reasoning*. Kluwer, 2001.

[9] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. of Man-Machine Studies*, 7: 1-13, 1975.

[10] R. Yager. An approach to inference in approximate reasoning. *Int. J. of Man-Machine Studies*, 13: 323-338, 1980.

[11] K. Kilic, B. A. Sproule, I. B. Türksen and C. A. Naranjo. A fuzzy system modeling algorithm for data analysis and approximate reasoning. https://research.sabanciuniv. edu/418/1/robotics _kemal. DOC.R.

[12] J. Kacprzyk. *Fuzzy sets in system analysis*. Warsaw: PWN, 1986.

[13] E. Rakus-Andersson. *Fuzzy and rough techniques in medical diagnosis and medication*. Berlin Heidelberg: Springer, 2007.

[14] E. Rakus-Andersson. *S*-truncated functions and rough sets in approximation and classification of polygons. *Proc. of Modeling Decisions for Artificial Intelligence – MDAI 2005,* Tsukuba, CD-ROM, paper nr 049, Consejo Superior de Investigaciones Cientificas, 2005.

ISBN: 978-989-95079-6-8

[15] O. Wolkenhauer. Fuzzy inference engines, composition and individual rule based on non-linear mappings. www. Systemsbiology .umist.ac. uk/dokumente/ dataengineering/ infengines .pdf

ISBN: 978-989-95079-6-8

Chaos induced by turbulent and erratic functions

H. Román-Flores[1] Y. Chalco-Cano[2] G. Nunes-Silva[3]

1.Instituto de Alta Investigación, Universidad de Tarapacá
Arica, Chile
2.Departamento de Matemática, Universidad de Tarapacá
Arica, Chile
3.Departamento de Ciência da Computação e Estatística, Universidade Estadual Paulista
São José do Rio Preto - SP, Brasil
Email: {hroman,ychalco}@uta.cl, gsilva@ibilce.unesp.br

***Abstract*— *Let (X,d) be a compact metric space and $f:X\to X$ a continuous function and consider the hyperspace $(\mathcal{K}(X),H)$ of all nonempty compact subsets of X endowed with the Hausdorff metric induced by d. Let $\bar{f}:\mathcal{K}(X)\to\mathcal{K}(X)$ be defined by $\bar{f}(A)=\{f(a)/a\in A\}$ the natural extension of f to $\mathcal{K}(X)$, then the aim of this work is to study the dynamics of $\bar{f}$ when f is turbulent (erratic, respectively) and its relationships.*

***Keywords*—** Chaos, dynamical systems, turbulent functions, erratic functions.

1 Introduction

Let (X,d) be a compact metric space and $f:X\to X$ a continuous function. If we consider the extension $\bar{f}:\mathcal{K}(X)\to\mathcal{K}(X)$, then an interesting problem is to analyze the chaotic relations between f and $\bar{f}$.

In this direction, Román-Flores in [1] proved that transitivity of $\bar{f}$ implies transitivity of f, whereas in [2] the author proved that sensitivity of $\bar{f}$ implies sensitivity of f and periodic density of f implies periodic density of $\bar{f}$. Also, Banks in [3] shows that transitivity of $\bar{f}$ is equivalent to f weakly mixing. Several other chaotic relations between f and $\bar{f}$ have also been investigated by many authors, for example Cánovas et al. [4] investigated some connections in relation to topological entropy, Gu in [5] studied Kato's chaos, and Ma et al. [6] investigated some aspects of topological entropy, Li-Yorke chaos and distributional chaos.

The aim of this work is to explore some turbulent and erratic connections between f and $\bar{f}$.

2 Preliminaries

Let (X,d) be a compact metric space and let $\mathcal{K}(X)$ be the class of all non-empty and compact subsets of X. If $A\in\mathcal{K}(X)$ we define the "ϵ-neighbourhood of A" as the set

$$N(A,\epsilon)=\{x\in X/\ d(x,A)<\epsilon\},$$

where $d(x,A)=\inf_{a\in A}d(x,a)$.

The Hausdorff metric on $\mathcal{K}(X)$ is defined as

$$H(A,B)=\inf\{\epsilon>0/\ A\subseteq N(B,\epsilon)\ and\ B\subseteq N(A,\epsilon)\},$$

and it is well known that $(\mathcal{K}(X),H)$ is a compact metric space (see [7]).

If $A\in\mathcal{K}(X)$ we denote by $B(A,\epsilon)$ the ball centered in A and radius ϵ in H-metric.

Remark 1 *An equivalent formula for H is given by*

$$H(A,B)=\max\left\{\sup_{a\in A}d(a,B),\ \sup_{b\in B}d(b,A)\right\}.$$

Remark 2 *If K_n,K_0 are nonemty compact subsets of X such that $K_n\xrightarrow{H}K_0$, then (see [7])*

$$K_0=\bigcap_{p=1}^{\infty}\overline{\bigcup_{n\geq p}K_n}. \tag{1}$$

If X has a real linear structure then, by using the Minkowski sum between two sets, a linear structure of convex cone is defined on $\mathcal{K}(X)$ by means

$$A+B=\{a+b,\ a\in A,b\in B\}\ \text{ and }\ \lambda A=\{\lambda a,\ a\in A\},$$

for all $A,B\in\mathcal{K}(X),\ \lambda\in\mathbb{R}$.

Let A be a subset of X. Then we define the extension of A to $\mathcal{K}(X)$ as

$$e(A)=\{K\in\mathcal{K}(X)/\ K\subseteq A\}.$$

The following results were proved by the author in [1].

Lemma 1 *Let A,B be two subsets of X and let $f:X\to X$ be a continuous function. Then,*

i) $e(A)=\emptyset$ *if and only if* $A=\emptyset$.

ii) $e(A\cap B)=e(A)\cap e(B)$

iii) If A is a nonempty open subset of X, then $e(A)$ is a nonempty open subset of $\mathcal{K}(X)$.

iv) $\bar{f}(e(A))\subseteq e(f(A))$

v) $\bar{f}^p=\overline{f^p}$ *, for every* $p\in\mathbb{N}$.

We remark that if f is a bijective map then

$$\bar{f}(e(A))=e(f(A)). \tag{2}$$

Also, as a direct consequence of (1) we have

Lemma 2 *If A is a nonempty closed subset of X, then $e(A)$ is a nonempty closed subset of $\mathcal{K}(X)$.*

ISBN: 978-989-95079-6-8

Proof. If K_0 is an acumulation point of $e(A)$ in $\mathcal{K}(X)$ then there exists a sequence (K_n) contained in $e(A)$ such that $K_n \xrightarrow{H} K_0$. Thus, $K_n \subseteq A$ for all $n \in \mathbb{N}$ and, due to (1), we have

$$K_0 = \bigcap_{p=1}^{\infty} \overline{\bigcup_{n \geq p} K_n} \subseteq A$$

which implies that $K_0 \in e(A)$.□

Lemma 3 *If A is a nonempty convex subset of X, then $e(A)$ is a nonempty convex subset of $\mathcal{K}(X)$.*

Proof. Let J, K be two nonempty compact subsets of X such that $J, K \in e(A)$. Thus, since $J, K \subseteq A$ and A is a convex set, then

$$\lambda J + (1-\lambda)K = \{\lambda x + (1-\lambda)y : x \in J, y \in K\} \subseteq A,$$

for all $\lambda \in [0,1]$, which implies that
$\lambda J + (1-\lambda)K \in e(A)$ and, consequently, $e(A)$ is a convex set.□

3 Erraticness and turbulence

In this section we will analyze some connections between the chaos induced by turbulent functions and erratic functions.

Definition 1 *Let $f : X \to X$ be a continuous function. We say that f is a turbulent function if there exist two disjoint nonempty compact subsets J, K of X such that*

$$J \cup K \subseteq f(J) \cap f(K)$$

Definition 2 *Let $f : X \to X$ be a continuous function. We say that f is chaotic in the sense of Block and Coppel (in short: B-C-chaotic) if one of its iterates is turbulent, i.e., there exist $n \geq 1$ and two disjoint nonempty compact subsets J, K of X such that $J \cup K \subseteq f^n(J) \cap f^n(K)$.*

Definition 3 *Suppose that X is a compact convex subset of a linear space and let $f : X \to X$ be a continuous function. We say that f is an erratic function if there exists a nonempty convex compact subset A of X such that*

a) $A \cap f(A) = \emptyset$;

b) $A \cup f(A) \subseteq f^2(A)$.

Theorem 1 *If $f : X \to X$ is a bijective turbulent function, then $\bar{f} : \mathcal{K}(X) \to \mathcal{K}(X)$ is also a turbulent function.*

Proof. Suppose that f is a turbulent function. Then, if J, K are two disjoint nonempty compact subsets of X such that $J \cup K \subseteq f(J) \cap f(K)$, then by Lemma 1 i)-ii) and Lemma 2 it is clear that $e(J), e(K)$ are two disjoint nonempty compact subsets of $\mathcal{K}(X)$.
Moreover, because $J \cup K \subseteq f(J) \cap f(K)$, we have $J \subseteq f(J) \cap f(K)$ and $K \subseteq f(J) \cap f(K)$, which implies that $e(J) \subseteq e(f(J) \cap f(K))$ and $e(K) \subseteq e(f(J) \cap f(K))$.
Thus, due to equality (2), we obtain

$$\begin{array}{rcl} e(J) \cup e(K) & \subseteq & e(f(J) \cap f(K)) \\ & = & e(f(J)) \cap e(f(K)) \\ & = & \bar{f}(e(J)) \cap \bar{f}(e(K)) \end{array}$$

and, consequently, $\bar{f}$ is a turbulent function.□

Corollary 1 *If the continuous bijection f is B-C-chaotic, then $\bar{f}$ is B-C-chaotic.*

Theorem 2 *If $f : X \to X$ is a bijective erratic function, then $\bar{f} : \mathcal{K}(X) \to \mathcal{K}(X)$ is also an erratic function.*

Proof. Suppose that f is an erratic function, and let A be a nonempty compact convex subset of X satisfying conditions a) and b) in Definition 3. Then, due to Lemmas 2 and 3, we have that $e(A)$ is a nonempty compact convex subset of $\mathcal{K}(X)$. Now we will prove that $e(A)$ satisfies conditions a) and b) for $\bar{f}$. In fact, by 2 and equality (2) we have:

$$\begin{array}{rcl} i)\ A \cap f(A) = \emptyset & \Rightarrow & e(A) \cap e(f(A)) = \emptyset \\ & \Rightarrow & e(A) \cap \bar{f}(e(A)) = \emptyset \end{array}$$

and, since $A \cup f(A) \subseteq f^2(A)$, we have:

$$\begin{array}{rcl} ii)\ e(A) \cup \bar{f}(e(A)) & = & e(A) \cup e(f(A)) \\ & \subseteq & e(A \cup f(A)) \\ & \subseteq & e(f^2(A)) = \bar{f}^2(e(A)) \end{array}$$

which implies that $\bar{f}$ is an erratic function.□

The following result shows that erraticness is stronger than B-C-chaos.

Theorem 3 *Let $f : X \to X$ be a continuous function. If f is an erratic function, then f is a B-C-chaotic function.*

Proof. Suppose that f is an erratic function, and let A be a nonempty compact convex subset of X satisfying conditions a) and b) in Definition 3 Then, due to a), taking $J = A$ and $K = f(A)$ we have $J \cap K = A \cap f(A) = \emptyset$, and by b):

$$\begin{array}{rcl} J \cup K & = & A \cup f(A) \\ & \subseteq & f^2(A) \\ & = & f(K) \\ & \subseteq & f(J \cup K) \\ & \subseteq & f(f^2(J)) = f^3(J) \end{array}$$

and

$$\begin{array}{rl} J \cup K & = A \cup f(A) \subseteq f^2(J) \\ & \subseteq f^2(J \cup K) \subseteq f^2(f(K)) = f^3(K) \end{array}$$

which implies that $J \cup K \subseteq f^3(J) \cap f^3(K)$ and, consequently, f is a B-C-chaotic function.□

Example 1 *If we consider the tent function $f : [0,1] \to [0,1]$ defined by*

$$f(x) = \begin{cases} 2x & if \quad 0 \leq x \leq \frac{1}{2} \\ 2(1-x) & if \quad \frac{1}{2} \leq x \leq 1, \end{cases}$$

then f is an erratic function. In fact, taking $A = [\frac{1}{4}, \frac{4}{9}]$ we have $f(A) = [\frac{1}{2}, \frac{8}{9}]$, which implies that $A \cap f(A) = \emptyset$ and :

$$A \cup f(A) = \left[\frac{1}{4}, \frac{4}{9}\right] \cup \left[\frac{1}{2}, \frac{8}{9}\right] \subseteq \left[\frac{2}{9}, 1\right] = f^2(A)$$

and, consequently, f is an erratic function.

On the other hand, if we take $J = A$ and $K = f(A)$ then we have $J \cap K = \emptyset$ and, since $f^2(J) = \left[\frac{2}{9}, 1\right]$ and $f^2(K) = [0,1]$, we obtain that $J \cup K \subseteq f^2(J) \cap f^2(K)$ which implies that f is B-C-chaotic.

ISBN: 978-989-95079-6-8

Acknowledgment

This work was supported by Conicyt-Chile by projects Fondecyt 1080438 and 1061244.

References

[1] H. Román-Flores. A note on transitivity in set-valued discrete systems *Chaos, Solitons & Fractals* 17, 99–104, 2003.

[2] H. Román-Flores and Y. Chalco-Cano Robinson's chaos in set-valued discrete systems *Chaos, Solitons & Fractals* 25, 33–42, 2005.

[3] J. Banks Chaos for induced hyperspace maps *Chaos, Solitons & Fractals* 25, 681–685, 2005.

[4] J. Cánovas and G. Soler. Topological entropy for induced hyperspace maps *Chaos, Solitons & Fractals* 28, 979–982, 2006.

[5] R. Gu Kato's chaos in set-valued discrete systems *Chaos, Solitons & Fractals* 31, 765–771, 2007.

[6] X. Ma, B. Hou and G. Liao Chaos in hyperspace system *Chaos, Solitons & Fractals* 40, 653–660, 2009.

[7] E. Klein and A. Thompson *Theory of Correspondences* Wiley-Interscience, New York, 1984.

ISBN: 978-989-95079-6-8

Verbalizing Time-series Data: With an Example of Stock Price Trends

Ichiro KOBAYASHI[1] Naoko OKUMURA[2]

1.Graduate School of Humanities and Sciences, Advanced Sciences, Ochanomizu University
2-1-1 Ootsuka Bunkyo-ku Tokyo, 112-8610 Japan
2.NTT Data Corporation Ltd.
Email: koba@is.ocha.ac.jp

Abstract— *We propose a method to verbalize time-series data and to generate texts that explain the behavior of the data expressed in a 2D chart. We use Nikkei stock price trends data as an example of time-series data, and then aim to generate a text which summarizes the behavior of the trends in a day with words as the one we often see on a newspaper or Web sites. The trends of the stock price are observed by numerical data expressed in a 2D chart. When observing the trends of time-series data, we usually observe the shape of the chart, and then verbalize the shape with proper words to express the behavior of the chart. In order to recognize the behavior in qualitative and quantitative ways, we use the least squares method to mathematically recognize the shape of a 2D chart and expressed it with words often used in news articles reporting the trends of stock prices. Our proposed method can change non-verbal information into verbal information — this provide us with high accessibility and usability for various kinds of information.*

Keywords— 2D chart, Stock price trends, Time-series data, Verbalizing

1 Introduction

We propose a method to verbalize time-series data. In particular, we focus on explaining the behavior of time-series data expressed in 2D charts. We often come across many documents with multimodal information like texts, graphs, tables, etc, therefore, technologies to deal with multimodal information have been increasingly required. In particular, if non-verbal information can be expressed with natural language, the information will be easily retrieved with the current text-based information retrieval system. Thus, changing non-verbal information into verbal information will provide us with high accessibility to various kinds of information, and also usability of non-verbal information can also be expanded. In this context, we propose a method to change modality through expressing non-verbal information with verbal information and to expand communication among different kinds of modalities by that. The proposed method is verified by applying it to a problem to generate a stock price trends report with the actual numerical data of Nikkei stock average.

2 Overview of system

First of all, we show the architecture of a system that generates verbal reports on stock prices trends in Figure 1. As the data for stock prices dealt with by the system, we use Nikkei stock average.

The system consists of several modules: a data base to manage numerical data of Nikkei stock average, a module to recognize the shape of 2D charts, a dictionary to express the shape of 2D line chart with words, and a module to provide syntactic rules for generating a text.

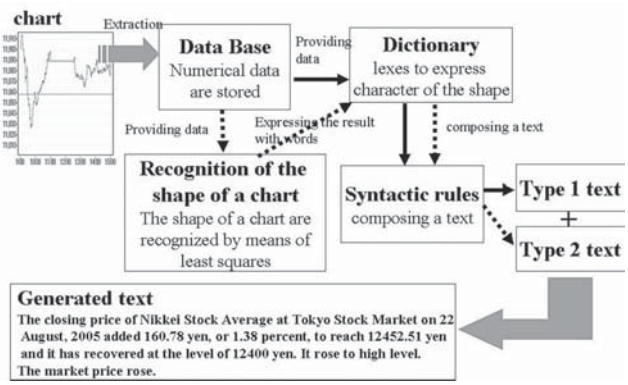


Figure 1: System architecture

The texts generated by the system are categorized into two types as follows:

type 1: The text which does not refer to the shape of 2D charts. The texts of this type are mainly confined to explaining the particular states or levels of numerical data.

type 2: The text which refers to the shape of 2D charts. The texts of this type explain the behavior of the 2D charts trends by recognizing the shape of the line charts.

2.1 Generation process

The algorithm of generating a type 2 text is as follows: (The algorithm of generating a text of type 1 merely omits *step 2* in the following steps.)

step 1. Extracting numerical data from a data base
Necessary numerical data to generate a text are extracted from a data base.

step 2. Recognition of the shape of a chart
The shape of chart is approximately recognized by means of the linear least squares.

step 3. Lexical selection for the partial shapes of a chart
To generate a type 1 text, proper lexes are selected to verbally express the numerical data obtained at step 1. And to generate a type 2 text, proper lexes are selected by recognizing the shape of a chart.

step 4. Providing templates and syntactic rules
As for type 1 text generation, the selected lexes are put

ISBN: 978-989-95079-6-8

into the slots of a prepared text template and short sentences to explain the trend of the price are added to output text. As for type 2 text generation, the short sentences to explain the behavior of a line chart are composed with syntactic rules, and then a text explaining the behavior of a line chart is generated.

Each component of the system will be explained in the following.

2.2 Data base

We use numerical data of Nikkei stock average from July 25, 2005 till August 30, 2005 as input information to the system. There are several kinds of input information which are 10 minutes interval data through a day, opening and closing price, and highest and the lowest prices data for past three days.

2.3 Recognition of the shape of a line chart

We apply the linear least squares to make an approximate line from the original line chart, and then verbally recognize the behavior of the original line chart from that of the approximate line chart. Figure 2 shows an example of applying the least squares to the original line chart. As we see from the example, we can recognize the behavior of the original line chart from the approximate line chart as its trend goes down at first and then it rises up.

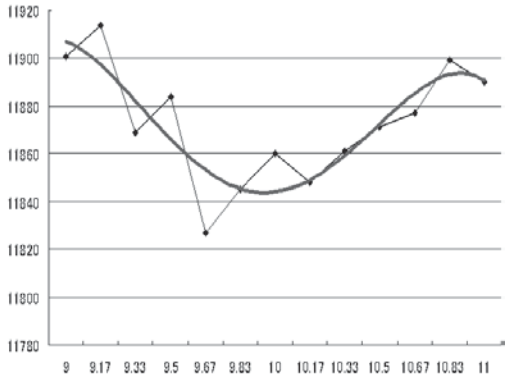


Figure 2: Example of the least squares

2.3.1 Least squares

The least squares method assumes that the best-fit curve of a given type is the curve that has the minimal sum of the deviations squared (least square error) from a given set of data. Suppose that the data points are,(x_1, y_1), (x_2, y_2),$\cdots$,(x_n, y_n) where x is the independent variable and y is the dependent variable. The fitting curve $f(x)$ has the deviation (error) d from each data point, i.e., $d_1 = y_1 - f(x_1)$, $d_2 = y_2 - f(x_2)$, $\cdots$, $d_n = y_n - f(x_n)$.

According to the least squares method, the best fitting curve has the property that:

$$F(x) = d_1^2 + d_2^2 + \cdots + d_n^2 = \sum_{i=1}^{n} d_i^2$$
$$= \sum_{i=1}^{n} [y_i - f(x_i)]^2 \rightarrow a\ minimum$$

Through analyzing corpus data — we used 27 news articles during the same period as numerical data — we adopted five dimensional polynomial function to properly approximate the shape of line charts which is because we found that the function is the most suitable polynomial function to be verbalized with the words often used in the news articles explaining the behavior of line charts.

2.3.2 Whole and partial shapes of charts

We have defined 11 types for the shape of a line chart as shown in Figure 3.

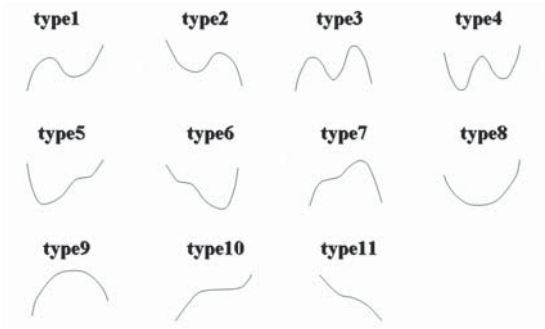


Figure 3: 11 types of whole shape

Furthermore, those 11 types are broken down into sub 13 types of partial shapes as shown in Figure 4.

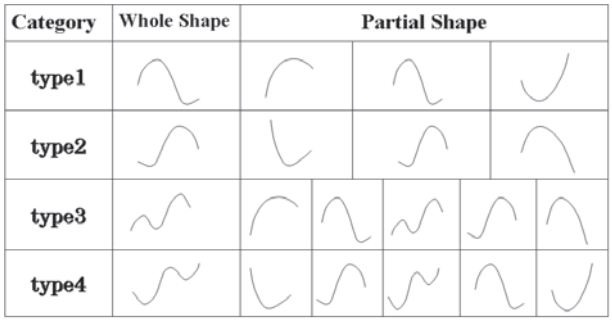


Figure 4: Example of partial shapes

This categorization for the shapes is decided based on the words often used to explain the behavior of a line chart in the corpus — we focus on the words used in explaining the behavior of a line chart, not on the characteristics of the shape of a chart, and extracted typical shapes expressed with words.

The algorithm we developed firstly selects a relevant type from the 11 types of whole shape and then the final shape of the line chart is decided through mathematical recognition of the partial shape of the selected whole type (see, Table 1 for example). Afterwards, the proper words to express the final shape are chosen to generate a report on stock price trends.

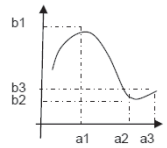
2.3.3 Target zones of recognizing the shape

There are two time zones in the trade of stocks, i.e., 'the morning session' and 'the afternoon session'(see, Fig. 5).

The texts explaining the trends of Nikkei stock price in a day on a news paper or Web sites are usually reported and summarized in each session. We follow this, and then recognize the shape of a chart with the approximate function in each session.

ISBN: 978-989-95079-6-8

Table 1: Partial shape and verbal expressions

Partial shape	Characteristics	Verbal expressions
	|b2-b1|/|MAX-MIN|>0.4 |a1-a2|/|max-min|<0.7	Sell order was dominant
	|a1-a2|/|max-min|>0.7	Sell order expanded
	|b2-b1|/|MAX-MIN|>0.4 |b2-b3|/|b2-b1|>0.5 |a1-a2|/|max-min|<0.7	Sell order was once dominant
	|a3-a2|/|max-min|<0.7 |a3 -a2|/|max-min|>0.5	Over the middle of the session
	|a3 -a1|/|max-min|<0.2 |a3 -a2|/|max-min|<0.2	Around the middle of the session
	|a3-a1|/|max-min|<0.6 |a3-a1|/|max-min|>0.45	After the middle of the session

MAX, MIN : the maximum and minimum prices at morning (afternoon) sessions.
max, min : the starting and closing time of morning (afternoon) sessions.

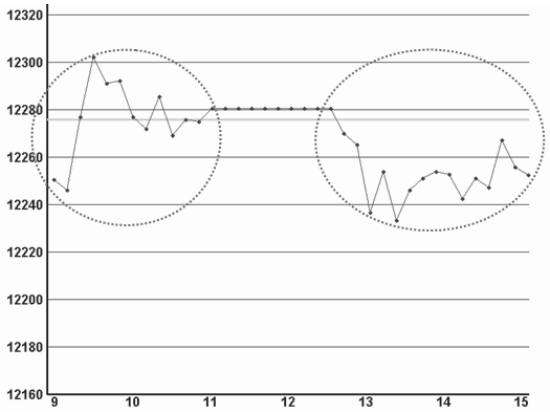


Figure 5: Target zones of recognizting the shape

2.4 *Dictionary*

In generating texts explaining the behavior of a chart, we start with the standing point where we use the words that are usually used to explain the behavior, therefore, we firstly analyze a news article corpus about the stock price trends and extract the words often used to explain the stock price trends. And then we observe the behavior of the trends by recognizing the shapes of the 2D chart of the trends from a viewpoint of explaining the behavior with the words. In concrete, we have analyzed 27 Japanese news articles about Nikkei stock average during the same period as the numerical data were obtained, i.e., from July 25, 2005 till August 30, 2005. We collected words and short sentences to explain the partial shapes of charts; words to explain particular conditions recognized from numerical data; words to express time and comparison between the current and past conditions; and conjunctions to bond those sentences.

Some examples registered in the dictionary are shown below.

- *Short sentences to explain the partial shapes of a chart*
  'Sell order was expanded', 'The price was edgingly rising', etc. 38 kinds of short sentences are registered in the dictionary.

- *Words to explain the states identified from a particular numerical data set*
  'rebound', 'continuously rise', 'The price was moving with high level through a day', etc. 19 kinds of words are registered in the dictionary.

- *Words to express time and comparison*
  'At morning (afternoon) session', 'At closing session', 'After the middle of the session', 'compared to the price of last weekend', etc. 10 kinds of words are registered in the dictionary.

- *Conjunction*
  4 kinds of conjunctions are registered in the dictionary. They are 'and then', 'therefore', 'afterwards', and 'however'.

2.5 *Syntactic rules for generated texts*

Syntactic rules to compose a text differ for each text type. Type 1 texts are generated basically by using templates. Type 2 texts are generated based on verbal expressions of the behavior of a line chart with some syntactic rules and short sentences.

2.5.1 *Type 1 text*

From the analysis of the 27 Japanese news articles, we have obtained typical three types of template to explain the Nikkei stock average trends. Type 1 text is normally generated based on one of the text templates shown below.

(a) The closing price of Nikkei stock average at Tokyo market on *DATE* added/lost *PRICE_VALUE*.

(b) The closing price of Nikkei stock average at Tokyo market on *DATE* added/lost *PRICE_VALUE*, and it got down to *PRICE_LEVEL*.

(c) The closing price of Nikkei stock average at Tokyo market on *DATE* added/lost *PRICE_VALUE*, and it has recovered at *PRICE_LEVEL*.

DATE, *PRICE_VALUE*, *PRICE_LEVEL* are the variables to be filled with numerical values obtained from the observed data stored in the data base.

And some phrases to explain the conditions of the price are usually added. For example, sentences as follows: '*The stock price was moving at high level throughout a day.*', '*The market closed with the high price of the day.*', etc.

2.5.2 *Type 2 text*

Type 2 texts are generated by composing words or short sentences to express the behavior of a chart with circumstantial information and conjunctions for the consistency among the sentences. Our system has 2 rules for adding circumstantial information and 4 rules about conjunctions. These rules reflect syntactic patterns used to explain the trends of the stock prices and are obtained through corpus analysis.

- Depending on time zone, the expression either 'in the morning session' or 'in the afternoon session' is put at the head of sentences.

- If the partial shapes of a chart can be expressed with the sentence either 'Buying is dominant' or 'Prices turn to be rising', one of the sentences expressing time interval: 'throughout the middle of trading', 'around the middle of

ISBN: 978-989-95079-6-8

trading', 'after the middle of trading' is put at the head of the sentence depending on the shape of an approximate function especially in terms of time interval.

- The partial shape expressed with the sentence 'The width of rising was small' comes with the conjunction either 'since' or 'but' at the head of the sentence, depending on the case whether or not there is a stock price trend expressed with the sentence 'It is likely to stay at a low level.' before the sentence.
- If there is not any short sentence expressing time or any conjunction before the sentence, furthermore, if there are sentences expressing the partial shape, e.g, 'Buying was ahead' before the sentence, the word 'afterwards' is inserted in between the two sentences.
- If there is a sentence describing 'Buying was happening' before the sentence 'It was hardly going down', a conjunction 'and then' is inserted in between them.
- The conjunction 'and then' is basically put at the head of the short sentences describing the width of rising or descending.
- Depending on the width of a shape expressed with a sentence 'Buying was increased', a short sentence describing 'after the middle of trading' is put at the head of the sentence.
- Depending on the width of a shape expressed with a sentence, 'the width of descending was gradually expanded', a sentence 'around the middle of trading' is put at the head of the sentence.

3 Generation Example

A screenshot of the developed system is shown in Fig. 6. The system generates type 1 and type 2 texts in Japanese, and displays the data of Nikkei stock average price on a particular day and also its 2D line chart.

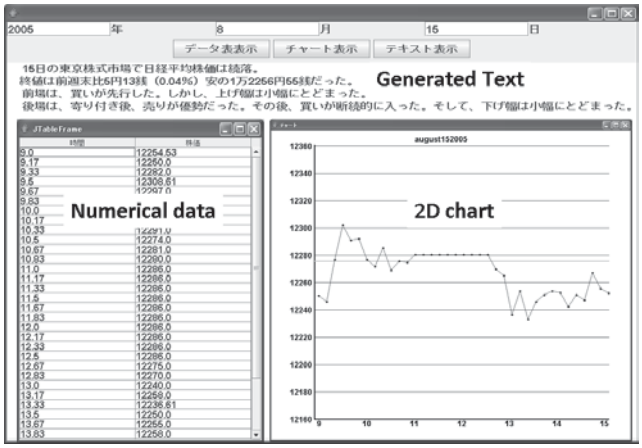


Figure 6: Screenshot of the system

3.1 Generation process of type 1 text

The generation process is explained by following generation steps explained in 2.1.

Here, let us assume that we want to get a report of the stock price trends on "August 15, 2005". This date information is the initial input information to the system.

step 1. Obtaining numerical data from a data base

Numerical data of the stock price on August 15, 2005 and the closing and the opening price, and the highest and the lowest prices data for the past three days are obtained from the data base.

step 2. Lexical selection for the shape of a chart

Proper lexes to express the obtained numerical data are selected from the dictionary.

- Date information is obtained from the input information, "August 15, 2005".
- From the obtained numerical data, the following events are found by comparing the states or levels of the points in the chart: The closing price of the stock is 'higher' compared to that of yesterday. The closing price of yesterday is 'cheaper' compared to that of the day before yesterday. The closing price of the day before yesterday is 'higher' compared to that of three days ago. Therefore, as a proper word to explain the trend of the stock price, '*rebound*' is selected.
- Compared the closing price to that of yesterday, the price of today is 160.78 yen higher (1.38 % higher), therefore, the actual values, in italic fonts, related to the price conditions were selected. 'Nikkei stock average added *160.78* yen, or *1.38* percent, to reach *12452.51* yen'.
- By the facts that the closing price of the day was over 12400 yen and the price of yesterday was under 12400 yen, a sentence, 'The closing prices has recovered at the level of *12400* yen' is generated.

step 3. Providing templates

- Compared the closing price of the day to that of the day before, the following template was selected — "The closing price of Nikkei stock average at Tokyo stock market on *DATE* added *PRICE_VALUE* and it has recovered at *PRICE_LEVEL*."
- The market closed with 25 yen higher of the closing price compared to the price of the day before, therefore, a sentence, '*It rose to a high level*' is added to a generated text.
- The closing price was higher than the opening price at both morning and afternoon sessions, and also, the closing price was higher than the opening price by 100 yen. Therefore, a sentence, '*The market price rose*' is added to a generated text.
- Compared to the closing price of the day before, the mean value between morning and afternoon sessions was over 25 yen, therefore, a sentence, '*The stock price was moving at a high level throughout the day*'.

The generated text of type 1 is as follows:
" The closing price of Nikkei stock average at Tokyo stock market on August 15, 2005 rebounded. It added 160.78 yen, or 1.38 percent, to reach 12452.51 yen and has recovered at the level of 12400 yen. It rose to high level. The market price rose. The stock price was moving at a high level throughout the day. "

ISBN: 978-989-95079-6-8

3.2 Generation process for type 2 text

step 1. Obtaining numerical data in a data base

The same process as shown in the generation process for type 1 text happens.

step 2. Recognition of the shape of a line chart

The shapes of a line chart at morning and afternoon sessions are recognized by using the least squares with the numerical data obtained at step 1.

In this example, the shapes of the chart at morning and afternoon sessions are recognized as type 1 and type 7 of whole the shape categories, respectively.

step 3. Lexical selection for the partial shapes of a chart

Proper sentences to express the behavior of a line chart are selected by recognizing the shape of the chart with verbal expressions (see, Figure 7).

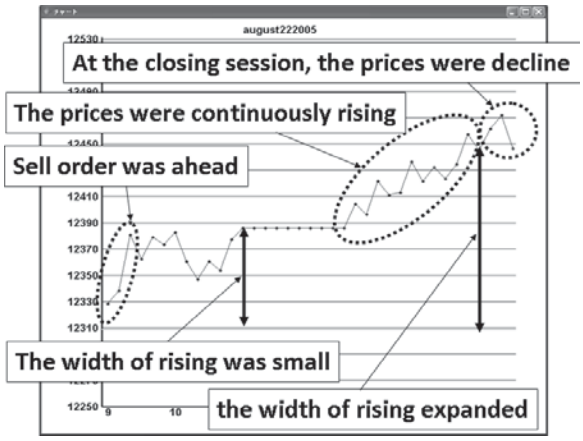


Figure 7: Shapes and their verbal expressions

- At morning session, '*Sell order was ahead.*', '*Trading was steady.*' '*The width of rising was small.*'
- At afternoon session, '*The prices were continuously rising*', '*The width of rising expanded*', '*The prices were decline*'.
- '*At the closing session*' is put ahead of '*the prices were decline*'.

step 4. Providing syntactic rules

Proper conjunctions and temporal information are added to the sentences generated at step 3 by following the syntactic rules for type 2 text generation.

- As for temporal information, '*At morning session*', and '*At afternoon session*' are added respectively at the head of the first sentence reporting an event of each session.
- A conjunction, '*afterwards*' is inserted between '*Sell order was ahead*' and '*Trading was steady.*'
- A conjunction, '*Therefore*' is put at the head of '*The width of rising was small.*'

The generated text of type 2 is as follows:
"At the morning session, sell order was ahead. Afterwards, trading was steady. therefore, the width of rising was small. At the afternoon session, trading was continuously rising. Therefore, The width of rising was expanded. At the closing session, the prices were decline. "

4 Related work

There are several studies on text generation for explaining the behavior of a line chart. Boyd [1] has applied Wavelet analysis to recognizing the characteristics of time-series data expressed with a line chart and proposed a method to express the characteristics of the chart with natural language based on the analysis result. Carberry et al. [2] is a work about summarizing graph information as well as textual information — here, summarizing graph information means expressing the graph information with natural language. Kobayashi [3] has investigated the relation between patterns of line chart behaviors and natural language to express the patterns, and has proposed an algorithm to generate a text explaining the behavior of a line chart. Furthermore, as studies to deal with time-series data, Goldberg et al. [4] and Coch [5] have developed systems that generate weather reports from weather data. As the studies that discover particular patterns from time-series data, Barndt et al. [6] dealt with economic data, Bakshi et al. [7] dealt with process control data, Himowitz et al. [8] dealt with medical data, etc.

5 Conclusions

As a method to transform non-verbal information into verbal information, we have proposed a method to generate a text by recognizing the shape of a 2D line chart. In particular, we used the linear least squares to make an approximate line chart against the original chart and then expressed the shape of the approximate line with words to generate a text reporting the stock price trends. This approach is, unlike the approach using Wavelet analysis, to verbalize the behavior of charts from the viewpoint of the words often used to verbalize time-series data, especially stock price trends, in the case of this paper. We verified our proposed method by comparing the generated reports with actual news articles. Although we used the same data for generating reports and for evaluation of the system, we have got 89.8 % precision rate between the generated report and news article for the same period of 27 days. This result shows us our system can properly verbalize the behavior of time-series data.

The proposed method is useful for reporting the trends of time-series data, and also for mining knowledge from the data with natural language. Furthermore, since the method changes visual information, i.e., 2D charts in the case of this paper, into verbal information, there is possibility to provide blind people with an interface which can report what is happening in time-series data via voice. In this context, the proposed method provides us with high accessibility and usability to various kinds of information. This provides high interaction among multimodal information.

References

[1] S. Boyd. Trend: A system for generating intelligent descriptions of time-series data. *Proc. IEEE-ICIPS.*, 1998.

[2] S. Carberry, S. Elzer, K. Green, N.and McCoy, and D. Chester. Extending document summarization to information graphics. *Proc. ACL Workshop on Text Summarization.*, 2004.

[3] I. Kobayashi. A study on text generation from non-verbal information on 2d charts. *In Computational Linguistics and Intelligent Text Processing*, pages 226–238, 2001.

ISBN: 978-989-95079-6-8

[4] E.Goldberg and N.Driedger. Using natural language processing to produce weather forecasts. *IEEE Expert.*, 1994.

[5] Jose Coch. Interactive generation and knowledge administration in multimeteo. *Proc. The 9th International Workshop on Natural Language Generation (INLG-98)*, pages 300–303., 1998.

[6] D.J.Berndt and J.Clifford. Finding patterns in time series: A dynamic programming approach. *Advances in Knowledge Discovery 1996*, 1996.

[7] B.R.Bakshi and G.Stephanopoulos. Reasoning in time. *Advances in Chemical Engineering*, 22:485–548, 1995.

[8] I.J.Himowitz, P. Phuc Le, and I.S.Kohane. Clinical monitoring using regression-based trend templates. *Artificial Intelligence in Medicine*, 7:473–496., 1995.

ISBN: 978-989-95079-6-8

Differential Equations based on Fuzzy Rules

Marina R. B. Dias, Laécio C. Barros[1]

1.Department of Applied Mathematics, IMECC, State University of Campinas
Campinas, SP - Brazil
Email: {madias, laeciocb}@ime.unicamp.br

***Abstract*— *The purpose of this paper is to use the Fuzzy Set Theory, specially fuzzy rule-based systems, as a mathematical tool for modeling dynamical systems given by differential equation, whose direction field is f (IVP-f). As fuzzy systems methodology produces input-output systems, that is, it produces m outputs from n inputs, one can see it as $f_r : \mathbb{R}^n \to \mathbb{R}^m$, where r is the number of rules of the rule base. Under certain assumptions, one can prove that the functions (f_r) have good analytical properties. Beyond it, they can approximate theoretical functions (f). The main result of this paper states that the solutions x of an ODE whose direction field is f (IVP-f) can be approximated by the solutions x_r of an ODE whose direction field is f_r (IVP-f_r). By utilizing important theorems as the Lebesgue's Dominated Convergence Theorem it was possible to prove the proposed theorem. Finally we use a numerical method (Runge-Kutta) to simulate solutions x_r^n that approximates x_r when n increases. The solutions x_r^n approximates the solution x of IVP-f also provided $\{x_r^n\} \overset{n\to\infty}{\longrightarrow} x_r \overset{r\to\infty}{\longrightarrow} x$.*

***Keywords*—** approximation theory, dynamical systems, fuzzy systems, initial value problem, numerical solutions.

1 Introduction

The investigation of a physical (biological) phenomena using differential equations needs knowledgement of the direction field. That is, it is necessary to adopt a theoretical function to represent the variation rates (derivatives) as function of the state variables. For deterministic differential equations, this direction field is a function of real variables that models its variation in time. Traditionally, aleatory uncertainties have been studied from Differential Equations in two ways: first, considering the state variables as aleatory variables. In this case it is necessary to define a variation rate (or derivative) of an aleatory variable. So, it is called stochastic differential equation. Second, considering only the aleatory parameters (therefore, it is not necessary a new definition of variation rate). These are called aleatory differential equations [1, 2]. Recently, the use of Fuzzy Set Theory to incorporate uncertainties in differential equations has been increased, so we have the fuzzy differential equations. Like the stochastic case, there are many proposals for studying these equations. The first one, using the notion of fuzzy derivative, is originated of the Hukuhara's derivative used to multifunctions and introduced in the fuzzy area by Puri and Ralescu [3]. Another notions do not require derivative to fuzzy variables. The variation rate is the same as the classical one, but some proposals have already been suggested: use differential inclusions [4, 5], use some fuzzification procedure for a deterministic solution [6], etc. In each one of these proposals there is an equation. That is, the derivatives are given by direction fields that depend on the state variables. Our research suggests a new methodology to study fuzzy differential equations using the fuzzy set theory. It will not be required to have the variation rates explicitly as a function of the state variables. We propose a fuzzy relation to model the association between the state variables and the direction field. We consider the variation rate and the state variables as linguistic variables. If no defuzzification procedure is used or if the used method produces an interval [7, 8], then the system can be studied with the aid of differential inclusion theory. The methodology we propose here uses some defuzzification method, which produces a crisp solution (real number or real vector) for each instant of time t. Regarding fuzzy differential equations, the methodology we proposed differs from previous methodologies in its formulation. While other methodologies use the concept of derivative (classical or fuzzy), our methodology uses a fuzzy formulation given by a rule base. More specifically, we adopt a fuzzy rule-based system (FRBS) to represent the differential equation. The FRBS methodology has been very important to solve modeling problems, since it is a universal approximator. For universal approximator we mean something that can approximate a continuous functions with the desired accuracy in compact sets. Thus, it is expected that fuzzy rule-based systems can model a direction field when the direction field is only partially known. Works such as Nguyen [9, 10] were extremely useful in our research. The main purpose of this work is to use fuzzy systems to obtain functions (f_r) in accordance with certain properties of the phenomenon being modeled. These families of functions will be constructed in such a way that they will be universal approximators of the function f that represents a theoretical field. Our greatest interest is to study the solutions of these new dynamical systems, given by the direction field (f_r) and establish conditions that ensure convergence of its solutions to the solution of the theoretical problem. The structure of this paper is as follows: in Section 2 we give all basic concepts and definitions the reader needs for a good understanding of our work. In Section 3 we present the problem we must solve, that is, the new IVP given by a fuzzy rule-based system. In Section 4 we present our main result as a theorem and then we prove it. Finally, in Section 5 we show an example to illustrate our result.

2 Basic Concepts and Definitions

2.1 Fuzzy Set Theory

An ordinary subset A of a set U is determined by its characteristic function χ_A defined by

$$\chi_A(x) = \begin{cases} 1, & if\ x \in A \\ 0, & if\ x \notin A \end{cases}.$$

ISBN: 978-989-95079-6-8

The characteristic function of a subset A of a set U specifies whether or not an element is in A. Zadeh [11] generalized this notion by allowing images of elements to be in the interval $[0, 1]$ rather than being restricted to the two element set $\{0, 1\}$.

Definition 1 *A fuzzy subset A of a set U is determined by a function $\varphi_A : U \to [0, 1]$, where the number $\varphi_A(x)$ represents the degree of membership of the element x to the fuzzy subset A.*

Definition 2 *A binary operation $\Delta : [0, 1] \times [0, 1] \to [0, 1]$ is a* t-norm *if it satisfies the following:*

1. $\Delta(1, x) = 1\Delta x = x$;
2. $\Delta(x, y) = x\Delta y = y\Delta x = \Delta(y, x)$;
3. $x\Delta(y\Delta z) = (x\Delta y)\Delta z$;
4. *If $x \leq u$ and $y \leq v$ then $x\Delta y \leq u\Delta v$.*

The operator t-norm models the connective "and".

Definition 3 *A binary operation $\nabla : [0, 1] \times [0, 1] \to [0, 1]$ is a* t-conorm *if and only if*

1. $\nabla(0, x) = 0\nabla x = x$;
2. $\nabla(x, y) = x\nabla y = y\nabla x = \nabla(y, x)$;
3. $x\nabla(y\nabla z) = (x\nabla y)\nabla z$;
4. *If $x \leq u$ and $y \leq v$ then $x\nabla y \leq u\nabla v$.*

The operator t-conorm models the connective "or".

Definition 4 *By a* defuzzification procedure $\mathcal{D}$ *we mean a mapping that transforms a membership function $\varphi(x)$ into a number and satisfies the following properties:*

1. *if $\varphi(x) = 0$ for all $x \in (-\infty, a)$, then $\mathcal{D}(\varphi) \geq a$;*
2. *if $\varphi(x) = 0$ for all $x \in (-\infty, a]$, then $\mathcal{D}(\varphi) > a$;*
3. *if $\varphi(x) = 0$ for all $x \in (a, +\infty)$, then $\mathcal{D}(\varphi) \leq a$;*
4. *if $\varphi(x) = 0$ for all $x \in [a, +\infty)$, then $\mathcal{D}(\varphi) < a$.*

2.2 Fuzzy Rule-Based System

A Fuzzy Rule-Based System (FRBS) (Fig. 1) is an input-output system based on fuzzy rules. It is composed of four components: an input processor; a collection of linguistic rules, called rule base; a fuzzy inference method and an output processor, which generates a real number as output.

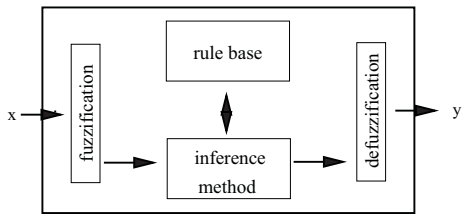


Figure 1: A FRBS scheme.

The fuzzification is the stage where the system's entries are modeled by fuzzy sets. The membership functions are constructed for each fuzzy set involved in the process and, even if the entry is crisp, it will be fuzzified by its characteristic function. The rule base is composed of a set of "if-then" rules describing an input-output relationship. In this component we can find the fuzzy propositions, which are provided in accordance with an expert. The defuzzification translates the result of fuzzy inference into a single numeric value of control (a real number) [12].

A rule base consisting of r rules has the form:

$$\begin{array}{c} R_1 : \text{"Fuzzy Proposition 1"} \\ \text{or} \\ R_2 : \text{"Fuzzy Proposition 2"} \\ \text{or} \\ \vdots \\ \text{or} \\ R_r : \text{"Fuzzy Proposition r"}. \end{array}$$

In general, each fuzzy proposition has the form

$$\text{If "}state\text{" then "}action\text{"},$$

where every "state" and every "action" are assumed as linguistic variables modeled by fuzzy sets.

In mathematical terms, each one of the r propositions is described as

$$\begin{array}{c} \text{"If } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \text{ and} \ldots x_n \text{ is } A_n \text{ then} \\ y_1 \text{ is } B_1 \text{ and } y_2 \text{ is } B_2 \text{ and} \ldots u_m \text{ is } B_m\text{"}, \end{array}$$

where $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_m)$ are the input and output variables, respectively. A_i, $1 \leq i \leq n$ and B_j, $1 \leq j \leq m$ are fuzzy sets which model each one of the terms assumed by the input and output variables, respectively.

It is important to note that a FRBS maps $\mathbb{R}^n$ in $\mathbb{R}^m$ of a specific manner. Given an n-tuples $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, the FRBS produces an output $y = y(y_1, y_2, \ldots, y_m) \in \mathbb{R}^m$. For a set of rules of the form "If x_i is A_{ij} then y is B_j, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, r$", where $x_i,\ y_i \in \mathbb{R}$, A_{ij} and B_j are fuzzy sets, n is the number of input variables and r is the number of rules, we obtain $y_r = f_r(x_1, x_2, \ldots, x_n)$. This map always depends on: a) the membership functions A_{ij} and B_j; b) the t-norm and t-conorm chosen and c) the defuzzification method.

Many authors proved that, for particular methodologies, a fuzzy rule-based system is a universal approximator [13, 14, 15, 16], but Nguyen [9, 10] generalized and extended these results. Due to importance of his work we will report some aspects of it in here.

Let a general class of functions $\mathcal{F}(\mathcal{M}, \mathcal{L}, \mathcal{D})$, where $\mathcal{M}$ consists of those membership functions φ such that $\varphi(x) = \varphi_0(ax + b)$ for some $a, b \in \mathbb{R}$, and $a \neq 0$, and $\varphi_0(x)$ is continuous, positive on some interval of $\mathbb{R}$, and 0 outside that interval. $\mathcal{L}$ consists of continuous t-norms and t-conorms. $\mathcal{D}$ is a defuzzification procedure transforming each membership function into a real number in such a way that if $\varphi(x) = 0$ outside an interval (α, β), then $\mathcal{D}(\varphi) \in [\alpha, \beta]$.

Theorem 1 *For any design methodology $(\mathcal{M}, \mathcal{L}, \mathcal{D})$ and any compact subset K of $\mathbb{R}^n$, $\mathcal{F}(\mathcal{M}, \mathcal{L}, \mathcal{D})|_K$ is dense in $C(K)$ with respect to the sup-norm.*

ISBN: 978-989-95079-6-8

Proof: See [10].

According to Nguyen [10], for a FRBS with a $(\mathcal{M}, \mathcal{L}, \mathcal{D})$ methodology, the output f_r is continuous (therefore integrable) and it can be written as

$$f_r(x) = \frac{\sum_{j=1}^{r} y_j \ \Delta(\varphi_{A_{1j}}(x_1) \ \Delta \ ... \ \Delta \ \varphi_{A_{nj}}(x_n))}{\sum_{j=1}^{r}(\varphi_{A_{1j}}(x_1) \ \Delta \ ... \ \Delta \ \varphi_{A_{nj}}(x_n))}, \quad (1)$$

where r is the number of rules of the rule base, $\varphi_{A_{ij}}(x_i)$ is the degree of membership of x_i in the fuzzy set A_{ij} and Δ is a t-norm (see Definition 2).

Besides, the function f_r is bounded, since it is continuous and defined in a compact set.

Note that $f_r(K)$ is a compact set of $\mathbb{R}$, that is, $f_r(K)$ is a closed interval of $\mathbb{R}$.

3 The Problem

3.1 Construction of an Initial Value Problem and its solution

Suppose that the Initial Value Problem (IVP) models a particular phenomenon. Moreover, suppose that it is an Autonomous Ordinary Differential Equation, that is, the rate of change does not depend explicitly of the time. The IVP is, therefore,

$$\begin{cases} \dfrac{dx}{dt} = f(x(t)) \\ x(t_0) = x_0 \end{cases}, \quad (2)$$

where f is a known function.

The solution of the IVP (2) is guaranteed by the following proposition:

Proposition 1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous. The function $x : [a, b] \to \mathbb{R}^n$ is a solution of the IVP (2) if, and only if, it is continuous and satisfies the integral equation*

$$x(t) = x_0 + \int_{t_0}^{t} f(x(s))ds,$$

$t \in [a, b]$.

In practice, depending on the complexity of the field f, the solution $x(t)$ may not have an analytical expression. However, there are several numerical methods that estimate the unknown solution with the desired accuracy, that is, we can use a numerical method to obtain a numerical solution to the IVP (2) so that $\{x^n\} \overset{n\to\infty}{\longrightarrow} x$. Note that knowing the direction field f (or at least knowing it in some points) is necessary to use a numerical method.

We propose an alternative way to obtain the solution of the IVP (2) or at least an approximation of it, without knowing the field f explicitly. Our goal is to take advantage of the qualitative information available to construct a rule base that will represent the properties that characterize the phenomenon. To obtain the output f_r we use the methodology presented in Section 2.2 and therefore we guarantee that f_r is a good approach to f.

Thus, we have a new IVP to represent the studied phenomenon, which is given by

$$\begin{cases} \dfrac{dx}{dt} = f_r(x(t)) \\ x_r(t_0) = x_0 \end{cases}, \quad (3)$$

whose solution will be denoted by x_r and it will be studied later.

In the next subsection we will make an analysis about the existence and uniqueness of the solution x_r of the IVP (3). Since we can guarantee its existence, we will describe the main result of this work: the theorem ensures that the solution of the IVP (3) is arbitrarily close to the solution of the IVP (2) as r increases.

3.2 Existence and uniqueness of the new IVP's solution, x_r

Based on Proposition 1, if $f_r : \mathbb{R}^n \to \mathbb{R}$ is continuous, then the solution of (3) is given by

$$x_r(t) = x_0 + \int_{t_0}^{t} f_r(x(s))ds,$$

$t \in [a, b]$.

To ensure the existence and uniqueness of the solution we have the following result:

Proposition 2
- *If $f \in C_0(\mathbb{R}^n, \mathbb{R})$ then, for all $x_0 \in \mathbb{R}^n$, there is an interval $I_{x_0} \equiv (\alpha_{x_0}, \beta_{x_0})$ containing t_0 and a solution $x(., x_0)$ of the initial value problem that satisfies the initial condition $x(t_0, x_0) = x_0$.*
- *Besides, if $f \in C^1(\mathbb{R}^n, \mathbb{R})$, then $x(., x_0)$ is unique in I_{x_0} and it is a function of C^1 class.*

As we have seen in section 2.2, f_r is continuous and integrable. Therefore, the function f_r satisfies both conditions of the Proposition 2 and so we can guarantee that the solution of the IVP (3), x_r, exists and it is unique.

4 Main Result

In this section we will state the theorem that ensures the convergence of solution of the IVP (3), x_r, to the solution of the IVP (2), x, as r increases.

For a demonstration of this result we will make use of the following Lebesgue's Dominated Convergence Theorem:

Theorem 2 *Lebesgue's Dominated Convergence Theorem (LDCT): Let X be a measure space with a σ-algebra $\mathcal{A}$ and a measure μ on it. Let $f, f_1, f_2, \ldots : X \to C$ be measurable functions such that $f_n \overset{n\to\infty}{\longrightarrow} f$ for almost everywhere (that is, $\mu(\{x \in X; f_n(x) \nrightarrow f(x)\}) = 0$). Suppose there is an integrable function $g : X \to [0, +\infty]$ such that $|f_n(x)| \leq g(x)$ for almost every $x \in X$. Then $\int f_n d\mu \overset{n\to\infty}{\longrightarrow} \int f d\mu$.*

Proof: See [17].

In the following we state the main result of this work.

Theorem 3 *Let f be a continuous function in a compact set $K \subset \mathbb{R}^n$. Let $\{f_r\}$ be class of functions defined in K such that $f_r \overset{r\to\infty}{\longrightarrow} f$, for example, the fuzzy system output as in (1). Consider the IVPs (2) and (3) with x and x_r their solutions, respectively.*

Then $x_r \overset{r\to\infty}{\longrightarrow} x$.

ISBN: 978-989-95079-6-8

Proof: We know that the solutions x and x_r are given, respectively, by the equations:

$$x(t) = x_0 + \int_{t_0}^{t} f(x(s))ds,$$

and

$$x_r(t) = x_0 + \int_{t_0}^{t} f_r(x(s))ds.$$

Using the LDCT we can guarantee that the solution x_r converges to x as r increases if there is a positive and integrable function g such that $|f_r(x)| \leq g(x)$ in μ - for almost $x \in K$.

Let $g : K \to [0, +\infty]$ be such that $g(x) = max\{|\alpha|, |\beta|\}$, where $[\alpha, \beta] \subset \mathbb{R}$ is an interval such that $\mathcal{D}(\varphi(x)) \in [\alpha, \beta]$, for all $x \in K$ and $\mathcal{D}$ is a defuzzification procedure as we have seen in Section 2.2. Moreover g is a constant function in a compact set, then it is integrable. Therefore we have $|f_r(x)| \leq g(x)$ for every $x \in K$ and all r.

So, the hypothesis of LDCT are satisfied and we can conclude that

$$\int_{t_0}^{t} f_r(x(s))ds \overset{r\to\infty}{\longrightarrow} \int_{t_0}^{t} f(x(s))ds$$

and, therefore,

$$x_r \overset{r\to\infty}{\longrightarrow} x.$$

4.1 Numerical Solution

In practice, the solution, x, of a theoretical dynamical system cannot have an analytical expression due to the complexity of the direction field f. In this case it is necessary to use numerical methods, and then find solutions $\{x^n\}$ such that $\{x^n\} \overset{n\to\infty}{\longrightarrow} x$.

However $\{x^n\}$ will only be obtained if the direction field f is known, that is, knowing the direction field directions f is a restriction of the numerical method to produce the estimated $\{x^n\}$.

When there is no explicit knowledge of the field f, we can use the fuzzy sets theory and, as we have seen before, we can use fuzzy rule-based system to produce a function f_r that can replace the field f, getting the IVP (3).

Knowing f_r in a table form (for each x, we have $f_r(x)$), we can use stable and convergent numerical methods to obtain $\{x_r^n\}$ such that $\{x_r^n\} \overset{n\to\infty}{\longrightarrow} x_r$.

Therefore, as long as the fuzzy rule-based system used to produce f_r satisfies the assumptions of the main theorem, we have

$$\{x_r^n\} \overset{n\to\infty}{\longrightarrow} x_r \overset{r\to\infty}{\longrightarrow} x,$$

that is

$$x_r^n \overset{\substack{n\to\infty \\ r\to\infty}}{\longrightarrow} x.$$

Now we will show a practical example of the theory discussed here. More details can be found in [18].

5 Example

Suppose a population that has its dynamics following the qualitative characteristics of the Verhulst model (Logistic Equation). Thus, this population grows as the IVP

$$\begin{cases} \dfrac{1}{x}\dfrac{dx}{dt} = a(P_\infty - x) \\ x(t_0) = x_0 \end{cases}, \tag{4}$$

where a is the relative growth rate and P_∞ is the carrying capacity of the population (inhibition factor).

The classical solutions of (4), which represent the population $x(t)$ at every instant t, are given by

$$x(t) = \frac{P_\infty}{\left(\dfrac{P_\infty}{x_0} - 1\right)e^{-at} + 1}, \tag{5}$$

from where we conclude that:

- x increases if $x_0 < P_\infty$
- x is constant if $x_0 = P_\infty$
- x decreases if $x_0 > P_\infty$,

and is illustrated in Fig. 2.

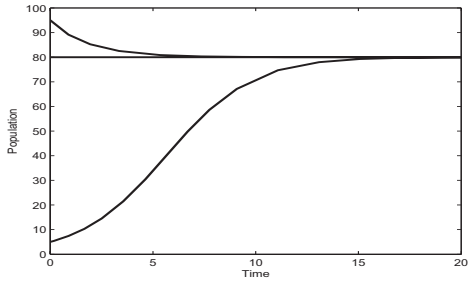


Figure 2: Possible solutions of the Verhulst model.

Considering the relative growth rate we have, rewriting (4) that

$$\frac{1}{x}\frac{dx}{dt} = a(P_\infty - x) = f(x), \tag{6}$$

where f is linear (decreasing).

Consider a FRBS with the Mamdani inference and the centroid as the defuzzification procedure. The input variable is *population* and the output variable is *relative growth rate* (here we will denote it by variation rate). The solution $x_r(t)$ is obtained using a Runge-Kutta method for each output $f_r(x)$.

Next, we have two simulations given by a fuzzy system with 3 and 8 rules in the rule base, respectively. Note that in all simulations we have a rule base that presents semantic opposition in the consequents, that is, there exists change of the consequents signals. That is what guarantees the "auto-inhibition" characteristic.

5.1 Simulation 1

In this first simulation we have three rules in the rule base. They are as follows:

R_1: If the population is *"small"* then the variation rate is *"positive medium"*
R_2: If the population is *"medium"* then the variation rate is *"positive small"*
R_3: If the population is *"large"* then the variation rate is *"negative small"*

The membership functions of the input and output variables are in Figs. 3 and 4, respectively.

The solution $x_r(t)$ is showed in Fig. 5.

ISBN: 978-989-95079-6-8

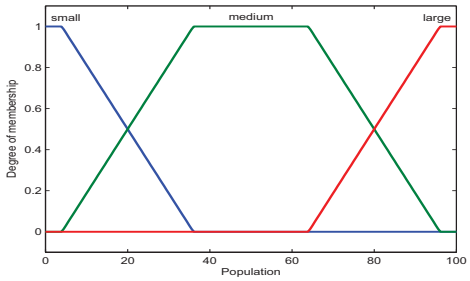


Figure 3: Input of variable *Population*.

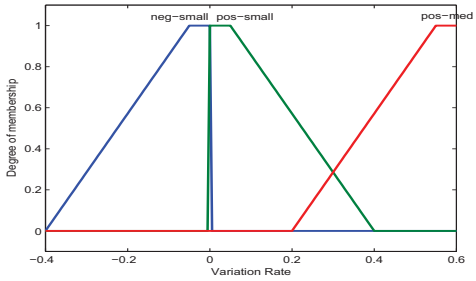


Figure 4: Input of variable *Variation Rate*.

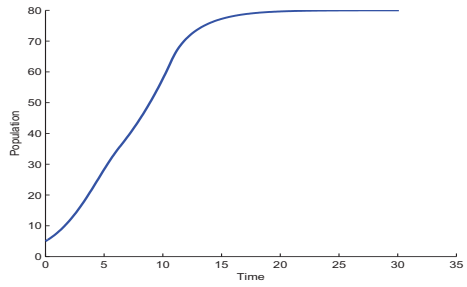


Figure 5: Solution $x_r(t)$ of simulation 1.

5.2 *Simulation 2*

In this second simulation we have eight rules in the rule base. They are as follows:

R_1: If the population is *"very small"* then the variation rate is *"positive very large"*
R_2: If the population is *"small A"* then the variation rate is *"positive large"*
R_3: If the population is *"small B"* then the variation rate is *"positive large"*
R_4: If the population is *"medium A"* then the variation rate is *"positive medium"*
R_5: If the population is *"medium B"* then the variation rate is *"positive medium"*
R_6: If the population is *"medium C"* then the variation rate is *"positive small"*
R_7: If the population is *"large"* then the variation rate is *"negative small"*
R_8: If the population is *"very large"* then the variation rate is *"negative medium"*

The membership functions of the input and output variables are in Figs. 6 and 7, respectively.

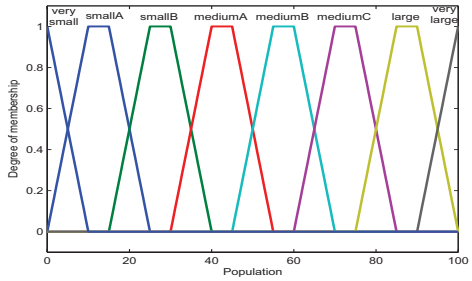


Figure 6: Input of variable *Population*.

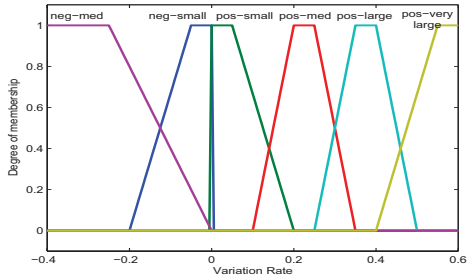


Figure 7: Input of variable *Variation Rate*.

The solution $x_r(t)$ is showed in Figure 8

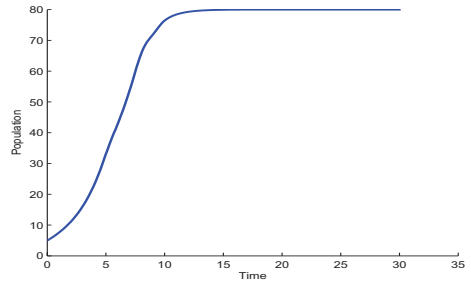


Figure 8: Solution $x_r(t)$ of simulation 2.

Fig. 9 shows the two solutions obtained in the simulations and the classical solution of the model, when adopted $K = 80$, $a = 0,006$ and $x_0 = 5$.

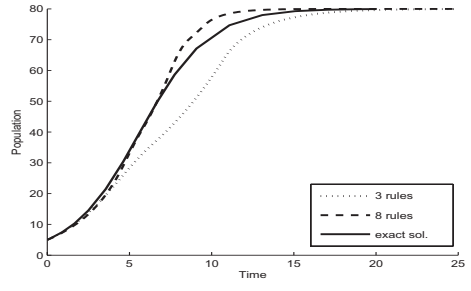


Figure 9: Comparison of the solutions obtained through simulations and the exact solution, using $x_0 = 5$.

Fig. 10 shows the two solutions obtained in the simulations

ISBN: 978-989-95079-6-8

and the classical solution of the model, when adopted $K = 80$, $a = 0,006$ and $x_0 = 90$.

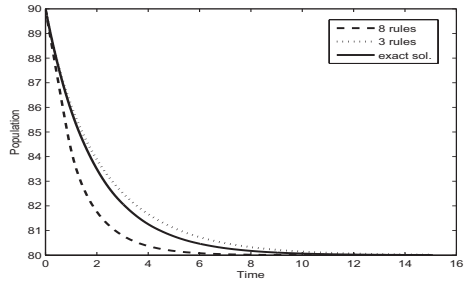


Figure 10: Comparison of the solutions obtained through simulations and the exact solution, using $x_0 = 90$.

Note that Figs. 9 and 10 show that the solutions of the IVP (3) converge to the exact solution of Verhulst Model. Moreover, the bigger the number of rules is, the better the approximation is. This is in accordance with Theorem 3, which ensures convergence as r increases. Table 1 shows the values obtained in the simulations and the value obtained by the solution of the Verhulst Model when $x_0 = 5$, for some values of t.

Table 1: Values obtained by simulation and by the equation.

Time	Theoretical Value (TV)	Sim. 1 (S1) (3 rules)	Sim. 2 (S2) (8 rules)
0	5	5	5
0,2	5,4152	5,4644	5,4460
0,9	7,4423	7,3920	7,2719
2,5	14,5304	13,7874	13,2925
4,6	30,4040	25,8578	28,7976
7,7	58,6421	42,7152	62,3510
11	74,6871	65,6782	78,3973
15,1	79,2707	77,3232	79,9304
17,1	79,7398	78,7905	79,9850
20	79,9407	79,6222	79,9984

The norm of the error of these approximations can be seen in Table 2. It is the difference between the exact value and the simulated values.

Table 2: Errors obtained comparing simulations 1 and 2 with the theoretical value.

Time	Error 1 ($\|TV - S1\|$)	Error 2 ($\|TV - S2\|$)
0	0	0
0,2	0,0491	0,0308
0,9	0,0504	0,1704
2,5	0,7430	2,2487
4,6	4,5416	1,6063
7,7	15,9269	3,7089
11	9,0059	3,7132
15,1	1,9475	0,6597
17,1	0,9493	0,2452
20	0,3185	0,0577

6 Conclusions

In this work we proposed a fuzzy rule-based system methodology to represent the direction field of differential equations. It is applicable when we have only partial knowledgement about the direction field in question. A typical case is when the informations about it are qualitative, as example, low, medium, high, conform the values of the state variable. As we have seen, this methodology was able to construct an input-output system from qualitative information by means of fuzzy sets. Thus, it was possible to adopt a direction field (f_r) (IVP (3)) obtained from theoretical direction field (f) (IVP (2)), that is given by Theorem 1, since f_r converges to f. Moreover, from theorem 3 and using the Lebesgue's Dominated Convergence Theorem we have proved that the solution of the IVP (3) converges to the solution of the IVP (2). Through a simple example we illustrated the methodology we proposed, confirming the obtained results. From the point of view of modeling this result is very important because in many cases, the available informations are qualitative. That is the case of Peixoto's work [19] that applies this methodology in a predator-prey model.

Acknowledgment

Authors thank CNPq (307890/2006-6 and 140297/2006-5) and FAPESP (06/05920-7).

References

[1] M. Turelli. Stochastic community theory: A partially guided tour. *Biomathematics*, 17:321 – 339, 1986.

[2] L. Arnold. *Random dynamical systems*. Springer, Verlag-Berlin-Heidelberg, Germany, 2nd edition.

[3] M. L. Puri and D. A. Ralescu. Differential of fuzzy functions. *Journ. Math. Analysis and Applications*, 91:552 – 558, 1983.

[4] H. Hullemeier. An approach to modelling and simulation of uncertain dynamical systems. *Intern. J. Uncertainty Fuzziness Knowledge-Based Systems*, 5:117–137, 1997.

[5] V. A. Baisodov. Fuzzy differential inclusions. *PMM USSR*, 54(1):8 – 13, 1990.

[6] M. T. Mizukoshi, L. C. Barros, Y. Chalco-Cano, Rman-Flores, and R. C. Bassanezi. Fuzzy differential equation and the extension principle. *Information Sciences*, 177, 2007.

[7] B. Bouchon-Meunier, D. Dubois, L. Godo, and H. Prade. *Fuzzy Sets in Approximate Reasoning and Information Systems*, chapter Fuzzy sets and possibility theory in approximate and plausible reasoning, pages 15–190. The Handbooks of Fuzzy Sets. Kluwer Academic Publisher, 1999.

[8] D. A. Ralescu, Y. Ogura, and S. Li. Set defuzzification and choquet integrals. *International J. Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:1 – 12, 2001.

[9] H. T. Nguyen, V. Kreinovich, and O. Sirisaengtaksin. Fuzzy control as a universal control tool. *Fuzzy Sets and Systems*, 80:71 – 86, 1996.

[10] H. T. Nguyen and E. A. Walker. *A First Course in Fuzzy Logic*. Chapman & Hall/CRC, 2000.

[11] L. A. Zadeh. Fuzzy sets. *Informat. Control*, 8:338 – 353, 1965.

[12] F. Gomide W. Pedrycz. *Fuzzy Systems Engineering - Toward Human-Centric Computing*. IEEE Press, John Wiley & Sons, Inc., 2007.

ISBN: 978-989-95079-6-8

[13] L. C. Barros and R. C. Bassanezi. *Tópicos de Lógica Fuzzy e Biomatemática*. Coleção IMECC Textos Didáticos, Campinas, 2006.

[14] J. J. Buckley and Y. Hayashi. Fuzzy input-output controllers are universal approximators. *Fuzzy Sets and Systems*, 58:273–278, 1993.

[15] B. Kosko. Fuzzy sets as universal approximators. *Proc. of IEEE Int. Conf. on Fuzzy Systems*, pages 1153–1162, Marc 8–12 1992.

[16] B. Kosko and J. A. Dickerson. *Theoretical Aspects of Fuzzy Control*, chapter Function Approximation with Additive Fuzzy Systems, pages 313–347. Wiley, 1995.

[17] A. A. Castro Jr. *Curso de teoria da Medida*. IMPA, Rio de Janeiro, 2004.

[18] M. R. B. Dias. Equações diferenciais ordinárias com campo de direções parcialmente conhecido. Master's thesis, IMECC - Unicamp, 2006.

[19] M. S. Peixoto, L. C. Barros, and R. C. Bassanezi. Predator-prey fuzzy model. *Ecological Modelling*, 214:39 – 44, 2008.

ISBN: 978-989-95079-6-8

Sparsely Connected Autoassociative Fuzzy Implicative Memories for the Reconstruction of Large Gray-Scale Images

Marcos Eduardo Valle
Department of Mathematics, University of Londrina
Londrina, Paraná, Brazil
Email: valle@uel.br

Abstract— *Autoassociative fuzzy implicative memories (AFIMs) are models that exhibit optimal absolute storage capacity and an excellent tolerance with respect to incomplete or eroded patterns. Thus, they can be effectively used for the reconstruction of gray-scale images. In practice, however, applications of AFIMs are confined to images of small size due to computational limitations. This paper introduces a class of sparsely connected AFIMs (SCAFIMs) that circumvent this computational overhead and, therefore, can be used for the reconstruction of large images. We show that SCAFIMs exhibit optimal absolute storage capacity and tolerance with respect to incomplete or eroded patterns. Furthermore, we compare the performance of SCAFIMs with their corresponding fully connected AFIM both theoretically and by means of computational experiments.*

Keywords— Fuzzy associative memory, reconstruction of large gray-scale images, sparse encoding, storage capacity, tolerance with respect to noise.

1 Introduction

Associative memories (AMs) are models inspirated in the human brain ability to recall by association [1, 2, 3, 4]. Here, a partial or approximate representation of a stored item is used to recall the full item. An example would be recalling a poem by knowing its first words as an initial clue. Applications of AM models include classification, prediction, control, and pattern recognition [5, 6, 7, 8, 9].

We speak of a *fuzzy associative memory* (FAM) if the AM model is used for the association of fuzzy sets [5, 8, 10, 11]. The *max-min* and *max-product FAMs* of Kosko, the *generalized FAMs* (GFAMs) of Chung and Lee, and the *implicative fuzzy associative memories* (IFAMs) are instances of FAMs [5, 8, 12, 13]. We would like to recall, however, that the FAMs of Kosko as well as the GFAMs usually fail to perfectly recall a memorized item due to crosstalk between the stored patterns [8, 12]. Therefore, despite their successful applications in a variety of problems [5, 12], the FAMs of Kosko and the GFAMs are not recommended for the reconstruction of gray-scale images.

In contrast, IFAMs exhibit *optimal absolute storage capacity* in the autoassociative case [13]. In other words, one can store has many patterns as desired in an *autoassociative fuzzy implicative memory* (AFIM). In addition, AFIMs exhibit one step-convergence and an excellent tolerance with respect incomplete or eroded patterns [8, 13]. As a consequence, AFIMs can be effectively used for the reconstruction of gray-scale images [8, 14].

Due to computational limitations, however, AFIMs can be used only for the storage and recall of images of small size. Specifically, such as several other AMs, including the *complex-valued neural network* [15, 16], the multi-state model of Costantini *et al.* [17, 18], and the *morphological associative memories* [14, 19], AFIMs require the storage of a large amount of synaptic junctions if used to process large images. For example, approximately 512 gigabytes of memory space would be consumed by an AFIM that is used for the storage of gray-scale images of size 512×512. In order to circumvent this computational overhead, we propose to remove a large amount of synaptic junctions of an AFIM. The resulting model is referred to as a *sparsely connected AFIM* (SCAFIM). We show that SCAFIMs inherit the optimal absolute storage capacity and the tolerance with respect to incomplete or eroded patterns of AFIMs. Therefore, they can be used for the reconstruction of large gray-scale images.

The paper is organized as follows. The next section reviews the basic concepts of AFIMs. SCAFIMs are introduced in Section 3. Some theoretical results concerning the storage capacity and noise tolerance of SCAFIMs, as well as two strategies for the construction of a SCAFIM from a (fully connected) AFIM, are also given in Section 3. Section 4 provides computational experiments. The paper finishes with the concluding remarks in Section 5.

2 A Brief Review on Autoassociative Fuzzy Implicative Memories

Associative memories (AMs) are input-output systems that store a set of associations $\{(\mathbf{x}^\xi, \mathbf{y}^\xi) : \xi = 1, \ldots, k\}$, called *fundamental memory set* [2, 3]. Formally, an AM corresponds to a mapping G such that $G(\mathbf{x}^\xi) = \mathbf{y}^\xi$ for every $\xi = 1, \ldots, k$. In addition, an AM model should be endowed with a certain tolerance with respect to noise. In other words, the mapping G should be such that $G(\tilde{\mathbf{x}}^\xi)$ equals $\mathbf{y}^\xi$ even for noisy or incomplete versions $\tilde{\mathbf{x}}^\xi$ of $\mathbf{x}^\xi$.

An associative memory can be classified as *autoassociative* or *heteroassociative* [3]. We have an *autoassociative memory* if $\mathbf{y}^\xi = \mathbf{x}^\xi$ for all $\xi = 1, \ldots, k$. We speak of an heteroassociative memory otherwise. This paper focus on the autoassociative case. The famous *Hopfield network* is an example of autoassociative memory for binary patterns [2, 4, 20].

We speak of a *fuzzy associative memory* (FAM) if the mapping G is given by a fuzzy neural network and the patterns $\mathbf{x}^\xi$ and $\mathbf{y}^\xi$ represent finite fuzzy sets for every $\xi = 1, \ldots, k$ [10, 11]. Applications of FAMs include backing up a truck and trailer, target tracking, and forecasting the average monthly streamflow of a large hydroelectric plant [5, 8]. This paper focus only on the class of *autoassociative fuzzy implicative memories* (AFIMs). The reader interested in a comprehensive survey on FAM models is invited to read [8].

ISBN: 978-989-95079-6-8

2.1 Autoassociative Fuzzy Implicative Memories

AFIMs are single-layer feedforward fuzzy neural networks equipped with neurons that compute the maximum of continuous triangular norms (t-norms) [13, 21]. Specifically, given a synaptic weight matrix $W \in [0,1]^{n\times n}$ and a threshold vector $\boldsymbol{\theta} \in [0,1]^n$, the output $\mathbf{y} \in [0,1]^m$ of an AFIM is given by

$$\mathbf{y} = \mathcal{W}(\mathbf{x}) = (W \circ \mathbf{x}) \vee \boldsymbol{\theta}\,, \tag{1}$$

where $\mathbf{x} \in [0,1]^n$ is the input pattern and the symbol "$\circ$" denotes a max-T product[1]. Recall that the max-T product of two matrices $A \in [0,1]^{n\times k}$ and $B \in [0,1]^{k\times n}$, denoted by $C = A \circ B \in [0,1]^{n\times m}$, is defined as follows [10]:

$$c_{ij} = \bigvee_{\xi=1}^{k} T(a_{i\xi}, b_{\xi j})\,. \tag{2}$$

The synaptic weight matrix W and the threshold vector $\boldsymbol{\theta}$ of an AFIM model are computed by means of a recording recipe called *implicative fuzzy learning* (IFL). Formally, given a fundamental memory set $\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}$, where each $\mathbf{x}^\xi \in [0,1]^n$, IFL defines $W \in [0,1]^{n\times n}$ and $\boldsymbol{\theta} \in [0,1]^n$ as follows [13, 21, 11]:

$$[W \,|\, \boldsymbol{\theta}] = \bigwedge \left\{ [A \,|\, \boldsymbol{\beta}] : (A \circ \mathbf{x}^\xi) \vee \boldsymbol{\beta} \leq \mathbf{x}^\xi,\ \forall\, \xi \in \mathcal{K} \right\}. \tag{3}$$

Here, $[W \,|\, \boldsymbol{\theta}]$ represents the $n \times (n+1)$ fuzzy matrix obtained by concatenating W and $\boldsymbol{\theta}$, and $\mathcal{K} = \{1, \ldots, k\}$.

It is important to note that IFL makes optimal use of the synaptic weights and thresholds of the AM model given by (1). In fact, if there exist $A \in [0,1]^{m\times n}$ and $\boldsymbol{\beta} \in [0,1]^n$ such that $(A \circ \mathbf{x}^\xi) \vee \boldsymbol{\beta} = \mathbf{x}^\xi$ for all $\xi \in \mathcal{K}$, then W and $\boldsymbol{\theta}$ given by (3) also satisfies $(W \circ \mathbf{x}^\xi) \vee \boldsymbol{\theta} = \mathbf{x}^\xi$ for all $\xi \in \mathcal{K}$. Furthermore, the inequalities $A \leq W$ and $\boldsymbol{\beta} \leq \boldsymbol{\theta}$ hold true.

Observe, in particular, that the identity matrix $I \in [0,1]^{n\times n}$ and the vector of zeros $\mathbf{0} = [0, \ldots, 0]^T \in [0,1]^n$ are such that $(I \circ \mathbf{x}^\xi) \vee \mathbf{0} = \mathbf{x}^\xi$ for every original pattern $\mathbf{x}^\xi$. Thus, the synaptic weight matrix W and the threshold vector $\boldsymbol{\theta}$ given by (3) satisfy $I \leq W$ and $(W \circ \mathbf{x}^\xi) \vee \boldsymbol{\theta} = \mathbf{x}^\xi$ for all $\xi \in \mathcal{K}$. As a consequence, one can store as many patterns as desired in an AFIM model [13].

The following proposition establishes a relationship between the output of an AFIM and the fixed points of the synaptic weight matrix W [13, 8, 11]. Recall that $\mathbf{z} \in [0,1]^n$ is a fixed point of $W \in [0,1]^{n\times n}$ if and only if $W \circ \mathbf{z} = \mathbf{z}$. We denote the set of all fixed points of W by $F(W)$, i.e., $F(W) = \{\mathbf{z} \in [0,1]^n : W \circ \mathbf{z} = \mathbf{z}\}$.

Proposition 1. *Given an input pattern $\mathbf{x} \in [0,1]^n$, the output $\mathcal{W}(\mathbf{x}) = (W \circ \mathbf{x}) \vee \boldsymbol{\theta}$ of an AFIM is the smallest fixed point $\mathbf{z}$ of W such that $\mathbf{z} \geq \mathbf{x}$ and $\mathbf{z} \geq \boldsymbol{\theta}$, i.e.,*

$$\mathcal{W}(\mathbf{x}) = \bigwedge \{\mathbf{z} \in F(W) : \mathbf{z} \geq (\mathbf{x} \vee \boldsymbol{\theta})\}\,. \tag{4}$$

Proposition 1 shows that AFIMs exhibit one-step convergence. Moreover, it gives useful insights on the noise tolerance of AFIMs. For example, Proposition 1 says that an AFIM recalls an original pattern $\mathbf{x}^\xi$ only if the input $\mathbf{x}$ is smaller than $\mathbf{x}^\xi$. In other words, AFIMs are suited for the reconstruction of patterns corrupted by erosive noise, but are incapable of handling dilative noise. Recall that a distorted version $\tilde{\mathbf{x}}^\xi$ of the original pattern $\mathbf{x}^\xi$ has undergone an *erosive change* if $\tilde{\mathbf{x}}^\xi \leq \mathbf{x}^\xi$ and a *dilative change* if $\tilde{\mathbf{x}}^\xi \geq \mathbf{x}^\xi$ [19]. Further information on the noise tolerance of AFIMs can be obtained by investigating $F(W)$. A complete characterization of the set of fixed points of a general class of FAMs can be found in [22].

Let us conclude this section by recalling that the synaptic weight matrix $W = (w_{ij}) \in [0,1]^{n\times n}$ and the threshold vector $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]^T \in [0,1]^n$ given by IFL can be easily computed by means of the following equations for every $i, j = 1, \ldots, n$:

$$w_{ij} = \bigwedge_{\xi=1}^{p} I_T(x_j^\xi, x_i^\xi) \quad \text{and} \quad \theta_i = \bigwedge_{\xi=1}^{p} x_i^\xi\,. \tag{5}$$

Here, the operator $I_T : [0,1] \times [0,1] \to [0,1]$ denotes the R-implication associated with the t-norm that is used in the recall phase [23, 10]. Recall that the R-implication associated with a continuous t-norm T is defined as follows:

$$I_T(x,y) = \bigvee \{z \in [0,1] : T(x,z) \leq y\}\,. \tag{6}$$

The following equations present the R-implications associated with the *minimum*, *product*, and *Lukasiewicz t-norm*, respectively [23, 10, 13]:

$$I_M(x,y) = \begin{cases} 1, & x \leq y \\ y, & x > y \end{cases} \qquad \text{(Gödel)} \tag{7}$$

$$I_P(x,y) = \begin{cases} 1, & x \leq y \\ y/x, & x > y \end{cases} \qquad \text{(Goguen)} \tag{8}$$

$$I_L(x,y) = 1 \wedge (1 - x + y) \qquad \text{(Lukasiewicz)} \tag{9}$$

3 Autoassociative Fuzzy Implicative Memories for Large Gray-Scale Images

First of all, recall that a gray-scale image $\mathbf{a}$ of size $M \times N$ can be identified with a finite fuzzy set $\mathbf{x} = [x_1, \ldots, x_n]^T$ of length $n = NM$. The fuzzy set $\mathbf{x}$ is obtained by confining the gray-scale values to the unit interval and by arranging the pixels in a column vector using the standard column-scan method. As a consequence, AFIMs can be used for the storage and retrieval of gray-scale images.

Note, however, that an AFIM needs a synaptic weight matrix $W \in [0,1]^{n\times n}$ and a threshold vector $\boldsymbol{\theta} \in [0,1]^n$. In other words, the model requires computation and storage of $n(n+1)$ values. The following example reveals that the required computational resources restrain the applications of AFIMs to small gray-scale images.

Example 1. Consider a gray-scale image of size 512×512 with 256 shades of gray. This image corresponds to a finite fuzzy set that can be represented by a vector of length $n = 512^2 = 262144$. As a consequence, an AFIM requires the computation and storage of approximately 6.9×10^9 values corresponding to the synaptic weights and thresholds. If these values are represented in a computer using 64-bit double precision floating points, then the model allocates approximately 512 gigabytes of memory space. Similar considerations reveal that AFIMs allocate approximately 2 and 32 gigabytes of memory spaces for the storage and recall of images of size 128×128 and 256×256, respectively.

[1] In this paper, the symbols "$\bigvee$" and "$\bigwedge$" denote the supremum (or maximum) and infimum (or minimum) operations, respectively.

ISBN: 978-989-95079-6-8

Concluding, due to computer memory limitations, applications of AFIMs are usually bounded to images of size less than or approximately equal to 128×128. The following subsections provide strategies that allow for the storage and recall of large gray-scale images in AFIM based models.

3.1 Sparsely Connected AFIMs

Several studies suggest that neurons in the parts of the human brain that exhibit functional properties of associative memory are connected to few other neurons [24, 25, 26]. Motivated by this biological remark, we will remove a considerable amount of synaptic junctions of an AFIM model, i.e., we will introduce zeros in the synaptic weight matrix given by (5). However, in view of the following theorem[2], we will not delete synaptic weights in the main diagonal of W.

Theorem 1. *Consider an arbitrary fundamental memory set $\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}$, where each $\mathbf{x}^\xi \in [0,1]^n$. If $V \in [0,1]^{n\times n}$ and $\boldsymbol{\vartheta} \in [0,1]^n$ satisfy the inequalities $I \leq V \leq W$ and $\boldsymbol{\vartheta} \leq \boldsymbol{\theta}$, where $W \in [0,1]^{n\times n}$ and $\boldsymbol{\theta} \in [0,1]^n$ are given by IFL, then the following equation holds true for every $\xi = 1, \ldots, k$:*

$$\mathbf{x}^\xi = (V \circ \mathbf{x}^\xi) \vee \boldsymbol{\vartheta}\,. \tag{10}$$

In view of Theorem 1, we define a *sparsely connected AFIM* (SCAFIM) as an autoassociative max-T FAM with threshold $\boldsymbol{\vartheta} \leq \boldsymbol{\theta}$ and a synaptic weight matrix $I \leq V \leq W$ that has very few nonzero elements. Since V has a sparse structure, SCAFIMs usually do not require large computational resources. Thus, they are suited for the storage and recall of large gray-scale images.

By definition, SCAFIMs exhibit optimal absolute storage capacity such as the fully connected AFIMs. However, in contrast to the latter, SCAFIMs do not necessarily exhibit one-step convergence. Therefore, in analogy to the famous *Hopfield network*, we may employ SCAFIMs with feed-back [2, 20]. In this case, given an input pattern $\mathbf{x} \in [0,1]^n$, we define the following sequence where $\mathbf{x}(0) = \mathbf{x}$:

$$\mathcal{V}_t(\mathbf{x}) = \mathbf{x}(t) = \big[V \circ \mathbf{x}(t-1)\big] \vee \boldsymbol{\vartheta}\,, \quad \forall\, t = 1, 2, \ldots \tag{11}$$

The resulting model is referred to as *dynamic or recursive SCAFIM*. Note that $\mathcal{V}_t(\mathbf{x})$ represents the output of a dynamic SCAFIM after t steps. In particular, $\mathcal{V}_1(\mathbf{x})$ corresponds to the output of a single-step model.

The following theorem shows that the sequence $\mathcal{V}_t(\mathbf{x})$ given by (11) converges for every input pattern $\mathbf{x} \in [0,1]^n$. Furthermore, Theorem 2 below relates the limit of the sequence $\mathcal{V}_t(\mathbf{x})$, denoted by $\mathcal{V}_*(\mathbf{x})$, with the fixed points of the synaptic weight matrix V. Recall that the set of fixed points of V is defined as $F(V) = \{\mathbf{z} \in [0,1]^n : V \circ \mathbf{z} = \mathbf{z}\}$.

Theorem 2. *Given a fundamental memory set $\{\mathbf{x}^1, \ldots, \mathbf{x}^k\}$, define $\boldsymbol{\vartheta}$ and V such that $\boldsymbol{\vartheta} \leq \boldsymbol{\theta}$ and $I \leq V \leq W$, where W and $\boldsymbol{\theta}$ are given by IFL. Then, for every input pattern $\mathbf{x} \in [0,1]^n$, the sequence given by* (11) *is monotonically increasing and converges to the smallest fixed point $\mathbf{z}$ of V such that $\mathbf{z} \geq \mathbf{x}$ and $\mathbf{z} \geq \boldsymbol{\vartheta}$, i.e.,*

$$\mathcal{V}_*(\mathbf{x}) = \lim_{t\to\infty} \mathbf{x}(t) = \bigwedge \{\mathbf{z} \in F(V) : \mathbf{z} \geq (\mathbf{x} \vee \boldsymbol{\vartheta})\}. \tag{12}$$

[2]We would like to point out that we intend to publish the proofs of Theorems 1, 2, and 3 in an upcoming journal paper.

Note that Theorem 2 is analogous to Proposition 1. Thus, such as AFIMs, SCAFIMs cannot recall an original pattern $\mathbf{x}^\xi$ if the input $\mathbf{x}$ is greater than $\mathbf{x}^\xi$. More importantly, both Proposition 1 and Theorem 2 relate noise tolerance with the set of fixed points of the synaptic weight matrix. The following theorem establishes a relationship between $F(V)$ and $F(W)$. As a consequence, we are able to relate the noise tolerance of a certain SCAFIM and its corresponding fully connected AFIM. Theorem 3 below also relates the set of fixed points of V with the set of fixed points of a matrix U that has a smaller number of non-zeros entries than V.

Theorem 3. *Let $F(U), F(V)$ and $F(W)$ denote the set of fixed points of matrices $U, V, W \in [0,1]^{n\times n}$ with respect to a certain max-T product. If $I \leq U \leq V \leq W$, then $F(W) \subseteq F(V) \subseteq F(U)$.*

Remark 1. Theorems 2 and 3 tell us that the noise tolerance of dynamic SCAFIMs with respect to erosive noise diminishes - because the number of fixed points increases - as we remove more and more synaptic junctions of a fully connected AFIM. In contrast, a dynamic SCAFIMs exhibit better noise tolerance with respect to dilative noise than its corresponding AFIM. Let us clarify this remark by means of an illustrative example.

Example 2. Let W and $\boldsymbol{\theta}$ denote the synaptic weight matrix and threshold vector of an AFIM. Suppose that U and V are sparsely connected synaptic weight matrices that satisfy the inequalities $I \leq U \leq V \leq W$, i.e., U has fewer non-zeros elements than V. Figure 1 depicts the sets of fixed points $F(W)$, $F(V)$, and $F(U)$. Here, we represented the set $[0,1]^n$ by a line from $\mathbf{0} = [0, 0, \ldots, 0]^T$ to $\mathbf{1} = [1, 1, \ldots, 1]^T$ although this set is not *totally ordered*[3]. The fixed points of W, V, and U are represented by short vertical lines. Note that, in agreement with Theorem 3, we have depicted $F(W) \subseteq F(V) \subseteq F(U)$.

Suppose that $\mathbf{x}^\xi$ represents an original pattern and $\check{\mathbf{x}}$ corresponds to an eroded version of $\mathbf{x}^\xi$. Let $\mathcal{W}$ denote the AFIM model whose synaptic weight matrix is W and threshold vector $\boldsymbol{\theta}$ is shown in Figure 1. By Proposition 1, $\mathcal{W}(\check{\mathbf{x}}) = \mathbf{x}^\xi$ because $\mathbf{x}^\xi$ is the smallest fixed point of W greather than $\check{\mathbf{x}}$ and $\boldsymbol{\theta}$. Similarly, let $\mathcal{V}_*$ and $\mathcal{U}_*$ denote dynamic SCAFIMs with threshold vector $\boldsymbol{\vartheta} = \boldsymbol{\theta}$ and synaptic weight matrices V and U, respectively. In contrast to the AFIM model, the dynamic SCAFIMs $\mathcal{V}_*$ and $\mathcal{U}_*$ fail to recall the original pattern $\mathbf{x}^\xi$ because there are fixed points in $F(V)$ and $F(U)$ between $\check{\mathbf{x}} \vee \boldsymbol{\vartheta}$ and $\mathbf{x}^\xi$. Furthermore, observe that $\mathcal{V}_*$ outperforms the SCAFIM $\mathcal{U}_*$ because U has a fixed point between $\check{\mathbf{x}} \vee \boldsymbol{\vartheta}$ and $\mathcal{V}_*(\check{\mathbf{x}})$.

Conversely, let us suppose that $\hat{\mathbf{x}}$ represents a corrupted version of $\mathbf{x}^\xi$ contaminated by a small dilative noise. In view of the relationship $F(W) \subseteq F(V) \subseteq F(U)$, the dynamic SCAFIM $\mathcal{U}_*$ and the AFIM $\mathcal{W}$ yield the best and the worst results, respectively.

3.2 Strategies for the Construction of SCAFIMs

This subsection briefly answers the question as to derive an SCAFIM model from an AFIM. First, note that the threshold

[3]Recall that $(X, \leq)$ is *totally ordered*, also called *linearly ordered*, if for any $x, y \in X$, either $x \leq y$ or $y \leq x$. Note, however, that there are patterns $\mathbf{x}, \mathbf{y} \in [0,1]^n$ such that neither inequalities $\mathbf{x} \leq \mathbf{y}$ nor $\mathbf{y} \leq \mathbf{x}$ hold true. Thus, Figure 1 is merely illustrative.

ISBN: 978-989-95079-6-8

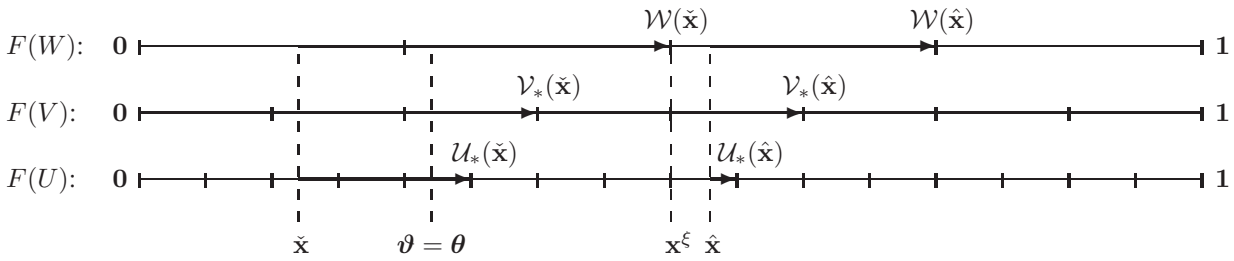


Figure 1: Illustrative example of the noise tolerance of an AFIM $\mathcal{W}$ and two dynamic SCAFIMs $\mathcal{V}_*$ and $\mathcal{U}_*$. Here, the synaptic weight matrices satisfy the inequalities $I \leq U \leq V \leq W$.

vector of an SCAFIM can be defined either as $\boldsymbol{\vartheta} = \boldsymbol{\theta}$ or $\boldsymbol{\vartheta} = \mathbf{0}$, depending on computational resources available. Let us consider $\boldsymbol{\vartheta} = \boldsymbol{\theta}$ in this paper. As far as we are concerned, the synaptic junctions of an AFIM can be removed in a strength-based manner or in a structured-based manner.

In the strength-based strategy, we delete synaptic weights by taking into account their strength. Specifically, we define the sparsely connected synaptic weight matrix $V \in [0,1]^{n \times n}$ as follows for some $\alpha \in (0,1]$:

$$v_{ij} = \begin{cases} w_{ij} & \text{if } w_{ij} \geq \alpha \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Note that the synaptic junctions with strength less than α are removed from the fully connected AFIM.

Alternatively, we say that we removed synaptic junctions in a structurated-based manner if we fix the non-zeros entries of the synaptic weight matrix before its computation. For example, we can set that the synaptic weight matrix V has a band or a block diagonal structure. We would like to point out that this strategy generalizes the widely used procedure of partitioning large gray-scale images in order to obtain several small-sized AM models [15, 16, 17].

4 Computational Experiments

This section provides some computational experiments concerning the reconstruction of noisy or incomplete gray-scale images. In order to compare the performance of SCAFIMs with the fully connected AFIM model, and due to page constrain, we only consider gray-scale images of small size. Specifically, let us consider the images of size 64×64 with 256 shades of gray displayed in the first row of Figure 2. For each of these image, we generated a vector $\mathbf{x}^\xi$ of length $n = 64 \times 64 = 4096$ using the standard column-scan method. Furthermore, we divided all entries of $\mathbf{x}^\xi$ by 255 in order to obtain patterns in the hypercube $[0,1]^{4096}$.

We stored the eight original patterns $\mathbf{x}^\xi$ in the Lukasiewicz AFIM, i.e., the AFIM based on the Lukasiewicz t-norm. We would like to recall that the Lukasiewicz AFIM outperformed several AM models, including the FAMs of Kosko and the GFAM based on the Lukasiewicz t-norm, in an experiment using faces images from the database of AT&T Laboratories, Cambridge [8, 13].

Next, we constructed six SCAFIM models, denoted by $\mathcal{V}^a, \mathcal{V}^b, \ldots, \mathcal{V}^f$, all of them with threshold vector $\boldsymbol{\vartheta} = \boldsymbol{\theta}$. The corresponding sparsely connected synaptic weight matrices $V^a, \ldots, V^f$ were derived as follows from W given by IFL: V^a has 16 blocks of size 256×256, V^b has 64 blocks of size 64×64, and V^c is a band matrix with bandwidth equals to 19. The synaptic weight matrices V^d, V^e, and V^f were obtained by means of (13) with $\alpha = 0.893, 0.961$, and 1, respectively. We would like to point out that we have chosen α and the bandwith of V^c in order to obtain pairs V^a and V^d, V^b and V^e, V^c and V^f with approximately the same number of non-zeros entries. Furthermore, the output of the SCAFIM $\mathcal{V}^b$ corresponds to the gray-scale image that is obtained by adopting the following procedure: Store each column of an original image in a separate Lukasiewicz AFIM. This results in 64 AFIMs, each of which has a 64×64 synaptic weight matrix that corresponds to a block of V^b. The recalled image is obtained by feeding each AFIM with its corresponding column of a given image and by concatenating the outputs.

The first row of Table 1 provides the density (i.e., percentage of non-zeros entries) of the synaptic weight matrices W, $V^a, \ldots, V^f$. Table 1 also provides the memory space required by the seven AM models if double precision numbers are used to represent the synaptic weights and threshold values. We would like to point out that we used the standart row-column-value representation to store the sparsely connected synaptic weight matrices in the computer.

Afterward, we confirmed that the eight original patterns represent fixed points of $W, V^a, \ldots, V^f$, i.e., the Lukasiewicz AFIM as well as the six SCAFIMs exhibit optimal absolute storage capacity.

In order to compare the noise tolerance of the seven AM models, we first introduced as input an image corrupted by *pepper noise* with probability 0.3. The second row of Figure 2 shows the corrupted image and the corresponding recalled patterns. Table 1 contains the *base-10 logarithm* applied to the *arithmetic mean* of the *relative normalized mean squared error* (RNMSE) in 100 experiments. Recall that the RNMSE is given by the following equation where x_i, $\tilde{x}_i$, and y_i denote the i-th entry of the original, noisy, and recalled patterns, respectively:

$$RNMSE = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \tilde{x}_i)^2}\,. \tag{14}$$

We would like to point out that we have corrupted an original pattern that was randomly selected in each trial. Moreover, the six dynamic SCAFIMs were iterated until convergence. The arithmetic means of the number of iterations required for convergence are shown between parentheses in Table 1.

We also conducted similar experiments using *salt noise* with probability 0.01, *salt and pepper noise* with probability 0.05, *additive Gaussian noise* with mean 0 and variance 0.01,

ISBN: 978-989-95079-6-8

	$\mathcal{W}$	$\mathcal{V}^a$	$\mathcal{V}^b$	$\mathcal{V}^c$	$\mathcal{V}^d$	$\mathcal{V}^e$	$\mathcal{V}^f$
Density of the AM model	100%	6.25%	1.56%	0.46%	6.3%	1.42%	0.48%
Required memory space	134.2 MB	16.8 MB	4.2 MB	1.3 MB	17.0 MB	4.2 MB	1.3 MB
Tolerance w.r.t. pepper noise	-3.05 (1)	-2.47 (1)	-1.93 (1)	-1.78 (2.3)	-2.37 (2.6)	-1.27 (2.7)	-0.62 (1)
Tolerance w.r.t. salt noise	1.85 (1)	1.55 (1)	1.13 (1)	0.97 (5.3)	1.80 (8.9)	1.51 (9.7)	0.96 (1)
Tol. w.r.t. salt & pepper noise	13.84 (1)	9.53 (1)	5.00 (1)	3.76 (3.7)	13.08 (7.6)	8.01 (6.7)	3.17 (1)
Tol. w.r.t. add. gaussian noise	6.40 (1)	3.50 (1)	2.18 (1)	1.64 (4.1)	4.97 (7.5)	2.78 (7.3)	1.35 (1)
Tolerance w.r.t. speckle noise	7.64 (1)	4.27 (1)	2.67 (1)	1.80 (4.8)	5.47 (8.7)	2.17 (8.1)	1.09 (1)

Table 1: Comparison of the noise tolerance, density of synaptic junctions, and required computational resources of Lukasiewicz AFIM and six dynamic SCAFIMs.

and *multiplicative gaussian noise (speckle)* with mean 1 and variance 0.04. Table 1 provides the base-10 logarithm of the mean of the RNMSE in 100 experiments. Again, the mean of the number of iterations required for convergence is shown between parentheses. Figure 2 shows instances of the noisy images and corresponding recalled patterns.

Note that, in accordance with Remark 1, the noise tolerance with respect to erosive (pepper) noise diminishes - whereas the tolerance with respect to dilative (salt) and mixed (gaussian and speckle) noise increase - as we remove more and more synaptic junctions from a fully connected AFIM.

The following compares the two strategies for the construction of dynamic SCAFIMs. Let us begin by considering $\mathcal{V}^a$ and $\mathcal{V}^d$. Both SCAFIMs have approximately the same number of synaptic junctions. Thus, they required almost the same amount of memory space. However, the SCAFIM $\mathcal{V}^a$ outperformed $\mathcal{V}^d$ with respect to erosive, dilative, and mixed noise. Furthermore, $\mathcal{V}^a$ is preferable than $\mathcal{V}^d$ in computational terms. First, because V^a converged to the fixed point with only one iteration whereas V^b required in mean 7 iterations. Secondly, in the structurated-based strategy, we only compute the synaptic junctions that are really stored in the computer. In contrast, in the strength-based strategy, we compute all synaptic junctions but we keep only those with strength greater than or equal to α. In addition, we can determine exactly the amount of bytes that the SCAFIM $\mathcal{V}^a$ will require before the computation of V^a, but we can only estimate the memory space that V^d will consume. We reach a similar conclusion by comparing the SCAFIMs $\mathcal{V}^b$ and $\mathcal{V}^e$.

Finally, let us compare $\mathcal{V}^c$ and $\mathcal{V}^f$. Note that both models consumed almost the same amount of computational resources. Moreover, they exhibited similar tolerance with respect to dilative or mixed noise. However, the SCAFIM $\mathcal{V}^c$ outperformed $\mathcal{V}^f$ with respect to erosive noise, but the former usually required more than one iteration to converge[4].

5 Concluding Remarks

This paper introduces the class of SCAFIMs. In few words, a SCAFIM is obtained by removing a considerable amount of synaptic junctions of an AFIM model. The novel models inherit the optimal storage capacity of AFIMs but do not necessarily exhibit one step convergence. Nevertheless, we showed that dynamic SCAFIMs, i.e., SCAFIMs with feedback, yield a monotonically increasing sequence of patterns that converges to a fixed point of the underlying (sparse) synaptic weight matrix. Thus, such as the AFIMs, SCAFIMs exhibit tolerance with respect to incomplete or eroded patterns. Furthermore, we pointed out that the noise tolerance of SCAFIMs decreases as the density of synaptic junctions. In contrast, the tolerance with respect to dilative and mixed noise increases as the density of synaptic junctions decreases.

In addition, we briefly investigated two strategies for the construction of SCAFIMs: the structurated-based and the strength-based strategies. In computational terms, the former is usually preferable than the latter due to the following reasons. First, because we compute only the synaptic junctions that will be really stored in the computer. Secondly, because we can determine exactly the amount of bytes that the sparse matrix will require. Finally, computational experiments revealed that the structurated-based strategy usually yields better noise tolerance than the strength-based strategy. However, further research is needed on how to choose the best strategy for the construction of SCAFIMs.

[4] We would like to point out that the SCAFIM $\mathcal{V}^f$ corresponds to the *Gaines fuzzy morphological associative memory* (FMAM) that was introduced and investigated in [22]. This model always converges to a fixed point with only one step.

References

[1] J. Anderson. *An Introduction to Neural Networks*. MIT Press, MA, 1995.

[2] M. Hassoun, editor. *Associative Neural Memories: Theory and Implementation*. Oxford University Press, Oxford, U.K., 1993.

[3] T. Kohonen. *Self-Organization and Associative Memory*. Springer Verlag, 1984.

[4] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, NJ, 1999.

[5] B. Kosko. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, Englewood Cliffs, NJ, 1992.

[6] B. Zhang, H. Zhang, and S. Ge. Face recognition by applying wavelet subband representation and kernel associative memory. *IEEE Transactions on Neural Networks*, 15(1):166–177, January 2004.

[7] D. Zhang and W. Zuo. Computational intelligence-based biometric technologies. *IEEE Computational Intelligence Magazine*, 2(2):26–36, May 2007.

[8] P. Sussner and M. Valle. Fuzzy associative memories and their relationship to mathematical morphology. In W. Pedrycz, A. Skowron, and V. Kreinovich, editors, *Handbook of Granular Computing*, chapter 33, pages 733–754. John Wiley and Sons, Inc., New York, 2008.

[9] S. Teddy, C. Quek, and E. Lai. PSECMAC: A novel self-organizing multiresolution associative memory architecture.

ISBN: 978-989-95079-6-8

Figure 2: First row displays the original gray-scale images of size 64×64 with 256 shades of gray. The first column of the following three rows shows patterns corrupted by *pepper noise* with probability 0.3, *salt noise* with probability 0.01, and *salt and pepper noise* with probability 0.05, respectively. The other seven columns (from left to right) exhibit the corresponding patterns that were recalled by the Lukasiewicz AFIM $\mathcal{W}$ and the SCAFIMs $\mathcal{V}^a$, $\mathcal{V}^b$, $\mathcal{V}^c$, $\mathcal{V}^d$, $\mathcal{V}^e$, and $\mathcal{V}^f$, respectively.

IEEE Transactions on Neural Networks, 19(4):689–712, April 2008.

[10] W. Pedrycz and F. Gomide. *Fuzzy Systems Engineering: Toward Human-Centric Computing*. Wiley-IEEE Press, New York, 2007.

[11] M. Valle and P. Sussner. A general framework for fuzzy morphological associative memories. *Fuzzy Sets and Systems*, 159(7):747–768, 2008.

[12] F. Chung and T. Lee. On fuzzy associative memory with multiple-rule storage capacity. *IEEE Transactions on Fuzzy Systems*, 4(3):375–384, 1996.

[13] P. Sussner and M. Valle. Implicative fuzzy associative memories. *IEEE Transactions on Fuzzy Systems*, 14(6):793–807, 2006.

[14] P. Sussner and M. Valle. Grayscale morphological associative memories. *IEEE Transactions on Neural Networks*, 17(3):559–570, May 2006.

[15] D. Lee. Improvements of complex-valued hopfield associative memory by using generalized projection rules. *IEEE Transactions on Neural Networks*, 17(5):1341–1347, September 2006.

[16] M. Müezzinoğlu, C. Güzeliş, and J. Zurada. A new design method for the complex-valued multistate Hopfield associative memory. *IEEE Transactions on Neural Networks*, 14(4):891–899, July 2003.

[17] G. Costantini, D. Casali, and R. Perfetti. Neural associative memory storing gray-coded gray-scale images. *IEEE Transactions on Neural Networks*, 14(3):703–707, May 2003.

[18] G. Costantini, D. Casali, and R. Perfetti. Associative memory design for 256 gray-level images using a multilayer neural network. *IEEE Transactions on Neural Networks*, 17(2):519–522, march 2006.

[19] G. Ritter, P. Sussner, and J. Diaz de Leon. Morphological associative memories. *IEEE Transactions on Neural Networks*, 9(2):281–293, 1998.

[20] J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79:2554–2558, April 1982.

[21] M. Valle, P., and F. Gomide. Introduction to implicative fuzzy associative memories. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, pages 925 – 931, Budapest, Hungary, 2004.

[22] M. Valle and P. Sussner. Storage and recall capabilities of fuzzy morphological associative memories with adjunction-based learning. 2007. Submitted for publication.

[23] H. Nguyen and E. Walker. *A First Course in Fuzzy Logic*. Chapman & Hall/CRC, Boca Raton, 2 edition, 2000.

[24] H. Bosch and F. Kurfess. Information storage capacity of incompletely connected associative memories. *Neural Networks*, 11(5):869–876, 1998.

[25] B. Olshausen and D. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–487, August 2004.

[26] R. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried. Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107, June 2005.

ISBN: 978-989-95079-6-8

A Restriction Level Approach for the Representation and Evaluation of Fuzzy Association Rules

Miguel Delgado[1] M. Dolores Ruiz[1] Daniel Sánchez[1]

1.Dept. Computer Science and Artificial Intelligence, University of Granada
ETSIIT, C/Periodista Daniel Saucedo Aranda s/n, 18071, Granada
Email: mdelgado@ugr.es, {mdruiz, daniel}@decsai.ugr.es

Abstract— *The goal of this paper is to present a logic model with very good properties for the representation and evaluation of fuzzy association rules. For that we are going to combine two approaches. The first one is a model for the representation and evaluation of crisp association rules. The second one is a new proposal for the representation of imprecise properties (in particular for fuzzy sets) by using restriction levels, which verifies all the crisp logic equivalences [1]. Combining both approaches we will achieve a solid model, with very good properties for the representation of fuzzy rules and a very simple framework for extending the crisp quality measures for the evaluation of fuzzy rules.*

Keywords— Logic model, restriction level, RL-probability, RL-numbers, fuzzy association rules.

1 Introduction

Fuzzy association rules have been very developed and have been applied in numerous situations since their appearance in the last nineties. The first works used the Fuzzy Subsets Theory proposed by L.A. Zadeh [2] for putting into groups imprecise values with a clear semantic and later they were used for representing the different types of imprecision found in a stored database. The advantages of using fuzzy sets are on the one hand making smother the bounds and on the other hand give a formal representation for the knowledge semantically significative and meaningful for the user. In [3], [4] we can find a description of some of the most important works in the field of extracting fuzzy association rules.

There are some approaches which offer a framework for generalizing the quality measures for fuzzy association rules (see [5] and references in it). Moreover, although the evaluation measures can be generalized by distinct ways, it should be necessary to search a logic support or an axiomatic base which reflects the choice of a specific measurement. This work tries to follow this line of research and it also tries to establish a logic model that generalizes and preserves in a natural way all the properties of crisp association rules, as well as the different interest measures used in their evaluation. For that we make use of two approaches: the logic model proposed in [6],[7] and the RL-representation theory proposed in [1], [8], [9].

The logic model we are dealing with can manage with different kinds of association rules which might be useful for the user. There are several examples in [10] where the model offers a good formalization for mining exception rules and double rules. Furthermore, we have implemented an algorithm for extracting any kind of rule having the quality measures used for assessing the validity of the rule represented by this model. The algorithm uses the items representation by means of bitsets [11] which reduces de time consuming when computing the conjunction between items and their associated support (bitset cardinality).

The rest of the paper follows with the approach for representing imprecise properties by means of restriction levels presented in [1]. Then we do a brief summary of a model proposed in [7] for the representation of crisp association rules which was widely developed in [6]. The last part of the paper contains our proposal for the representation and the evaluation of fuzzy association rules using a combination of the approaches presented in the previous sections. We finish with the conclusions and possible lines for future research.

2 Representation by Restriction Levels

An imprecise property in an universe X can be represented by a collection of crisp realizations. The approach of [1] for representing imprecise properties by means of restriction levels extends the usual operations from the crisp to the fuzzy case satisfying the logic equivalences between that operations.

Definition 1. [1] A *RL-set* Λ is a finite set of restriction levels $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying that $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$, $m \geq 1$.

In general, the *RL*-set of an atomic property represented by means of a fuzzy set A is defined as follows.

Definition 2. [1] Let be A a fuzzy set defined on the referencial X. Then the *RL*-set associated to A is given by:

$$\Lambda_A = \{A(x) \,|\, x \in X\} \cup \{1\} \tag{1}$$

where $A(x)$ is the grade of membership of x to the fuzzy set A.

The *RL*-set employed for representing an imprecise property is obtained by the union of the *RL*-sets associated to the atomic properties which define that property.

For representing an imprecise property in X by means of restriction levels we are going to use a *RL-representation* defined like a pair (Λ, ρ) where Λ is a *RL*-set and $\rho : \Lambda \to \mathcal{P}(X)$ is a function which applies each restriction level into a crisp realization in this level. For example, the *RL*-representation of an imprecise atomic property defined by a fuzzy set A will be the pair (Λ_A, ρ_A), where Λ_A is given by the equation (1) and $\rho_A(\alpha) = A_\alpha = \{x \in X \,|\, A(x) \geq \alpha\}$ for all $\alpha \in \Lambda_A$.

ISBN: 978-989-95079-6-8

Given an imprecise property P represented by (Λ_P, ρ_P), the set of crisp representatives of P, Ω_P is defined as

$$\Omega_P = \{\rho_P(\alpha) \mid \alpha \in \Lambda_P\}. \quad (2)$$

Definition 3. [1] Let be (Λ, ρ) a *RL*-representation with $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying that $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$. Let $\alpha \in (0,1]$ and $\alpha_i, \alpha_{i+1} \in \Lambda$ satisfying that $\alpha_i > \alpha > \alpha_{i+1}$. Then we define

$$\rho(\alpha) = \rho(\alpha_i). \quad (3)$$

If we look to this definition, this extension for values that there are not in the *RL*-set of the function ρ, is the natural extension if we think in a fuzzy set A and its α-cuts. Using this definition the concept of equivalence between two *RL*-representations is defined.

Definition 4. [1] Let (Λ, ρ) and (Λ', ρ') be two *RL*-representations on X. We will say that both representations are *equivalent* and will be noted by $(\Lambda, \rho) \equiv (\Lambda', \rho')$, if and only if, $\forall \alpha \in (0,1]$

$$\rho(\alpha) = \rho(\alpha'). \quad (4)$$

2.1 Logic Operations

In this section we present a brief overview of the logic operations necessary for the understanding of the rest of the paper. In particular we present the logic operations of disjunction, conjunction and negation that we will need for the generalization of the logic model to the case of fuzzy association rules. The basic ideas of how they are defined can be found in [1].

Definition 5. Let P, Q be two imprecise properties with *RL*-representations (Λ_P, ρ_P), (Λ_Q, ρ_Q). Then, $P \wedge Q$, $P \vee Q$ and $\neg P$ are imprecise properties represented by $(\Lambda_{P\wedge Q}, \rho_{P\wedge Q})$, $(\Lambda_{P\vee Q}, \rho_{P\vee Q})$ and $(\Lambda_{\neg P}, \rho_{\neg P})$ respectively, where

$$\begin{aligned} \Lambda_{P\wedge Q} &= \Lambda_{P\vee Q} = \Lambda_P \cup \Lambda_Q \\ \Lambda_{\neg P} &= \Lambda_P \end{aligned} \quad (5)$$

and, for all $\alpha \in (0,1]$,

$$\begin{aligned} \rho_{P\wedge Q}(\alpha) &= \rho_P(\alpha) \cap \rho_Q(\alpha), \\ \rho_{P\vee Q}(\alpha) &= \rho_P(\alpha) \cup \rho_Q(\alpha), \\ \rho_{\neg P}(\alpha) &= \overline{\rho_P(\alpha)}, \end{aligned} \quad (6)$$

where $\overline{Y}$ is the usual complement of a crisp set Y.

Proposition 1. [1] The operations $\wedge, \vee, \neg$ between *RL*-representations verify the ordinary properties of logic equivalence as for example $\neg\neg A \equiv A$, the Morgan's laws $(\neg(A \wedge B) \equiv (\neg A \vee \neg B)$, one of them) and the law of excluded middle that can be expressed as $A \wedge \neg A \equiv \bot$ or $A \vee \neg A \equiv \top$, where $\top$ and $\bot$ are the atomic properties which represent the tautology (whose *RL*-representation is obtained from the referencial X) and the contradiction (obtained from $\emptyset$).

2.2 RL-numbers

On the basis of *RL*-representations and operations, we introduced in [8] the *RL*-numbers as a representation of fuzzy quantities. This approach offers two main advantages: (1) *RL*-numbers are representations of fuzzy quantities that can be easily obtained by extending usual crisp measurements to fuzzy sets. (2) Arithmetic and logical operations on *RL*-numbers are straightforward and unique extensions of the operations on crisp numbers, verifying the usual properties of crisp arithmetic and logical operations. In addition, the imprecision does not necessarily increase through operations, and it can even diminish. The following definitions and properties are from [8]:

Definition 6. A *RL*-real number is a pair $(\Lambda, \mathcal{R})$ where Λ is a *RL*-set and $\mathcal{R} : (0,1] \to \mathbb{R}$.

We shall note $\mathbb{R}_{RL}$ the set of *RL*-real numbers. The *RL*-real number R_x is the representation of a (precise) real number x iff $\forall \alpha \in \Lambda_{R_x}$, $\mathcal{R}_{R_x}(\alpha) = x$. We shall denote such *RL*-real number as R_x or, equivalently, x, since in the crisp case, the set Λ_{R_x} is not important. Operations are extended as follows:

Definition 7. Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $R_1, \ldots, R_n$ be *RL*-real numbers. Then $f(R_1, \ldots, R_n)$ is a *RL*-real number with

$$\Lambda_{f(R_1,\ldots,R_n)} = \bigcup_{1 \le i \le n} \Lambda_{R_i} \quad (7)$$

and, $\forall \alpha \in \Lambda_{f(R_1,\ldots,R_n)}$

$$\mathcal{R}_{f(R_1,\ldots,R_n)}(\alpha) = f(\mathcal{R}_{R_1}(\alpha), \ldots, \mathcal{R}_{R_n}(\alpha)) \quad (8)$$

It is obvious that operations defined in this way are consistent extensions of crisp operations. We want to remark that operations not defined for certain combinations of real values are not defined for *RL*-numbers that verify that combination in at least one restriction level. This is the case of division by 0 for example, i.e., R/R' is defined iff $0 \notin \Omega_{R'}$.

2.3 Probabilities based on RL-numbers

Let us note $[0,1]_{RL} \subseteq \mathbb{R}_{RL}$ the set of *RL*-real numbers verifying $0 \le R \le 1$ $\forall R \in [0,1]_{RL}$ (see definition of *RL*-ranking in [9]). We introduce the notions of *RL*-probability space and *RL*-probability as in [9]:

Definition 8. A *RL*-probability space is a triple (X, Σ, P) where

1. X is a crisp set.
2. Σ is a collection of imprecise events defined by *RL*-representations on X, closed under complement and countable unions, and verifying $X \in \Sigma$.
3. $P : \Sigma \to [0,1]_{RL}$ verifying Kolmogorov's Axioms, i.e., $P(X) = 1$, $P(E) \ge 0$ $\forall E \in \Sigma$, and for any finite collection of disjoint representations $E_1, \ldots, E_n$, $P(E_1 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i)$.

In the previous definition, P is a *RL*-probability measure. In particular, *RL*-probability measures can be obtained from ordinary probability measures in an easy way as follows: let (U, F, p) be a (crisp) probability space and let U_F^{RL} be the set of all the *RL*-representations on U such that for any $A \in U_F^{RL}$, $\Omega_A \subseteq F$. By the properties of F, U_F^{RL} is closed under complement and countable unions, and $U \in U_F^{RL}$. For any $A \in U_F^{RL}$, let $P(A) \in [0,1]_{RL}$ defined by $\rho_{P(A)}(\alpha) = p(\rho_A(\alpha))\forall \alpha \in \Lambda_A$. It is easy to show that (U, U_F^{RL}, P) is a *RL*-probability space and P is a *RL*-probability measure.

ISBN: 978-989-95079-6-8

Also easy to show, *RL*-probabilities verify the addition law $P(A\cup B)=P(A)+P(B)-P(A\cap B)$ and the property $P(\neg A)=1-P(A)$. Finally, by definition 8, it is obvious that crisp probabilities are a particular case of *RL*-probabilities since crisp events are particular cases of imprecise events as described by *RL*-representations.

It is easy to check that the usual arithmetic relations between probabilities are preserved in each restriction level and, hence, they are preserved by the arithmetic of *RL*-numbers. For example, $p(A)+p(\neg A)=1$.

The concept of conditional probability is straightforwardly extended as follows:

Definition 9. Let A and B be two imprecise events defined on U by *RL*-representations $A=(\Lambda_A,\rho_A)$ and $B=(\Lambda_B,\rho_B)$ with $\rho_{p(B)}>0\ \forall\alpha\in(0,1]$. The *RL*-probability of A given B in U is

$$p(A|B)=\frac{p(A\wedge B)}{p(B)} \tag{9}$$

Proposition 2. [9] Let A and B be two imprecise events defined on U by *RL*-representations $A=(\Lambda_A,\rho_A)$ and $B=(\Lambda_B,\rho_B)$ with $\rho_{p(B)}>0\ \forall\,\alpha\in(0,1]$. Then

$$\rho_{p(A|B)}(\alpha)=\frac{\rho_{p(A\wedge B)}(\alpha)}{\rho_{p(B)}(\alpha)}. \tag{10}$$

Example 1. Let X and Y be atomic imprecise events defined by the following fuzzy sets on $U=\{u_1,\ldots,u_6\}$:

$$\begin{aligned} X&=1/u_1+0.8/u_2+0.5/u_3+0.4/u_5\\ Y&=0.9/u_1+0.6/u_3+0.5/u_4 \end{aligned} \tag{11}$$

Then we have $\Lambda_X=\Lambda_{\neg X}=\{1,0.8,0.5,0.4\}$ and $\Lambda_Y=\Lambda_{\neg Y}=\{1,0.9,0.6,0.5\}$. The *RL*-set for any operation between X and Y is $\Lambda_X\cup\Lambda_Y$. Table 1 shows the *RL*-representation of X, $\neg X$, Y, $X\wedge Y$, $X\vee Y$ and $X\wedge\neg Y$. Assuming that every u_i is equally probable, table 2 shows the corresponding *RL*-probabilities of imprecise events $p(X)$, $p(\neg X)$, $p(Y)$, $p(X\wedge Y)$, $p(X\vee Y)$ and $p(X\wedge\neg Y)$ and $p(Y|X)$. Remark that the *RL*-probability $p(X|Y)$ is undefined since $\rho_{p(Y)}(1)=0$.

3 A Logic Model for Association Rules

The logic model we are going to use is based in a method developed in the sixties by Háyek et al. [6]. This method calls GUHA (General Unary Hypotheses Automaton) and it has a good logic and statistical base which contributes to a better understanding of two important aspects of association rules: their nature and the properties of the interest measures used for their evaluation. Recently, several authors have implemented a good and fast algorithm [11] based on a data representation using bit strings. We also use a similar algorithm based on bitsets (an item representation by means of a set of bits) that is low time consuming for computing the items conjunction and their cardinality (support). Having the representation of every item into bitsets computing the contingency table which is based this model is straightforward and low time consuming [11].

Let M be a data matrix where its rows $O_1,\ldots,O_n$ are associated to a set of observed objects, and its columns $A_1,\ldots,A_K$ are the associated attributes which describe the objects. In this way, the entry (i,j) of M will be 1 when the object O_i satisfies the attribute A_j and 0 in other case. In the association rules framework, each matrix M will represent a transaction and the set of all matrixes (transactions) will be the database called D.

D	$<O_1,A_1>$	$\ldots$	$<O_n,A_K>$
t_1	1	$\ldots$	0
t_2	0	$\ldots$	1
$\vdots$	$\vdots$	$\ddots$	$\vdots$
t_n	1	$\ldots$	1

For the logic model we are presenting, an *item* will be a pair of the form $<O_1,A_1>$, and an *itemset* will be an aggregation of items using the usual logic connectors: $\wedge,\vee,\neg$. An association rule in the model proposed in [6] is an expression of the type $\varphi\approx\psi$ where φ and ψ represent itemsets (in the sense before) derived from a database D, and the symbol $\approx$ called *quantifier* is an evaluation or condition for the fulfilment of the association rule which will depend on the interest measure used and on the *four fold table*, $4ft$ associated to the itemsets φ and ψ. An example of association rule could be $<O_1,A_1>\wedge<O_3,A_2>\approx<O_2,A_5>\wedge\neg<O_3,A_7>$.

For any pair of attributes φ and ψ the so called *four fold table* may be constructed from the database D as follows:

$\mathcal{M}$	ψ	$\neg\psi$
φ	a	b
$\neg\varphi$	c	d

This table will be represented by $\mathcal{M}=4ft(\varphi,\psi,D)=\langle a,b,c,d\rangle$ where a,b,c and d will be non negative integers satisfying that a is the number of objects (i.e. the number of rows of D) which contain at the same time the itemsets φ and ψ, b the number of objects satisfying φ and not ψ, and analogously for c and d. It is obvious that the inequality $a+b+c+d>0$ is always satisfied.

The association rule $\varphi\approx\psi$ will be true in the database D (or in the matrix $\mathcal{M}$) if and only if the condition associated to the $4ft$-quantifier $\approx$ is satisfied for the four fold table $4ft(\varphi,\psi,D)$.

Depending on the type of $4ft$-quantifier we can express different kinds of associations between the itemsets φ and ψ. In [6] and [7] we can find some examples. The classical framework of support and confidence can be modeled by means of two quantifiers: the *support* and the *implication* $4ft$-quantifiers as follows:

$$\approx_{supp}(a,b,c,d)=\frac{a+b}{a+b+c+d},\quad\Rightarrow(a,b)=\frac{a}{a+b} \tag{12}$$

which must exceed the $0<minsupp,minconf<1$ thresholds respectively imposed by the user. In [12] we explain the existing relation between the $4ft$-quantifiers and the interestingness measures used in the evaluation and validation of association rules.

4 A Logic Model for Fuzzy Association Rules

In [13] the concepts of transaction and association rule are generalized to the fuzzy case.

Definition 10. [13] Let $I=\{i_1,\ldots,i_m\}$ be a finite set of items. A fuzzy transaction is a non empty fuzzy subset $\tilde{\tau}\subseteq I$.

ISBN: 978-989-95079-6-8

Table 1: *RL*-representations associated to several imprecise events derived from the atomic properties X and Y.

α_i	$\rho_X(\alpha)$	$\rho_{\neg X}(\alpha)$	$\rho_Y(\alpha)$	$\rho_{X\wedge Y}(\alpha)$	$\rho_{X\vee Y}(\alpha)$	$\rho_{X\wedge\neg Y}(\alpha)$
1	$\{u_1\}$	$\{u_2,u_3,u_4,u_5,u_6\}$	$\emptyset$	$\emptyset$	$\{u_1\}$	$\{u_1\}$
0.9	$\{u_1\}$	$\{u_2,u_3,u_4,u_5,u_6\}$	$\{u_1\}$	$\{u_1\}$	$\{u_1\}$	$\emptyset$
0.8	$\{u_1,u_2\}$	$\{u_3,u_4,u_5,u_6\}$	$\{u_1\}$	$\{u_1\}$	$\{u_1,u_2\}$	$\{u_2\}$
0.6	$\{u_1,u_2\}$	$\{u_3,u_4,u_5,u_6\}$	$\{u_1,u_3\}$	$\{u_1\}$	$\{u_1,u_2,u_3\}$	$\{u_2\}$
0.5	$\{u_1,u_2,u_3\}$	$\{u_4,u_5,u_6\}$	$\{u_1,u_3,u_4\}$	$\{u_1,u_3\}$	$\{u_1,u_2,u_3,u_4\}$	$\{u_2\}$
0.4	$\{u_1,u_2,u_3,u_5\}$	$\{u_4,u_6\}$	$\{u_1,u_3,u_4\}$	$\{u_1,u_3\}$	$\{u_1,u_2,u_3,u_4,u_5\}$	$\{u_2,u_5\}$

Table 2: *RL*-probabilities in U of the imprecise events in table 1 and the conditional *RL*-probability $p(Y|X)$.

α_i	$\rho_{p(X)}(\alpha)$	$\rho_{p(\neg X)}(\alpha)$	$\rho_{p(Y)}(\alpha)$	$\rho_{p(X\wedge Y)}(\alpha)$	$\rho_{p(X\vee Y)}(\alpha)$	$\rho_{p(X\wedge\neg Y)}(\alpha)$	$\rho_{p(Y\|X)}(\alpha)$
1	1/6	5/6	0	0	1/6	1/6	0
0.9	1/6	5/6	1/6	1/6	1/6	0	1
0.8	1/3	2/3	1/6	1/6	1/3	1/6	1/2
0.6	1/3	2/3	1/3	1/6	1/2	1/6	1/2
0.5	1/2	1/2	1/2	1/3	2/3	1/6	2/3
0.4	2/3	1/3	1/2	1/3	5/6	1/3	1/2

For every item $i \in I$ and every transaction $\tilde{\tau}$, an item i will belong to $\tilde{\tau}$ with grade[1] $\tilde{\tau}(i)$ where $\tilde{\tau}(i)$ is a real number in the interval $[0,1]$.

Let $A \subseteq I$ be an itemset. The membership grade of A to the fuzzy transaction $\tilde{\tau}$ is defined as

$$\tilde{\tau}(A) = \min_{i\in A} \tilde{\tau}(i).$$

Using the definition 10 a crisp transaction will be a special case of fuzzy transaction where every item in the transaction will have membership grade equal to 1 or 0 depending on if they are in the transaction or not.

Example 2. We consider the set of items $I = \{i_1, i_2, i_3, i_4, i_5\}$ and the set of fuzzy transactions given by the table 3.

Table 3: Set of fuzzy transactions $\tilde{D}_1$

	i_1	i_2	i_3	i_4	i_5
$\tilde{\tau}_1$	1	0.2	1	0.9	0.9
$\tilde{\tau}_2$	1	1	0.8	0	0
$\tilde{\tau}_3$	0.5	0.1	0.7	0.6	0
$\tilde{\tau}_4$	0.6	0	0	0.5	0.5
$\tilde{\tau}_5$	0.4	0.1	0.6	0	0
$\tilde{\tau}_6$	0	1	0	0	0

In particular, we can see that $\tilde{\tau}_6$ is a crisp transaction. Some membership grade could be: $\tilde{\tau}_1(\{i_3,i_4\})$ =0.9, $\tilde{\tau}_1(\{i_2,i_3,i_4\})$ =0.2 and $\tilde{\tau}_2(\{i_1,i_2\})$ =1.

Definition 11. [13] Let I be a set of itmes, $\tilde{D}$ a set of fuzzy transactions and $A,B \in I$ two disjoint itemsets, i.e. $A \cap B = \emptyset$. A fuzzy association rule is satisfied in $\tilde{D}$ if and only if, $\tilde{\tau}(A) \leq \tilde{\tau}(B)$ for all $\tilde{\tau} \in \tilde{D}$, that is, the membership grade of B is higher than the membership grade of A for all fuzzy transactions $\tilde{\tau}$ in $\tilde{D}$.

[1]Note that $\tilde{\tau}(i)$ is $\mu_{\tilde{\tau}}(i)$ where $\mu_{\tilde{\tau}} : I \longrightarrow [0,1]$ is the membership function associated to the fuzzy set $\tilde{\tau}$ on the referencial $I = \{$ set of items $\}$.

This definition maintains the meaning of crisp association rules because if we need that $A \subseteq \tilde{\tau}$ is satisfied, we also need that $B \subseteq \tilde{\tau}$ be satisfied, in our case this can be translated to $\tilde{\tau}(A) \leq \tilde{\tau}(B)$. In this way, as a crisp transaction is a special case of fuzzy transaction, a crisp association rule will be a special case of fuzzy association rule.

Let $\tilde{\Gamma}_A$ and $\tilde{\Gamma}_B$ two fuzzy sets defined on $\tilde{D}$ as $\tilde{\Gamma}_A(\tilde{\tau}) = \tilde{\tau}(A)$ and $\tilde{\Gamma}_B(\tilde{\tau}) = \tilde{\tau}(B)$ respectively. As both are fuzzy sets, we can obtain their own *RL*-representations that we note[2] as $(\Lambda_{\tilde{A}}, \rho_{\tilde{A}})$, $(\Lambda_{\tilde{B}}, \rho_{\tilde{B}})$, defined as the equations (1) and (3) show.

For the crisp case, some of the suggested measures for the validation of association rules are defined using the concept of probability. This is the case of support and confidence. Using the *RL*-probabilities we can easily extend the classical framework of support and confidence.

Definition 12. (Support of an itemset) Let $A \subseteq I$ be an itemset and $(\Lambda_{p(\tilde{A})}, \rho_{p(\tilde{A})})$ the *RL*-representation of the *RL*-probability associated to the fuzzy set $\tilde{\Gamma}_A$ in $\tilde{D}$. Then, the support of A in the fuzzy database $\tilde{D}$ is defined as

$$supp(A) = \sum_{\alpha_i \in \Lambda_{p(\tilde{A})}} (\alpha_i - \alpha_{i+1}) \left(\rho_{p(\tilde{A})}(\alpha_i)\right). \qquad (13)$$

Following a similar reasoning we will define the support and the confidence for a fuzzy association rule $A \rightarrow B$.

Definition 13. Let be $A,B \subseteq I$ two disjoint itemsets and $(\Lambda_{p(\tilde{A}\wedge\tilde{B})}, \rho_{p(\tilde{A}\wedge\tilde{B})})$, $(\Lambda_{p(\tilde{B}|\tilde{A})}, \rho_{p(\tilde{B}|\tilde{A})})$ the *RL*-representations of the *RL*-probabilities $p(\tilde{\Gamma}_A \wedge \tilde{\Gamma}_B)$ and $p(\tilde{\Gamma}_B|\tilde{\Gamma}_A)$ in $\tilde{D}$. Then, the support and the confidence of a fuzzy association rule $A \rightarrow B$ on $\tilde{D}$ are defined as

$$Supp(A \rightarrow B) = \sum_{\alpha_i \in \Lambda_{p(\tilde{A}\wedge\tilde{B})}} (\alpha_i - \alpha_{i+1}) \left(\rho_{p(\tilde{A}\wedge\tilde{B})}(\alpha_i)\right) \qquad (14)$$

$$Conf(A \rightarrow B) = \sum_{\alpha_i \in \Lambda_{p(\tilde{B}|\tilde{A})}} (\alpha_i - \alpha_{i+1}) \left(\rho_{p(\tilde{B}|\tilde{A})}(\alpha_i)\right) \qquad (15)$$

[2]A and B are crisp sets of disjoint items and $\tilde{\Gamma}_A$, $\tilde{\Gamma}_B$ are fuzzy sets defined on $\tilde{D}$

ISBN: 978-989-95079-6-8

We want to remark that in the previous definitions $\Lambda_{p(\tilde{A}\wedge\tilde{B})}$ and $\Lambda_{p(\tilde{B}|\tilde{A})}$ coincides and are equal to $\Lambda_{\tilde{A}}\cup\Lambda_{\tilde{B}}$.

But some other interesting measures are not in terms of probability. For those, the model is necessary for the generalization of those crisp measures of interest.

The following itemsets: $A\wedge B, A\wedge\neg B, \neg A\wedge B, \neg A\wedge\neg B$ are a partition of the database $\tilde{D}$ (in the same way as $\varphi\wedge\psi, \varphi\wedge\neg\psi, \neg\varphi\wedge\psi$ and $\neg\varphi\wedge\neg\psi$ are a partition for the crisp case). We take their associated fuzzy sets defined on $\tilde{D}$ with their own *RL*-representations that we are going to call $(\Lambda_{\tilde{A}\wedge\tilde{B}}, \rho_{\tilde{A}\wedge\tilde{B}})$, $(\Lambda_{\tilde{A}\wedge\neg\tilde{B}}, \rho_{\tilde{A}\wedge\neg\tilde{B}})$, etc. Note that the *RL*-sets obtained contain the same restriction levels, that is, the set of restriction levels is common and equal to $\Lambda_{\tilde{A}}\cup\Lambda_{\tilde{B}}$.

As for every $\alpha\in\Lambda_Y$, $\rho_Y(\alpha)$ is a crisp set, we can compute its cardinality as usual, and we note it by $|\rho_Y(\alpha)|$.

In this way, for every restriction level $\alpha_i\in\Lambda_{\tilde{A}}\cup\Lambda_{\tilde{B}}$ we define the associated four fold table $\mathcal{M}_{\alpha_i} = 4ft(\tilde{\Gamma}_A, \tilde{\Gamma}_B, \tilde{D}, \alpha_i)$ as follows:

$\mathcal{M}_{\alpha_i}$	$\tilde{\Gamma}_B$	$\neg\tilde{\Gamma}_B$
$\tilde{\Gamma}_A$	a_i	b_i
$\neg\tilde{\Gamma}_A$	c_i	d_i

where a_i, b_i, c_i and d_i are non negative integers such that

$$a_i = |\rho_{\tilde{A}\wedge\tilde{B}}(\alpha_i)|$$
$$b_i = |\rho_{\tilde{A}\wedge\neg\tilde{B}}(\alpha_i)|$$

and analogously with c_i and d_i.

Note that $\forall\alpha_i\in\Lambda_{\tilde{A}}\cup\Lambda_{\tilde{B}}$ the following equality is satisfied

$$a_i+b_i+c_i+d_i = n = |\tilde{D}|. \tag{16}$$

We are going to prove that this definition is a good generalization to the crisp case. Suppose that A and B are two itemsets in a crisp database D, in this case, $\tilde{\Gamma}_A(\tilde{\tau}) = \tilde{\tau}(A)\in[0,1]$ is equivalent to $t(A)\in\{0,1\}$ where $t(A)$ is given by the following indicator function:

$$t(A) = \begin{cases} 1 & \text{si } A\in t \\ 0 & \text{si } A\notin t \end{cases} \tag{17}$$

where t is a crisp transaction of D. The associated *RL*-sets to $\tilde{\Gamma}_A$ and to $\tilde{\Gamma}_B$ will be $\Lambda_{\tilde{A}} = \Lambda_{\tilde{B}} = \{1\}$. And the associated *RL*-representations will be: $(\Lambda_{\tilde{A}}, \rho_{\tilde{A}})$ y $(\Lambda_{\tilde{B}}, \rho_{\tilde{B}})$ where

$$\begin{aligned} \rho_{\tilde{A}}(1) &= (\tilde{\Gamma}_A)_1 = \{t\in D\,|\,t(A)\geq 1\} = \{t\in D\,|\,A\in t\} \\ \rho_{\tilde{B}}(1) &= (\tilde{\Gamma}_B)_1 = \{t\in D\,|\,t(B)\geq 1\} = \{t\in D\,|\,B\in t\} \end{aligned} \tag{18}$$

Following the same process, we can compute the associated *RL*-representations for the sets $\tilde{\Gamma}_A\wedge\tilde{\Gamma}_B$, $\tilde{\Gamma}_A\wedge\neg\tilde{\Gamma}_B$, $\neg\tilde{\Gamma}_A\wedge\tilde{\Gamma}_B$, $\neg\tilde{\Gamma}_A\wedge\neg\tilde{\Gamma}_B$. Therefore, the $4ft$ table for the restriction level $\alpha=1$ is the same as the $4ft$ table for the itemsets A and B seen in section 3:

$\mathcal{M}_1\equiv\mathcal{M}$	B	$\neg B$
A	a_1	b_1
$\neg A$	c_1	d_1

Using this representation it is easy to generalize every kind of interest measure used in the crisp case, in particular, we can generalize every $4ft$-quantifier.

If we look again to the classical framework of support and confidence, we can extend it as we show from now on.

Let be $A, B\subseteq I$ two disjoint itemsets and $(\Lambda_{\tilde{A}}, \rho_{\tilde{A}}), (\Lambda_{\tilde{B}}, \rho_{\tilde{B}})$ the *RL*-representations associated to the fuzzy sets $\tilde{\Gamma}_A$ and $\tilde{\Gamma}_B$ in $\tilde{D}$. Then, the support and the confidence of a fuzzy association rule $A\rightarrow B$ on $\tilde{D}$ are defined as

$$Supp(A\rightarrow B) = \sum_{\alpha_i\in\Lambda_{\tilde{A}}} (\alpha_i-\alpha_{i+1})\,(\approx_{supp}(a_i,b_i,c_i,d_i)), \tag{19}$$

$$Conf(A\rightarrow B) = \sum_{\alpha_i\in\Lambda_{\tilde{A}}} (\alpha_i-\alpha_{i+1})\,(\Rightarrow(a_i,b_i)). \tag{20}$$

The support and the confidence defined above coincide with those defined by means of probabilities based on RL-numbers. In this last equation (20) we can find an indetermination of the type "$\frac{0}{0}$", in this case, we will take the value 1 for preserving the definition 11 of fuzzy association rule, because this indetermination happens when there is no transactions satisfying the antecedent and the consequent of the rule among zero transactions satisfying only the antecedent. Again, it is easy to see that for every restriction level α_i the last part of the equations (19) and (20) are the same as the defined support and confidence for a crisp association rule (see quantifiers in (12)).

This process can be done for every measure of interest used in the validation and evaluation of crisp association rules. In particular we use the certainty factor instead of confidence because of its proven good properties: (1) it reduces the number of obtained rules and (2) it provides an useful meaning to the measurement of the validity of association rules. In [12], we prove by means of the logic model that the certainty factor can be seen as an equivalence $4ft$-quantifier as follows:

$$\equiv_{CF}(a,b,c,d) = \begin{cases} \dfrac{ad-bc}{(a+b)(b+d)} & \text{if } ad>bc \\ 0 & \text{if } ad=bc \\ \dfrac{ad-bc}{(a+b)(a+c)} & \text{if } ad<bc. \end{cases} \tag{21}$$

Its generalization for fuzzy rules is straightforward:

$$CF(A\rightarrow B) = \sum_{\alpha_i\in\Lambda_{\tilde{A}}} (\alpha_i-\alpha_{i+1})\,(\equiv_{CF}(a_i,b_i,c_i,d_i)). \tag{22}$$

We want to remark that the previous concepts of support, confidence and the certainty factor perfectly extends the support, confidence and certainty factor measures used for crisp association rules. This can be immediately proved because the measures obtained in equations from 19 to 21 coincide with the ones proposed in [13] where it is used a semantic approach based on the evaluation of quantified sentences (see also [14]) using the *GD* method and the fuzzy relative quantifier the majority $Q_M(x)=x$ for evaluating the sentences. In these papers it is proved that using the *GD* method and the quantifier Q_M, the obtained measures in equations (19)-(22) are the ordinary support, confidence and certainty factor for the rule $A\rightarrow B$ respectively.

In addition, using the proposed logic model which uses the representation by means of the $4ft$ tables for each restriction

ISBN: 978-989-95079-6-8

level, is immediate to prove that if A and B are two itemsets in a set of crisp transactions, the concepts of support, confidence and certainty factor for instance (19)-(22), are equal to the usual crisp concepts of support, confidence and certainty factor (see again equations (12) and (21)).

5 Concluding Remarks

We have presented a simple model with very good properties for the representation and evaluation of fuzzy association rules which extends in a natural way the evaluation of crisp association rules.

We have combined two existing approaches: a logic model for crisp association rules [7] and the recently developed RL-representation theory for imprecise properties [1]. In particular, using the philosophy of RL-representations we define the RL-numbers for the representation of fuzzy quantities [8] and the associated probabilities which are useful for extending crisp quality measures like support and confidence. RL-numbers and probabilities have been also introduced in [15],[16] named gradual numbers. Both approaches are similar with slight differences that mainly affect how is the negation considered [15]. The main similarities and differences between both approaches is treated in [17]. But this is beyond the scope of this paper.

We used RL-numbers and their associated probabilities for the special cases of computing support and confidence of fuzzy association rules but there are many cases that the rule's assessment is made by other measures which are not in terms of probabilities. For these cases we also provide a very simple and systematic approach for extending the crisp measures for mining fuzzy association rules, for instance the certainty factor is extended by equation (22).

In this line of research we have found works like the one in [5]. They develop a systematic approach to the assessment of fuzzy association rules. To this end, they proceed from the idea of partitioning the data stored in a database into examples of the given rules, counterexamples and irrelevant data. Maybe what they called irrelevant data is not so "irrelevant" for measuring the accomplishment of an association rule. See for example the case of the certainty factor in equation (21) which uses the so called irrelevant data (divided into two different frequencies, c and d) for measuring the strength of the rule, and in addition, its good properties are experimentally proved in [13] and formally proved in [12] using the introduced logic model.

Our approach has some similarities with the approach in [5] as we also divide the data, but we use a conjunction-based model, taking all the possible conjunctions between rule's antecedent and consequent. The main difference between the former and our proposal remains in the way of generalizing the quality measures for fuzzy associations, as we use a four fold table's RL-representation.

Since algorithmic aspects of fuzzy rule mining have not been addressed in this paper, let us mention that the method used in [11] for mining crisp rules using the bitset approach can be easily adapted to the fuzzy case and the complexity will increase depending on the number of restrictions levels taken into account.

Acknowledgment

This work has been partially supported by the projects TIN2006-15041-C04-01 and TIN2006-07262.

References

[1] D. Sánchez, M. Delgado, and M.A. Vila. A restriction level approach to the representation of imprecise properties. In *IPMU'08*, pages 153–159, Málaga, Spain, 2008.

[2] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[3] T-P Hong and Y-C Lee. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, chapter An Overview of Mining Fuzzy Association Rules, pages 397–410. Springer, 2008.

[4] E. Hüllermeier. Fuzzy sets in machine learning and data mining. *Applied Soft Computing, to appear*, 2008.

[5] D. Dubois, E. Hüllermeier, and H. Prade. A systematic approach to the assessment of fuzzy association rules. *Data Min. Knowl. Discov.*, 13(2):167–192, 2006.

[6] P. Hájek and T. Havránek. *Mechanising Hypothesis Formation-Mathematical Foundations for a General Theory*. Springer-Verlag: Berlin, Heidelberg and New York, 1978.

[7] J. Rauch. Logic of association rules. *Applied Intelligence*, 22:9–28, 2005.

[8] D. Sánchez, M. Delgado, and M.A. Vila. Rl-numbers: An alternative to fuzzy numbers for the representation of imprecise quantities. In *WCCI 2008*, Hong Kong, China, To appear.

[9] D. Sánchez, M. Delgado, and M.A. Vila. A restriction level approach to probability and statistics. In *ESTYLF'08*, pages 107–112, Asturias, Spain, 2008.

[10] M. Delgado, M.D. Ruiz, and D. Sánchez. Analyzing exception rules. In *IPMU'08*, Málaga, Spain.

[11] J. Rauch and M. Šimunek. An alternative approach to mining association rules. *Studies in Computational Intelligence (SCI)*, 6:211–231, 2005.

[12] M. Delgado, M.D. Ruiz, and D. Sánchez. Studying interest measures for association rules through a logical model. *IJUFKS*, Submitted, 2007.

[13] M. Delgado, N. Marín, D. Sánchez, and M.A. Vila. Fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems*, 11(2):214–225, 2003.

[14] M. Delgado, M.D. Ruiz, and D. Sánchez. Fuzzy association rules: new challenges (in spanish). In *ESTYLF'08*, Langreo-Mieres, Spain.

[15] D. Dubois and H. Prade. Gradual elements in a fuzzy set. *Soft Computing*, 12:165–175, 2008.

[16] D. Dubois and H. Prade. Fuzzy elements in a fuzzy set. In *Proc. 10th Int. Fuzzy Systems Assoc. (IFSA)*, pages 55–60, Beijing, 2005.

[17] D. Sánchez, M. Delgado, and M.A. Vila. A discussion about representation models for imprecise quantities and probabilities. In *30th Linz Seminar on Fuzzy Set Theory*, Linz, Austria, 2009.

ISBN: 978-989-95079-6-8

Semantic Proximity Between Queries and the Empty Answer Problem

Patrick Bosc Carmen Brando Allel Hadjali Hélène Jaudoin Olivier Pivert

IRISA/ENSSAT, Université de Rennes 1
6, Rue de Kerampont, BP 80518, 22305 Lannion Cedex, France
Email: {bosc, brando, hadjali, jaudoin, pivert}@enssat.fr

Abstract— This paper proposes an approach aimed at obviating empty answers for a family of conjunctive queries involving Boolean or fuzzy value constraints. Contrary to the approaches based on a relaxation of the predicates involved in the query, the principle suggested here consists in replacing the query by another one similar which has been processed previously and whose answer is known to be non-empty. This technique thus avoids the combinatory explosion induced by classical relaxation-based approaches.

Keywords— Databases, fuzzy queries, empty answers, proximity.

1 Introduction

Since the late 80's, there is an increasing interest in designing intelligent information systems endowed with some cooperative behavior [12]. The most well-known issue approached in this field is the empty answer problem, that is, the problem of providing the user with some alternative data when there is no data fitting his/her query. Several approaches have been proposed to deal with this issue. The relaxation paradigm [13] is one of the basic cooperative techniques used in most of such approaches. Query relaxation aims at expanding the scope of a query searching for answers which are in the neighborhood of the user's query and consists in replacing some query conditions by more general conditions or in just eliminating some of them.

Let us note that manually relaxing failing queries is a tedious and time-consuming task because, in the worst case, one must consider an exponential number of possible relaxations [21]. Hence, several automated approaches to query relaxation have been proposed (see, e.g., [8, 11, 15, 17, 20, 21]). The main objective of those approaches is to modify a failing user query into a relaxed query whose answer is non-empty, or at least to identify the cause of the failure. Some of those works rely on the key concept of false presuppositions (a presupposition of a statement is any statement entailed by the original, for instance, the statement "the king of France is bald" has as presupposition "there is a king of France" which is a false presupposition). Motro [20] has addressed the issue of empty answers, i.e. when a query fails to produce any answers, by proposing a relaxation method which focuses on finding the false presuppositions of a failing query. A related approach has been proposed by Godfrey [15], who considers any subquery as a presupposition of the query itself. The focus of this work is the search for Minimal Failing Subqueries (MFSs) of a failing query.

In this paper, we propose an approach for dealing with failing conjunctive queries (Boolean or fuzzy), according to another approach than relaxation as described above. The idea that we advocate is not to modify/suppress some predicates from a failing query Q using this sole query, but rather to find a "good" global substitute to Q among queries previously submitted to the system whose answer is known to be non-empty. We thus consider a context where the system stores the non-failing queries in a repository D^+. One also assumes available a resemblance measure over every attribute domain involved in the database considered. The approach raises the question of defining the notion of semantic proximity between queries. With respect to related works, which will be briefly presented farther, the main originality of the approach introduced here is to take into account queries involving Boolean or fuzzy value constraints in an explicit form.

An important gain brought by this method, relatively to a classical relaxation-based approach, lies in the fact that it avoids the combinatory explosion induced by the relaxation of the predicates from a conjunctive query. Indeed, there exists in general a high number of relaxed queries and one cannot know whether these queries provide a non-empty answer before processing them. With the approach proposed here, one has the guarantee to obtain a non-empty answer in one step because only one query needs to be processed. In this paper, we limit the scope to conjunctive selection queries involving value constraints which can be represented either by crisp sets, intervals or fuzzy sets. The remainder of the paper is organized as follows. Section 2 presents a query substitution approach in the case where the predicates expressing value constraints are Boolean. Section 3 generalizes this approach to the case of conjunctive fuzzy queries. In Section 4, we describe the principle of a mechanism aimed at providing the user with some (at least partial) explanations about the causes of the original empty answer. The technique proposed uses both the repository D^+ and a second one denoted by D^- which contains the failing queries previously submitted to the system. Section 5 is devoted to a comparison of the approach with some related works. Finally, the conclusion recalls the main contributions of the paper and outlines some perspectives for future work.

2 Boolean Queries

2.1 Single-Predicate Queries

Let $Q = (A \text{ in } E)$ denote the user query where A is an attribute and E a set or an interval, and Q' a query from the repository D^+. We denote by r the database relation concerned by Q and by res the proximity relation defined over the domain of A.

2.1.1 Case where E is a finite set

It is assumed that the user query Q returns an empty answer. One thus has to search through D^+ so as to retrieve the queries

ISBN: 978-989-95079-6-8

involving a predicate of the form (A in E').

Remark 1. Any set E' present in the queries from D^+ is such that $E' \not\subseteq E$. Otherwise the associated query Q' would have returned an empty answer and Q' would not be in D^+.

The emptiness of the answer to Q means that none of the elements from E is present as an A-value in relation r. In order to obtain a substitute to Q which returns a non-empty answer, it is thus necessary to find a query Q' bearing on r and involving a predicate (A in E') such that E' contains at least one value absent from E. However, so as not to drift too far away from the initial user need, it is desirable that the elements from E' absent from E be sufficiently close to at least one element from E. Consequently, one looks for the set E' which is as close as possible to E, so as to replace E by $(E' - E)$ in Q. In order to find this "best" E', a measure is needed. It is not strictly speaking a proximity measure, since the symmetry property is not desired here. Indeed one wants to know whether E' is a good substitute to E, but not necessarily the reciprocal. Several possible substitutivity measures (denoted by sbs_i later on) are discussed hereafter.

1st idea: one assesses the extent to which every element from $(E' - E)$ resembles at least one element from E:

$$sbs_1(E,\ E') = inf_{x' \in (E'-E)}\ sup_{x \in E}\ res(x,\ x'). \quad (1)$$

The problem with this measure is that the worst element "masks" the others, as illustrated in the next example. In the following, we assume available the following subset of a resemblance relation on animals:

res(rooster, hen) = 0.9, res(rooster, duck) = 0.6, res(rooster, turkey) = 0.7, res(hen, duck) = 0.6, res(hen, turkey) = 0.7, res(cow, hen) = res(cow, duck) = res(cow, turkey) = 0.

Example 1. E = {hen, duck, turkey}, E'_1 = {hen, turkey, cow}, E'_2 = {cow, rooster}. We get $sbs_1(E,\ E'_1) = sbs_1(E,\ E'_2) = 0$ but since there are neither hens nor turkeys in the database — otherwise Q would not be failing —, it seems reasonable to claim that E'_2 should be a better substitute than E'_1. However, in the computation of $sbs_1(E,\ E'_2)$, the element cow "masks" rooster.◇

2nd idea: one assesses the extent to which there is an element from $(E' - E)$ which resembles at least an element from E:

$$sbs_2(E,\ E') = sup_{x' \in (E'-E)}\ sup_{x \in E}\ res(x,\ x'). \quad (2)$$

Here, the difficulty is that the "winning set" may include elements which are very distant from those desired by the user.

Example 2. E = {hen}, E'_1 = {hen, cow, turkey}, E'_2 = {duck}. Here, E'_2 should win, since it includes only elements close to the desired ones, contrary to E'_1, which includes "cow". However, it is E'_1 which wins since $sbs_2(E,\ E'_1) = 0.7$ while $sbs_2(E,\ E'_2) = 0.6$.◇

3rd idea: one mixes the quantitative and the qualitative aspects by measuring the average resemblance degree between an element from $(E' - E)$ and an element from E. For each element x' from $(E' - E)$, the corresponding measure looks for the maximal proximity between x' and an element x from E, computes the sum of these maximal proximities, and divides this sum by the number of elements present in $(E' - E)$:

$$sbs_3(E,\ E') = \frac{\sum_{x' \in (E'-E)}\ sup_{x \in E}\ res(x,\ x')}{|E' - E|}. \quad (3)$$

Example 3. E = {hen, duck, turkey}, E'_1 = {hen, turkey, cow}, E'_2 = {cow, rooster}. We get: $sbs_3(E,\ E'_1) = 0$ and $sbs_3(E,\ E'_2) = 0.45$.◇

Since measure sbs_3 appears the most satisfactory, it will be used in the following.

2.1.2 *Case where E is an interval*

It is quite straightforward to extend measure sbs_3 defined previously so as to make it work with intervals instead of sets: one just has to replace the sum by an integral. The calculus is rather simple since it boils down to computing areas of rectangles or trapezoids.

Let us first consider the simple case where resemblance is defined in a Boolean manner:

$$res(x,\ y) = 1 \text{ if } |x - y| \leq \delta, 0 \text{ otherwise.} \quad (4)$$

Let us consider two intervals: $I = [m,\ M]$, that from the user query, and $I' = [m',\ M']$, that from the candidate substitute. Let us assume that $m \leq m'$, the dual case can be obtained straightforwardly from this one. One gets:

$$sbs_3(I, I') = \begin{cases} 0 \text{ if } M + \delta \leq m', \\ 1 \text{ if } M' \leq M + \delta, \\ (M + \delta - p)/(M' - p) \text{ otherwise,} \end{cases} \quad (5)$$

where $p = m'$ if $m' \geq M$, M otherwise. A slightly more complex case is that where resemblance is defined by means of a fuzzy tolerance indicator Z with a trapezoidal membership function of support $[-\alpha,\ \alpha]$ and of core $[-\beta,\ \beta]$ [4]. In other words:

$$res(x,\ y) = \mu_Z(|x - y|). \quad (6)$$

The principle is the same as for Boolean resemblance, except that one has to compute areas of trapezoids instead of rectangles. The case where the predicate from Q involves an interval I and that from Q' involves a set E can be managed by rewriting the definition of sbs_3 the following way:

$$sbs_3(E,\ E') = \frac{\sum_{x' \in E \wedge x' \notin I}\ sup_{x \in I}\ res(x,\ x')}{|\{x \in E \mid x \notin I\}|}. \quad (7)$$

On the other hand, the dual case (set in Q and interval in Q') is more tricky and cannot be captured by the formula defining sbs_3 when the attribute domain is continuous. Consequently, we introduce the constraint that a set can only be replaced by another set.

2.2 *Conjunctive Queries*

Let Q be the user query and Q' a query from the repository D^+.

ISBN: 978-989-95079-6-8

Remark 2. Even if query Q' involves a predicate which is subsumed by the corresponding one in Q, query Q' can be an interesting substitute to Q since the other predicates must also be taken into account. For instance, if query Q = (A in {rabbit, hen} and B in {wheat, cabbage}) returns an empty answer, it is still possible that query Q' = (A in {rabbit} and B in {wheat, oats}) returns a non-empty one whereas the predicate on A in Q' is subsumed by that in Q. For a query Q' to be a possible substitute, it is necessary that Q' involves at least one predicate which is not subsumed by the corresponding one in Q (but notice that if it were not the case, the answer to Q' would be empty — since the answer to Q is — and Q' would therefore not be in D^+).

Remark 3. The predicates from Q' which are strictly subsumed by those from Q can be replaced by the latter ones.

The substitution process that we propose consists of the following three steps:

1. select the candidate queries (and adapt these queries, see algorithm below),
2. compute the proximity degrees between the queries retained and the user query Q through the measure sbs_3,
3. determine the closest substitute to Q and process it.

Remark 4. For every predicate P from Q which is not "covered" by Q', i.e., which concerns an attribute on which there is no constraint in Q', one may compute the proximity between P and the entire domain of the attribute considered.

The conjunctive combination of the proximities related to the atomic predicates can be performed by means of a triangular norm, so as to obtain the overall proximity between two queries. Notice that alternative solutions could also be possible, for instance one might use a mean operator. The substitution algorithm is outlined hereafter.

Algorithm:

Let Q' be a query from D^+ which shares at least one attribute from its "where" clause with that from Q. The five steps of the algorithm are:

1. replace the "select" clause from Q' by that from Q;
2. remove from Q' every predicate that concerns an attribute absent from the "where" clause from Q;
3. replace every predicate from Q' which is strictly subsumed by the corresponding one from Q by the latter;
4. for the other predicates, compute the proximity between the predicate from Q' and the corresponding one from Q, by means of measure sbs_3, and replace the predicate from Q' by its union with that from Q. As to the predicates from Q which are not covered by Q', one computes their substitutivity degree relatively to the entire domain of the attribute involved;
5. aggregate the local proximities by means of a triangular norm (the idea is to assess the extent to which every predicate of the substitute query is close to the corresponding predicate from the initial failing query).

Table 1: Resemblance relation over attribute veg

	co	ra	su	wh	ca	br	po	ru
co	1	0.4	0.3	0.8	0.1	0.1	0.6	0.4
ra	0.4	1	0.9	0.6	0.2	0.2	0.1	0.1
su	0.3	0.9	1	0.5	0.1	0.1	0.3	0.3
wh	0.8	0.6	0.5	1	0.2	0.1	0.5	0.4
ca	0.1	0.2	0.1	0.2	1	0.9	0.6	0.7
br	0.1	0.2	0.1	0.1	0.9	1	0.4	0.6
po	0.6	0.1	0.3	0.5	0.6	0.4	1	0.8
ru	0.4	0.1	0.3	0.4	0.7	0.6	0.8	1

Example 4. Let Q be the following failing user query:

select #id **from** F
where veg **in** {corn, rapeseed} **and**
city **in** {Lannion, Caouennec, Prat} **and**
area **in** [60, 100].

Let us assume that the domain of "veg" is: {corn, rapeseed, sunflower, wheat, cabbage, broccoli, potato, rutabaga} and that the associated resemblance relation is given in Table 1. Let Q'_1 be the following query from D^+:

select #name **from** F
where veg **in** {wheat, rapeseed, sunflower} **and**
city **in** {Lannion, Prat} **and** area = 125 **and**
animal **in** {cow, pig}.

The query Q''_1 obtained by adapting Q'_1 according to the algorithm above is:

select #id **from** F
where veg **in** {corn, rapeseed, wheat, sunflower} and
city **in** {Lannion, Caouennec, Prat} and
(area **in** [60, 100] **or** area = 125).

The degree computed by sbs_3 for the substitution of {corn, rapeseed} by {wheat, rapeseed, sunflower} equals:

$$\frac{max(0.8,\ 0.6) + max(0.3,\ 0.9)}{2} = 0.85.$$

Let us assume that the proximity over the areas is based on a fuzzy tolerance indicator Z with a triangular membership function of support $[-50,\ 50]$. The substitution of [60, 100] by 125 is assigned the degree 0.5 (i.e., the proximity degree between 100 and 125). Finally, the degree computed for Q''_1 using the t-norm minimum is:

$$min(0.85, 0.5) = 0.5.$$

Let us now consider another query, denoted by Q'_2, from D^+:

select #name **from** F
where city = Caouennec **and**
area **in** [80, 180] **and** animal **in** {sheep, goat}.

Altering Q'_2 according to the algorithm yields Q''_2:

ISBN: 978-989-95079-6-8

select #id **from** F
where city **in** {Lannion, Caouennec, Prat} **and**
area **in** [60, 180].

As to the condition on attribute *veg* we get:
sbs_3({corn, rapeseed}, *domain(veg)*) =
(0.9 + 0.8 + 0.2 + 0.2 + 0.6 + 0.4)/6 = 0.52.

As to the condition on attribute *area*, we get:
sbs3([60, 100], [80, 180]) =
$((150 - 100)/2)/(180 - 100) = 25/80 = 0.31$.

Thus, the degree attached to Q_2'' is: $min(0.52, 0.31) = 0.31$ and Q_1'' is a better substitute to Q than Q_2''.$\diamond$

Remark 5. In case of ties, one could take into account the cardinality of the result of each candidate query so as to break these ties, provided that these cardinalities are stored in D^+.

3 Fuzzy Queries

Let us now move to the case where value constraints are expressed by means of fuzzy predicates. Let us consider a conjunctive fuzzy query $Q = P_1$ and ... and P_n where any predicate P_i is of the form (A_i is T_i) and T_i is a fuzzy term. Here, the fact that Q returns an empty answer means that there does not exist any element x in the database such that $\top_{i=1..n}(\mu_{T_i}(x)) > 0$, where $\top$ denotes a triangular norm generalizing the conjunction. This state of fact can be expressed by saying that the support of the query relatively to the database is empty.

In order to deal with this kind of queries, one needs to generalize measure sbs_3 by replacing the arithmetic mean by a weighted mean, and by taking into account the resemblance between the degrees coming from the two fuzzy terms considered. The generalized measure obtained, which can also be seen as a variant of the interchangeability measure proposed in [5], is defined as:

$$sbs_3(E,\ E') = \frac{\sum_{x' \in sp(E'-E)} w(x') \times \Gamma(x')}{\sum_{x' \in sp(E'-E)} w(x')} \qquad (8)$$

where:

$$\Gamma(x') = sup_{x \in E}\, min(res(x,\ x'),\ \Psi(\mu_{E'}(x'),\ \mu_E(x))), \quad (9)$$

$sp(E)$ denotes the support of E, function Ψ assesses the resemblance between two degrees in the unit interval — it can be defined e.g. as $\Psi(a,\ b) = 1 - |a - b|$ — and:

$$w(x') = (E' - E)(x') = min(E'(x'),\ 1 - E(x')). \quad (10)$$

The weight $w(x')$ captures the fact that it is all the more important to find a good substitute to x' as x' strongly belongs to $E' - E$. It is straightforward to show that if the sets are crisp, this formula reduces to that given in Section 2.

The definition above can be directly extended to the case of continuous fuzzy sets by replacing the sum by an integral.

Example 5. Let us consider the fuzzy sets: E = {1/rapeseed, 0.8/cabbage, 0.3/wheat} and E' = {0.4/rapeseed, 0.3/cabbage, 0.4/corn, 0.7/broccoli}. With the most commonly used definition of the difference between fuzzy sets, i.e.,

$$\mu_{(A-B)}(x) = min(\mu_A(x),\ 1 - \mu_B(x)), \qquad (11)$$

one gets: $E' - E$ = {0.2/cabbage, 0.4/corn, 0.7/broccoli}.

For "cabbage", the supremum equals:
sup(min(0.2, 0.3), min(1, 0.5), min(0.2, 1) = 0.5,

for "corn", it equals:
sup(min(0.4, 0.4), min(0.1, 0.6), min(0.8, 0.9)) = 0.8,

and for "broccoli" we get the degree 0.9. Hence, the final substitutivity degree equals:

(0.2×0.5 + 0.4×0.8 + 0.7×0.9)/(0.2 + 0.4 + 0.7) = 0.81.$\diamond$

The algorithm given in Subsection 2.2 can be adapted straightforwardly. The notion of subsumption between fuzzy predicates can be based on the inclusion between fuzzy sets proposed by Zadeh, i.e., $E \subseteq F \Leftrightarrow \forall x,\ \mu_E(x) \leq \mu_F(x)$.

4 Explaining the Emptiness of the Answer

Besides providing the user with a non-empty answer, it is important to also give him/her some explanations about:

- the reasons why the answer was originally empty,
- the way his/her query has been modified.

This section deals with the first point and gives the principle of a mechanism for explaining (at least partly) the emptiness of the original answer, somewhat in the spirit of [20].

4.1 Boolean Queries

Let us assume that one has available not only D^+ but also a repository D^- containing the failing queries submitted previously to the system. We suggest using both D^+ and D^- to identify some of the failing subqueries (in the sense of [15], i.e., the subsets of atomic conditions from the original query) which cause the empty answer.

Example 6. Let us consider the following failing query:

Q: veg ∈ {corn, wheat} ∧ animal ∈ {cow, pig}.

Let Q_1 and Q_2 be two queries from D^+:

Q_1: veg ∈ {corn, rapeseed} ∧ animal ∈ {sheep, pig},

Q_2: animal = pig,

and Q_3 a query from D^-: veg ∈ {rapeseed, broccoli}.

From $Q_3 \in D^-$, one can deduce that there is no rapeseed in the database relation considered. From this result and $Q_1 \in D^+$, one infers that there are some farms growing corn in the relation. From $Q_2 \in D^+$, one deduces that there are farms breeding pigs. From these two results, one can conclude that none of the terms from query Q leads to an empty answer. Consequently, the "minimal failing

ISBN: 978-989-95079-6-8

subquery" is Q itself, which means that there is some sort of "incompatibility" between (corn, wheat) and (cow, pig).$\diamond$

In order to provide the user with explanations, we suggest to partition the subqueries of Q into three classes: L_1 (those which are known to produce non-empty answers), L_2 (those which are known to be failing), L_3 (the others). Let us first consider the case of set-based predicates. A query from D^- (or D^+) is expressed: P_1 **and** ... **and** P_n where every P_i is a condition of the type (A_i **in** E_i), which corresponds to a disjunction ($A_i = v_{1,1}$ **or** ... **or** $A_i = v_{1,p}$). The first step is to transform every query from D^- into a rule:

($A_i = v_{1,1}$ **or** ... **or** $A_i = v_{1,p}$) **and** ...
and ($A_n = v_{n,1}$ or ... or $A_n = v_{n,k}$) $\rightarrow false$.

Symmetrically, every query from D^+ gives birth to a set of rules:

($A_i = v_{1,1}$ **or** ... **or** $A_i = v_{1,p}$) $\rightarrow true$
...
($A_n = v_{n,1}$ **or** ... **or** $A_n = v_{n,k}$) $\rightarrow true$.

These two types of rules constitute the rule base of the reasoning system used further.

Let us now consider a failing query Q. For each subquery of Q, one checks whether it is possible to infer *true* or *false* using the rule base. If it is possible to infer *true*, the subquery belongs to L_1, if one can infer *false* it belongs to L_2, otherwise it belongs to L_3. The subqueries have to be examined in increasing order of their size, starting with the atomic predicates, considering that a subquery including a failing subquery is itself failing (it is then useless to examine it). The corresponding algorithm, which exploits the concept of Minimal Failing Subqueries (MFS), is thus somewhat analogous to that proposed in [15], but here, we do not have to process any subquery to know if it is failing or not, due to the existence of the repositories D^+ and D^-.

In the case of intervals, one cannot replace a predicate by an explicit disjunction. One thus needs an inference engine able to deal with interval constraints.

4.2 Fuzzy queries

For discrete fuzzy sets, one uses the same principle as for Boolean queries. In the rules, the terms of a disjunction correspond to the values from the support of a fuzzy predicate. For continuous fuzzy sets, one cannot express rules involving explicit disjunctions. A query $Q = P_1$ and ... and P_n from D^- produces a rule:

A_1 **in** *support*(P_1) **and** ... **and** A_n **in** *support*(P_n) $\rightarrow false$

and a query $Q = P_1$ **and** ... **and** P_n from D^+ gives birth to the set of rules:

A_1 **in** *support*(P_1) $\rightarrow true$
...
A_n **in** *support*(P_n) $\rightarrow true$.

Here again, the reasoning engine must be able to deal with interval constraints.

5 Related Work

Some related work can be found in both domains of databases and information retrieval, including web search.

The semantic caching approach proposes to keep in a cache some previously executed queries along with their results and checks whether the answers to a given user query can be retrieved from the cache for optimization purposes. It uses the notion of query containment (i.e. a query Q is contained in a query G, if all answers to Q are also answers to G) to find the answers in the cache (see, e.g., [1, 7, 16, 19]). In a similar spirit, Ghosh *et al.* [14] propose a query clustering approach aimed at optimizing queries by reusing execution plans computed for similar queries.

On the other hand, the approach of query rewriting using views aims at finding view-based queries which are equivalent to (or contained in) a given user query. View-based query rewriting was first introduced using materialized views for query optimization purposes [6]. Afterwards, it was brought to the domain of data integration systems [2, 18] where the objective is to find certain answers to a query in a decentralized database context.

However, none of these approaches deals with the empty answer problem: they are concerned either in optimizing the access to information or in computing the set of certain answers to a query from distributed data sources. In our case, given a failing query, we are not interested in finding neither contained nor equivalent queries, since those would produce empty answers too.

In the information retrieval domain, other techniques are based on similarity measures and make use of previous executed queries to improve web search (see, e.g., [22, 23]). The measures underlying these approaches are strongly based on relationships between keywords, whereas we deal with a more general type of conditions than those expressed by a set of keywords, namely value constraints.

There are also relevant work in the domain of case-based reasoning, which integrates past user experience to solve current queries. In particular, Fouqué *et al.*[10] propose to include additional information in the answer to a user query, using an approach that they call associative query answering. The basic idea is to extend the "select" clause of a user query with attributes which appeared in similar queries previously submitted by other users. As we do, the authors use a nearness measure between every pair of values of each attribute, which is used to evaluate similarity between the user query and those previously executed. However, that work does not deal with failing queries.

In contrast, Bidault *et al.* [3] tackle the empty answer issue and use a repository of predefined queries in the context of mediation systems. Although their approach shows some similarity with ours, their solution is still quite different. They build a set of predefined successful queries for each source to offer a substitute for a failing user query. Similarity degrees between the initial user query and predefined successful queries are computed on the basis of a hierarchy of concepts. Finally, some heuristics make it possible to select the "best" substitute to the initial user query. In this approach, the similarity degree concerns the extent to which two concepts share

ISBN: 978-989-95079-6-8

the same characteristics whereas the substitutivity measure we defined is based on the proximity between the domain values of the attributes involved in the queries.

Let us also emphasize that none of the aforementioned approaches deals with failing fuzzy queries.

6 Conclusion

In this paper, we have outlined an approach aimed at obviating empty answers to Boolean or fuzzy queries involving value constraints. The method proposed uses a query repository and is based on the adaptation of the past non-failing query which is the most similar to the user query considered. Moreover, a technique aimed at providing the user with some explanations about the emptiness of the result of his/her original query has been briefly outlined.

The perspectives for future work are manifold. First, it would be useful to tackle the implementation of the query repository and to devise access methods for retrieving the candidate queries as efficiently as possible. The logical formalism proposed in [16] to represent information in cache and the index implementation presented in [10] to access it could be of interest for that purpose. Secondly, it would be worth investigating the possibility of reusing the execution plans of past queries to optimize the evaluation of the selected substitute query, in a spirit similar to [14]. Another point worthy to study concerns the substitutivity measure which is at the heart of the approach. The measure advocated here should be compared with some others adapted from classical similarity measures, cf. [9]. Finally, we intend to generalize this approach to a broader class of queries, e.g., those involving fuzzy joins.

References

[1] Y. Arens and C.A. Knoblock. Intelligent caching: selecting, representing and reusing data in an information server. In *Proc. of CIKM 1994*, pages 433–438, 1994.

[2] C. Beeri, A. Halevy, and M.-C. Rousset. Rewriting queries using views in description logics. In *Proc. of PODS'97*, pages 99–108, 1997.

[3] A. Bidault, C. Froidevaux, and B. Safar. Similarity between queries in a mediator. In *Proc. of ECAI 2002*, pages 235–239, 2002.

[4] P. Bosc, A. Hadjali, and O. Pivert. Weakening of fuzzy relational queries: an absolute proximity relation-based approach. *Mathware & Soft Computing*, 14(1):35–55, 2007.

[5] P. Bosc and O. Pivert. On the comparison of imprecise values in fuzzy databases. In *Proc. of the 6th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'97)*, pages 707–712, 1997.

[6] S. Chaudhuri, R. Krishnamurthy, S. Potamianos, and K. Shim. Optimizing queries with materialized views. In *Proc. of ICDE 1995*, pages 190–200, 1995.

[7] B. Chidlovskii and U.M. Borghoff. Semantic caching of web queries. *VLDB Journal*, 9:2–17, 2000.

[8] W. Chu, H. Yang, K. Chiang, M. Minock, G. Chow, and C. Larson. Cobase: A scalable and extensible cooperative information system. *Journal of Intelligent Information Systems*, 6(2-3):223–259, 1996.

[9] V.V. Cross and T.A. Sudkamp. *Similarity and Compatibility in Fuzzy Set Theory – Assessment and Applications*. Physica-Verlag, Heidelberg, 2002.

[10] G. Fouqué, W.W. Chu, and H. Yau. A case-based reasoning approach for associative query answering. In *Proc. of ISMIS'94*, pages 183–192, 1994.

[11] T. Gaasterland. Cooperative answering through controlled query relaxation. *IEEE Expert*, 12:48–59, 1997.

[12] T. Gaasterland, P. Godfrey, and J. Minker. An overview of cooperative answering. *Journal of Intelligent Information Systems*, 1(2):123–157, 1992.

[13] T. Gaasterland, P. Godfrey, and J. Minker. Relaxation as a platform for cooperative answering. *Journal of Intelligent Information Systems*, 1:293–321, 1992.

[14] A. Ghosh, J. Parikh, V.S. Sengar, and J.R. Haritsa. Plan selection based on query clustering. In *Proc. of VLDB 2002*, pages 179–190, 2002.

[15] P. Godfrey. Minimization in cooperative response to failing database queries. *Int. Journal of Cooperative Information Systems*, 6(2):95–149, 1997.

[16] P. Godfrey and J. Gryz. Answering queries by semantic caches. In *Proc. of DEXA 1999*, pages 485–498, 1999.

[17] S.Y. Huh, K.H. Moon, and H. Lee. A data abstraction approach for query relaxation. *Information and Software Technology*, 42:407–418, 2000.

[18] H. Jaudoin, J.-M. Petit, C. Rey, M. Schneider, and F. Toumani. Query rewriting using views in presence of value constraints. In *Proc. of the Int. Workshop on Description Logics (DL'05)*, pages 250–262, 2005.

[19] D.P. Miranker, M.C. Taylor, and A. Padmanaban. A tractable query cache by approximation. In *Proc. of SARA 2002*, pages 140–151, 2002.

[20] A. Motro. SEAVE: A mechanism for verifying user presuppositions in query systems. *ACM Trans. on Off. Inf. Syst.*, 4(4):312–330, 1986.

[21] I. Muslea. Machine learning for online query relaxation. In *Proc. of the Int. Conf. of Knowledge and Discovery and Data mining (KDD'04)*, pages 246–255, 2004.

[22] V. Raghavan and H. Sever. On the reuse of past optimal queries. In *Proc. of ACM SIGIR*, pages 344–350, 1995.

[23] J.R. Wen, J.-Y. Nie, and H.-J. Zhang. Query clustering using user logs. *ACM Transactions on Information Systems*, 20(1):59–81, 2002.

ISBN: 978-989-95079-6-8

A Fuzzy Delay Differential Equation Model for HIV Dynamics

Rosana M. Jafelice[1] Laécio C. Barros[2] Rodney C. Bassanezi[3]

1.Faculty of Mathematics, Federal University of Uberlândia
Uberlândia, Brazil
2.Department of Applied Mathematics, State University of Campinas
Campinas, Brazil
3. Center of Mathematics, Computation and Cognition, Federal University of ABC
Santo André, Brazil
Email: rmotta@ufu.br, laeciocb@ime.unicamp.br, rodney.bassanezi@ufabc.edu.br

Abstract— *This paper presents a process to obtain the solution (or flux) of a fuzzy delay system and to determine the fuzzy expected curve for the HIV (human immunodeficiency virus) when HIV-positive individuals receive antiretroviral therapy. This delay is defined as the time between the infection of a CD4+ type T-lymphocyte cell by the virus and the production of new virus particles. The intracellular delay is represented by an uncertainty parameter that depends on the individual characteristics of HIV-positive patients. A fuzzy rule-based system is used to model this parameter. The solution of the system of delay differential equations, which is a fuzzy process, is obtained from Zadeh's Extension Principle. Lastly, for each instant t, we calculate the fuzzy expected value obtained by the Center of Gravity.*

Keywords— Delay Differential Equation; Epidemiological Modeling; Fuzzy Expected Value; Fuzzy Set; HIV Dynamics; Mamdani Controller.

1 Introduction

HIV is an uncommon type of virus called a retrovirus and drugs developed to disrupt the action of HIV are known as antiretrovirals, or ARVs. The AIDS (Acquired Immunodeficiency Syndrome) virus mutates rapidly, which makes it extremely skillful at developing resistance to drugs. To minimize this risk, people with HIV are generally treated with a combination of ARVs that attack the virus on several fronts at once. The advent of ARVs in 1996 transformed the treatment of HIV and AIDS, improving the quality and greatly prolonging the lives of many infected people in places where the drugs are available. Of the estimated 6.5 million people in need of antiretroviral treatment in June 2006, 1.65 million people were reported to have had access to ARV treatment in low- and middle-income countries (World Health Organization, Jun 2006) [1]. In HIV-1 infection, treatment with reverse transcriptase or protease inhibitors results in a decline of free virus in several distinct phases. Fig. 1 summarizes the results of clinical studies, representing the different phases of viral decline after *in vivo* [2] treatment. Initially, the plasma viral load remains at approximately pretreatment levels, which, in the asymptomatic stage of the infection, are almost stationary on a time scale of weeks. This phase, which lasts from 0 to 1 day, is called the quasi-stationary initial stage. This initial phase is followed by a transition phase of 1 or 2 days, which is explained by the combination of pharmacological and intracellular delay effects that cause the disappearance of the free particles of the virus and the decline of infected cells. Another phase observed is the rapid decline of the plasma virus, which lasts from 2 to 7 days. Finally, the decline flattens out and virus levels may even rise again [2].

In this paper, we propose a process to determine the fuzzy flux of HIV decline as a function of time and, for each t, we determine the fuzzy expected value when the HIV-positive individual is receives antiretroviral therapy, considering the intracellular delay of the viral life cycle. Intracellular delay is defined as the time between the infection of a T lymphocyte cell of the $CD4+$ type by the virus and the production of new virus particles. The main lymphocyte that HIV attaches to when it reaches the bloodstream is the $CD4+$ type T lymphocyte. Based on the solution of a delay differential equation system [2], considering that the antiretroviral therapy is 100% effective, we introduced the delay τ as a fuzzy triangular number, where the degree of pertinence 1 is assumed to be $\tau = 12$h. Starting from a fuzzy rule-based system (FRBS), we determined the mortality rate of the virus as a function of delay. The fuzzy rule base was obtained from information given in [2]. With this fuzzy parameter [3], and by means of Zadeh's extension principle, we obtained a solution for the system. With this methodology applied to each instant t, the solution was a fuzzy set. Hence, we found the degree of pertinence for each solution of the deterministic system as a function of time. We then calculated the fuzzy expected value for the solutions of the system at each instant t, using the center of gravity. The curve of the expected value shows a profile similar to the curves of the clinical data when antiretroviral therapy is applied, as illustrated in Fig. 1.

2 Classic Models

In [2] was introduced two microscopic models for HIV infection dynamics in the individual organism undergoing antiretroviral therapy. These models are used in this paper.

The first is a model for viral dynamics with three variables: the population of uninfected cells, $x(t)$; the population of infected cells that produce the virus, $y(t)$, and the plasma virus population, $v(t)$. In a first approximation of the dynamics, we assumed that the influx λ and death rate d of uninfected cells are constant. Uninfected cells and free virus produce infected cells at a rate of $\beta(t)x(t)v(t)$. Infected cells produce free virus particles at a rate of $k(t)$ and die at a rate of a. Free virus

ISBN: 978-989-95079-6-8

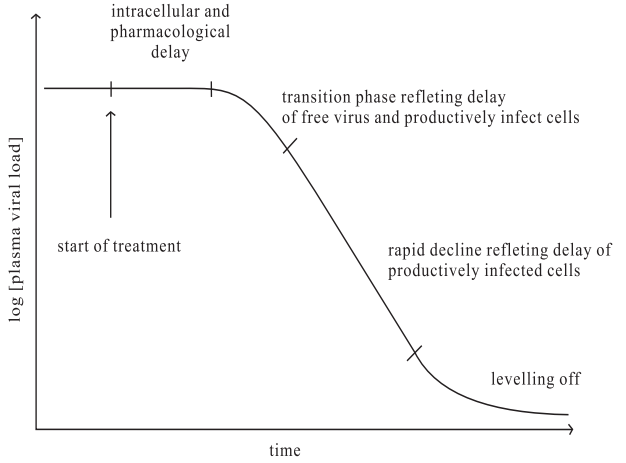


Figure 1: Schematic illustration of the different phases of plasma virus decline following antiretroviral therapy [2].

particles are cleared at a rate of u. To describe the effects of various drug therapies, the parameters $\beta(t)$ and $k(t)$ are time-dependent. Based on these assumptions, the differential equation system is given by:

$$\frac{dx(t)}{dt} = \lambda - dx(t) - \beta(t)x(t)v(t) \quad (1)$$

$$\frac{dy(t)}{dt} = \beta(t)x(t)v(t) - ay(t) \quad (2)$$

$$\frac{dv(t)}{dt} = k(t)y(t) - uv(t) \quad (3)$$

The model does not contain an intracellular time delay between infection of a cell and production of new virus particles. The model (4)-(6) incorporates the intracellular phase of the virus life cycle. In [2], it was assumed that virus production lags by a delay τ behind the infection of a cell. This implies that the recruitment of virus-producing cells, at time t, is given by the density of cells that were newly infected at time $t-\tau$ and are still alive at time t. Moreover, we assumed a constant death rate $\tilde{a}$ for infected but not yet virus-producing cells. The probability of survival from $t-\tau$ to time t is only $e^{-\tilde{a}\tau}$. More generally, the probability of survival is given by a nonincreasing function $f(\tau)$ with $0 \le f(\tau) \le 1$. Thus, (1), (2) and (3) can be written as:

$$\frac{dx(t)}{dt} = \lambda - dx(t) - \beta(t)x(t)v(t) \quad (4)$$

$$\frac{dy(t)}{dt} = \beta(t-\tau)x(t-\tau)v(t-\tau)e^{-\tilde{a}\tau} - ay(t) \quad (5)$$

$$\frac{dv(t)}{dt} = k(t)y(t) - uv(t) \quad (6)$$

The equation (5) has become a 'delay-differential equation τ'. Analytical solutions of this type of equation are generally difficult to find. However, for the problem in question, the populations of uninfected cells, infected virus-producing cells and free virus are at a steady-state level before treatment sets in [2]. This situation facilitates the mathematical analysis and enables us to derive simple analytical solutions. The nontrivial steady state of the system is given by:

$$x_0 = \frac{au}{\beta k}e^{\tilde{a}\tau}$$

$$y_0 = \frac{\lambda}{a}e^{\tilde{a}\tau} - \frac{du}{\beta k}$$

$$v_0 = \frac{ky_0}{u} \quad (7)$$

where β and k are constant pretreatment rates.

2.1 Protease Inhibitor Therapy

Reference [2] includes treatment with HIV protease inhibitors that block the production of new infectious virus v_I from already infected cells, allowing only noninfectious virus to be generated. As previously reported, the infectious virus declines but continues to infect cells [4] and [5].

According to [2], within the present framework, equation (6) also describes the dynamics of the total free virus. Infectious virus v_I, however, is not produced in $t > 0$ and declines according to $\frac{dv_I(t)}{dt} = -uv_I(t)$. Also, equations (4) and (5) remain valid if one replaces v with v_I.

In [2], it is assumed that the uninfected cell population remains constant, $x(t) = x_0$ within the time scale under consideration. For $x(t) = x_0$ and exponentially declining $v_I(t)$, equation (5) is solved by:

$$y(t) = \frac{y_0}{a-u}[ae^{-u(t-\tau)} - ue^{-a(t-\tau)}] \quad for \quad t > \tau. \quad (8)$$

From (6), the time evolution of free virus is then given by $v(t) = v_0$ for $0 < t \le \tau$ and

$$v(t) = v_0 e^{-u(t-\tau)} + \frac{uv_0}{a-u}\left\{\frac{u}{a-u}\left[e^{-a(t-\tau)} - e^{-u(t-\tau)}\right]\right\}$$

$$+\frac{uv_0}{a-u}\left\{a(t-\tau)e^{-u(t-\tau)}\right\} \quad for \quad t > \tau. \quad (9)$$

Combined reverse transcriptase inhibitor and protease inhibitor therapy may, in the long term, delay the evolution of drug-resistant virus strains [2].

After application of any antiviral drug, there is a short delay in the pharmacological effect due to the time required for drug absorption, distribution, and penetration into target cells. In the initial phase, the plasma virus load remains constant. The duration of this phase is the sum of the pharmacological delay [2], defined as the time span the drug requires to reach an effective concentration, and the intracellular delay, defined as the time between infection of a cell and production of new virus particles.

Thus, differences in plasma virus decline for a fixed half-life of infected cells ($a = 0.5$/day), with a delay of about 0.9 days, are shown in Fig. 2 for protease inhibitors. Intracellular delays were chosen to ensure that the long-term decline is identical in all cases. The curves differ mostly at the end of the shoulder phase, but even for strongly varying parameters, the differences are minor, see Fig. 3.

It is important to note that for different values of intracellular delay (τ) and rate of decline in the plasma virus concentration (u), the behavior of the plasma viral load (Fig. 3) is similar to that of the initial phases of antiretroviral treatment,

ISBN: 978-989-95079-6-8

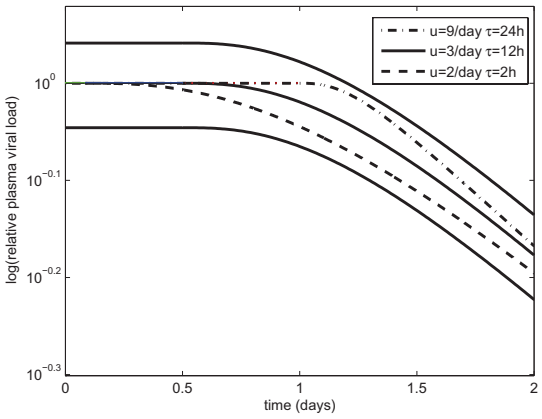


Figure 2: Sensitivity parameter of the protease inhibitor equation (9). Parameter a is fixed ($a = 0.5/day$), whereas u and τ vary. The two thin lines represent 10% deviations from the mean solution to illustrate expected measurement errors under ideal experimental conditions [2].

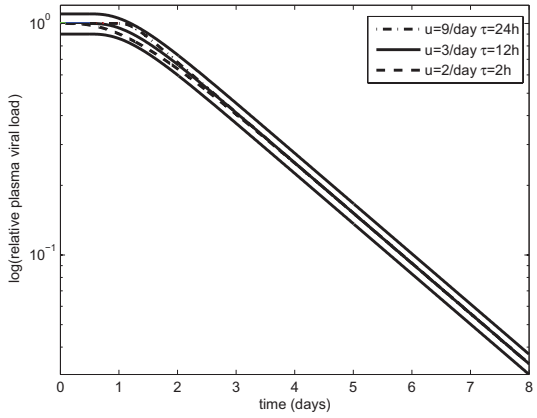


Figure 3: Range of HIV decline in response to protease inhibitor therapy [2].

as illustrated in Fig. 1, which summarizes the clinical studies, rendering the model credible (4)-(6). Reference [6] states that the length of delays is not directly measurable in vivo. Thus, additional in vitro data would prove valuable in expanding our knowledge of the characteristics of delay. Although these in vitro data are fraught with uncertainties, they clearly indicate that there is a delay between the initial infection and the production of new virus [7].

Reference [6] also states the need to explore the effects of different types of delay, such as distributed delays. Reference [7] assumes that the delay is given by a probability distribution. We propose a process to study the decline of the viral load after the beginning of protease inhibitor therapy. To this end, we consider that the intracellular and pharmacological delay is modeled by a fuzzy number, which takes into account a possibility distribution that can be given by a specialist.

The next section introduces the fuzzy model, which is the main object of this paper.

3 Fuzzy Model

Mathematical models for biological phenomena are normally fraught with uncertainties, in both the state variables and the parameters of the model's equations [8]. In our specific case, for reasons set forth in the introduction of this paper, we will consider that the delay time (τ) is uncertain and is modeled by the fuzzy set theory. Therefore, to study HIV decline over time, we developed a model which is a combination of classic delay differential equations and fuzzy logic.

A mathematical model in [4] was developed for protease inhibitor therapy, which fitted theoretically predicted curves of plasma virus decline to empirical data. The authors determined an average of about 0.9 days for the intracellular phase.

The next subsection presents a summary of the fuzzy set theory that was adopted in this work.

3.1 Fuzzy Sets

A *fuzzy subset* F of the universe set $\mathcal{U}$ is defined in terms of a function of *pertinence* u , which, at each element x of $\mathcal{U}$, associates a number $u(x)$ from *zero* and *one* called degree of pertinence of x to F. Thus, the fuzzy set F is symbolically indicated by its pertinence function

$$u_F : \mathcal{U} \rightarrow [0,1].$$

The values $u_F(x) = 1$ and $u_F(x) = 0$ indicate, respectively, the complete pertinence and the nonpertinence of element x to F.

A concept that will play a key role in this paper is fuzzy rule-based systems (FRBS) that have four components: an input processor, a collection of fuzzy rules called rule base, a fuzzy inference machine, and an output processor [9]. As the fuzzy inference machine we adopted the Mamdani Inference Method, while the output processor we used was the Center of Gravity.

Another concept used in this paper is Zadeh's Extension Principle.

Let X and Y be sets and f na application of X in Y. Let Abe a fuzzy set in X. The extension principle states that the image of A by the function f is a fuzzy set $B = f(A)$ in Y, whose pertinence function is given by

$$u_B(y) = \sup_x u_A(x) \tag{10}$$

for each $y \in Y$ with $x \in X$ and $y = f(x)$.

3.2 Methodology

Using an FBRS with the Mamdani Inference Method and center of gravity defuzzification, we obtained the defuzzified values of u as a function of τ. The least-square regression yielded $u(\tau)$. To represent the distributed delay, we adopted τ as a fuzzy triangular number. We then applied Zadeh's extension principle in two stages: first, to obtain the fuzzy u, and then to fuzzify the vital load. Finally, for each instant t, we calculated the fuzzy expected value by means of the Center of Gravity.

3.3 Fuzzy Rules

To determine the relationship between u and τ ($u = u(\tau)$), we adopted a FBRS, taking advantage of the fact that these variables are positively correlated [2].

ISBN: 978-989-95079-6-8

The intracellular delay (τ) and rate of decline in plasma virus concentration (u) are linguistic variables: the values of intracellular delay are expressed as $\{$*low*, *medium*, *high medium*, *high*, and *very high*$\}$, while the rate of decline in plasma virus concentration is expressed in the term set $\{$*low*, *medium*, *high medium*, *high*, *very high* $\}$.

Based on the values

- $u = 9/\text{day}$ and τ=0.08 day (approximately 2h);
- $u = 3/\text{day}$ and τ=0.5 day (12h);
- $u = 2/\text{day}$ and τ=1 day (24 h).

employed in the studies of [2], we adopted the domains of the pertinence functions of the intracellular delay (τ) and of the rate of decline in plasma virus concentration (u) as intervals that contain these values.

Following the positive correlation, the rule base that encodes the correlation between τ and u is given by:

- If τ is *low* then u is *low*.
- If τ is *medium* then u is *medium*.
- If τ is *high medium* then u is *high medium*.
- If τ is *high* then u is *high*.
- If τ is *very high* then u is *very high*.

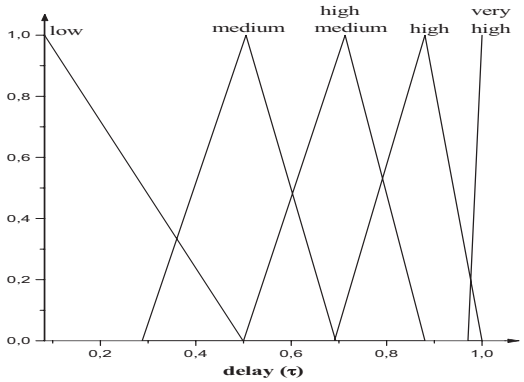


Figure 4: Membership functions for delay (τ).

The membership functions for each fuzzy set that specify the meaning of the linguistic variables for intracellular delay (τ) and rate of decline in plasma virus concentration (u) are shown in Fig. 4 and 5, respectively.

The data on virus mortality rate as a function of delay presented in Fig. 6 were obtained from the defuzzified values in the $[0.08, 1]$ interval of the FRBS. Using least-squares regression, we obtained the curve of the mortality rate of the virus as a function of delay, which is given by

$$u(\tau) = 1.67e^{1.63\tau} \tag{11}$$

with a determination coefficient (R^2) of 0.9873, as indicated in Fig. 6.

For biologic reasons, Mittler et al. [7] used a probabi-

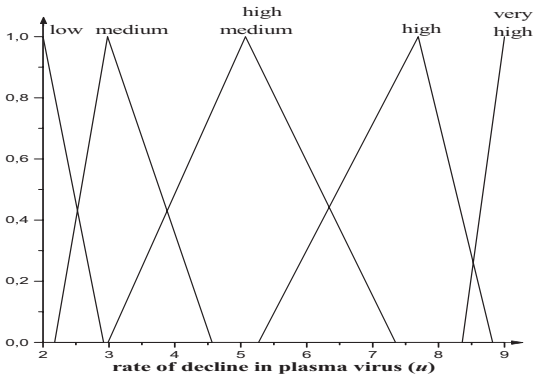


Figure 5: Membership functions for rate of decline in plasma virus concentration (u).

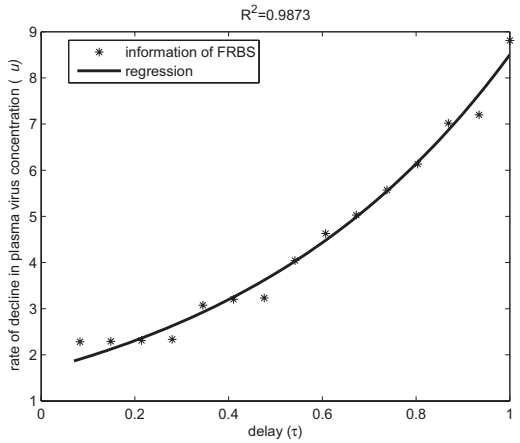


Figure 6: Regression of the rate of decline in plasma virus concentration

lity distribution function for τ. Here, in this paper, we will consider a possibility distribution instead a probability distribution function for τ. The rate of decline (u) will be obtained from a fuzzy rule-based system combined with the extension principle. Another method could be used for obtaining u. For example, the fuzzy systems given by gradual rules [10]. However, in this last case, we are not sure if these systems, in fact, approximate real functions.

3.4 Fuzzy Delay Distribution

We considered τ as a fuzzy parameter given by the fuzzy triangular number, whose support is [0.08,1] and maximum pertinence in 0.5, Fig.7.

The extension principle was used to obtain the image u of the fuzzy set τ, through function $u(\tau) = 1.67e^{1.63\tau}$. More specifically, from the formula:

$$\mu_{u(\tau)}(y) = \sup_{\{s:u(s)=y\}} \mu_{\tau}(s) \tag{12}$$

where τ is the fuzzy set whose membership function is u_τ, and $u(\tau)$ is the corresponding fuzzy set with membership function $\mu_{u(\tau)}$. Fig. 8 shows the extension principle for $u(\tau)$, assuming

ISBN: 978-989-95079-6-8

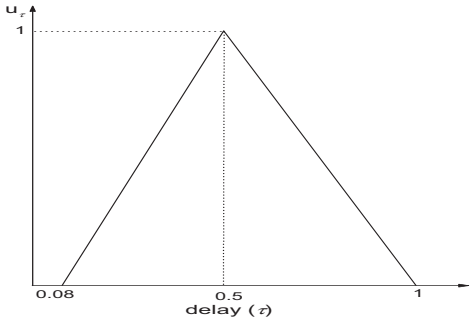


Figure 7: Fuzzy parameter (τ).

μ_τ, as in Fig. 7.

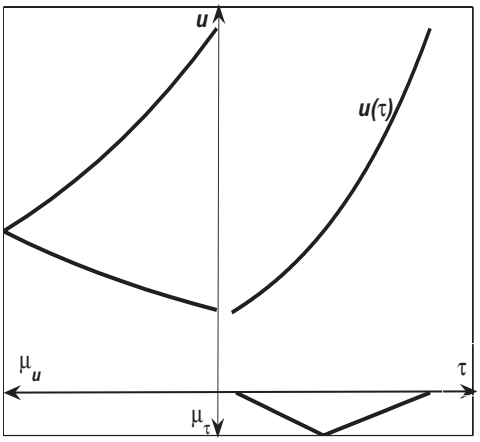


Figure 8: Computation μ_u.

3.5 *Fuzzy Solution*

The fuzzy solution for the viral load was obtained by fuzzification of the deterministic solution given by:
$v(t;\tau) = v_0$ for $0 < t \leq \tau$ and

$$v(t;\tau) = v_0 e^{-u(\tau)(t-\tau)} + \frac{u(\tau)v_0}{a-u(\tau)}\left\{\frac{u(\tau)}{a-u(\tau)}\left[e^{-a(t-\tau)} - e^{-u(\tau)(t-\tau)}\right]\right\}$$

$$+\frac{u(\tau)v_0}{a-u(\tau)}\left\{a(t-\tau)e^{-u(\tau)(t-\tau)}\right\} \quad for \quad t > \tau \qquad (13)$$

since we are admitting τ, and hence, u, as a fuzzy number.

At each instant t, Zadeh's Extension Principle is adopted for the function $v(t;\tau)$ applied in τ.

Fig. 9 presents the graphic solution of the fuzzy model. The light region shows the curves with the highest possibility of representing the phenomenon.

3.6 *Defuzzification of the Fuzzy Solution*

A common defuzzification scheme is the center of gravity method. For a fixed t, denote $v(t;\tau)$ by $v_t(\tau)$ to simplify the notation. Let $\mu_{v_t(\tau)}$ be the membership function of $v_t(\tau)$. Thus,

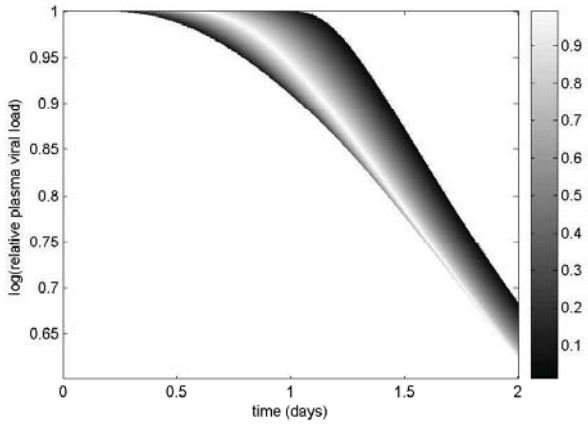


Figure 9: Solution of the fuzzy viral load for t varying from 0 to 2 days

a real-valued output $\bar{v}(t)$ is obtained at each time instant t, according to the formula:

$$\bar{v}(t) = \frac{\int\limits_{supp(v_t(\tau))} v_t(\tau)\mu_{v_t(\tau)}(v_t(\tau))dv_t(\tau)}{\int\limits_{supp(v_t(\tau))} \mu_{v_t(\tau)}(v_t(\tau))dv_t(\tau)} \qquad (14)$$

Given the fuzzy solution shown in Fig.9 and using the center of gravity, we obtained the defuzzified solution depicted in Fig.10.

The method generates a fuzzy expected curve which we have called a (defuzzified) solution $\bar{v}(t)$.

Reference [11] showed that this curve is the solution of the nonautonomous differential equation and may be viewed as the expected value of $v_t(\tau)$.

The solution $v(t;\tau)$ is obtained from a family of classic differential equations (Fig.11). However, it does not coincide with any solution for a fixed τ. For each t, $v(t;\tau)$ (14) is a value that belongs to the solution of the family of equations parameterized by $u(\tau)$. What differs here from the derived curve (14) suggested by the deterministic solutions is that, in deterministic solutions, all the uncertainties are excluded at the beginning (defuzzified at $t = 0$ and solved), whereas here the uncertainties evolve and defuzzification occurs at the time instant of interest (defuzzified when necessary) [11] and [12]. Hullermeier [13] suggests that the solution of a fuzzy differential equation is a fuzzy function obtained from a family of differential inclusions. Mizukoshi et al [14] have shown that when parameters (coefficients and/or initial conditions) are fuzzy, the Hullermeier solution is the same as that obtained from the extension principle. Therefore, we first obtained the solution of the classical differential equation and then used the extension principle to obtain fuzzy solutions. To obtain a real-valued trajectory, we adopted the center of gravity to defuzzify the fuzzy solution.

By means of equation (14), we obtained the fuzzy expected value $\bar{v}(t)$ at each instant t, which at first intercepts curves that present greater delays and, over time, intercepts curves with smaller delays, Fig. 11.

ISBN: 978-989-95079-6-8

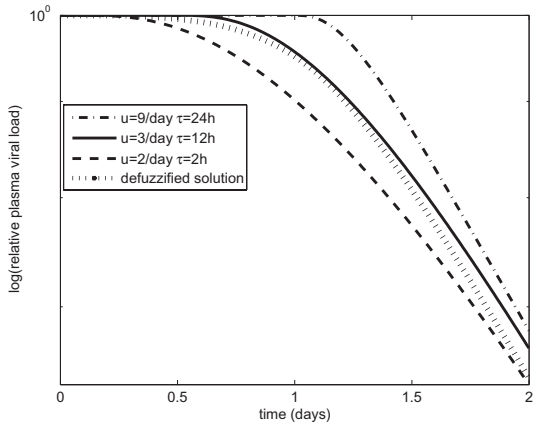


Figure 10: Defuzzified solution at each time instant t

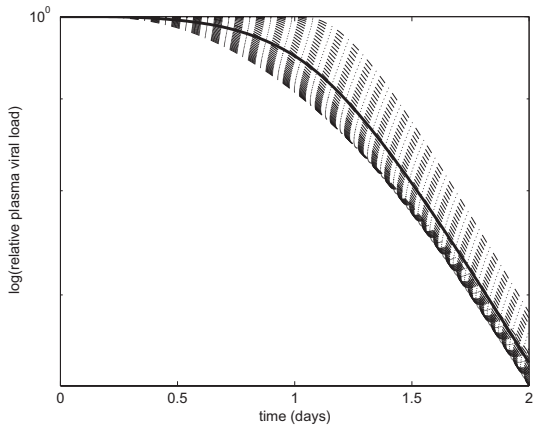


Figure 11: In the graph, the solid line represents the fuzzy expected value $\overline{v}(t)$.

4 Conclusions

In this work, emphasis was given to the study of the initial phase of treatment, since the plasma concentration of HIV declined by approximately 90% in the first two weeks of therapy due to the rapid elimination of free virus and the decline in infected cell productivity [15], see Fig. 10.

The work of [2] shows that intracellular delay may affect the magnitude of the range observed in Fig. 3, which represents variability or uncertainty. The adoption of distributed delay has yielded good results in the study of the dynamics of HIV under treatment [7]. In this paper, we adopted a fuzzy number to represent the distributed delay, which enabled us to obtain curves with different degrees of possibility to represent the phenomenon, Fig. 9. The degrees of pertinence of the curves come close to 1, as the curves approach the light region. This region is the one that best represents the phenomenon from the standpoint of credibility.

The defuzzified curve $\overline{v}(t)$ represents the amount of virus at each instant t. This curve is an average of the deterministic solutions given in (13), weighted by their respective degrees of possibility, which renders the curve more representative than each of the solutions of (13), adopting a value for τ.

Acknowledgements

The first author gratefully acknowledges CNPq, the Brazilian National Research Council, for granting a postdoctoral fellowship (152068/2007-4). The second author thanks FAPESP, the Research Foundation of the State of São Paulo, for grant no. 06/05920-7, and CNPq for grant no. 307890/2006-6.

References

[1] Joint United Nations Programme on HIV/AIDS. *www.unaids.org*, Acessed: June 18 2008.

[2] A.V.M. Herz, S. Bonhoeffer, R.M. Anderson, R.M. May, and M.A. Nowak. Viray dynamics *in vivo*: Limitations on estimates of intracellular delay and virus decay. In *Proc. Natl. Acad. Sci (Medical Sciences)*, pages 7247–7251, USA, 1996.

[3] R.M. Jafelice, L.C. Barros, R.C. Bassanezi, and F. Gomide. Fuzzy modeling in asymptomatic HIV virus infected population. *Bulletin of Mathematical Biology*, 66(6):1463–1942, 2004.

[4] A.S. Perelson, A.U. Neummann, M. Markowitz, J.M. Leonard, and D.D. Ho. HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span and viral generation time. *Science*, 271:1582–1586, 1996.

[5] D.L. Winslow and M.J. Otto. HIV protease inhitors. *AIDS 9 (Suppl.A)*, pages S183–S192, 1995.

[6] P.W. Nelson, J.D. Murray, and A.S. Perelson. A model of HIV-1 pathogenesis that includes an intracellular delay. *Mathematical Biosciences*, 163:201–215, 2000.

[7] J.E. Mittler, B. Sulzer, A.U. Neumann, and A.S. Perelson. Influence of delayed viral production on viral dynamics in HIV-1 infected patients. *Mathematical Biosciences*, 152:143–163, 1998.

[8] E. Massad, N.R.S. Ortega, L.C. Barros, and C.J. Struchiner. *Fuzzy Logic in Action: Applications in Epidemiology and Beyond*. Springer-Verlag Berlin, 2008.

[9] W. Pedrycz and F. Gomide. *An Introduction to Fuzzy Sets: Analysis and Design*. MIT Press, Cambridge, EUA, 1998.

[10] B. Bouchon-Meunier, D. Dubois, L. Godo, and H. Prade. *Fuzzy Sets and Possibility Theory in Approximate and Plausible Reasoning. Fuzzy Sets in Approximate Reasoning and Information Systems*. The Handbooks of Fuzzy Sets. Kluwer Academic Publishers, Kluwer, Boston, Mass., 1999.

[11] R.M. Jafelice, L.C. Barros, R.C. Bassanezi, and F. Gomide. Methodology to determine the evolution of asymptomatic HIV population using fuzzy set theory.

[12] L.C. Barros, R.C. Bassanezi, and M.B. Leite. The epidemiological models SI with fuzzy parameter of transmission. *Computers and Mathematics with Applications*, 45:1619–1628, 2003.

[13] E. Hüllermeier. An approach to modelling and simulation of uncertain dynamical systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5(2):117–137, 1997.

[14] M.T. Mizukoshi, L.C. Barros, Y. Chalco-Cano, H. Román-Flores, and R.C. Bassanezi. Fuzzy differential equation and the extension principle. *Information Sciences*, 177:3627–3635, 2007.

[15] A.S. Perelson, P. Essunger, Y. Cao, M. Vesanen, A. Hurley, K. Saksela, M. Markowitz, and D.D. Ho. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*, 387:188–191, 1997.

ISBN: 978-989-95079-6-8

Dispersal of Leaf-Cutting Ants: Fuzzy Mathematical Modeling, Numerical Approximation and Simulations

Rosana M. Jafelice, César G. Almeida[1] J.F.C.A. Meyer[2] Heraldo L. Vasconcelos[3]

1.Faculty of Mathematics, Federal University of Uberlândia
Uberlândia, Brazil
2.Department of Applied Mathematics, State University of Campinas
Campinas, Brazil
3.Institute of Biology, Federal University of Uberlândia
Uberlândia , Brazil
Email: {rmotta,cesargui}@ufu.br, joni@ime.unicamp.br, heraldo@umuarama.ufu.br

Abstract— *The objective of this paper is to propose a model for the occupation, by leaf-cutting ants, of an attractive area using a form of the diffusion-advection partial differential equation, in which the population dispersal is modeled using a fuzzy parameter. In the studied domain there is an attractive region, which can represent areas with an elevated concentration of palatable host-plants with better nutritious qualities. The developed algorithm uses information about the foraging behavior of a leaf-cutting ant colony from the Amazon region, in northern Brazil. The solution of the chosen partial differential equation is approximated using first order finite triangular elements and determines population dispersal with the use of a fuzzy rule-based system, which depends upon the number of individuals in the population and terrain characteristics. The numerical algorithm was developed in a MATLAB environment. The calculated solution is compatible with ant behavior when there is an attraction region nearby, as reported in the literature.*

Keywords— Fuzzy Set; Leaf-Cutting Ants; Mathematical Modeling; Numerical Methods; Partial Differential Equation;

1 Introduction

Ants of the genera *Atta* and *Acromyrmex* (*Hymenoptera: Formicidae*), collectively known as leaf-cutting ants, have the unique habit of culturing fungus on fresh plant materials [1]. These ants are generalist foragers that exploit a large number of plant species, although they usually focus on a subset of these species [2], especially those with low levels of toxic secondary compounds and high nutrient content [3, 4]. Their system of foraging trails influence the spatial organization of the foraging activity. Spatial and temporal heterogeneity in plant resources within the vicinity of the ants' nest produces changes in trail direction and modifications in the geometry of the foraging territories [5, 6]. Trails often lead to sites where, at a given time, plant resources are more attractive and abundant. For instance, a study in Costa Rica showed that, when leaf-cutting ant colonies have access to different plant communities, they concentrate their foraging efforts on the community with the greatest density of their preferred plant species (a monoculture of cassava) [7].

This paper proposes a model for the occupation of an attractive area by leaf-cutting ants, using a form of the diffusion-advection partial differential equation to model population dispersal [8] using a fuzzy parameter. The mathematical model for population dispersal is based on the balance of mass in phenomena of fluid dynamics and their physical principles. In this case, Fick's law; a conservation equation - such as conservation of mass -; and the notion of advection and diffusive flux arising from directed motion (convection) and effective random motion, respectively, are fundamental for the development and understanding of the model. In this work, the dispersion depends on the number of individuals of the population and on the characteristics of the terrain over which ants travel. This dispersion is determined by means of a system based on fuzzy rules. The model is constructed using expert entomological knowledge on behavior of leaf-cutting ants. The domain of this study contains an attractive region, which can represent areas with a high concentration of palatable host plants with good nutritious qualities. In literature, this research usually involve the use of a deterministic or stochastic model. However, mathematical literature on uncertainty has grown considerably over the last decade, especially in system modeling, optimization, control, and pattern recognition areas, to mention just a few. Recently, several authors have proposed the use of fuzzy set theory in epidemiology problems [9, 10, 11, 12] and population dynamics [13]. In this paper, we suggest fuzzy set theory, introduced in the 1960s by Lotfi Zadeh [14], to deal with the uncertain nature of ant population dispersal. In [15], the authors propose fuzzy rules that have input variables given by the difference (D) in the pheromone concentrations of the ants on the left and right trail branches. Based on these rules, the authors are able to determine the probability of choosing the left branch. The behavior of ants has also inspired several optimization techniques, the most exhaustively studied of which is known as ant colony optimization [16], which is also based on ant pheromone. The present study is based on the fact that the availability of plant resources to leaf-cutting ant colonies varies considerably in space and time. Thus, colonies must continually gather information about the current availability or stock of resources within their territories. Once a patch of palatable vegetation (i.e., an attractive area) is found, a chemically and physically marked trail is established and the ant foragers follow this trail in order to retrieve the newly found resource.

The next section presents information about the area where the research on ants was conducted [17].

ISBN: 978-989-95079-6-8

2 Ant Study Area in Amazonas

This study was carried out in a forest reserve located about 70 kilometers north of Manaus, state of Amazonas, Brazil. The vegetation is that of a primary inland rain forest, which, unlike other types of forest in the Amazon basin, is not subject to annual flooding. The climate is tropical and the annual precipitation, which is about 2,100 mm, is seasonal in distribution [18]. The rainy season lasts from November to May, and the dry season from June to October.

Vasconcelos's research [17] involved a study of the foraging activity of an *Atta cephalotes* colony from July 1985 to January 1986 and from September 1986 to March 1987. This study determined the spatial distribution of ant foraging trails and of the plant species exploited by the ants (i.e., their foraging sites).

Fig. 1 shows that the foraging sites of the *Atta cephalotes* were scattered over the entire extent of the trails, though mainly at intermediate distances from the nest (40 - 60m).

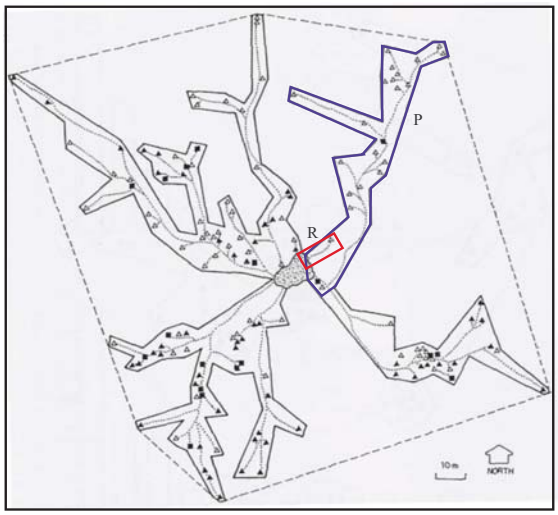


Figure 1: Foraging range (broken line) and foraging territory (continuous line) of the *Atta cephalotes* colony. Solid triangles represent the plants attacked between July 1985 and January 1986, and empty triangles those attacked between September 1986 and March 1987. Solid squares represent plants attacked during both periods. Dotted lines represent the foraging trails radiating from the nest (shaded area) [17].

The next section presents a partial differential equation with a fuzzy parameter for the occupation of leaf-cutting ants.

3 Fuzzy Model

The study of population dynamics involves several uncertain variables, such as the number of individuals in the population, the landscape, the speed at which the population travels, and population dispersal. Fuzzy set theory is a mathematical tool for modeling uncertain phenomena [19], for example, provides key notions for modeling epidemiological phenomena. Thus, the aim of this research is to find a numerical solution for modeling the presence of ant populations, treating dispersal as a fuzzy parameter [20].

The model proposed for the occupation by leaf-cutting ants will be studied by means of the diffusion-advection partial differential equation, given by:

$$\frac{\partial P}{\partial t} + v \cdot \nabla P - \nabla \cdot (\alpha(P, loc_{tot}) \nabla P) = 0. \qquad (1)$$

The functional variable $P = P(\mathbf{x}, t)$ indicates the population at the instant $t \in [0, T]$ and at the point $\mathbf{x} \in \Omega_1 \subset \mathbf{R}^2$ (see Fig. 2). The speed of occupation of the region v is a constant vector (v_1, v_2). Thus, we assume that the population dispersal is represented by parameter $\alpha(P, loc_{tot})$, where total locomotion (loc_{tot}) is determined by a fuzzy rule-based system that depends on the input variables loc_x (horizontal movement) and loc_y (vertical movement). This dispersal depends on the subregions of the domain, indicating the difficulty of locomotion of the ants over the terrain. The fuzzy parameters are modeled as follows:

- the domain is divided into subregions;
- the characteristics of the environment in which the individuals travel are represented by trapezoidal membership functions that indicate the degree of difficulty in locomotion through the domain;

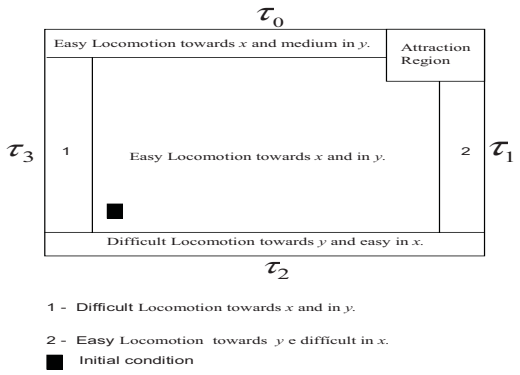


Figure 2: Domain Ω_1.

The adopted boundary conditions considered are:

$$\alpha(P, loc_{tot}) \frac{\partial P}{\partial y} = P v_2, \qquad \forall \mathbf{x} \in \tau_0, \qquad \forall t \in [0, T] \quad (2)$$

$$\alpha(P, loc_{tot}) \frac{\partial P}{\partial x} = P v_1, \qquad \forall \mathbf{x} \in \tau_1, \qquad \forall t \in [0, T] \quad (3)$$

$$\alpha(P, loc_{tot}) \frac{\partial P}{\partial y} = P v_2, \qquad \forall \mathbf{x} \in \tau_2, \qquad \forall t \in [0, T] \quad (4)$$

$$\alpha(P, loc_{tot}) \frac{\partial P}{\partial x} = P v_1, \qquad \forall \mathbf{x} \in \tau_3, \qquad \forall t \in [0, T] \quad (5)$$

The initial condition for this problem is given by:

$$P(\mathbf{x}, 0) = P_0(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega_1. \qquad (6)$$

To determine the numerical solution for (1), (2)-(5), the finite element method was used with a triangular uniform grid and linear functions [8]. The Crank-Nicolson Method was employed for the time discretization.

ISBN: 978-989-95079-6-8

3.1 Boundary Conditions

The spatial domain of the partial differential equation is rectangular: $\Omega_1 = [0, X] \times [0, Y]$, where X is the longitudinal length (axis x) and Y is the transversal length (axis y) of domain Ω_1. The domain boundary, $\partial\Omega_1$, is divided into parts which are indicated by: τ_0: the upper edge; τ_1: the right edge; τ_2: the lower edge and τ_3: the left edge. The unitary vector with an external direction normal to $\partial\Omega_1$ is given by: $n = (0, 1)$, in τ_0; in τ_1, $n = (1, 0)$, at the right edge; in τ_2, $n = (0, -1)$, at the lower edge; and $n = (-1, 0)$, in τ_3. In order to preserve the population mass as a function of time $M(t)$, the proper boundary conditions must be identified [**?**, 21]. Initially, the equation (1) is written as follows:

$$\frac{\partial P}{\partial t} + div(vP) - div(\alpha(P, loc_{tot})\nabla P) = 0. \quad (7)$$

Integrating equation (7) throughout the domain Ω_1, with $\int_{\Omega_1} Pd\mathbf{x} = \mathbf{M(t)}$ and using the Divergence Theorem, one has:

$$\frac{d}{dt}M(t) + \int_{\partial\Omega_1} (Pv \cdot n - \alpha(P, loc_{tot})\nabla P \cdot n)dS = 0 \quad (8)$$

To ensure that $M(t)$ is constant, one must have: $\frac{d}{dt}M(t) = 0$. Therefore:

$$(Pv \cdot n - \alpha(P, loc_{tot})\nabla P \cdot n) = 0. \quad (9)$$

Thus, one obtains the boundary conditions for the Ω_1, from (2) to (5).

3.2 Parameters for Computational Simulation

Fig. 2 represents domain Ω_1. To find the solution for equation (1) through first-order finite elements, this domain was subdivided into small triangles. The discrete solution will be obtained at the vertices (or nodes) of these triangles. The objective of computational simulations was to describe region R, given in Fig. 1. The solution for equation (1) was obtained from the finite triangular elements, using 3600 elements with 1860 nodes. The region of attraction has eight triangles and is located in the upper right-hand side of domain Ω_1, Fig. 2. The *A. cephalotes* colony is represented by four initial nodes close to boundaries τ_2 and τ_3, where an initial population of 400 ants was placed, Fig. 10. At each time step, two individuals are added to the calculated population, simulating ants present at the entrance of the ant nest at all times.

To calculate the dispersal several calculations were made based on Fig. 1. Initially, we measured the dotted trails in region P (large polygon) of Fig. 1. These trails were about 173 meters (m) in length, and the ants took approximately 40 days to occupy this area [17]. The dotted trail in region R (small rectangle) of Fig. 1 was $12m$ long. The rectangle of region R is approximately $13m$ by $5m$ and, to increase the performance of the computational program, let us consider that domain Ω_1 is $6m$ by $3m$. Because the computational program was implemented in a region whose sides measured half of the original region R, we considered that the trail was also half as long as the original one, i.e., $6m$. Therefore, we determined that the ants took 1.3873 days to occupy the $6m$ trail. According to the entomologist, ant trails have a width of approximately $0.15m$. Thus, the area occupied by the ants in the region for the implementation is approximately $4.7971m^2$. Since the dispersion (α) is determined by the area/time ratio, one has: $\alpha \cong 0.14m^2$/hour. We considered the speed in direction y to be equal to $0.07m$/hour. The speed in direction x was considered to be equal to $0.14m$/hour, i.e., double the speed in direction y.

3.3 Linguistic variables and rule base

Fuzzy sets are a way to represent imprecise information and knowledge. The values of locomotion along the x and y directions are expressed as {*constant*, *difficult*, *medium*, *easy*} while those of the population (P) are expressed in term sets {*very small*, *small*, *medium*, *large*, *very large*}. The membership functions that specify the meaning of the linguistic values are shown in Fig. 3, 4, 6, 7 and 8 for locomotion along the x direction, locomotion along the y direction, total locomotion, population and dispersal, respectively. Total locomotion is the output variable in the fuzzy rule-based system (FRBS 1) that depends on the variables of locomotion along the x and y directions. Dispersal is the output variable in the fuzzy rule-based system (FRBS 2) that depends on the variables of total locomotion and population number, Fig. 5. Population dispersal varies from 0 to $0.14m^2$/hours, according to the result of the calculations of this part of region R of Fig. 1. Total locomotion is expressed in a term set {*constant*, *very difficult*, *difficult*, *medium*, *easy*}. The membership functions that designate the meaning of the linguistic values are given in Figs. 6 and 8 for total locomotion and dispersal, respectively. The rule base that encodes the relationship of locomotion along the x and y directions; and total locomotion is summarized in Table 1. The rule base that encodes the relationship of total locomotion, population and dispersal is shown in Table 2.

All the rule bases are processed using Mamdani's Inference Method with center of gravity defuzzification [22]. There are other inference methods that could be adopted, such as Takagi-Sugeno's Inference Method [23]. Nevertheless, we could not use this method because the FRBS output variables are not precise and also they can not be written as input variables.

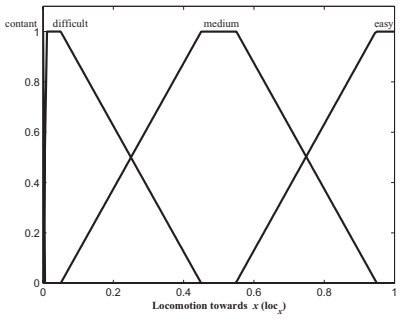


Figure 3: Membership functions for locomotion along the x direction (loc_x).

3.4 Fuzzy Computational Simulation

In order to define the type of ant locomotion as a function of the type of ground (e.g., rivers, hills, mountains and rocks)

ISBN: 978-989-95079-6-8

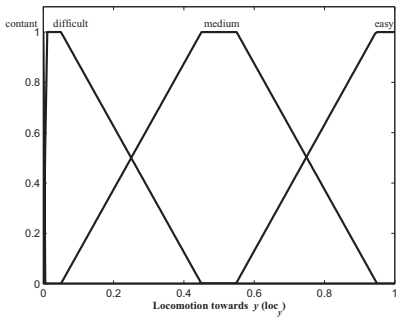


Figure 4: Membership functions for locomotion along the y direction (loc_y).

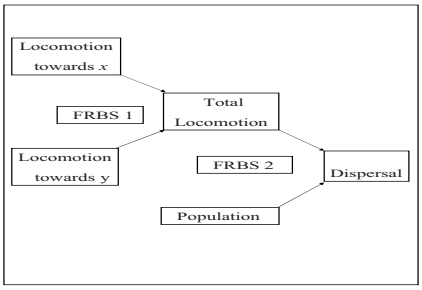


Figure 5: Fuzzy rule-based systems 1 and 2.

in the area of study to each node is allotted a value $[0, 1]$ for movement in x and y according to its location in the area, see Fig. 2.

The chosen trapezoidal membership function, which is associated to a fuzzy set A, $u_A(x)$, involving four parameters $[c_1\ c_2\ c_3\ c_4]$, see Fig. 9, is given by:

$$u_A(x) = \begin{cases} \dfrac{x - c_1}{c_2 - c_1} & \text{if } c_1 \leq x < c_2 \\ 1 & \text{if } c_2 \leq x \leq c_3 \\ \dfrac{-x + c_4}{c_4 - c_3} & \text{if } c_3 < x \leq c_4 \end{cases} \quad (10)$$

If the triangular elements are located in the area of difficult locomotion, the value for locomotion in those triangular elements is calculated by $c_1 + (c_4 - c_1) * rand$, where rand is equal to a random value in interval $(0, 1)$, $c_1 = 0$ e $c_4 = 0.45$. If the triangular elements are located in the area of medium locomotion, the value for locomotion in those triangular elements is calculated by $c_1 + (c_4 - c_1) * rand$, where rand is equal to a random value of $(0, 1)$, $c_1 = 0.05$ and $c_4 = 0.95$. If the triangular elements are located in the area of easy locomotion, the value for locomotion in those triangular elements is calculated by $c_1 + (1 - c_1) * rand$, since the values of x that represent the degree of membership between 0 and 1 for easy locomotion vary from $c_1 = 0.548$ to 1. Thus, locomotion in directions x and y in each triangle is determined randomly depending on the domain of the function of pertinence corresponding to the classification of the region of Fig.2. For horizontal and vertical locomotion, the triangular elements of the region of attraction are allotted the value of zero (0), which has a unitary fuzzy set. Therefore, each triangular element received a value for locomotion along x and y.

To obtain the fuzzy output for each triangular element, for FRBS 2, we use the average of population values on each of three nodes.

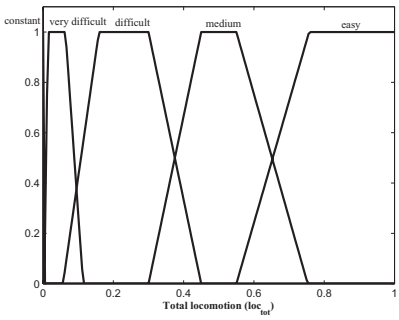


Figure 6: Membership functions for total locomotion (loc_{tot}).

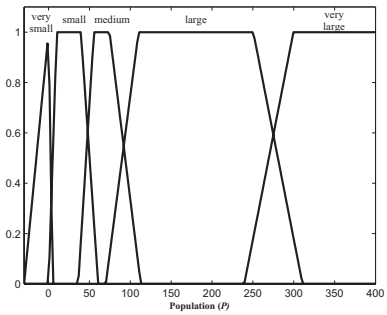


Figure 7: Membership functions for population (P).

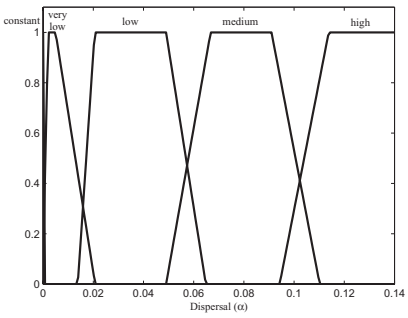


Figure 8: Membership functions for dispersal (α).

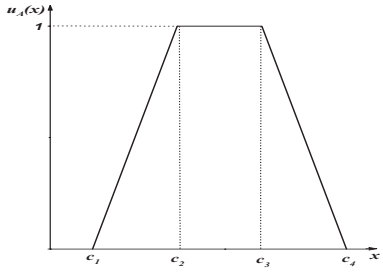


Figure 9: Parameters of trapezoidal membership functions.

ISBN: 978-989-95079-6-8

Table 1: Fuzzy rules for total locomotion (loc_{tot}).

(loc_x) \ (loc_y)	*constant*	*difficult*	*medium*	*easy*
constant	*constant*	*very difficult*	*difficult*	*medium*
difficult	*very difficult*	*difficult*	*difficult*	*medium*
medium	*difficult*	*difficult*	*medium*	*medium*
easy	*medium*	*medium*	*medium*	*medium*

Table 2: Fuzzy rules for dispersal (α).

(loc_{tot}) \ (P)	*very small*	*small*	*medium*	*large*	*very large*
constant	*constant*	*very low*	*very low*	*very low*	*very low*
very difficult	*constant*	*constant*	*very low*	*low*	*low*
difficult	*constant*	*very low*	*very low*	*medium*	*low*
medium	*constant*	*very low*	*very low*	*medium*	*medium*
easy	*constant*	*low*	*very low*	*medium*	*high*

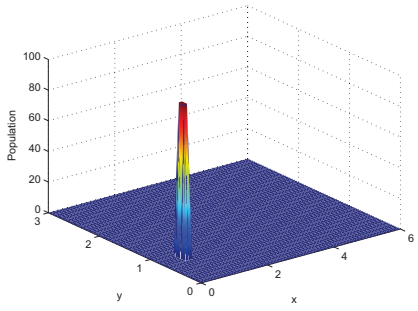


Figure 10: Initial population.

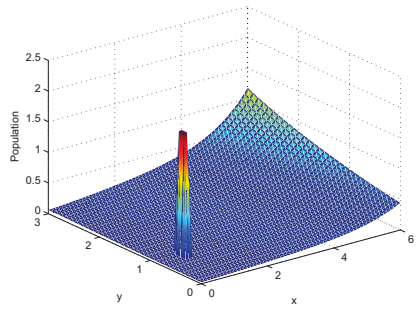


Figure 11: Population after 33 hours.

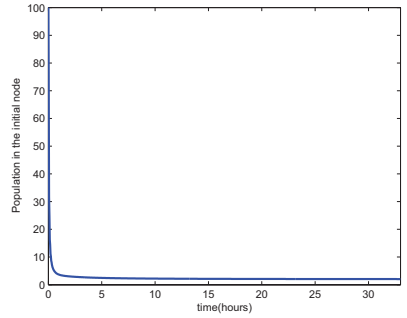


Figure 12: Population in the initial node.

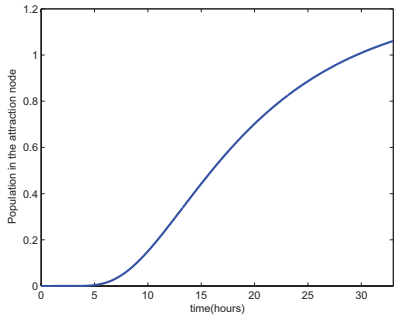


Figure 13: Population in one of the nodes of the region of attraction.

To determine the dispersal in each triangular element, the two FBRS were used: FRBS 1 and FRBS 2, Fig.5. Using FRBS 1, the value for total locomotion in each triangle was determined based on the values of locomotion along x and y. The dispersal in the node of each triangle was determined based on the values of total locomotion and medium population, using FRBS 2.

Thus, x and y components of locomotion were determined only once for each triangle of each region, prior to the first iteration, and total locomotion was determined based on the speed in directions x and y. The dispersion was calculated in each iteration for each triangle, since it depends on the total locomotion and on the population. The latter varies in each iteration.

Fig.10 illustrates the initial condition, i.e., the *A. cephalotes* colony before it occupied the region of attraction. The graph in Fig.11 shows the population after 33 hours, since ants take 33 hours to occupy the considered region of Fig.2. Fig.12 represents ant population in one of the initial nodes as a function of time. As it can be seen, the population in this node is close to two, since there will always be insects at the entrance to the ant nest. Fig. 13 shows the behavior of the number of ants as a function of time in one of the nodes of the region of attraction. Note that the population in this region grows logistically over time.

4 Conclusions

This paper proposes a model for the occupation of an attractive area by leaf-cutting ants, using a form of the diffusion-advection partial differential equation that models the population dispersal using a fuzzy parameter. Dispersion is this fuzzy parameter that depends on the number of the population and on its total locomotion, which is associated with the ease or difficulty of locomotion of the ants as a function of the terrain they traverse. The main difference between the classic models and the fuzzy model (1) is that the latter model exploits uncertainty parameters whereas the classic model does not. Thus, authors believe that, due to the uncertainty of biological phenomena, the combination of differential equations and fuzzy set theory enables computational simulations to much more faithfully portray the phenomenon under study.

Acknowledgement

The first author acknowledges CNPq, the Brazilian National Research Council, for its financial support in the form of a

ISBN: 978-989-95079-6-8

postdoctoral fellowship (152068/2007-4).

References

[1] B. Hölldobler and E.O. Wilson. *The Ants*. Belknap Press, Cambridge, MA, USA, 1990.

[2] H.L. Vasconcelos and H.G. Fowler. *Applied Myrmecology: a World Perspective*, chapter Foraging and fungal substrate selection by leaf-cutting ants, pages 410–419. Westview Press, Boulder, CO, USA, 1990.

[3] S.P. Hubbel and D.F. Wiemer. Host plant selection by an attine ant. In P. Jaisson, editor, *Social insects in the tropics*, volume 2, pages 133–154, Paris: University of Paris Press, 1983.

[4] F.M. Mundim, A.N. Costa, and H.L. Vasconcelos. Leaf nutrient content and host-plant selection by leaf-cutter ants, atta laevigata, in a neotropical savanna. *Entomologia Experimentalis et Applicata*, 130:47–54, 2009.

[5] H.G. Fowler and E.W. Stiles. Conservative resource management by leaf-cutting ants? the role of foraging territories and trails, and environmental patchiness. *Sociobiology*, 5:25–41, 1980.

[6] J.D. Shepherd. Trunk trails and searching strategy of a leaf-cutting ant, atta colombica. *Behavioural Ecology and Sociobiology*, 11:77–84, 1982.

[7] C.M. Blanton and J.J. Ewel. Leaf-cutting ant herbivory in successional and agricultural tropical ecosystems. *Ecology*, 66(3):861–869, 1985.

[8] R.S. Cantrell and C. Cosner. *Spatial Ecology via Reaction-Diffusion Equations*. John Wiley and Sons, Ltd, England, 2003.

[9] L.C. Barros, R.C. Bassanezi, and M.B. Leite. The epidemiological models SI with fuzzy parameter of transmission. *Computers and Mathematics with Applications*, 45:1619–1628, 2003.

[10] R.M. Jafelice, L.C. Barros, R.C. Bassanezi, and F. Gomide. Fuzzy modeling in asymptomatic HIV virus infected population. *Bulletin of Mathematical Biology*, 66(6):1597–1620, 2004.

[11] R.M. Jafelice, L.C. Barros, R.C. Bassanezi, and F. Gomide. Methodology to determine the evolution of asymptomatic HIV population using fuzzy set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 13:39–58, 2005.

[12] N. Ortega, L.C. Barros, and E. Massad. Fuzzy gradual rules in epidemiology. *Kybernetes: The International Journal of Systems and Cybernetics*, 32(4):460–477, 2003.

[13] V. Krivan and G. Colombo. A non-stochastic approach for modeling uncertainty in population dynamics. *Bulletin of Mathematical Biology*, 60, 1998.

[14] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[15] V. Rozin and M. Margaliot. The fuzzy ant. *Computational Intelligence Magazine, IEEE*, 2(4):18–28, 2007.

[16] M. Dorigo, M. Birattari, and T. Stützle. Ant colony optimization: Artificial ants as a computational intelligence technique. *IEEE Computational Intelligence Magazine*, pages 28–39, November 2006.

[17] H.L. Vasconcelos. Foraging activity of two species of leaf-cutting ants (atta) in a primary forest of the central amazon. *Insectes Sociaux*, 37(2):131–145, 1990.

[18] M.N.G. Ribeiro. Aspectos climatológicos de manaus. *Acta Amazonia*, 6:229–233, 1976.

[19] D. Dubois and H. Prade. *Fuzzy sets and systems - Theory and applications*. Academic press, New York, 1980.

[20] C.G. Almeida, R.S.M. Jafelice, and J.F.C.A. Meyer. Modelagem fuzzy em dinâmica populacional. In *Proccedings XXIX Congresso Nacional de Matemática Aplicada e Computacional*, Campinas, Brazil, 2006 (in Portuguese).

[21] L.C. Paraíba and P. Pulino. Simulação numérica da dispersão-advecção de pesticidas sob efeito da temperatura do solo: o modelo da peste. Relatório de Pesquisa 04, IMECC, UNICAMP, Campinas, Brazil, 2003 (in Portuguese).

[22] W. Pedrycz and F. Gomide. *An Introduction to Fuzzy Sets: Analysis and Design*. MIT Press, Cambridge, EUA, 1998.

[23] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. on Systems, Man, and Cybernetics*, 15(1):116–132, 1985.

ISBN: 978-989-95079-6-8

Continuous R-implications

Balasubramaniam Jayaram[1] Michał Baczyński[2]

1. Department of Mathematics, Indian Institute of Technology Madras, Chennai 600 036, India
2. Institute of Mathematics, University of Silesia, 40-007 Katowice, ul. Bankowa 14, Poland
Email: jbala@iitm.ac.in, michal.baczynski@us.edu.pl

Abstract*— *In this work we have solved an open problem related to the continuity of R-implications. We have fully characterized the class of continuous R-implications obtained from any arbitrary t-norm. Using this result, we also determine the exact intersection between the continuous subsets of R-implications and (S,N)-implications.

***Keywords*—** R-implication, (S,N)-implication, Łukasiewicz implication, t-norm.

1 Introduction

R-implications and (S,N)-implications are two of the most established families of fuzzy implications. Still, many open problems remain unsolved, see [3, 4]. One of them is related to the continuous subsets of these families. Only recently a characterization of continuous (S,N)-implications was given by the authors in [2].

However, a similar complete characterization regarding the continuous subset of R-implications has not been available so far. It is only known that in the class of R-implications obtained from *left-continuous* t-norms, the only continuous elements are those that are isomorphic to the Łukasiewicz implication, i.e., those R-implications obtained as residuals of nilpotent t-norms. In particular, the following question has remained unanswered so far:

> Does there exist a continuous R-implication obtained from a non-left continuous t-norm?

In this note we show that an R-implication I_T obtained from a t-norm T is continuous if and only if T is a nilpotent t-norm.

Using this result, we are also able to resolve another question related to the intersections between these two families, which is also a generalization of an original result of Smets and Magrez [13], see also [7, 9]. We show that the only continuous (S,N)-implication that is also an R-implication obtained from any t-norm, not necessarily left-continuous, is the Łukasiewicz implication up to an isomorphism.

2 Preliminaries

We assume that the reader is familiar with the classical results concerning basic fuzzy logic connectives, but to make this work more self-contained, we introduce basic notations used in the text and we briefly mention some of the concepts and results employed in the rest of the work.

Definition 1 (cf. [7, 10, 9])**.**

(i) A function $N\colon [0,1] \to [0,1]$ is called a *fuzzy negation*, if it is decreasing and satisfies the boundary conditions $N(1) = 0$ and $N(0) = 1$.

(ii) A fuzzy negation N is called *strong*, if it is an involution, i.e., $N \circ N = \mathrm{id}_{[0,1]}$.

(iii) A function $T\colon [0,1]^2 \to [0,1]$ is called a *t-norm*, if it is increasing in both variables, commutative, associative and has 1 as the neutral element.

(iv) A function $S\colon [0,1]^2 \to [0,1]$ is called a *t-conorm*, if it is increasing in both variables, commutative, associative and has 0 as the neutral element.

(v) A t-norm T is said to be *border continuous*, if it is continuous on the boundary of the unit square $[0,1]^2$, i.e., on the set $[0,1]^2 \setminus (0,1)^2$.

(vi) A t-norm T is said to be *left-continuous*, if it is left-continuous in each component.

(vii) A t-norm T is said to be *nilpotent*, if it is continuous and if each $x \in (0,1)$ is a nilpotent element of T, i.e., if there exists $n \in \mathbb{N}$ such that $x_T^{[n]} = 0$, where

$$x_T^{[n]} := \begin{cases} 1, & \text{if } n = 0, \\ x, & \text{if } n = 1, \\ T(x, x_T^{[n-1]}), & \text{if } n > 1. \end{cases}$$

(viii) A t-norm T is said to be *Archimedean* if for every $x, y \in (0,1)$ there exists $n \in \mathbb{N}$ such that $x_T^{[n]} < y$.

Remark 1 (see. [10, p. 17])**.** For the border continuity of a t-norm T, it is sufficient to require the continuity on the upper right boundary, since from the monotonicity we get

$$\lim_{x\to 0^+} T(x,y) \le \lim_{x\to 0^+} T(x,1) = \lim_{x\to 0^+} x = 0 = T(0,y),$$

for any $y \in [0,1]$.

Remark 2. From the commutativity, the left-continuity of a t-norm T is equivalent to the left-continuity of T with respect to the first or the second variable. Moreover, $T(x,1) = 1$ and $T(x,0) = 0$ for every $x \in [0,1]$, thus a t-norm T is left-continuous if and only if for any $y \in (0,1)$ and every increasing sequence $(x_n)_{n\in\mathbb{N}}$, where $x_n \in [0,1)$, we have

$$\lim_{n\to\infty} T(x_n, y) = T(\lim_{n\to\infty} x_n, y).$$

Proposition 1. *If T is an Archimedean t-norm, then $T(x,y) < \min(x,y)$, for all $x, y \in (0,1)$.*

ISBN: 978-989-95079-6-8

Proof. Let T be an Archimedean t-norm. If, on the contrary, there exist some $x_0, y_0 \in (0,1)$ such that $x_0 \geq y_0$ and $T(x_0, y_0) = y_0 = \min(x_0, y_0)$, then we will prove, by induction, that for every $n \in \mathbb{N}$ we have

$$x_0{}_T^{[n]} \geq y_0. \tag{1}$$

Indeed, firstly see that

$$x_0{}_T^{[0]} = 1 > T(x_0, y_0)) = y_0,$$
$$x_0{}_T^{[1]} = x_0 \geq T(x_0, y_0)) = y_0.$$

Let us assume that (1) holds for some $n \in \mathbb{N}$. Then by the monotonicity of T and our inductive assumption we get

$$x_0{}_T^{[n+1]} = T(x_0, x_0{}_T^{[n]}) \geq T(x_0, y_0) = y_0,$$

which implies that T is not Archimedean, a contradiction. □

By Φ we denote the family of all increasing bijections $\varphi\colon [0,1] \to [0,1]$. We say that two functions $f, g\colon [0,1]^n \to [0,1]$, where $n \in \mathbb{N}$, are Φ-conjugate, if there exists $\varphi \in \Phi$ such that $g = f_\varphi$, where

$$f_\varphi(x_1, \ldots, x_n) := \varphi^{-1}\left(f(\varphi(x_1), \ldots, \varphi(x_n))\right),$$

for all $x_1, \ldots, x_n \in [0,1]$. Equivalently, g is said to be the Φ-conjugate of f.

Definition 2 ([7, 4])**.** A function $I\colon [0,1]^2 \to [0,1]$ is called a *fuzzy implication* if it satisfies the following conditions:

I is decreasing in the first variable, (I1)

I is increasing in the second variable, (I2)

$I(0,0) = 1, \qquad I(1,1) = 1, \qquad I(1,0) = 0.$ (I3)

The set of all fuzzy implications will be denoted by $\mathcal{FI}$.

3 R-implications

Definition 3 (cf. [15, 7, 9])**.** A function $I\colon [0,1]^2 \to [0,1]$ is called an *R-implication*, if there exists a t-norm T such that

$$I(x,y) = \sup\{t \in [0,1] \mid T(x,t) \leq y\}, \tag{2}$$

for all $x, y \in [0,1]$. If an R-implication is generated from a t-norm T, then we will often denote this by I_T.

It is important to note that the name 'R-implication' is a short version of 'residual implication', and I_T is also called as the residuum of T (see e.g. [7, 9, 10]).

Example 1. The Łukasiewicz implication

$$I_{\mathbf{LK}}(x,y) = \min(1, 1 - x + y), \qquad x, y \in [0,1],$$

is an R-implication obtained from the nilpotent (Łukasiewicz) t-norm

$$T_{\mathbf{LK}}(x,y) = \max(x + y - 1, 0), \qquad x, y \in [0,1].$$

For more well-known R-implications along with their t-norms from which they have been obtained, we refer the readers to other sources, notably [7, 10, 4].

Theorem 1 (cf. [7], [3, Theorem 5.5])**.** *If T is any t-norm, then $I_T \in \mathcal{FI}$ and it satisfies the left neutrality property, i.e.,*

$$I_T(1,y) = y, \qquad y \in [0,1], \tag{NP}$$

and the identity principle, i.e.,

$$I_T(x,x) = 1, \qquad x \in [0,1]. \tag{IP}$$

Moreover, if T is left-continuous, then I_T satisfies the exchange principle, i.e.,

$$I_T(x, I_T(y,z)) = I_T(y, I_T(x,z)), \qquad x, y, z \in [0,1],$$

and the ordering property, i.e.,

$$x \leq y \iff I_T(x,y) = 1, \qquad x, y \in [0,1].$$

Proposition 2 ([3, Proposition 5.8])**.** *For a t-norm T the following statements are equivalent:*

(i) T is border continuous.

(ii) I_T satisfies the ordering property.

For R-implications generated form left-continuous t-norms we have the following results.

Proposition 3 (cf. [9, Proposition 5.4.2 and Corollary 5.4.1])**.** *For a t-norm T the following statements are equivalent:*

(i) T is left-continuous.

(ii) T and I_T form an adjoint pair, i.e., they satisfy the residuation property

$$T(x,t) \leq y \iff I_T(x,y) \geq t, \qquad x, y, t \in [0,1].$$

(iii) The supremum in (2) *is the maximum, i.e.,*

$$I_T(x,y) = \max\{t \in [0,1] \mid T(x,t) \leq y\},$$

where the right side exists for all $x, y \in [0,1]$.

Using the above result we are able to obtain the following characterization.

Theorem 2 (cf. [8, Corollary 2])**.** *For a function $I\colon [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i) I is a continuous R-implication based on some left-continuous t-norm.

(ii) I is Φ-conjugate with the Łukasiewicz implication, i.e., there exists $\varphi \in \Phi$, which is uniquely determined, such that

$$I(x,y) = \varphi^{-1}(\min(1 - \varphi(x) + \varphi(y), 1)), \tag{3}$$

for all $x, y \in [0,1]$.

For more facts related to R-implication see e.g. [7, 3, 4].

ISBN: 978-989-95079-6-8

4 Continuous Partial Functions of R-implications

Note that from Theorem 1 we can consider, for any fixed $\alpha \in [0,1)$, the decreasing partial function $I_T(\cdot,\alpha)\colon [\alpha,1] \to [\alpha,1]$, which we will denote by g_α^T. Observe that g_α^T is decreasing and such that $g_\alpha^T(\alpha) = 1$ and $g_\alpha^T(1) = \alpha$.

Remark 3. If the domain of g_α^T is extended to $[0,1]$, then this is exactly what are called contour lines by Maes and De Baets in [11, 5]. If $\alpha = 0$, then g_0^T is the natural negation associated with the t-norm T (see [3]).

Theorem 3. *Let T be any t-norm and $\alpha \in [0,1)$. If g_α^T is continuous, then g_α^T is strictly decreasing.*

Proof. Let T be any t-norm and $\alpha \in [0,1)$ be fixed. We know that g_α^T is decreasing. On the contrary, let us assume that g_α^T is constant on some interval $[x_0, y_0]$ for some $\alpha < x_0 < y_0 < 1$, i.e., there exists $p \in [\alpha, 1]$ such that

$$g_\alpha^T(x_0) = g_\alpha^T(y_0) = p.$$

Let us fix arbitrarily $z \in (x_0, y_0)$.

Firstly, consider the case $p = 1$. Then

$$g_\alpha^T(z) = I_T(z,\alpha) = \sup\{t \in [0,1] \mid T(z,t) \le \alpha\} = 1,$$

thus $T(z, 1-\varepsilon) \le \alpha$ for any $\varepsilon \in (0,1)$. Hence

$$g_\alpha^T(1-\varepsilon) = \sup\{t \in [0,1] \mid T(1-\varepsilon, t) \le \alpha\} \ge z,$$

for all $\varepsilon \in (0, 1-\alpha)$. However, by the continuity of g_α^T, as $\varepsilon \to 0^+$, we get

$$\alpha = g_\alpha^T(1) = g_\alpha^T(\lim_{\varepsilon\to 0^+} 1-\varepsilon) = \lim_{\varepsilon\to 0^+} g_\alpha^T(1-\varepsilon) \ge \lim_{\varepsilon\to 0^+} z = z,$$

a contradiction to the fact that $\alpha < x_0 < z$.

If $p = \alpha$, then

$$g_\alpha^T(z) = I_T(z,\alpha) = \sup\{t \in [0,1] \mid T(z,t) \le \alpha\} = \alpha,$$

thus $T(z, \alpha+\varepsilon) > \alpha$ for all $\varepsilon \in (0, 1-\alpha)$. Hence

$$g_\alpha^T(\alpha+\varepsilon) = \sup\{t \in [0,1] \mid T(\alpha+\varepsilon, t) \le \alpha\} \le z,$$

for all $\varepsilon \in (0, 1-\alpha)$. Once again, by the continuity of g_α^T we have, as $\varepsilon \to 0^+$, that

$$1 = g_\alpha^T(\alpha) = g_\alpha^T(\lim_{\varepsilon\to 0^+} \alpha+\varepsilon) = \lim_{\varepsilon\to 0^+} g_\alpha^T(\alpha+\varepsilon) \le \lim_{\varepsilon\to 0^+} z = z,$$

a contradiction to the fact that $z < 1$.

Finally, let $p \in (\alpha, 1)$. Then, by the definition of g_α^T, we have

$$T(z, p+\varepsilon) > \alpha \ge T(z, p-\varepsilon),$$

for any $\varepsilon > 0$ such that $p + \varepsilon \le 1$ and $p - \varepsilon \ge \alpha$. Therefore

$$I_T(p+\varepsilon, \alpha) \le z \le I_T(p-\varepsilon, \alpha),$$

hence

$$g_\alpha^T(p+\varepsilon) \le z \le g_\alpha^T(p-\varepsilon).$$

Since g_α^T is continuous, we have, as $\varepsilon \to 0^+$, that

$$g_\alpha^T(p) = z.$$

Now this happens for every $z \in (x, y)$, which contradicts the fact that g_α^T is a function itself. Hence g_α^T is strictly decreasing. □

5 Main results: continuous R-implications

The main result of this work is the generalization of Theorem 2, viz., we show that the *left-continuity* of the underlying t-norm is implied and need not be assumed. Thus we give a complete characterization of the class of all continuous R-implications by showing that it is equivalent to the class of fuzzy implications which are Φ-conjugate to the Łukasiewicz implication.

Theorem 4. *Let T be a t-norm and I_T the R-implication obtained from it. If I_T is continuous, then T is border continuous.*

Proof. On the contrary, let us assume that I_T is continuous and T is not border continuous. Then, by Remark 1, there exist $y_0 \in (0,1)$ and an increasing sequence $(x_n)_{n\in\mathbb{N}}$, where $x_n \in [0,1)$, such that $\lim_{n\to\infty} x_n = 1$, but

$$\lim_{n\to\infty} T(x_n, y_0) = y' < y_0 = T(1, y_0).$$

This implies, in particular, that

$$I_T(y_0, y') = \sup\{t \in [0,1] \mid T(y_0, t) \le y'\} = 1.$$

Now, by (I1) and (IP) of I_T (cf. Theorem 1) we have that

$$1 = I_T(y_0, y') \le I_T(y', y') = 1,$$

i.e., $I_T(x, y') = 1$ for all $x \in [y', y_0]$. Note that $I_T(\cdot, y') = g_{y'}^T$. Since I_T is continuous we have that $g_{y'}^T$ is also continuous and from Theorem 3 we see that it is strictly decreasing. However, from the above, we see that $g_{y'}^T$ is constant on $[y', y_0]$, a contradiction. Thus T is border continuous. □

Theorem 5. *Let T be a t-norm and I_T the R-implication obtained from it. If I_T is continuous, then T is Archimedean.*

Proof. Let T be a t-norm. On the contrary, let us assume that I_T is continuous and T is non-Archimedean. Then, by the Definition 1(viii) there exist $x_0, y_0 \in (0,1)$ such that for all $n \in \mathbb{N}$ we have that $x_0{}_T^{[n]} \ge y_0$.

Let us denote

$$X_0 := \{z \in [0,1] \mid x_0{}_T^{[n]} > z \text{ for all } n \in \mathbb{N}\}.$$

Observe, that $X_0 \neq \emptyset$ since for all $y < y_0$ we have that $x_0{}_T^{[n]} > y$ for all $n \in \mathbb{N}$. Further, let

$$z_0 := \sup X_0.$$

See that $0 < z_0 \le x_0$ and $z_0 - \varepsilon \in X_0$ for all $\varepsilon \in (0, z_0]$. Also, if $t > z_0$, then there exists $m \in \mathbb{N}$ such that $x_0{}_T^{[m]} \le t$, which implies that

$$z_0 - \varepsilon < x_0{}_T^{[m+1]} = T(x_0, x_0{}_T^{[m]}) \le T(x_0, t),$$

for any $t > z_0$. Hence

$$I_T(x_0, z_0-\varepsilon) = \sup\{t \in [0,1] \mid T(x_0, t) \le z_0 - \varepsilon\} \le z_0,$$

for all $\varepsilon \in (0, z_0]$. From the continuity of I_T we get

$$I_T(x_0, z_0) = I_T(x_0, \lim_{\varepsilon\to 0^+} z_0 - \varepsilon) = \lim_{\varepsilon\to 0^+} I_T(x_0, z_0-\varepsilon) \le \lim_{\varepsilon\to 0^+} z_0 = z_0.$$

ISBN: 978-989-95079-6-8

Now, by (I1) and (NP) of I_T (cf. Theorem 1) we have that

$$z_0 \geq I_T(x_0, z_0) \geq I_T(1, z_0) = z_0,$$

i.e., $I_T(x, z_0) = z_0$ for all $x \in [x_0, 1]$. Note that $I_T(\cdot, z_0) = g^T_{z_0}$. Since I_T is continuous we have that $g^T_{z_0}$ is also continuous and from Theorem 3 we see that it is strictly decreasing. However, from the above, we have that $g^T_{z_0}$ is constant on $[x_0, 1]$, a contradiction. Thus T is Archimedean. □

Theorem 6. *Let T be a t-norm and I_T the R-implication obtained from it. If I_T is continuous, then T is left-continuous.*

Proof. Let T be a t-norm such that I_T is continuous. From Theorems 4 and 5 we see that T is border continuous and Archimedean.

On the contrary, let us assume that T is non-left-continuous. From Remark 2 there exist $x_0 \in (0, 1]$, $y_0 \in (0, 1)$ and an increasing sequence $(x_n)_{n\in\mathbb{N}}$, where $x_n \in [0, 1)$, such that $\lim_{n\to\infty} x_n = x_0$, but

$$\lim_{n\to\infty} T(x_n, y_0) = z' < z_0 = T(x_0, y_0).$$

Since T is border continuous it suffices to consider the case when $x_0 \in (0, 1)$.

Firstly observe that

$$I_T(y_0, z') = \sup\{t \in [0, 1] \mid T(y_0, t) \leq z'\} = x_0, \qquad (4)$$

since from the monotonicity of T we have $T(y_0, x_n) \leq z'$ for every $n \in \mathbb{N}$ and $T(y_0, x_0) = z_0 > z'$.

Next, from Proposition 1, by the Archimedeanness and monotonicity of T, we get that for any arbitrary $\varepsilon \in (0, 1-x_0)$ the following inequality holds

$$T(x_0, 1 - \varepsilon) < \min(x_0, 1 - \varepsilon) = x_0. \qquad (5)$$

Now, by (4) and (5) we get

$$T(x_0, 1 - \varepsilon) < I_T(y_0, z'),$$

for any $\varepsilon \in (0, 1 - x_0)$, thus

$$T(x_0, 1 - \varepsilon) < \sup\{t \in [0, 1] \mid T(y_0, t) \leq z'\},$$

hence

$$T(y_0, T(x_0, 1 - \varepsilon)) \leq z'$$

By the associativity of T we get

$$T(T(x_0, y_0), 1 - \varepsilon) \leq z',$$

i.e.,

$$T(z_0, 1 - \varepsilon) \leq z'.$$

for any $\varepsilon \in (0, 1 - x_0)$. This implies that

$$\lim_{\varepsilon\to 0^+} T(z_0, 1 - \varepsilon) \leq z' < z_0 = T(z_0, 1),$$

i.e., T is not border continuous, a contradiction to Theorem 4 and hence T is left-continuous. □

From Theorems 2 and 6 we obtain the following result.

Corollary 1. *For a function $I\colon [0, 1]^2 \to [0, 1]$ the following statements are equivalent:*

(i) I is a continuous R-implication based on some t-norm.

(ii) I is Φ-conjugate with the Łukasiewicz implication, i.e., there exists $\varphi \in \Phi$, which is uniquely determined, such that I has the form (3) *for all $x, y \in [0, 1]$.*

Let us denote the different families of fuzzy implications as follows:

- $\mathbb{I}_{\mathbb{T}}$ – the family of all R-implications;
- $\mathbb{I}_{\mathbb{T}_{\mathbb{LC}}}$ – the family of all R-implications obtained from left-continuous t-norms;
- ${}^{\mathbb{C}}\mathbb{I}_{\mathbb{T}}$ – the family of all continuous R-implications;
- ${}^{\mathbb{C}}\mathbb{I}_{\mathbb{T}_{\mathbb{LC}}}$ – the family of all continuous R-implications obtained from left-continuous t-norms;
- $\mathbb{I}_{\mathbb{LK}}$ – the family of all fuzzy implications Φ-conjugate with the Łukasiewicz implication $I_{\mathbf{LK}}$.

From Corollary 1 we have the following equalities between the above sets:

$${}^{\mathbb{C}}\mathbb{I}_{\mathbb{T}_{\mathbb{LC}}} = {}^{\mathbb{C}}\mathbb{I}_{\mathbb{T}} = \mathbb{I}_{\mathbb{LK}}.$$

6 (S,N)-implications

Definition 4 (cf. [15, 7, 1, 2])**.** A function $I\colon [0, 1]^2 \to [0, 1]$ is called an *(S,N)-implication*, if there exist a t-conorm S and a fuzzy negation N such that

$$I(x, y) = S(N(x), y), \qquad x, y \in [0, 1].$$

If N is a strong negation, then I is called an S-implication. Moreover, if an (S,N)-implication is generated from S and N, then we will often denote this by $I_{S,N}$.

Firstly note that $I_{S,N} \in \mathcal{FI}$ for any t-conorm S and any fuzzy negation N. In the class of continuous (S,N)-implications we have the following important result.

Proposition 4 ([2, Proposition 5.4])**.** *For a function $I\colon [0, 1]^2 \to [0, 1]$ the following statements are equivalent:*

(i) I is a continuous (S,N)-implication.

(ii) I is an (S,N)-implication generated from some continuous t-conorm S and some continuous fuzzy negation N.

Let us denote the different families of fuzzy implications as follows:

- $\mathbb{I}_{\mathbb{S},\mathbb{N}}$ – the family of all (S,N)-implications;
- ${}^{\mathbb{C}}\mathbb{I}_{\mathbb{S},\mathbb{N}}$ – the family of all continuous (S,N)-implications;
- $\mathbb{I}_{\mathbb{S}_{\mathbb{C}},\mathbb{N}_{\mathbb{C}}}$ – the family of all (S,N)-implications obtained from continuous t-conorms and continuous negations.

Observe, that from Proposition 4 we get

$${}^{\mathbb{C}}\mathbb{I}_{\mathbb{S},\mathbb{N}} = \mathbb{I}_{\mathbb{S}_{\mathbb{C}},\mathbb{N}_{\mathbb{C}}}.$$

ISBN: 978-989-95079-6-8

7 Intersection between continuous R- and (S,N)-implications

The intersections between the families and subfamilies of R- and (S,N)-implications have been studied by many authors, see e.g. [6, 13, 7, 3]. As regards the intersection between their continuous subsets only the following result has been known so far.

Theorem 7. *The only continuous (S,N)-implications that are also R-implications obtained from left-continuous t-norms are the fuzzy implications which are Φ-conjugate with the Łukasiewicz implication, i.e.,*

$$^{\mathrm{C}}\mathbb{I}_{\mathrm{S,N}} \cap \mathbb{I}_{\mathrm{T_{LC}}} = \mathbb{I}_{\mathrm{LK}}.$$

Now, from Corollary 1, we can prove the following equivalences.

Theorem 8. *For a function $I\colon [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i) I is a continuous (S,N)-implication that is also an R-implication obtained from a left-continuous t-norm.

(ii) I is a continuous (S,N)-implication that is also an R-implication.

(iii) I is an (S,N)-implication that is also a continuous R-implication.

(iv) I is Φ-conjugate with the Łukasiewicz implication, i.e., there exists $\varphi \in \Phi$, which is uniquely determined, such that I has the form (3) *for all $x, y \in [0,1]$.*

In summary, we have

$$^{\mathrm{C}}\mathbb{I}_{\mathrm{S,N}} \cap \mathbb{I}_{\mathrm{T_{LC}}} = {}^{\mathrm{C}}\mathbb{I}_{\mathrm{S,N}} \cap \mathbb{I}_{\mathrm{T}} = \mathbb{I}_{\mathrm{S,N}} \cap {}^{\mathrm{C}}\mathbb{I}_{\mathrm{T}} = {}^{\mathrm{C}}\mathbb{I}_{\mathrm{T}} = \mathbb{I}_{\mathrm{LK}}.$$

8 Conclusions

In this paper, we have shown that continuous R-implications cannot be obtained from purely non-left-continuous t-norms and that the only continuous R-implications are those that are Φ-conjugate with the Łukasiewicz implication. Using this result we have been able to answer another question related to the intersection between the continuous sub-families of (S,N)- and R-implications. We believe that this work will further help in solving some of the open problems still remaining with regards these two basic families of fuzzy implications.

References

[1] C. Alsina and E. Trillas. When (S,N)-implications are (T, T_1)-conditional functions? *Fuzzy Sets and Systems* 134 (2003) 305–310.

[2] M. Baczyński and B. Jayaram. On the characterizations of (S,N)-implications. *Fuzzy Sets and Systems* 158 (2007) 1713–1727.

[3] M. Baczyński and B. Jayaram. (S,N)- and R-implications: A state-of-the-art survey. *Fuzzy Sets and Systems* 159 (2008) 1836–1859.

[4] M. Baczyński and B. Jayaram. *Fuzzy implications.* (Studies in Fuzziness and Soft Computing, Vol. 231), Springer, Berlin, 2008.

[5] B. De Baets and K.C. Maes. Orthosymmetrical monotone functions. *Bull. Belg. Math. Soc. Simon Stevin* 14 (2007) 99–116.

[6] D. Dubois and H. Prade. A theorem on implication functions defined from triangular norms. *Stochastica* 8 (1984), 267–279.

[7] J. Fodor and M. Roubens. *Fuzzy preference modelling and multicriteria decision support.* Kluwer, Dordrecht, 1994.

[8] J.C. Fodor. Contrapositive symmetry of fuzzy implications. *Fuzzy Sets and Systems* 69 (1995) 141–156.

[9] S. Gottwald. *A treatise on many-valued logic.* Research Studies Press, Baldock, 2001.

[10] E.P. Klement, R. Mesiar and E. Pap. *Triangular norms.* Kluwer, Dordrecht, 2000.

[11] K.C. Maes and B. De Baets. A contour view on uninorm properties. *Kybernetika* 42 (2006) 303–318.

[12] B. Schweizer and A. Sklar. *Probabilistic metric spaces.* North-Holland, New York, 1983.

[13] P. Smets and P. Magrez. Implication in fuzzy logic. *Internat. J. Approx. Reason.* 1 (1987) 327–347.

[14] E. Trillas and L. Valverde, On some functionally expressable implications for fuzzy set theory. In: E.P. Klement (Ed.) *Proc. 3rd Inter. Seminar on Fuzzy Set Theory.* Linz, Austria, 1981, pp. 173–190.

[15] E. Trillas and L. Valverde. On implication and indistinguishability in the setting of fuzzy logic. In: J. Kacprzyk and R.R. Yager (Eds.), *Management decision support systems using fuzzy sets and possibility theory.* TÜV-Rhineland, Cologne, 1985, pp. 198–212.

ISBN: 978-989-95079-6-8

Fuzzy Voxel Object

Derek T. Anderson, Robert H. Luke, Erik Stone, and James M. Keller

Department of Electrical and Computer Engineering
University of Missouri, Columbia, USA
Email: {dtaxtd, rhl3db, erikstone}@mizzou.edu and kellerj@missouri.edu

Abstract—*In this paper, computer vision and fuzzy set theory are merged for the robust construction of three-dimensional objects using a small number of cameras and minimal a priori knowledge about the objects. This work extends our previously defined crisp model, which has been successfully used for recognizing and linguistically summarizing human activity. The objects true features more closely resemble the fuzzy object than those of the crisp object. This is demonstrated both empirically and through the comparison of features used in tracking human activity.*

Keywords—computer vision, human activity analysis, fuzzy objects, fuzzy voxel person.

1 Introduction

Information gathered from three-dimensional models is significantly more useful and robust than that gathered from two-dimensional sources. This is quite apparent in the area of human surveillance. Features, such as the centroid, height, velocity, orientation, etc, in three-dimensional space are not camera view dependent. Object segmentation can also be improved using three-dimensional information about the environment, such as floors, walls, and objects for the removal of additional erroneous artefacts, such as shadows and specular highlights [1][2]. At each step in computer vision, be it construction, shape refinement, etc, various forms of uncertainty are present and should be utilized.

We previously demonstrated a method for constructing a three-dimensional human model, voxel person, in voxel (volume element) space (figure 1) through the use of privacy protected images of the human called silhouettes [1][2].

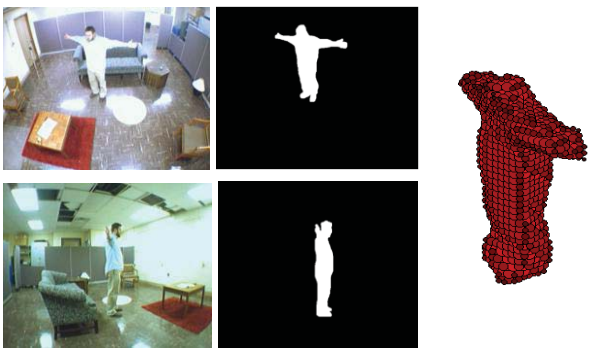


Fig. 1. Crisp voxel person created at a voxel resolution of 1.25"x1.25"x1.25", using 2 cameras and image silhouettes.

A silhouette is a binary map that distinguishes an individual (foreground) from his or her background. Well-known algorithms for computing the silhouette include Mixtures of Gaussians [3], Eigen Backgrounds [4], and Wallflower [5]. Silhouettes are used and we extract linguistic summarizations of human state based on crisp voxel person and fuzzy logic [1]. The resulting information is in a natural format for humans (linguistic) and it also yields succinct descriptions for managing complexity. We extended the work in [1] using a hierarchy of fuzzy logic and linguistic summarization for the inference of activity [2]. The knowledge (rules) is designed under the supervision of nurses for the recognition of falls in elderly populations.

Construction of three-dimensional objects, both solid representations and hulls, from two-dimensional images has been studied intensely in computer graphics, computer vision, biomedicine, and even in the activity analysis domain [6][7][8][9][10]. What separates our object construction related work from most, besides the use of silhouettes for back-projection, is the way in which a small number of low cost cameras, typically two per environment, are used to build a low, but rich for tracking, resolution voxel object; how its shape is refined using environment knowledge; and how features are extracted and used for fuzzy-based activity recognition [1][2]. Segmentation of the object into body parts (torso, head, etc) is not attempted, such as in [11][12]. While advances have been made in body segmentation, no approach to date is either real-time or mature enough to be included in a real-world eldercare system that runs unsupervised for long time periods.

This paper is divided as such. Section 2 describes the crisp voxel model, section 3 is the fuzzy model, and section 4 is the assesment of the quality of construction. Feature extraction is described next, followed by an alpha-cut procedure for an improved crisp voxel person, and section 7 is experiments and qualitative and quantitative results.

2 Crisp Voxel Object

An environment used for tracking is first converted to a voxel representation. Voxels are non-overlapping cubes that discretize a volume. The voxel centered at position $< i, j, l >$ is $v_{(i,j,l)}$. For each pixel in each camera, a list of intersected voxels is pre-computed, $L_{c,(n,m)}$, where c is the camera index, $1 \le c \le C$, and (n,m) is the pixel index [1]. The lists are sorted according to their distance (Euclidean) to the camera. This way, voxels that correspond to regions of change (the foreground, e.g. the human) from the viewpoint of a given camera can be quickly identified. For camera c at time t, the human voxel object is the union of all voxel lists

ISBN: 978-989-95079-6-8

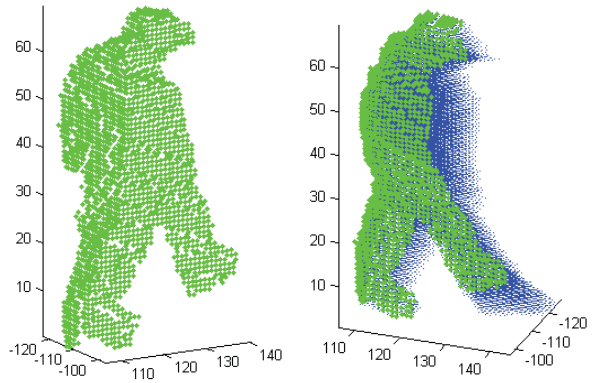
for the foreground pixel set, V_t^c. The object's shape is refined by intersecting the voxel objects acquired from each camera in the scene, i.e. $\{V_t^1,\ldots,V_t^C\}$, thus

$$V'_t = \wedge_{c=1}^{C} V_t^c . \quad (1)$$

The object is then subject to additional morphological operators for noise removal and voxel space processing for removing additional shadows and specular highlights [1].

Another useful object of interest is the visible shell, S_t, which is the set of all voxels that are directly visible according to the C cameras. The algorithm for computing the shell (shown in figure 2) at time step t is

Initialize the set S_t to empty
Compute voxel person, i.e. $V'_t = \wedge_{c=1}^{C} V_t^c$
For each camera $(1 \le c \le C)$
  For each pixel in the foreground set
    Find the closest voxel, $v_{(i,j,l)}$, from $L_{c,(n,m)}$ in V'_t
    Add $v_{(i,j,l)}$ to S_t.



(a) (b)

Fig. 2. (a) Shell (shown in green) and (b) shell along with the rest of voxel object (shown in blue) for a two camera setup.

3 Fuzzy Voxel Object

The quality of voxel person construction varies over the space depending on the object's position and the location of the cameras. As a result, information gathered from the crisp model can be inaccurate. The crisp voxel object is a good real-time initial technique to produce three-dimensional models for tracking. The technique works quite well when there are several cameras viewing a space with overlapping fields of view. Unfortunately, it is rare to have more than two cameras in a given area, due to factors such as cost, processing time of subsequent algorithms, data transmission and storage, and in the case of eldercare, limited space and installation locations, for the seamless integration of a passive video sensor network into the home.

Fuzzy voxel person is an attempt to model and use the different types of uncertainties related to the construction of the object. Each voxel is assigned a membership value that reflects how much it belongs to the actual object. The two types of uncertainty identified and fused in this paper are (a) how reasonable it is to infer that a voxel is part of the object based on the shell and (b) where a voxel is located relative to the distribution of mass across the object.

The first measure considers the minimum distance one must "step" in voxel space until a voxel on the shell is reached. More importance is placed on voxels near the shell. However, this measure has a natural tendency to favor the shell and pull confidence away from the central mass of the object. The second measure determines how confident one is in a voxel, based on how close it is to the central mass. This measure obviously favors the central mass and in return pulls confidence away from the shell. When these two sources of information are combined, the final value is a measure of how dense the region is around the voxel and how reasonable it is to infer that location given its distance to the observed shell. This combined confidence is low for voxels in the tail of the object (the furthest part from the shell), high in the dense close to the shell regions, and relatively higher (than the tail) for voxels in non-dense volumes close to the shell.

The first value is inversely related to a voxel's distance from the shell. The greatest certainty of object intersection occurs at the shell and decreases outwardly. The distance value is found using mathematical morphology, specifically dilation, $\oplus$. The kernel used is K, and $V'_t \oplus K$ is the dilation of the model V'_t by kernel K. A K of 3x3x3 of all ones is used (i.e. the immediate 26 adjacent voxel neighborhood). The distance value is computed quickly by repeatedly dilating the voxels in the shell, then subtracting that set from the set of remaining voxels in the intersected object. All surviving voxels in the intersected set have a related value, $m_{(i,j,l)}$ for $v_{(i,j,l)}$, that is incremented each iteration it survives. Formally, the algorithm is

Initialize all $m_{(i,j,l)}$ to 0
$V' = V'_t$
$S' = S_t$
while $|V'| > 0$
  for each $v'_{(i,j,l)} \in V'$
    add one to $m_{(i,j,l)}$
  end
  $V' = V' - S'$
  $S' = S' \oplus K$
end

The confidence for each voxel is

$$m'_{(i,j,l)} = 1 - \left(\frac{m_{(i,j,l)}}{\max_{m_{(a,b,c)}} m_{(a,b,c)}} \right). \quad (2)$$

The density value is computed using erosion, Θ. Again, a K of 3x3x3 of all ones is used. The voxel object is continually eroded until no voxels remain in the set. For each erosion step that a voxel remains in the set, a separate value, $e_{(i,j,l)}$ for $v_{(i,j,l)}$, is incremented. The algorithm is

ISBN: 978-989-95079-6-8

Initialize all $e_{(i,j,l)}$ to 0
$V' = V'_t$
while $|V'| > 0$
  for each $v'_{(i,j,l)} \in V'$
    add one to $e_{(i,j,l)}$
  end
  $V' = V' \ominus K$
end.

The most direct method for calculating the membership value per voxel is

$$e'_{(i,j,l)} = \frac{e_{(i,j,l)}}{\max_{e_{(a,b,c)}} e_{(a,b,c)}}. \quad (3)$$

However, this assigns low confidence to the object's entire outer shell. Up to this point, no a priori knowledge about the object being tracked has been assumed. For some objects, such as a human with appendages (arms, legs, chest, etc), this calculation might be too drastic, resulting in too little of importance assigned to the appendages (which are less dense than the chest). In order to accommodate such cases, the confidence values can instead be calculated as

$$e'_{(i,j,l)} = (1-\beta)\frac{e_{(i,j,l)}}{\max_{e_{(a,b,c)}} e_{(a,b,c)}} + \beta. \quad (4)$$

This maps the value into [β , 1], where ($0 \le \beta \le 1$) is a free parameter, either user defined or learned. The value β can be thought of as the minimum support one is willing to assign to voxels on, and subsequently near, the shell.

The final value per voxel, $\mu_{(i,j,l)} \in [0,1]$, is found using a t-norm (we use the product),

$$\mu_{(i,j,l)} = m'_{(i,j,l)} \wedge e'_{(i,j,l)}. \quad (5)$$

4 Object Construction Quality

As alluded to earlier, the quality of construction varies over the space with respect to the object's location, the installation locations of the cameras, and their respective orientations. Quantitatively representing the construction quality is of importance for two reasons. First, the value informs us about the reliability of the lower level results. For example, knowledge about the ability to properly construct the human affects the features calculated, and subsequently our decisions regarding activity inferred using those features. Secondly, there may be many locations in a space for which it is impossible to properly construct a voxel object given a particular camera configuration. In many instances, suppression of any object construction in those particular areas is of value.

The best construction generally occurs when the intersecting view vectors are orthogonal. This means that each individual voxel has a different construction quality. The quality of construction for voxel $v_{(i,j,l)}$ is

$$Q_{(i,j,l)} = 1 - \max_{\substack{C_g, C_h \\ g \neq h}} \left| (v_{(i,j,l)} - C_g)^T (v_{(i,j,l)} - C_h) \right|, \quad (6)$$

where C_g and C_h are the center locations for two cameras. Thus, the best case per voxel per camera pair with respect to orthogonality is considered. If the intrinsic camera parameters have been estimated, for example using [13], then the pixel ray in camera c that intersects voxel $v_{(i,j,l)}$ can be used instead. The environment is again converted into a voxel representation, but this time the resolution is lower. We empirically determined that a sampling of once every foot was a good resolution for our application. Figure 3 shows 9 horizontal slices (x-y planes for a varying z) of the sampled voxel space and their respective qualities.

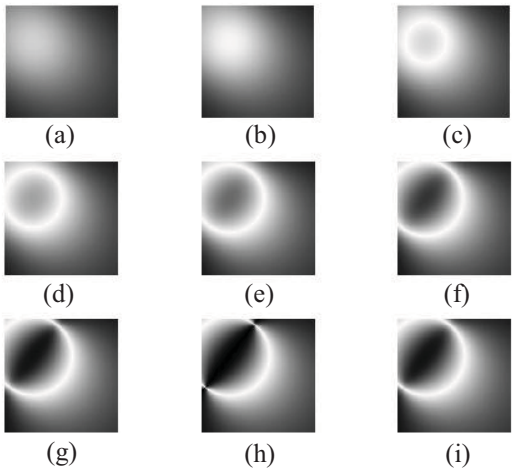


Fig. 3. Nine horizontal (x-y plane) slices of voxel object construction quality for the camera configuration {(a),(b)} shown in figure 4. Brighter values represent higher quality. Map (a) above is at a height of 1 foot, and each consecutive map is 1 foot higher in the z dimension (world up direction).

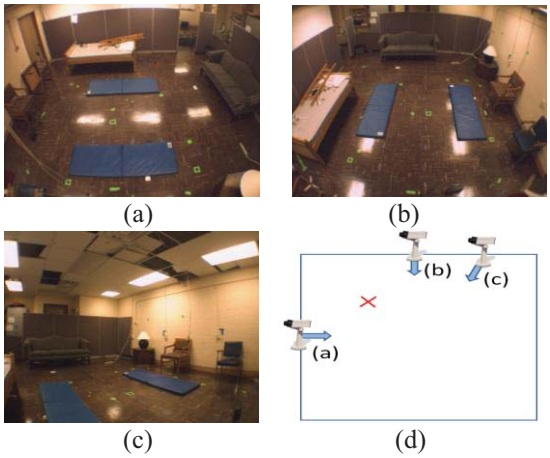


Fig. 4. Camera installation locations (d) and images showing their respective views of the monitored space (a-c).

The next matter is the determination of the quality of construction for an object. One can compute the mentioned

ISBN: 978-989-95079-6-8

quality measure for each voxel and make a decision based on the set of all memberships, or the object can be summarized, according to its centroid, height, or some other measurement, and that point estimate can be used to make a decision for the entire object. We perform the latter. We convert the monitored space into a low resolution voxel configuration (such as demonstrated in figure 3), then the height domain is collapsed and a single voxel plane in the x-y dimension is produced. The voxel qualities are combined using a t-norm (we used the min),

$$Q_{(i,j)} = \min_k Q_{(i,j,k)}. \qquad (7)$$

This pessimistically selects the worst construction case per element in the final x-y plane voxel set. Figure 5 shows the final x-y voxel plane of combined confidences for figure 3.

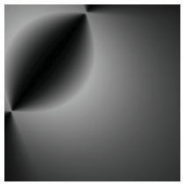
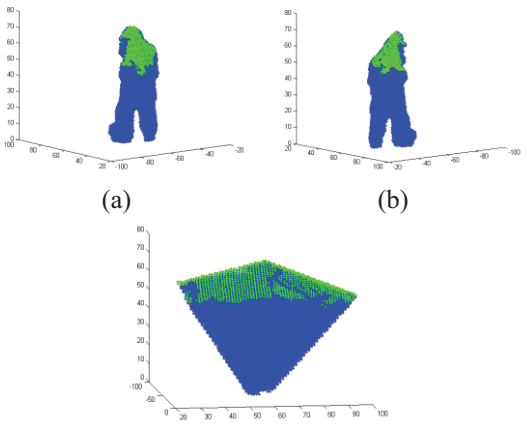


Fig. 5. T-norm produced x-y voxel plane of combined object construction confidences for the nine slices in figure 3.

The voxel object's centroid is mapped to the closest (Euclidean) voxel in the quality map (according to its x-y plane distance), and the respective quality value, ϕ_t, is retrieved. If $\phi_t < \delta_1$ (we empirically determined $\delta_1 = 0.2$), then the object is not constructed and ultimately not tracked. The parameter δ_1 is not specific to any one single tracking case, but it is rather a general parameter related to ray-based back-projection object construction. Figure 6 shows an example construction in which the quality value is 0.05.



(a) (b)

(c)

Fig. 6. Example low quality construction (value of 0.05) using camera configuration {(a),(b)} (the human's location is marked by an X in figure 4). The object appears to be constructed correctly when viewed independently with respect to the two cameras, images (a) and (b). However, its actual shape from another position (c) in three-dimensional space shows that the object was not built correctly.

5 Fuzzy Feature Extraction

After fuzzy voxel person is created, and its quality is assessed and determined to be adequate, one of two possible decisions can be made. The object can remain fuzzy and fuzzy feature extraction can be performed for activity analysis, or the fuzzy object can be used to acquire a better crisp voxel person and standard crisp feature extraction can be performed. This section details the prior.

The fuzzy feature selected for analysis in this paper is the centroid, which is most often used for determining object interaction, velocity, direction, position, and more generally, inferring the overall state of the tracked person (standing, kneeling, on the ground, etc) [1]. The fuzzy centroid is

$$M_{fuzzy} = \frac{\sum_{\dot{v}_{(i,j,l)} \in V'_t} \mu_{(i,j,l)} \dot{v}_{(i,j,l)}}{\sum_{\dot{v}_{(i,j,l)} \in V'_t} \mu_{(i,j,l)}}. \qquad (8)$$

6 Crisp Voxel Object Improvement

An alternative to directly using fuzzy voxel person for tracking is if the volume of the object is known ahead of time, or if it can be sufficiently approximated online, fuzzy voxel person can be used to acquire an improved crisp voxel object. Once the crisp object is obtained, standard crisp tracking and feature extraction can be used. The concept here is that fuzzy voxel person is useful for obtaining the global relative set of memberships, and a search for a better crisp object is guided by the evidence contained in fuzzy voxel person. By using the known or approximated volume, P_{real}, the highest confidence areas can be identified and kept. One possibility is to look for an alpha cut such that the crisp volume (cardinality) of the resultant set is closest to P_{real}. Formally, the search is for an α such that

$$\arg\min_{0 \le \alpha \le 1} \left| \left| {}^{\alpha}V' \right| - P_{real} \right|, \qquad (9)$$

where ${}^{\alpha}V'$ is

$${}^{\alpha}V' = \left\{ \dot{v}_{(i,j,l)} \mid \mu_{(i,j,l)} \ge \alpha \right\}. \qquad (10)$$

One can easily solve this by sorting all voxel memberships in descending order and then selecting the first P_{real} voxels (in the case of a tie, include all ties).

In the case that P_{real} must be approximated, the procedure is as follows. For a user defined time window, T frames, compute crisp voxel person and its associated quality value at each time step. Find the maximum ϕ_t value for the T time steps. If this value is above some threshold δ_2, again this is a global parameter not specific to any one camera configuration, use the crisp voxel-based volume to approximate P_{real}. Figure 7 shows a human built using an approximated P_{real}.

ISBN: 978-989-95079-6-8

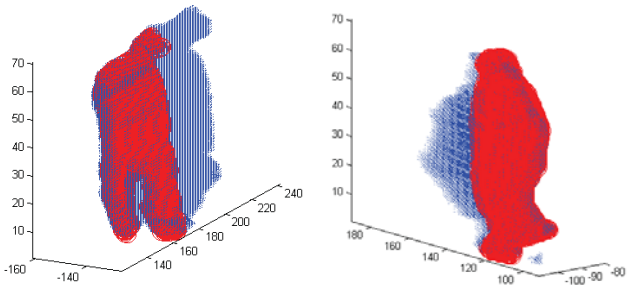


Fig. 7. The proposed alpha cut-based procedure for fuzzy voxel person to obtain an improved crisp object. Red areas are the improved voxel person and the blue areas are the rest of the original crisp voxel person.

In order to quantitatively demonstrate the improvement in the general shape and object's orientation, the following metrics for tracking humans are presented. The voxel covariance matrix is calculated, and its major orientation, the eigenvector with the largest respective eigenvalue, is compared to the world up vector ($<0,0,1>$). For a human standing and for a file cabinet used in the results section the major orientation should be very similar to the world up vector. Next, the difference, according to the variation associated with each eigenvector, between the new and real object is compared. These values indicate how the mass was distributed across the object according to its primary orthonormal basis with respect to variation. The eigen-information is extracted from the covariance matrix of voxel person, and the eigenvalues (φ) at 3 standard deviation are calculated, $\omega = \sqrt{\varphi} * 3$. The eigenvalues are sorted in decreasing order, and the difference between the real object (ω_{real}) and the crisp voxel (ω_{voxel}) is computed, $\omega' = |\omega_{real} - \omega_{voxel}|$. The feature used in the results section is the sum of the three components in ω', denoted by ω_{diff}.

7 Experiments and Quantitative Results

While figure 7 qualitatively demonstrated the benefits of the techniques presented in this paper, comparison between the (a) fuzzy and crisp centroids, (b) eigen-based analysis of the voxel spread (variance), and (c) comparison between the expected and the observed major orientations using eigen-analysis is performed. The subject in this case is a file cabinet that was placed in unique locations and orientations throughout the scene. A file cabinet has easy-to-recognize shape, orientation, known volume, and centroid.

For the file cabinet and the density-based measure, a value of $\beta = 0.5$ is used. Combinations of camera pairs, {(a),(b)} and {(b),(c)} were used. Camera combination {(a),(b)} is an ideal situation, given the near orthogonal installation, while {(b),(c)} is less than ideal. The later configuration results in much larger inaccuracies due to the trailing tail of intersected voxels. Tables 1-6 numerically display the accuracy of the fuzzy-based approaches, figures 8 and 9 show top-down view of the file cabinet at several locations, and figure 10 shows a few perspective views of the constructed object to give a better understanding of camera view induced warping and the resultant object shape.

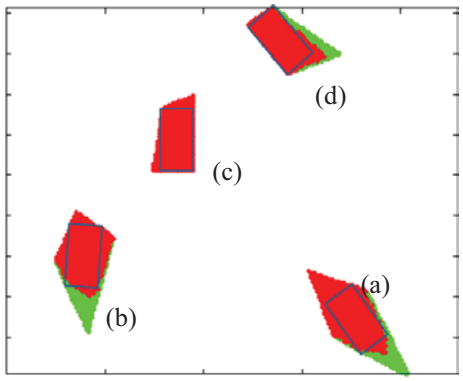


Fig. 8. Crisp (red and green areas) and alpha-cut based crisp objects (red areas) for camera configurations {(a),(b)}. The real object's approximate shape and orientation are overlaid using a rectangular outline.

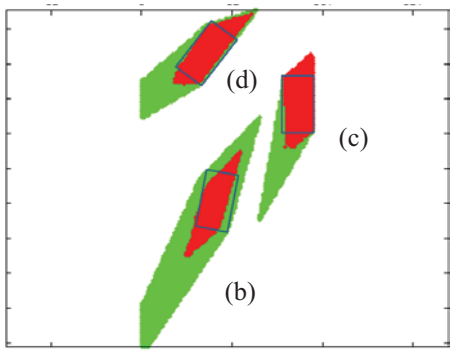


Fig. 9. Crisp (red and green areas) and alpha-cut based crisp objects (red areas) for camera configurations {(b),(c)}. The real object's approximate shape and orientation are overlaid using a rectangular outline.

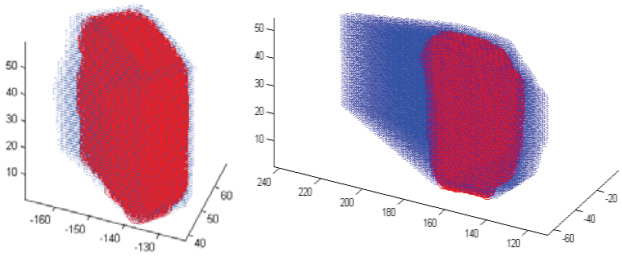


Fig. 10. Improved construction. Red areas are the improved voxel object and the blue areas are the rest of the original crisp voxel object. The left image is for (d) in figure 8, i.e. cameras {(a),(b)}, and the right image is for (b) in figure 9, an extreme case, using cameras {(b),(c)}.

Table 1: Difference between real and computed object centroid. Camera configuration {(a),(b)}. Units are specified in feet. Smaller numbers are better.

	a	b	c	d
Crisp	0.46	0.51	0.49	0.21
Fuzzy	0.19	0.29	0.47	0.27

ISBN: 978-989-95079-6-8

Table 2: Difference between real and computed object centroid. Camera configuration {(b),(c)}. Units are specified in feet. Smaller numbers are better.

	b	c	d
Crisp	2.24	1.18	1.13
Fuzzy	1.47	0.56	0.79

Table 3: Camera configuration {(a),(b)}. Value of ω_{diff}. Smaller numbers are better.

	a	b	c	d
Crisp	87.79	58.60	16.72	42.78
α -cut	4.60	1.72	0.82	0.07

Table 4: Camera configuration {(b),(c)}. Value of ω_{diff}. Smaller numbers are better.

	b	c	d
Crisp	3776.99	411.02	658.13
α -cut	807.52	55.71	17.11

Table 5: Camera configuration {(a),(b)}. Values are the dot product between the object's major orientation (eigenvector with the largest eigenvalue) with the world up vector. Larger numbers are better.

	a	b	c	d
Crisp	0.91	0.96	1.00	0.99
α -cut	1.00	1.00	1.00	1.00

Table 6: Camera configuration {(b),(c)}. Values are the dot product between the object's major orientation (eigenvector with the largest eigenvalue) with the world up vector. Larger numbers are better.

	b	c	d
Crisp	0.15	0.32	0.31
α -cut	0.93	0.96	0.79

The fuzzy centroid is more accurate than the crisp centroid in every case for configuration {(b),(c)} (i.e. extreme viewing condition). For configuration {(a),(b)}, very ideal condition, the centroids are very close. Fuzzy is better overall, however, in one instance the fuzzy centroid is slightly worse (Table 2.d). In that situation, camera (b) was looking at the object from a top down view, and the visible shell resulted in a greater vertical (z direction) pull. Tables 3 and 4 show that the objects shape (according to ω_{diff}) is closer to the real object. Lastly, tables 5 and 6 show that the objects major direction is always equal or more accurate (significant in Table 6) to the world up direction.

8 Conclusion and Future Work

In this paper, a fuzzy-based computer vision approach to the robust construction of three-dimensional objects for human activity analysis using only a few cameras and minimal a priori knowledge of the object is introduced. Extreme joint viewing conditions were considered and the fuzzy acquired results are better than the crisp counterpart in both the quantitative and qualitative regard.

The shell is currently approximated using the visible parts of the voxel object, created using back-projection. As the figures illustrate, this results in some warping in the objects shape. Stereo vision will be used to improve the estimation of the visible shell from depth maps (which will also require the t-norm operator in (5) to be revisited). Additionally, the use of a priori knowledge about the object being tracked will be used, for example, attempting region segmentation, and using that information for improved membership calculation.

Acknowledgment

Robert Luke and Derek Anderson are pre-doctoral biomedical informatics research fellows funded by the National Library of Medicine (T15 LM07089). This work is also supported by the National Science Foundation (ITR award llS-0428420) and the Administration on Aging (90AM3013).

References

[1] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, M Aud, "Linguistic Summarization of Video for Fall Detection Using Voxel Person and Fuzzy Logic," *Comp. Vision and Image Understanding*, doi:10.1016/j.cviu.2008.07.006, Vol. 113, pp. 80-89, 2008.

[2] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, M Aud, "Modeling Human Activity From Voxel Person Using Fuzzy Logic," *IEEE Transactions on Fuzzy Systems*, Accepted July 20th 2008.

[3] C. Stauffer, W.E.L. Grimson, Adaptive Background mixture Models for Real-time Tracking, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2), 243-252, 1999

[4] N. Oliver, B. Rosario, A. Pentland, A bayesian computer vision system for modeling human interactions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8), 831-843, 2000

[5] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, Wallflower: Principles and Practice of Background Maintenance. *In Proceedings of ICCV'1999*, 255-261 1999

[6] B. G. Baumgart, Geometric Modeling for Computer Vision. Technical Report AIM-249, Artificial Intelligence Laboratory, Stanford University, 1974.

[7] B. C. Vemuri and J. K. Aggarwal, "3-D model construction from multiple views using range and intensity data," in *Proceedings of IEEE Conf. Computer Vision and Pattern Recognition*, pp. 435-437, 1986.

[8] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, pp. 150-162, 1994.

[9] G. Dudek and D. Daum, "On 3-D surface reconstruction using shape from shadows," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 461-468, 1998.

[10] M. Pardas and J. Landabaso, "Foreground regions extraction and characterization towards real-time object tracking," in *Proc. of Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, Lecture Notes in Computer Science. Springer, 2005.

[11] G. Cheung, T. Kanade, J. Bouguet, and M. Holler, "A Real Time System for Robust 3D Voxel Reconstruction of Human Motions", in *Proc. of Computer Vision and Pattern Recognition*, Vol. 2, pp. 714-720, 2000.

[12] N. Thome, D. Merad, and S. Miguet, "Human Body Part Labeling and Tracking Using Graph Matching Theory", in *Proc. of IEEE Intl. Conf. on Video and Signal Based Surveillance*, 2006.

[13] Jean-Yves Bouguet, "Camera Calibration Toolbox for Matlab," Intel Corp, 2004, http://www.vision.caltech.edu/bouguetj/calib_doc/.

ISBN: 978-989-95079-6-8

Fuzzy Logic-Based Image Processing Using Graphics Processor Units

R. H. Luke, D. T. Anderson, J. M. Keller
Department of Computer Engineering
University of Missouri
Columbia, Missouri, USA
{rhl3db, dtaxtd}@mizzou.edu, kellerj@missouri.edu

S. Coupland
Centre for Computational Intelligence
De Montfort University
Leicester, LE1 9BH, UK
simonc@dmu.ac.uk

Abstract—**This paper introduces a parallelization of fuzzy logic-based image processing using Graphics Processor Units (GPUs). Using an NVIDIA 8800 Ultra, a 126 time speed improvement can be made to fuzzy edge extraction making its processing real-time. The GPU can process approximately 42 frames per second at 640x480 image resolution, thus 307,200 inference processes per frame. With a computational speed improvement of over two orders of magnitude, more time can be allocated to higher level computer vision algorithms. This GPU solution is described using NVIDIAs Compute Unified Device Architecture (CUDA).**

Keywords- ***Graphics Processor Units, Fuzzy Image Processing, Fuzzy Edge Extraction, Compute Unified Device Architecture***

I. INTRODUCTION

Edge extraction is a low level technique used in image processing and computer vision. There are many algorithms to perform edge extraction with a wide range of computational complexities and performance. Common approaches include the Sobel [1] and Prewitt [2] convolution kernels and the Canny Edge Detector [3]. Edges are most often used as low level information fed to a higher level process such as segmentation or recognition.

Some more recent algorithms have merged the strengths of image processing and fuzzy logic. These algorithms span a wide range of applications such as image enhancement [4], edge extraction [5], segmentation [6], and contrast adjustment [7]. An edge detection model based on Fuzzy Logic, called Fuzzy Inference Ruled by Else-Action (FIRE), was designed by Russo and Ramponi in [8]. This algorithm checks eight unique edge cases and outputs a value related to the confidence of the best matched case. The significance of this specific edge detection algorithm is its robustness to noise.

Even with simple algorithms, there is a large amount of work required to perform edge detection on moderately sized images. The workload is further compounded for real-time systems processing multiple frames per-second. A stream processing solution based on Graphics Processor Units (GPUs) and NVIDIAs Compute Unified Device Architecture (CUDA) is presented. While the work in this paper focuses on edge detection as the parallel execution of a single Fuzzy Inference System (FIS) for multiple different inputs, the GPU solution can be easily modified to fit a wide range of batch-processing tasks that utilize the standard Mandani-type FIS [9][10][11]. The system's extension is discussed in a later section.

It should be noted that we could have implemented any fuzzy logic-based image processing technique. We chose to demonstrate the proposed GPU-based architecture using FIRE because it is an easily understood algorithm. Our goal is not to show the performance of FIRE, but rather to present a framework for the parallel execution of fuzzy logic based image processing.

Several hardware accelerated FIS solutions have been developed to improve processing speed. For example, FPGA [12][13], and VLSI [14] solutions exist, however the significant portion of these are related to Fuzzy Control, not Image Processing. These specialized solutions are generally expensive and time consuming to modify and adapt to a new problem. In contrast, GPUs are relatively inexpensive, easily integrated with PCs and flexible to general purpose programming.

Using GPUs for general purpose computing is not a new idea. Well-known graphics API-based approaches include: Fast Fourier Transform (FFT) [15], linear algebra operators [16], protein folding [17], Fuzzy C-Means [18][19], and the execution of a single FIS [20]. Researchers have also begun to recently explore CUDA in Image Processing, such as: optical flow estimation [21], segmentation of medical images [22], and the Canny Edge Detector [23].

II. EDGE EXTRACTION USING FIRE

An input image I has luminance values from $\{0,1,\ldots,255\}$. For pixel P_{ij}, let $W_{ij} = \{P_{(i-e)(j-f)},\ldots,P_{(i+e)(j+f)}\}$, where $e,f \in \lfloor 0,\ldots,K/2 \rfloor$ and K is the window size, be the set of pixels in the neighborhood of P_{ij}, with the exception of P_{ij} (hence, $|W_{ij}| = K*K-1$). The inputs to the fuzzy edge extractor are the luminance differences between pixels in W_{ij} to P_{ij},

$$X_{(ij,m)} = W_{(ij,m)} - P_{ij}, \quad (1)$$

$$1 \le m \le (K*K-1),$$

ISBN: 978-989-95079-6-8

where $W_{(ij,m)}$ is the m[th] element in the window centered at P_{ij}. All subsequent operations performed on $X_{(ij,m)}$ are independent of all other pixels, therefore, $X_{(ij,m)}$ will be notationally reduced to X_m, to simplify the indexing and as all discussion is about a specific P_{ij} and its W_{ij}. The domain of X_m is the set $\{-255,\ldots,255\}$.

As stated previously, FIRE is a case-based system used to test eight unique edge orientations on a window. The size of the window used in this system is 3x3. Therefore, X has a cardinality, $|X|$, of 8.

An overly simplified rule is: *If a pixel lies upon an edge Then make it black Else make it white.* Eight unique cases, representing an edge at every 45 degrees of rotation, figure 1, is represented as a rule. Each rule has six antecedents, coinciding with six of the eight pixels in X. Fuzzy sets are used to represent the linguistic terms negative and positive. Similarly, the linguistic uncertainties of the consequent terms white and black are also modeled as fuzzy sets.

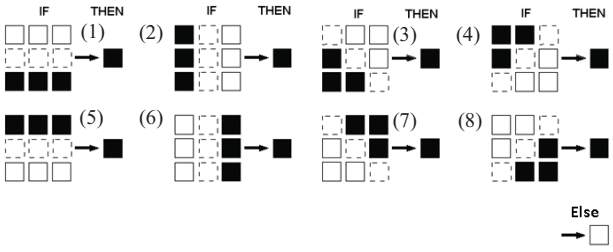


Figure 1. A graphical representation of the eight rules in the fuzzy rule base. Each rule has six antecedents represented as negative (black boxes) and positive (white boxes) luminance differences to the center pixel. The consequent is the same for each rule, a black pixel. A single else case handles the situation of none of the antecedents being matched.

For each pixel, each of the eight rules consisting of six antecedents are fired using the values in X. For each rule, the average of the six confidences is used as the total antecedent firing, f_i, where $1 \le i \le 8$. In this algorithm, it is assumed that a given pixel can belong to, at most, one of the eight cases. Therefore, the rule with the largest firing f_l is used as the antecedent firing strength to the fuzzy consequent set representing black.

$$f_l = \max(f_1,\ldots,f_8) \quad (2)$$

This is a departure from the more common Mamdani FIS which uses all rules in its decision making. The implication of the white fuzzy set uses $f_s = 1 - f_l$, i.e. the complement of the antecedent. The centroid of the aggregation of implication using f_l and f_s is the output of the algorithm. It should be noted that the result of the algorithm, $\{P'_{11},\ldots,P'_{ij},\ldots\}$, is in [0, 127.5], not [0, 255]. This is due to the setup of the antecedent fuzzy sets. In the extreme case, if all pixels in the 3x3 window are the same intensity, (no edge), all antecedent sets for all rules are fired with a 0.5 membership. The output of equation 2 will be 0.5 and therefore, f_s will also be 0.5. The consequents will have the same mass, and the final result will be 127.5. The domain of P'_{ij} is discretized and its range expanded to $\{0,1,\ldots,255\}$ by multiplying a constant, C, by P'_{ij}. In this paper, $C = 2$. Figure 2 displays an input and output image using this system.

Figure 2: Example output image of the algorithm described in [8].

III. CUDA INTRODUCTION

Because the calculation of the edge confidence at each pixel is independent of all other pixel calculations, this algorithm is a good candidate for parallelization. A single FIS is run for each pixel in an image. For a 640x480 image this results in the execution of a single FIS for each of the 307,200 pixels. Parallelization and implementation of the FIRE system to take advantage of the large number of co-processors on a GPU greatly increases the speed over sequentially processing the elements on the CPU.

NVIDIA defines CUDA as "a hardware and software architecture for issuing and managing computations on the GPU as a data-parallel computing device" [24]. The GPU is a highly parallel general purpose co-processor, not just a graphics processor. The CUDA API extends the C language and allows the largest portion of the programming community a quick transition to its use.

CUDA allows multiple programs, kernels, to run sequentially on a single GPU. CUDA organizes a kernel into a grid. A grid is subdivided into blocks. All blocks run the same kernel, but each runs independently from all others. Each block is made up of threads, the smallest divisible unit on the GPU. The actual work of the kernel is performed at the thread level.

The hierarchical structure of CUDA provides a mechanism to take advantage of the underlying hardware structure. This is mainly due to the grouping of processors into multiprocessors and that the memory types are unique to the grid, blocks and threads. Each multiprocessor is comprised of Q processors, Q=8 for the NVIDIA 8800, and each block is loaded into a single multiprocessor. Each multiprocessor is a Single Instruction Multiple Data (SIMD) set, meaning that multiple elements, Q, will be processed in parallel using the same instruction but different data. Threads in a block are partitioned into warps, where a warp size in the NVIDIA 8800 is 32. All threads in a warp are executed using the same

ISBN: 978-989-95079-6-8

instruction. Conditional logic can cause warp divergence and result in performance degredation. Code must be designed to take the warp size and control flow into account.

There are several types of memory in CUDA available to the programmer. Using the proper memory is vital to the efficiency of a program. Therefore, developers must create algorithms with memory access in mind.

The Global memory type can be read from and written to by any thread. Global memory access is the slowest of all types and should be used sparingly. The Constant and Texture memories can be read from any thread in a kernel, but not written to. The access times of texture memory are substantially smaller than Global memory.

Shared memory can be read from and written to by any thread associated with the block. This memory is local to the specific block and cannot be accessed by threads of any other block. Memory access times are shorter than the types mentioned above. Values from Global memory are often brought into shared memory at the beginning of a kernel, all threads are synchronized, and future repetitive access to this data is very efficient.

At the thread level, there is Local Memory and Registers which have the smallest access times. These memory locations are only accessible from a single thread. The entire memory layout is shown in figure 3.

IV. CUDA FIRE IMPLEMENTATION

This GPU implementation of the FIRE algorithm consists of two kernels. The first kernel performs all pixel neighborhood differencing, fuzzification for all rules, computes the average of the antecedent firings for each rule, and finds the maximum rule firing. The second kernel uses the minimum operator for implication, a maximum operator for rule aggregation and two sum reductions [25] to determine the centroid of the aggregated consequent sets for defuzzification. Though a simpler type of defuzzification could have been performed, we elected to use the standard general centroid defuzzification so that readers could easily extend this work to other more complex fuzzy-logic problems.

Even though more parallelism could be performed at each step, this two kernel approach is faster for three reasons. First, if the first kernel is subdivided into more parallel stages, each stage does too little computation relative to the amount of memory access. The GPU is most efficient when there is a large arithmetic to memory access ratio. Secondly, the second stage, implication, takes advantage of shared memory, making it more efficient to be a single kernel. Finally, most fuzzy systems use a small number of rules, less than 32, which does not justify further kernel subdivision.

The first kernel divides the input image into 8x8 non-overlapping pixel blocks. Each image block is processed in a CUDA block. This is done to make the processing of pixels faster by loading blocks of pixel data into shared memory for threads in a CUDA block to work on. As described earlier, the algorithm works on 3x3 windows of data for each pixel. In order to perform these operations on the pixels of the boundary of the 8x8 image blocks, a one pixel border must also be brought into memory. So, a 10x10 window of pixels is brought from texture memory into a block's shared.

Each thread can now perform the operations of the algorithm. An array of length eight, local to each thread, is defined and filled with the pixel neighborhood difference values (X_m , i.e. equation 1). The six antecedents of the eight rules are fired using the values in X . The antecedent firings are then summed across each rule. The max of these eight values is found and divided by six to find the average. This value is then written to the output of this kernel.

The output of the first kernel, f_1 , is the average of the maximum fired antecedent for each pixel of the input image and is in [0, 1]. For the second kernel, a separate CUDA block is created for each f_1 . Each of the T threads in a block performs inference aggregation, and is used to find the centroid of the aggregated consequent sets. Each consequent set is discretized into 256 elements. We empirically picked a T of 128, however, the selection of T is related to reduction, which is detailed later.

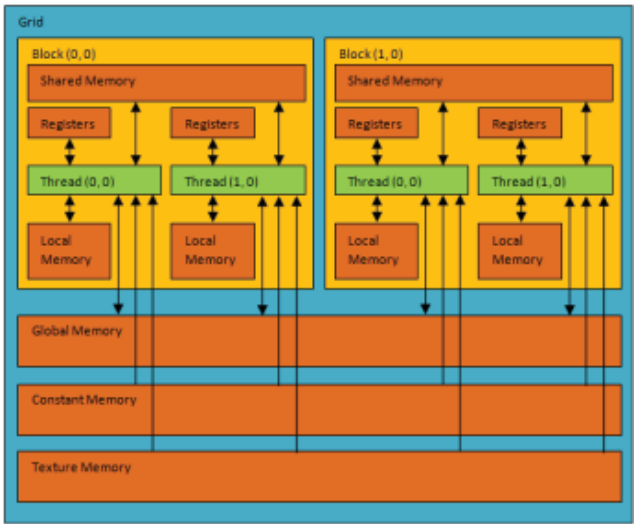


Figure 3: Processing organization scheme and memory layout for the GPU using CUDA. Global, Constant and Texture Memories are accessible by all threads. Shared Memory has is faster, but is local to each Block. Registers and Local Memory are the fastest, but are local to each thread.

Each thread performs inference (min) of f_1 with its corresponding discrete consequent domain membership value in the fuzzy set "black". The minimum of f_s and the appropriate discrete consequent domain membership value in the fuzzy set "white" is calculated, and each thread stores their aggregation (max) value in a shared memory array. Each thread is assigned an identifier in CUDA and this identifier is what is used to select the index into the discrete consequent domain fuzzy sets and the shared memory region. For a discrete consequent domain of size 256, a T of 128 was selected, where T was empirically determined. This means each thread calculates two discrete consequent domain elements instead of one. The reason for doing this is related to reduction, which is used for calculating the centroid.

ISBN: 978-989-95079-6-8

Reduction is the next step and there are many guidelines to follow, [25]. One rule states that it is generally better to calculate a few loads and reduction steps at the beginning versus allocating more threads, which become unused as reduction steps proceed. Other guidelines include sequential addressing and unrolling. Each block calculates the numerator and denominator separately and in parallel. The kernel must now compute the sum of the numerator and denominator. Reduction [25], which is the collapsing of a set of elements by some operator into one scalar, (such as the average, sum or max), is a common GPU operation. This procedure usually takes $\log_2(N)$ time, where N is the number of elements to reduce. For large values of N, such as N=1,000,000, a multiple block reduction would be used. However for 256 elements, a single block method keeping all in shared memory is the most efficient. An example of sum reduction in shown in figure 4.

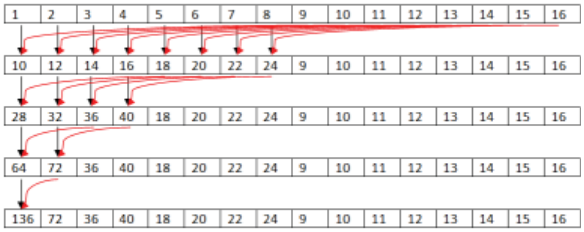


Figure 4: An example of sum reduction. In the first row, eight threads sample two locations in memory, add their values and output the result to a specified location. In the next row, four threads perform the operations. Then, two and one in the final two rows. The total sum is in the first index of the array in the bottom row.

It should be noted that GPU reduction is faster than a single CPU, however, not all processing cores can be used efficiently. This is because an increasing number of processors go unused with each reduction step. For example, reducing 256 elements with 128 threads will reduce to 128 elements in the second step. However, in the second step, only 64 of the threads are involved in reduction. In the third step only 32 are performing work, 16 in the next, and so on.

The final operation of the kernel is the calculation of the centroid. As previously discussed, this value is in [0, 127.5] and is multiplied by 2 (C). It is then discretized so that the final values of the output image, P'_{ij}, are in $\{0,1,\ldots,255\}$. The entire GPU solution is displayed in figure 5.

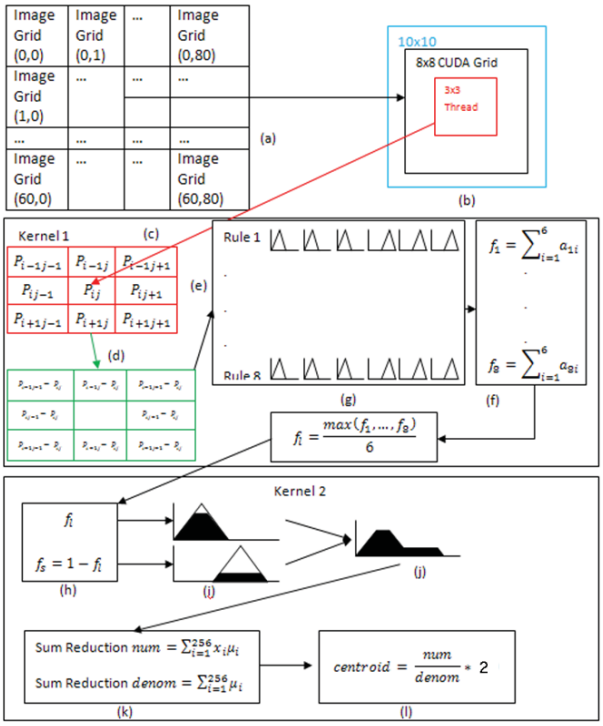


Figure 5: A graphical representation of the system defined in this paper. (a) The 640x480 input image divided into 8x8 blocks. (b) A single CUDA block with a 10x10 shared memory. (c) 3x3 input to a thread. (d) The difference values computed from (c). (e) Antecedent firing over eight rules. (f) Find the average of each rule. (g) Output the largest antecedent firing for kernel 1. (h) Compute else firing. (i) Perform implication on the consequent sets. (j) Aggregate consequent sets. (k) Sum reduction of the numerator and denominator values for centroid computation. (l) Compute the centroid of the aggregated consequents and scale to image output.

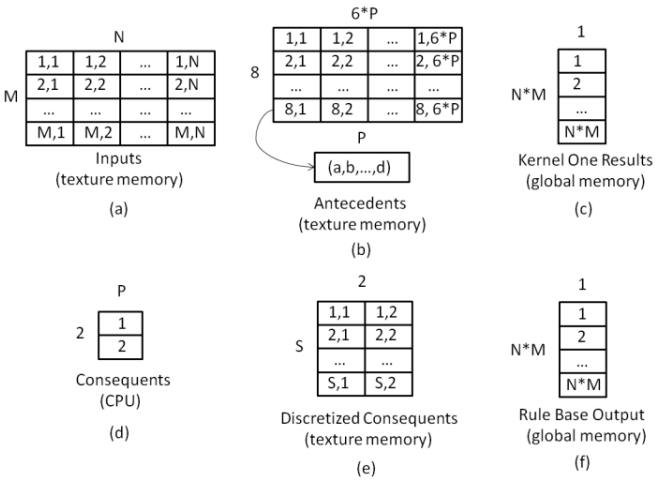


Figure 6: CPU memory and GPU global and texture memory organization for the GPU FIRE solution. For this paper, M = 640, N = 480 and S = 256, however, M, N and S can take on different profiles. (a) The input image in texture memory, (b) antecedents in texture memory, (c) global memory is used to store the output of the first kernel, (d) CPU side parameters of the two consequent sets, (e) GPU texture memory for the discretized consequent sets, and (f) global memory is used to store the output of the second kernel (image of edge confidences).

ISBN: 978-989-95079-6-8

As mentioned above, memory usage and organization is vital to CUDA efficiency. The input image is moved from the CPU to GPU texture memory. In this case, the image size is 640x480. Texture memory is used instead of the slower global memory because the image is read-only on the GPU. Secondly, the antecedent sets for the rule base are stored in texture memory. There are 8 rules with 6 antecedents per rule, and each antecedent set is made up of 4 ordered points, {a,b,c,d} for a trapezoid. Because the rules are read-only, texture memory is used. The output of the first kernel is stored in a one dimensional global memory segment. Next, the discretized consequent sets are transferred from the CPU memory to GPU texture memory. Finally, the output of kernel 2 is stored in a one dimensional global memory segment. The memory organization is displayed in figure 6.

RESULTS

The CPU used for testing was an AMD Athlon 64 FX-55 running at 2.6 GHz. The system had 2GB of RAM and the operating environment was Windows XP. The GPU was an NVIDIA 8800 BFG Ultra with 768 MB of memory [26]. This GPU has 128 processors and connects to the motherboard using PCI express 16X .

The CPU and GPU implementations were then run over a series of different image sizes. Table 1 shows the setups and the associated timings

TABLE I. PROCESSING TIMES AND SPEED INCREASE FROM CPU TO GPU

	320x240 Image 76,800 Pixels	640x480 Image 307,200 Pixels	960x720 Image 691,200 Pixels	1280x960 Image 1,228,800 Pixels	1280x1024 Image 1,310,720 Pixels
CPU	0.73	2.86	6.6	11.5	12.3
GPU	0.006	0.024	0.052	0.091	0.098
Speed Increase	121X	119X	126X	126X	125X

Figure 7: Images processed using the algorithm defined in this paper.

Two output images are shown in figure 7. These images are consistent with the output images from the original paper [8].

As it relates to global memory access, it was found that in the second kernel, the block format and global memory access pattern (index calculation per block) has a large performance impact. For example, two possible block formats for a 640x480 image are (38400, 8) and (8, 38400). A global memory read index (sampling of the f_l value) would then be $b_x * 8 + b_y$ for (38400, 8) and $b_x * 38400 + b_y$ for (8, 38400), where b_x and b_y are the block x and y indices respectively. It was noticed that the (38400, 8) format was three times faster than the (8, 38400) format. In fact, for all profiles (N_{b_x}, N_{b_y}), where N_{b_x} and N_{b_y} are the number of blocks in the x and y dimensions of the grid, it holds that greater speeds are produced if N_{b_y} is a small power of 2 (hence 8, 16, 32), resulting in a smaller calculated offset by $b_x * N_{b_y} + b_y$.

EXTENSION OF SYSTEM

It should be noted that the FIRE system is an unconventional FIS. The Mamdani FIS is the most common model used for Fuzzy Logic. A wide range of Image Processing operations, not just edge extraction, could be performed using a generalized GPU Mamdani FIS implementation.

The implementation defined in this paper requires only a few modifications to perform generalized Mamdani inference. First, instead of finding the mean value of all antecedent firings for a single rule, only the minimum firing would be used. Second, instead of using only the maximum antecedent firing to perform implication on a single consequent set, all fuzzy rules firings would be used. All that is needed is one global memory segment that is $(N * M) \times R$ in size (the FIRE implementation uses an $(N * M) \times 1$ memory segment, fig 6 c), where R is the number of rules. We showed in [27] that fuzzy consequent set aggregation can be efficiently performed as the first step in the reduction kernel, i.e. no need for a separate pass. This is done by storing the discrete consequents in texture memory, a memory segment of size $S \times R$ (the FIRE implementation uses a memory segment of size $S \times 2$, fig 6 e). In the first step of the reduction kernel for implication, aggregation and defuzzification, each thread samples its corresponding discrete consequent values for the R rules. Next, the minimum of each value fetched and their respective rule firings are computed. The final step is to compute the maximum over these minimums. This is a practical assumption, given that the number of rules in a typical rule base is low, e.g., 32 or less, and performing reduction on a set of elements this small is not ideal on the GPU.

CONCLUSION

This paper described a parallelization of fuzzy-logic based image processing on the GPU. The FIRE edge extractor was demonstrated specifically. Using the NVidia 8800 Ultra GPU,

ISBN: 978-989-95079-6-8

a 126 time speed improvement can be made to the original algorithm making its processing real-time. The GPU can process approximately 42 frames per second at 640x480 resolution using this algorithm.

With such large processing power, more time can be spent on higher level processes. The output of this system could be fed into a larger real-time system that performs more complex operations such as object recognition or tracking. The low price of GPUs and the ease of learning and using the CUDA API make this type of parallel programming a legitimate possibility for a large portion of the programming community.

REFERENCES

[1] K. Pringle. "Visual perception by a computer," *In Automatic Interpretation and Classification of Images*, pages 277-284. Academic Press, New York, 1969.

[2] J. Prewitt, "Object enhancement and extraction", in *Picture Processing and Psychopictorics*, Academic Press, pp. 75-149, 1970.

[3] J. Canny, "A computational approach to edge detection," *in IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 8, pp. 679–698, 1986.

[4] Y. S. Choi and R. Krishnapuram, "A robust approach to image enhancement based on fuzzy logic," *in IEEE Trans. Image Process*, 6 (6), pp. 808-825, 1997.

[5] H. R. Tizhoosh, "Fast fuzzy edge detection," in Proceedings NAFIPS 2002 Annual Meeting of the North American Fuzzy Information Processing Society, pp. 239-242, 27-29 June 2002.

[6] L. O. Hall and A. Namasivayam, "Using adaptive fuzzy rules for image segmentation," *in IEEE World Congress on Computational Intelligence*, Vol. 2, pp. 1560-1565, 1998.

[7] M. Craig and M. Schneider, "On the use of fuzzy sets in histogram equalization," *in Fuzzy Sets and Systems*, Vol. 45, pp. 271-278, 1992.

[8] F. Russo and G. Ramponi, "Edge extraction by FIRE operators", *in Proceedings of the Third IEEE International Conference on Fuzzy Systems*, 1994, pp. 249-253.

[9] L. Zadeh, 'Fuzzy sets,' *Information Control*, pp. 338-353, 1965.

[10] L. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes", *in IEEE Trans. on SMC*, Vol. 3, pp. 28-44, 1973.

[11] E. Mamdani, "Application of Fuzzy Algorithms for Simple Dynamic Plant", *in Proc. IEEE*, Vol. 121, pp. 1585-1588, 1974.

[12] M. Cirstea, J. Khor and M. McCormick, "FPGA fuzzy logic controller for variable speed generators," *Proceedings of the 2001 IEEE International Conference on Control Applications*, pp. 2001.

[13] C. Lynch, H. Hagras and V. Callaghan, "Parallel Type-2 Fuzzy Logic Co-Processors for Engine Management," *Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International*, 23-26 July 2007 pp.1-6.

[14] Hiroyuki Watanabe, Wayne D. Dettloff, and Kathy E. Yount, "A VLSI Fuzzy Logic Controller with Reconfigurable, Cascadable Architecture," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, 1990, pp. 376-382.

[15] K. Moreland and E. Angel, 'The FFT on a GPU,' in *Proc. SIGGRAPH/EUROGRAPHICS Conf. on Graphics hardware*, 112–119, 2003.

[16] J. Krüger and R. Westermann, "Linear algebra operators for GPU implementation of numerical algorithms," in *Intl. Conf. on Computer Graphics and Interactive Techniques*, 908-916, 2003.

[17] S. Larson, C. Snow, M. Shirts, and V. Pande, 'Folding@Home and Genome@Home: Using distributed computing to tackle previously intractable problems in computational biology,' *Comp. Genomics,* 2002.

[18] D. Anderson, R. H. Luke, and J. M. Keller, "Speedup of Fuzzy Clustering Through Stream Processing on Graphics Processing Units", in *IEEE Transactions on Fuzzy Systems*, 2007.

[19] D. Anderson, R. H. Luke, and J. M. Keller, "Incorporation of Non-Euclidean Distance Metrics into Fuzzy Clustering on Graphics Processing Units", in *Proceedings, International Fuzzy Systems Association Conference*, 2007.

[20] N. Harvey, R. H. Luke, J. M. Keller, and D. Anderson, "Speedup of Fuzzy Logic through Stream Processing on Graphics Processing Units", *in IEEE Congress on Evolutionary Computation (CEC), WCCI,* 2008.

[21] Y. Mizukami And K. Tadamura, "Optical Flow Computation on Compute Unified Device Architecture," in *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, 10-14 Sept. 2007 pp. 179-184.

[22] L. Pan, L. Gu nad J. Xu, "Implementation of medical image segmentation in CUDA," *Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on*, 30-31 May 2008 pp. 82-85.

[23] Y. Luo andR. Duraiswami, "Canny edge detection on NVIDIA CUDA," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPR Workshops 2008. IEEE Computer Society Conference on*, 23-28 June 2008 pp. 1-8.

[24] Nvidia Corp., 'NVIDIA CUDA Compute Unified Device Architecture,' June 23, 2007, http://developer.download.nvidia.com/compute/cuda/1_0/NVIDIA_CUDA_Programming_Guide_1.0.pdf

[25] M. Harris, "Optimizing Parallel Reduction in CUDA", NVIDIA Whitepaper, March 18, 2008, http://developer.download.nvidia.com/compute/cuda/1_1/Website/projects/reduction/doc/reduction.pdf

[26] Nvidia Corp., 'NVIDIA GeForce 8800 GPU Architecture Overview,' GeForce_8800_GPU_Architecture_Technical_Brief.pdf, November, 2006, http://www.nvidia.com/object/IO_37100.html

[27] D. Anderson and S. Coupland, "Parallelisation of Fuzzy Inference on a Graphics Processor Unit Using the Compute Unified Device Architecture", *UKCI 2008, the 8th Annual Workshop on Computational Intelligence*

ISBN: 978-989-95079-6-8

Dominance-based Rough Set Analysis of Uncertain Data Tables

Tuan-Fang Fan[1] Churn-Jung Liau[2] Duen-Ren Liu[3]

1.Department of Computer Science and Information Engineering, National Penghu University, Penghu 880, Taiwan
2.Institute of Information Science, Academia Sinica, Taipei 115, Taiwan
3.Institute of Information Management, National Chiao-Tung University, Hsinchu 300, Taiwan
Email: dffan@npu.edu.tw, liaucj@iis.sinica.edu.tw, dliu@iim.nctu.edu.tw

Abstract— *In this paper, we propose a dominance-based rough set approach for the decision analysis of a preference-ordered uncertain data table, which is comprised of a finite set of objects described by a finite set of criteria. The domains of the criteria may have ordinal properties that express preference scales. In the proposed approach, we first compute the degree of dominance between any two objects based on their imprecise evaluations with respect to each criterion. This results in a fuzzy dominance relation on the universe. Then, we define the degree of adherence to the dominance principle by every pair of objects and the degree of consistency of each object. The consistency degrees of all objects are aggregated to derive the quality of the classification, with which we can define the reducts of an uncertain data table. In addition, the upward and downward unions of decision classes are fuzzy subsets of the universe. The lower and upper approximations of the decision classes based on the fuzzy dominance relation are thus fuzzy rough sets. By using the lower approximations of the decision classes, we can derive two types of decision rules that can be applied to new decision cases.*

Keywords— Dominance-based rough set approach, multi-criteria decision analysis, preference-ordered data tables, rough set theory, uncertain data tables.

1 Introduction

The theory of knowledge has long been an important topic in many academic disciplines, such as philosophy, psychology, economics, and artificial intelligence, whereas the storage and retrieval of data is the main concern of information science. In modern experimental science, knowledge is usually acquired from observed data, which is a valuable resource for researchers and decision-makers. However, when the amount of data is large, it is difficult to analyze the data and extract knowledge from it. With the aid of computers, the vast amount of data stored in relational data tables can be transformed into symbolic knowledge automatically. Thus, intelligent data analysis has received a great deal of attention in recent years. The rough set theory proposed in [19] provides an effective tool for extracting knowledge from data tables.

When rough set theory is applied to *multi-criteria decision analysis* (MCDA), it is crucial to deal with preference-ordered attribute domains and decision classes [6, 7, 8, 9, 10, 11, 12, 24]. The original rough set theory cannot handle inconsistencies arising from violations of the dominance principle due to its use of the indiscernibility relation. Therefore, in the above-mentioned works, the indiscernibility relation is replaced by a dominance relation to solve the multi-criteria sorting problem; and the data table is replaced by a pairwise comparison table to solve multi-criteria choice and ranking problems. The approach is called the *dominance-based rough set approach* (DRSA). For MCDA problems, DRSA can induce a set of decision rules from sample decisions provided by decision-makers. The induced rules form a comprehensive preference model and can provide recommendations about a new decision-making environment.

A strong assumption about data tables is that each object takes exactly one value with respect to an attribute. However, in practice, we may only have incomplete information about the values of an object's attributes. Thus, more general data tables are needed to represent incomplete information. For example, set-valued and interval-valued data tables have been introduced to represent incomplete information [15, 16, 17, 18, 25]. DRSA has also been extended to deal with missing values in MCDA problems [10, 24]. Since a data table with missing values is a special case of uncertain data tables, we propose further extending DRSA to the decision analysis of uncertain data tables. In this paper, we investigate such an extension based on the fuzzy dominance principle.

In the proposed approach, we first compute the degree of dominance between any two objects based on their imprecise evaluations with respect to each criterion. This results in a fuzzy dominance relation on the universe. Then, we define the degree of adherence to the dominance principle by every pair of objects and the degree of consistency of each object. The consistency degrees of all objects are aggregated to derive the quality of the classification, with which we can define the reducts of the uncertain data tables. In addition, the upward and downward unions of decision classes are fuzzy subsets of the universe. The lower and upper approximations of the decision classes based on the fuzzy dominance relation are thus fuzzy rough sets. By using the lower approximations of the decision classes, we can derive two types of decision rules that can be applied in new decision-making environments.

In the next section, we review the dominance-based rough set approach. Then, in Section 3, we present the extension of DRSA for decision analysis of uncertain data tables. Section 4 contains some concluding remarks.

2 Dominance-based Rough Set Approach

2.1 Rough set theory

The basic construct of rough set theory is an *approximation space*, which is defined as a pair (U, R), where U is a finite universe and $R \subseteq U \times U$ is an equivalence relation on U. A binary relation R is an equivalence relation if it is (1)reflexive (i.e., $(x, x) \in R$ for all $x \in U$); (2) symmetric (i.e., for all $x, y \in U$, if $(x, y) \in R$, then $(y, x) \in R$); and (3) transitive

ISBN: 978-989-95079-6-8

(i.e., for all $x, y, z \in U$, if $(x, y) \in R$ and $(y, z) \in R$, then $(x, z) \in R$). An equivalence relation partitions the universe U into a family of equivalence classes so that each element of U belongs to exactly one of these equivalence classes. In other words, there exist $U_1, U_2, \cdots, U_k \subseteq U$ such that $U = \cup_{i=1}^{k} U_i$, $U_i \cap U_j = \emptyset$ for $i \neq j$; and for $x, y \in U$, $(x, y) \in R$ iff there exists i such that both x and $y \in U_i$. Thus, we can write an equivalence class of R as $[x]_R$ if it contains the element x. Note that $[x]_R = [y]_R$ iff $(x, y) \in R$.

According to Pawlak, knowledge is deep-seated in the classification capabilities of human beings. A classification is simply a partition of the universe, so an approximation space can construct basic knowledge about the objects in the universe. In philosophy, the *extension* of a concept is defined as the objects that are instances of the concept. For example, the extension of the concept "bird" is simply the set of all birds in the universe. Pawlak identified a concept by its extension. Thus, a subset of the universe is called a *concept* or a *category* in rough set theory.

Given an approximation space (U, R), each equivalence class of R is called an *R-basic category* or *R-basic concept*, and any union of R-basic categories is called an R-category. Now, for an arbitrary concept $X \subseteq U$, we are interested in defining X by using R-basic categories. We say that X is *R-definable*, if X is an R-category; otherwise X is *R-undefinable*. The R-definable concepts are also called *R-exact* sets, whereas R-undefinable concepts are said to be *R-inexact* or *R-rough*. When the approximation space is explicit from the context, we simply omit the qualifier R and call a set an exact set or a rough set.

A rough set can be approximated from below and above by two exact sets. The lower approximation and upper approximation of X are denoted by $\underline{R}X$ and $\overline{R}X$ respectively, and defined as follows:

$$\underline{R}X = \{x \in U \mid [x]_R \subseteq X\}, \tag{1}$$

$$\overline{R}X = \{x \in U \mid [x]_R \cap X \neq \emptyset\}. \tag{2}$$

2.2 Classical data tables

Although an approximation space is an abstract framework used to represent classification knowledge, it can easily be derived from a concrete data table (DT). The following formal definition of a data table is given in [20].

Definition 1 *A data table*[1] *is a tuple*

$$T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\}), \tag{3}$$

where U is a nonempty finite set, called the universe; A is a nonempty finite set of primitive attributes; for each $i \in A$, V_i is the domain of values for i; and for each $i \in A$, $f_i : U \rightarrow V_i$ is a total function.

An attribute in A is usually denoted by the lower-case letters i or a. In decision analysis, we assume the set of attributes is partitioned into $\{d\} \cup (A - \{d\})$, where d is called the *decision attribute*, and the remaining attributes in $A - \{d\}$ are called *condition attributes*.

[1] Also called knowledge representation systems, information systems, or attribute-value systems

Let $T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\})$ be a data table and $B \subseteq A$ be a subset of attributes. Then, we can define an equivalence relation, called the *indiscernibility relation* based on B, as

$$ind(B) = \{(x, y) \mid x, y \in U, f_i(x) = f_i(y) \forall i \in B\}. \tag{4}$$

In other words, x and y are B-indiscernible if they have the same values with respect to all the attributes in B. Consequently, for each $B \subseteq A$, $(U, ind(B))$ is an approximation space.

Let B be a subset of attributes. Then, an object x is *B-consistent* (with respect to the decision attribute d) if $[x]_B \subseteq [x]_d$, where $[x]_B = [x]_{ind(B)}$ and $[x]_d = [x]_{ind(\{d\})}$; otherwise, x is *B-inconsistent*. In other words, x is a B-consistent object in a data table if it satisfies the following *indiscernibility principle* for all $y \in U$:

$$(x, y) \in ind(B) \Rightarrow (x, y) \in ind(\{d\}). \tag{5}$$

That is, the objects that have the same evaluations as x on the condition attributes should have the same decision class assignment as x. Note that a *B-consistent* object corresponds to a decision rule whose decision class can be determined consistently based on the values of the attributes in B.

In [20], a decision logic (DL) is proposed for the representation of knowledge discovered from data tables. The basic alphabet of a DL consists of a finite set of attribute symbols A, and a finite set of value symbols V_i for $i \in \mathsf{A}$. Thus, the syntax of DL can be defined as follows.

Definition 2

1. *An atomic formula of DL is a descriptor (i, v), where $i \in \mathsf{A}$ and $v \in \mathsf{V}_i$.*
2. *The set of well-formed formulas (wffs) of a DL is the smallest set that contains the atomic formulas and is closed under the Boolean connectives $\neg$, $\wedge$, and $\vee$.*
3. *If φ and ψ are wffs of a DL, then $\varphi \longrightarrow \psi$ is a rule in the DL, where φ is called the antecedent of the rule and ψ is the consequent.*

A data table $T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\})$ relates to a given DL if there is a bijection $\tau : \mathsf{A} \rightarrow A$ such that, for every $a \in \mathsf{A}$, $V_{\tau(a)} = \mathsf{V}_a$. Thus, by somewhat abusing the notation, we usually denote an atomic formula as (i, v), where $i \in A$ and $v \in V_i$ if the data tables are clear from the context. Intuitively, each element in the universe of a data table corresponds to a data record, and an atomic formula (which is in fact an attribute-value pair) describes the value of some attribute in the data record. Thus, the atomic formulas (and therefore the wffs) can be satisfied or not satisfied by each data record. This results in a satisfaction relation between the universe and the set of wffs.

Definition 3 *Given a DL and a data table $T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\})$ related to it, the satisfaction relation $\models_T$ between U and the wffs of the DL is defined inductively as follows (the subscript T is omitted for brevity).*

1. *$x \models (i, v)$ iff $f_i(x) = v$,*
2. *$x \models \neg\varphi$ iff $x \not\models \varphi$,*

ISBN: 978-989-95079-6-8

3. $x \models \varphi \wedge \psi$ *iff* $x \models \varphi$ *and* $x \models \psi$,

4. $x \models \varphi \vee \psi$ *iff* $x \models \varphi$ *or* $x \models \psi$.

If φ is a DL wff, the set $m_T(\varphi)$ defined by

$$m_T(\varphi) = \{x \in U \mid x \models \varphi\} \tag{6}$$

is called the meaning set of the formula φ in T. If T is understood, we simply write $m(\varphi)$.

In terms of DL, each equivalence class of $ind(B)$ is characterized by a DL formula $\wedge_{i\in B}(i, v_i)$ and any formula φ of DL can be considered as a concept $m_T(\varphi)$. Thus, for each B-consistent object x, we can derive a valid rule $\wedge_{i\in B}(i, f_i(x)) \longrightarrow (d, f_d(x))$.

2.3 Preference-ordered data tables

In this subsection, we consider the DRSA in [10]. For MCDA problems, each object in a data table or decision table can be seen as a sample decision, and each condition attribute is a criterion for the decision. Since the domain of values of a criterion is usually ordered according to the decision-maker's preferences, we define a preference-ordered data table (PODT) as a tuple

$$T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\}), \tag{7}$$

where $T = (U, A, \{V_i \mid i \in A\}, \{f_i \mid i \in A\})$ is a classical data table; and for each $i \in A$, $\succeq_i \subseteq V_i \times V_i$ is a binary relation over V_i. The relation $\succeq_i$ is called a *weak preference relation* or *outranking* on V_i, and represents a preference over the set of objects with respect to the criterion i [24]. For $x, y \in U$, $f_i(x) \succeq_i f_i(y)$ means "x is at least as good as y with respect to criterion i". The weak preference relation $\succeq_i$ is supposed to be a complete preorder, i.e., a complete, reflexive, and transitive relation. In addition, we assume that the domain of the decision attribute is a finite set $V_d = \{1, 2, \cdots, n\}$ such that r is strictly preferred to s if $r > s$ for any $r, s \in V_d$. Thus, the weak preference relation $\succeq_d$ is defined as $r \succeq_d s$ iff $r \geq s$; consequently, $\succeq_d$ is a total order.

For a condition criterion i and an object x, $f_i(x)$ denotes the *evaluation* of the object with respect to the criterion i; and for the decision attribute d, $f_d(x)$ is the *assignment* of x to a decision label in V_d. Let P be a subset of criteria. We can then define the *P-dominance relation* $D_P \subseteq U \times U$ as follows:

$$(x, y) \in D_P \Leftrightarrow f_i(x) \succeq_i f_i(y) \forall i \in P. \tag{8}$$

When $(x, y) \in D_P$, we say that x P-dominates y, and that y is P-dominated by x. We usually use the infix notation xD_Py to denote $(x, y) \in D_P$. Although each $\succeq_i$ is a complete preorder, the dominance relation may simply be a preorder. If $P = \{i\}$ is a singleton, we write D_i instead of $D_{\{i\}}$. The most basic principle underlying DRSA is called the *dominance principle*. Let P denote a subset of condition criteria. Then, the dominance principle with respect to P can be expressed for $x, y \in U$ as follows:

$$xD_Py \Rightarrow xD_dy. \tag{9}$$

The principle states that if x P-dominates y (i.e., x is at least as good as y with respect to all criteria in P), then x should be assigned to a decision class at least as good as the class assigned to y.

In the classical rough set approach, a consistent object must satisfy the indiscernibility principle; however, in DRSA, we also require adherence to the dominance principle. Thus, an object x is *P-consistent* in the PODT $T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\})$ if for all $y \in U$, we have

$$(xD_Py \Rightarrow xD_dy) \wedge (yD_Px \Rightarrow yD_dx); \tag{10}$$

otherwise, x is *P-inconsistent*. Note that the dominance principle implies the indiscernibility principle because of the reflexivity of the dominance relation and the antisymmetry of $\succeq_d$.

Given the dominance relation D_P, the P-dominating set and P-dominated set of x are defined, respectively, as

$$D_P^+(x) = \{y \in U \mid yD_Px\} \tag{11}$$

and

$$D_P^-(x) = \{y \in U \mid xD_Py\}. \tag{12}$$

In addition, for each $t \in V_d$, we define the decision class Cl_t as $\{x \in U \mid f_d(x) = t\}$. Then, the *upward and downward unions of classes* are defined as

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s \tag{13}$$

and

$$Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s \tag{14}$$

respectively. Based on the P-dominating sets and P-dominated sets, we can define the P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ for each $t \in V_d$ as follows:

$$\underline{P}(Cl_t^{\geq}) = \{x \in U \mid D_P^+(x) \subseteq Cl_t^{\geq}\}, \tag{15}$$

$$\overline{P}(Cl_t^{\geq}) = \{x \in U \mid D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\}, \tag{16}$$

$$\underline{P}(Cl_t^{\leq}) = \{x \in U \mid D_P^-(x) \subseteq Cl_t^{\geq}\}, \tag{17}$$

$$\overline{P}(Cl_t^{\leq}) = \{x \in U \mid D_P^+(x) \cap Cl_t^{\geq} \neq \emptyset\}. \tag{18}$$

The P-boundaries of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are then defined as

$$Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}) \tag{19}$$

and

$$Bn_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}) \tag{20}$$

respectively. Let $\mathbf{Cl} = \{Cl_t \mid t \in V_d\}$ denote the partition of the universe into decision classes. Then, the *quality of the approximation* of the partition $\mathbf{Cl}$ based on the set of criteria P is defined as the ratio

$$\gamma_P(\mathbf{Cl}) = \frac{|U - (\bigcup_{t\in V_d} Bn_P(Cl_t^{\geq}) \cup \bigcup_{t\in V_d} Bn_P(Cl_t^{\leq}))|}{|U|}. \tag{21}$$

Note that $\gamma_P(\mathbf{Cl})$ is equal to the ratio of P-consistent objects in the universe U. Let $C = A - \{d\}$ be the set of all condition criteria; then every minimal subset $P \subseteq C$ such that $\gamma_P(\mathbf{Cl}) = \gamma_C(\mathbf{Cl})$ is a *reduct* of C.

ISBN: 978-989-95079-6-8

2.4 Preference-ordered decision logic

To represent the rules induced from a PODT, preference-ordered decision logic (PODL) is proposed in [5]. The syntax of PODL is the same as that of DL, except for the form of the atomic formulas. An atomic formula in PODL is a descriptor in the form of $(\geq_i, v)$ or $(\leq_i, v)$, where $i \in A$ and $v \in V_i$. The satisfaction relation between U and the set of PODL wffs is defined in the same way as the relation for DL wffs, except that the satisfaction of an atomic formula is defined by $x \models (\geq_i, v)$ iff $f_i(x) \succeq v$, and by $x \models (\leq_i, v)$ iff $v \succeq f_i(x)$.

By using DRSA, two kinds of rules can be induced from a PODT explicitly. Let P be a reduct and $t \in V_d$. Then, for an object $x \in \underline{P}(Cl_t^{\geq})$, we can derive

$$\bigwedge_{i \in P} (\geq_i, f_i(x)) \longrightarrow (\geq_d, t), \tag{22}$$

and for an object $x \in \underline{P}(Cl_t^{\leq})$, we can derive

$$\bigwedge_{i \in P} (\leq_i, f_i(x)) \longrightarrow (\leq_d, t). \tag{23}$$

Note that if x is P-consistent, these two formulas also hold for $t = f_d(x)$.

3 DRSA for Uncertain Data Tables

3.1 Preference-ordered Uncertain Data Tables

Although the PODT can represent multi-criteria decision cases effectively, it inherits the restriction of the classical DT so that uncertain information can not be represented. An uncertain data table is a generalization of a DT such that the values of some or all of its attributes are imprecise [4, 1]. An analogous generalization can be applied to PODT to define preference-ordered uncertain data tables (POUDT). Formally, a POUDT is a tuple

$$T = (U, A, \{(V_i, \succeq_i) \mid i \in A\}, \{f_i \mid i \in A\}), \tag{24}$$

where $U, A, \{(V_i, \succeq_i) \mid i \in A\}$ are defined as above, and for each $i \in A$, $f_i : U \to 2^{V_i} - \{\emptyset\}$. The intuition about a POUDT is that the evaluation of criterion i for an object x belongs to $f_i(x)$, although the evaluation is not known exactly. When $f_i(x)$ is a singleton, we say that the evaluation is precise. If all evaluations of T are precise, then T is said to be precise. Furthermore, we assume that for each criterion i, the Cartesian plane $V_i \times V_i$ is endowed with a uniform measure μ_i. Thus, for each subset $S \subseteq V_i \times V_i$, $\mu_i(S)$ is a non-negative real number. When V_i is a finite set, we take $\mu_i(S)$ as the cardinality of S; and when V_i is a real interval, we take $\mu_i(S)$ as the area of S.

3.2 Fuzzy dominance relation

In a POUDT, the objects may have imprecise evaluations with respect to the condition criteria and imprecise assignments to decision classes. Thus, the dominance relation between objects can not be determined with certainty. Instead, we define a degree of dominance between two objects with respect to each criterion i based on the associated measures μ_i. Formally, the dominance relation with respect to the criterion i is a fuzzy relation $D_i : U \times U \to [0, 1]$ such that for all $x \neq y$,

$$D_i(x, y) = \frac{\mu_i(\{(v_1, v_2) \mid v_1 \succeq_i v_2, v_1 \in f_i(x), v_2 \in f_i(y)\})}{\mu_i(f_i(x) \times f_i(y))}, \tag{25}$$

and $D_i(x, x) = 1$ for any $x \in U$.

An example of computing the degree of dominance is shown in Figure 1, where the evaluation of x with respect to criterion i, denoted by $s(x)$, is in a continuous interval $f_i(x) = [l_x, u_x]$. In this example, $D_i(x, y)$ is the ratio of the area of ABC over the area of $ABDE$, i.e., $\frac{u_x - l_y}{2(u_y - l_y)}$.

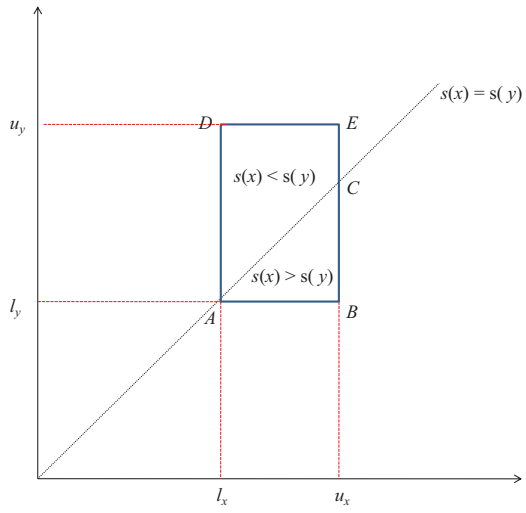


Figure 1: The degree of dominance between x and y

After deriving the fuzzy dominance relation for each criterion, we can aggregate all the relations into P-dominance relations for any subset of criteria P. Let $\otimes$, $\oplus$ and $\rightarrow$ denote, respectively, a t-norm, an s-norm and an implication operation[2] on $[0, 1]$. Then, the fuzzy P dominance relation $D_P : U \times U \to [0, 1]$ is defined as

$$D_P(x, y) = \bigotimes_{i \in P} D_i(x, y). \tag{26}$$

Since the dominance relation is a fuzzy relation, the satisfaction of the dominance principle is a matter of degree. Thus, the *degree of adherence* of (x, y) to the dominance principle with respect to a subset of condition criteria P is defined as

$$\delta_P(x, y) = D_P(x, y) \rightarrow D_d(x, y), \tag{27}$$

and the degree of P-consistency of x is defined as

$$\delta_P(x) = \bigotimes_{y \in U} (\delta_P(x, y) \otimes \delta_P(y, x)). \tag{28}$$

Let T be a POUDT. Then, the *quality of the classification* of T based on the set of criteria P is defined as

$$\gamma_P(T) = \frac{\sum_{x \in U} \delta_P(x)}{|U|}. \tag{29}$$

Note that $\gamma_P(T)$ is monotonic with respect to P, i.e., $\gamma_Q(T) \leq \gamma_P(T)$ if $Q \subseteq P$. Thus, we can define every minimal subset $P \subseteq C$ such that $\gamma_P(T) = \gamma_C(T)$ as a *reduct* of C, where

[2] For the properties of these operations, see a standard reference on fuzzy logic, such as [14]

ISBN: 978-989-95079-6-8

$C = A - \{d\}$ is the set of all condition criteria. In addition, the degree of P-consistency is monotonic with respect to P, so a reduct is also a minimal subset $P \subseteq C$ such that $\delta_P(x) = \delta_C(x)$ for all $x \in U$. However, because $\delta_P(x)$ is less sensitive to individual changes in $\delta_P(x, y)$, we can not guarantee that a reduct will preserve the degree of adherence to the dominance principle for each pair of objects. Thus, an alternative definition of the quality of the classification is

$$\eta_P(T) = \frac{\sum_{x,y \in U} \delta_P(x, y)}{|U|^2}. \tag{30}$$

The reducts can also be defined with respect to this kind of definition.

3.3 Dominance-based fuzzy rough approximations

In a POUDT, the assignment of a decision label to an object may be imprecise, so the decision classes may be fuzzy subsets of the universe. First, for each decision label $t \in V_d$, the decision class $Cl_t : U \to [0, 1]$ is defined by

$$Cl_t(x) = \begin{cases} \frac{1}{|f_d(x)|}, & \text{if } t \in f_d(x), \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

Second, the upward and downward unions of classes are defined by

$$Cl_t^{\geq}(x) = \frac{|f_d(x) \cap \{v \in V_d : v \geq t\}|}{|f_d(x)|} \tag{32}$$

and

$$Cl_t^{\leq}(x) = \frac{|f_d(x) \cap \{v \in V_d : v \leq t\}|}{|f_d(x)|} \tag{33}$$

respectively. Note that $Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$ and $Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$ only hold when we take the Łukasiewicz s-norm as the union operation, i.e., only when $(F \cup G)(x) = F(x) \oplus G(x)$, where $a \oplus b = \min(1, a + b)$. Finally, since our dominance relation is a fuzzy relation and the decision classes are fuzzy sets, the lower and upper approximations of these classes are defined in the same way as those for fuzzy rough sets[3, 21]. More specifically, the P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ for each $t \in V_d$ are defined as fuzzy subsets of U with the following membership functions:

$$\underline{P}(Cl_t^{\geq})(x) = \bigotimes_{y \in U} (D_P(y, x) \to Cl_t^{\geq}(y)), \tag{34}$$

$$\overline{P}(Cl_t^{\geq})(x) = \bigoplus_{y \in U} (D_P(x, y) \otimes Cl_t^{\geq}(y)), \tag{35}$$

$$\underline{P}(Cl_t^{\leq})(x) = \bigotimes_{y \in U} (D_P(x, y) \to Cl_t^{\leq}(y)), \tag{36}$$

$$\overline{P}(Cl_t^{\leq})(x) = \bigoplus_{y \in U} (D_P(y, x) \otimes Cl_t^{\leq}(y)). \tag{37}$$

3.4 Decision rules

To represent knowledge discovered from a POUDT, we generalize PODL to a kind of preference-ordered uncertain decision logic (POUDL). The syntax of POUDL is same as that of PODL, except that its atomic formulas are of the form $(\geq_i, s_i)$ or $(\leq_i, s_i)$, where $i \in A$ and $s_i \subseteq V_i$. When $s_i = \{v_i\}$ is a singleton, we abbreviate $(\geq_i, s_i)$ (resp. $(\leq_i, s_i)$) as $(\geq_i, v_i)$ (resp. $(\leq_i, v_i)$).

Let P denote a reduct of a POUDT and $t \in V_d$. Then, for each object x where $\underline{P}(Cl_t^{\geq})(x) > 0$ (or above some predetermined threshold), we can derive the first type of fuzzy rule:

$$\underline{P}(Cl_t^{\geq})(x) : \bigwedge_{i \in P} (\geq_i, f_i(x)) \longrightarrow (\geq_d, t); \tag{38}$$

and for each object x where $\underline{P}(Cl_t^{\leq})(x) > 0$ (or above some predetermined threshold), we can derive the second type of fuzzy rule:

$$\underline{P}(Cl_t^{\leq})(x) : \bigwedge_{i \in P} (\leq_i, f_i(x)) \longrightarrow (\leq_d, t), \tag{39}$$

where $\underline{P}(Cl_t^{\geq})(x)$ and $\underline{P}(Cl_t^{\leq})(x)$ are the respective degrees of truth of the rules.

Now, for a new decision case with (possibly imprecise) evaluations on the condition criteria P, we can apply these two types of rules to derive its decision label assignment. Specifically, let x be a new object such that for each criterion $i \in P$, $f_i(x) \subseteq V_i$ is given; and let α be a rule $c : \bigwedge_{i \in P} (\geq_i, s_i) \longrightarrow (\geq_d, t)$ discovered by the above-mentioned approach. Then, we can derive that the degree of satisfaction of $f_d(x) \succeq_d t$, according to the rule α and denoted by $\varepsilon(\alpha, f_d(x) \succeq_d t)$, is

$$c \otimes \bigotimes_{i \in P} \frac{\mu_i(\{(v_1, v_2) \mid v_1 \succeq_i v_2, v_1 \in f_i(x), v_2 \in s_i\})}{\mu_i(f_i(x) \times s_i)}. \tag{40}$$

Let $\mathcal{R}_t^{\geq}$ denote the set of all rules with the consequent $(\geq_d, t)$. Then, the final degree of $f_d(x) \succeq_d t$ is

$$\bigoplus_{\alpha \in \mathcal{R}_t^{\geq}} \varepsilon(\alpha, f_d(x) \succeq_d t). \tag{41}$$

Analogously, we can derive the degree of $f_d(x) \preceq_d t$ from the second type of rule.

4 Conclusion

The work reported in this paper extends DRSA to a dominance-based fuzzy rough set approach (DFRSA), which can be applied to the reduction of criteria and the induction of rules for decision analysis in a POUDT. Unlike other approaches that deal with imprecise evaluations and assignments, DFRSA induces fuzzy rules instead of qualitative rules. Thus, it would be worthwhile to compare DFRSA with other extensions of DRSA for handling uncertain data tables, like those proposed in [2, 13, 22]. In addition, the proposed approach may be useful for sparse data sets[23][3], so we will explore possible applications of DFRSA to such data sets in a future work. Moreover, since DFRSA is a general framework, we do not specify the t-norms used in the aggregation of consistency degrees or the implications used in the definition of adherence to the dominance principle. Consequently, we do not present detailed algorithms for the computation of reducts. The computational aspects of DFRSA for specialized t-norms and implication operations will aslo be addressed in a future work.

[3] We would like to thank the anonymous referee for pointing out this possibility.

ISBN: 978-989-95079-6-8

Acknowledgment

This work was partially supported by the National Science Council of Taiwan under grant 95-2221-E-001-029-MY3.

References

[1] K. Dembczynski, S. Greco, and R. Slowinski. Methodology of rough-set-based classification and sorting with hierarchical structure of attributes and criteria. *Control & Cybernetics*, 31:891–920, 2002.

[2] K. Dembczynski, S. Greco, and R. Slowinski. Second-order rough approximations in multi-criteria classification with imprecise evaluations and assignments. In D. Slezak, G. Wang, M.S. Szczuka, I. Duntsch, and Y. Yao, editors, *Proceedings of The Tenth International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, LNAI 3641, pages 54–63. Springer-Verlag, 2005.

[3] D. Dubois and H. Prade. Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, 17:191–209, 1990.

[4] T.F. Fan, W.C. Hu, and C.J. Liau. Decision logics for knowledge representation in data mining. In *Proceedings of the 25th Annual International Conference on Computer Software and Applications*, pages 626–631. IEEE Press, 2001.

[5] T.F. Fan, D.R. Liu, and G.H. Tzeng. Rough set-based logics for multicriteria decision analysis. *European Journal of Operational Research*, 182(1):340–355, 2007.

[6] S. Greco, B. Matarazzo, and R. Slowinski. Rough set approach to multi-attribute choice and ranking problems. In *Proceedings of the 12th International Conference on Multiple Criteria Decision Making*, pages 318–329, 1997.

[7] S. Greco, B. Matarazzo, and R. Slowinski. Rough approximation of a preference relation in a pairwise comparison table. In L. Polkowski and A. Skowron, editors, *Rough Sets in Data Mining and Knowledge Discovery*, pages 13–36. Physica-Verlag, 1998.

[8] S. Greco, B. Matarazzo, and R. Slowinski. Rough approximation of a preference relation by dominance relations. *European Journal of Operational Research*, 117(1):63–83, 1999.

[9] S. Greco, B. Matarazzo, and R. Slowinski. Extension of the rough set approach to multicriteria decision support. *INFOR Journal: Information Systems & Operational Research*, 38(3):161–195, 2000.

[10] S. Greco, B. Matarazzo, and R. Slowinski. Rough set theory for multicriteria decision analysis. *European Journal of Operational Research*, 129(1):1–47, 2001.

[11] S. Greco, B. Matarazzo, and R. Slowinski. Rough sets methodology for sorting problems in presence of multiple attributes and criteria. *European Journal of Operational Research*, 138(2):247–259, 2002.

[12] S. Greco, B. Matarazzo, and R. Slowinski. Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. *European Journal of Operational Research*, 158(2):271–292, 2004.

[13] S. Greco, B. Matarazzo, and R. Slowinski. Dominance-based rough set approach as a proper way of handling graduality in rough set theory. *Transactions on Rough sets*, VII:36–52, 2007.

[14] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, 1998.

[15] M. Kryszkiewicz. Properties of incomplete information systems in the framework of rough sets. In L. Polkowski and A. Skowron, editors, *Rough Sets in Knowledge Discovery*, pages 422–450. Physica-Verlag, 1998.

[16] M. Kryszkiewicz and H. Rybiński. Reducing information systems with uncertain attributes. In Z. W. Raś and M. Michalewicz, editors, *Proceedings of the 9th ISMIS*, LNAI 1079, pages 285–294. Springer-Verlag, 1996.

[17] M. Kryszkiewicz and H. Rybiński. Reducing information systems with uncertain real value attributes. In *Proceedings of the 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 1165–1169, 1996.

[18] W. Lipski. On databases with incomplete information. *Journal of the ACM*, 28(1):41–70, 1981.

[19] Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, 11(15):341–356, 1982.

[20] Z. Pawlak. *Rough Sets–Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, 1991.

[21] A.M. Radzikowska and E.E. Kerre. A comparative study of fuzzy rough sets. *Fuzzy Sets and Systems*, 126(2):137–155, 2002.

[22] H. Sakai, R. Ishibashi, and M. Nakata. Lower and upper approximations of rules in non-deterministic information systems. In C.C. Chan, J.W. Grzymala-Busse, and W. Ziarko, editors, *Proceedings of The 6th International Conference on Rough Sets and Current Trends in Computing*, LNAI 5306, pages 299–309. Springer-Verlag, 2008.

[23] D. Slezak. Rough sets and few-objects-many-attributes problem: The case study of analysis of gene expression data sets. In *Proceedings of the International Conference on Frontiers in the Convergence of Bioscience and Information Technologies*, pages 437–442. IEEE Press, 2007.

[24] R. Slowinski, S. Greco, and B. Matarazzo. Rough set analysis of preference-ordered data. In J.J. Alpigini, J.F. Peters, A. Skowron, and N. Zhong, editors, *Proceedings of the 3rd International Conference on Rough Sets and Current Trends in Computing*, LNAI 2475, pages 44–59. Springer-Verlag, 2002.

[25] Y.Y. Yao and Q. Liu. A generalized decision logic in interval-set-valued information tables. In N. Zhong, A. Skowron, and S. Ohsuga, editors, *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing*, LNAI 1711, pages 285–293. Springer-Verlag, 1999.

ISBN: 978-989-95079-6-8

Overlap index, overlap functions and migrativity

H. Bustince[1] J. Fernandez[1] R. Mesiar[2] J. Montero[3] R. Orduna[1]

1.Dpto. Automatica y Computacion, Universidad Publica de Navarra
Pamplona, Spain
2.Department of Mathematics and Descriptive Geometry, Slovak University of Technology,
SK-813 68 Bratislava, Slovakia
Institute of Information Theory and Automation, Czech Academy of Sciences,
CZ-182 08 Prague, Czech Republic
3.Facultad de Matematicas, Universidad Complutense de Madrid
Madrid, Spain
Email: {bustince,fcojavier.fernandez,raul.orduna}@unavarra.es, mesiar@math.sk, monty@mat.ucm.es

Abstract— *In this work we study overlap degrees expressed in terms of overlap functions. We present the basic properties that from our point of view must satisfy these overlap functions. We study a construction method, we analyze which t-norms are also overlap functions and we prove that if we apply particular aggregations to such functions we recover the overlap index between fuzzy sets as defined by Dubois, and the consistency index of Zadeh. We also consider some properties that can be required to overlap functions, as k-Lipschitzianity or migrativity .*

Keywords— Overlap degree, Overlap function, Overlap index, t-norm, Migrativity.

1 Overlap function. Definition and properties

Zadeh's fuzzy sets theory has been very useful for solving problems which are described by imprecise models and with a large amount of noise. In particular, this theory is very appropriate to study the problem of identifying the objects in an image (see [18, 26]).

To separate the object from the background in an image, the first thing to do is to represent the object by means of a fuzzy set and the background by means of another one. The success of the separation method lies on the correct choice of those fuzzy sets, which do not need to be disjoint in the sense of Ruspini [27] (see [1, 2, 17]).

To build these sets it is necessary to know the exact property that characterizes the pixels belonging to the object (background). This property determines the expression of the membership function associated to the fuzzy set representing the object (background)(see [8, 9]). Usually, this function is not precisely known. There are some pixels for which the expert is sure they belong to the object or the background, but there are other pixels for which the expert hesitates. It is for these last ones that the value of the membership function is not accurately known.

So, suppose that for a given an image, we ask an expert to assign to each pixel of intensity q the following values:
$\mu_B(q)$, representing the membership of the pixel to the background.
$\mu_O(q)$, representing the membership of the pixel to the object.

In Fig.1 we show the two membership functions provided by the expert to represent an image in an L gray-level scale. We can deduce that, for intensities less than q_i, the expert is sure that the pixels do not belong to the object. For intensities greater than q_j he is sure that the pixels do not belong to the background. However, for intensities between q_i and q_j the expert is not sure about the membership of the pixels, with intensity q_k corresponding to the maximum lack of knowledge.

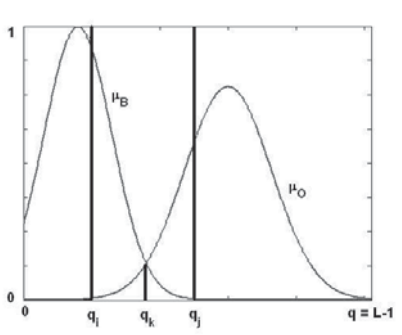


Figure 1: Overlap between two functions

From this analysis we deduce that the overlap degree between the two functions can be understood as a representation of the lack of knowledge of the expert when he has to settle if a given pixel belongs to the background or to the object. So we can define the overlap degree between $\mu_B(q)$ and $\mu_O(q)$ by means of an overlap function

$$G_S : [0,1] \times [0,1] \to [0,1]$$

such that:
(G_S0) G_S depends only on $\mu_B(q)$ and $\mu_O(q)$.
(G_S1) G_S is symmetric. The overlap degree does not depend on the order we consider the membership degrees.
(G_S2)
$G_S(\mu_B(q), \mu_O(q)) = 0$ if and only if
$\mu_B(q) = 0$ or $\mu_O(q) = 0$, (i.e., $\min(\mu_B(q), \mu_O(q)) = 0$) .
(G_S3)
$G_S(\mu_B(q), \mu_O(q)) = 1$ if and only if
$\mu_B(q) = 1$ and $\mu_O(q) = 1$, (i.e., $\min(\mu_B(q), \mu_O(q)) = 1$).
(G_S4) If the membership degrees increase, so does the overlap degree.
(G_S5) Continuity. The overlap degree between the two memberships associated to a given pixel must not react chaotically

ISBN: 978-989-95079-6-8

under small variations of the values of the membership to the background or to the object.

Apart from these five necessary properties, we consider it is also natural to require the following migrative property.

(G_S6) Migrativity. If we decrease $\mu_B(q)$ in a proportion $\alpha \in (0,1]$, the overlap degree G_S should decrease in the same amount as if we decrease $\mu_O(q)$ in the same proportion α. That is:

$$G_S(\alpha\mu_B(q), \mu_O(q)) = G_S(\mu_B(q), \alpha\mu_O(q)) \text{ for all } \alpha \in (0,1].$$

Definition 1 *A mapping $G_S : [0,1]^2 \to [0,1]$ is an overlap function if and only if it satisfies $(G_S0) - (G_S5)$. If G_S satisfies (G_S6) for any $0 < \alpha \leq 1$, we say that it is a migrative overlap function.*

Let's denote by $\mathcal{G}$ the set of overlap functions in the sense of Definition 1. Then the following result is immediate.

Theorem 1 *$(\mathcal{G}, \leq_{\mathcal{G}})$ with the ordering $\leq_{\mathcal{G}}$ defined for $G_1, G_2 \in \mathcal{G}$ by*

$$G_1 \leq_{\mathcal{G}} G_2 \text{ if and only if } G_1(x,y) \leq G_2(x,y) \qquad (1)$$

for all $x, y \in [0,1]$ is a lattice.

It is clear that the lattice $(\mathcal{G}, \leq_{\mathcal{G}})$ is not complete, neither has it top or bottom elements. On the other hand, it is closed under the action of appropriate aggregation functions, as shown next.

Definition 2 *An aggregation function of dimension n ([7, 11, 20, 16]) is a mapping $M : [0,1]^n \to [0,1]$ satisfying (see also [12, 3]):*
$M1$. $M(0, \cdots, 0) = 0$ and $M(1, \ldots, 1) = 1$.
$M2$. For any $(x_1, \cdots, x_n)$, $(y_1, \cdots, y_n) \in [0,1]^n$, if $x_i \leq y_i$ for any $i \in \{1, \cdots, n\}$, then $M(x_1, \cdots, x_n) \leq M(y_1, \cdots, y_n)$; that is, M is monotone increasing in all its arguments.

Theorem 2 *Let M be a binary aggregation function without zero divisors (that is, $M(x,y) = 0$ implies $min(x,y) = 0$) and such that $M(x,y) = 1$ only if $\max(x,y) = 1$. Then, $M(G_1, G_2) \in \mathcal{G}$ for any $G_1, G_2 \in \mathcal{G}$.*

2 Construction

The following theorem provides both a characterization and a construction method of overlap functions.

Theorem 3 *A mapping $G_S : [0,1]^2 \to [0,1]$ is an overlap function if and only if it can be written as*

$$G_S(x,y) = \frac{f(x,y)}{f(x,y) + h(x,y)} \qquad (2)$$

for some $f, h : [0,1]^2 \to [0,1]$ such that
1) f and h are symmetric;
2) f is non decreasing and h is non increasing;
3) $f(x,y) = 0$ if and only if $\min(x,y) = 0$;
4) $h(x,y) = 0$ if and only if $\min(x,y) = 1$;
5) f and h are continuous;

Example 1.
1) If $f(x,y) = \min(x,y)$ and $h(x,y) = \max(1-x, 1-y)$, then $G_S(x,y) = \min(x,y)$ is an overlap function.
2) If we take $f(x,y) = \sqrt{x.y}$ and $h(x,y) = \max(1-x, 1-y)$, then the construction proposed in Theorem 3 provides an overlap function.
3) If $f(x,y) = \sqrt{x \cdot y}$ and $h(x,y) = 1 - x \cdot y$, expression

$$G_S(x,y) = \frac{\sqrt{x \cdot y}}{\sqrt{x \cdot y} + 1 - x \cdot y} \qquad (3)$$

is an overlap function.

Corollary 1 *In the setting of Theorem 3, assume that G_S can be expressed in two different ways:*

$$G_S(x,y) = \frac{f_1(x,y)}{f_1(x,y) + h_1(x,y)} = \frac{f_2(x,y)}{f_2(x,y) + h_2(x,y)} \qquad (4)$$

for any $x, y \in [0,1]$. Let M be a binary continuous aggregation function that is homogeneous of order one, that is, such that

$$M(\lambda x, \lambda y) = \lambda M(x,y) \qquad (5)$$

for any $x, y \in [0,1]$ and any $\lambda \geq 0$ such that $\lambda x, \lambda y \in [0,1]$. Then, if we define $f(x,y) = M(f_1(x,y), f_2(x,y))$ and $h(x,y) = M(h_1(x,y), h_2(x,y))$ it also holds that

$$G_S(x,y) = \frac{f(x,y)}{f(x,y) + h(x,y)} \, . \qquad (6)$$

Proof. First observe that $f_i = h_i \frac{G_S}{1-G_S}$ for $i = 1, 2$. By the homogeneity condition on M, also $f = h \frac{G_S}{1-G_S}$ and the result follows.

2.1 Specific case: t-norms

In this subsection we study under which conditions we can assure a t-norm satisfies the properties required to an overlap function.

We know that a t-norm is a commutative, associative, increasing mapping $T : [0,1]^2 \to [0,1]$ such that $T(x,1) = x$ for any $x \in [0,1]$. So t-norms satisfy (G_S1) and (G_S4). They also satisfy (G_S3), since if $T(x,y) = 1$, as $T(x,y) \leq \min(x,y)$, it must be $x = y = 1$. The reciprocal is direct, taking into account that 1 is the neutral element of any t-norm. So it is only required to study condition (G_S2). Observe that when dealing with t-norms, the necessary condition in this property coincides with the definition of positive t-norm. Besides, the analysis of the conditions under which a t-norm satisfies (G_S2) leads to the following classification result.

Theorem 4 *If a t-norm T is an overlap function, then T is of one of the following three types:*
1) $T = \min$;
2) T is strict;
3) T is the ordinal sum of the family $\{([a_m, b_m], T_m)\}$, with all the T_m continuous Archimedean t-norms and such that if for some m_0 we have $a_{m_0} = 0$ then necessarily T_{m_0} is a strict t-norm.

Proof. By hypothesis, T is continuous. From the classification of continuous t-norms given in page 11 of [15] (see [25],

ISBN: 978-989-95079-6-8

[21]), we know that for a continuous t-norm T there are three possibilities:
1.- $T = \min$;
2.- T is Archimedean;
3.- There exists a family $\{([a_m, b_m], T_m)\}$ such that T is the ordinal sum of this family in the sense of [15].

As, by hypothesis, T is an overlap function, (G_S2) holds. If T is also Archimedean, we have that T is strict.

Suppose now that T is the ordinal sum of the family $\{([a_m, b_m], T_m)\}$; that is:

$$T(x,y) = \begin{cases} a_m + (b_m - a_m)T_m(\frac{x-a_m}{b_m-a_m}, \frac{y-a_m}{b_m-a_m}) \\ \qquad \text{if } (x,y) \in [a_m, b_m]^2 \ ; \\ \min(x,y) \text{ otherwise.} \end{cases} \tag{7}$$

We know that for any t-norm T, if $\min(x,y) = 0$ then $T(x,y) = 0$. Since our t-norm is an overlap function, also the reciprocal is true. So, if $T(x,y) = 0$ two things can happen:
a) (x,y) does not belong to any $[a_m, b_m]^2$. Then we have $T(x,y) = \min(x,y)$ for that (x,y).
b) (x,y) belongs to $[a_m, b_m]^2$. As $T(x,y) = 0 = a_m + (b_m - a_m)T_m(\frac{x-a_m}{b_m-a_m}, \frac{y-a_m}{b_m-a_m})$, we have $a_m = 0$ and $b_m \neq 0$ since otherwise the interval would be $[0,0]$ and $x = y = 0$. We know that T verifies (G_S2) and if $T(x,y) = 0$, then $T_m(\frac{x}{b_m}, \frac{y}{b_m}) = 0$, so T_m also verifies (G_S2). Hence, the continuous and Archimedean t-norm T_m associated to the interval $[0, b_m]$ also satisfies (G_S2), so it is strict.

Example 2. 1) In the construction of the following overlap function we use item 3) of Theorem 4 . We take as t-norm for the corresponding interval $[0, b_m]$ the product, which is strict, continuous and Archimedean.

$$G_S(x,y) = \begin{cases} 2xy \text{ if } (x,y) \in [0, 0.5]^2 \ ; \\ \min(x,y) \text{ in other case.} \end{cases} \tag{8}$$

2) In the construction of the following overlap function we take the Lukasiewicz and the product t-norms (see page 84 in [19]). Observe that now we do not consider any interval of the type $[0, b_m]$.

$$G_S(x,y) = \\ \begin{cases} 0.1 + 2.5(x - 0.1)(y - 0.1) \text{ if } (x,y) \in [0.1, 0.5]^2; \\ 0.7 + \max(x + y - 1.6, 0) \text{ if } (x,y) \in [0.7, 0.9]^2; \\ \min(x,y) \text{ in other case.} \end{cases} \tag{9}$$

3) The following t-norm does not satisfy (G_S2). This is due to the fact that in $[0, 0.25]^2$ we take Lukasiewicz t-norm, which is continuous and Archimedean, but it is not strict.

$$T(x,y) = \begin{cases} \max(x + y - 0.25, 0) \text{ if } (x,y) \in [0, 0.25]^2 \\ \min(x,y) \text{ in other case.} \end{cases} \tag{10}$$

3 Overlap functions and k-Lipschitzianity

In this section we consider a particular type of overlap functions, satisfying a sort of stronger continuity. We start adapting the definition of k-Lipschitz functions to the overlap function case.

Definition 3 *Let $k \geq 1$. An overlap function G_S is k-Lipschitz if for any $x, y, z, t \in [0,1]$ it holds*

$$|G_S(x,y) - G_S(z,t)| \leq k(|x - z| + |y - t|) \ . \tag{11}$$

It is worth remarking that the usual definition of k-Lipschitziniaty allows any value of k greater than zero. But, in the case of overlap functions, just by taking $x = y = z = 1$ and $t = 0$ the restriction to $k \geq 1$ becomes justified.

The set of k-Lipschitz overlap functions is bounded and its supremum can be easily determined, as the next result shows.

Theorem 5 *Let $k \geq 1$. Then the supremum of the set of k-Lipschitz overlap functions is given by the mapping $\min(kx, ky, 1)$, whereas the infimum is given by $\max(kx + ky - 2k + 1, 0)$.*

Proof. Suppose that $G_S(x,y) > \min(kx, ky, 1)$ for some $x, y \in [0,1]$. Since $G_S(x,y) \leq 1$, this means that $\min(kx, ky, 1) = kx$ or $\min(kx, ky, 1) = ky$. In the first case, $y = t = 1$ and $z = 0$ in Eq. (11), we arrive at

$$kx < G_S(x, 1) \leq kx \tag{12}$$

which is a contradiction. The second case is analogous. On the other hand, by defining for $\epsilon > 0$ the mappings

$$\max(x \cdot y, (1 - \epsilon))(\min(kx, ky, 1)) \tag{13}$$

we get a sequence of overlap functions which converges uniformly to $\min(kx, ky, 1)$ as $\epsilon \to 0$. The proof for the lower bound is similar.

The mapping $\max(kx + ky - 2k + 1, 0)$ is never an overlap function. On the contrary, although in general, the mapping $\min(kx, ky, 1)$ for $k > 1$ and $x, y \in [0,1]$ such that $kx, ky \in [0,1]$ does not define an overlap function, (since by taking $x = y = \frac{1}{k}$ we see that it does not fulfill condition (G_S3)), $\min(x,y)$ is an overlap function, so we have the following corollary.

Corollary 2 *The mapping $\min(x,y)$ is the strongest 1-Lipschitz overlap function, in the sense that for any other 1-Lipschitz overlap function G_S the inequality*

$$G_S(x,y) \leq \min(x,y) \tag{14}$$

holds for any $x, y \in [0,1]$.

For associative k-Lipschitz overlap functions we have the next result which can be derived from [23, 24].

Theorem 6 *If G_S is an associative k-Lipschitz overlap function, then G_S is a t-norm of the form given in Theorem 4, where each involved strict t-norm T (see item 2) or item 3)) has a k-convex additive generator t, i.e.,*

$$t(y + k\epsilon) - t(y) \leq t(x + \epsilon) - t(x) \tag{15}$$

for all $0 \leq y \leq x < 1$ and $\epsilon \in]0, \min(1 - x, (1 - y)/k)[$.

4 Overlap index. Construction from overlap functions

In this section we are going to build overlap indexes by aggregating overlap degrees. We start recalling the concepts of overlap index and consistency. Then we justify why most of the overlap indexes in the literature do not satisfy one of the four conditions required by Dubois and Prade. Next we show the properties we have to demand to aggregation functions so

ISBN: 978-989-95079-6-8

that, when applied to the previously studied overlap degrees, we get overlap indexes. We finish showing how to construct such indexes and, in particular, how to recover the two most used expressions.

We denote by $F(U)$ the set of all fuzzy sets over the finite, non empty referential U ($Cardinal(U) = n$). We are going to represent the fuzzy sets over U in the following way:

$$A = \{(u, \mu_A(u)) | u \in U\} \tag{16}$$

In 1978 Zadeh [29] presented the natural extension to the fuzzy set theory of the classical concept of overlap, which he called consistency:

$$O(A, B) = \sup_{i=1}^{n}(\min(\mu_A(u_i), \mu_B(u_i))). \tag{17}$$

Clearly, $O(A, B) = 0$ if A and B are completely disjoint, and $O(A, B) = 1$ if there is $u_i \in U$ such that $\mu_A(u_i) = \mu_B(u_i) = 1$.

In 1982 Dubois and Prade [13] presented the following axiomatization for the overlap index:

Definition 4 *An overlap index is a function $O(A, B)$ from $F(U) \times F(U)$ on the unit interval such that:*
$(O1)$ $O(A, B) = 0$ if and only if A and B have disjoint supports;
$(O2)$ $O(A, B) = 1$, if ($\mu_A(u_i) = 0$ or $\mu_B(u_i) = 1$) or ($\mu_A(u_i) = 1$ or $\mu_B(u_i) = 0$);
$(O3)$ $O(A, B) = O(B, A)$;
$(O4)$ If $B \leq C$, then $O(A, B) \leq O(A, C)$.

Condition $(O2)$ in this definition presents the advantage of that, if A is not fuzzy, then $O(A, A) = 1$. But Dubois, Ostasiewicz and Prade in [14] settled the following:

1. For subnormal fuzzy sets (i.e., $\mu_A(u_i) < 1$ and $\mu_B(u_i) < 1$ for any $u_i \in U$), $(O2)$ *must be ignored.*
2. The ROC index(see [13]) does not fulfill $(O2)$.

It is also interesting to notice that if $A = \{(u_i, \mu_A(u_i) = 0) | u_i \in U\}$, then $\min(A, A) = \{(u_i, \mu_{\min(A,A)}(u_i) = 0) | u_i \in U\}$ and from $(O1)$, $O(A, A) = 0$. If we also impose $(O2)$, then $O(A, A) = 1$. So we get a contradiction.

Due to all these considerations, usually only conditions $(O1)$,$(O3)$ and $(O4)$ from Def. 4 are required to overlap indexes.

In the following theorem we present a construction method of overlap indexes, by means of aggregation functions.

Theorem 7 *Let $M : [0, 1]^n \to [0, 1]$ be an aggregation function being idempotent and such that $M(x_1, \ldots, x_n) = 0$ if and only if $x_1 = \ldots = x_n = 0$.*
Let $G_S : [0, 1]^2 \to [0, 1]$ be a mapping and consider:
$O : F(U) \times F(U) \to [0, 1]$ *defined as*
$O(A, B) = M_{i=1}^{n}(G_S(\mu_A(u_i), \mu_B(u_i)))$
Then the following items hold:
i) O verifies $(O1)$ if and only if G_S verifies (G_S2);
ii) O verifies $(O3)$ if and only if G_S verifies (G_S1);
iii) If G_S verifies (G_S4), then O verifies $(O4)$.

Example 3.

1. $O(A, B) = \max_{i=1}^{n} G_S(\mu_A(u_i), \mu_B(u_i))$
2. $O(A, B) = \left(\frac{1}{n}\sum_{i=1}^{n}(G_S(\mu_A(u_i), \mu_B(u_i))^\beta\right)^{\frac{1}{\beta}}, \beta \neq 0$

Notes for the Example

1. If in 1. we take $G_S(x, y) = \min(x, y)$, then we recover Zadeh's consistency index (see [29]).
2. If there is a single u_i such that $G_S(\mu_A(u_i), \mu_B(u_i)) = 1$, then, by 1. we have $O(A, B) = 1$. This fact of a single element making $O(A, B) = 1$ suggest us to use 2. instead of 1.
3. Expression 2. satisfies:

   $O(A, B) = 1$ if and only if $G_S(\mu_A(u_i), \mu_B(u_i)) = 1$

   for all $i = 1, \ldots, n$. In these conditions, if $G_S(x, y) = \min(x, y)$ or $G_S(x, y) = x \cdot y$ or $G_S(x, y) = \sqrt{x \cdot y}$, then $O(A, B) = 1$ if and only if

   $$A = B = \{(u_i, \mu_A(u_i) = \mu_B(u_i) = 1) | u_i \in U\} \tag{18}$$

   so 2., as most of the expressions of overlap indexes (see [4, 5, 6, 14, 13]), does not fulfill condition $(O2)$ in Definition 2.

Corollary 3 *In the setting of Theorem 7, if we demand M to satisfy $M(x_1, \ldots, x_n) = 1$ if and only if $x_1 = \ldots = x_n = 1$, then, if G_S verifies (G_S3) we have*
$O(A, B) = 1$ if and only if $A = B = \{(u_i, \mu_A(u_i) = 1) | u_i \in U\}$.

5 Construction of migrative overlap functions

As we have already said, migrativity seems to be quite a natural property to be required to overlap functions. In [10], an in-deep study of the migrativity property is carried on for general aggregation functions. In this paper we use the following results that are proved there.

Lemma 1 *Let $H : [0, 1]^2 \to [0, 1]$ be a binary function. Then H is migrative if and only if $H(x, y) = H(1, x \cdot y)$, for all $x, y \in [0, 1]$.*

Lemma 2 *A function $H : [0, 1]^2 \to [0, 1]$ is migrative if and only if there exists $h : [0, 1] \to [0, 1]$ such that $H(x, y) = h(x \cdot y)$, for all $x, y \in [0, 1]$.*

In this section we present a characterization theorem of migrative overlap functions. Clearly there are overlap functions that are not migrative (for instance, those in Ex. 1). In the following results we prove that there exist also overlap functions which are migrative.

Theorem 8 *Let $H : [0, 1]^2 \to [0, 1]$ be a migrative mapping (not necessarily an overlap function). Then*

$$H(x, 1) = H(\sqrt{x}, \sqrt{x}) \tag{19}$$

for each $x \in [0, 1]$.

ISBN: 978-989-95079-6-8

Proof. By Lemmas 1 and 2 there exists a mapping $h : [0,1] \to [0,1]$ such that $H(x,1) = h(x) = h(\sqrt{x} \cdot \sqrt{x}) = H(\sqrt{x}, \sqrt{x})$.

Theorem 9 *A mapping $G_S : [0,1] \to [0,1]$ is a migrative overlap function if and only if there exists a non decreasing function $g : [0,1] \to [0,1]$ satisfying $g^{-1}((0,1)) = (0,1)$ such that*

$$G_S(x,y) = g(x \cdot y). \tag{20}$$

Proof. (Necessity) Since G_S is migrative, by Lemma 2 we know that there exists a function $g : [0,1] \to [0,1]$ such that $G_S(x,y) = g(x \cdot y)$ for all $x, y \in [0,1]$. As G_S is an overlap function, g is not decreasing and continuous. Besides G_S satisfies $(G_S 2)$ and $(G_S 3)$, so:

$$g(x) = g(x \cdot 1) = G_S(x,1) = 0 \text{ if and only if } x = 0 \tag{21}$$

$$g(x) = g(x \cdot 1) = G_S(x,1) = 1 \text{ if and only if } x = 1 \tag{22}$$

(Sufficiency) By Lemma 2 we have that $G_S(x,y) = g(x,y)$ satisfies $(G_S 6)$. From the migrativity it is clear that $(G_S 1)$ holds. On the other hand:

$$G_S(x,y) = 0 = g(x \cdot y) \text{ if and only if } x \cdot y = 0 \tag{23}$$

$$G_S(x,y) = 1 = g(x \cdot y) \text{ if and only if } x \cdot y = 1 \tag{24}$$

Clearly, G_S satisfies $(G_S 4)$ and $(G_S 5)$ since g is non decreasing and continuous.

Acknowledgements. H. Bustince, J. Fernandez and R. Orduna have been supported by project TIN2007-65981. R. Mesiar has been supported by grants APVV-0012-07 and GACR 402/08/0618. J. Montero has been supported by project TIN2006-06190.

References

[1] A. Amo, D. Gómez and J. Montero, Spectral fuzzy classification: a supervised approach, *Mathware and Soft Computing*, **10**, Pages 141-154, 2003.

[2] A. Amo, J. Montero, G. Biging and V. Cutello, Fuzzy classification systems, *European Journal of Operational Research*, **156**, Pages 459-507, 2004.

[3] A. Amo, J. Montero, D. Gómez and E. Molina, Representation of consistent recursive rules, *European Journal of Operational Research*, **130**, Pages 29-53, 2001.

[4] H. Bustince, M. Pagola, E. Barrenechea, Construction of fuzzy indices from fuzzy DI-subsethood measures: Application to the global comparison of images, *Information Sciences*, 177, **3**, Pages 906-929, 2007.

[5] H. Bustince, E. Barrenechea, M. Pagola, et al., Weak fuzzy S-subsethood measures. Overlap index, *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, 14 **5**, Pages. 537-560, 2006.

[6] H. Bustince, V. Mohedano, E. Barrenechea, et al. Definition and construction of fuzzy DI-subsethood measures, *Information Sciences*, 176, **21**, Pages 3190-3231, 2006.

[7] H. Bustince H., J. Montero, E. Barrenechea and M. Pagola, Semiautoduality in a restricted family of aggregation operators, *Fuzzy Sets and Systems*, 158, **12**, Pages 1360-1377, 2007.

[8] H. Bustince, E. Barrenechea, M. Pagola, Restricted equivalence functions, *Fuzzy Sets and Systems*, 157 **17**, Pages 2333-2346, 2006.

[9] H. Bustince, E. Barrenechea, M. Pagola, Image thresholding using restricted equivalence functions and maximizing the measures of similarity, *Fuzzy Sets and Systems*, 158, **5**, Pages 496-516, 2007.

[10] Bustince H., Montero J. and Mesiar R., Migrativity of Aggregation Operators, *Fuzzy Sets and Systems*, 160, **6**, Pages 766-777, 2008.

[11] T. Calvo, A. Kolesárová, M. Komorníková and R. Mesiar: Aggregation operators: properties, classes and construction methods. In T. Calvo, G. Mayor and R. Mesiar (Eds.): *Aggregation Operators New Trends and Applications* (Physica-Verlag, Heidelberg); Pages 3-104, 2002.

[12] V. Cutello and J. Montero, Recursive connective rules, *International Journal of Intelligent Systems*, 14, Pages 3-20, 1999.

[13] D. Dubois and J.L. Koning, Social choice axioms for fuzzy set aggregation. *Fuzzy Sets and Systems* 58, Pages 339-342, 1991.

[14] D. Dubois, W. Ostasiewicz, H. Prade, Fuzzy Sets: History and Basic Notions, in: D. Dubois, H. Prade (Eds), *Fundamentals of Fuzzy Sets*, Kluwer, Boston, MA, Pages 21-124, 2000.

[15] J. Fodor and M. Roubens, Fuzzy Preference Modelling and Multicriteria Decision Support, in: *Theory and Decision Library* (Kluwer Academic Publishers) 1994.

[16] D. Gómez and J. Montero: A discussion on aggregation operators. *Kybernetika* 40, Pages 107-120, 2004.

[17] D. Gómez, J. Montero and H. Bustince, On the relevance of some families of fuzzy sets, *Fuzzy Sets and Systems*, 158, Pages 2429-2442, 2007.

[18] L.K. Huang and M.J. Wang, Image thresholding by minimizing the measure of fuzziness, *Pattern recognition*, 28, **1**, Pages 41-51, 1995.

[19] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms, Trends in Logic. *Studia Logica Library*, Kluwer Academic Publishers, Dordrecht, 2000.

[20] G.J. Klir and T.A. Folger: *Fuzzy Sets, Uncertainty and Information.* Prentice Hall, Englewood Cliffs, NJ, 1988.

[21] C.H. Ling, Representation of associative functions, *Publ. Math. Debrecen*, 12, Pages 189-212, 1965.

[22] R. Mesiar, V. Novák, Open problems from the 2nd International conference of fuzzy sets theory and its applications, *Fuzzy Sets and Systems*, 81, Pages 185-190, 1996.

ISBN: 978-989-95079-6-8

[23] A.Mesiarová, A note on two open problems of Alsina, Frank and Schweizer. *Aequationes Math.*, 72, no. 1-2, Pages 41-46, 2006.

[24] A. Mesiarová, $k - l_p$-Lipschitz t-norms. *Int. J. Approximate Reasoning*, 46, Pages 596-604, 2007.

[25] P.S. Mostert, A.L. Shields, On the structure of semigroups on a compact manifold with boundary, *Ann. of Math.*, 65, Pages 117-143, 1957.

[26] N. R. Pal and S. K. Pal, A review of image segmentation techniques, *Pattern recognition*, 26, Pages 1277-1294, 1993.

[27] E.H. Ruspini, A new approach to clustering, *Information and Control*, 15, Pages 22-32, 1969.

[28] L.A. Zadeh, Fuzzy sets, *Information Control*, 8, Pages 338-353, 1965.

[29] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems*, 1,Pages 3–28, 1978.

ISBN: 978-989-95079-6-8

On 'family resemblances' with fuzzy sets*

E. Trillas, C. Moraga[1] A. Sobrino[2]

[1] European Centre for Soft Computing
Mieres (Asturias), Spain
[2] University of Santiago de Compostela
Santiago de Compostela, Spain
Email: {enric.trillas, claudio.moraga}@softcomputing.es, alejandro.sobrino@usc.es

Abstract— *This paper takes into account the Wittgenstein's idea on family resemblances as a particular crisp relation between some fuzzy sets, that is, between some predicates representable from its use. It is shown that all uses of the same predicate actually do have family resemblance, that some pairs of predicates cannot, and a numerical degree of family resemblance is introduced.*

Keywords— Degree of family resemblance, Family resemblance, Fuzzy sets.

1 Introduction

Ludwig Wittgenstein headed two influential traditions in the so-called philosophy of language, that were originated by his famous books, *Tractatus logico-philosophicus* (1922, [1]), and *Philosophical Investigations* (1953, [2]), respectively.

The *Tractatus* does not properly deal with ordinary language, but with the logical analysis of propositions built up from atomic propositions, considered as 'pictures' of facts and keeping a strict correspondence with the world, understood through the totality of facts. Instead, the *Philosophical Investigations* in which Wittgenstein abandoned logical analysis, meant a shift in conferring a main role to the ways of designating facts, as an activity-oriented perspective on language. What is central at this respect, is that language does not primarily consist on describing the facts, but on playing 'language's games', or ways of dynamically using words to define their meaning, and that are to be described. In order to fix the meaning of a word, to show how a word works, it should be placed in the context and environment it is used. Language is not yet defined through propositions that, now, in Wittgenstein's view, come from their function in a language's game, and to note the absence of boundaries for describing such use of words, Wittgenstein introduced the term 'family resemblances'. Of course, Wittgenstein's idea on 'family resemblances', is broader than the relation of *family resemblance* in next section.

Fuzzy logic manages the extensional meaning of predicates through its use, once captured by the corresponding membership functions. This paper is nothing else than a first approach, in the path towards Zadeh's *Computing With Words*, to introduce 'family resemblances' between full-normalized fuzzy sets, and a numerical degree for measuring such relation.

Usually, full-normalized fuzzy sets, $\mu_P \in [0,1]^X$, those for which there are $x, y \in X$ such that $\mu_P(x) = 1$, $\mu_P(y) = 0$, appear as 'data' in the modeling of fuzzy systems, and they are neither self-contradictory ($\mu_P \leq \mu'_P$), nor negatively self contradictory ($\mu'_P \leq \mu_P$). Of course, μ'_P denotes the fuzzy set corresponding to 'not p'. Points x verifying $\mu_P(x) = 1$ can be taken as *the prototypes of* P *in* X, and points y verifying $\mu_P(y) = 0$ as *the anti-prototypes of* P *in* X.

In its own nature, this paper is not a conclusive one, but only a tentative to reflect the potentially interesting subject of the family resemblances shown by 'data' fuzzy sets. That is, by fuzzy sets with prototypes and anti-prototypes.

2 Family resemblance of fuzzy sets

Let $P, Q, ...$ be predicates on a universe of discourse X, such that their use, or meaning, is described by fuzzy sets $\mathbb{P}, \mathbb{Q}, ...$, given by membership functions $\mu_P, \mu_Q, ...$ in $[0,1]^X$.
For each membership function μ in $[0,1]^X$, define the sets of

- its 0-points, $Z(\mu) = \{x \in X; \mu(x) = 0\}$
- its 1-points, $S(\mu) = \{x \in X; \mu(x) = 1\}$

Definition 2.1 *With* $X \subset \mathbb{R}$*, the relation of family resemblance,* **fr** $\subset [0,1]^X \times [0,1]^X$*, is defined by* $(\mu, \sigma) \in$ **fr** *if and only if,*

1. $Z(\mu) \cap Z(\sigma) \neq \emptyset$, $S(\mu) \cap S(\sigma) \neq \emptyset$
2. μ *is non-decreasing in* $A \subset X$ *iff* σ *is non-decreasing in* A.
3. μ *is decreasing in* $A \subset X$ *iff* σ *is decreasing in* A.

In this definition, both 'decreasing' and 'non-decreasing', are not in strict sense, but allowing some constant pieces that can be taken as the first, or the second, by following what happens before or after.

Notice that the binary relation **fr** is only predicable between full-normalized fuzzy sets, that is, such that $Z(\mu) \neq \emptyset$, and $S(\mu) \neq \emptyset$. Denote

$$\mathfrak{F}^*(X) = \{\mu \in [0,1]^X - \{0,1\}^X; Z(\mu) \neq \emptyset, S(\mu) \neq \emptyset\}.$$

*This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias, Spain), and CICYT (Spain) under project TIN2008-06890-C02-01

ISBN: 978-989-95079-6-8

It should be pointed out that fuzzy sets μ in $\{0,1\}^X$ (crisp sets) are excluded, since such functions μ are neither decreasing, nor non-decreasing, but only piecewise constant with the values 0 or 1. In addition, in many cases if $(\mu,\sigma) \in \mathbf{fr}$, with $\mu,\sigma \in \{0,1\}^X$, it should be $\mu = \sigma$. Obviously, also constant fuzzy sets μ_r ($\mu_r(x) = r$, with $r \in [0,1], x \in X$) are not in $\mathfrak{F}^*(X)$, since $Z(\mu_r) = \emptyset$, or $S(\mu_r) = \emptyset$.
Of course, if σ results from a translation of μ keeping $Z(\mu) \cap Z(\sigma) \neq \emptyset$, and $S(\mu) \cap S(\sigma) \neq \emptyset$, it is obvious that $(\mu,\sigma) \in \mathbf{fr}$.

Remark 2.2 *Any non full-normalized fuzzy set, represented by a non-constant but continuous membership function $\mu \in [0,1]^{\mathbb{R}}$, can be re-scaled to a full-normalized one μ^* by*

$$\mu^* = \frac{\mu - \min(\mu)}{\max(\mu) - \min(\mu)}$$

Obviously, it is $Z(\mu^*) \neq \emptyset, S(\mu^*) \neq 0$, that is $\mu^* \in \mathfrak{F}^*(\mathbb{R})$, and μ^* is non-decreasing (decreasing) if and only if μ is non-decreasing (decreasing), but μ^* and μ can differ in the respective slopes (provided they have derivatives, it is $\mu^{*\prime} = \mu' / \max(\mu) - \min(\mu)$.

Example 2.3

1. *Fuzzy sets μ,σ in figure 1 verify $(\mu,\sigma) \in \mathbf{fr}$, since $Z(\mu) \cap Z(\sigma) = [0,2] \neq \emptyset$, and $S(\mu) \cap S(\sigma) = \{10\} \neq \emptyset$, and both are non-decreasing in $X = [0,10]$.*
2. *Fuzzy set μ in figure 1, and fuzzy set λ in figure 2, verify $(\mu,\lambda) \notin \mathbf{fr}$, since $S(\mu) \cap S(\lambda) = \emptyset$, and λ is decreasing in $[0,2]$, but μ is not.*

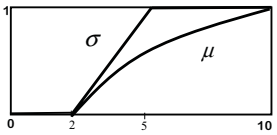


Figure 1:

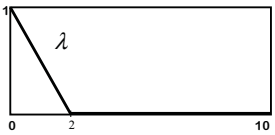


Figure 2:

Theorem 2.4 **fr** *is a reflexive and symmetric relation in $\mathfrak{F}^*(X)$.*

Proof. It is immediate to check that **fr** does verify the reflexive and symmetric properties. □

As it is intuitive, crisp relation **fr** is not transitive. A counterexample is given by membership functions μ,σ,λ in the figures

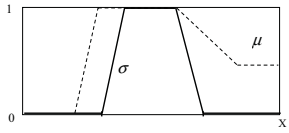


Figure 3:

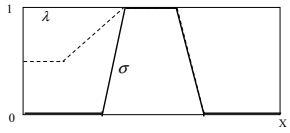


Figure 4:

that verify $(\mu,\sigma) \in \mathbf{fr}$, and $(\sigma,\lambda) \in \mathbf{fr}$, but $(\mu,\lambda) \notin \mathbf{fr}$, since $Z(\mu) \cap Z(\lambda) = \emptyset$.

Hence, if sets $[\mu] = \{\sigma \in \mathfrak{F}^*(X); (\mu,\sigma) \in \mathbf{fr}\}$ are not empty and cover $\mathfrak{F}^*(X)$, because it is $\bigcup_{\mu \in \mathfrak{F}^*(X)} [\mu] = \mathfrak{F}^*(X)$, they do not give a partition of $\mathfrak{F}^*(X)$, and the quotient $\mathfrak{F}^*(X)/\mathbf{fr}$ does not exist. For example, in the above figures, it is $\sigma \in [\mu] \cap [\lambda']$. Anyway, $[\mu]$ can be called the *family of* $\mu \in \mathfrak{F}^*(X)$, and $\sigma \in [\mu]$ a *relative* of μ, although μ could have relatives in other families, like it happens in people's families.

Theorem 2.5 *For no complement, μ' of $\mu \in \mathfrak{F}^*(X)$, is $(\mu,\mu') \in \mathbf{fr}$.*

Proof. If for $x,y \in A$, $x \leq y$, and $\mu(x) \leq \mu(y)$, it is $\mu'(y) \leq \mu'(x)$. □

Theorem 2.6 *For no opposite, or antonym, $'\mu$ of $\mu \in \mathfrak{F}^*(X)$, given by a symmetry α in X, $'\mu = \mu \circ \alpha$, is $(\mu, {}'\mu) \in \mathbf{fr}$.*

Proof. If $x \leq y$, and $\mu(x) \leq \mu(y)$, it follows $\alpha(y) \leq \alpha(x)$, and $\mu(\alpha(y)) \leq \mu(\alpha(x))$, or $\mu \circ \alpha(y) \leq \mu \circ \alpha(x)$, that is ${}'\mu(y) \leq' \mu(x)$. (see [3]) □

Theorem 2.7 *Let $u : X \to X$ be a bijective mapping such that*

- *If $x \leq y$, then $u(x) \leq u(y)$*
- $u^{-1}(Z(\mu)) \subset Z(\mu)$
- $u^{-1}(S(\mu)) \subset S(\mu)$

for $\mu \in \mathfrak{F}^*(X)$, $X \subset \mathbb{R}$. It is $(\mu, \mu \circ u^{-1}) \in \mathbf{fr}$.

Proof. It is $Z(\mu) \cap Z(\mu \circ u^{-1}) \neq \emptyset$, since $x \in Z(\mu)$, or $\mu(x) = 0$, implies $u^{-1}(x) \in Z(\mu)$, or $\mu(u^{-1}(x)) = \mu \circ u^{-1}(x) = 0$, it is, $x \in Z(\mu \circ u^{-1})$. Analogously, $x \in S(\mu)$, or $\mu(x) = 1$, that is, $x \in S(\mu \circ u^{-1})$. Hence, $Z(\mu) \cap Z(\mu \circ u^{-1}) \neq \emptyset$, and $S(\mu) \cap S(\mu \circ u^{-1}) \neq \emptyset$. If μ is non-decreasing, '$x \leq y \Rightarrow \mu(x) \leq \mu(y)$, from $u^{-1}(x) \leq u^{-1}(y)$, follows $\mu(u^{-1}(x)) \leq \mu(u^{-1}(y))$. Analogously, if μ is decreasing, so it is $\mu \circ u^{-1}$. □

In particular, if Q represented by $\mu_Q = \mu_P \circ u^{-1}$ is a synonym of P (see[4]), then $(\mu_P, \mu_Q) \in \mathbf{fr}$.

Example 2.8

1. *Fuzzy sets μ,σ in figure 5 verify $\sigma = \mu \circ \alpha$, with the symmetry $\alpha(x) = 10 - x$, and represent two antonyms. Obviously, μ and $\mu \circ \alpha$ do not show family resemblance.*

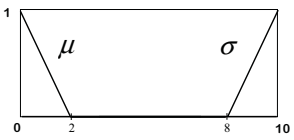


Figure 5:

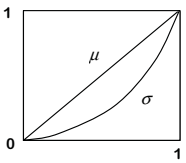


Figure 6:

2. *Fuzzy sets μ,σ in figure 6 verify $\sigma = \mu \circ u^{-1}$, with $u(x) = \sqrt{x}$, and show family resemblance.*

3 Degree of family resemblance

Obviously, if σ_1, σ_2 are in the family $[\mu]$, they do verify $(\mu, \sigma_1) \in$ **fr**, and $(\mu, \sigma_2) \in$ **fr**, but this does not mean that σ_1, σ_2 show the same extent of family resemblance with μ. A way of measuring such extent is by means of a convenient T-indistinguishability (see([5])) DR : **fr** $\rightarrow [0,1]$, defining,

If $(\mu, \sigma) \in$ **fr**, the degree up to which μ resembles σ is $DR(\mu, \sigma) \in [0,1]$.

3.1

Any prod-indistinguishability E in $\mathfrak{F}^*(X)$, can be represented by Ovchinnikov's theorem (see [6])

$$E(\mu,\sigma) = \inf_{f\in F}(\frac{f(\mu)}{f(\sigma)}, \frac{f(\sigma)}{f(\mu)}),$$

with F a family of functions $f : \mathfrak{F}^*(X) \rightarrow \mathbb{R}^+ - \{0\}$.

Provided X is a closed interval in $\mathbb{R}$, consider only the functions $\mu \in \mathfrak{F}^*(X)$ that are Riemann-integrable in X, and such that both $Z(\mu)$ and $S(\mu)$ are intervals in X. Denote by $\mathfrak{F}^{**}(X)$ this subset of functions in $\mathfrak{F}^*(X)$, and take

$$f_1(\mu) = \int_X \mu(x)dx, \quad f_2(\mu) = \text{length of the interval } X - Z(\mu).$$

Then,

$$Ovch(\mu,\sigma) = \min(\frac{f_1(\mu)}{f_1(\sigma)}, \frac{f_1(\sigma)}{f_1(\mu)}, \frac{f_2(\mu)}{f_2(\sigma)}, \frac{f_2(\sigma)}{f_2(\mu)}),$$

for all $\mu, \sigma \in \mathfrak{F}^{**}(X)$, is a prod-indistinguishability in the part of **fr** in $\mathfrak{F}^{**}(X) \times \mathfrak{F}^{**}(X)$.

3.2

If $X = [a,b] \subset \mathbb{R}$, define

$$I(\mu,\sigma) = 1 - \frac{1}{b-a}|\int_X \mu(x)dx - \int_X \sigma(x)dx|,$$

for all $\mu, \sigma \in \mathfrak{F}^{**}(X)$.

Function I is a W-indistinguishability (W is the Łukasiewicz t-norm), since:

- $I(\mu,\mu) = 1$, and $I(\mu,\sigma) = I(\sigma,\mu)$, for all $\mu, \sigma \in \mathfrak{F}^{**}(X)$.
- $W(I(\mu,\sigma), I(\sigma,\lambda)) = \max(0, 1 - \frac{1}{b-a}[|\int_X \mu(x)dx - \int_X \sigma(x)dx| + |\int_X \sigma(x)dx - \int_X \lambda(x)dx|]) \leq \max(0, 1 - \frac{1}{b-a}|\int_X \mu(x)dx - \int_X \lambda(x)dx|) = 1 - \frac{1}{b-a}|\int_X \mu(x)dx - \int_X \lambda(x)dx| = I(\mu,\lambda)$, for all $\mu, \sigma, \lambda \in \mathfrak{F}^{**}(X)$, because of the triangular inequality, $|\int_X \mu(x)dx - \int_X \lambda(x)dx| \leq |\int_X \mu(x)dx - \int_X \sigma(x)dx| + |\int_X \sigma(x)dx - \int_X \lambda(x)dx|$.

3.3

Since it is $Ovch(\mu,\sigma) \neq I(\mu,\sigma)$ and, in general, $Ovch(\mu,\sigma) < I(\mu,\sigma)$, a better degree could be obtained with a mean like,

$$DR(\mu,\sigma) = \frac{r_1 Ovch(\mu,\sigma) + r_2 I(\mu,\sigma)}{r_1 + r_2},$$

$$\text{with } r_1, r_2 \in \mathbb{R}^+ - \{0\},$$

of which the only DR that is symmetric is the one with $r_1 = r_2 = \frac{1}{2}$ (arithmetic mean). In addition, and since $W \leq prod$, $Ovch$ is also W-transitive: $W(Ovch(\mu,\sigma), Ovch(\sigma,\lambda)) \leq Ovch(\mu,\sigma) \cdot Ovch(\sigma,\lambda) \leq Ovch(\mu,\lambda)$. Thus, also DR with $r_1 = r_2 = \frac{1}{2}$ is W-transitive, since:

$$W(DR(\mu,\sigma), DR(\sigma,\lambda)) =$$

$$\max(0, \frac{Ovch(\mu,\sigma) + Ovch(\sigma,\lambda) - 1}{2} +$$

$$\frac{I(\mu,\sigma) + I(\sigma,\lambda) - 1}{2}) \leq$$

$$\max(0, \frac{Ovch(\mu+\lambda)}{2} + \frac{I(\mu+\lambda)}{2}) \leq$$

$$\max(0, DR(\mu,\lambda)) = DR(\mu,\lambda),$$

because $Ovch(\mu,\sigma) + Ovch(\sigma,\lambda) - 1 \leq \max(0, Ovch(\mu,\sigma) + Ovch(\sigma,\lambda) - 1)$, and $I(\mu,\sigma) + I(\sigma,\lambda) - 1 \leq \max(0, I(\mu,\sigma) + I(\sigma,\lambda) - 1)$. Hence, we will take

$$DR(\mu,\sigma) = \frac{Ovch(\mu,\sigma) + I(\mu,\sigma)}{2},$$

as the *index of family resemblance*.

Remark 3.1 *Obviously, functions DR are not only applicable to pairs in* **fr***, but also to all pairs in $\mathfrak{F}^{**}(X) \times \mathfrak{F}^{**}(X)$*

3.4

Let E be a T-indistinguishability relation. The crisp relation

$$\mu \equiv \sigma \Leftrightarrow E(\mu,\sigma) > 0,$$

is reflexive, and symmetric. Obviously, and provided $T = \min$, or $T = prod$, it is also transitive. Nevertheless, if $T = W$, since "$E(\mu,\sigma) > 0$ and $E(\sigma,\lambda) > 0$" is equivalent to the existence of $\varepsilon > 0$ such that "$E(\mu,\sigma) \geq \varepsilon$ and $E(\sigma,\lambda) \geq \varepsilon$", it is

$$W(E(\mu,\sigma), E(\sigma,\lambda)) \geq W(\varepsilon,\varepsilon) = \max(0, 2\varepsilon - 1), \text{ or}$$

$$E(\mu,\lambda) \geq \max(0, 2\varepsilon - 1).$$

To have $E(\mu,\lambda) > 0$, it is sufficient that $2\varepsilon - 1 > 0$, or $\varepsilon > 0.5$. In this case, $\mu \equiv \sigma$ and $\sigma \equiv \lambda$ imply $\mu \equiv \lambda$ (transitivity). Hence $E(\mu,\sigma) > 0.5$ allows to take μ and σ as 'equivalent'.

If $0.5 < Ovch(\mu,\sigma)$, and $0.5 < I(\mu,\sigma)$, it is $\mu \equiv \sigma$ for both T-indistinguishabilities $E = Ovch$ and $E = I$. Thus, if in addition to $(\mu,\sigma) \in$ **fr**, is $0.5 < \varepsilon \leq Ovch(\mu,\sigma)$, and $0.5 < \delta \leq I(\mu,\sigma)$, it is also $0.5 < \frac{\varepsilon+\delta}{2} \leq DR(\mu,\sigma)$, implying $\mu \equiv \sigma$ for $E = DR$. For example if $0.7 \leq Ovch(\mu,\sigma)$, and $0.7 \leq I(\mu,\sigma)$, it is $0.7 \leq DR(\mu,\sigma)$. In these cases, *it can be said that μ and σ show high family resemblance.*

ISBN: 978-989-95079-6-8

When $(\mu,\sigma) \in$ **fr**, if e.gr., $Ovch(\mu,\sigma) \geq 0.8$ and $I(\mu,\sigma) \geq 0.8$ implying $DR(\mu,\sigma) \geq 0.8$, there is a so high family resemblance between μ and σ that μ and σ can be taken as strongly representing similar uses of the same predicate (linguistic label).

Example 3.2

1. *Functions μ,σ in figure 7, do obviously verify $(\mu,\sigma) \in$* **fr**. *For them, $f_1(\mu) = 5$, $f_2(\mu) = 8$, $f_1(\sigma) = 4$ and $f_2(\sigma) = 7$,*

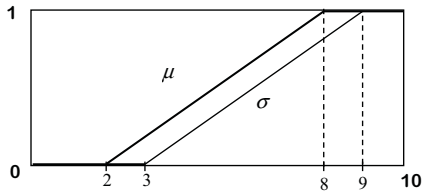


Figure 7:

hence, $Ovch(\mu,\sigma) = \min(\frac{5}{4}, \frac{4}{5}, \frac{8}{7}, \frac{7}{8}) = 0.8$.
Since, $I(\mu,\sigma) = 1 - \frac{1}{10}|5-4| = 0.9$, it results

$$DR(\mu,\sigma) = \frac{0.8+0.9}{2} = 0.85,$$

μ *and* σ show *high family resemblance.*

2. *Functions μ,σ in figure 8, verify $(\mu,\sigma) \in$* **fr**. *For then, $f_1(\mu) = 4 + 2{\cdot}0.5 = 5$, $f_2(\mu) = 6$, $f_1(\sigma) = 2$, $f_2(\sigma) = 4$,*

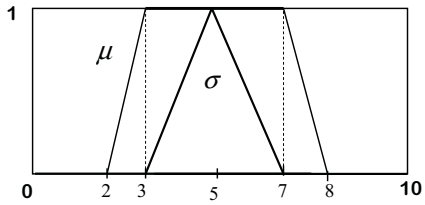


Figure 8:

hence, $Ovch(\mu,\sigma) = \min(\frac{5}{2}, \frac{2}{5}, \frac{6}{4}, \frac{4}{6}) = \frac{2}{5} = 0.4$.
Since, $I(\mu,\sigma) = 1 - \frac{1}{10}(|5-2|) = 0.7$, it results

$$DR(\mu,\sigma) = \frac{0.4+0.7}{2} = 0.55,$$

μ and σ *do not show* high family resemblance.

3. *Functions μ,σ in figure 9, verify $(\mu,\sigma) \in$* **fr**. *For them $f_1(\mu) = 5+1 = 6$, $f_2(\mu) = 7$, $f_1(\sigma) = 4+2 = 6$ and $f_2(\sigma) = 8$, hence, $Ovch(\mu,\sigma) = \min(\frac{6}{6}, \frac{7}{8}, \frac{8}{7}) = \frac{7}{8} = 0.88$.*

Since $I(\mu,\sigma) = 1 - \frac{1}{10}(|6-6|) = 1$, it results

$$DR(\mu,\sigma) = \frac{0.88+1}{2} = 0.94,$$

μ and σ *show* high family resemblance.

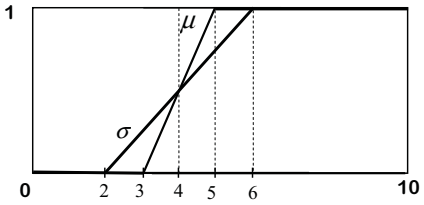


Figure 9:

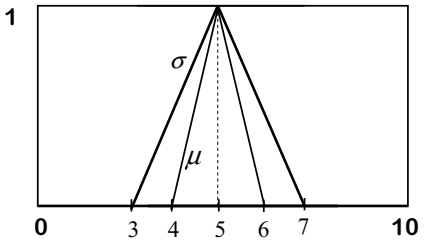


Figure 10:

4. *Functions μ,σ in figure 10, verify $(\mu,\sigma) \in$* **fr**. *For them $f_1(\mu) = 1$, $f_2(\mu) = 2$, $f_1(\sigma) = 2$ and $f_2(\sigma) = 4$,*
hence, $Ovch(\mu,\sigma) = \min(\frac{1}{2}, \frac{2}{1}, \frac{2}{4}, \frac{4}{2}) = 0.5$.
Since $I(\mu,\sigma) = 1 - \frac{1}{10}(2-1) = 0.9$, it results

$$DR(\mu,\sigma) = \frac{0.5+0.9}{2} = 0.7,$$

μ and σ *show* high family resemblance.

5. *Functions μ,σ in figure 11, verify $(\mu,\sigma) \in$* **fr**. *For them $f_1(\mu) = 3 + 2{\cdot}0.5 + 1 = 5$, $f_2(\mu) = f_2(\sigma) = 10$ and $f_1(\sigma) = 0.5 + 3{\cdot}0.5 + \frac{3{\cdot}0.5}{2} + 5 = 7.75$, hence, $Ovch(\mu,\sigma) = \frac{5}{7.75} = 0.64$.*

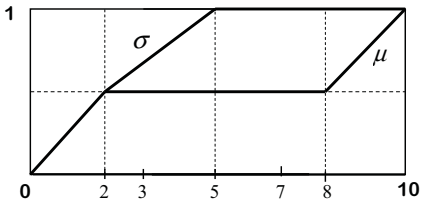


Figure 11:

Since $I(\mu,\sigma) = 1 - \frac{1}{10}(7.75-5) = 0.725$, it results

$$DR(\mu,\sigma) = \frac{0.64+0.725}{2} = 0.6825,$$

μ and σ *do not show* high family resemblance.

6. *Functions μ,σ in figure 12, verify $(\mu,\sigma) \in$* **fr**. *It is:*
$f_1(\mu) = 3$, $f_2(\mu) = 4$, $f_1(\sigma) = 3$ *and* $f_2(\sigma) = 5$.*Then,*

$$Ovch(\mu,\sigma) = \frac{4}{5}, \;\; I(\mu,\sigma) = 1$$

That is,

$$DR(\mu,\sigma) = 0.9,$$

and μ, σ *show* high family resemblance.

ISBN: 978-989-95079-6-8

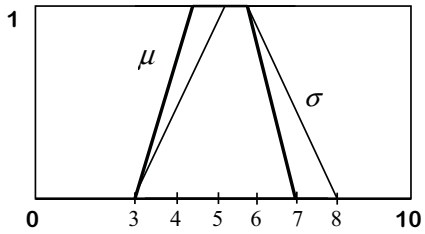


Figure 12:

4 The case of membership functions with the same predicate.

Given a predicate, or linguistic label, P on X, there is not a single fuzzy set generated by P, but each use of P in X, once reflected by its membership function μ_P (see [4]), defines a fuzzy set $\mathbb{P}$ by

$$x \in_r \mathbb{P} \Leftrightarrow \mu_P(x) = r, \text{ for all } x \in X, r \in [0,1].$$

Nevertheless, all functions μ_P do have some 0-points, and some 1-points, in common, as well as they should either decrease or non-decrease simultaneously in the same parts of X. Hence, *all μ_P should show some degree of family resemblance.*

This is the case for example, with P =*small* in $X = [0, 10]$, if two of its uses are represented by the membership functions μ_P^1 and μ_P^2 in figure 13.

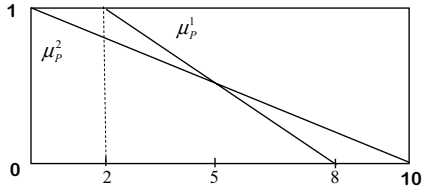


Figure 13:

It is clear that $(\mu_P^1, \mu_P^2) \in$ **fr**, and:

- $Ovch(\mu_P^1, \mu_P^2) = \min(\frac{5}{5}, \frac{10}{8}, \frac{8}{10}) = 0.8$
- $I(\mu_P^1, \mu_P^2) = 1 - \frac{1}{10}(|5-5|) = 1.$

  Hence,

$$DR(\mu_P^1, \mu_P^2) = \frac{0.8+1}{2} = 0.9,$$

  and μ, σ *show high family resemblance*.

Examples 1 and 4 in section 3, do correspond to uses of P =*big* and P =*around five*, respectively.

Remark 4.1 *In natural language, families $[\mu]$ must be 'open', but not 'closed' like they were defined in section 2. These families are here static (sets), but in natural language they should be dynamical. With time, an element σ that was not in $[\mu]$, but possibly with a not too low degree $DR(\sigma, \mu)$, could be included in $[\mu]$ and thus generating a new family, indeed, changing the relation* **fr**. *Actually,* **fr** *is not a permanent relation, in natural language it is a changing one. Only with families of resemblance taken as classical sets, human thought seems to be impossible (see [7]).*

5 Last remarks

5.1

In his *Philosophical Investigations*, Wittgenstein conceived language in a way close to how people manages it. Thus, the meaning of an imprecise predicate is not given by necessary and sufficient conditions, but is built up by similarity with its prototypes in the universe of discourse, like in the case of *big* in [0,10] with, at least, the prototype 10. Notwithstanding, for more complex predicates like P =*beautiful* in a set of art's objects, it could be not clear the existence of prototypes and, since in such cases it could be $S(\mu_P) = \emptyset$, the study of the family resemblance for these predicates remains an open problem.

5.2

Certainly, to consider $X \subset \mathbb{R}$ is a restriction for this paper's results. Nevertheless, most of the predicates fuzzy logic considers are those exhibiting a numerical characteristic, allowing to translate them into an interval in $\mathbb{R}$. This is the case, for example, of P = *tall* in a big population X, that is translated into the interval $[0, 2]$, in meters, by the numerical characteristic 'Height', and the predicate $Q = big$ in such interval. That is, by the identification $\mu_{tall} = \mu_{big} \circ Height$, or $\mu_{tall}(x) = \mu_{big}(Height(x))$, for all x in X, once the use of *big* in $[0, 2]$ is chosen accordingly with that of *tall* interpreted, as it turn, by *big height*.

There are not so obvious cases that, also, can be translated into an interval through a more complex process, consisting in identifying each x in X with an n-tuple of significative parts in x, $x := (x_1, \ldots, x_n)$, with numerical characteristics $Ch_i(x_i) \in [a_i, b_i]$, and taking $Ch(x) = A(Ch_1(x_1), \ldots, Ch_n(x_n)) \in [a, b]$, with some n-place aggregation function A. An example is given by P= *beautiful* in a set X of paintings, if each painting x can be partitioned in $x = x_1 \cup \ldots \cup x_n (x_i \cap x_j = \emptyset, i \neq j)$, allowing to interpret the statement 'x is P' as the composite one (x_1 is P,$\ldots, x_n$ is P). Provided each component 'x_i is P' is numerically evaluable by a clearly explicitable characteristic $Ch_i(x_i) \in \mathbb{R}^+$, the values of μ_P could be obtained by

$$\mu_P(x) = \frac{r_1 Ch_1(x_1) + \ldots + r_n Ch_n(x_n)}{r_1 + \ldots + r_n} (r_i \geq 0),$$

that are in the interval $[\min \mu_P, \max \mu_P]$.

Of course, $A(a_1, \ldots, a_n) = \frac{r_1 a_1 + \ldots + r_n a_n}{r_1 + \ldots + r_n}$, is not the only way of reasonably aggregating the n-tuples $(Ch_1(x_1), \ldots, Ch_n(x_n)) \in \mathbb{R}^n$. For example, if '$x$ is P' is interpretable by 'x_1 is P' *and* $\ldots$ *and* 'x_n is P', A can be taken as a t-norm, provided $Ch_i(x_i) \in [0, 1], 1 \leq i \leq n$.

6 Conclusions

6.1

The concept of family resemblances is here introduced as a crisp-binary relation for only a particular type of fuzzy sets in the real line, with 0-points and 1-points. Even more restrictive is the class of Riemann-integrable membership functions, to which pairs a degree of family resemblance is assigned. Anyway, this paper should be viewed as only a first trial to consider Wittgenstein's idea on family resemblances with fuzzy

ISBN: 978-989-95079-6-8

sets . For example, the definition of **fr** could be, perhaps, extended to other fuzzy sets without 1-points (non-normalized fuzzy sets), or without 0-points, as well as to pairs (μ, σ) with either μ, or σ in $\{0,1\}^X$. That is, to non-data fuzzy sets, to fuzzy sets resulting from computations with data ones. In addition, the selected degree DR cannot be considered as the definitive definition to measure to what extent there is family resemblance. More study on the subject is deserved since, for example, DR can be applied to pairs $(\mu, \sigma) \notin$ **fr**, and more accurate values can be obtained in some cases with non symmetric functions $DR(\mu,\sigma) = \frac{r_1 \cdot Ovch(\mu,\sigma) + r_2 \cdot I(\mu,\sigma)}{r_1 + r_2}$.

6.2

Of course, definition 2.1 could also be applied to crisp sets by simply allowing either μ or σ to only have 'constant pieces'. At this respect, examples like the one in figure 14 are of some interest. In such figure, μ is the trapezoidal fuzzy set $(3.8, 4, 6, 6.2)$, and σ does represent the crisp subset $[4, 6]$.

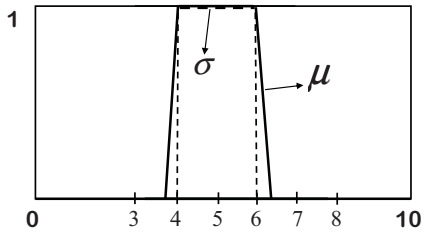


Figure 14:

Since $Ovch(\mu,\sigma) = 0.83$, $I(\mu,\sigma) = 0.98$, it follows $DR(\mu,\sigma) = 0.90$, a big value suggesting a so high resemblance between μ and σ that, perhaps, could allow to identify μ with μ_P, $P =$ *Almost between 4 and 6*. That is, μ could be 'linguistically approached' by the imprecise predicate P.

6.3

Maybe what is here introduced could be useful to approach the (pending) problem of 'linguistic approximation'. Namely, given the output $f : X \to [0,1]$ of a system of fuzzy rules, how to find a predicate P in X such that f could be identified with μ_P? A possible way could come once identified a membership function μ_P (representing a use of the known predicate P), such that, $DR(f, \mu_P) > \varepsilon$, and $Sup|f - \mu_P| < \delta$ (for fixed $\varepsilon > 0, \delta > 0$), even if $(f, \mu_P) \notin$ **fr**. Then 'P' can be called an (ε, δ)-linguistic approximation of f. For example, in 6.2, *Almost between 4 and 6* is an $(0.89, 0.2)$-linguistic approximation of μ.

6.4

In this paper, the threshold 0.7 is taken for the sake of illustrating the idea of 'high family resemblance'. Nevertheless, and although 0.7 seems to be a good enough value for the examples shown in figures 7 to 14, such threshold's value still remains to be studied.

Acknowledgment

The authors express their thanks to the two anonymous reviewers for their interesting and constructive remarks.

References

[1] L. Wittgenstein. *Tractatus logico-philosophicus*. Routledge & Kegan Paul, London, 1961.

[2] L. Wittgenstein. *Philosophical Investigations*. Basil Blackwell, London, 1973.

[3] E. Trillas, C. Moraga, S. Guadarrama, S. Cubillo, and E. Castiñeira. Computing with Antonyms. In *Forging New Frontiers: Fuzzy Pioneers I*, volume 217 of *Studies in fuzziness and soft computing*, pages 133–153. Springer, Berlin, 2007.

[4] E. Trillas. On a model for the meaning of predicates. A naïve approach to the genesis of fuzzy sets. In Rudolf Seising, editor, *Fuzzy sets theory. Philosophy and criticism*, Studies in fuzziness and soft computing. Springer Verlag, forthcoming.

[5] L. Valverde E. Trillas. An inquiry into indistinguishability operators. *Aspects of Vagueness. H.J. Skala, S. Termini, E. Trillas (Eds.)*, pages 231–256. Reidel, Dordrecht, 1984.

[6] S.V. Ovchinnikov. Representation of transitive fuzzy relations. *Aspects of Vagueness. H.J. Skala, S. Termini, E. Trillas (Eds.)*, pages 105–118. Reidel, Dordrecht, 1984.

[7] S. Pinker. *Words and Rules*. Basic Books, N.Y., 1999.

ISBN: 978-989-95079-6-8

Convergence theorems for generalized random variables and martingales

Andrew L Pinchuck

Department of Mathematics (Pure and Applied),

Rhodes University, PO Box 94,Grahamstown, 6140,

South Africa

Email: a.pinchuck@ru.ac.za.

Abstract— We examine generalizations of random variables and martingales. We prove a new convergence theorem for set-valued martingales. We also generalize a well known characterization of set-valued random variables to the fuzzy setting.

Keywords— Banach space, fuzzy, martingale, random variable, set-valued.

1 Set-valued random variables and martingales

The theory of conditional expectations and martingales has been established for Banach space-valued, Bochner-integrable functions. Hiai and Umegaki have then generalized random variables, conditional expectations and martingales to the set-valued (multivalued) setting in [1]. We will assume that the reader is familiar with the basic ideas of random variables and provide a brief introduction to the set-valued generalization and also to the related concept of conditional expectations which has played an important role in probability theory, ergodic theory and quantum statistical mechanics. We also contribute to the theory of set-valued martingales and since martingales are important in probability theory, we feel that there are potential applications to be developed from this work. This paper can be divided into two main results. The first result is an original convergence theorem for set-valued martingales. The second result is a generalization of a useful characterization of measurable set-valued random variables by Hiai and Umegaki in the fuzzy setting. The original theorem by Hiai and Umegaki was an essential tool used to obtain many of the subsequent results in [1] and it is clear that we can apply this new generalized theorem in an analagous way in the fuzzy setting.

We will be again considering a measure space (Ω, Σ, μ), a separable Banach space $(X, \|\cdot\|)$ with K a field of scalars (either $\mathbb{R}$ or $\mathbb{C}$). We remind the reader that functions f, g from Ω into X are said to be equal *almost everywhere*, denoted $f = g$ a.e. (μ) or $f(\omega) = g(\omega)$ $(\forall \omega \in \Omega)$ a.e. (μ) if $f(\omega) = g(\omega)$ for all $\omega \in \Omega$ except on a set of measure zero. We can omit reference to μ and Ω if there is no confusion. That is $f = g$ a.e. if $\mu(\{\omega \in \Omega : f(\omega) \neq g(\omega)\}) = 0$. We will denote by $M[\Omega, X]$ and $L^p[\Omega, X]$ the collections of measurable and p-integrable functions $f : \Omega \to X$ respectively, for $1 \leq p \leq \infty$. Due to the completeness of X we have several characterizations of this class of functions (see [1] Theorem 1.0). As in [1], $K(X)$ shall denote the collection of nonempty compact subsets of X and $K_k(X)$ the collection of convex sets in $K(X)$. We denote the collection of (weakly) measurable set-valued functions $F : \Omega \to K_k(X)$ by $\mathcal{M}[\Omega, K_k(X)]$. That is $F \in \mathcal{M}[\Omega, K_k(X)]$ if and only if for all O open, $O \subset X$, $F^{-1}(O) \in \Sigma$. Let p be chosen such that $1 \leq p \leq \infty$ then we denote by $\mathcal{L}^p[\Omega, K_k(X)] = \mathcal{L}^p[\Omega, \Sigma, \mu, K_k(X)]$ the space of all p-integrable functions in $\mathcal{M}[\Omega, K_k(X)]$, where two functions $F_1, F_2 \in \mathcal{L}^1[\Omega, K_k(X)]$ will be considered identical if $F_1(\omega) = F_2(\omega)$ a.e.

The topology on $K_k(X)$ is induced by the Hausdorff metric defined by

$$d_H(A, B) = \max\Big\{ \sup_{a\in A} \inf_{b\in B} \|a - b\|, \sup_{b\in B} \inf_{a\in A} \|a - b\| \Big\}, \tag{1.1}$$

for $A, B \subseteq X$. It is well known $(K_k(X), d_H)$ is a complete separable metric space. It now follows that in this topology $(F_n)_{n\in N} \subseteq \mathcal{M}[\Omega, K_k(X)]$ then $F_n \longrightarrow F$ as $n \longrightarrow \infty$ if and only if $d_H(F_n(\omega), F(\omega)) \longrightarrow 0$ for all $\omega \in \Omega$ a.e.

We can consider $(X, \|\cdot\|)$ to be a topological space since a norm defines a natural metric on X which induces the *metric topology* on X. We define the *closure* of a set $A \subseteq X$, denoted $\mathrm{cl}(A)$ to be the *norm closure* of A. That is, the smallest closed set containing A with respect to the metric topology induced by the norm $\|\cdot\|$.

A measurable function $f : \Omega \to X$ is called a *measurable selection* of F if $f(\omega) \in F(\omega)$ for all $\omega \in \Omega$ a.e. That is f is a measurable selection of F if $f(\omega) \in F(\omega)$ for all $\omega \in \Omega$ except on a set of measure zero. We define the set $S^p_F = \{f \in L^p[\Omega, X] : f(\omega) \in F(\omega)\ a.e.\}$. It is easy to show that S^p_F is a closed subset of $L^p[\Omega, X]$ for $1 \leq p \leq \infty$.

Lemma 1.1 *[1] Let $F \in \mathcal{M}[\Omega, K_k(X)]$ and $1 \leq p \leq \infty$. If S^p_F is nonempty, then there exists a sequence $(f_i)_{i\in N}$ contained in S^p_F such that $F(\omega) = \mathrm{cl}\{f_i(\omega)\}$ for $\omega \in \Omega$.*

Corollary 1.2 *[1] Let $F_1, F_2 \in \mathcal{M}[\Omega, K_k(X)]$ and $1 \leq p \leq \infty$. If $S^p_{F_1} = S^p_{F_2} \neq \emptyset$ then $F_1(\omega) = F_2(\omega)$, for all $\omega \in \Omega$.*

A brief discussion of the generalization of conditional expectations is necessary before we reach our first main result. This material is covered more comprehensively in [1].

Let W be a sub-σ-algebra of Σ,

$$S^1_F(W) = \{f \in L^1[\Omega, W, \mu, X] : f(\omega) \in F(\omega)\, a.e.\}. \tag{1.2}$$

We define

ISBN: 978-989-95079-6-8

$$\int_{\Omega}^{(W)} F d\mu = \Big\{ \int_{\Omega} f d\mu : f \in S_F^1(W) \Big\}. \quad (1.3)$$

If $f \in L^1[\Omega, X]$ then the *conditional expectation* $E[f|W]$ *of* f *relative to* W is defined as a W-measurable function $E[f|W] \in L^1[\Omega, W, \mu, X]$ such that

$$\int_A E[f|W] d\mu = \int_A f d\mu, \quad (1.4)$$

for all $A \in W$.

The following theorem provides a natural extension of the notion of a conditional expectation to the set-valued setting.

Theorem 1.3 *[1] Let $F \in \mathcal{L}^1[\Omega, K_k(X)]$. Then there exists a unique $E[F|W] \in \mathcal{L}^1[\Omega, W, \mu, K_k(X)]$ such that*

$$S^1_{E[F|W]}(W) = \text{cl}\{E[f|W] : f \in S_F^1\} \quad (1.5)$$

where the closure is taken with respect to the norm in $\mathcal{L}^1[\Omega, K_k(X)]$. $E[F|W]$ is called the conditional expectation of F relative to W.

The concept of martingale in probability theory was introduced by Paul Pierre Lévy. Part of the motivation for that work was to show the impossibility of successful betting strategies. A martingale was originally devised as an indexed sequence of random variables with the index representing time. If t is a later time than s then the idea is that the conditional expected value at time t given the same observations as at time s will be equal to the expected value at time s. The notion of a martingale has proved to be a useful tool in modeling various events in probability theory. We present a precise definition of the classical notion followed by the generalization to the set-valued setting.

We remind the reader that a *filtration* on (Ω, Σ, μ) is a sequence of σ-algebras $(\Sigma_n)_{n\in N}$ such that $\Sigma_n \subseteq \Sigma$ and $\Sigma_n \subseteq \Sigma_{n+1}$ for all $n \in \mathbb{N}$.

Definition 1.4 *If $(f_n)_{n\in N}$ is a sequence of random variables and $(\Sigma_n)_{n\in N}$ a filtration then the sequence $(f_n, \Sigma_n)_{n\in N}$ is called a martingale if we have for each $n \in \mathbb{N}$:*

(a) f_n is Σ_n-measurable and

$$\int_{\Omega}^{(\Sigma_n)} \|f_n\| d\mu < \infty, \quad (1.6)$$

(b)

$$E[f_{n+1}|\Sigma_n] = f_n. \quad (1.7)$$

Alternatively, if property (b) is replaced by

(b')

$$E[f_{n+1}|\Sigma_n] \geq f_n (E[f_{n+1}|\Sigma_n] \leq f_n) \quad (1.8)$$

then $(f_n, \Sigma_n)_{n\in N}$ is called a submartingale (supermartingale), respectively.

Definition 1.5 *Let $(F_n)_{n\in N}$ be a sequence of set-valued random variables and $(\Sigma_n)_{n\in N}$ a filtration then the sequence $(F_n, \Sigma_n)_{n\in N}$ is called a set-valued martingale if we have, for each $n \geq 1$:*

(a) F_n is Σ_n-measurable and

$$\int_{\Omega}^{(\Sigma_n)} \|F_n\| d\mu < \infty, \quad (1.9)$$

(b)

$$E[F_{n+1}|\Sigma_n] = F_n. \quad (1.10)$$

Alternatively, if property (b) is replaced by

(b')

$$E[F_{n+1}|\Sigma_n] \geq F_n (E[F_{n+1}|\Sigma_n] \leq F_n) \quad (1.11)$$

then $(F_n, \Sigma_n)_{n\in N}$ is called a fuzzy submartingale (supermartingale), respectively.

Theorem 1.6 *If $(F_n, \Sigma_n)_{n\in N}$ is a set-valued martingale and $1 \leq p \leq \infty$ then $F_n \longrightarrow F$ as $n \longrightarrow \infty$ if and only if for each choice of $f_n \in S^p_{F_n}$, $n \in N$, we have that*

$$\lim_{n\longrightarrow\infty} f_n \in S_F^p. \quad (1.12)$$

Proof: Let p be such that $1 \leq p \leq \infty$. For brevity, we will henceforth denote $\lim_{n\longrightarrow\infty} f_n(\omega)$ by $f(\omega)$ for each $\omega \in \Omega$. Now $f_n \in S^p_{F_n}$ for all $n \in \mathbb{N}$ but $f \notin S_F^p \iff$ for all $n \in \mathbb{N}$, there exists $A \in \Sigma$, $\mu(A) \neq 0$ such that $f(\omega) \notin F(\omega)$ for all $\omega \in A$. Now for a given $A \in \Sigma$ we have $\forall \omega \in A$, $f_n(\omega) \neq f(\omega) \iff \forall \omega \in A$, $f(\omega) - f_n(\omega) \neq 0 \iff \forall \omega \in A$, $\|f(\omega) - f_n(\omega)\| \geq \delta$ for some $\delta > 0$. Thus $f \notin S_F^p \iff \exists A \in \Sigma, \mu(A) \neq 0, \forall \omega \in A, \forall n \in \mathbb{N}$,

$$d_H(F(\omega), F_n(\omega)) =$$

$$\max\{ \sup_{f(\omega)\in F(\omega)} \inf_{f_n(\omega)\in F_n(\omega)} \|f(\omega) - f_n(\omega)\|,$$

$$\sup_{f_n(\omega)\in F_n(\omega)} \inf_{f(\omega)\in F(\omega)} \|f(\omega) - f_n(\omega)\| \}$$

$\geq \delta \iff d_H(F(\omega), F_n(\omega) \not\longrightarrow 0$ for all $\omega \in A \in \Sigma$, $\mu(A) \neq 0 \iff d_H(F, F_n) \not\longrightarrow 0 \iff F_n \not\longrightarrow F$ as $n \longrightarrow \infty$.

2 Fuzzy sets

We now provide a brief introduction to fuzzy sets with a particular focus on the tools that we will need in the subsequent section. These notions are all well known and standard. For further reading on the topic, the reader is referred to [2, 3, 4, 5, 8](by no means an exhaustive list). Throughout this paper, we will be considering fuzzy sets $A \in I^X$ where X is a Banach Space and I the unit interval $[0, 1]$.

For a given fuzzy set we associate collections of crisp subsets of X with it. This concept is highly useful in relating statements in the fuzzy setting to the classical setting.

ISBN: 978-989-95079-6-8

Definition 2.1 *If $A \in I^X$ and $\alpha \in I$ we define,*

$$A^\alpha = \{x \in X : A(x) > \alpha\}; \tag{2.13}$$

$$A_\alpha = \{x \in X : A(x) \geq \alpha\}. \tag{2.14}$$

These crisp sets are referred to as α-levels (or cuts), strong and weak respectively. We call $\text{supp}A = A^0$ *the support of A.*

We shall work mainly with the weak cuts in this paper although similar results can be obtained using the strong cuts.

Throughout this paper we will denote the characteristic function of a crisp set $A \subseteq X$ by χ_A. That is $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ elsewhere. We immediately obtain a useful characterization of fuzzy sets.

Theorem 2.2 *If $A \in I^X$ and $x \in X$ then*

$$A(x) = \sup_{n \in N}\{\alpha_n \cdot \chi_{A_{\alpha_n}}(x)\} \tag{2.15}$$

where $\{\alpha_n : n \in \mathbb{N}\}$ is dense in I.

We remind the reader of the standard definitions of addition and scalar multiplication of crisp sets. $A + B$ denotes the collection $\{a + b : a \in A, b \in B\}$ and for a scalar α, we denote by $\alpha \cdot A$ the collection $\{\alpha \cdot a : a \in A\}$. There are natural extensions of these concepts in the fuzzy setting and it should be noted that the following two definitions are consequences of the well known concept of a fuzzy image.

Definition 2.3 *Let $A, B \in I^X$ we define*

$$(A + B)(x) = \sup_{a+b=x} \{A(a) \wedge B(b)\} \tag{2.16}$$

for all $x \in X$.

Definition 2.4 *Let $A \in I^X, t \in K$ and $x \in X$. Then we define $t \odot A(x) = A(\frac{x}{t})$ for $t \neq 0$ and*

$$t \odot A(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ \sup A & \text{if } x = 0 \end{cases} \tag{2.17}$$

if $t = 0$.

The $\odot$ is simply used to distinguish between this type of scalar multiplication and the $\cdot$ representing ordinary pointwise multiplication of functions as used in Theorem 2.2.

Due to the lattice structure of I, we have that pointwise ordering is a natural partial ordering that can be defined on I^X. That is $\leq$ given by

$$A \leq B \Leftrightarrow A(x) \leq B(x) \tag{2.18}$$

for all $x \in X$ and $A, B \in I^X$. This leads to the following straightforward relationships between the concepts of α-levels and this ordering respectively. The proofs follow trivially.

Lemma 2.5 *If $A, B \in I^X$ then for all $\alpha \in [0, 1]$ the following statements hold*

1.

$$[A \vee B]_\alpha = A_\alpha \cup B_\alpha, \tag{2.19}$$

2.

$$[A \wedge B]_\alpha = A_\alpha \cap B_\alpha, \tag{2.20}$$

3.

$$A \leq B \Longleftrightarrow A_\alpha \subseteq B_\alpha. \tag{2.21}$$

We have the following immediate consequence of the previous lemma.

Lemma 2.6 *Let A and B be fuzzy sets. Then*

$$A = B \Longleftrightarrow A_\alpha = B_\alpha \tag{2.22}$$

for all $\alpha \in [0, 1]$.

Definition 2.7 *A fuzzy set A on X is convex if $A(kx+(1-k)y) \geq A(x) \wedge A(y)$* whenever *$x, y \in X$ and $0 \leq k \leq 1$.*

The following proposition is a useful characterization of fuzzy convexity which justifies the definition and relates it to the corresponding crisp definition.

Proposition 2.8 *Let A be a fuzzy set in X then the following three assertions are equivalent:*

1. *A is convex.*
2. *$\forall k \in [0, 1], kA + (1 - k)A \leq A$.*
3. *$\forall \alpha \in I, A_\alpha$ is convex.*

We will later require the next theorem which illustrates the relationship between the scalar multiplication and addition of fuzzy sets and the corresponding concepts in the crisp setting.

Lastly we discuss the natural concept of fuzzy points since this will be necessary for the next section. We use $\mathcal{P}(X)$ to denote the *power set* of X - that is, $\mathcal{P}(X) = \{A : A \subseteq X\}$.

If $A \in \mathcal{P}(X)$ and $\alpha \in I$ we define

$$\alpha \cdot \chi_{\{x\}} = \begin{cases} \alpha & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \tag{2.23}$$

$$\alpha \cdot \chi_{\{x\}} = \begin{cases} \alpha & \text{on } x \\ 0 & \text{elsewhere} \end{cases} \tag{2.24}$$

So

We call $\alpha \cdot \chi_{\{x\}}$ a *fuzzy point* with *support* at x and *value* α. We will denote the set of fuzzy points in I^X by $\hat{X}$.

3 Fuzzy set-valued random variables

Our next focus is on the generalization of set-valued random variables to the fuzzy setting. The following overview of fuzzy set-valued random variables is taken mainly from [5] and [9]. These concepts are also discussed in [6, 7] amongst other works.

Definition 3.1 *Let* $\mathrm{F(X)}$ *denote the collection of fuzzy sets such that*

(a) A is uppersemicontinuous,

(b) $\{x \in X : A(x) = 1\} \neq \emptyset$,

(c) $\text{supp}A \in K(X)$.

ISBN: 978-989-95079-6-8

We define $F_k(X)$ to be the collection of fuzzy sets in $F(X)$ that are fuzzy convex. We denote the collection of measurable functions $f : \Omega \to \hat{X}$ by $M[\Omega, \hat{X}]$, that is, $f \in M[\Omega, \hat{X}]$ if and only if for each $\alpha \in (0, 1]$, $f_\alpha \in M[\Omega, X]$ with $f_\alpha(\omega) = [f(\omega)]_\alpha$ for all $\omega \in \Omega$. Let p be chosen such that $1 \leq p \leq \infty$ then $L^p[\Omega, \hat{X}]$ will denote the p-integrable functions from Ω into $\hat{X}$. A fuzzy set-valued random variable (f.r.v.) or a fuzzy random set is a function $F : \Omega \to \mathrm{F_k(X)}$ such that $F_\alpha(\omega) = \{x \in X : F(\omega)(x) \geq \alpha\}$ is a set-valued random variable for all $\alpha \in (0, 1]$. A fuzzy set-valued random variable F is called measurable (integrable) if for each $\alpha \in (0, 1]$, F_α is measurable (integrable). Let $\mathcal{M}[\Omega, F_k(X)]$ be the collection of all measurable fuzzy random variables and $\mathcal{L}^p[\Omega, \mathrm{F_k(X)}] = \mathcal{L}^p[\Omega, \Sigma, \mu, \mathrm{F_k(X)}]$ denote the set of all p-integrable fuzzy random variables for $1 \leq p \leq \infty$. Clearly, a fuzzy set-valued random variable can be considered to be a generalized set-valued random variable. F_1, F_2 fuzzy set-valued random variables then $(F_1 + F_2)(\omega) = F_1(\omega) + F_2(\omega)$ for all $\omega \in \Omega$ (the normal addition of fuzzy sets defined earlier). Similarly for a fuzzy set-valued random variable F and measurable real valued function ζ on Ω, $(\zeta F)(\omega) = \zeta(\omega)F(\omega)$ for all $\omega \in \Omega$. We will also generalize Theorem 2.8 to the fuzzy setting. There are several ways to generalize d_H to the fuzzy setting but due to certain desirable properties we will restrict our attention to d_∞ defined by

$$d_\infty(F_1, F_2) = d_H(\mathrm{supp}F_1, \mathrm{supp}F_2). \tag{3.25}$$

It follows trivially from our earlier discussion of the Hausdorff metric that $(F_k(X), d_\infty)$ is a metric space. Under the topology induced by d_∞ we have that $F_n \longrightarrow F$ if and only if $d_\infty(F_n(\omega), F(\omega)) \longrightarrow 0$ as $n \longrightarrow \infty$ for all $\omega \in \Omega$ a.e.

Based on the notation used in [1] we now generalize the definition of S_F^p to the fuzzy setting.

Definition 3.2 *A measurable function $f \in M[\Omega, \hat{X}]$ such that $f(\omega) \in F(\omega)$ for all $\omega \in \Omega$ a.e. is called a measurable selection of F. Let $F \in \mathcal{M}[\Omega, F_k(X)]$ and let $1 \leq p \leq \infty$ then $S_F^p = \{f \in L^p[\Omega, \hat{X}] : f(\omega) \in F(\omega)$ a.e. $\}$.*

We now state the second of our main results. This is a new generalization of Lemma 1.1 - notice that the collection of fuzzy points $\hat{X}$ replaces the set X in this version of the theorem. This theorem opens the way for similar results to those obtained in [1].

Theorem 3.3 *If $F \in \mathcal{M}[\Omega, F_k(X)]$ then there exists a countable collection $\{f^i\}_{i \in N} \subseteq S_F^p$, such that $F(\omega) = \mathrm{cl}_{i \in N}\{f^i(\omega)\}$ for all $\omega \in \Omega$.*

Proof: Let $F \in \mathcal{M}[\Omega, F_k(X)]$ and let $\omega \in \Omega$. Then for $n \in \mathbb{N}$ we have $F_{\alpha_n} \in \mathcal{M}[\Omega, K_k(X)]$ with $\{\alpha_n\}_{n \in N}$ a countable set dense in I as in Theorem 2.2. Therefore by Lemma 1.1, for each $n \in \mathbb{N}$, $\exists \{f^i_{\alpha_n}\}_{i \in N}$ such that

$$\mathrm{cl}_{i \in N}\{f^i_{\alpha_n}(\omega)\}(x) = F_{\alpha_n}(\omega).$$

For each $i \in \mathbb{N}$ we now define

$$f^i(\omega) = \bigvee_{n \in N} \alpha_n \cdot \chi_{f^i_{\alpha_n}(\omega)}.$$

Then $\{f^i\}_{i \in N}$ is the required collection. Let $x \in X$ and now notice that for each $i \in \mathbb{N}$, $f^i : \Omega \to \hat{X}$. We now have that $\mathrm{cl}_{i \in N}\{[f_i(\omega)]\}(x) = \mathrm{cl}_{i \in N}\{[f_i(\omega)](x)\} = \mathrm{cl}_{i \in N}\{\bigvee_{n \in N} \alpha_n \cdot \chi_{f^i_{\alpha_n}(\omega)}(x)\} = \lim_{j \longrightarrow \infty} b_j$ where for each $j \in \mathbb{N}$, $b_j \in \left\{ \bigvee_{n \in N} \alpha_n \cdot \chi_{f^i_{\alpha_n}(\omega)}(x) : i \in \mathbb{N} \right\}$. Now for each $j \in \mathbb{N}$, $b_j = \lim_{k \longrightarrow \infty} b_j^k$ with $b_j^k \in \{\alpha_n \cdot \chi_{f^i_{\alpha_n}(\omega)}(x) : i, n \in \mathbb{N}\} \subseteq I$. Thus

$$\mathrm{cl}_{i \in N}\{f_i(\omega)\}(x) = \lim_{j \longrightarrow \infty} \lim_{k \longrightarrow \infty} b_j^k, \tag{3.26}$$

for $\{b_j^k\}_{j,k \in N} \subseteq \{\alpha_n \cdot \chi_{f^i_{\alpha_n}(\omega)}(x) : i, n \in \mathbb{N}\} = \left\{[f^i_{\alpha_n}(\omega)](x) : i, n \in \mathbb{N}\right\}$ and b_j^k is nondecreasing with respect to k.

We also have by Theorem 2.2 that $[F(\omega)](x) = \left(\bigvee_{n \in N} \alpha_n \cdot \chi_{F_{\alpha_n}(\omega)}\right)(x)$ $= \left(\bigvee_{n \in N} \alpha_n \cdot \chi_{\mathrm{cl}_{\{f^i_{\alpha_n}(\omega)\}}}\right)(x) = \left(\bigvee_{n \in N} \alpha_n \cdot \chi_{\mathrm{cl}_{\{f^i_{\alpha_n}(\omega)\}}}(x)\right) \in I$.

By definition $\left(\bigvee_{n \in N} \alpha_n \cdot \chi_{\mathrm{cl}_{\{f^i_{\alpha_n}(\omega)\}}}(x)\right) = \lim_{k \longrightarrow \infty} b^k$ where $(b^k)_{k \in N} \subseteq \{[f^i_{\alpha_n}(\omega)](x) : n \in \mathbb{N}\} \subseteq I$ and b^k is nondecreasing with respect to k. Also for each $k \in \mathbb{N}$ we have that $b^k = \lim_{j \longrightarrow \infty} b_j^k$ with $b_j^k \in \left\{[f^i_{\alpha_n}(\omega)](x) : i, n \in \mathbb{N}\right\}$ for each $j, k \in \mathbb{N}$. Therefore we have

$$[F(\omega)](x) = \lim_{k \longrightarrow \infty} \{ \lim_{j \longrightarrow \infty} b_j^k\} \tag{3.27}$$

This concludes the proof since it is well known from real analysis that

$$\lim_{k \longrightarrow \infty} \{ \lim_{j \longrightarrow \infty} b_j^k\} = \lim_{j \longrightarrow \infty} \{ \lim_{k \longrightarrow \infty} b_j^k\}$$

for all $b_j^k \in I$ with $j, k \in \mathbb{N}$.

References

[1] F. Hiai, H Umegaki. Integrals, Conditional Expectations, and Martingales of Multivalued Functions, *J. Multivariate Anal 7*, no. 1 (1977), 149–182.

[2] A. K. Katsaras & D. B. Liu. Fuzzy Vector Spaces and Fuzzy Topological Vector Spaces. J.Math.Anal.Appl. 58 (1977), 135-146.

[3] A. K. Katsaras. Fuzzy Topological Vector Spaces I. Fuzzy Sets and Systems 6 (1981) 85-95.

[4] A. K. Katsaras. Fuzzy Topological Vector Spaces II. Fuzzy Sets and Systems 12 (1984) 143-154.

[5] S. Li, Y. Ogura, V. Kreinovich. *Limit theorems and applications of set-valued and fuzzy set-valued random variables*, Kluwer Academic Press, Dordrecht-Boston-London, 2002.

ISBN: 978-989-95079-6-8

[6] S. Li, Y. Ogura. Convergence of set-valued and fuzzy-valued martingales. Fuzzy Sets and Systems 101 (1999) 453-461.

[7] S. Li, Y. Ogura. A Convergence theorem of fuzzy-valued martingales in the extended Hausdorff metric H_∞. Fuzzy Sets and Systems 135 (2003) 391-399.

[8] A. L. Pinchuck. Extension Theorems on L-Topological Spaces and L-Fuzzy Vector Spaces. Masters thesis, Rhodes University (2001).

[9] M. L. Puri, D. A. Ralescu. Convergence Theorem for Fuzzy Martingales. *Journal of Mathematical Analysis and Applications.* **160** (1991), 107-122.

ISBN: 978-989-95079-6-8

Production and Transportation Planning – A fuzzy Approach for Minimizing the Total Cost

Heinrich J. Rommelfanger

Goethe University Frankfurt am Main
Faculty of Economics and Business Administration
Rommelfanger@wiwi.uni-frankfurt.de

Abstract— *In this paper, we deal with the production and transportation planning of a household appliances manufacturer that has production facilities and central stores for resellers in several sites in Europe. Each store can receive products from all production plants and it is not necessary that all products are produced in all production units. The transport between any two bases is done by trucks. For simplicity we assume, that each truck has the same capacity of M EURO-pallets, and for each product the unit is EURO-pallet. The target of this paper is to determine a production and transportation plan that minimize the total sum of the production cost and the transportation cost. For working in a realistic environment we assume that the production capacities in the different plants and the demand in the sales bases are not known exactly but the management can describe the data in form of fuzzy numbers. By using an inter-active algorithm for solving the fuzzy linear programming system we achieve a stable production and a satisfactory supply of the products. Moreover, we demonstrate that this integer programming problem can adequately be solved without using computation-intensive integer programming algorithms. Additionally, in the course of the inter-active solution process the production bottlenecks get clearly visible. A numerical example illustrates the efficiency of the proposed procedure.*

Keywords — *Fuzzy LP-systems, integer programming, FULPAL, Production and transportation planning, inter-active process*.

1 Introduction

Due to globalisation and the involved international expansion of companies numerous firms in Europe produce their products in several European and Overseas countries. On the one hand they take advantage of the different production cost and of higher lot sizes, on the other hand they endeavour to minimize the total cost of manufacturing and transportation.

In literature it is usually proposed to determine a minimal cost production plan by means of an integer programming system. But this procedure is very computationally intensive. Moreover, it is neglected that real production capacities in the different plants and the demand for the products are not known exactly. In this paper we look on the more realistic case that the management can describe the production capacities in the different plants and the demand in the sales bases in form of fuzzy numbers.

The paper is organized as follows: In Section 2 the problem is formulated in detail. In Section 3 we demonstrate that the integer programming problem can adequately be solved by using an inter-active algorithm for solving the fuzzy linear programming system. A numerical example in Section 4 illustrates the efficiency of the proposed procedure. Finally, conclusions and possible extensions are presented in Section 5.

2 Problem formulation

We deal with the production and transportation planning of a household appliances manufacturer HAM that has production facilities and central stores for resellers in several sites in Europe. Each store can receive products from all production plants and it is not necessary that all products are produced in all production units.

- The number of production and sales bases of HAM is N. Obviously, it is not necessary to differ between production plants and sales stores. A base without any demand is a pure production location and a base where nothing is produced is a pure sales shop.
- The number of products of HAM is K.
- The transport between any two bases is done by trucks. For simplicity we assume, that each truck has the same capacity of M EURO-pallets independent of the sort of products.
- For each product the unit is EURO-pallet.

In the production and transportation planning model, we use the following notations:

x_{ki} output of the product k at the base i,

y_{kij} total number of units of the product k transported from the base i to the base j,

w_{ij} number of the trucks from base i to base j,

$\tilde{D}_{ki}$ demand for the product k at the base i,

$\tilde{P}_{ki}$ production capacity for the product k at the base i,

C_{ki} production cost of one unit of product k at the base i,

T_{ij} cost for transporting a truck from the base i to the base j, where $T_{ii} = 0$.

$k \in \{1,\ldots,K\}$, $i, j \in \{1,\ldots,N\}$

The items $x_{ki}, y_{kij}, w_{ij}, \tilde{D}_{ki}, \tilde{P}_{ki}$ are referred to the same time period, e. g. one week or one month.

Now, a production and transportation plan that minimize the total sum of the production cost and the transportation cost can be determined by solving the following integer programming problem:

ISBN: 978-989-95079-6-8

$$z(\mathbf{x},\mathbf{w}) = \sum_{i=1}^{N}\sum_{k=1}^{K} C_{ki}x_{ki} + \sum_{j=1}^{N}\sum_{i=1}^{N} T_{ji}w_{ji} \;\rightarrow\; \text{Min}$$

subject to (1)

$$x_{ki} \mathrel{\tilde{\leq}} \tilde{P}_{ki}, \quad k=1,...,K;\ i=1,...,N$$

$$s_{ik} = x_{ki} - \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kji} \mathrel{\tilde{\geq}} \tilde{D}_{ki}, \quad k=1,...,K;\ i=1,...,N$$

$$\sum_{k=1}^{K} y_{kij} \leq M\, w_{ij}, \quad i,j=1,...,m$$

$$x_{ij}, y_{kij}, w_{ij} \in \mathbf{N}\cup\{0\}, \quad k=1...,K,\ i,j=1,...,N$$

$$\mathbf{x} = (x_{11},\ldots,x_{K1},x_{12},\ldots,x_{K2},\ldots,x_{1N},\ldots,x_{KN})$$

$$\mathbf{w} = (w_{11},\ldots,w_{N1},w_{12},\ldots,w_{N2},\ldots,w_{1N},\ldots,w_{NN})$$

The first constraint expresses that the output of the product k at the base i is smaller than or equal to its production capacity $\tilde{P}_{ki}$; the second constraint means that the supply s_{ki} of the product k at the base i is larger than or equal to the its demand $\tilde{D}_{ki}$; the third constraint indicates that the total amount of products transported from the base i to the base j is smaller than or equal to the transportation capacity of w_{ij} trucks; the fourth constraint means that the output of the product k at the base i, the number of products transported from the base i to the base j and the number of trucks from the base i to base j are nonnegative integer.

Concerning the imprecise right-hand sides $\tilde{P}_{ki}$ and $\tilde{D}_{ki}$ we assume that management is able to describe the production capacities in the different plants and the demand in the sales bases in form of fuzzy numbers $\tilde{P}_{ki} = (p_{ki}, 0, \pi_{ki})$ and $\tilde{D}_{ki} = (d_{ki}, \delta_{ki}, 0)$. Here, p_{ki} and $d_{ki} - \delta_{ki}$ d_{ki} are the production capacities and the demands respectively that are expected in any case. Furthermore the management estimate the maximal production capacities and the maximal demands as $p_{ki} + \pi_{ki}$ and d_{ki} respectively.

3 Solution process

As shown in Rommelfanger [3], fuzzy integer programming LP-problem can be effectively solved by interactive algorithms for solving fuzzy LP-systems. The present system (1) is a relative simple model with one crisp objective function and soft constraints. Therefore we can use a special form of the algorithm FULPAL (**FU**zzy **L**inear **P**rogramming Based on **A**spiration **L**evels) for a stepwise calculation of an efficient compromise solution of the system (1).

At first, we have to calculate the smallest and the highest total cost by solving the two crisp LP-systems:

$$\underline{z} = \text{Min}\,(\sum_{i=1}^{N}\sum_{k=1}^{K} C_{ki}x_{ki} + \sum_{i=1}^{N}\sum_{j=1}^{N} T_{ji}w_{ji})$$

subject to (2)

$$x_{ki} \leq p_{ki} + \pi_{ki}, \quad k=1,...,K;\ i=1,...,N$$

$$s_{ki} = x_{ki} - \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kji} \geq d_{ki} - \delta_{ki}, \quad \begin{matrix} k=1,...,K \\ i=1,...,N \end{matrix}$$

$$\sum_{k=1}^{K} y_{kij} - M\, w_{ij} \leq 0, \quad i,j=1,...,m$$

$$x_{ij}, y_{kij}, w_{ij} \geq 0, \quad k=1...,K,\ i,j=1,...,N$$

and

$$\overline{z} = \text{Min}\,(\sum_{i=1}^{N}\sum_{k=1}^{K} C_{ki}x_{ki} + \sum_{i=1}^{N}\sum_{j=1}^{N} T_{ji}w_{ji})$$

subject to (3)

$$x_{ki} \leq p_{ki}, \quad k=1,...,K;\ i=1,...,N$$

(In case of $\sum_{i=1}^{N} p_{ki} < \sum_{i=1}^{N} d_{ki}$ the capacities p_{ki} must be increased up to the existence of an feasible solution of (2), starting with the plants that have the highest production costs; $k=1,\ldots,K$.)

$$s_{ki} = x_{ki} - \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1\\ j\neq i}}^{N} y_{kji} \geq d_{ki}, \quad k=1,...,K;\ i=1,...,N$$

$$\sum_{k=1}^{K} y_{kij} - M\, w_{ij} \leq 0, \quad i,j=1,...,N$$

$$x_{ij}, y_{kij}, w_{ij} \geq 0, \quad k=1...,K,\ i,j=1,...,N$$

In accordance with FULPAL, the objective function and the soft constraints of the system (1) are transformed in utility functions, where $z^A[r], p_{ki}^A[r], d_{ki}^A[r]$ with $r=1$ are the crisp aspiration levels that are specified by the management for the time being.

In the course of the inter-active solution process the management can change the aspiration levels step by step. For getting an effective comparability of the utilities, the same utility (membership degree) λ_A is assigned to all crisp aspiration levels. The goal of the management is to get a solution that satisfies all aspiration levels.

$$\mu_z(\mathbf{x},\mathbf{w}) = \quad (4)$$

$$\begin{cases} 1 & \text{if} \quad z(\mathbf{x},\mathbf{w}) < \underline{z} \\ 1 - \dfrac{z(\mathbf{x},\mathbf{w}) - \underline{z}}{z^A[r] - \underline{z}} \cdot (1-\lambda_A) & \text{if} \quad \underline{z} \leq z(\mathbf{x},\mathbf{w}) < z^A[r] \\ \lambda_A + \dfrac{z(\mathbf{x},\mathbf{w}) - z^A[r]}{\overline{z} - z^A[r]} \cdot (1-\lambda_A) & \text{if} \quad z^A[r] \leq z(\mathbf{x},\mathbf{w}) \leq \overline{z} \\ 0 & \text{if} \quad \overline{z} < z(\mathbf{x},\mathbf{w}) \end{cases}$$

$$\mu_{xki}(x_{ki}) = \quad (5)$$

$$\begin{cases} 1 & \text{if} \quad x_{ki} < p_{ki} \\ 1 - \dfrac{x_{ki} - p_{ki}}{p_{ki}^A[r] - p_{ki}} \cdot (1-\lambda_A) & \text{if} \quad p_{ki} \leq x_{ki} \\ \lambda_A + \dfrac{x_{ki} - p_{ki}^A[r]}{p_{ki} + \pi_{ki} - p_{ki}^A[r]} \cdot (1-\lambda_A) & \text{if} \quad p_{ki}^A[r] < x_{ki} \\ 0 & \text{if} \quad p_{ki} + \pi_{ki} < x_{ki} \end{cases}$$

ISBN: 978-989-95079-6-8

$$\mu_{ski}(s_{ki}) = \qquad (6)$$

$$\begin{cases} 0 & \text{if} \quad s_{ki} < d_{ki} - \delta_{ki} \\ \dfrac{s_{ki} - (d_{ki} - \delta_{ki})}{d^A_{ki}[r] - (d_{ki} - \delta_{ki})} \cdot \lambda_A & \text{if} \quad d_{ki} - \delta_{ki} \le s_{ki} \le d^A_{ki}[r] \\ \lambda_A + \dfrac{s_{ki} - d^A_{ki}[r]}{d_{ki} - d^A_{ki}[r]} \cdot (1 - \lambda_A) & \text{if} \quad d^A_{ki}[r] < s_{ki} \le d_{ki} \\ 1 & \text{if} \quad d_{ki} < s_{ki} \end{cases}$$

$$k = 1,...,K;\ i = 1,...,N$$

For calculating a compromise solution of the multi-objective system

$$(\mu_Z(\mathbf{x},\mathbf{w}), \mu_{x11}(x_{11}),\ldots,\mu_{xKN}(x_{KN}), \mu_{s11}(s_{11}),\ldots,\mu_{sKN}(s_{KN})) \rightarrow \text{Max}$$

subject to (7)

$$x_{ki} \le p_{ki} + \pi_{ki}, \quad k = 1,...,K;\ i = 1,...,N$$

$$s_{ki} = x_{ki} - \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kji} \ge d_{ki} - \delta_{ki},$$

$$k = 1,...,K;\ i = 1,...,N$$

$$\sum_{k=1}^{K} y_{kij} - M w_{ij} \le 0, \ i, j = 1,...,m$$

$$x_{ij}, y_{kij}, w_{ij} \ge 0, \ k = 1...,K, \ i, j = 1,...,N$$

we use the compromise objective function

$$\mu = \text{Min}\ (\mu_Z(\mathbf{x},\mathbf{w}), \mu_{x11}(x_{11}),\ldots,\mu_{xKN}(x_{KN}), \mu_{s11}(s_{11}),\ldots,\mu_{sKN}(s_{KN})) \qquad (8)$$

Then, we get a pareto-optimal compromise solution of (7) by solving the crisp mathematical programming system (9).

$$\lambda \rightarrow \text{Max}$$

subject to (9)

$$\lambda \le \mu_Z(\mathbf{x},\mathbf{w})$$

$$\lambda \le \mu_{xki}(x_{ki}), \quad k = 1,...,K;\ i = 1,...,N$$

$$\lambda \le \mu_{ski}(s_{ki}), \quad k = 1,...,K;\ i = 1,...,N$$

$$x_{ki} \le p_{ki} + \pi_{ki}, \quad k = 1,...,K;\ i = 1,...,N$$

$$s_{ki} = x_{ki} - \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kji} \ge d_{ki} + \delta_{ki},$$

$$k = 1,...,K;\ i = 1,...,N$$

$$\sum_{k=1}^{K} y_{kij} - M w_{ij} \le 0, \ i, j = 1,...,m$$

$$x_{ij}, y_{kij}, w_{ij} \ge 0, \ k = 1...,K, \ i, j = 1,...,N$$

As the management is only interested in a solution that satisfies all aspiration levels, it is sufficient to solve the following crisp linear LP-system, where $r = 1$; for details see [1; 2].

$$\lambda \rightarrow \text{Max}$$

subject to (10)

$$(z^A[r] - \underline{z})\lambda + (1 - \lambda_A)\left(\sum_{i=1}^{N}\sum_{k=1}^{K} C_{ki}x_{ki} + \sum_{i=1}^{N}\sum_{j=1}^{K} T_{ji}w_{ji}\right) \le z^A[r] - \lambda_A \underline{z}$$

$$(p^A_{ki}[r] - p_{ki})\lambda + (1 - \lambda_A)x_{ki} \le p^A_{ki}[r] - \lambda_A p_{ki}, \quad k = 1,...,K;\ i = 1,...,N$$

$$(d_{ki} - d^A_{ki}[r])\lambda - (1 - \lambda_A)\left(x_{ki} - \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kij} + \sum_{\substack{j=1 \\ j\neq i}}^{N} y_{kji}\right) \le \lambda_A d_{ki} - d^A_{ki}[r]$$

$$k = 1,...,K;\ i = 1,...,N$$

$$\sum_{k=1}^{K} y_{kij} - M w_{ij} \le 0, \ i, j = 1,...,m$$

$$\lambda, x_{ij}, y_{kij}, w_{ij} \ge 0, \ k = 1...,K, \ i, j = 1,...,N$$

Obviously, a solution of the system (10) with $\lambda \ge \lambda_A$ is a pareto-optimal solution of the systems (7) and (1) and fulfils all aspiration levels $z^A[1]$, $p^A_{ki}[1]$, $d^A_{ki}[1]$.

Moreover, if λ is greater than λ_A, the management can increase some of the aspiration levels to $z^A[2]$, $p^A_{ki}[2]$, $d^A_{ki}[2]$ and can calculate a new pareto-optimal solution by means of the revised system (10), and so on. When the management is satisfied with the non-integer solution, it is time for looking for an integer solution. Due to the soft constraints several integer solution will exist in the neighborhood of the non-integer solution.

4 Numerical example

The manufacturer HAM produces four products P1, P2, P3 and P4 in three product plants in **E**ssen, **K**rakow and **L**yon.

Table 1: Production costs in € per unit

	P1	**P2**	**P3**	**P4**
E	200	400	330	500
K	180	300	280	
L	220		300	570

Table 2: Transportation costs in € per truck

	E	**K**	**L**
E		1000	600
K	1000		1200
L	600	1200	

Table 3: Secure production capacities p_{ki}

	p_{1i}	p_{2i}	p_{3i}	p_{4i}
E	600	450	250	250
K	350	200	250	
L	250		400	150

ISBN: 978-989-95079-6-8

Table 4: Maximal production capacities $p_{ki} + \pi_{ki}$

	$p_{1i} + \pi_{1i}$	$p_{2i} + \pi_{2i}$	$p_{3i} + \pi_{3i}$	$p_{4i} + \pi_{4i}$
E	750	500	300	300
K	500	300	300	
L	300		500	200

Table 5: Minimal demands $d_{ki} - \delta_{ki}$

	$d_{1i} - \delta_{1i}$	$d_{2i} - \delta_{2i}$	$d_{3i} - \delta_{3i}$	$d_{4i} - \delta_{4i}$
E	650	300	300	150
K	450	100	200	80
L	150	150	250	100

Table 6: Maximal demands d_{ki}

	d_{1i}	d_{2i}	d_{3i}	d_{4i}
E	750	400	350	200
K	350	150	200	120
L	200	200	400	130

Using the LP-systems (2) and (3), we get the minimal total cost $\underline{z}$ = 864,100 € and the maximal cost $\overline{z}$ = 1,091,700 €. With all these information the management of HAM specifies for the total cost the aspiration level z[1] = 950,000 € and for the outputs and the demands the following aspiration levels.

Table 7: Aspiration levels p_{ki} [1]

output	p_{1i} [1]	p_{2i} [1]	p_{3i} [1]	p_{4i} [1]
E	720	480	280	280
K	450	270	280	
L	280		470	180

Table 8: Aspiration levels d_{ki} [1]

demand	d_{1i} [1]	d_{2i} [1]	d_{3i} [1]	d_{4i} [1]
E	670	350	330	180
K	280	120	200	100
L	180	170	300	110

As $\lambda = 0.53 > \lambda_A = 0.5$, the solution of the corresponding LP-system (10) fulfills all aspiration levels. E. g. the total cost of this plan is 945,082.50 €.
Moreover this result indicates that it is possible to improve the aspiration levels. For simplification we assume that the management is only interested in lower cost and decides to reduce the total cost to z[2] = 940,000 €.

Even the solution of the revised system (10) fulfils all aspiration levels. We have $\lambda = 0.5077$, z = 938,712.20 € and the following production and transportation plan.

Table 9: Production and transportation plan

x1e	x1k	x1l	x2e	x2k	x3e
671,05	280,92	180,26	372,24	269,08	250,79
x3k	x3l	x4e	x4l	wek	wel
279,61	301,32	279,61	111,18	4,97	1,08
wke	wkl	wle	wlk	y1ek	y1el
3,98	7,43	0,00	0,05	0,00	0,00
y1ke	y1kl	y1le	y1lk	y2ek	y2el
0,00	0,00	0,00	0,00	0,00	21,58
y2ke	y2kl	y3ek	y3el	y3ke	y3kl
0,00	148,68	0,00	0,00	79,61	0,00
y3le	y3lk	y4ek	y4el	y4le	y4lk
0,00	0,00	99,34	0,00	0,00	0,92

As we have soft constraints it is very simple to derive an integer plan that fulfils all aspiration levels and actually leads to lower total cost z= 937,100 €.

Table 10: Integer production and transportation plan

x1e	x1k	x1l	x2e	x2k	x3e
670	280	180	370	270	250
x3k	x3l	x4e	x4l	wek	wel
280	300	280	110	5	1
wke	wkl	wle	wlk	y1ek	y1el
4	8	0	0	0	0
y1ke	y1kl	y1le	y1lk	y2ek	y2el
0	0	0	0	0	20
y2ke	y2kl	y3ek	y3el	y3ke	y3kl
0	150	0	0	80	0
y3le	y3lk	y4ek	y4el	y4le	y4lk
0	0	100	0	0	0

5 Conclusions

Fuzzy mathematical programming systems offer the possibility to model real problems as precisely as a decision maker is able to describe them. In doing so non-satisfying modeling can be avoided. Another advantage of fuzzy models is the fact that (mixed) integer programming problems can be solved very easily because the boundaries are not crisp but fuzzy and points with integer variables in the neighborhood of an optimal solution are feasible in general.

In this paper we have assumed that only the right-hand sides of some constraints are not known exactly. The model can be extended to the case that additionally coefficients of the objective function or of the constraints are described in form of fuzzy intervals. Moreover, it is possible to allow several objective functions. All these systems can adequately be solved by means of the inter-active algorithm FULPAL, see [1; 2; 3].

References

[1] H. Rommelfanger. FULPAL 2.0 - An Interactive Algorithm for Solving Multicriteria Fuzzy Linear Programs Controlled by Aspiration Levels. In: Scheigert D. (Ed.) Methods of Multicriteria Decision Theory. Pfalzakademie Lamprecht, 21-34, 1995.

[2] H. Rommelfanger and R. Slowinski. Fuzzy Linear Programming with single or multiple Objective Functions. In: Slowinski R. (Ed.): Fuzzy Sets in Decision Analysis, Operations Research and Statistics, Kluwer Academic Publishers Norwell, Massachusetts, USA,179-213, 1998.

[3] H. Rommelfanger. The Advantages of Fuzzy Optimization Models in Practical Use. *Fuzzy Optimization and Decision Making 3*, 295-309, 2004.

ISBN: 978-989-95079-6-8

The Use of Interval-Valued Probability Measures in Fuzzy Linear Programming: A Constraint Set Approach

Phantipa Thipwiwatpotjana
Department of Mathematical sciences
University of Colorado at Denver
Phantipa.Thip@ucdenver.edu

Weldon A. Lodwick
Department of Mathematical sciences
University of Colorado at Denver
Weldon.Lodwick@ucdenver.edu

***Abstract*—*This paper uses a constraint set approach to linear programming problems with equality constraints whose coefficients and/or right-hand side values could be uncertain. We consider three types of uncertainty: probabilistic, fuzzy/possibilistic, and interval. The constraint set generated by the linear constraints under uncertainty is ill-defined and difficult to generate. Our approach computes an inner constraint set and an outer constraint set. Optimization is then carried out using these two sets using interval-valued probability approaches. We discuss the requisite associated semantics.*

Keywords—inverval-valued probability measure, random sets, inner constraint set, outer constraint set.

I. Introduction

In order to solve linear programming problems under uncertainty, many researchers convert the problem into a sequence of deterministic problems. The approach in this paper is to work with the constraint set of the problem.

It is difficult (exponentially hard) to find the exact shape of the feasible set of uncertain constraints, as we shall see. Therefore we create two sets (boxes); one represents the set of solutions guaranteed to have a corresponding data to satisfy the constraints while the other is the set that contains all solutions. They are called the *inner set* and the *outer set*, respectively. We use these two sets in our optimization problem.

We present how to obtain these two sets in section III after giving some useful definitions. In section IV, we create an interval-valued probability measure (IVPM) by a pair of possibility and necessity measures on the space between the inner and outer sets.

In section V, we solve an uncertain optimization problem, at each α-level of the uncertain data, on the inner box and give a guaranteed bound on the objective value. We use an IVPM to get the bound on the objective value with some γ degree of occurrence. The final section articulates some conclusions.

II. Background and Definitions

We consider linear programming (LP) problems with uncertain parameters of the form

$$\min_x \breve{c}^{\mathrm{T}} x \quad \text{s.t.} \quad \breve{A}x = \breve{b}. \tag{1}$$

$\breve{A}, \breve{b}$ and $\breve{c}$ may be (or are) uncertain. Without knowing the type of uncertainty, we use $\breve{b}_j$ and $\breve{c}_j$ as the j^{th} component of the uncertain vectors $\breve{b}$ and $\breve{c}$, respectively. Similarly, $\breve{a}_{ij}$ is the (i,j) component of the uncertain matrix $\breve{A}$.

The main types of uncertainties are probabilistic, fuzzy/possibilistic and interval. When we know the type of uncertain component $\breve{a}_{ij}$, we can specify the uncertainty by the following symbols. (The same symbols also apply for components $\breve{b}_j$ and $\breve{c}_j$.)

- $\tilde{a}_{ij} \equiv$ a fuzzy component a_{ij},
- $\hat{a}_{ij} \equiv$ a possibilistic component a_{ij},
- $\breve{a}_{ij} \equiv$ a probabilistic component a_{ij},
- $[a_{ij}] = \left[\underline{a}_{ij}, \overline{a}_{ij}\right] \equiv$ an interval component a_{ij}.

Definition 2.1: Given an LP problem (1), the constraint set $\Omega_{\exists\exists} := \left\{ x_{Ab} \mid \exists A \in \breve{A}, \exists b \in \breve{b} \ : \ Ax_{Ab} = b \right\}$ is said to be the set of all solution x's for the constraints of (1).
The set $\Omega_{\exists\exists}$ may be (usually is) a non-convex set.

Definition 2.2: A finite *random set* on S is a pair $(\mathscr{F}, m)$ where $\mathscr{F}$ is a finite family of distinct non-empty subsets of S and m is a mapping $\mathscr{F} \to [0,1]$ such that $\sum_{A \in \mathscr{F}} m(A) = 1$.

We can also define a random set on S when S is infinite by using the measure space $(S, \mathscr{S}, m)$, where $\mathscr{S}$ is a σ-algebra of S. Therefore, a random set on an infinite set S is a pair $(\mathscr{F}, m)$ where $\mathscr{F}$ is a subset of $\mathscr{S}$ such that $\sum_{A \in \mathscr{F}} m(A) = 1$.

In real applications, we might not know (with certainty) the probability measure for our problems. Lodwick and Jamison [5] use an IVPM, $i_{\breve{m}}(A) = \left[i^-_{\breve{m}}(A), i^+_{\breve{m}}(A)\right]$, to measure a partial representation for an unknown probability measure. The original paper for the idea of IVPM is Weichselberger [12]. We use the following notation and information throughout the paper unless stated otherwise:

- The arithmetic operations applied to intervals are those of interval arithmetic [7] and where tractable, constraint interval arithmetic [3], [6].
- The set of all intervals contained in $[0,1]$ is denoted as

$$\mathrm{Int}_{[0,1]} \equiv \{[a,b] \mid 0 \le a \le b \le 1\}.$$

- $\mathcal{S}$ denotes the universal set and $\mathcal{A}$ is a σ-algebra of $\mathcal{S}$.

Definition 2.3: (Weichselberger [12]) Given measurable space $(\mathcal{S}, \mathcal{A})$, an interval valued function $i_{\breve{m}} : A \subseteq \mathcal{A} \to \mathrm{Int}_{[0,1]}$ is called an *R-probability* if:

- $i_{\breve{m}}(A) = \left[i^-_{\breve{m}}(A), i^+_{\breve{m}}(A)\right] \subseteq [0,1]$,

ISBN: 978-989-95079-6-8

- $\exists$ a probability measure, Pr, on $\mathcal{A}$ such that $\forall A \in \mathcal{A}$, $\Pr(A) \in i_{\check{m}}(A)$.

Definition 2.4: A function $p : \mathcal{S} \to [0,1]$ is called a *regular possibility distribution function* if it is a possibility such that

$$\sup\{p(x) \mid x \in \mathcal{S}\} = 1. \tag{2}$$

Possibility distribution functions (see [11]) define a possibility measure, $Pos : \mathcal{S} \to [0,1]$ where

$$Pos(A) = \sup\{p(x) \mid x \in A\} \tag{3}$$

and its dual necessity measure is

$$Nec(A) = 1 - Pos(A^c), \tag{4}$$

where $\sup\{p(x) \mid x \in \emptyset\} = 0$.
In [4], it is shown that possibility distributions can be constructed which satisfy the following consistency definition.

Definition 2.5: Let $p : \mathcal{S} \to [0,1]$ be a regular possibility distribution function with associated possibility measure Pos and necessity measure Nec. Then p is said to be *consistent* with random variable X if for every measurable set A, $Nec(A) \leq \Pr(X \in A) \leq Pos(A)$. Note that this necessity measure Nec may not be the dual of Pos in this definition.

An R-probability from definition 2.3 is an *interval-valued probability measure* (IVPM) where $i_{\check{m}}^-$ and $i_{\check{m}}^+$ are constructed from a possibility (fuzzy) distribution function. To see this is an IVPM see [4].

The R-probability function $i_{\check{m}}$ in definition 2.3 is used to define IVPMs. A possibility and necessity pair, $i_{\hat{m}}(A) = [Nec(A), Pos(A)]$, constructed by definition 2.5 is able to bound an unknown probability of interest. Therefore it can be used to define an IVPM.

Definition 2.6: The Interval-Valued Probability Measure (IVPM) defined from possibility and necessity measures is $i_{\hat{m}}(A) = [Nec(A), Pos(A)]$.
The reader could find more explanations, examples and a construction of an IVPM in [5] .

III. Generating Two Sets from Equality Constraint Set

In this section, we consider only constraints of (1),

$$\breve{A}x = \breve{b}. \tag{5}$$

Although the set of all solution x's, $\Omega_{\exists\exists}$, for (5) is difficult to construct, it is relatively easy to generate two random sets for each α-level called *inner set*, B_α^I, and *outer set*, B_α^O, such that $B_\alpha^I \subseteq B_\alpha^O$. Each x in B_α^I has a particular matrix $A_x \in A_\alpha$ and vector $b_x \in b_\alpha$, such that $A_x x = b_x$. An outer set, B_α^O, represents a space where x should be. This implies that $\Omega_{\exists\exists} \subseteq B^O$.

A. Interval case

We drop the α subscript when we work with interval version of equation (5).

A system

$$[A]x = [b] \tag{6}$$

is called *solvable* if $Ax = b$ has a solution for each $A \in [A]$ and for each $b \in [b]$, where $[A] = \left[\underline{A}, \overline{A}\right]$ and $[b] = \left[\underline{b}, \overline{b}\right]$.

Gay [2] gives a method for solving interval linear equations when $[A]$ is a regular square interval uncertain matrix, i.e., A is invertible, $\forall A \in [A]$. Rohn [8] provides a powerful theorem for any $m \times n$ interval matrix $[A]$ that guarantees solvability by testing nonnegative solutions of a finite number of linear systems.

The following notations are useful for theorem 3.1 below.

- $A_c = \frac{(\underline{A}+\overline{A})}{2}$, the center matrix.
- $\Delta = \frac{(\overline{A}-\underline{A})}{2}$, the radius matrix.
- $b_c = \frac{(\underline{b}+\overline{b})}{2}$ and $\delta = \frac{(\overline{b}-\underline{b})}{2}$.
- $Y_m = \{y \in \mathrm{R}^m \mid y_j \in \{-1, 1\}, \forall j = 1, 2, \ldots m\}$.
- $T_y = \mathrm{diag}(y_1, y_2, \ldots, y_m)$, for each $y \in Y_m$.
- $\mathrm{Conv}(X) = \{\gamma p_1 + (1-\gamma)p_2 \mid p_1, p_2 \in X,\ \gamma \geq 0\}$.

Theorem 3.1 (see Rohn [8]): A system of linear equations $[A]x = [b]$ is solvable if and only if for each $y \in Y_m$ the system

$$\left.\begin{array}{c}(A_c - T_y\Delta)\,x^1 - (A_c + T_y\Delta)\,x^2 = b_c + T_y\delta, \\ x^1 \geq 0,\ \ x^2 \geq 0\end{array}\right\} \tag{7}$$

has a solution x_y^1, x_y^2. Moreover, if this is the case, then for each $A \in [A], b \in [b]$ the system $Ax = b$ has a solution in the set $\mathbf{C} := \mathrm{Conv}\left\{x_y^1 - x_y^2 \mid y \in Y_m\right\}$.

Theorem 3.1 tells us that there is a solution for the system $Ax = b$ in the set $\mathbf{C}$. However, it does not mean that all the solutions of the system $Ax = b$ are in $\mathbf{C}$, especially when A is not a square matrix.

Corollary 3.1 (see [1]): A system of linear equations $[A]x = [b],\ x \geq 0$ is solvable if and only if for each $y \in Y_m$ the system

$$(A_c - T_y\Delta)\,x = b_c + T_y\delta, \tag{8}$$

has a nonnegative solution x_y. Moreover, if this is the case, then for each $A \in [A], b \in [b]$ the system $Ax = b,\ x \geq 0$ has a solution in the set $\mathrm{Conv}\{x_y \mid y \in Y_m\}$.

Now we are going to generate outer and inner sets given the assumption that our constraint set is solvable.

1) An outer set, B^O: We will consider two cases when $m = n$ and when $m < n$.

- When $m = n$. Suppose that a system of linear equation (6) is solvable by theorem 3.1 and $[A]$ is a regular matrix, we can create a rectangular box that contains all the solution points $x_y^1 - x_y^2$ of the system (7). To see this simply take $\max / \min$ of components. We prefer to use the rectangular box as an outer set over the set $\mathrm{Conv}\left\{x_y^1 - x_y^2;\ y \in Y_m\right\}$ because it is easier to construct. We can use $2^{m^2}2^m$ endpoint equations to generate $\mathbf{C}$, hence creating a rectangular box, B^O.
  We will not consider the case when $[A]$ is not a regular matrix. If there exists a singular matrix $A \in [A]$, then B^O is an unbounded set.
- When $m < n$. Each system $Ax = b$ in the constraint set (6) provides infinitely many solutions in the form of $n - r$ parameters where $\mathrm{rank}(A) = r$. For

ISBN: 978-989-95079-6-8

example, $2x_1 + x_2 = 3$ has solutions as set $S = \{(t, 3-2t) \mid t \in \mathrm{R}\}$. This leads unbounded objective value for the linear programming (1). However, if the constraint set is in the form

$$[A]x = [b];\ x \geq 0, \tag{9}$$

the LP might be bounded (even if it has unbounded feasible set).

Suppose that a system of linear equation (9) is solvable by corollary 3.1. Each system (8) has infinitely many solutions of dimension $n - r$ which are bounded below by $x = 0$. For each $y \in Y_m$, let $S_y = \{x_y \mid (A_c - T_y\Delta)\, x_y = b_c + T_y\delta,\ x_y \geq 0\}$. Then an outer set $B^O = \mathrm{Conv}\{S_y \mid y \in Y_m\}$.

2) An inner set, B^I: To build an inner set, B^I, we also consider two cases when $m = n$ and when $m < n$.

- When $m = n$, we work through the following steps.
  - Step 1 : Choose x_* as one of the point $x_y^1 - x_y^2$ in theorem 3.1 that creates B^O.
  - Step 2 : Find A_* and b_* such that $A_*x_* = b_*$ The matrix A_* and vector b_* is one of the endpoint matrices and vectors created by endpoints of $[A]_{m\times n}$ and $[b]_{n\times 1}$, respectively.
  - Step 3 : An inner set, B^I, is the set $\mathrm{Conv}\{x^* \mid A_*x^* = b^*\}$. There are $\begin{pmatrix} m \\ 0 \end{pmatrix} + \begin{pmatrix} m \\ 1 \end{pmatrix} + \begin{pmatrix} m \\ 2 \end{pmatrix} + \ldots + \begin{pmatrix} m \\ m \end{pmatrix}$ different b^*'s created by changing any component(s) of b_* to the other endpoint(s). For example, if $[b] = \begin{bmatrix} [1,3] \\ [0,2] \end{bmatrix}$ and $b_* = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, then the 4 choices of b^* are b_*, $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$.
- When $m < n$, we work through the following steps.
  - Step 1 : Choose two elements y^1 and y^2 in Y_m.
  - Step 2 : $B^I = \mathrm{Conv}\left\{S_{y^1}, S_{y^2}\right\}$.

To see that these B^O and B^I are random sets, we define $m(B^I) = 1$ and $m(K) = 0$ for other σ-algebra K's of B^O.

B. General case

When we have mixed uncertain information for (5); for instance, when $\breve{A}$ is an uncertain matrix of 1×2 dimension such that $\breve{a}_{11} := \hat{a}_{11}$ with some possibility distribution and $\breve{a}_{12} := [\underline{a}_{12}, \overline{a}_{12}]$, we need to consider each component separately by using α-levels. The semantics for each α-level of different types of uncertainty is as follow. (Similar idea applies for $\breve{b}_i$).

1) Fuzzy uncertainty: An α-level set, $\left[\underline{a}_{ij_\alpha}, \overline{a}_{ij_\alpha}\right]$, of a fuzzy number $\tilde{a}_{ij}$ represents an interval that every element in it has degree of satisfaction equal to or greater than α.

2) Possibilistic uncertainty: An α-level set, $\left[\underline{a}_{ij_\alpha}, \overline{a}_{ij_\alpha}\right]$, of a possibility distribution $\hat{a}_{ij}$ expresses that every element in the set has degree of occurrence equal to or greater than α.

3) Probabilistic uncertainty: A probabilistic component $\check{a}_{ij}$ with the corresponding density function f. Define $h := \max\{f(a) \mid a \in \check{a}_{ij}\}$, an interval $\left[\underline{a}_{ij_{\alpha h}}, \overline{a}_{ij_{\alpha h}}\right]$ coincides with the α-level set of two previous cases and it means that $\forall a \in \left[\underline{a}_{ij_{\alpha h}}, \overline{a}_{ij_{\alpha h}}\right],\ f(a) \geq \alpha h$.

At each α-level, the uncertainty acts as an interval. However, each interval has a history of the original type of uncertainty, (III-B.1, III-B.2 or III-B.3). So the semantics must be different from the interval case.

We present B^I_α and B^O_α when $\alpha = 0, \frac{1}{k}, \frac{2}{k}, \ldots, 1,\ k \in \mathrm{N}$. We define $[A]_\alpha$ as a matrix uncertainty whose each component is restricted to that particual α-level. (Similar definition applies to $[b]_\alpha$). Moreover, if system $[A]_\alpha x = [b]_\alpha$ is solvable, we write $\mathbf{C}$ in the result of the theorem 3.1 as $\mathbf{C}_\alpha$.

To generate B^O_α, we start with $\alpha = 0$. That means we consider a system $[A]_0 x = [b]_0$. Suppose that this system is solvable, we can create B^O_0 as shown in the interval case III-A.1. Therefore it is not hard to see that $B^O_{\frac{i}{k}} \supseteq B^O_{\frac{j}{k}}, \forall\ i \leq j \leq k,\ i, j, k \in \mathrm{N}$.

In the case of B^I_α, we adopt the construction in III-A.2 as follows.

- When $m = n$:
  - Step 1 : When $\alpha = 0, (i = 0)$, we deal with $[A]_0 x = [b]_0$. We use the method in III-A.2 for finding B^I_0. We also get x_*, A_* and b_* s.t. $A_*x_* = b_*$.
  - Step 2 : Set $c := x_*$ as a reference point.
  - Step 3 : Set $i := i + 1$. Now we deal with $[A]_{\frac{i}{k}} x = [b]_{\frac{i}{k}}$. Choose the point x_* that creates $B^O_{\frac{i}{k}}$ so that $\|x_* - c\| = \min_x \{\|x - c\|\}$, where x is the other point $x_y^1 - x_y^2$ of $B^O_{\frac{i}{k}}$. Follow Step 2 in III-A.2.
  - Step 4 : An inner set $B^I_{\frac{i}{k}} = B^I_{\frac{i-1}{k}} \cap \mathrm{Conv}\{x^* \mid A_*x^* = b^*\}$ where b^*'s are the different endpoints of $[b]_{\frac{i}{k}}$.
  - Step 5 : Go back to Step 2 , or quit when $i = k$.
- When $m < n$: At each i^{th} iteration, we need to choose $y^1, y^2 \in Y_m$ so that $B^I_{\frac{i-1}{k}} \cap \mathrm{Conv}\left\{S_{y^1}, S_{y^2}\right\} \neq \emptyset$. Then $B^I_{\frac{i}{k}} = B^I_{\frac{i-1}{k}} \cap \mathrm{Conv}\left\{S_{y^1}, S_{y^2}\right\}$.

Using this construction, we have $B^I_{\frac{i}{k}} \supseteq B^I_{\frac{j}{k}};\ \forall i \leq j \leq k,\ i, j, k \in \mathrm{N}$.

Therefore the outer and inner sets maintain the following properties.

Theorem 3.2: For all $i \leq j \leq k$ and $i, j, k \in \mathrm{N}$,

$$B^O_{\frac{j}{k}} \subseteq B^O_{\frac{i}{k}} \text{ and } B^I_{\frac{j}{k}} \subseteq B^I_{\frac{i}{k}}.$$

Proof: Follow directly from the construction of the sets. $\diamond$

Using $m(B^I_{\frac{i}{k}}) = \frac{1}{k}, \forall\ i = 1, 2, \ldots, k$ and $m(K) = 0$ for other σ-algebra K's of B^O_0, we create the random set $\left\{B^0_{\frac{i}{k}}, i = 0, 1, \ldots, k\right\}$ and the random set $\left\{B^I_{\frac{i}{k}}, i = 0, 1, \ldots, k\right\}$.

ISBN: 978-989-95079-6-8

IV. Interval-valued probability measure on the set $B_\alpha^O \backslash B_\alpha^I$

In this section we consider each α-level separately. For convenience, we drop the subscript α of B_α^O and B_α^I.

After we have an inner set B^I and an outer set B^O, we consider the space between these two sets, $B^O \backslash B^I$, as our universal space of consideration. Let X be a random variable with an unknown distribution. For any space $B \subseteq B^O \backslash B^I$, there is a probability that $X = x \in \Omega_{\exists\exists}$ is in B. We write this probability as $\Pr(x \in B)$. However, we lack information to justify for sure what the value is.

Each of the figures 1(a) and 1(b) illustrates a geometric interpretation of B^I and B^O in R^2 of a system $[A]x = [b]$ where $[A]$ is regular with $m = n = 2$. Inner sets are the black areas while outer sets are the areas of rectangular boxes (that look like squares in the figures below). It can be difficult to give a geometric interpretation of B^I and B^O, especially in higher dimension.

The points p, q, r and s that create B^O in each figure are $x_y^1 - x_y^2$ where (x_y^1, x_y^2) is a solution of each system (7) in theorem 3.1. As we said in the construction of an outer set that it is easier to get a rectangular box in practice. Therefore, we use the rectangular box in 1(b) to represent B^O instead of Conv $\{p, q, r, s\}$.

When we choose $x_* = p$ in the construction of an inner set, we obtain an inner set as the parallelogram black set in both 1(a) and 1(b) as can be seen.

A point z is more likely to be in the set $\Omega_{\exists\exists}$ than a point y that is further away from B^I. We can create an possibility distribution of a particular nested sets by using this fact.

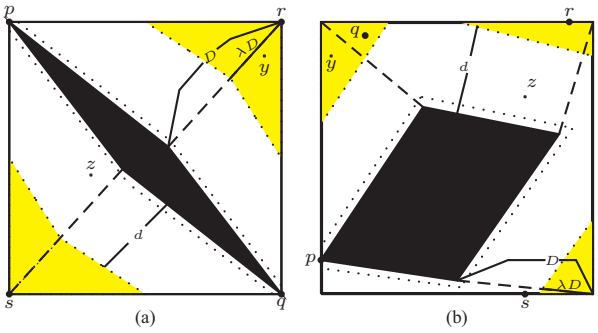


Fig. 1. Inner and outer boxes: inner boxes are the black areas and outer boxes are the rectangular sets.

We choose two corresponding corner points of B^I and B^O that make the longest distance, i.e., the length of one of dashed lines in figure 1(a) (or (b)). We use D to represent this longest distance. Let $0 \leq \lambda \leq 1$ be a scalar. An easy possibility distribution of the nested sets $V(\lambda D)$ could be a function in the term of λ, for instance,

$$\text{Poss}\,(V(\lambda D)) = \lambda, \quad \text{when} \;\; 0 \leq \lambda \leq 1. \tag{10}$$

$V(\lambda D)$ in the domain of the function (10) represents the yellow areas (volume for higher dimensions) in the Figure 1. For $\lambda_1 \leq \lambda_2$, we have $V(\lambda_1 D) \subseteq V(\lambda_2 D)$. When λ is increasing, the possibility that $V(\lambda D)$ contains solutions $x \in \Omega_{\exists\exists}$ is also increasing.

By restricting to the nested domain, we would be able to work in the white space between B^I and B^O. The white space has the width from the edge of the black and yellow areas to be d, as shown in figure 1. The distance d is less than or equal to $(1 - \lambda)D$. We also can use an appropriate necessity distribution, Nec $(V(\lambda D))$.

However, in order to satisfy the definition of an IVPM, we need to make sure that the unknown probability measure is in the bound of our necessity and possibility measures. Therefore, the equation (10) may not be appropriate for most of the systems. We leave this part to the decision maker to come up with a reasonable nestedness necessity and possibility distributions that (s)he thinks they are an lower and upper bounds on the unknown probability.

Thus an IVPM with the restricted nested domain $V(\lambda D)$ is

$$i_{\text{m}}(V(\lambda D)) = [\text{Nec}\,(V(\lambda D)), \text{Poss}\,(V(\lambda D))] \tag{11}$$

where Poss $(V(\lambda D))$ and Nec $(V(\lambda D))$ are defined to cover the unknown probability by the expert's opinion. In the next section, we show the use this IVPM in LP problems.

V. Optimization on boxes using IVPMs

To deal with an optimization problem (1), it would make sense to consider on each α-level of each uncertainty.

As we know for sure that the solutions in B_α^I have their correspondent matrix A and vector b, one approach for this problem (1), restricted on the constraint set B_α^I, can be

$$\min_{x \in B_\alpha^I} \check{c}_\alpha x, \tag{12}$$

where $\check{c}_\alpha = [\underline{c}_\alpha, \overline{c}_\alpha]$. By solving $\min \underline{c}_\alpha x$ and $\min \overline{c}_\alpha x$ on the convex polyhedral B_α^I, we obtain the guaranteed bound on the objective value.

Moreover, since $\Omega_{\exists\exists_\alpha} \supseteq B_\alpha^I$, we use the idea in IV to deal with the solution between inner and outer sets. We use a nestedness possibility distribution in a restricted IVPM obtained form IV to determine the space of consideration. We illustrate through the following examples.

Example 5.1:

$$\left.\begin{array}{ll} \min & [2,4]x_1 + [1,3]x_2 \\ \text{s.t.} & \widehat{3}x_1 + [-2,1]x_2 = [-2,2] \\ & \widetilde{0.5}x_1 + \check{3}x_2 = \widehat{0}, \end{array}\right\} \tag{13}$$

where the uncertain data are presented in the figure 2 above. For this example, we work at the α-level when $\alpha = 0.5$. Therefore, the problem (13) with $\alpha = 0.5$ becomes

$$\left.\begin{array}{ll} \min & [2,4]x_1 + [1,3]x_2 \\ \text{s.t.} & [2,4]x_1 + [-2,1]x_2 = [-2,2] \\ & [-1,2]x_1 + [2,4]x_2 = [-2,2]. \end{array}\right\} \tag{14}$$

Figure 3 shows the inner box and outer box for (14). So the guaranteed bound on the objective value of

$$\min_{x \in B_{0.5}^I} [2,4]x_1 + [1,3]x_2,$$

is $[-25, 25]$. From figure 3, we get $D = 5$ units. By using possibility distribution as in (10) for simplicity, if we want

ISBN: 978-989-95079-6-8

degree of possibility of an IVPM (as constructed in IV) to be 0.6, then

$$\min_{x\in\left(B^O_{0.5}\setminus B^I_{0.5}\right)_{0.6}} [2,4]x_1 + [1,3]x_2$$

would have the guaranteed bound, $[-28, 28]$, on the

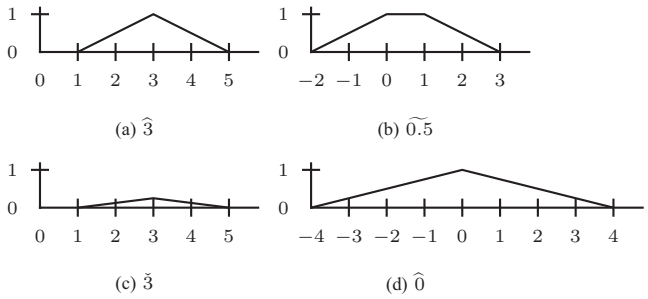


Fig. 2. Uncertain data for example 5.1.

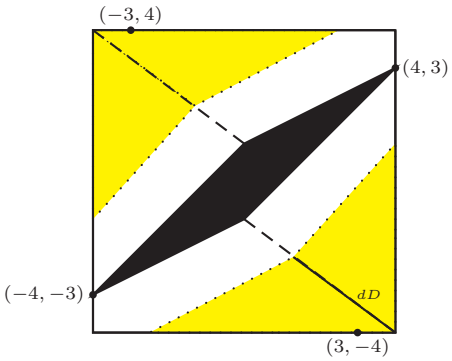


Fig. 3. Inner and outer boxes for the problem (14). Inner box is the black area. Outer box is the square generated by points (-3,4), (4,3), (3,-4) and (-4,-3).

objective value with 0.6 degree of occurrence. Similarly, we can have the guaranteed bound with a certain degree of belief by using an appropriate necessity measure of the IVPM.

Example 5.2: A drug company wants to create two supplement products X and Y that extract mainly from a super food F. This super food F has a fuzzy color 'orange-red-pink'. Some of fully grown F may have the color red to pink or could be more in the orange shade. The company will be totally satisfied if each of the plant F has red color. So the fuzzy color (from the company's point of view) has the membership function as in figure 4(a). There is only one place P that grows F. So P can charge the price that he wants. Although P can not predict his harvest's color, he also prefers F to be red to please his customer. P sets the price of each F relating to its fuzzy color as in figure 4(b).

The company found out that each 1g of F (depend on the color) contains an amount of Omega III and Zinc (in mg) with possibility distribution in figure 4(c) and 4(d), respectively. The company has a secret formula for X and Y. To produce X, the company need to make sure that the basic ingredient has Omega III in the range of [1280,1300]mg. The product Y need to have $\widehat{586}$mg of Zinc, (Fig. 5(a)).

Therefore the company needs to add pure Omega III and Zinc from a substance T if F does not provide enough as required. One gram of T obtains 5mg of Omega III and $\check{1.8}$mg of Zinc , (Fig. 5(b)), which means Zinc in T is not certain. The price of T depends on the market which is given by $\$\check{3}$ per 1g of T, (Fig. 5(c)). We can formulate a

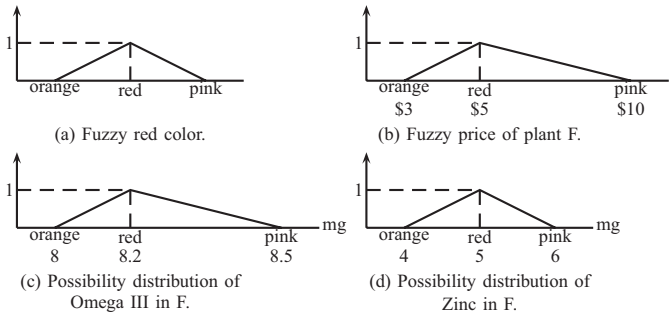


Fig. 4. Membership functions and possibility distributions.

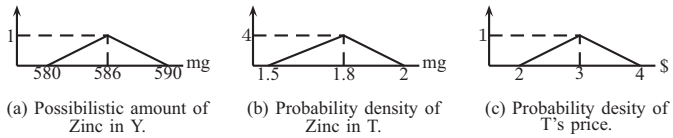


Fig. 5. Possibility distribution and probability density functions.

linear program to minimize the cost to produce these two supplements.

$$\begin{aligned} \min \quad & \widetilde{5}\mathrm{f} + \check{3}\mathrm{t} \\ \text{s.t.} \quad & \widehat{8.2}\mathrm{f} + 5\mathrm{t} = [1280, 1300] \\ & \widehat{5}\mathrm{f} + \check{1.8}\mathrm{t} = \widehat{586} \\ & \mathrm{f}, \mathrm{t} \geq 0, \end{aligned}$$

where f is the amount of grams of plant F and t is the amount of grams of substance T. When $\alpha = 0.5$, the problem becomes

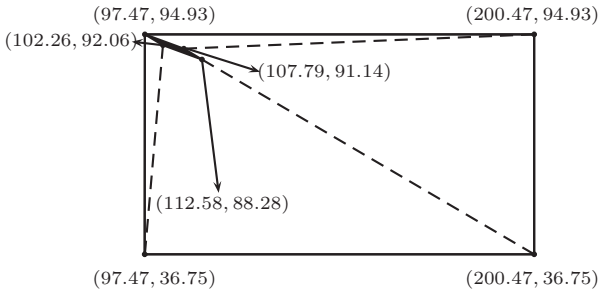


Fig. 6. Inner and outer boxes for the problem (15). Each point represents (t,f).

$$\left.\begin{array}{ll} \min & [4, 7.5]\mathrm{f} + [2.5, 3.5]\mathrm{t} \\ \text{s.t.} & [8.1, 8.35]\mathrm{f} + 5\mathrm{t} = [1280, 1300] \\ & [4.5, 5.5]\mathrm{f} + [1.65, 1.9]\mathrm{t} = [583, 588]. \end{array}\right\} \quad (15)$$

The inner and outer sets of the system (15) is shown in figure 6. The inner set is the black space generated by points (97.47,94.93), (102.26,92.06), (112.58,88.28) and (107.79,91.14). The outer set is the rectangular box. However, points (200.47,94.93) and (97.47,36.75) are not feasible for any $A \in [A]$ and $b \in [B]$.

ISBN: 978-989-95079-6-8

The semantics explanation for this example is as follows. The company sets the degree of satisfaction of F's price to be at least 0.5 together with 75% confidence interval of probabilistic price of T. It also wants 0.5 or more degree of occurrences of Omega III and Zinc in the food F and 93.75% confidence interval of probabilistic amount of Zinc in T. By its secret formula, the company sets at least 0.5 degree of occurrence of Zinc in product Y. Then the guaranteed bound on the objective value with $\alpha = 0.5$ is $\$[623.40, 1060.82]$ by using the corner point (97.47,94.93) and (107.79,91.14) of the black space. We also can find the guaranteed bound for some curtain degree of occurrence in the similar way as presented in the example 5.1.

VI. Conclusion and Further Research

From the construction of B^I_α and B^O_α, we can use B^I_α to find a guaranteed set of solutions and bound on an objective value of an optimization problem. We use the set $B^O_\alpha \setminus B^I_\alpha$ to create an IVPM, then use a possibility to get a set of solutions with some level γ of occurrence. This paper considered merely equality constraints. The next step is to consider the inequality case.

References

[1] M. Fiedler, J. Nedoma, J. Ramík, J. Rohn and K. Zimmermann, "Linear optimization problems with inexact data" *Springer,* 2006.

[2] D. M. Gay, "Solving interval linear equation" *SIAM J. Number Anal.* no. 4, vol.19, pp. 858-870, 1982.

[3] Edited by P. W. Hawkes, "Advances in Imaging and electron physics" *Elsevier Academic Press, NY* vol. 148, pp. 75-192, 2007.

[4] W. A. Lodwick and K. D. Jamison, "The construction of consistent possibility and necessity measures" *Fuzzy Sets and System.* vol. 312, pp. 1-10, 2002.

[5] W. A. Lodwick and K. D. Jamison, "Interval valued probability in the analysis of problems containing a mixture of possibilistic, probabilistic and interval uncertainty" *Fuzzy Information Processing Society, 2006. NAFIPS 2006. Annual meeting of the North American.* pp. 617-620.

[6] W. A. Lodwick, "Constrained interval arithmetic" *CCM Report, 138.*

[7] R. E. Moore, *Methods and Applications of Interval Analysis.* SIAM, Philadelphia, 1979.

[8] J. Rohn, "Solvability of systems of linear interval equations" *SIAM J. Matrix Anal. Appl.* no.1, vol. 25, pp. 237-245, 2003.

[9] P. Thipwiwatpotjana and W. A. Lodwick, "Algorithm for solving optimization problems using interval valued probability measure" *2008 NAFIPS Conference Proceedings*

[10] P. Walley, *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

[11] Z. Wang and G. J. Klir, *Fuzzy Measure Theory.* Plenum Press, New York, 1992.

[12] K. Weichselberger, "The theory of interval-probability as a unifying concept for uncertainty," *International Journal of Approximate Reasoning.* no.24, pp. 149-170, 2000.

ISBN: 978-989-95079-6-8

Human perceptions versus computational perceptions in Computational Theory of Perceptions

Gracian Trivino[1] Alejandro Sobrino[2]

1.European Centre for Soft Computing
Asturias, Spain
2.University of Santiago de Compostela
Galicia, Spain
Email: gracian.trivino@softcomputing.es, alejandro.sobrino@usc.es

Abstract— *The development of a computational system based on the Computational Theory of Perceptions is a challenging task that requires taking into account knowledge extended to several scientific disciplines such as Fuzzy Logic, Linguistics and Software Engineering among others. In this paper we contribute to the Computational Theory of Perceptions by providing a set of ideas and definitions which could be useful as components of a future methodology for the development of practical applications of this theory.*

Keywords— Computational Theory of Perceptions, Computing with Words, Fuzzy Logic.

1 Introduction

The Computational Theory of Perceptions (CTP), outlined in the Zadeh's seminal paper [1] and further developed in subsequent papers [2], deals with the automated processing of human knowledge expressed in Natural Language (NL). This theory is inspired in the way human beings use NL to make rational decisions and communicate their experience in an environment of imprecision, uncertainty and partial truth.
The general goal of CTP is to develop computational systems with the capacity of computing with the meaning of NL propositions, i.e. with the capacity of computing with imprecise descriptions of the world in the same way that humans do.
When we face the project of design and development of a computational system we must deal with knowledge extended to several disciplines. Typically the designer must solve problems in the fields of Mechanics, Electricity, Electronics and Software Engineering but in this type of project the multidisciplinary scope extends in different degrees to the disciplines grouped around Cognitive Science, namely, Philosophy, Psychology, Artificial Intelligence, Linguistics, Anthropology, Neurophysiology and Education [3].
Unfortunately it is not easy to obtain a practical perspective of CTP in such a wide context. However ignoring this global perspective could lead to misunderstandings and delays in undertaking the development of this new theory.
This paper tries to contribute with a brief analysis of several concepts of CTP from the point of view of its practical application. We present an ordered set of concepts where the description of the essence of CTP is complemented with ideas taken from Systemic Functional Linguistics and Software Engineering among other disciplines.
We analyze five concepts going from the most general to the most practically applicable.
In section 2 we analyze briefly the computational use of NL from both perspectives, namely, as a tool for describing human experience and as a tool for enacting social processes. Usually, the implications of this analysis are not explicitly accounted in the development of the current computational systems based on NL.
In section 3 we analyze the concept of computational perception versus human perception. We use this concept in our search of the "basic unit of meaning" and the different ways of representing it.
In section 4 we propose a practical hierarchy of levels of Precisiation. The concept of Precisiation by Zadeh is interpreted as a process of aggregating constraints to go from the most general meaning to the most specific.
In section 5 we introduce the practical concept of Linguistic Fuzzy Transducer as a computational module with capacity of translating between two different descriptions of the reality.
Section 6 provides our conclusions

2 Computational use of Natural Language

NL has evolved in parallel with the human brain as a product of natural evolution. NL is not the result of the rational design of human beings as is the case of Mathematics or Music Notation. NL is a dynamic system that can not be explained simply as the sum of its parts [4]. Typically, we do not use a set of formal rules to produce our NL utterances. We could say that a human speaker produces his/her discourse following a chaotic procedure similar to the one used by a tree to build branches and leaves.
Taking the perspective of the developer of CTP applications let us briefly analyze the two complementary functions of language: construing experience and enacting social processes.

2.1 Experience

Regarding the first function, in agreement with Systemic Functional Linguistics (SFL), NL is a resource with which human beings construct the mental maps of their phenomenal world, of their experience of process: what goes on out there and what goes on in the realms of their own consciousness. The experience is a resource of meaning, a potential of understanding, representing and acting on the world. The particulars of our daily life make sense because they are instantiations of this potential. The experience includes descriptions of two layers of reality: First order phenomena related directly with the environment and the second order phenomena formed of

ISBN: 978-989-95079-6-8

the meanings and wordings that perceptions of first order phenomena bring into being [5].
During the last thousands of years, human evolution has been supported not only by genetic transmission but essentially by cultural learning [6]. It is well known that, in this process, the invention of the printing press produced a revolutionary push. During recent decades, computers have acquired an important role as a medium of storing and processing information boosting this process, i.e. computers and the Internet are used as the new encyclopedia.
But we believe computers will have a more important role as contributors to mankind evolution. The next challenge consists of using computers as a tool where NL does not have the general meaning provided in an encyclopedia but the particular meaning that a specific user, with a specific personal experience, requires in a specific context of use.
The achievement of this goal will convert computers into personal assistants, into personalized tools to help us to go far in the construction of our personal experience, i.e. in our personal understanding of the world.
From the perspective of computational systems developers, we will use the term "World Model" for a computational representation of reality, a representation of the context where the computer must perform its function. We can see this World Model as a projection of the programmer's experience into the computer memory.

2.2 Human Computer Interface

Regarding the second function of NL, enacting social processes, the computer can be considered as a new type of partner in the NL communication process typically performed between human beings.
When using a computer with this role, we must consider three modes to communicate meaning with NL [5]:
1. The domain of language of the computational application. We should use the concepts and associated words of the culturally recognized repertory of social practices and concerns. Moreover, we should take into account the personal use of NL, i.e. the personal experience of the specific user at whom the application is aimed.
2. The relationship established between user and computer. When using NL we should use the culturally recognized repertory of role relationships and interactive patterns corresponding with a specific speaker - listener relationship. We should choose and design the role the user expects to be adopted by the computer. That is to say, we must assign the computer the role of "secretary", "weather broadcast speaker", or perhaps a new type of role that still should be carefully designed and explained.
It is important to realize that, when humans communicate, they establish a theory about the mind of the other, about the other's experience and intentions, which is absolutely relevant to interpret of the meaning of the NL utterances [6].
3. The mode of expression. We must consider both, the medium (written, spoken) and the rhetorical function (persuasive, didactical, informative, etc.) of the language used. Additionally, we should consider the possibility of complementing the use of NL with graphics and sounds, and the possibility of expressing emotions with the tone of voice and the face expressions of an avatar.

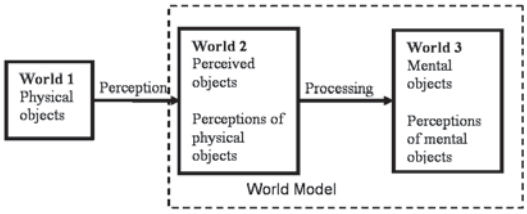


Figure 1: The three worlds by Popper.

3 Computational Perceptions

In agreement with Popper, we can distinguish three worlds: The first is the unknown world of what is going on out there, i.e. the world of the objects in our environment; the second is the world of the perceived objects, i.e. the world of information that we acquire thanks to our senses; and the third is the world of the mental objects that we create in the realm of our consciousness, i.e. they are abstract objects built using NL and that constitute an important part of our Experience (See Fig. 1) [7].
In our approach we can say that we construe our experience using perceptions. Moreover we can say that, remembering or figuring out a mental object is a form of perception. This is because NL is basically a system for describing perceptions [8].
If we define *to perceive* as: "to attain awareness or understanding of elements in the environment" [9] then we can say that *to perceive using a computational system* is: the process performed using a computer to obtain information and to produce a representation useful to the system designer, and therefore for the human operator, to attain awareness or understanding of elements in the environment".
Therefore the programmer and the final user of the computational system are the interpreters of the information provided by the computer. The human uses the computer as a tool to perceive the reality in a comparable way that he/she uses a clock to know about the passage of time or a microscope to detect the presence of bacteria in a drop of water.

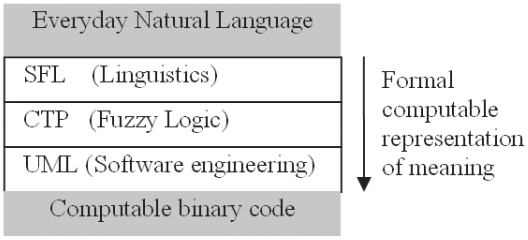


Figure 2: CTP occupies a position between Linguistics and Software Engineering

3.1 Granule

In CTP, granules are used to describe perceptions. The basic unit of meaning is a granule. A granule is a clump of elements which are drawn together by indistinguishability, similarity, proximity or functionality [10].
The perception of an object can be described either only as a granule, or composed by a collection of granules. A granule is a node in a network of relations with other granules. The

ISBN: 978-989-95079-6-8

granularity of the description of objects allows us to deal with real world problems considering different degrees of detail. For example, we summarize the description of an object by hiding the irrelevant granules and remarking on the relevant ones. Indeed, in this sense, the description of an object, i.e. a perception, is always a summary.
Fig. 2 shows three disciplines devoted to NL meaning representation where CTP occupies a relevant position between Linguistics and Software Engineering. Linguistics is needed to face the complexity of a normal conversation, a tale, etc. Unified Modelling Language (UML) is a language developed in Software Engineering during the last decades to develop computational models of reality and it is a type of tool needed when we deal with the development of complex computational applications.
In SFL the basic unit of meaning is a *figure* (figure with the meaning of understanding of the verb figure out). A figure has three types of elements: participants, processes and circumstances [5].
In UML the basic unit of meaning is a *class*. A class has associated attributes and operations [11].
Figures and classes are types of granule and we can use them to computationally describe perceptions.

3.2 Information as constraint

In CTP, granules have fuzzy attributes, i.e. the borders of a granule are fuzzy. Fuzziness of granules, their attributes and their values allow us to model the way in which human concepts are formed, organized and manipulated.
Granules are defined using fuzzy constraints on the possible values of their attributes. In this sense, in CTP information is expressed as a set of constraints [12]. The elasticity of constraints is obtained by using Fuzzy Logic to define them. This elasticity is intended to be a reflection of the same elasticity found in the meaning of words in NL.

3.3 Computational Perception

As mentioned above, a computational perception (perception for short) is a representation of the information obtained about an object with a level of granularity useful for the programmer's purposes.
In CTP there are two different ways of describing a perception: on the one hand using NL and on the other hand using a formalized language which computers can process.
The first description is the natural way of describing the world. NL has evolved with the human being acquiring capacity to understand and communicate complex world descriptions [5], i.e. NL contains all the required resources to describe the objects that we can perceive or imagine, their properties and behaviour.
The second description is a challenge that still lacks of a general solution. It consists of a formalized computable representation with equivalent meaning to the NL propositions. In CTP this language is named Generalized Constraint Language (GCL) [13].
A key element of GCL is the concept of linguistic variable [14]. The values of a linguistic variable are granules. Assigning value to a linguistic variable consists of defining a constraint on the set of its possible values.
In GCL the meaning of a proposition p is expressed as:

$$XisrR$$

where

- *X* is the constrained linguistic variable whose structure of values ranges from the simple linguistic label required to represent a perception associated with the value of a sensor to a complex structure of information, namely, a vector of propositions, a fuzzy graph, a function, etc.
- *R* is the constraining relation
- *isr* is a copula in which *r* is a variable indicating the type of constraint. Primary constraints are formalizations of three basic perceptions: perception of possibility; perception of likelihood; and perception of truth.

4 Precisiation

Granularity allows us to order perceptions by their meaning. At one extreme we have the set of all possible meanings and at the other extreme a simple text with a the meaning of a specific perception, i.e. using constraints we construe a text as an instance of the whole available potential of meaning.
Precisiation consists of aggregating constraints until a description is provided of a perception useful for a specific propose. Precisiation is a form of modelling perceptions by aggregating constraints .
As mentioned above we can see experience as a network of interrelated granules. We describe a perception by defining a sub-network of related granules. Precisiation requires determining a level of granularity, i.e. the composition of this sub-network.
From a practical point of view, one important relation, that we could consider the first in the Precisiation process, is the relation "a type of". This relation organizes the potential of meaning into the hierarchal order of Instantiation - Generalization.
Once determined the position of the perception in this dimension we can continue introducing constraints in other dimensions depending on every specific application such as: "is part of", "causality", "dependency", etc.
In SFL this network of interrelated granules is called a System Network. A System Network is an acyclic directed graph, consisting of elements of meaning partially ordered in hierarchies using relations of meaning [5].

4.1 Levels of Precisiation

Precisiation is a multidimensional process and the ranking is continuous rather than discrete, i.e. there are so many levels of Precisiation as practical uses of NL. From the practical point of view of organizing the design of a computational system based on CTP we have defined the hierarchy of levels represented in Fig. 3.

Domain of Experience/Language
This first level of Precisiation consists of delimiting a Domain of Experience. A Domain of Experience is a closed domain where the process we like to describe takes place. A Domain of Experience defines a specific context, a set of possible situations and corresponds with a Domain of Language. The

ISBN: 978-989-95079-6-8

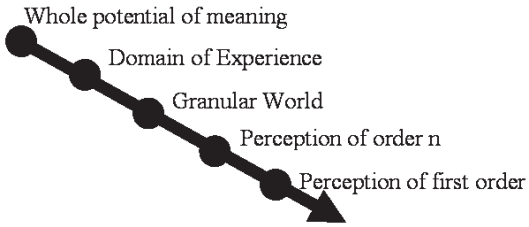


Figure 3: A hierarchy of levels of precisiation

concept of Domain of Language is related with the concept of language-games by Wittgenstein [15].
A Domain of Language contains a subset of possible meanings, each domain has associated a corpus of NL expressions and linguistic patterns. They are at intermediate level in the ordered structure of granularity. For example, the "language of cooking" is an instantiation of the "whole language", the "language of cooking deserts" is a finer level of instantiation, and "the language of cooking apple cakes" is even a more detailed instantiation.

Granular World

A Granular World (GW) contains a subset of meanings belonging to a Domain of Experience. See in [16] for interesting research on the concept of GW.
In our approach this is a quite technical concept taken from Control Theory. We can see GW as an n-dimensional State Space where we must define the needed variables and procedures to create a computational model of the system we want to describe.
A basic element of a GW is the Vector of Linguistic State Variables (SV). Every perception in a GW can be described as a combination of constraints on the possible values of these variables. The application of Fuzzy Logic in Systems Theory was introduced by Zadeh several decades ago [17].
Complex perceptions in a GW are granules built with sets of lower order granules. In a GW the explanation of a perception is a set of more detailed perceptions that describes its meaning with deeper degree of granularity.
The variables belonging to SV provide the most detailed possible explanation of a perception in a GW.

Perception of first order

Perception of first order (p^1) in a computational system is a perception directly associated with the value provided by a sensor at an instant in time.
The basic form of a perception of first order is a proposition that describes the subjective value of a measure, e.g. "The temperature is High" is a subjective description, i. e. for a specific user in a specific use, when the sensor measure reads 45^{o}C.
There are two forms of representing p^1:
The linguistic representation, e.g.:

$$p^1\text{: "The Temperature is High"}$$

And the formal representation:

$$p^1\text{: } X = \mu_{A_i}(x)$$

where:

- X is a linguistic variable (e.g. Temperature).
- A_i is a linguistic term belonging to the set of possible linguistic values of X (e.g. {Low, Medium, High}).
- $\mu_{A_i}(x)$ is the membership function associated with the linguistic term A_i.
- x is a numerical value obtained from the sensor (e.g. 45^{o}C).

We say that the meaning of a perception of first order p^1 can be *explained* using numerical values obtained from sensors and a set of membership functions that covers the domain of these possible values. Because perceptions of the first order work directly upon sensor values, we say these perceptions constitute the most detailed linguistic description of a signal, i.e. it has the finest granularity. Obviously, the linguistic variables associated with first order perceptions are good candidates to belong to the SV of a GW.

Perception of order n

Granularity of perceptions allows us to create a hierarchy describing complex perceptions using sets of more detailed ones.
For example, two first order propositions
p_1^1: "The Temperature is Warm"
and
p_2^1: "The Humidity is Medium"
could be used to explain the meaning of
p^2: "The Room is Comfortable".
Typically this explanation has the form of a rule:
IF p_1^1 AND p_2^1 THEN p^2.
We could extend easily this example by considering other perceptions such as "Acoustic noise" or "Number of persons in the room" to construe the *explanation* of a more complex n-order perception of Comfort.

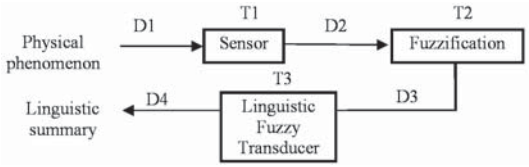


Figure 4: A chain of transducers (T) to translate between several Domains of Language (D).

5 Linguistic Fuzzy Transducer

In our context a Linguistic Fuzzy Transducer is a computational system capable of translating between two different representations of a perception.
Humans are good translators between perceptions described in different domains of language, e.g. when a teacher explains his matter he is translating his perceptions to the domain of language that he expects the students have. A system with the capacity of translating between two natural languages is a transducer. A question answering system could be considered as a transducer with the capacity of providing detailed explanations of a given perception. A system with the capacity of summarize information is a transducer.
Unfortunately to build a computational version of a Linguistic

ISBN: 978-989-95079-6-8

Transducer is still a challenge without a general solution.
Fig. 4 shows a sequence of transducers which convert a perception of the physical environment into representations in different domains of language with different levels of granularity. Indeed, the figure represents a family of computational systems with the capacity of generating a linguistic summary of data provided by sensors. D1 is the domain of language of Physics, D2 is the domain of numerical values provided by sensors, D3 is essentially a domain of first order perceptions obtained after a process of fuzzification and D4 is the domain of language where the final user will use the information.
A Linguistic Fuzzy Transducer has the following basic elements:

- GW1: The Granular World of the input represented by its associated SV1.
- GW2: The Granular World of the Output represented by its SV2.
- Linguistic Fuzzy Model: A model is a simplification of reality constructed using our perceptions. Usually we need to analyze the system at different levels of granularity to understand its functioning. In Fuzzy Logic this type of model is called Linguistic Fuzzy Model that typically can be implemented as a set of fuzzy rules that control the system evolution in time. [18].

The rest of this section describes a simple example of the type of computational system represented in Fig. 4.

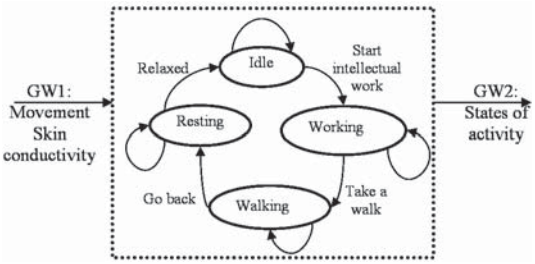


Figure 5: Transducer in the example of application

5.1 Example

This experiment has been performed in the setting of the daily routine of a person dedicated to sedentary work: reading, writing and managing a computer in an office environment. A typical situation that we would like to model is the following:

1. The subject starts his/her day relaxed at the desk.
2. The subject turns on the computer, reads and answers emails, reads papers, writes notes and so on.
3. Initially the subject will remain seated and tension levels rise slowly.
4. After time the tension starts to provoke difficulties in concentration. The subject starts to move in his/her chair and shows behavior such as going to the toilet or moving paperwork around the office.
5. Then the subject decides to take a small walk outside 'to clear his/her mind'.
6. Finally the subject returns to his/her office desk and continues working, again in a less stressful state (similar to the starting point).

Fig. 5 shows a diagram that represents the evolution of the activity and state of the subject during the experiment. Note that we consider the subject able to recover part of the initial 'relaxed state' by doing some physical exercise.
The goal consists of monitoring the physical activity, the levels of stress and mental tiredness of the subject and to create a linguistic report summarizing the temporal evolution of these parameters. This experiment provides the first steps into the development of a computational system with the objective to help the user to increase his/her effectiveness and satisfaction [19].
Let us see briefly how to obtain, in this case, the main components of the last transducer (T3) in Fig. 4:
GW1: The subject were wearing two sensors: A skin conductivity meter fixed attached to the left wrist, with electrodes attached to the index finger and the middle finger. The accelerometer was kept in the chest pocket of the subject's shirt. The subject was instructed to follow the steps of the experiment as described above.
After the fuzzification process we have two fuzzy linguistic variables which constitute the vector: SV1=(Activity, Skin Conductivity).
GW2: is represented by the vector SV2=(Idle, Working, Walking, Relaxing).
LFM: We have built this LFM using a Fuzzy Finite Machine (FFSM) (see [19] for more details) where the states were labeled as: q_1: Idle, q_2: Working, q_3: Walking and q_4: Relaxing. The evolution of this FFSM is controlled by the following set of fuzzy rules:
R_{11}: IF (Q IS q_1) AND (Activity IS Low) AND (Conductivity IS Low) THEN Q IS q_1
R_{12}: IF (Q IS q_1) AND (Activity IS Low) AND (Conductivity IS Medium) THEN Q IS q_2
R_{22}: IF (Q IS q_2) AND (Activity IS Medium) AND (Conductivity IS Medium) THEN Q IS q_2
R_{23}: IF (Q IS q_2) AND (Activity IS Medium) AND (Conductivity IS high) THEN Q IS q_3
R_{33}: IF (Q IS q_3) AND (Activity IS High) THEN Q IS q_3
R_{34}: IF (Q IS q_3) AND (Activity IS Low) THEN Q IS q_4
R_{44}: IF (Q IS q_4) AND (Activity IS Medium) AND (Conductivity IS High) THEN Q IS q_4
R_{41}: IF (Q IS q_4) AND (Activity IS Low) AND (Conductivity IS Medium) THEN Q IS q_1
Where R_{ii} are rules to remain in the state and R_{ij} are rules to change.
The output is the template of a simple report as in the following example:
"The subject started **RELAXED** at the desk at **0 minutes**. Then the subject was working during **15 minutes** without signs of tiredness. Around **25 minutes** the subject was moving and appears uncomfortable. At **60 minutes** the subject decided to take a walk outside and went back at **75 minutes**. Around **80 minutes** the subject was relaxing until to be **RELAXED**. About **90 minutes** the subject started to work again."

ISBN: 978-989-95079-6-8

6 Conclusions

The authors are involved in a long term project aimed to contribute to Zadeh's CTP.
The main contribution of this paper consist in providing an ordered set of definitions obtained from different fields that could help the software engineer to understand and organize the project of a computational system based on CTP. These definitions can be used as the base of a future methodology for developing this type of projects.

Acknowledgment

This work has been partially funded by the Foundation for the Advancement of Soft Computing (Asturias, Spain) and Spanish government (CICYT) under grant: TIN2008-06890-C02-01

References

[1] Lotfi A. Zadeh. From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems*, 1:1–10, 1999.

[2] Lotfi A. Zadeh. Toward human level machine intelligence - is it achievable? the need for a paradigm shift. *IEEE Computational Intelligence Magazine*, 1:1–10, 2008.

[3] J.C. Gardner. *The Minds New Science. A History of the Cognitive Revolution*. Basic Books, New York, 1985.

[4] M.A.K. Halliday and M.I.M Matthiessen. *An Introduction to Functional Grammar*. Oxford University, New York, 2004.

[5] M.A.K. Halliday and M.I.M Matthiessen. *Construing experience through meaning*. Continuum, New York, 1999.

[6] M. Tomasello. *The cultural origins of human cognition*. Harvard, New York, 1999.

[7] K.R. Popper and J.C. Eccles. *The Self and Its Brain*. Springer-Verlag, Berlin, 1977.

[8] Lotfi A. Zadeh. Precisiated natural language (pnl). *AI Magazine*, 25(3):74–91, October 2004.

[9] http://www.merriam webster.com. 2008.

[10] Lotfi A. Zadeh. *Fuzzy sets and information granularity*. In: Gupta M. Ragade R, Yager R (eds) Advances in Fuzzy Set Theory and Applications. North-Holland Publishing CO, 1979.

[11] J. Rumbaugh G. Booch and I. Jacobson. *The Unified Modelling Language user guide*. Addison Wesley, 1999.

[12] Lotfi A. Zadeh. Generalized theory of uncertainty (gtu) - principal concepts and ideas. *Computational Statistics & Data Analysis*, 51:15–46, 2006.

[13] Lotfi A. Zadeh. Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, 4:103–110, 1996.

[14] Lotfi A. Zadeh. The concept of linguistic variable and its application to approximate reasoning. *Information Sciences*, 1975.

[15] L. Wittgenstein. *Philosophical investigations*. Blackwell Publishing, 1953/2001.

[16] A. Bargiela and W.G. Pedrycz. *Granular Computing. An Introduction.* Kluwer Academic Publishers, 2003.

[17] Lotfi A. Zadeh. *Toward a theory of fuzzy systems*. In: Klir G.J., Yuan B. (eds) *Fuzzy sets, Fuzzy logic, and Fuzzy systems: selected papers by Lotfi A. Zadeh*. 1996 - World Scientific Publishing Co., Inc. River Edge, NJ, USA, 1971.

[18] R. R. Yager and D. P. Filev. *Essentials of Fuzzy Modelling and Control*. John Wiley & Sons, Inc, 1994.

[19] G. Trivino and A. van der Heide. Linguistic summarization of the human activity using skin conductivity and accelerometers. *Proceedings of the Conference Information Processing and Management of Uncertainty in Knowledge Based Systems. (IPMU2008), June 22-27, Malaga, Spain*, 2008.

ISBN: 978-989-95079-6-8

A characterization of residual implications derived from uninorms

Isabel Aguiló Jaume Suñer Joan Torrens

Departament de Ciències Matemàtiques i Informàtica
Universitat de les Illes Balears
07122 Palma de Mallorca, Spain
Email: {isabel.aguilo,jaume.sunyer,dmijts0}@uib.es

Abstract*— *In this paper, a set of axioms is given that characterizes those functions $I : [0,1]^2 \to [0,1]$ for which a left-continuous uninorm U exists in such a way that I is the residual implication derived from U. A characterization for the particular case when U is representable is also given.

***Keywords*—** Implication function, left-continuity, residual implication, uninorm.

1 Introduction

Implication functions are probably the most important operations in fuzzy logic, approximate reasoning and fuzzy control, because they are used not only to model fuzzy conditionals, but also to make inferences in any fuzzy rule based system (see for instance [1] or [2]). Moreover, they are useful not only in approximate reasoning and fuzzy control, but also in many other fields like fuzzy relational equations and fuzzy mathematical morphology ([3]), fuzzy DI-subsethood measures and image processing ([4] and [5]), and data mining ([6]). An excellent and very recent book on fuzzy implications is [7].

Among other models (see [3]), residual implications and strong implications are the most usual ones. They are usually derived from t-norms and t-conorms respectively as follows:

- Given a t-norm T, the residual implication or R-implication derived from T is given by

$$I_T(x,y) = \sup\{z \in [0,1] \mid T(x,z) \leq y\} \qquad (1)$$

  for all $x, y \in [0,1]$.

  R-implications come from residuated lattices based on the so-called residuation property that can be written as

$$T(x,y) \leq z \qquad \Longleftrightarrow \qquad I_T(x,z) \geq y.$$

  Note that this property is satisfied if and only if T is left-continuous and then the supremum in (1) can be substituted by maximum.

- Given a t-conorm S and a strong negation N, the strong implication or (S,N)-implication derived from S, N is given by

$$I_{S,N}(x,y) = S(N(x),y) \qquad \text{for all} \quad x, y \in [0,1].$$

  In this case, (S,N)-implications appear as a generalization of the classical boolean implication $p \to q \equiv \neg p \vee q$.

Axiomatic characterizations for R-implications derived from left-continuous t-norms, as well as for (S,N)-implications (even when N is only a continuous fuzzy negation, not necessarily strong), have appeared along the time (see for instance [8], [9], [10], [11], [12], [13], and the recent surveys [3] and [14]).

On the other hand, conjunctive and disjunctive uninorms (see [15]) are a class of associative binary aggregation functions that generalize both t-norms and t-conorms, and that has been successfully used in many application fields where t-norms and t-conorms apply. In particular, R and (S,N)-implications have also been derived from uninorms, obtaining in this way new implication functions with nice and interesting properties (see [16], [17] and [18]). However, an axiomatic characterization for these new classes of implications has been done for the case of (S,N)-implications derived from disjunctive uninorms in [19] (again in the general case when N is not necessarily strong), but not for R-implications.

In this paper, we deal with this problem and we give characterizations for the class of R-implications derived from left-continuous conjunctive uninorms in a similar way as it is done for left-continuous t-norms. Moreover, a characterization for the special case of left-continuous representable uninorms is given. From this case we derive a characterization of a kind of quasi-continuous implications (continuous except at two points) that can be viewed as a generalization of the well known Smets-Magrez Theorem (see the Preliminaries).

2 Preliminaries

We will suppose the reader to be familiar with the theory of t-norms, t-conorms and fuzzy negations (all necessary results and notations can be found in [20]). We recall here only some facts on implications.

Definition 1 *A binary operator $I : [0,1] \times [0,1] \to [0,1]$ is said to be an* implication function*, or an* implication*, if it satisfies:*

(I1) $I(x,z) \geq I(y,z)$ *when* $x \leq y$*, for all* $z \in [0,1]$.

(I2) $I(x,y) \leq I(x,z)$ *when* $y \leq z$*, for all* $x \in [0,1]$.

(I3) $I(0,0) = I(1,1) = 1$ *and* $I(1,0) = 0$.

Note that, from the definition, it follows that $I(0,x) = 1$ and $I(x,1) = 1$ for all $x \in [0,1]$ whereas the symmetrical values $I(x,0)$ and $I(1,x)$ are not derived from the definition.

Special interesting properties for implication functions are:

- The *ordering property*,

$$x \leq y \Longleftrightarrow I(x,y) = 1, \quad \text{for all} \quad x, y \in [0,1]. \qquad (\mathbf{OP})$$

ISBN: 978-989-95079-6-8

- The *exchange principle*,

$$I(x, I(y, z)) = I(y, I(x, z)), \quad \text{for all} \quad x, y, z \in [0, 1]. \tag{EP}$$

With respect to R-implications derived from left-continuous t-norms (see for instance [9] and the survey [14] with the references therein), they satisfy both properties (OP) and (EP). Moreover, we have the following characterization.

Theorem 1 *Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is an R-implication derived from a left-continuous t-norm, if and only if, I satisfies $(I2), (OP), (EP)$ and I is right-continuous with respect to the second variable.*

With the assumption of continuity we have a characterization of the following subclass of R-implications, that are also (S, N)-implications, known as the Smets-Magrez Theorem, see [21] and see also [12] for the current version.

Theorem 2 *Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is a continuous function satisfying $(OP), (EP)$, if and only if, I is conjugate with the Łukasiewicz implication, that is, there exists an increasing bijection $\varphi : [0,1] \to [0,1]$ such that*

$$I(x, y) = \varphi^{-1}(\min(1, 1 - \varphi(x) + \varphi(y)))$$

for all $x, y \in [0,1]$.

We also suppose that some basic facts on uninorms are known as well as the class of representable ones (see for instance [15]). Let us only recall here the definition.

Definition 2 *A* uninorm *is a two-place function $U : [0,1]^2 \longrightarrow [0,1]$ which is associative, commutative, increasing in each place and such that there exists some element $e \in [0,1]$, called* neutral element*, such that $U(e, x) = x$ for all $x \in [0,1]$.*

It is clear that uninorms generalize both t-norms and t-conorms, since they are retrieved from uninorms just taking $e = 1$ and $e = 0$, respectively. Moreover, for a uninorm U, it is always $U(1,0) \in \{0,1\}$ and U is said to be *conjunctive* when $U(1,0) = 0$, and *disjunctive* when $U(1,0) = 1$.

On the other hand, residual implications derived from uninorms have been also studied (see [16]).

Definition 3 *Let U be a uninorm. The* residual operation *derived from U is the binary operation given by*

$$I_U(x, y) = \sup\{z \in [0,1] \mid U(x, z) \leq y\} \tag{2}$$

for all $x, y \in [0,1]$.

Proposition 1 *([16]) Let U be a uninorm and I_U its residual operation. Then I_U is an R-implication if and only if the following condition holds*

$$U(x, 0) = 0 \quad \textit{for all} \quad x < 1. \tag{3}$$

This includes all conjunctive uninorms but also many disjunctive ones, for instance in the classes of representable and idempotent uninorms (see [16] and [17]). However, when we deal with left-continuous uninorms U we clearly have that U satisfies condition (3) if and only if it is conjunctive.

3 R-implications derived from left-continuous uninorms

There are some properties of R-implications derived from uninorms, that can be deduced directly from the definition, or can be proved in the same way as for those derived from t-norms (see [16]).

Proposition 2 *Let U be a uninorm with neutral element e satisfying condition(3) and I_U its residual implication. Then*

i) *$I_U(e, y) = y$ for all $y \in [0,1]$ counterpart for uninorms of the* neutrality principle*, that will be denoted by* $(\mathbf{NP_U})$.

ii) *$I_U(x, y) \geq e$ for all $x, y \in [0,1]$ such that $x \leq y$.*

iii) *$y \leq I_U(x, U(x, y))$ for all $x, y \in [0,1]$.*

In this section we will deal specially with R-implications derived from left-continuous uninorms. Recall that in this case condition (3) is equivalent to the uninorm U be conjunctive and so, we will refer only to left-continuous conjunctive uninorms. As for t-norms, this case of left-continuity is specially important because then (and only then) I_U satisfies the residuation property. The proof of this fact is similar to the case of t-norms (see [2]) and thus we do not include it.

Proposition 3 *Let U be a conjunctive uninorm and I_U its residual implication. Then U is left-continuous if and only if I_U satisfies the residuation property:*

$$U(x, y) \leq z \iff I_U(x, z) \geq y \tag{RP}$$

for all $x, y, z \in [0,1]$.

Moreover, when U is left-continuous additional properties are satisfied.

Proposition 4 *Let U be a left-continuous conjunctive uninorm with neutral element $e \in (0,1]$. Then I_U satisfies*

i) *Counterpart for uninorms of the ordering property:*

$$x \leq y \iff I_U(x, y) \geq e \quad \textit{for all } x, y \in [0,1]. \tag{OP$_\mathbf{U}$}$$

ii) *Exchange principle,*

$$I_U(x, I_U(y, z)) = I_U(y, I_U(x, z)) \tag{EP}$$

for all $x, y, z \in [0,1]$.

iii) *The* modus ponens property*:*

$$U(x, I_U(x, y)) \leq y \tag{MP}$$

for all $x, y \in [0,1]$.

iv) *$I_U(x, -)$ is right-continuous for all $x \in [0,1]$.*

Proof: Part $i)$ is straightforward from the definition, part $iv)$ follows as for the case of t-norms and the other properties were already proved in [16]. ∎

ISBN: 978-989-95079-6-8

In order to characterize all functions $I : [0,1]^2 \to [0,1]$ that are R-implications derived from left-continuous uninorms, we firstly study those functions with properties (OP_U) and (EP).

Proposition 5 *Let $I : [0,1]^2 \to [0,1]$ be a function satisfying properties (OP_U) and (EP). Then*

i) $I(x,x) \geq e$ for all $x \in [0,1]$.

ii) I satisfies property (I1): $I(x,z) \geq I(y,z)$ when $x \leq y$, for all $z \in [0,1]$.

iii) I satisfies (NP_U). In particular, $I(e,e) = e$.

iv) $I(0,y) = 1$ for all $y \in [0,1]$.

v) If $I(1,e) = 0$ then $N(x) = I(x,e)$ is a fuzzy negation with $N(e) = e$.

Proof:

i) It is obvious from condition (OP_U).

ii) Let us consider $x, y, z \in [0,1]$ with $x \leq y$. Then we have

$$I(y, I(I(y,z),z)) = I(I(y,z), I(y,z)) \geq e$$

$$\Longrightarrow y \leq I(I(y,z),z)$$

by i) and conditions (OP_U) and (EP). Then,

$$x \leq y \leq I(I(y,z),z)$$

and this implies similarly that

$$e \leq I(x, I(I(y,z),z) = I(I(y,z), I(x,z))$$

and thus $I(y,z) \leq I(x,z)$.

iii) First of all,

$$I(x, I(e,x)) = I(e, I(x,x)) \geq e$$

where the last inequality holds because $e \leq I(x,x)$ by i). But the equation above ensures that $x \leq I(e,x)$. To prove the other inequality, it is sufficient to apply i) and conditions (OP_U), (EP) to obtain

$$I(e, I(I(e,x),x)) = I(I(e,x), I(e,x)) \geq e$$

$$\Longrightarrow I(I(e,x),x) \geq e$$

and thus $I(e,x) \leq x$.

iv) We have for all $y \in [0,1]$

$$\begin{array}{rcl} 0 \leq I(1,y) & \Longrightarrow & I(0, I(1,y)) \geq e \\ & \Longrightarrow & I(1, I(0,y)) \geq e \\ & \Longrightarrow & 1 \leq I(0,y) \end{array}$$

and thus $I(0,y) = 1 \; \forall y \in [0,1]$.

v) From ii) we have that N is decreasing and iv) implies that $N(0) = I(0,e) = 1$. Thus, since $N(1) = I(1,e) = 0$, N is a fuzzy negation. Finally, by iii) $N(e) = e$. ∎

Now, we can already give the axiomatic characterization of residual implications derived from left-continuous uninorms. Recall that in this case, since the uninorm must satisfy $U(x,0) = 0$ for all $x < 1$, it must be conjunctive.

Theorem 3 *Let $I : [0,1]^2 \to [0,1]$ be a function. The following statements are equivalent:*

i) *I is an R-implication derived from a left-continuous uninorm U with neutral element $e \in (0,1]$.*

ii) *I satisfies $I2)$, (OP_U), (EP) and $I(x,-)$ is right-continuous for all $x \in [0,1]$.*

Moreover, in this case the uninorm U must be conjunctive and it is given by:

$$U(x,y) = \inf\{z \in [0,1] \mid I(x,z) \geq y\}.$$

Proof: Suppose first that

$$I(x,y) = I_U(x,y) = \sup\{z \in [0,1] \mid U(x,z) \leq y\}$$

for all $x, y \in [0,1]$ where U is a left-continuous conjunctive uninorm with neutral element $e \in (0,1]$. Then we already know that I_U is an implication and consequently satisfies $(I2)$. The other properties follow from Proposition 4 and Proposition 2-iv).

Suppose now that I satisfies the required conditions. We have to prove that the function $U : [0,1]^2 \to [0,1]$ defined by

$$U(x,y) = \inf\{z \in [0,1] \mid I(x,z) \geq y\}$$

for all $x, y \in [0,1]$, is a left-continuous conjunctive uninorm U with neutral element e and also that $I = I_U$.

- $U(x,e) = x \; \forall x \in [0,1]$:
  We have by condition (OP_U) that

$$\begin{array}{rcl} U(x,e) & = & \inf\{z \in [0,1] \mid I(x,z) \geq e\} \\ & = & \inf\{z \in [0,1] \mid z \geq x\} = x. \end{array}$$

- $U(x,y) = U(y,x) \; \forall x, y \in [0,1]$:
  Observe that

$$\begin{array}{rcl} I(y,z) \geq x & \Longleftrightarrow & I(x, I(y,z)) \geq e \\ & \Longleftrightarrow & I(y, I(x,z)) \geq e \\ & \Longleftrightarrow & I(x,z) \geq y \end{array}$$

  by conditions (OP_U) and (EP). Thus

$$\begin{array}{rcl} U(y,x) & = & \inf\{z \in [0,1] \mid I(y,z) \geq x\} \\ & = & \inf\{z \in [0,1] \mid I(x,z) \geq y\} \\ & = & U(x,y) \end{array}$$

- $x_1 \leq x_2 \Longrightarrow U(x_1,y) \leq U(x_2,y)$:
  Let $x_1 \leq x_2$. Now if $I(x_2,z) \geq y$, then $I(x_1,z) \geq I(x_2,z) \geq y$ by property ii) of Proposition 5, and thus $U(x_1,y) \leq U(x_2,y)$.

- $y_1 \leq y_2 \Longrightarrow U(x,y_1) \leq U(x,y_2)$:
  It follows immediately by commutativity and the previous step.

ISBN: 978-989-95079-6-8

- $U(U(x,y),z) = U(x,U(y,z))\ \forall x,y,z \in [0,1]$:

  We have that $U(U(x,y),z) = U(z,U(x,y))$ by commutativity, and thus $U(U(x,y),z)$ is given by

  $$U(U(x,y),z) = \inf\{t \in [0,1] \mid I(z,t) \geq U(x,y)\}.$$

  On the other hand we have that

  $$U(x,U(y,z)) = \inf\{t \in [0,1] \mid I(x,t) \geq U(y,z)\},$$

  and consequently it is sufficient to prove that for all $t \in [0,1]$,

  $$I(z,t) \geq U(x,y) \Longleftrightarrow I(x,t) \geq U(y,z).$$

  Let us prove the left-to-right implication. Since $I(x,-)$ is right-continuous, we have from the expression of $U(x,y)$ that $I(x,U(x,y)) \geq y$. Then

  $$I(z,t) \geq U(x,y) \Longrightarrow I(x,I(z,t)) \geq I(x,U(x,y)) \geq y.$$

  Now, the increasingness with respect to the second component gives that:

  $$\begin{array}{rcl} U(y,z) = U(z,y) & \leq & U(z,I(x,I(z,t))) \\ & = & U(z,I(z,I(x,t))) \\ & \leq & I(x,t) \end{array}$$

  where the last inequality follows from the definition of U. The right-to-left implication follows similarly.

- U is conjunctive and left-continuous:

  Obviously since

  $$U(1,0) = \inf\{z \in [0,1] \mid I(1,z) \geq 0\} = 0,$$

  and left-continuity also follows easily from the definition of U.

- $I(x,y) = I_U(x,y) = \sup\{z \in [0,1] \mid U(x,z) \leq y\}$:

  First of all, observe that the residuation property for I_U allows to write

  $I_U(z, I_U(x,y)) =$

  $$\begin{array}{rcl} & = & \sup\{t \in [0,1] \mid U(z,t) \leq I_U(x,y)\} \\ & = & \sup\{t \in [0,1] \mid U(x,U(z,t)) \leq y\} \\ & = & \sup\{t \in [0,1] \mid U(U(x,z),t) \leq y\} \\ & = & I_U(U(x,z),y) \end{array}$$

  Now, for all $z \in [0,1]$ we have

  $$U(x,z) \leq y \Longleftrightarrow I_U(U(x,z),y) = I_U(z,I_U(x,y)) \geq e$$

  which implies $I_U(x,y) \geq z$. Thus, taking $z = I(x,y)$ we know, by definition of U, that $U(x,I(x,y)) \leq y$ and consequently we will have $I_U(x,y) \geq I(x,y)$.

  On the other hand, right-continuity of $I(x,-)$ implies that $I(x,U(x,z)) \geq z$. Now if we take $z = I_U(x,y)$,

  $$I_U(x,y) \leq I(x,U(x,I_U(x,y))) \leq I(x,y)$$

  again by the residuation property of I_U and the increasingness of I in the second variable. ∎

Remark 1 *Note that when $e = 1$, condition (OP_U) becomes the ordering property (OP) and so, the axiomatic characterization of R-implications derived from left-continuous t-norms (Theorem 1) follows as a particular case.*

The mutual-independence of the properties in the theorem above is an open problem as it already is for the case of t-norms in Theorem 1 (see for instance [7], Remark 2.5.18). However, there are some of these properties that are independent to each other. The same examples for the case of t-norms work here as follows.

Example 1 *i) The Rescher implication given by*

$$I(x,y) = \begin{cases} 1 & if\ x \leq y \\ 0 & if\ x > y \end{cases}$$

satisfies (I2), (OP_U) with respect to any $e \in (0,1]$, and $I(x,-)$ is right-continuous for all $x \in [0,1]$, but it does not satisfy (EP).

ii) The Kleene-Dienes implication given by $I(x,y) = \max(1-x,y)$ satisfies (I2), (EP) and $I(x,-)$ is right-continuous for all $x \in [0,1]$, but there is no $e \in (0,1)$ for which I satisfies (OP_U).

On the other hand, the previous theorem gives the axiomatic characterization of R-implications derived from left-continuous conjunctive uninorms in general. Some particular characterizations can also be given for special classes of left-continuous conjunctive uninorms. Let us study the case of representable uninorms in the next section.

3.1 Residual implications derived from representable uninorms

Recall that a uninorm U with neutral element $e \in (0,1)$ is said to be representable (see [15]) if there exists a strictly increasing and continuous function $h : [0,1] \to [-\infty,+\infty]$ with $h(0) = -\infty$, $h(e) = 0$ and $h(1) = +\infty$ such that U is given by

$$U(x,y) = \begin{cases} h^{-1}(h(x)+h(y)) & \text{if } (x,y) \notin \{(0,1),(1,0)\} \\ 0 \text{ (or } 1) & \text{otherwise.} \end{cases}$$

Function h is called an additive generator of U. Thus it is clear that there are two different representable uninorms with the same generator h, a conjunctive one (which is then left-continuous) and a disjunctive one (which is right-continuous). Representable uninorms are clearly continuous in $[0,1]^2 \setminus \{(1,0),(0,1)\}$ and this is in fact a characterization of such class of uninorms.

Proposition 6 *([22]) A uninorm U with neutral element $e \in (0,1)$ is representable if and only if it is continuous in $[0,1]^2 \setminus \{(1,0),(0,1)\}$*

Note that in both cases, conjunctive and disjunctive, it is $U(0,y) = 0$ for all $y < 1$ and so the residual implication from U can be derived. However, when we consider left-continuous representable uninorms they will be necessarily conjunctive as in the general case.

ISBN: 978-989-95079-6-8

Proposition 7 *([16]) Let U be a representable uninorm with additive generator h. Then its residual implication I_U is given by*

$$I_U(x,y) = \begin{cases} h^{-1}(h(y)-h(x)) & \text{if } (x,y) \notin \{(0,0),(1,1)\} \\ 1 & \text{otherwise.} \end{cases}$$

Let us now prove the characterization theorem for residual implications derived from this kind of uninorms.

Theorem 4 *Let $I : [0,1]^2 \to [0,1]$ be a function. The following statements are equivalent:*

i) *I is an R-implication derived from a left-continuous representable uninorm with neutral element $e \in (0,1)$.*

ii) *I satisfies (OP_U), (EP), $I(1,e) = 0$ and I is continuous except at points $(0,0)$ and $(1,1)$.*

Proof: Suppose first that I is an R-implication derived from a left-continuous representable uninorm with neutral element $e \in (0,1)$. In this case U is conjunctive and by Theorem 3, I satisfies (OP_U) and (EP). Moreover, since U is representable, I must be continuous except at points $(0,0)$ and $(1,1)$ from Proposition 7.

Now suppose that the required conditions for I are satisfied. First consider $N(x) = I(x,e)$ and let us prove that N is a strong negation with fixed point e. Since $I(1,e) = 0$, we already know from Proposition 5 that N is a fuzzy negation with $N(e) = e$. Moreover, since $e \neq 0, 1$, $N(x) = I(x,e)$ is continuous.

On the other hand, again from Proposition 5 we have for all $x \in [0,1]$

$$I(x, I(I(x,e),e)) = I(I(x,e), I(x,e)) \geq e$$

and consequently

$$x \leq I(I(x,e),e) = N(N(x)) = N^2(x) \quad \text{for all} \quad x \in [0,1].$$

Thus, by decreasingness of N we have $N(x) \geq N(N(N(x))) = N^3(x)$. On the other hand,

$$I(I(x,e), I(I(I(x,e),e),e)) =$$

$$= I(I(I(x,e),e), I(I(x,e),e)) \geq e$$

and, thus $N(x) = I(x,e) \leq I(I(I(x,e),e),e) = N^3(x)$. From both inequalities we deduce $N(x) = N^3(x)$ for all $x \in [0,1]$. Now, since $N(x) = I(x,e)$ is continuous in [0,1] by hypothesis, given $x \in [0,1]$, there exists $y \in [0,1]$ such that $N(y) = x$. Then $N(N(x)) = N(N(N(y))) = N(y) = x$, that is, N is involutive and so is a strong negation.

Now, define

$$U(x,y) = N(I(x,N(y))) \quad \text{for all} \quad x,y \in [0,1]$$

and let us prove that U is a uninorm with neutral element e.

- Commutativity comes from condition (EP),

$$\begin{aligned} U(y,x) &= N(I(y,N(x))) \\ &= I(I(y,N(x)),e) \\ &= I(I(y,I(x,e)),e) \\ &= I(I(x,I(y,e)),e) \\ &= N(I(x,N(y))) \\ &= U(x,y) \end{aligned}$$

- From condition (EP) we have

$$\begin{aligned} I(N(y),N(x)) &= I(I(y,e),I(x,e)) \\ &= I(x,I(I(y,e),e)) = I(x,y) \end{aligned}$$

since $I(I(y,e),e) = N^2(y) = y$.

Now, associativity can be proved as folows

$$\begin{aligned} U(x,U(y,z)) &= N(I(x,N(U(y,z)))) \\ &= N(I(x,N(N(I(y,N(z)))))) \\ &= N(I(x,I(y,N(z)))) \\ &= N(I(y,I(x,N(z)))) \\ &= N(I(y,I(z,N(x)))) \\ &= N(I(z,I(y,N(x)))) \\ &= N(I(N(I(y,N(x))),N(z))) \\ &= N(I(N(I(x,N(y))),N(z))) \\ &= N(I(U(x,y),N(z))) \\ &= U(U(x,y),z) \end{aligned}$$

- The increasingness of U comes from Proposition 5 and the decreasingness of N. Given $x_1 \leq x_2$ and $y \in [0,1]$, we have

$$I(x_1,N(y)) \geq I(x_2,N(y)).$$

Thus,

$$N(I(x_1,N(y))) \leq N(I(x_2,N(y)))$$

$$\Longrightarrow U(x_1,y) \leq U(x_2,y)$$

- Finally,

$$\begin{aligned} U(x,e) &= N(I(x,N(e)) = N(I(x,e)) \\ &= N^2(x) = x. \end{aligned}$$

Now, since I is continuous except at points $(0,0)$ and $(1,1)$, U will be continuous except at points $(0,1)$ and $(1,0)$. Moreover, since $I(1,1) \geq e$ we have

$$U(1,0) = N(I(1,1)) \leq e$$

and consequently $U(1,0) = 0$. This proves that U is a conjunctive representable uninorm (in particular it is left-continuous).

Finally, to see that $I = I_U$, it is enough to prove the residuation property:

$$U(x,y) \leq z \Longleftrightarrow I(x,z) \geq y$$

since then $I(x,z)$ will be given by the supremum of $y \in [0,1]$ such that $U(x,y) \leq z$. We have

$$\begin{aligned} U(x,y) \leq z &\iff N(I(x,N(y))) \leq z \\ &\iff I(x,N(y)) \geq N(z) \\ &\iff I(N(z),I(x,N(y))) \geq e \\ &\iff I(x,I(N(z),N(y))) \geq e \\ &\iff I(x,I(y,z)) \geq e \\ &\iff I(y,I(x,z)) \geq e \\ &\iff I(x,z) \geq y \end{aligned}$$

■

ISBN: 978-989-95079-6-8

With respect to the mutual independence of the properties in the theorem above, we already know that (OP_U), and (EP) can not be derived from the others (see the counterexamples below). Now, we are currently working in the independence of the other two conditions.

Example 2 *Consider* $e \in (0,1)$ *and let* $h : [0,1] \to [-\infty, +\infty]$ *be a strictly increasing continuous function with* $h(0) = -\infty, h(e) = 0$ *and* $h(1) = +\infty$. *Then*

- *The function*

$$I(x,y) = \begin{cases} e & if\ x \leq y \\ h^{-1}(h(y) - h(x)) & if\ x > y \end{cases}$$

satisfies all conditions in Theorem 4 except (EP).

- *The function*

$$I(x,y) = \begin{cases} 0 & if\ x = y = 0 \\ & or\ x = y = 1 \\ h^{-1}(h(y) - h(x)) & otherwise \end{cases}$$

satisfies all conditions in Theorem 4 except (OP_U).

Corollary 1 *Let* $I : [0,1]^2 \to [0,1]$ *be a function and* $e \in (0,1)$. *Then there exists a strictly increasing continuous function* $h : [0,1] \to [-\infty, +\infty]$ *with* $h(0) = -\infty, h(e) = 0$ *and* $h(1) = +\infty$ *such that*

$$I(x,y) = \begin{cases} h^{-1}(h(y) - h(x)) & \text{if } (x,y) \notin \{(0,0),(1,1)\} \\ 1 & \text{otherwise.} \end{cases}$$

if and only if I *satisfies* (OP_U), (EP), $I(1,e) = 0$ *and it is continuous except at points* $(0,0)$ *and* $(1,1)$.

This corollary can be viewed as a generalization of the well known Smets-Magrez Theorem (Theorem 2) since it characterizes a kind of implications that are continuous except at two points.

Finally, we want to claim that other characterization theorems can be given for residual implications derived from other classes of left-continuous conjunctive uninorms, and we are currently working in this direction. For instance, from left-continuous conjunctive idempotent uninorms or from left-continuous conjunctive uninorms continuous at the open unit square.

Acknowledgment

This paper has been partially supported by the Spanish grants MTM2006-05540 and MTM2006-08322 (with FEDER support), and the Government of the Balearic Islands grant PCTIB-2005GC1-07.

References

[1] E.E. Kerre, C. Huang, and D. Ruan. *Fuzzy Set Theory and Approximate Reasoning*. Wu Han University Press, Wu Chang, 2004.

[2] S. Gottwald. *A Treatise on Many-Valued Logic*. Research Studies Press, Baldock, 2001.

[3] M. Mas, M. Monserrat, J. Torrens, and E. Trillas. A survey on fuzzy implication functions. *IEEE Transactions on Fuzzy Systems*, 15(6):1107–1121, 2007.

[4] H. Bustince, V. Mohedano, E. Barrenechea, and M. Pagola. Definition and construction of fuzzy DI-subsethood measures. *Information Sciences*, 176:3190–3231, 2006.

[5] H. Bustince, M. Pagola, and E. Barrenechea. Construction of fuzzy indices from fuzzy DI-subsethood measures: application to the global comparison of images. *Information Sciences*, 177:906–929, 2007.

[6] P. Yan and G. Chen. Discovering a cover set of ARsi with hierarchy from quantitative databases. *Information Sciences*, 173:319–336, 2005.

[7] M. Baczyński and B. Jayaram. *Fuzzy Implications*. Springer, Berlin Heidelberg, 2008.

[8] C. Alsina et al. On the characterization of S and R implications. *Proceedings of* VI^{th} *IFSA World Congress*, pages 317–319, 1995.

[9] J.C. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers, Dordrecht, 1994.

[10] J. C. Fodor. Contrapositive symmetry on fuzzy implications. *Fuzzy Sets and Systems*, 69:141–156, 1995.

[11] H. Bustince, P. Burillo, and F. Soria. Automorphisms, negation and implication operators. *Fuzzy Sets and Systems*, 134:209–229, 2003.

[12] M. Baczyński. Residual implications revisited. notes on the smets-magrez theorem. *Fuzzy Sets and Systems*, 145:267–277, 2004.

[13] M. Baczyński and B. Jayaram. On the characterization of (S,N)-implications. *Fuzzy Sets and Systems*, 158:1713–1727, 2007.

[14] M. Baczyński and B. Jayaram. (S,N)- and R-implications: A state-of-the-art survey. *Fuzzy Sets and Systems*, 159:1836–1859, 2008.

[15] J. C. Fodor, R. R. Yager, and A. Rybalov. Structure of uninorms. *Int. J. of Uncertainty, Fuzziness and Knowledge-based Systems*, 5:411–127, 1997.

[16] B. De Baets and J. C. Fodor. Residual operators of uninorms. *Soft Computing*, 3:89–100, 1999.

[17] D. Ruiz and J. Torrens. Residual implications and co-implications from idempotent uninorms. *Kybernetika*, 40:21–38, 2004.

[18] D. Ruiz-Aguilera and J. Torrens. S-implications and R-implications from uninorms continuous in $]0,1[^2$. *Fuzzy Sets and Systems, doi: 10.1016/j.fss.2008.05.015.*, 2008.

[19] M. Baczyński and B. Jayaram. (U,N)-implications and their characterizations. *Fuzzy Sets and Systems. doi:10.1016/j.fss.2008.11.001.*, 2008.

[20] E.P. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Publishers, Dordrecht, 2000.

[21] P. Smets and P. Magrez. Implications in fuzzy logic. *International Journal of Approximate Reasoning*, 1:327–347, 1987.

[22] D. Ruiz-Aguilera and J. Torrens. Distributivity and conditional distributivity of a uninorm and a continuous t-conorm. *IEEE Transactions on Fuzzy Systems*, 14(2):180–190, 2006.

ISBN: 978-989-95079-6-8

A logic-based approach for evaluating interpretability of fuzzy rule-based classifiers

Corrado Mencar Ciro Castiello Anna M. Fanelli

Dept. of Informatics, University of Bari
Bari, Italy
Email: {mencar, castiello, fanelli}@di.uniba.it

Abstract— *We describe an automatic approach for evaluating interpretability of fuzzy rule-based classifiers. The approach is based on the logical view of fuzzy rules, which are interpreted as rows in truth tables. These truth tables are subject of a minimization procedure based on a variant of the Quine-McCluskey algorithm. The minimized truth tables are used to build new fuzzy rules, which are compared with the original ones in terms of classification accuracy. If the two sets of rules have similar performances, we deduce that the logical view of rules is applicable to the fuzzy classifier, which is hence considered interpretable. On the other hand, a significant difference in classification ability shows that fuzzy rules cannot be interpreted in logical terms, hence linguistic labelling may not be significant. Two illustrative examples show both the cases.*

Keywords— fuzzy rule-based classifiers, interpretability assessment, logic minimization, Quine McCluskey algorithm

1 Introduction

Interpretability is one of the most important driving forces for the adoption of fuzzy rule-based systems . The reason is clear: several models exist that can perform intelligent tasks (such as prediction, classification, etc.) but fuzzy systems allow for a representation of knowledge that can be easily read and understood by their users. Interpretability, however, is *not* given for granted when fuzzy models are used, especially when they are acquired from data. The main problem is that data-driven design has a great number of degrees of freedom (number of fuzzy sets, their shape, position, etc.) and may end up with fuzzy models that are very accurate but very hard to comprehend. For this reason, interpretability constraints have been defined so as to bind data-driven design in order to derive interpretable fuzzy models [1]. This usually comes to a price that is a lower accuracy with respect to unconstrained design. Furthermore, often interpretability is accounted without taking care of accuracy. This approach has been criticized, since interpretable but inaccurate models are as useless as very accurate but not interpretable ones [2].

Interpretability can be viewed at different levels of a fuzzy model. Interpretability of fuzzy sets involves the possibility of tagging them with linguistic labels. On a higher level, interpretability of partitions concerns the possibility of linguistically describe the domain of data. On the rule-level, which is of our concern in this paper, interpretability is viewed as the ability of explaining relationships between input and output variables in a linguistic form.

An important issue concerns the evaluation of interpretability. Interpretability assessment is important because several fuzzy rule-based models can be derived from the same data. Many of these models could provide for high accuracies but can vary in their interpretability. As a consequence, a tool for interpretability assessment could be helpful in the choice of the final model or, to a greater extent, the design technique. However, interpretability assessment is an ill-posed problem because the definition of interpretability eludes any formal characterization.

In [3], Michalski gives a referential definition of interpretability, the so-called "comprehensibility postulate", reported in the following.

> The results of computer induction should be symbolic descriptions of given entities, semantically and structurally similar to those a human expert might produce observing the same entities. Components of these descriptions should be comprehensible as single "chunks" of information, directly interpretable in natural language, and should relate quantitative and qualitative concepts in an integrated fashion.

This postulate justifies the use of linguistic values in rule-based fuzzy systems, but that is not enough to guarantee interpretability (for a more extensive discussion of the postulate, see [4]). In [5] Zadeh introduces the notion of *co-intension*, a semantic relation between concepts. Roughly speaking, two concepts are co-intensive if they refer to almost the same entities. In fuzzy rule-based systems, rules are defined by composition of linguistic terms, which are related to two different semantics. The first one is defined by the fuzzy model and the second one is implicitly designated by the linguistic label. By merging the notion of co-intension with the comprehensibility postulate, we derive a formulation of interpretability that can be more helpful for designing assessing techniques: *a rule base is interpretable if the two semantics related to each linguistic term are co-intensive*.

On the basis of this definition, we propose an automatic technique for evaluating interpretability. Our approach evaluates interpretability by assessing the co-intension of the semantics of the rule base of a fuzzy model with the intrinsic semantics designated by linguistic labels. The core of the evaluation technique relies on the propositional view of rules and on logical operations. We expect that, for interpretable knowledge bases, logical operations on rules do not change their semantics and, hence, do not affect accuracy. If this is assumption is violated, we deduce a lack of interpretability.

We focus our research on fuzzy rule-based classifiers, described in Section 2. The proposed approach is then described, by first focusing on its rationale and then on the methodology

ISBN: 978-989-95079-6-8

(Sections 3.1 and 3.2, respectively). Two illustrative examples are reported in Section 4, to show how the proposed approach can be useful in detecting lacks of interpretability for two simple knowledge bases. The paper is concluded with some final remarks.

2 Fuzzy rule-based classifiers

We consider a classifier as a system computing a function of type:

$$f : \mathbf{X} \longrightarrow \Lambda, \tag{1}$$

where $\mathbf{X} \subseteq \mathbf{R}^n$ is an n-dimensional input space, and $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_c\}$ is a set of class labels. If a dataset D of pre-classified data is given, i.e.

$$D = \{(\mathbf{x}_i, l_i) | \mathbf{x}_i \in \mathbf{X}, l_i \in \Lambda, i = 1, 2, \ldots, N\}, \tag{2}$$

then the classification error can be computed as:

$$E(f, D) = \frac{1}{N} \sum_{i=1}^{N} (1 - \chi(l_i, f(\mathbf{x}_i))), \tag{3}$$

being $\chi(a, b) = 1$ iff $a = b$ and 0 otherwise.

A fuzzy rule-based classifier (FRBC) is a system that carries out classification (1) through inference on a knowledge base. The knowledge base includes the definition of a linguistic variable for each input. Thus, for each $j = 1, 2, \ldots, n$, linguistic variables[1] are defined as:

$$V_j = (v_j, X_j, Q_j, S_j, I_j), \tag{4}$$

being:

- v_j the name of the variable;
- X_j the domain of the variable (it is assumed that $\mathbf{X} = X_1 \times X_2 \times \cdots \times X_n$);
- $Q_j = \{q_{j1}, q_{j2}, \ldots, q_{jm_j}, \text{ANY}\}$ is a set of labels denoting linguistic values for the variable (e.g. SMALL, MEDIUM, LARGE);
- $S_j = \{s_{j1}, s_{j2}, \ldots, s_{jm_j+1}\}$ is a set of fuzzy sets on X_j, $s_{jk} : X_j \to [0, 1]$;
- I_j associates each linguistic value q_{jk} to a fuzzy set s_{jk}. We will assume that $I_j(q_{jk}) = s_{jk}$.

We assume that each linguistic variable contains the linguistic value "ANY" associated to a special fuzzy set $s \in S_j$ such that $s(x) = 1, \forall x \in X_j$.

The knowledge base of a FRBC is defined by a set of R rules. Each rule can be represented by the schema:

$$\text{IF } v_1 \text{ IS } [\text{NOT}]\ q_1^{(r)} \text{ AND } \cdots \text{ AND } v_n \text{ IS } [\text{NOT}]\ q_n^{(r)} \text{ THEN } \lambda^{(r)} \tag{5}$$

being $q_j^{(r)} \in Q_j$ and $\lambda^{(r)} \in \Lambda$. Symbol NOT is optional for each linguistic value. If for some j, $q_j^{(r)} = \text{ANY}$, then the corresponding atom "v_j IS ANY" can be removed from the representation of the rule.[2]

Inference is carried out as follows. When an input $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is available, the strength of each rule is calculated as:

$$\mu_r(\mathbf{x}) = s_1^{(r)}(x_1) \otimes s_2^{(r)}(x_2) \otimes \cdots \otimes s_n^{(r)}(x_n), \tag{6}$$

being $s_j^{(r)} = \nu_j^{(r)}(I_j(q_j^{(r)}))$, $j = 1, 2, \ldots, n$, $r = 1, 2, \ldots, R$. Function $\nu_j^{(r)}(t)$ is $1 - t$ if NOT occurs before $q_j^{(r)}$, otherwise it is defined as t. The operator $\otimes : [0, 1]^2 \to [0, 1]$ is usually a t-norm, such as minimum or product.

The degree of membership of input $\mathbf{x}$ to class λ_i is computed by considering all the rules of the FRBC as:

$$\mu_{\lambda_i}(\mathbf{x}) = \frac{\sum_{r=1}^{R} \mu_r(\mathbf{x}) \chi(\lambda_i, \lambda^{(r)})}{\sum_{r=1}^{R} \mu_r(\mathbf{x})}. \tag{7}$$

Finally, since just one class label has to be assigned for the input $\mathbf{x}$, the FRBC assigns the class label with highest membership (ties are solved randomly):

$$f_{FRBC}(\mathbf{x}) = \lambda \Rightarrow \mu_\lambda(\mathbf{x}) = \max_{i=1,2,\ldots,c} \mu_{\lambda_i}(\mathbf{x}). \tag{8}$$

3 Interpretability assessment

We assume the availability of an interpretable FRBC, verifying a number of interpretability constraints so that the rule base is described in terms of linguistic values.

Interpretability of FRBC is usually achieved by imposing a number of constraints on the knowledge base. For our purposes, it is necessary that labels denoting linguistic values for each variable can be interpreted as well-distinguished elementary concepts. To achieve this goal, normality, convexity and distinguishability are the basic interpretability constraints to be imposed on the FRBC. Normality is indeed required to represent consistent concepts (i.e. concepts that are fully applicable to at least one object), convexity is necessary to represent elementary concepts (i.e. concepts that cannot be split in two or more sub-concepts) and distinguishability allows the use of different labels to denote different fuzzy sets (for a detailed explanation, the reader is referred to [1]).

3.1 Rationale

The proposed approach for interpretability assessment relies on the formal structure of the FRBC. The rationale behind this approach comes from the observation that the rule base is the linguistic interface of the FRBC to the user. For an interpretable knowledge base, the user should be able to understand the classification rules by simply observing their linguistic representation. All the semantic information (fuzzy sets attached to linguistic values, t-norm used for conjunction, etc.) should be hidden to the user because – this is the key point of interpretability – the semantics of FRBC knowledge should be co-intensive with the user's knowledge, recalled by the linguistic terms.

The proposed approach tries to evaluate how much co-intensive is the semantics of the FRBC knowledge w.r.t. user knowledge. As a further requirement, we want the evaluation

[1] For the sake of clarity, in this paper we use a simplified definition of linguistic variable.

[2] The sequence NOT ANY is not allowed.

ISBN: 978-989-95079-6-8

process to be carried out automatically. To clarify its rationale, we consider the example of two communicating actors, A and B. When A communicates some piece of knowledge to B, he/she uses a linguistic structure that (at least partially) represents his/her knowledge. To be comprehensible, A chooses linguistic terms whose semantics is deemed as co-intensive as possible with B's knowledge. It is not necessary that the semantics of linguistic terms in A's mind perfectly match the semantics in B's mind: a high "overlapping" of semantics is enough for knowledge communication. The actor A can be almost sure of the co-intension with B if they share a similar environment, language, culture, experiences, etc. Thus, co-intension can be achieved if A and B share similar cognitive structures.

To assess interpretability, we exploit the cognitive structures that are shared by users and FRBC. In particular, we observe a strict affinity of a FRBC rule base to logical propositions. Actually, rules are formed so as to resemble propositions, so that they can be understood by users. In consequence of this, FRBC and users share this propositional view of rules. Being like propositions, rules could be transformed by truth-preserving operators without any change of the semantics. This is not completely true since the application of such operators may distort the semantics of rule (which is fuzzy); however, we should expect only small distortions.

The core of our approach is the following: given a rule base of a FRBC, we represent it as a collection of propositions, and then we transform it through a truth-preserving operator, thus obtaining a new set of propositions, that represents a new rule base. We then compare the two rule bases on the basis of their classification ability: if they differ too much, then we conclude that the logical view of rules is wrong. Also, since rules are defined so as to resemble logical propositions, we derive that the semantics of rules (which is responsible of classification) is such that logical view is not possible. This means that the semantics of rules is not co-intensive with user knowledge, since users expect that truth-preserving transformations of propositions do not change (too much) their semantics.

A simple example may further clarify the rationale of our approach. Suppose to have a trivial FRBC with the following two rules:

```
IF cell_size is Large AND
   cell_shape is Irregular THEN malign
IF cell_size is Small THEN benign
```

Any user reading this rule base understands that the size of a cell is enough to discriminate between malign and benign classes. As a consequence, a truth-preserving transformation of the rule base could lead to the following set of rules:

```
IF cell_size is Large THEN malign
IF cell_size is Not Large THEN benign
```

This understanding is legitimate under the Closed World Assumption (CWA), according to which any input satisfies the condition of at least one rule. CWA is generally valid in most fuzzy systems and FRBC in particular, hence it is taken for grant in this work. Open World Assumption (OWA), which leads to non-monotonic reasoning, is out of the scope of this paper.

The two rule bases can be compared in terms of classification errors. If the two classification errors are very similar, we derive that the semantics of the rule base is compatible with the applied transformation and, hence, to the logical view of rules. In this sense, we state that the knowledge base of the FRBC is co-intensive with user knowledge. On the other hand, if the classification errors differ too much, then the semantics of the knowledge base cannot be represented with a set of propositions, hence, it is not co-intensive with user mind. In other words, it is not interpretable.

We cannot expect that classification errors are always identical, even for very interpretable FRBC, because the law of excluded middle ($A\bar{A} = 0$) is not valid in fuzzy logic but it holds in Boolean logic. On the other hand, we should expect that the violation of excluded middle does not influence too much the inference process, otherwise we should conclude that the FRBC is heavily based on inconsistent knowledge. This latter situation clashes with the property of interpretability.

An issue arises in the choice of the truth-preserving transformation. Actually, several transformations can be considered, but we choose to apply a transformation that minimizes the number of linguistic terms used. This choice has a twofold advantage. First, by eliminating as many terms as possible, we test if logical view of rules is preserved with the minimum required information. Second, if assessment leads to positive results, we could retain the simplified rule base because it is easier to read than the original.

3.2 Methodology

The proposed approach for interpretability assessment is based on a four-stage strategy. It is similar to the approach proposed in [8], however our strategy is able to deal with negative information to achieve more compact representations of the rule bases.

3.2.1 Definition of truth tables

Each rule of the FRBC is seen as a proposition, i.e. a combination of propositional variables that is considered true for a class.

For each class label $\lambda_i \in \Lambda$ and for each rule r, a truth function π_i is defined on the propositional variables defined for the FRBC as:

$$\pi_i(\chi_{11}^{(r)}, \ldots, \chi_{1m_1}^{(r)}, \chi_{n1}^{(r)}, \ldots, \chi_{nm_n}^{(r)}) = \chi(\lambda_i, \lambda^{(r)}), \quad (9)$$

being $\chi_{jk}^{(r)} = \chi(q_j^{(r)}, q_{jk})$. Inputs $\chi_{jk}^{(r)}$ assume value DC ("don't care") if the corresponding linguistic value is ANY or, in other words, the linguistic variable V_j does not occur in the r-th rule. For any other combination of inputs, the output of π_i is undefined, i.e. any truth value is possible (again, this condition is usually referred as "don't care").

Each truth function π_i can be represented as a truth table, which enumerates any combination of assignments to the propositional variables of the FRBC and associates the value of π_i to each combination. Combinations associated to undefined values of π_i are not included in the table. The number of rows of each truth table matches the number of rules of the FRBC. This prevents the combinatorial explosion of rows that

ISBN: 978-989-95079-6-8

Small	Large	Regular	Irregular	Malign	Benign
0	1	0	1	1	0
1	0	DC	DC	0	1

Table 1: The truth tables of the simple rulebase

Small	Large	Regular	Irregular	Malign	Benign
DC	1	DC	DC	1	0
DC	0	DC	DC	0	1

Table 2: The minimized truth tables of the simple rulebase

would be expected in the general case of truth function representation. Table 1 shows the truth tables for classes "malign" and "benign" of the previous example.

3.2.2 Minimization

Once each truth table has been built, it can be processed so as to be minimized. The minimization procedure produces a new truth table without modifying the truth function (where it is defined). The new truth table has a number of rows not greater than the original truth table. It also has a number of DC values in its inputs not smaller than in the original truth table. Furthermore, minimization guarantees that any further simplification (in terms of rows and inputs) provides for a truth function different from the original.

The Quine-McCluskey (QMC) algorithm represents an effective mechanism for minimization of truth tables [7]. It is mainly based on the distributive property, which simplifies propositions according to the law: $ABC + A\bar{B}C \equiv AC$.

The QMC algorithm works in two stages:

1. Merge rows with output 1 or DC that differ in only one input;
2. Find the minimum number of merged rows that cover all rows of the original truth table.

Although computationally expensive, the QMC algorithm can be implemented by an efficient procedure that exploits the peculiar structure of truth tables derived from FRBC rules to perform minimization quickly. A specific implementation of the first stage of QMC avoids the generation of all input combinations, thus saving time for merging. This can be achieved because the number of rows of the truth tables representing rule bases is often very small. Also, the second stage can be optimized by using heuristic procedures that drive the minimization process without expensive search. Table 2 shows the minimized truth table for the example FRBC.

3.2.3 Reconstruction

After minimization, a new FRBC is built from the rows of the minimized truth table. For each class label $\lambda_i \in \Lambda$ we consider the minimized table associated to the truth function π_i. A rule is built for each row with output equal to 1. It is easy to show that for each j there is at most one k such that $\chi_{jk}^{(r)} \neq DC$. Therefore, the antecedent of the rule can be defined by atoms v_j IS [NOT] $q_j^{(r)}$ where:

- $q_j^{(r)} = q_{jk}$ if $\chi_{jk}^{(r)} \neq DC$;
- NOT occurs if $\chi_{jk}^{(r)} = 0$ and it does not occur if $\chi_{jk}^{(r)} = 1$;
- $q_j^{(r)} =$ ANY if $\forall k : \chi_{jk}^{(r)} = DC$;

Atoms with ANY are removed to improve the readability of the rule. The consequent of the rule is λ_i.

At the end of reconstruction, each rule is formed by a sequence of atoms tied together by conjunction (AND). In this paper we do not deal with more complex representations (such as conjunctions of disjunctions of atoms) that could provide for more compact representations of rules but deserve further investigation that will be object of future research.

3.2.4 Comparison

The two FRBCs, the first with the original rule base and the second with the minimized version, are compared in terms of classification error on the same data.

If the two errors differ too much, we conclude that the original FRBC lacks of interpretability. Its accuracy is mainly due to the semantic definition of linguistic values, which do not correspond to the propositional view of rules. The FRBC can be used for classification as a "grey box", but its labelling is arbitrary and not co-intensive with user knowledge. Attaching natural language terms to such FRBC is useless and potentially misleading.

If the two errors are very similar, we conclude that the original FRBC is interpretable. The semantic definition of linguistic values is coherent with the logic operators used in minimization. In this sense, the semantic of linguistic values is co-intensive with user knowledge. We could retain the simplified FRBC because its interpretability is greater than the original (due to its higher simplicity) while its accuracy is almost the same of the original.

There is no threshold to decide if a FRBC is interpretable or not, but rather a continuous spectrum of possibilities. Interpretability – as expected – is a matter of degree, and the degree of interpretability is, in our approach, inversely proportional to the difference of accuracies. Even with small variations of accuracy, some considerations can be drawn on the interpretability of the FRBC, as shown in the next Section.

4 Illustrative examples

In this Section we evaluate two different FRBC derived from the same data set with two different methods. The first method is oriented toward interpretability while the second is more inclined to the design of accurate models, albeit taking into account some kind of readability of the knowledge base.

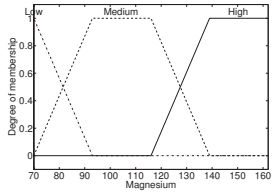
4.1 Interpretable FRBC

We consider a FRBC obtained from the application of HILK [6], a tool for building interpretable fuzzy rule-based systems. HILK allows for the definition of fuzzy rule-based systems from empirical learning, expert knowledge, or both. The resulting systems are considered highly interpretable because fuzzy sets defined for each input satisfy a number of interpretability constraints. Furthermore, the number of fuzzy sets per input, the number of inputs and the number of rules are kept as small as possible since, in principle, the simplest is the knowledge representation, the easier is to read and understand.

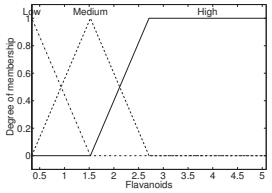
ISBN: 978-989-95079-6-8

If Flavonoids is Low THEN Class 1
If Flavonoids is Medium THEN Class 2
If Flavonoids is High AND Proline is Low THEN Class 2
If Flavonoids is High AND Proline is High THEN Class 3
If Magnesium is Medium AND Flavonoids is High AND Proline is Medium THEN Class 3

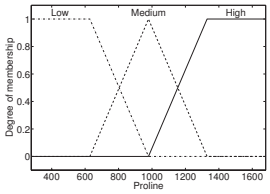
Figure 1: The rule base of the FRBC considered in the first example



(a) Magnesium



(b) Flavonoids



(c) Proline

Figure 2: The linguistic variables used in the first example

In our experimentation, the FRBC obtained from HILK was acquired from data, in order to classify Wine data, a well-known benchmark dataset, freely available from UCI repository [10]. The knowledge base of the FRBC is reported in fig. 1, while the linguistic variables are shown in fig. 2. The FRBC provided 10.67% of classification error on the entire dataset.

We transformed the rule base of the FRBC into three truth tables – one for each class – minimizing them with QMC algorithm; then we rebuilt the FRBC obtaining the simplified rule base reported in fig. 3.

We observe that the number of rules has been reduced to four, and the linguistic variable "Magnesium" has been removed. From the logical viewpoint, the two rule bases are equivalent since they compute the same truth functions. However, after applying the minimized rule base to the dataset, we obtained 11.24% of classification error, i.e. an increase of $+0.57\%$, corresponding to one more misclassified pattern over 178, which is defined as (Magnesium=162, Flavonoids=2.27, Proline=937). For the misclassified pattern we observe that, according to the definitions of the linguistic variables, Magnesium is high, Proline is medium, and Flavonoids is both medium and high to a significant degree. However, in the original rule base, there is not any classification rule activated when Magnesium is high, Proline is medium and Flavonoids is high. From the logical viewpoint, the truth functions are undefined for this condition. As a consequence, the simplified rule base subsumes this condition in the fourth rule, assigning the pattern to class 3, while the original FRBC arbitrarily assigns the pattern to class 2. The correct classification of the pattern performed by the original FRBC

If Flavonoids is Low THEN Class 1
If Flavonoids is Medium THEN Class 2
If Flavonoids is High AND Proline is Low THEN Class 2
If Flavonoids is High AND Proline is NOT Low THEN Class 3

Figure 3: The rule base of the FRBC in the first example after minimization

If Alc. is High And Fla. is High And A.Ash is Low And OD2 is High And Pro. is High Then class 1
If Alc. is Low And Pro. is Low Then class 2
If Alc is Medium And A.Ash is High And Fla. is Low And OD2 is Low And Pro. is Low Then class 3

Figure 4: The original rule base of 2nd example (Alc. = Alcohol, Fla. = Flavonoids, A. Ash = Alcalinity of Ash, OD2 = OD280/315, Pro. = Proline)

is due to the semantic specification of linguistic values, which does not emerge from the propositional view of rules. In this sense, the original FRBC lacks of interpretability for a specific case. However, due to the reduced increase of classification error, we conclude that the original FRBC is highly interpretable and we can retain the simplified version, which offers a further comprehensible knowledge base.

4.2 Non-interpretable FRBC

As a second example, we consider the FRBC obtained through the approach proposed in [9] aimed at optimizing accuracy by taking into account interpretability. In brief, the approach consists of an iterative procedure for developing the FRBC. The initial model is derived from the data and, subsequently, feature selection and rule-base simplification are applied to reduce the model, while a genetic algorithm is used for parameter optimization.

This approach has been applied to Wine data showing very good classification accuracy (1.1% of classification error[3]). One of the best rule bases derived from data is reported in fig. 4. We observe from fig. 5 that in some cases interpretability constraints are violated. As an example, we observe in fig. 5(b) that the rightmost fuzzy set is highly overlapping with the other fuzzy sets of the variable. Notwithstanding, the three fuzzy sets have different labels, thus recalling distinct concepts.

After simplification, we obtain the very small rule base reported in fig. 6. The reconstructed FRBC shows an unacceptable classification error of about 32%. In consequence of this considerable decay of classification performance, we deduce that classification accuracy of this FRBC is mainly based on the parameters of the fuzzy sets that cannot be expressed in linguistic terms. In consequence of this, we state that the logical view of rules is not applicable to the original rule base. In this sense, we conclude that the FRBC is not interpretable.

5 Conclusion

Assessment of interpretability is not an easy task. Difficulty is mainly due to the blurry definition of interpretability, which requires co-intension with human knowledge. Evaluating interpretability is a subjective task, which could be tedious and even ill-posed, therefore automatic techniques for

[3] We use the results reported in the original paper

ISBN: 978-989-95079-6-8

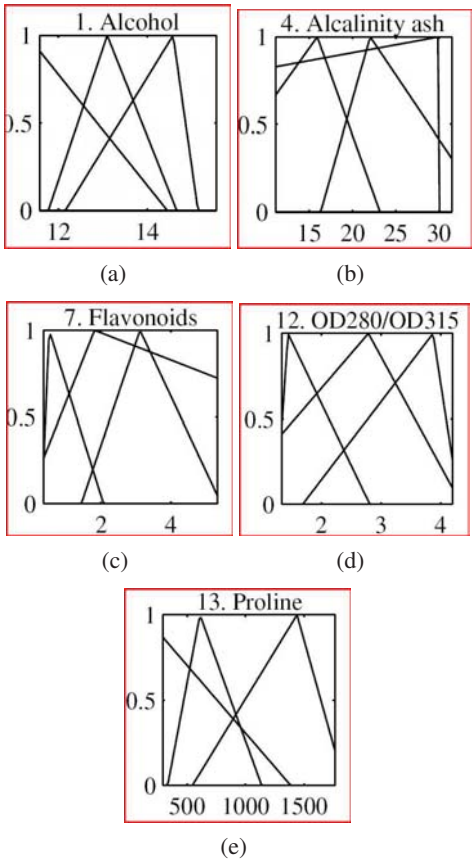


Figure 5: The fuzzy sets used in the second example (for each variable, the leftmost fuzzy set is labelled “Low”, the middle one “Medium” and the rightmost “High”)

If Alcohol is High Then class 1
If Alcohol is Low Then class 2
If Alcohol is Medium Then class 3

Figure 6: The minimized rule base of 2nd example

interpretability assessment are useful, but they should embody some information on semantic co-intension. In this paper, we have described an approach for automatically evaluating interpretability of rule-based fuzzy classifiers, exploiting the propositional view of rules as a mean to define co-intension. On the basis of this hypothesis, a procedure has been devised, so as to evaluate how much the semantics of fuzzy rules is coherent with their logical view.

The illustrative examples show the effects of the proposed approach for a simple knowledge base. Actually, we were able positively assess the interpretability for a FRBC that verifies several interpretability constraints. Also, we have verified that a FRBC is not interpretable even though it is described in linguistic terms. By analyzing the relative performances of the two FRBCs we immediately observe that interpretability preservation comes to the cost of accuracy reduction and, on the other hand, an accurate description of data (preserving interpretability constraints) would require a great number of rules. This, however, would clash with the Comprehensibility Postulate.

Research on this topic is in progress, especially in the direction of enhancing the efficiency of the minimization algorithm. That will allow an extensive experimentation over a wider class of knowledge bases, in order to promote deeper insights of the proposed technique. Further enhancements could spring from the use of additional information to refine the definition of co-intension in computational terms.

References

[1] C. Mencar. *Theory of Fuzzy Information Granulation: Contributions to Interpretability Issues.* PhD thesis, University of Bari, Italy, 2005

[2] A. Riid, E. Rustern. Interpretability of Fuzzy Systems and Its Application to Process Control. *In: Proc. of IEEE International Fuzzy Systems Conference FUZZ-IEEE 2007*, pp. 1–6, 2007

[3] R.S. Michalski. A theory and methodology of inductive learning. *Technical Report UIUCDCS-R-83-1122*, Department of Computer Science, University of Illinois, Urbana, 1983

[4] C. Mencar, G. Castellano, A. M. Fanelli. On the role of interpretability in fuzzy data mining. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS)* 15(5):521 – 537, 2007

[5] L.A. Zadeh. Is there a need for fuzzy logic? *Information Sciences* 178:2751–2779, 2008

[6] J.M. Alonso, L. Magdalena, S. Guillaume. HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *International Journal of Intelligent Systems* 23:761–794, 2008

[7] E.J. McCluskey. Minimization of Boolean functions. *Bell Labs Technical Journal* 35(5):1417–1444,1956

[8] R. Rovatti, R. Guerrieri, G. Baccarani. An enhanced two-level Boolean synthesis methodology for fuzzy rules minimization, *Fuzzy Systems, IEEE Transactions on* 3(3):288–299, 1995

[9] J.A. Roubos, M. Setnes, J. Abonyi. Learning fuzzy classification rules from labeled data. *Information Sciences* 150:77–93,2003

[10] *UCI Machine Learning Repository Content Summary:* http://www.ics.uci.edu/ ˜mlearn/MLSummary.html

ISBN: 978-989-95079-6-8

Cloud based design optimization

Martin Fuchs[12]

1. University of Vienna, Faculty of Mathematics, Vienna, Austria
2. CERFACS, Toulouse, France
Email: martin.fuchs81@gmail.com
www.martin-fuchs.net

Abstract— *Cloud based design optimization (CBDO) is an approach to significantly improve robustness and optimality of solutions sought in engineering design. One of the main features is the possibility to capture and model high-dimensional uncertainty information, even in the case that the information available is incomplete or unformalized.*

Continuing our past studies we present the graphical user interface for CBDO in this paper. Also we mention the latest improvements of our methods, we give an illustrative example demonstrating how unformalized knowledge can be captured, and we highlight relations to different uncertainty models, such as p-boxes, Dempster-Shafer structures, and α-level optimization for fuzzy sets.

Keywords— confidence regions, design optimization, higher dimensions, incomplete information, potential clouds

1 Introduction

Design optimization is frequently affected by uncertainties originating from several different sources. Being already a complicated task in absence of uncertainties, design optimization under uncertainty imposes an additional class of difficulties. We have developed a framework dividing design optimization under uncertainty in its two inherent components, i.e., uncertainty modeling and optimization.

The most critical problems in real-life uncertainty modeling are caused by the well-known curse of dimensionality (cf., e.g., [1]), and by lack of information. While in lower dimensions, lack of information can be handled with several tools (e.g., p-boxes [2], Dempster-Shafer structures [3]), in higher dimensions (say, greater than 10) there exist only very few. Often simulation techniques are used which, however, fail to be reliable in many cases, see, e.g., [4]. The clouds formalism [5] is one possibility to deal with both incomplete and higher dimensional information in a reliable and computationally tractable fashion.

The design optimization phase (cf., e.g., [6]) is the second major subject in our framework, loosely linked with the uncertainty modeling. One typically faces problems like strongly nonlinear, discontinuous, or black box objective functions, or mixed integer design variables. We have developed heuristics to solve these problems, e.g., using separable underestimation [7], or convex relaxation based splitting [8].

Since our approach can be considered as design optimization based on uncertainty modeling with clouds, we call the software *cloud based design optimization* (CBDO). We have implemented an interface for our methods that will be presented later in this paper. The implementation was motivated by the need of expert engineers of an easy-to-use tool, a framework respecting their working habits, and demonstrating usefulness in capturing and modeling incomplete, unformalized knowledge. Current research is focussed on improving both optimization and uncertainty modeling phase, and on capturing more types of information virtually, e.g., linguistic expressions. Of course, we are constantly looking for possible real-life applications of the methods. CBDO has already been successfully used in space system design applications, cf. [9, 10].

This paper is organized as follows. We introduce the formal background of CBDO in Section 2 also giving an illustrative example how we capture unformalized knowledge. In Section 3 we summarize relations of the potential clouds formalism to different uncertainty models. Finally, we present our software implementation, a MATLAB package for CBDO, in Section 4.

2 Clouds and robust optimization

Let ε be an n-dimensional random vector. A *potential cloud* is an interval-valued mapping $x \to [\underline{\alpha}(V(x)), \overline{\alpha}(V(x))]$, where the potential function $V : \mathbb{R}^n \to \mathbb{R}$ is bounded below, and $\underline{\alpha}, \overline{\alpha} : V(\mathbb{R}^n) \to [0,1]$ are functions constructed to be a lower and upper bound, respectively, for the cumulative distribution function (CDF) F of $V(\varepsilon)$, $\underline{\alpha}$ continuous from the left and monotone, $\overline{\alpha}$ continuous from the right and monotone. We define $\underline{C}_\alpha := \{x \in \mathbb{R}^n \mid V(x) \le \underline{V}_\alpha\}$ if $\underline{V}_\alpha := \min\{V_\alpha \in \mathbb{R} \mid \overline{\alpha}(V_\alpha) = \alpha\}$ exists, and $\underline{C}_\alpha := \emptyset$ otherwise; analogously $\overline{C}_\alpha := \{x \in \mathbb{R}^n \mid V(x) \le \overline{V}_\alpha\}$ if $\overline{V}_\alpha := \max\{V_\alpha \in \mathbb{R} \mid \underline{\alpha}(V_\alpha) = \alpha\}$ exists, and $\overline{C}_\alpha := \mathbb{R}^n$ otherwise. Thus we find a nested collection of lower and upper confidence regions in the sense that $\Pr(\varepsilon \in \underline{C}_\alpha) \le \alpha$, $\Pr(\varepsilon \in \overline{C}_\alpha) \ge \alpha$, $\underline{C}_\alpha \subseteq \overline{C}_\alpha$.

Note that lower and upper confidence regions $\underline{C}_\alpha, \overline{C}_\alpha$ – also called α-cuts of the cloud – are level sets of V. By choosing the potential function V reasonably one gets an uncertainty representation of high-dimensional, incomplete, and/or unformalized knowledge, cf. [11].

Our framework of cloud based design optimization consists of three essential parts, described in the following sections: uncertainty elicitation, uncertainty modeling, and robust optimization.

2.1 Uncertainty elicitation and modeling

We assume that the initially available uncertainty information consists of both formalized and unformalized knowledge. The formalized knowledge can be given as marginal CDFs, interval bounds on single variables, or real sample data. In real-life situations there is often only interval information, sometimes marginal CDFs without any correlation information, available for the uncertain variables. Moreover, there is typically a significant amount of unformalized knowledge available based

ISBN: 978-989-95079-6-8

on expert experience, e.g., knowledge about the dependence of variables.

Potential clouds enable to capture and formally represent this kind of information. We illustrate this by a simple example: First, we generate a data set from an $N(0,\Sigma)$ distribution with covariance matrix $\Sigma = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}$.

Assume that this data belongs to 2 random variables with a physical meaning, and that the data was given to an expert, without any information about the actual probability distribution of the random variables. Still, the expert may be able to provide vague, unformalized information about the dependence of the variables (opposed to formal knowledge, e.g., correlation information) from his knowledge about the physical relationship between the variables. We model this knowledge by polyhedral constraints on the variables, see, e.g., Fig. 3. We choose the potential function V according to these constraints, i.e., the lower and upper confidence regions $\underline{C}_\alpha, \overline{C}_\alpha$ constructed with clouds become polyhedra. The polyhedra reasonably approximate confidence regions of the true, but unknown distribution linearly, as shown in Fig. 1, although the information was vague and unformalized.

In more than 2 dimensions the polyhedral constraints are provided for projections to 1-dimensional or 2-dimensional subspaces.

It should also be highlighted that this approach for providing unformalized knowledge also allows for information updating, simply by adding further polyhedral constraints.

On the basis of the given information we use the confidence regions constructed by clouds in order to search for worst-case scenarios of certain design points via optimization techniques. The construction of the confidence regions is possible even in case of scarce, high-dimensional data, incomplete information, unformalized knowledge.

For further details on the construction of potential clouds the interested reader is referred to [11]. A comparison of different existing uncertainty models can be found in [12], and Section 3 gives a short summary.

2.2 *Robust optimization*

Assume that we wish to find the design point $\theta = (\theta^1, \theta^2, \ldots, \theta^{n_0})$ with the minimal design objective function value g under uncertainty of the n-dimensional random vector ε. Let $\mathbf{T}$ be the set of possible selections for the design point θ. Assume that the function G models the functional relationship between different design components and the objective function. Also assume that the uncertainty of ε is described by a convex set $\mathcal{C}$, in our case a polyhedral α-cut from the cloud.

We embed the confidence regions constructed above in a problem formulation for robust design optimization as follows:

$$\begin{aligned} \min_{\theta} \quad & \max_{\varepsilon} \; g(x) \\ \text{s.t.} \quad & x = G(\theta, \varepsilon), \\ & \varepsilon \in \mathcal{C}, \\ & \theta \in \mathbf{T}, \end{aligned} \tag{1}$$

where $g : \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^{n_0} \times \mathbb{R}^n \to \mathbb{R}^m$.

The optimization phase minimizes a certain objective function g (e.g., cost, mass of the design) subject to safety constraints $\varepsilon \in \mathcal{C}$ to account for the robustness of the design, and subject to the functional constraints which are represented by the underlying system model G.

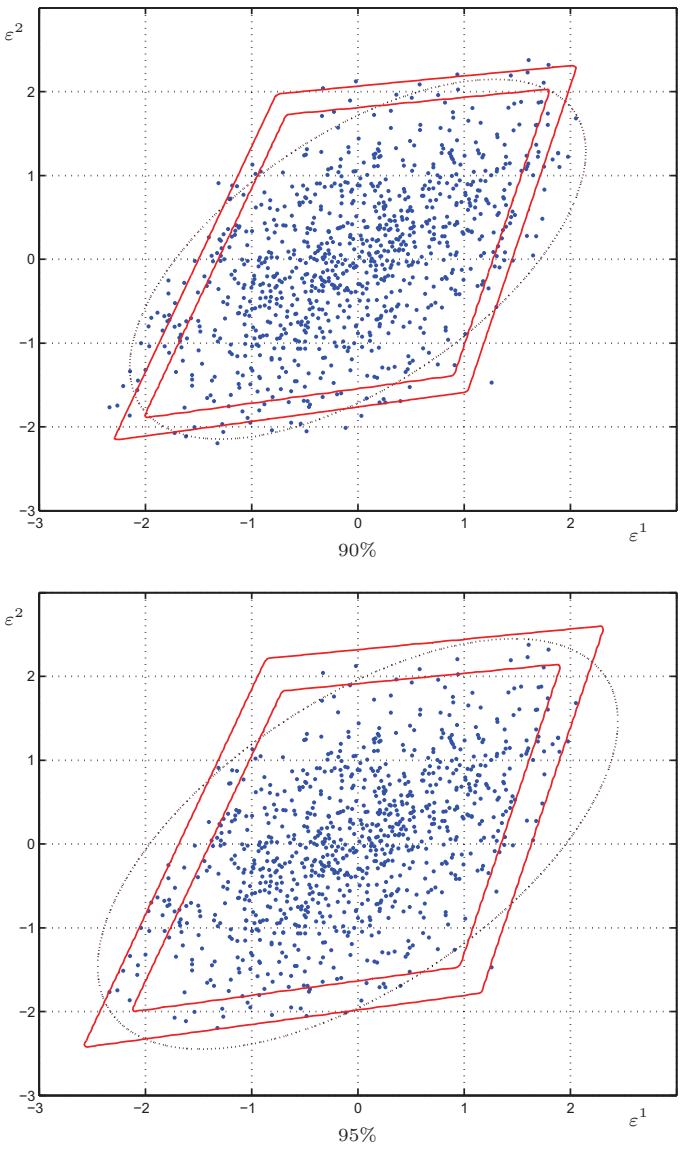


Figure 1: Approximation of confidence regions by 90% and 95% α-cuts, respectively: The polyhedral cloud results in confidence regions that reasonably approximate confidence regions of the true $N(0,\Sigma)$ distribution although the information was given unformalized.

The main difficulties arising from (1) are imposed by the bilevel structure in the objective function, by the mixed integer formulation (since θ^i can be either a discrete or continuous variable), and by the fact that G may comprise strong nonlinearities, or discontinuities, or may be given as a black box.

We have developed multiple techniques to tackle these difficulties and find a solution of (1). For details on approaches to solve such problems of design optimization under uncertainty the interested reader is referred to [7]. The latest improvements of the methods can be found in [8].

3 Relations to different uncertainty models

This section illustrates relations and differences of the potential clouds formalism to three other existing uncertainty mod-

ISBN: 978-989-95079-6-8

els: p-boxes, Dempster-Shafer structures, and α-level optimization for fuzzy sets.

3.1 Relation to p-boxes

A p-box – or p-bound, or probability bound – is a rigorous enclosure of the CDF F of a univariate random variable X, $F_l \leq F \leq F_u$, in case of partial ignorance about specifications of F. Such an enclosure enables, e.g., to compute lower and upper bounds on expectation values or failure probabilities.

There are different ways to construct a p-box depending on the available information about X, cf. [13]. Moreover, it is possible to construct p-boxes from different uncertainty models like Dempster-Shafer structures (cf. Section 3.2). The studies on p-boxes have already lead to successful software implementations, cf. [14, 2].

Higher order moment information on X (e.g., correlation bounds) cannot be handled or processed yet. This is a current research field, cf., e.g., [15]. In higher dimensions, the definition of p-boxes can be generalized similar to the definition of higher dimensional CDFs, cf. [16].

The problem of rigorously quantifying probabilities given incomplete information – as done with p-boxes – is highly complex, even for simple problems, e.g., [17]. Applications of the methods are rather restricted to lower dimensions and non-complex system models G. Black box functions G cannot be handled as one requires knowledge about the involved arithmetic operations. All in all, the methods often appear not to be reasonably applicable in many real-life situations.

The relation to potential clouds becomes obvious, regarding $V(\varepsilon)$ as a 1-dimensional random variable and the functions $\underline{\alpha}$, $\overline{\alpha}$ as a p-box for $V(\varepsilon)$. Thus the potential clouds approach extends the p-box concept to the case of multidimensional ε, without the exponential growth of work in the conventional p-box approach.

3.2 Relation to Dempster-Shafer structures

Dempster-Shafer theory [3] enables to process incomplete and even conflicting uncertainty information. Let $\varepsilon : \Omega \to \mathbb{R}^n$ be an n-dimensional random vector. One formalizes the available information by a so-called *basic probability assignment* $m : 2^\Omega \to [0,1]$ on a finite set $\mathcal{A} \subseteq 2^\Omega$ of non-empty subsets A of Ω, such that

$$m(A) \begin{cases} > 0 & \text{if } A \in \mathcal{A}, \\ = 0 & \text{otherwise,} \end{cases} \tag{2}$$

and the normalization condition $\sum_{A\in\mathcal{A}} m(A) = 1$ holds.

The basic probability assignment m is interpreted as the exact belief focussed on A, and not in any strict subset of A. The sets $A \in \mathcal{A}$ are called *focal sets*. The structure $(m, \mathcal{A})$, i.e., a basic probability assignment together with the related set of focal sets, is called a *Dempster-Shafer structure* (DS structure).

Given a DS structure $(m, \mathcal{A})$ one constructs two fuzzy measures Bel and Pl by

$$\mathrm{Bel}(B) = \sum_{\{A\in\mathcal{A}|A\subseteq B\}} m(A), \tag{3}$$

$$\mathrm{Pl}(B) = \sum_{\{A\in\mathcal{A}|A\cap B\neq\emptyset\}} m(A), \tag{4}$$

for $B \in 2^\Omega$. The fuzzy measures Bel and Pl have the property $\mathrm{Bel} \leq \mathrm{Pr} \leq \mathrm{Pl}$ by construction, where Pr is the probability measure that is unknown due to lack of information.

DS structures can be obtained from expert knowledge or in lower dimensions from histograms, or from the Chebyshev inequality given expectation value μ and variance σ^2 of a univariate random variable X, see, e.g., [18].

To combine different, possibly conflicting DS structures $(m_1, \mathcal{A}_1)$, $(m_2, \mathcal{A}_2)$ (in case of multiple bodies of evidence, e.g., several different expert opinions) to a new basic probability assignment m_{new} one uses Dempster's rule of combination [19].

The complexity of the rule is strongly increasing in higher dimensions, and in many cases requires independence assumptions for simplicity reasons avoiding problems with interacting variables. It is not yet understood how the dimensionality issue can be solved. Working towards more efficient computational implementations of evidence theory it can be attempted to decompose the high-dimensional case in lower dimensional components which leads to so-called compositional models, cf., e.g., [20].

The extension of a function $G(\varepsilon)$ is based on the joint DS structure $(m, \mathcal{A})$ for ε. The new focal sets of the extension are $B_i = G(A_i)$, $A_i \in \mathcal{A}$, the new basic probability assignment is $m_{\mathrm{new}}(B_i) = \sum_{\{A_i\in\mathcal{A}|G(A_i)=B_i\}} m(A_i)$.

To embed DS theory in design optimization one formulates a constraint on the upper bound of the failure probability p_f which should be smaller than an admissible failure probability p_a, i.e., $\mathrm{Pl}(\mathbb{F}) \leq p_a$, for a failure set $\mathbb{F}$. This is similar to the safety constraint in (1). It can be studied in more detail in [21] as evidence based design optimization (EBDO).

It is possible to generate a DS structure that approximates a given potential cloud discretely. Fix some confidence levels $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_N = 1$ of the potential cloud, then generate focal sets and the associated basic probability assignment by

$$A_i := \overline{C}_{\alpha_i} \backslash \underline{C}_{\alpha_i}, \tag{5}$$

$$m(A_1) = \alpha_1, m(A_i) = \alpha_i - \alpha_{i-1}, i = 2, \ldots, N. \tag{6}$$

Thus the focal sets are determined by the level sets of V. An analogous recipe works for approximating p-boxes by DS structures, cf. [13]. Note that focal sets A_i in this construction are not nested, so the fuzzy measures Bel and Pl belonging to $(m, \mathcal{A})$ are not equivalent to possibility and necessity measures.

Conversely, assume that one has a DS structure and the associated fuzzy measures Bel and Pl for the random variable $X := V(\varepsilon)$. Then

$$\underline{\alpha}(t) := \mathrm{Bel}(\{X \leq t\}), \tag{7}$$

$$\overline{\alpha}(t) := \mathrm{Pl}(\{X \leq t\}) \tag{8}$$

give bounds on the CDF of $V(\varepsilon)$ and thus construct a potential cloud.

3.3 Relation to fuzzy sets and α-level optimization

To see an interpretation of potential clouds in terms of fuzzy sets one may consider $\underline{C}_\alpha$, $\overline{C}_\alpha$ as α-cuts of a multidimensional interval valued membership function defined by $\underline{\alpha}$ and $\overline{\alpha}$. The major difference is given by the fact that clouds allow

ISBN: 978-989-95079-6-8

for probabilistic statements, i.e., one cannot go back in the other direction and construct a cloud from a multidimensional interval valued membership function because of the lack of the probabilistic properties mentioned in Section 2. If the interval valued membership function does have these probabilistic properties, it corresponds to consistent possibility and necessity measures [22] which are related to interval probabilities [23].

However, the interpretation of a potential cloud as a fuzzy set with such a membership function shows strong links to α-level optimization for fuzzy sets [24].

The α-level optimization method combines the extension principle and the α-cut representation of a membership function μ of an uncertain variable ε, i.e.,

$$\mu(x) = \sup_{\alpha} \min(\alpha, 1_{C_\alpha}(x)), \tag{9}$$

where 1_A denotes the characteristic function of the set A, $C_\alpha := \{x \mid \mu(x) \geq \alpha\}$ denotes the α-cut of the fuzzy set, in order to determine the membership function μ_f of a function $f(\varepsilon)$, $f : \mathbb{R}^n \to \mathbb{R}$. This is achieved by constructing the α-cuts $C_{f\,\alpha_i}$ belonging to μ_f from the α-cuts C_{α_i} belonging to μ. To this end one solves the optimization problems

$$\min_{\varepsilon \in C_{\alpha_i}} f(\varepsilon), \tag{10}$$

$$\max_{\varepsilon \in C_{\alpha_i}} f(\varepsilon), \tag{11}$$

for different discrete values α_i. Finally from the solution f_{i_*} of (10) and f_i^* of (11) one constructs the α-cuts belonging to μ_f by $C_{f\,\alpha_i} = [f_{i_*}, f_i^*]$.

To simplify the optimization step one assumes sufficiently nice behaving functions f and computationally nice fuzzy sets, i.e., convex fuzzy sets, typically triangular shaped fuzzy numbers.

In n dimensions one optimizes over a hypercube, obtained by the Cartesian product of the α-cuts, i.e., $C_{\alpha_i} = C^1_{\alpha_i} \times C^2_{\alpha_i} \times \cdots \times C^n_{\alpha_i}$, where $C^j_{\alpha_i} := \{\varepsilon^j \mid \mu^j(\varepsilon^j) \geq \alpha_i\}$, $\mu^j(\varepsilon^j) := \sup_{\varepsilon^k, k \neq j} \mu(\varepsilon)$, $\varepsilon = (\varepsilon^1, \varepsilon^2, \ldots, \varepsilon^n)$. Here one has to assume non-interactivity of the uncertain variables $\varepsilon^1, \ldots, \varepsilon^n$.

Using a discretization of the α-levels by a finite choice of α_i the computational effort for this methods becomes tractable. From (9) one gets a step function for μ_f which is usually linearly approximated through the points f_{i_*} and f_i^* to generate a triangular fuzzy number.

Now interpret $\underline{\alpha}$ and $\overline{\alpha}$ from a potential cloud as a multidimensional interval valued membership function and consider a system model $f(\varepsilon) := g(G(\theta, \varepsilon))$ with fixed θ (cf. Section 2.2). Similar to (10,11), optimization of f over $\underline{C}_{\alpha_i}$ for discrete values α_i would give a discretized version of $\underline{\alpha}_f$, i.e., the function belonging to the cloud for $f(\varepsilon)$ given the cloud for ε. Analogously, optimization of f over $\overline{C}_{\alpha_i}$ would give a discretized version of $\overline{\alpha}_f$.

This idea leads to the calculation of functions of clouds which is a current research topic. Also note that in α-level optimization one optimizes over boxes C_{α_i}, that means one assumes that the uncertain variables do not interact. Here a similar idea like interactive polyhedral constraints as described in Section 2.1 could also apply to model unformalized knowledge about interaction of the variables.

4 Cloud based design optimization GUI

We have realized the methods for CBDO in a *graphical user interface* (GUI). To install the software go to the CBDO website [25] and download the CBDO package. A quickstart guide helps through the first steps of the simple installation. A more detailed user manual is also included. How to set up a MATLAB file containing a user defined model is illustrated by an example included in the package.

We have developed the GUI using a sequential, iterative structure. The first and second step represent the uncertainty elicitation. In the first step, the user provides an underlying system model and all formal uncertainty information on the input variables of the model available at the current design stage. In the second step, polyhedral dependence constraints between the variables can be added, cf. Section 2.1. In the third step, the initially available information is processed to generate a cloud that provides a nested collection of confidence regions parameterized by the confidence level α. Thus we produce safety constraints for the optimization (cf. Section 2.2) which is the next step in the GUI. The results of the optimization, i.e., the optimal design point found and the associated worst-case analysis, are returned to the user. In an iterative step the user is eventually given an interactive possibility of adaptively refining the uncertainty information and rerunning the procedure until satisfaction.

4.1 Uncertainty elicitation

After starting the GUI with `cbdogui` from the CBDO folder in MATLAB it asks whether to load the last state to the workspace unless it is run for the first time. In the latter case one should first configure the options to set up the model file and inputs declaration file names, and other user-defined parameters after clicking *Options*/*Edit Options*. The notation – if not self-explanatory – is described in the user manual. Tooltips are given for each option in the GUI to guide the user through the configuration.

Having set up the options one returns to the uncertainty elicitation clicking *Back*. The initially available information can be specified in an inputs declaration file and is modified choosing a variable's name and specifying its associated marginal CDF, or interval bound, respectively, in the first step of the GUI, cf. Fig. 2.

The *Next* button leads to the next step which is scenario exclusion.

4.2 Scenario exclusion

From the information given in the first step the program generates a sample as described in [11]. The second step enables the user to exclude scenarios by polyhedral constraints as shown in Section 2.1, illustrating the great advantage of this approach in modeling unformalized knowledge.

To this end the user selects a 1-dimensional or 2-dimensional projection of the generated sample using the field *Projection* on the right. To add a constraint one hits the *Add constraint* button and defines a linear exclusion by two clicks into the sample projection on the left. All linear constraints can be selected from the *Constraint Selection* box to revise and possibly remove them via the *Remove constraint* button. Fig. 3 shows a possible exclusion in two dimensions.

ISBN: 978-989-95079-6-8

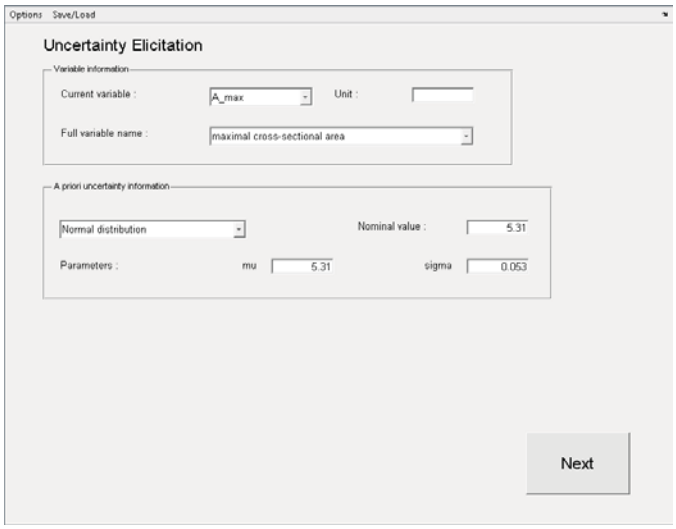


Figure 2: Example for the uncertainty elicitation GUI.

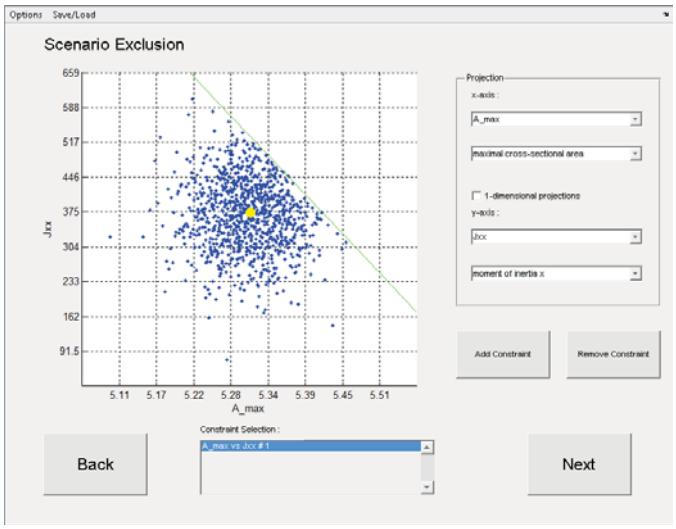


Figure 3: Example for scenario exclusion.

After the exclusion the *Next* button leads to the optimization phase.

4.3 Optimization

The *Start* button initiates two computations: potential cloud generation for (1), and optimization, cf. [7, 8]. As a result one gets the optimal design point found by the program, and the associated objective function value, cf. Fig. 4. It should be remarked that the workspace of the optimization including all results is stored as .mat files in the `cbdo` directory.

The user now has the possibility for the adaptive analysis of the results. Thus the *Next* button leads back to the uncertainty elicitation to be refined.

4.4 Adaptive step

The GUI determining the *a priori* information is not modifiable anymore at this stage of the program. Meanwhile, observe that in the lower part of the GUI a histogram illustrates weighted marginal distributions of the sample.

Hitting the *Next* button makes the scenario exclusion appear again and enables the *a posteriori* adaption of the uncertainty

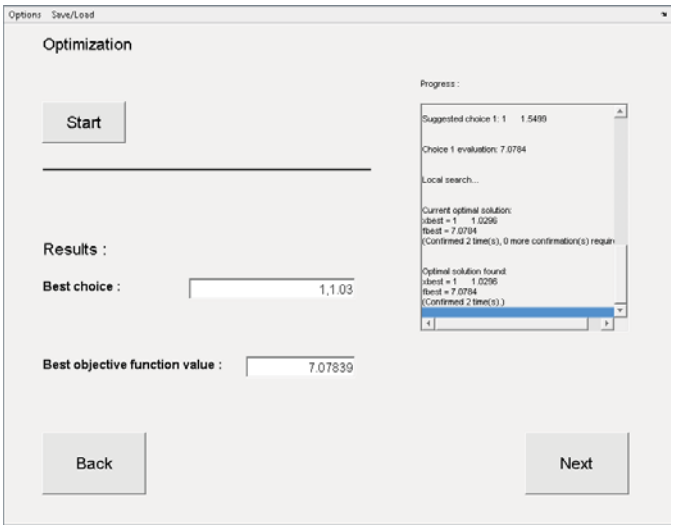


Figure 4: Example for the optimization phase.

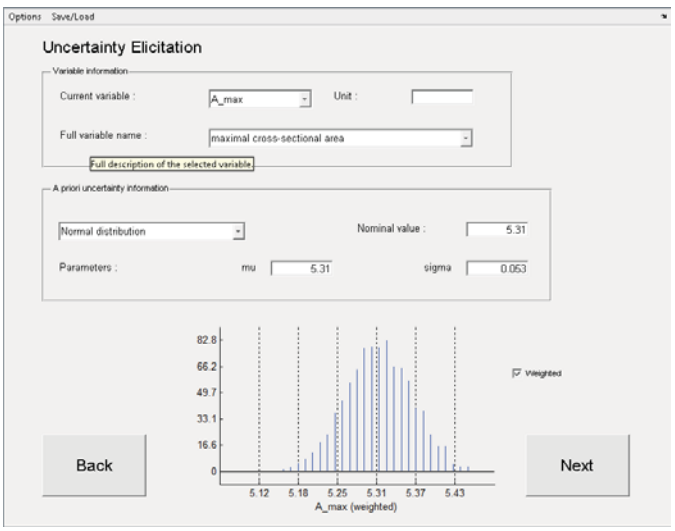


Figure 5: Example for uncertainty elicitation in the adaptive step.

information. For example, the user can consider the worst-case analysis (the worst-case scenario is highlighted with a red dot) to be too pessimistic and exclude it, cf. Fig. 6. Note that this approach is very much imitating real-life working habits of engineers! In early design phases little information is available and safety margins are refined or coarsened iteratively.

The *Next* button leads to the optimization phase again and the user can rerun the procedure until satisfaction.

Acknowledgments

I would like to thank Arnold Neumaier for the valuable discussions during creation of this paper. Also I would like to thank the anonymous reviewers for their constructive comments.

References

[1] P.N. Koch, T.W. Simpson, J.K. Allen, and F. Mistree. Statistical approximations for multidisciplinary optimization: The problem of size. *Special Issue on Multidisciplinary Design Optimization of Journal of Aircraft*, 36(1):275–286, 1999.

ISBN: 978-989-95079-6-8

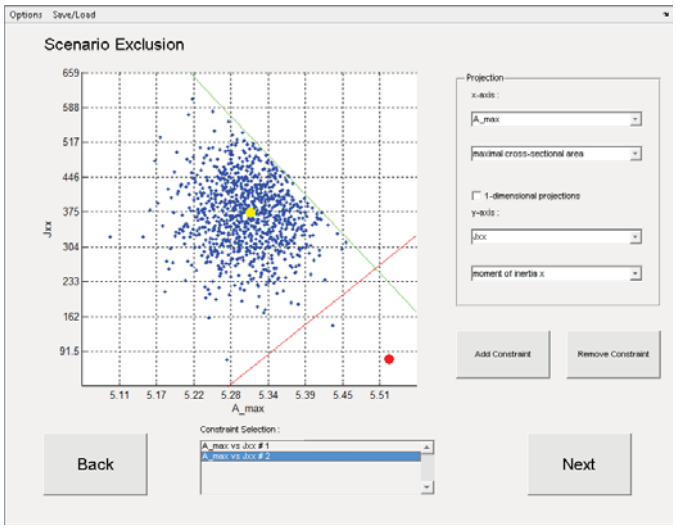


Figure 6: Example for *a posteriori* scenario exclusion.

[2] S. Ferson. *Ramas Risk Calc 4.0 Software: Risk Assessment with Uncertain Numbers*. Lewis Publishers, U.S., 2002.

[3] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[4] S. Ferson, L. Ginzburg, and R. Akcakaya. Whereof one cannot speak: When input distributions are unknown. *Risk Analysis*, 1996. In press, available on-line at: `www.ramas.com/whereof.pdf`.

[5] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10(4):249–272, 2004.

[6] N.M. Alexandrov and M.Y. Hussaini. Multidisciplinary design optimization: State of the art. In *Proceedings of the ICASE/NASA Langley Workshop on Multidisciplinary Design Optimization*, Hampton, Virginia, USA, 1997.

[7] M. Fuchs and A. Neumaier. Autonomous robust design optimization with potential clouds. *International Journal of Reliability and Safety, Special Issue on Reliable Engineering Computing*, 2008. Accepted, preprint available on-line at: `www.martin-fuchs.net/publications.php`.

[8] M. Fuchs and A. Neumaier. A splitting technique for discrete search based on convex relaxation. Submitted, 2009. Preprint available on-line at: `www.martin-fuchs.net/publications.php`.

[9] M. Fuchs, D. Girimonte, D. Izzo, and A. Neumaier. *Robust Intelligent Systems*, chapter Robust and automated space system design, pages 251–272. Springer, 2008.

[10] A. Neumaier, M. Fuchs, E. Dolejsi, T. Csendes, J. Dombi, B. Banhelyi, and Z. Gera. Application of clouds for modeling uncertainties in robust space system design. ACT Ariadna Research ACT-RPT-05-5201, European Space Agency, 2007.

[11] M. Fuchs and A. Neumaier. Potential based clouds in robust design optimization. *Journal of Statistical Theory and Practice, Special Issue on Imprecision*, 2008. Accepted, preprint available on-line at: `www.martin-fuchs.net/publications.php`.

[12] M. Fuchs. Clouds, p-boxes, fuzzy sets, and other uncertainty representations in higher dimensions. *Acta Cybernetica, CSCS Special Issue*, 2009. Accepted, preprint available on-line at: `www.martin-fuchs.net/publications.php`.

[13] S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Sand Report SAND2002-4015, Sandia National Laboratories, 2003. Available on-line at `www.sandia.gov/epistemic/Reports/SAND2002-4015.pdf`.

[14] D. Berleant and L. Xie. An interval-based tool for verified arithmetic on random variables of unknown dependency. Manuscript, 2005.

[15] S. Ferson, L. Ginzburg, V. Kreinovich, and J. Lopez. Absolute bounds on the mean of sum, product, etc.: A probabilistic extension of interval arithmetic. In *Extended Abstracts of the 2002 SIAM Workshop on Validated Computing*, pages 70–72, Toronto, Canada, 2002.

[16] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations–I: Generalized p-boxes. *International Journal of Approximate Reasoning*, 49(3):649–663, 2008.

[17] V. Kreinovich, S. Ferson, and L. Ginzburg. Exact upper bound on the mean of the product of many random variables with known expectations. *Reliable Computing*, 9(6):441–463, 2003.

[18] M. Oberguggenberger, J. King, and B. Schmelzer. Imprecise probability methods for sensitivity analysis in engineering. In *Proceedings of the 5th International Symposium on Imprecise Probability: Theories and Applications*, pages 317–325, Prague, Czech Republic, 2007.

[19] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38(2):325–339, 1967.

[20] R. Jirousek, J. Vejnarova, and M. Daniel. Compositional models of belief functions. In *Proceedings of the 5th International Symposium on Imprecise Probability: Theories and Applications*, pages 243–251, Prague, Czech Republic, 2007.

[21] Z.P. Mourelatos and J. Zhou. A design optimization method using evidence theory. *Journal of Mechanical Design*, 128(4):901–908, 2006.

[22] W.A. Lodwick and K.D. Jamison. Interval-valued probability in the analysis of problems containing a mixture of possibilistic, probabilistic, and interval uncertainty. *Fuzzy Sets and Systems*, 159(21):2845–2858, 2008.

[23] K. Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2–3):149–170, 2000.

[24] B. Möller, W. Graf, and M. Beer. Fuzzy structural analysis using α-level optimization. *Computational Mechanics*, 26(6):547–565, 2000.

[25] CBDO. Cloud Based Design Optimization webpage `www.martin-fuchs.net/cbdo.php`, 2009.

ISBN: 978-989-95079-6-8

Local models for the analysis of spatially varying relationships in a lignite deposit

Bulent Tutmez[1] A. Erhan Tercan[2] Uzay Kaymak[3] Christopher D. Lloyd[4]

1 School of Engineering, Inonu University, Malatya 44280, Turkey
2 School of Engineering, Hacettepe University, Ankara 06532, Turkey
3 Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, the Netherlands
4 School of Geography, Archaeology and Palaeoecology, Queen's University, Belfast, Northern Ireland, UK
Email: btutmez@inonu.edu.tr

Abstract *— Relationships between geographically referenced variables are usually spatially heterogeneous and, to account for such variations, local models are necessary. This paper compares the Geographically Weighted Regression (GWR) model, usually used to integrate and examine the spatial heterogeneity of a relationship, and the Fuzzy Clustering-Based Least Squares (FCBLS) model for the analysis of spatially varying relationships. Both models use the same model parameters and bandwidth values derived from the Akaike Information Criterion. The results show that FCBLS outperforms the GWR model.*

Keywords— *GWR, local model, weighted regression, fuzzy clustering, lignite.*

1 Introduction

Spatial measures contain both attribute and locational information. By their nature, local analyses focus on differences across space whereas global analyses focus on similarities across space. Recently, a variety of useful regression models have been developed to explore the spatial nature of variables [1].

Local modelling has been employed widely in some disciplines for several decades. For example, in image processing local filters have long been used to smooth or sharpen images. However, in some disciplines, like the geosciences in general and geography in particular, a focus on methods that account for local variation has been a comparatively recent development [2]. A variety of localized modelling techniques have been proposed to capture spatial heterogeneity. Geographically Weighted Regression (GWR) is one technique which is being increasingly widely used to explore local spatial variations in relationships, [3]. GWR is a useful and effective technique for locally modelling relationships by calibrating a spatially varying coefficient regression model.

Spatial statistics was developed based on probability and classical statistics. However, many spatial datasets have high levels of uncertainty, and in some cases analyses are dependent on 'soft' data, which may be more qualitative than quantitative in nature [4]. On the other hand, soft approaches like fuzzy computing have desirable features for spatial data analysis [5]: they are based on less restrictive assumptions, and are flexible in modelling non-linearity and non-constant variable structures.

Lignite quality parameters considered in this study have crucial importance in the production of energy in power plants and thus modeling of these parameters is useful in making decisions and in planning future production levels of operating plants [6]. In this study, the spatial relationships between the lignite quality parameters are investigated by local models. To meet this aim, the two local regression methods, GWR and fuzzy clustering based Least Squares, have been applied for local modelling of lignite reserve parameters derived from a deposit. First, both the methods have been used for developing weighted local models and then the performance comparisons have been conducted using a variety of performance indices.

2 Problem and methods

2.1 Description of problem

In general, spatial models and methods of spatial evaluation have been applied at a 'global' level, meaning that one set of results is generated from the models, representing one set of relationships, which is assumed to apply equally across the study area [7]. Although global approaches have proved useful they have the shortcoming that they can mask geographical variations in relationships. Owing to this realization, local regression model approaches have been developed that permit the exploration of spatially varying relationships in datasets [8].

The problem of local estimation considered in this study can be formulated as follows: we have a region and lignite quality values $\{y_\alpha\}, \alpha = 1,...,n,$ at n sampled locations. We aim to estimate the values $\{y_\beta\}, \beta = 1,...,N$ at unsampled locations. To model a system by a spatially weighted method, the identification data and the weights could be arranged as in the following matrices:

ISBN: 978-989-95079-6-8

$$\mathbf{X}=\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix},\ \mathbf{y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix},\ \mathbf{W}_i=\begin{bmatrix} w_{i1} & 0 & 0 & 0 \\ 0 & w_{i2} & 0 & 0 \\ 0 & 0 & w_{i3} & 0 \\ 0 & 0 & 0 & w_{iN} \end{bmatrix} \quad (1)$$

2.2 Geographically Weighted Regression (GWR)

An assumption of global regression is that the relationship under study is spatially constant, and thus, the relationships being measured are assumed to be 'stationary' over space. However, in most cases, the relationship varies in space. GWR, as a refinement to traditional regression methods, explicitly deals with the spatial non-stationary of empirical relationships. The method provides a weighting of information that is locally specific, and allows regression model parameters to vary in space [3].

Classical regression equation in matrix form can be given by:

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\varepsilon \quad (2)$$

where the vector of parameters to be estimated, β, is constant over space and is estimated by

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}. \quad (3)$$

The local estimation of the parameters with GWR is as given by (3), but the difference is that the observations used in GWR are weighted in accordance with their distance from the kernel centre. Fig. 1 indicates a spatial kernel in GWR. The parameters for GWR may be estimated by solving

$$\hat{\beta}(u_i,v_i)=\left[\mathbf{X}^T\mathbf{W}(u_i,v_{i,})\mathbf{X}\right]^{-1}\mathbf{X}^T\mathbf{W}(u_i,v_i)\mathbf{Y}, \quad (4)$$

where $\hat{\beta}$ represents an estimate of β, and $\mathbf{W}(u_i,v_i)$ is an n by n matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of each of the n observed data for regression point i [9]. There are many weighting schemes which express w_{ij} as a continuous function of distance between i and j, d_{ij}. In this study, the following Gaussian function has been used.

$$w_{ij}=\exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right] \quad (5)$$

where d_{ij} is the Euclidean distance between the location of measurement i and the centre of the kernel and b is the bandwidth of the kernel.

2.3 Fuzzy Clustering-Based Least Squares

A class of fuzzy clustering algorithms can be used to approximate a set of data by local linear models. Each of these models is represented by a fuzzy subset in the data set available for identification [10]. Most analytical fuzzy clustering algorithms are based on optimization of the basic c-means objective function, or some modification of it.

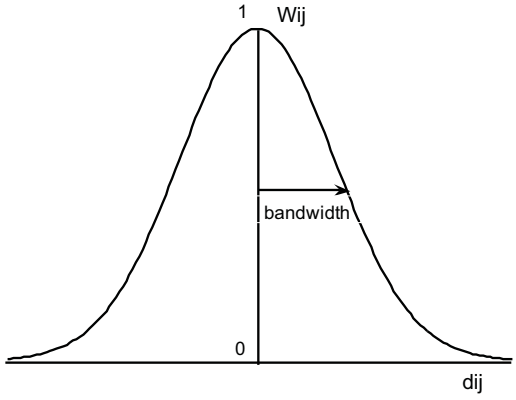


Figure 1: A Gaussian kernel in GWR.

In the present study, we used a general structure which is the fuzzy c-means functional [11] for constructing the weighted least squares model from the fuzzy partitions.

In the FCBLS model, the output parameters for the ith cluster, a_i and b_i are connected by a single parameter vector $\boldsymbol{\theta}_\mathbf{i}$ as follows:

$$\boldsymbol{\theta}_\mathbf{i}=\left[\mathbf{a}_i^T,b_i\right]^T. \quad (6)$$

Appending a unitary column to $\mathbf{X}$ gives the extended regressor matrix $\mathbf{X_e}$:

$$\mathbf{X_e}=[\mathbf{X},\mathbf{1}]. \quad (7)$$

Assuming that each cluster represents a local linear model of the system, the consequent parameter vectors $\boldsymbol{\theta}_\mathbf{i}$, $i=1,2,...,c$, can be estimated independently by the least-squares method. The membership degrees of the fuzzy partition serve as the weights expressing the relevance of the data pair x_k, y_k to that local model [10]. If the columns of $\mathbf{X}_e$ are linearly independent, then

$$\boldsymbol{\theta}_i=\left[\mathbf{X}_e^T\mathbf{W}_i\mathbf{X}_e\right]^{-1}\mathbf{X}_e^T\mathbf{W}_i\mathbf{y} \quad (8)$$

is the least-squares solution of $\mathbf{y}=\mathbf{X_e}\boldsymbol{\theta}+\boldsymbol{\varepsilon}$ where the kth data pair (x_k,y_k) is weighted by w_{ik}. The parameters $\mathbf{a}_i$ and b_i are given by:

$$\mathbf{a}_i=\left\lfloor\theta_1,\theta_2,...,\theta_p\right\rfloor,\ b_i=\theta_{p+1}. \quad (9)$$

3 Implementation

3.1 Data set

The study area, the Sivas-Kalburcayiri field, is one of the most important lignite reserves in Turkey [6]. Lignite seams in this field are used to feed coal to a power plant. In this study, one of the lignite quality parameters, Calorific Value (CV), has been estimated from the independent parameters which are spatial coordinates (x,y), ash and sulphur contents.

ISBN: 978-989-95079-6-8

The locations of the 67 records of the field have been employed in the model. The data set was divided into two subsets randomly: the training set (60%: 40 records) and the validation set (40%:27 records), respectively. For the analyses, data conditioning is necessary. In the present study, scaling was carried out by the local metric (L-metric) rescaling, in which the minimum and maximum values of x_{ij} for each j are respectively mapped onto zero and one respectively [7],

$$x_{ij}^{L} = \frac{x_{ij} - \min_j(x_{ij})}{\max_j(x_{ij}) - \min_j(x_{ij})}. \quad (10)$$

3.2. Local Models

GWR model

For the local analysis of the lignite parameters, GWR analyses were undertaken using a fixed spatial kernel with Gaussian function in which the bandwidth was determined by minimization of the Akaike Information Criterion (AIC) [12].

In the literature, various approaches have been used for ascertaining an optimal bandwidth. A method of deriving the bandwidth which provides a trade-off between goodness-of-fit and degrees of freedom is to minimize the AIC [12], used in this study. In [9], the AIC has been modified for GWR as follows:

$$AIC_c = 2n\log_e\left(\hat{\sigma}\right) + n\log_e\left(2\prod\right) + n\left\{\frac{n + tr(\mathbf{S})}{n - 2 - tr(\mathbf{S})}\right\} \quad (11)$$

where n is the sample size, $\hat{\sigma}$ is the estimated standard deviation of the error term, and tr($\mathbf{S}$) denotes the trace of the hat matrix $\mathbf{S}$ $\left(\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}\right)$. The matrix S maps $\hat{\mathbf{y}}$ on to $\mathbf{y}$.

In this application, the GWR model was fitted using a computer software program, *SAM 3.0.* [13]. In addition, the fixed bandwidth value has been determined from AIC using '*spgwr*' package in *R* [14]. Table 1 summarizes the model fitting statistics for training data. Similarly, the test statistics which were obtained from the fixed bandwidth are given in Table 2. In these tables, r^2 relates to observations against their estimates and OLS denotes the Ordinary Least Squares optimization.

Table 1. GWR fitting statistics for training data.

Spatial function:	Gaussian
Fixed Bandwidth (h):	0.191 distance units.
Number of Locations to Fit Model (n):	40
Coefficient of Determination (r^2):	0.811 (OLS: 0.785)

Table 2. GWR fitting statistics for testing data.

Spatial function:	Gaussian
Fixed Bandwidth (h):	0.191 distance units.
Number of Locations to Fit Model (n):	27
Coefficient of Determination (r^2):	0.718 (OLS: 0.655)

FCBLS model

To analyse the spatial relationships by clustering based least-squared optimization, firstly a clustering operation was conducted. The optimal number of clusters was determined experimentally using an index method which has been presented in [6]. Because the data variability is crucial in spatial data modelling, the method aims to reproduce variability of the sample data in the CV of cluster centres with a minimum number of clusters as follows:

$$\text{Minimize } n_c \quad under \quad Std\left[CV(x)\right] \approx Std\left[CV(c)\right] \quad (12)$$

where n_c is the optimal number of clusters, *Std* is the standard deviation of CV values. In this application, the appropriate numbers of clusters was determined as two. The cluster centres are depicted in Fig. 2.

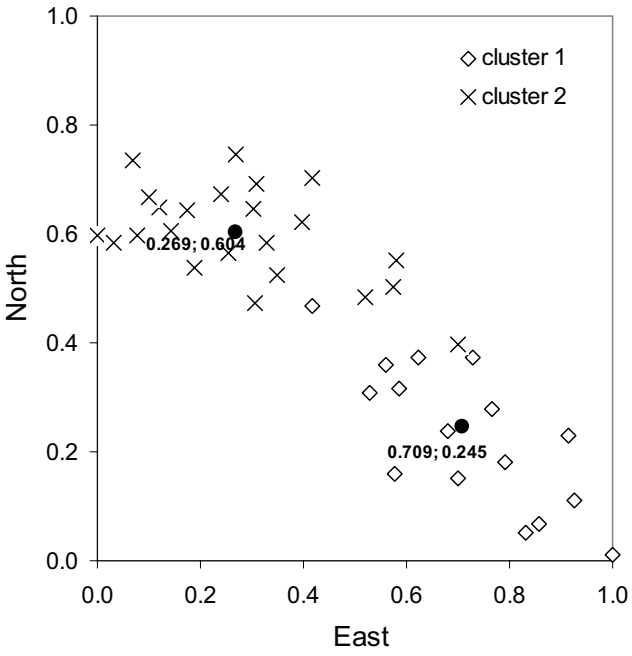


Figure 2: Data and cluster centres.

By using the information taken from the clustering, Gaussian type membership (weighting) functions were adopted. Fig.3 shows the input memberships considered in the model. To optimize the least square systems, weights taken from the Gaussian memberships have been used. Note that the Gaussian functions developed in the model used the same bandwidths derived from AIC. The consequent parameters are summarized in Table 3.

ISBN: 978-989-95079-6-8

Table 3: Consequent parameters for FCBLS model.

Cluster	*Ash*	*Sulphur*	*Constant*
Cluster 1	-0.045	0.957	0.031
Cluster 2	-0.403	0.532	0.0247

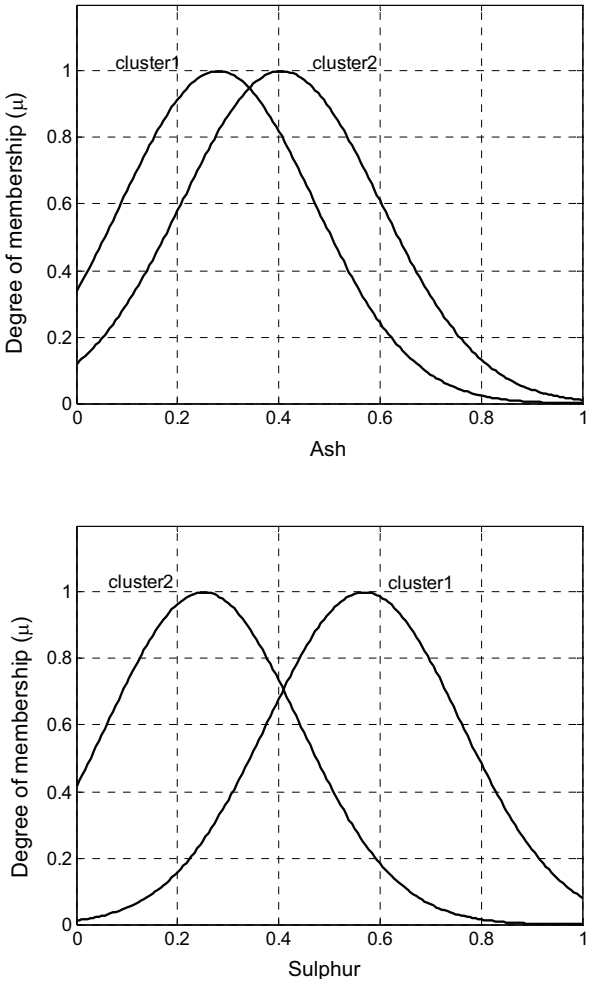


Figure 3: Gaussian memberships.

4 Results and conclusions

To assess the performance of the local models detailed in section 3.2, we have plotted the observed calorific values against the estimated calorific values. Fig. 4 illustrates the GWR model results together with the cross-correlations between estimated and observed data both for the training set and the validation set, respectively. Similarly, the results of the FCBLS model are depicted in Fig. 5.

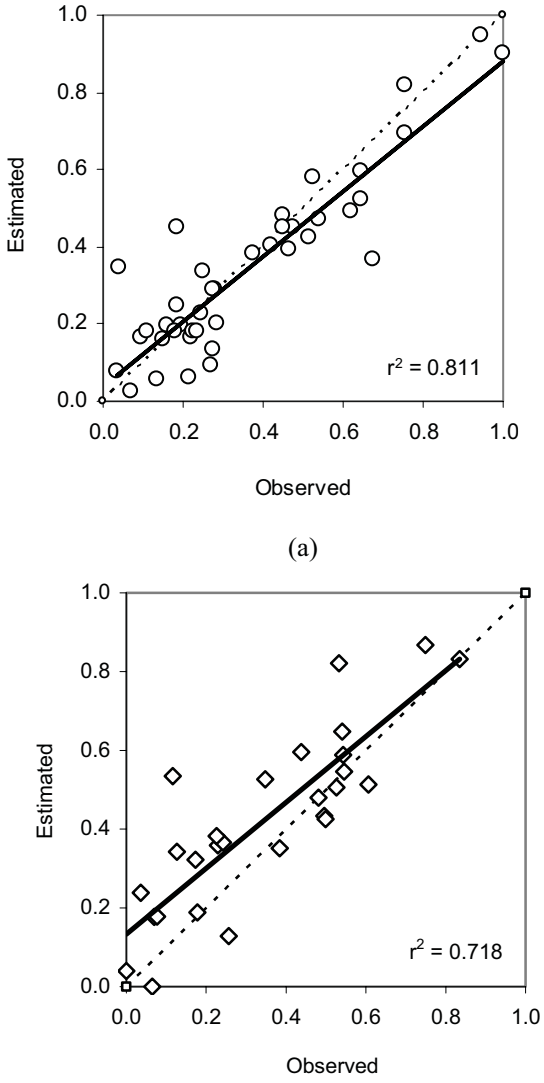
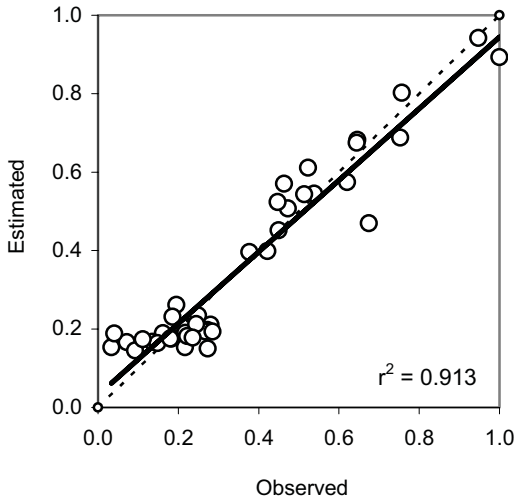


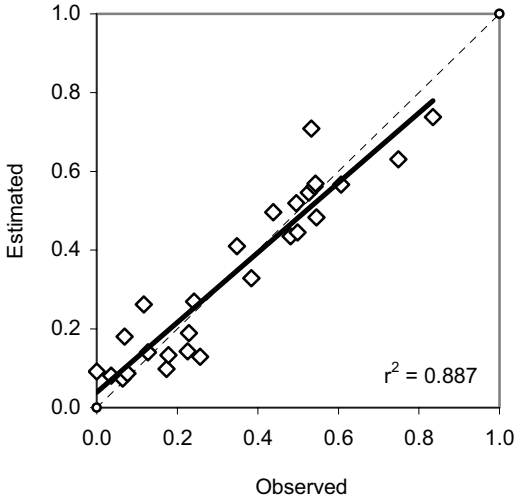
Figure 4: Scatter plot for GWR model: (a) training data and (b) testing data.

The large determination coefficient (r^2) shows that the model has good estimation capability. As observed from the determination coefficient, the FCBLS method outperforms the method based on GWR. In addition to r^2, the performance of the models has been assessed using the performance index, namely, the Variance Accounted For (VAF). In multivariate analysis, the measure of VAF plays a central role. Table 4 gives the VAF values.

ISBN: 978-989-95079-6-8



(a)



(b)

Figure 5: Scatter plot for FCBLS model: (a) training data and (b) testing data.

Table 4: VAF measures for the models.

Model	Training	Testing
GWR	82.06	72.57
FCBLS	91.30	88.73

In addition to performance indices, the estimates can be presented in map form. Figs. 6-8 show the observed values, GWR estimations and FCBLS estimations, respectively. Based on the performance evaluations and the maps, it can be argued that the fuzzy clustering based regression model (FCBLS) outperforms the geographically weighted regression model in this case.

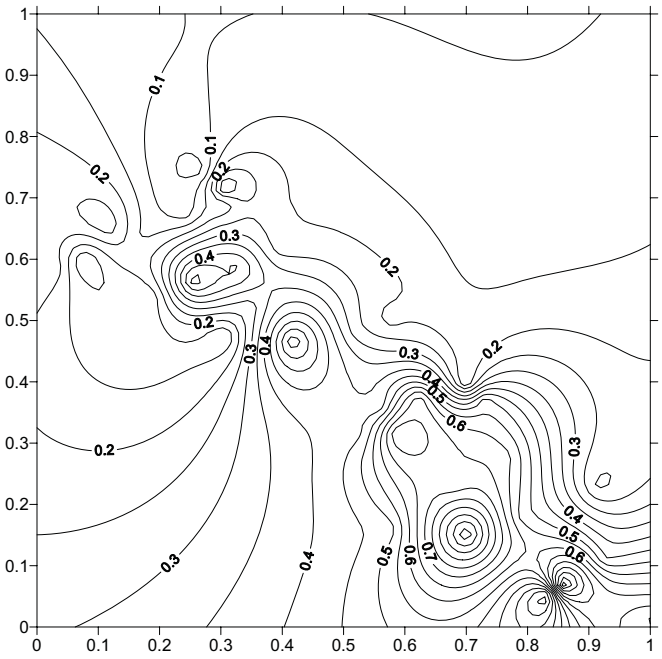


Figure 6: Contour map for observed data.

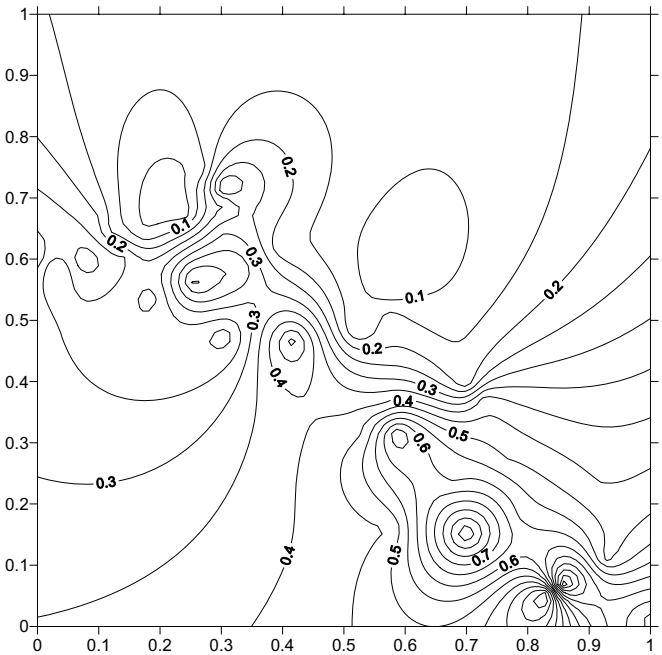


Figure 7: Contour map for the GWR model.

ISBN: 978-989-95079-6-8

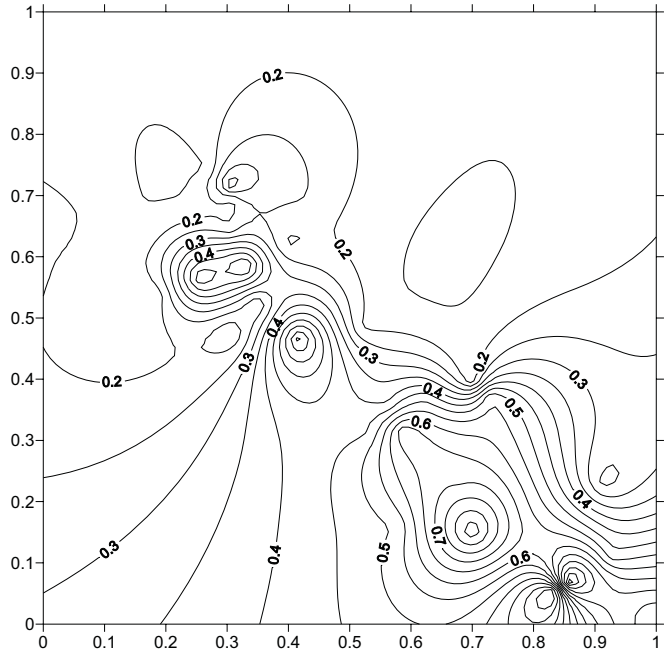


Figure 8: Contour map for FCBLS model.

Acknowledgements

This paper is part of the project (108M393) which is granted by Scientific and Technological Research Council of Turkey (TUBITAK).

References

[1] X. Gao, Y. Asami, CJ. F. Chung. An empirical evaluation of spatial regression models, *Computers &Geosciences*, 32:1040-1051,2006.

[2] C.D. Lloyd, *Local Models for Spatial Analysis*, Boca Raton: CRC Press, 2006.

[3] A.S. Fotheringham, M.E. Charlton, C. Brunsdon, Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis, *Environment and Planning A*, 30: 1905-1927, 1998.

[4] E. S. Lee, Neuro-Fuzzy estimation in spatial statistics, *J Mathematical Analysis and Applications*, 249:221-231, 2000.

[5] P. Wong, F. Aminzadeh, M. Nikravesh, *Soft Computing for Reservoir Characterization and Modeling*, Heidelberg:Physica-Verlag, 2001.

[6] B. Tutmez, A.E. Tercan, U. Kaymak. Fuzzy modeling for reserve estimation based on spatial variability, *Mathematical Geology*, 39(1):87-111, 2007.

[7] A.S. Fotheringham, C. Brunsdon, M.E. Charlton, *Qualitative Geography: Perspectives on Spatial Data Analysis*, London: SAGE Publications, 2000.

[8] C.D. Lloyd and I. Shuttleworth, Analysing commuting using local regression techniques: scale, sensitivity, and geographical patterning, *Environment and Planning A*, 37:81-113, 2005.

[9] A.S. Fotheringham, C. Brunsdon, M.E. Charlton, *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships*, Chichester:Wiley, 2002.

[10] R. Babuska, *Fuzzy Modeling for Control*, USA:Kluwer Academic, 1998.

[11] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy c-means clustering algorithm, Computers&Geosciences, 10(2-3):191-203, 1984.

[12] H. Akaike, Information theory and an extension of the maximum likelihood principle, In B. Petrov, F. Csaki (eds) *2nd Symposium on Information Theory*, Budapest, 267-281, 1973.

[13] Thiago F. L.V.B. Rangel, Jose A.F. Diniz-Filho, L.M. Bini, Towards and integrated computational tool for spatial analysis in macroecology and biogeography, *Global Ecology and Biogeography*, 15:321-327, 2006.

[14] R.S. Bivand, E.J. Pebesma, V. Gomez-Rubio, Applied Spatial Data Analysis with R, New York: Springer, 2008.

Fuzzy Integrals over Complete Residuated Lattices

Antonín Dvořák, Michal Holčapek

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava,
30. dubna 22, 701 03 Ostrava 1, Czech Republic
Email: {antonin.dvorak, michal.holcapek}@osu.cz

Abstract— *The aim of this paper is to introduce two new types of fuzzy integrals, namely, $\otimes$-fuzzy integral and $\rightarrow$-fuzzy integral, where $\otimes$ and $\rightarrow$ are the multiplication and residuum of a complete residuated lattice, respectively. The first integral is based on a fuzzy measure of* **L***-fuzzy sets and the second one on a complementary fuzzy measure of* **L***-fuzzy sets, where* **L** *is a complete residuated lattice. Some of their properties and a relation to the fuzzy (Sugeno) integral are investigated.*

Keywords— fuzzy measure, fuzzy integral, fuzzy quantifier.

1 Introduction

Let us consider two time series $t_1 = (t_{1k})_{k\in T}$, $t_2 = (t_{2k})_{k\in T}$ displayed on Fig. 1 and suppose that our goal is to compare them and to find the "better" time series, where a greater value of time series at some time point k means a better value. To

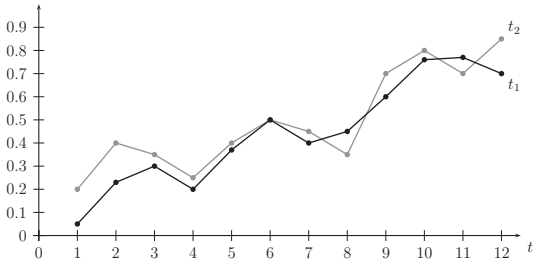


Figure 1: Time series.

solve this task it is reasonable, firstly, to determine degrees of truth saying how much formulas

$$\varphi(i,j,k) := \text{"}\textit{the value } t_{ik} \textit{ is better than the value } t_{jk}\text{"}$$

are true, where, first, $i = 1$ and $j = 2$, and then $i = 2$ and $j = 1$. Obviously, the degrees of truth of formula $\varphi(i,j,k)$ may be modeled by a fuzzy relation $R : [0,1]^2 \to [0,1]$, e.g.

$$R(a,b) = \max(a - b, 0), \tag{1}$$

where $R(a,b)$ determines the degree that a is better than b. Note that $R(a,b) = \neg(a \to b)$, where $a \to b$ is the operation of residuum and $\neg$ is the operation of negation in the Łukasiewicz algebra (see Example 2.1 and consider $\neg a = 1 - a$). If we know how one value of time series is better than the other one for each time k, then we have to solve a task how to aggregate the obtained values to find a degree in which one times series is better than the second one. One of the natural approaches could be to evaluate the following formula of the second order logic

$$\varphi(i,j) := (\exists Y \in \mathcal{P}(T) \setminus \{\emptyset\})(\forall k \in Y)(\varphi(i,j,k) \& \psi(Y)),$$

where $\mathcal{P}(T)$ is the power set of T and

$$\psi(Y) := \textit{"the set } Y \textit{ is a big subset of } T\textit{"}.$$

Note that formula $\varphi(i,j)$ may be defined using the formula $(Qk \in T)\varphi(i,j,k)$ with a fuzzy quantifier Q like *for nearly all* or *many* etc. Some types of fuzzy quantifiers could be determined by the interpretations of the formula ψ (cf. [1] and also see [2]). For example, if we consider the Łukasiewicz algebra as the structure of truth values for our logic, the truth value of the formula $\varphi(i,j,k)$ is defined as $R(t_{ik}, t_{jk})$ from (1), i.e., as the degree that t_{ik} is better than t_{jk}, and the truth value of the formula $\psi(Y)$ is interpreted by the value $\mu(Y)$, where $\mu : \mathcal{P}(T) \to [0,1]$ is a fuzzy measure (see Definition 3.2), then the evaluation of the formula $\varphi(i,j)$ is given by

$$||\varphi(i,j)|| = \bigvee_{Y\in\mathcal{P}(T)\setminus\{\emptyset\}} \bigwedge_{k\in Y} (R(t_{ik}, t_{jk}) \otimes \mu(Y)), \tag{2}$$

where $\otimes$ is the operation of multiplication in the Łukasiewicz algebra (see Example 2.1) interpreting the logical connective $\&$. Finally, we can conclude that, for example, the time series t_1 is better than t_2, if $||\varphi(1,2)|| > ||\varphi(2,1)||$.

Let us define a mapping $I_\mu : \mathcal{F}(T) \to [0,1]$ by

$$I_\mu(A) = \bigvee_{Y\in\mathcal{P}(T)\setminus\{\emptyset\}} \bigwedge_{k\in Y} (A(k) \otimes \mu(Y)), \tag{3}$$

where $\mathcal{F}(T)$ denotes the set of all fuzzy sets over T and $\mu : \mathcal{P}(T) \to [0,1]$ is a fuzzy measure. One could simply verify that if $c \in [0,1]$ is a constant and $A(k) = c$ for any $k \in T$, then $I_\mu(A) = c$, and if $A(k) \le B(k)$, then $I_\mu(A) \le I_\mu(B)$. Hence, I_μ is a fuzzy measure (in the sense of Definition 3.2) which could be understood, according to Mesiar [3], as an example of (fuzzy) integral. Putting $A(k) = R(t_{ik}, t_{jk})$ for any $k \in T$, we can write $||\varphi(i,j)|| = I_\mu(A)$ or also $||(Qk \in T)\varphi(i,j,k)|| = I_\mu(A)$ and, hence, we can see that fuzzy integrals may be used to model some types of fuzzy quantifiers. This idea is not new (see e.g. [4] or [5]), but some disadvantage of proposed approaches is that fuzzy quantifiers are defined as mappings from the set of all measurable (fuzzy) sets over a set M to $[0,1]$, although, the classical definition introduces fuzzy quantifiers as mappings from the set of all (fuzzy) sets over a set M to $[0,1]$ (see e.g. [6] or [7]). It will be clear that this drawback vanishes when fuzzy quantifiers are modeled by fuzzy integrals in a form similar to (3).

The aim of this contribution is to generalize the fuzzy integral defined in (3), namely, to introduce a $\otimes$-fuzzy integral that could be used for modeling fuzzy quantifiers like *all*, *some*,

ISBN: 978-989-95079-6-8

for nearly all, or *many*, etc. and then to introduce a $\rightarrow$-fuzzy integral that could be used for modeling quantifiers like *no*, *not all*, etc. Both types of fuzzy integrals are defined over a complete residuated lattice and it could be shown than, in general, one type of fuzzy integral cannot be expressed by another one. Nevertheless, if the structure of truth values is complete MV-algebra, then it is possible to define the $\rightarrow$-fuzzy integral from the $\otimes$-fuzzy integral using negation (see Theorem 4.14). Moreover, it is surprising that we are able to show that the well-known Sugeno integral [8] is, under certain conditions, a special case of our fuzzy integral (see Theorem 4.7).

2 Preliminaries

2.1 Structures of truth values

Let us suppose that the structure of truth values is a *complete residuated lattice* (see e.g. [9]), i.e., an algebra $\mathbf{L} = \langle L, \wedge, \vee, \rightarrow, \otimes, \bot, \top \rangle$ with four binary operations and two constants such that $\langle L, \wedge, \vee, \bot, \top \rangle$ is a complete lattice, where $\bot$ is the least element and $\top$ is the greatest element of L, respectively, $\langle L, \otimes, \top \rangle$ is a commutative monoid (i.e., $\otimes$ is associative, commutative and the identity $a \otimes \top = a$ holds for any $a \in L$) and the adjointness property is satisfied, i.e.,

$$a \leq b \rightarrow c \quad \text{iff} \quad a \otimes b \leq c \tag{4}$$

holds for each $a, b, c \in L$, where $\leq$ denotes the corresponding lattice ordering. The operations $\otimes$ and $\rightarrow$ are usually called the multiplication and residuum, respectively. A residuated lattice is *divisible*, if $a \otimes (a \rightarrow b) = a \wedge b$ holds for arbitrary $a, b \in L$, and satisfies the *law of double negation*, if $(a \rightarrow \bot) \rightarrow \bot = a$ holds for any $a \in L$. A divisible residuated lattice satisfying the law of double negation is called an *MV-algebra*. For other information about residuated lattices we refer to [9].

Example 2.1. It is easy to prove (see e.g. [10]) that the algebra

$$\mathbf{L}_{\mathrm{T}} = \langle [0,1], \min, \max, T, \rightarrow_T, 0, 1 \rangle,$$

where T is a left continuous t-norm and $a \rightarrow_T b = \bigvee \{c \in [0,1] \mid T(a,c) \leq b\}$ defines the residuum, is a complete residuated lattice. Moreover, if T is the Łukasiewicz t-norm, i.e., $T(a,b) = \max(a+b-1, 0)$ for all $a, b \in [0,1]$, then $\mathbf{L}_{\mathrm{T}}$ is a complete MV-algebra called a *Łukasiewicz algebra* (on $[0,1]$). One checks easily that $a \rightarrow_T b = \max(1-a+b, 0)$ is the residuum in the Łukasiewicz algebra.

Example 2.2. One checks easily that

$$\mathbf{L}_{[0,\infty]} = \langle [0,\infty], \min, \max, \rightarrow, 0, \infty \rangle,$$

where $\otimes = \min$ and

$$a \rightarrow b = \begin{cases} b, & \text{if } b < a, \\ \infty, & \text{otherwise,} \end{cases} \tag{5}$$

is a complete residuated lattice. Note that $\mathbf{L}_{[0,\infty]}$ is a special example of more general residuated lattice called a *Heyting algebra*.[1]

Let us define the following additional operations for all $a, b \in L$:

$$a \leftrightarrow b = (a \rightarrow b) \wedge (b \rightarrow a) \qquad (\textit{biresiduum})$$
$$\neg a = a \rightarrow \bot. \qquad (\textit{negation})$$

[1] A Heyting algebra is a residuated lattice with $\otimes = \wedge$.

2.2 L-fuzzy sets

Let $\mathbf{L}$ be a complete residuated lattice and M be a universe of discourse. A mapping $A : M \rightarrow L$ is called an $\mathbf{L}$-*fuzzy set on* M. A value $A(m)$ is called a *membership degree of* m *in the* $\mathbf{L}$-*fuzzy set* A. The set of all $\mathbf{L}$-fuzzy sets on M is denoted by $\mathcal{F}_{\mathbf{L}}(M)$. An $\mathbf{L}$-fuzzy set A on M is called *crisp*, if there is a subset X of M such that $A = 1_X$, where 1_X denotes the characteristic function of X. Particularly, $1_\emptyset$ denotes the empty $\mathbf{L}$-fuzzy set on M, i.e. $1_\emptyset(m) = \bot$ for any $m \in M$. The set of all crisp $\mathbf{L}$-fuzzy sets on M is denoted by $\mathcal{P}_{\mathbf{L}}(M)$. An $\mathbf{L}$-fuzzy set A is *constant*, if there is $c \in L$ such that $A(m) = c$ for any $m \in M$. For simplicity, a constant $\mathbf{L}$-fuzzy set is denoted by the corresponding element of L, e.g., a, b, c.[2] Let us denote $\mathrm{Supp}(A) = \{m \mid m \in M \;\&\; A(m) > \bot\}$ and $\mathrm{core}(A) = \{m \mid m \in M \;\&\; A(m) = \top\}$ the *support* and *core* of an $\mathbf{L}$-fuzzy set A, respectively. Obviously, $\mathrm{Supp}(1_X) = \mathrm{core}(1_X) = X$ for any crisp $\mathbf{L}$-fuzzy set. An $\mathbf{L}$-fuzzy set A is called *normal*, if $\mathrm{core}(A) \neq \emptyset$.

Let $\{A_i \mid i \in I\}$ be a non-empty family of $\mathbf{L}$-fuzzy sets on M. Then the *union of* A_i is defined by

$$\left(\bigcup_{i \in I} A_i\right)(m) = \bigvee_{i \in I} A_i(m) \tag{6}$$

for any $m \in M$ and the *intersection of* A_i is defined by

$$\left(\bigcap_{i \in I} A_i\right)(m) = \bigwedge_{i \in I} A_i(m) \tag{7}$$

for any $m \in M$. Let A be an $\mathbf{L}$-fuzzy set on M. The *complement* of A is an $\mathbf{L}$-fuzzy set $\overline{A}$ on M defined by $\overline{A}(m) = \neg A(m)$ for any $m \in M$. Finally, an extension of the operations $\otimes$ and $\rightarrow$ on L to the operations on $\mathcal{F}_{\mathbf{L}}(M)$ is given by

$$(A \otimes B)(m) = A(m) \otimes B(m) \tag{8}$$
$$(A \rightarrow B)(m) = A(m) \rightarrow B(m) \tag{9}$$

for any $A, B \in \mathcal{F}_{\mathbf{L}}(M)$ and $m \in M$, respectively. The following theorem shows the well-known relation between the operations of the union and intersection of sets which also holds for $\mathbf{L}$-fuzzy sets, if we restrict ourselves to a special class of complete residuated lattices.

Theorem 2.1. *Let* $\mathbf{L}$ *be a complete residuated lattice satisfying the law of double negation and* $\{A_i \mid i \in I\}$ *be a non-empty family of* $\mathbf{L}$*-fuzzy sets on* M*. Then*

$$\bigcup_{i \in I} A_i = \overline{\bigcap_{i \in I} \overline{A_i}} \quad \textit{and} \quad \bigcap_{i \in I} A_i = \overline{\bigcup_{i \in I} \overline{A_i}}. \tag{10}$$

We say that an $\mathbf{L}$-fuzzy set A is *less than or equal to* an $\mathbf{L}$-fuzzy set B and denote by $A \subseteq B$, if, for any $m \in M$, we have $A(m) \leq B(m)$. Let $f : M \rightarrow M'$ be a mapping. Then $f^{\rightarrow}(A)(m) = \bigvee_{m' \in f^{-1}(m)} A(m')$ defines a mapping $f^{\rightarrow} : \mathcal{F}_{\mathbf{L}}(M) \rightarrow \mathcal{F}_{\mathbf{L}}(M')$. Obviously, if f is a bijective mapping, then $f^{\rightarrow}(A)(f(m)) = A(m)$ for any $m \in M$.

[2] We suppose that the meaning of this symbol will be unmistakable from the context, that is, it should be clear when an element of L is considered and when a constant $\mathbf{L}$-fuzzy set is assumed.

ISBN: 978-989-95079-6-8

3 Fuzzy measures

In this section, we will introduce a notion of fuzzy measure and complementary fuzzy measure of **L**-fuzzy sets. More information about fuzzy measures could be found in [11, 12].

For our considerations we will consider algebras of **L**-fuzzy sets as a base for defining fuzzy measures of **L**-fuzzy sets.

Definition 3.1 ([11])**.** Let M be a non-empty universe of discourse. A subset $\mathcal{M}$ of $\mathcal{F}_{\mathbf{L}}(M)$ is an *algebra of* **L***-fuzzy sets on* M, if the following conditions are satisfied

(i) $1_\emptyset, 1_M \in \mathcal{M}$,

(ii) if $A \in \mathcal{M}$, then $\overline{A} \in \mathcal{M}$,

(iii) if $A, B \in \mathcal{M}$, then $A \cup B \in \mathcal{M}$.

A couple $(M, \mathcal{M})$ is called a *fuzzy measurable space*, if $\mathcal{M}$ is an algebra of **L**-fuzzy sets on M.

Example 3.1. The sets $\{1_\emptyset, 1_M\}$, $\mathcal{P}_{\mathbf{L}}(M)$, σ-algebras on M, or $\mathcal{F}_{\mathbf{L}}(M)$ are algebras of **L**-fuzzy sets on M.

Example 3.2. Let us say that an **L**-fuzzy set A on M is a *simple* **L***-fuzzy set* on M, if there exists a family of sets $\{M_i \mid i = 1, \ldots, n\}$ such that $\bigcup_{i=1}^n M_i = M$, $M_i \neq M_j$ for any $i \neq j$ and $A(m) = A(m')$ holds for each $m, m' \in M_i$, where $i = 1, \ldots, n$. Obviously, the set of all simple **L**-fuzzy sets on M is an algebra of **L**-fuzzy sets on M.

Example 3.3. Let **L** be the Łukasiewicz algebra on $[0, 1]$ (see Example 2.1) and $M = [0, 1]$. Then the set of all continuous mappings $A : [0, 1] \to [0, 1]$ is an algebra of **L**-fuzzy sets in M.[3]

Let us introduce the concepts of fuzzy measure and complementary fuzzy measure as follows. The first definition is a modification of the definition of a normed fuzzy measure with respect to truth values (see e.g. [12, 13]).

Definition 3.2. Let $(M, \mathcal{M})$ be a fuzzy measurable space. A mapping $\mu : \mathcal{M} \to L$ is called a *fuzzy measure* on $(M, \mathcal{M})$, if

(i) $\mu(1_\emptyset) = \bot$ and $\mu(1_M) = \top$,

(ii) if $A, B \in \mathcal{M}$ such that $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

A triplet $(M, \mathcal{M}, \mu)$ is called a *fuzzy measure space*, if $(M, \mathcal{M})$ is a fuzzy measurable space and μ is a fuzzy measure on $(M, \mathcal{M})$.

Definition 3.3. Let $(M, \mathcal{M})$ be a fuzzy measurable space. A mapping $\nu : \mathcal{M} \to L$ is called a *complementary fuzzy measure* on $(M, \mathcal{M})$, if

(i) $\nu(1_\emptyset) = \top$ and $\nu(1_M) = \bot$,

(ii) if $A, B \in \mathcal{M}$ such that $A \leq B$, then $\nu(A) \geq \nu(B)$.

A triplet $(M, \mathcal{M}, \nu)$ is called a *complementary fuzzy measure space*, if $(M, \mathcal{M})$ is a fuzzy measurable space and ν is a complementary fuzzy measure on $(M, \mathcal{M})$.

[3]Note that the set of all continuous mappings need not be an algebra of **L**-fuzzy sets for other residuated lattices determined by left continuous T-norms, because the negation is not a continuous mapping in general.

Example 3.4. Let $(M, \mathcal{M})$ be the fuzzy measurable space of all continuous mappings from Example 3.3. It is easy to see that

$$\mu(A) = \int_0^1 A(m)\, dm,$$

where $\int_0^1 A(m)\, dm$ denotes the Riemann integral, defines a fuzzy measure on $(M, \mathcal{M})$.

Example 3.5. Let **L** be a complete residuated lattice with the support $[0, 1]$ and $\mathbb{N}$ be the set of natural numbers with 0. For any non-empty countable (i.e., finite or denumerable) universe M, injective mapping $f : M \to \mathbb{N}$, $n \in \mathbb{N}$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$, denote

$$A_{f,n}(m) = \begin{cases} A(m), & \text{if } f(m) \leq n; \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Further, for any injective mapping $f : M \to \mathbb{N}$ and $n \in \mathbb{N}$, define $\mu_{f,n} : \mathcal{F}_{\mathbf{L}}(M) \to [0, 1]$ as follows

$$\mu_{f,n}(A) = \frac{\sum_{m \in \mathrm{Supp}(A_{f,n})} A_{f,n}(m)}{|\mathrm{Supp}(1_{M_{f,n}})|} \tag{12}$$

and, finally, define $\underline{\mu}_f, \overline{\mu}_f : \mathcal{F}_{\mathbf{L}}(M) \to [0, 1]$ as follows

$$\underline{\mu}_f = \liminf_{n \to \infty} \mu_{f,n}(A), \tag{13}$$

$$\overline{\mu}_f = \limsup_{n \to \infty} \mu_{f,n}(A). \tag{14}$$

It is easy to see that $\mu_{f,n}$, $\underline{\mu}_f$ and $\overline{\mu}_f$ are fuzzy measures on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by an injective mapping f.[4] If, for example, $M = \mathbb{N}$ and $f = \mathrm{id}$, then $\underline{\mu}_f(A) = \overline{\mu}_f(A) = \bot$ for any **L**-fuzzy set with finite universe. For the set of all even or odd numbers, both fuzzy measures give $\frac{1}{2}$ and, for the set of all prime numbers, we obtain 0.

If M is finite, then $\underline{\mu}_f = \underline{\mu}_g = \overline{\mu}_f = \overline{\mu}_g$ for any injective mappings $f, g : M \to \mathbb{N}$ and

$$\underline{\mu}_f(A) = \overline{\mu}_f(A) = \frac{\sum_{m \in M} A(m)}{|M|}. \tag{15}$$

Hence, it is easy to see that $\mu_f(A) = \mu_f(h^{\rightarrow}(A))$ holds for any non-empty finite universe M, $A \in \mathcal{F}_{\mathbf{L}}(M)$, injective mapping $f : M \to \mathbb{N}$ and bijective mapping $h : M \to M$. Unfortunately, this equality fails for denumerable universes in general. In fact, consider $M = \mathbb{N}$, $f = \mathrm{id}$ and a bijective mapping $h : \mathbb{N} \to \mathbb{N}$ such that the image of all even numbers is the set of prime numbers. Then both fuzzy measures give $\frac{1}{2}$ for the set of all even numbers, however, 0 for the set of all prime numbers.

Example 3.6. Let μ_f be one of the fuzzy measures on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by f defined in (13) and (14). If $h : [0, 1] \to [0, 1]$ is a non-decreasing mapping with $h(0) = 0$ and $h(1) = 1$, then $h \circ \mu_f$ is a fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by μ_f and h. If $h : [0, 1] \to [0, 1]$ is a non-increasing mapping with $h(0) = 1$ and $h(1) = 0$, then $h \circ \mu_f$ is a complementary fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by μ_f and h.

[4]Note that $\underline{\mu}_f$ and $\overline{\mu}_f$ could be understood as a generalization of lower and upper weighted densities well known in the number theory which are examples of so-called lower and upper asymptotic fuzzy measures (see [14]).

ISBN: 978-989-95079-6-8

Theorem 3.1. *Let $(M, \mathcal{M})$ be a fuzzy measurable space. If μ (ν) is a fuzzy measure (a complementary fuzzy measure) on $(M, \mathcal{M})$, then $\nu'(A) = \neg\mu(A)$ ($\mu'(A) = \neg\nu(A)$) defines a complementary fuzzy measure (a fuzzy measure) on $(M, \mathcal{M})$.*

Definition 3.4. Let $(M, \mathcal{M})$ be a fuzzy measurable space and $X \in \mathcal{F}_{\mathbf{L}}(M)$. We say that X is $\mathcal{M}$*-fuzzy measurable*, if $X \in \mathcal{M}$.

Let $(M, \mathcal{M})$ be a fuzzy measurable space and $X \in \mathcal{F}_{\mathbf{L}}(M)$. Denote $\mathcal{M}_X$ the set of all $\mathcal{M}$-fuzzy measurable sets which are contained in X, i.e.,

$$\mathcal{M}_X = \{A \mid A \in \mathcal{M} \text{ and } A \subseteq X\}. \tag{16}$$

Note that $1_\emptyset \in \mathcal{M}_X$ for each $X \in \mathcal{F}_{\mathbf{L}}(M)$ and if X is $\mathcal{M}$-fuzzy measurable set, then also $X \in \mathcal{M}_X$. If $X = M$, then we will write only $\mathcal{M}$ instead of $\mathcal{M}_M$.

Theorem 3.2. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. A mapping $\mu^* : \mathcal{F}_{\mathbf{L}}(M) \to L$ defined by*

$$\mu^*(X) = \bigvee_{A \in \mathcal{M}_X} \mu(A) \tag{17}$$

is a fuzzy measure on the fuzzy measurable space $(M, \mathcal{F}_{\mathbf{L}}(M))$. We say that μ^* is the inner fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by μ.

Example 3.7. Let $(M, \mathcal{P}_{\mathbf{L}}(M), \mu)$ be an arbitrary fuzzy measurable space (recall that $\mathcal{P}_{\mathbf{L}}(M)$ is the power set of M). Then the inner fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$ is defined by

$$\mu^*(A) = \begin{cases} \mu(A'), & \text{if } 1_{\mathrm{core}(A)} = A', \\ \bot, & \text{otherwise.} \end{cases} \tag{18}$$

Thus all $\mathbf{L}$-fuzzy sets that are not normal have the inner fuzzy measure equal to $\bot$.

Example 3.8. Let $\mathbf{L}$ be the Łukasiewicz algebra on $[0, 1]$, $(M, \mathcal{M}, \mu)$ be the fuzzy measure space of continuous $\mathbf{L}$-fuzzy sets from Example 3.4. Then, for example, we have $\mu^*(1_{[a,b]}) = b - a$, however, $1_{[a,b]} \notin \mathcal{M}$.

Theorem 3.3. *Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space. A mapping $\nu^* : \mathcal{F}_{\mathbf{L}}(M) \to L$ defined by*

$$\nu^*(X) = \bigwedge_{A \in \mathcal{M}_X} \nu(A) \tag{19}$$

is a complementary fuzzy measure on the fuzzy measurable space $(M, \mathcal{F}_{\mathbf{L}}(M))$. We say that ν is the inner complementary fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by ν.

In the following part we will define an isomorphism between fuzzy measure spaces and then between complementary fuzzy measure spaces.

Definition 3.5. Let $(M, \mathcal{M})$ and $(M', \mathcal{M}')$ be fuzzy measurable spaces. We say that a mapping $g : \mathcal{M} \to \mathcal{M}'$ is an *isomorphism between $(M, \mathcal{M})$ and $(M', \mathcal{M}')$*, if

(i) g is a bijective mapping with $g(1_\emptyset) = 1_\emptyset$,

(ii) $g(A \cup B) = g(A) \cup g(B)$ and $g(\overline{A}) = \overline{g(A)}$ hold for any $A, B \in \mathcal{M}$,

(iii) there exists a bijective mapping $f : M \to M'$ with $A(m) = g(A)(f(m))$ for any $A \in \mathcal{M}$ and $m \in M$.

Theorem 3.4. *Let $(M, \mathcal{M})$, $(M', \mathcal{M}')$ be fuzzy measurable spaces and $g : \mathcal{M} \to \mathcal{M}'$ be a surjective mapping. Then g is an isomorphism between $(M, \mathcal{M})$ and $(M', \mathcal{M}')$ if and only if there exists a bijective mapping $f : M \to M'$ such that $g = f^{\to}$.*

Definition 3.6. Let $(M, \mathcal{M})$ and $(M', \mathcal{M}')$ be fuzzy measurable spaces. We say that a mapping $g : \mathcal{M} \to \mathcal{M}'$ is an *isomorphism between $(M, \mathcal{M}, \mu)$ and $(M', \mathcal{M}', \mu')$* (or *between $(M, \mathcal{M}, \nu)$ and $(M', \mathcal{M}', \nu')$*), if

(i) g is an isomorphism between $(M, \mathcal{M})$ and $(M', \mathcal{M}')$,

(ii) $\mu(A) = \mu'(g(A))$ (or $\nu(A) = \nu'(g(A))$) for any $A \in \mathcal{M}$.

If g is an isomorphism between fuzzy measure spaces $(M, \mathcal{M}, \mu)$ and $(M', \mathcal{M}', \mu')$ or between complementary fuzzy measure spaces $(M, \mathcal{M}, \nu)$ and $(M', \mathcal{M}', \nu')$, then we write $g(M, \mathcal{M}, \mu) = (M', \mathcal{M}', \mu')$ or $g(M, \mathcal{M}, \nu) = (M', \mathcal{M}', \nu')$, respectively.

Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. If $f : M \to M'$ is a bijective mapping, then $(M', f^{\to}(\mathcal{M}), \mu_{f^{\to}})$, where

$$\mu_{f^{\to}}(f^{\to}(A)) = \mu(A) \tag{20}$$

holds for any $A \in \mathcal{M}$, is a fuzzy measure space isomorphic to $(M, \mathcal{M}, \mu)$. A simple consequence of Theorem 3.4 is the fact that each fuzzy measure space $(M', \mathcal{M}', \mu')$ isomorphic to $(M, \mathcal{M}, \mu)$ has the form $(M', f^{\to}(\mathcal{M}), \mu_{f^{\to}})$ for a suitable bijective mapping $f : M \to M'$. Analogously, to every couple of isomorphic complementary fuzzy measure spaces $(M, \mathcal{M}, \nu)$ and $(M', \mathcal{M}', \nu')$ there is a bijective mapping $f : M \to M'$ such that $(M', f^{\to}(\mathcal{M}), \nu_{f^{\to}}) = (M', \mathcal{M}', \nu')$.

4 Fuzzy integrals

In this section, we will introduce two types of fuzzy integrals. The first one is a generalization of the formula (3) derived in Introduction. For more information about fuzzy integrals we refer to [11, 12].

4.1 $\otimes$-fuzzy integral

In this part, we will introduce a type of fuzzy integral that can be defined on an arbitrary fuzzy measure space $(M, \mathcal{M}, \mu)$. The form of this integral is motivated by our need to describe a class of models of $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1 \rangle$. In [2], we show that this class of models is bounded by the models of determiners *all* and *some*. Note that models of *all* and *some* are the same as the interpretations of quantifiers $\forall$ and $\exists$, respectively, in fuzzy logic (see e.g. [6, 7, 15, 4]).

Definition 4.1. Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and X be a $\mathcal{M}$-fuzzy measurable $\mathbf{L}$-fuzzy set. The $\otimes$*-fuzzy integral of A on X* is given by

$$\int_X^{\otimes} A \, d\mu = \bigvee_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \bigwedge_{m \in \mathrm{Supp}(Y)} (A(m) \otimes \mu(Y)). \tag{21}$$

If $X = 1_M$, then we write $\int^{\otimes} A \, d\mu$.

ISBN: 978-989-95079-6-8

Remark 4.1. It is easy to see that $\int^{\otimes}_{1_\emptyset} A\, d\mu = \bigvee \emptyset = \bot$ for any $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $\int^{\otimes}_X A\, d\mu \leq \int^{\otimes}_Y A\, d\mu$, whenever $X \subseteq Y$. Since $\int^{\otimes}_{1_M} A\, d\mu \neq \top$ in general, $\mu_A(X) = \int^{\otimes}_X A\, d\mu$ does not define a fuzzy measure on $(M, \mathcal{M})$ in the sense of Definition 3.2.

Remark 4.2. One can also define a $\wedge$-fuzzy integral of A on X in such way that $\otimes$ is replaced by $\wedge$ in (21). Since $\otimes$ and $\wedge$ have many common properties, both types of fuzzy integral will have similar properties. Nevertheless, we prefer the $\otimes$-fuzzy integral in this paper, because it is closely related (due to the adjointness property) to $\rightarrow$-fuzzy integral that will be introduced in the following subsection.

Theorem 4.1. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. Then $\mu' : \mathcal{F}_{\mathbf{L}}(M) \rightarrow L$ defined by*

$$\mu'(A) = \int^{\otimes} A\, d\mu \tag{22}$$

is a fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$.

Theorem 4.2. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. Then*

(i) $\int^{\otimes}_X (A \cap B)\, d\mu \leq \int^{\otimes}_X A\, d\mu \wedge \int^{\otimes}_X B\, d\mu$,

(ii) $\int^{\otimes}_X (A \cup B)\, d\mu \geq \int^{\otimes}_X A\, d\mu \vee \int^{\otimes}_X B\, d\mu$,

(iii) $\int^{\otimes}_X (c \otimes A)\, d\mu \geq c \otimes \int^{\otimes}_X A\, d\mu$,

(iv) $\int^{\otimes}_X (c \rightarrow A)\, d\mu \leq c \rightarrow \int^{\otimes}_X A\, d\mu$,

hold for any $X \in \mathcal{M}$, $A, B \in \mathcal{F}_{\mathbf{L}}(M)$ and $c \in L$.

Theorem 4.3. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space and $c \in L$. Then we have*

(i) $\int^{\otimes} (c \otimes 1_X)\, d\mu = c \otimes \mu^*(1_X)$ *for any $X \subseteq M$,*

(ii) $\int^{\otimes} (c \otimes 1_X)\, d\mu = c \otimes \mu(1_X)$ *for any $X \subseteq M$ such that $1_X \in \mathcal{M}$,*

(iii) $\int^{\otimes} 1_X\, d\mu = \mu(1_X)$ *for any $X \subseteq M$ such that $1_X \in \mathcal{M}$,*

(iv) $\int^{\otimes} c\, d\mu = c$.

Theorem 4.4. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. If $X \in \mathcal{M}$ is such that $1_{\mathrm{Supp}(Y)} \in \mathcal{M}_X$ for any $Y \in \mathcal{M}_X$, then, for any $A \in \mathcal{F}_{\mathbf{L}}(M)$, we have*

$$\int^{\otimes}_X A\, d\mu = \bigvee_{1_Y \in \mathcal{P}_X \setminus \{1_\emptyset\}} \bigwedge_{m \in Y} (A(m) \otimes \mu(1_Y)), \tag{23}$$

where $\mathcal{P}_X = \{1_{\mathrm{Supp}(Z)} \mid Z \in \mathcal{M}_X\}$.

Theorem 4.5. *Let $\mathbf{L}$ be a complete MV-algebra, $(M, \mathcal{M}, \mu)$ be a fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $X \in \mathcal{M}$. Then*

$$\int^{\otimes}_X A\, d\mu = \bigvee_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \big(\mu(Y) \otimes \bigwedge_{m \in \mathrm{Supp}(Y)} A(m)\big). \tag{24}$$

Moreover,

$$\int^{\otimes}_X (c \otimes A)\, d\mu = c \otimes \int^{\otimes}_X A\, d\mu \tag{25}$$

for any $c \in L$.

Theorem 4.6. *Let g be an isomorphism between fuzzy measure spaces $(M, \mathcal{M}, \mu)$ and $(M', \mathcal{M}', \mu')$ and $X \in \mathcal{M}$. Then we have*

$$\int^{\otimes}_X A\, d\mu = \int^{\otimes}_{g(X)} g(A)\, d\mu' \tag{26}$$

for any $A \in \mathcal{F}_{\mathbf{L}}(M)$.

In the end of this part, we will show that the Sugeno integral is a special case of our proposed integral. For this purpose we will use a slight modification of the usual Sugeno integral definition with respect to the fuzzy measurable spaces over complete residuated lattices.

Let $\mathbf{L}$ be a complete residuated lattice and $(M, \mathcal{M})$ be a fuzzy measurable space such that $A \cap B \in \mathcal{M}$ for any $A, B \in \mathcal{M}$.[5] Denote $A_a = \{m \mid m \in M \;\&\; A(m) \geq a\}$. We say that an $\mathbf{L}$-fuzzy set A is *$\mathcal{M}$-Sugeno measurable*, if $1_{A_a} \in \mathcal{M}$ for any $a \in L$. The *Sugeno integral* is given, for any fuzzy measure space $(M, \mathcal{M}, \mu)$ with $B \cap C \in \mathcal{M}$ for any $B, C \in \mathcal{M}$, for any $\mathcal{M}$-Sugeno measurable $\mathbf{L}$-fuzzy set A and for any $X \in \mathcal{M}$, by

$$\int_X A\, d\mu = \bigvee_{a \in L} (a \wedge \mu(1_{A_a} \cap X)). \tag{27}$$

Theorem 4.7. *Let $\mathbf{L}$ be a complete Heyting algebra, $(M, \mathcal{M}, \mu)$ be a fuzzy measure space with $B \cap C \in \mathcal{M}$ for any $B, C \in \mathcal{M}$, A be a $\mathcal{M}$-Sugeno measurable $\mathbf{L}$-fuzzy set and $X \in \mathcal{M}$. Then $\int_X A\, d\mu = \int^{\otimes}_X A\, d\mu$.*

4.2 $\rightarrow$-fuzzy integral

In this part, we will introduce another type of fuzzy integral that can be defined on an arbitrary complementary fuzzy measure space $(M, \mathcal{M}, \nu)$. The form of this integral is motivated by our need to describe another class of models of $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1 \rangle$ which are kind of negations of the previous ones.

Definition 4.2. Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and X be a $\mathcal{M}$-fuzzy measurable $\mathbf{L}$-fuzzy set. The *$\rightarrow$-fuzzy integral of A on X* is given by

$$\int^{\rightarrow}_X A\, d\nu = \bigwedge_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \bigvee_{m \in \mathrm{Supp}(Y)} (A(m) \rightarrow \nu(Y)). \tag{28}$$

If $X = 1_M$, then we write $\int^{\rightarrow} A\, d\nu$.

Remark 4.3. It is easy to see that $\int^{\rightarrow}_{1_\emptyset} A\, d\nu = \bigwedge \emptyset = \top$ for any $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $\int^{\rightarrow}_X A\, d\nu \leq \int^{\rightarrow}_Y A\, d\nu$, whenever $Y \subseteq X$. Since $\int^{\rightarrow}_{1_M} A\, d\nu \neq \bot$ in general, $\nu_A(X) = \int^{\rightarrow}_X A\, d\nu$ does not define a complementary fuzzy measure on $(M, \mathcal{M})$ in the sense of Definition 3.3.

[5] Note that, according to Theorem 2.1, p. 2, each complete residuated lattice satisfying the law of double negation has this property. Nevertheless, there are fuzzy measurable spaces which keep this property, but $\mathbf{L}$ does not satisfy the law of double negation. A simple example is a fuzzy measurable space $(M, \mathcal{M})$ such that $\mathcal{M} \subseteq \mathcal{P}_{\mathbf{L}}(M)$ and $\mathbf{L}$ is an arbitrary complete residuated lattice (e.g. $L_{[0,\infty]}$ from Example 2.2).

ISBN: 978-989-95079-6-8

Theorem 4.8. *Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space. Then $\nu' : \mathcal{F}_{\mathbf{L}}(M) \to L$ defined by*

$$\nu'(A) = \int^{\to} A \, d\nu \tag{29}$$

is a complementary fuzzy measure on $(M, \mathcal{F}_{\mathbf{L}}(M))$.

Theorem 4.9. *Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space. Then*

(i) $\int_X^{\to}(A \cap B)\, d\nu \geq \int_X^{\to} A \, d\nu \vee \int_X^{\to} B \, d\nu$,

(ii) $\int_X^{\to}(A \cup B)\, d\nu \leq \int_X^{\to} A \, d\nu \wedge \int_X^{\to} B \, d\nu$,

(iii) $\int_X^{\to}(c \otimes A)\, d\nu \leq c \to \int_X^{\to} A \, d\nu$,

(iv) $\int_X^{\to}(c \to A)\, d\nu \geq c \otimes \int_X^{\to} A \, d\nu$

hold for any $X \in \mathcal{M}$, $A, B \in \mathcal{F}_{\mathbf{L}}(M)$ and $c \in L$.

Theorem 4.10. *Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space and $c \in L$. Then we have*

(i) $\int^{\to}(c \otimes 1_X)\, d\nu = c \to \nu^*(1_X)$ *for any $X \subseteq M$,*

(ii) $\int^{\to}(c \otimes 1_X)\, d\nu = c \to \nu(1_X)$ *for any $X \subseteq M$ such that $1_X \in \mathcal{M}$,*

(iii) $\int^{\to} 1_X \, d\nu = \nu(1_X)$ *for any $X \subseteq M$ such that $1_X \in \mathcal{M}$,*

(iv) $\int^{\to} c \, d\nu = \neg c$.

Theorem 4.11. *Let $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space. If $X \in \mathcal{M}$ is such that $1_{\mathrm{Supp}(A)} \in \mathcal{M}_X$ for any $A \in \mathcal{M}_X$, then, for any $A \in \mathcal{F}_{\mathbf{L}}(M)$, we have*

$$\int_X^{\to} A \, d\nu = \bigvee_{1_Y \in \mathcal{P}_X \setminus \{1_\emptyset\}} \bigwedge_{m \in Y} (A(m) \to \nu(1_Y)), \tag{30}$$

where $\mathcal{P}_X = \{1_{\mathrm{Supp}(A)} \mid A \in \mathcal{M}_X\}$.

Theorem 4.12. *Let $\mathbf{L}$ be a complete MV-algebra, $(M, \mathcal{M}, \nu)$ be a complementary fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $X \in \mathcal{M}$. Then*

$$\int_X^{\to} A \, d\nu = \bigwedge_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \big((\bigwedge_{m \in \mathrm{Supp}(Y)} A(m)) \to \nu(Y) \big). \tag{31}$$

Moreover,

$$\int_X^{\to} (c \otimes A)\, d\nu = c \to \int_X^{\to} A \, d\nu \tag{32}$$

for any $c \in L$.

Theorem 4.13. *Let g be an isomorphism between complementary fuzzy measure spaces $(M, \mathcal{M}, \nu)$ and $(M', \mathcal{M}', \nu')$ and $X \in \mathcal{M}$. Then we have*

$$\int_X^{\to} A \, d\nu = \int_{g(X)}^{\to} g(A) \, d\nu' \tag{33}$$

for any $A \in \mathcal{F}_{\mathbf{L}}(M)$.

The following statement shows that if we consider a complete MV-algebra, then we can restrict ourselves, for example, to $\otimes$-fuzzy integrals, since each $\to$-fuzzy integral is uniquely determined by the negation of a suitable $\otimes$-fuzzy integral.

Theorem 4.14. *Let $\mathbf{L}$ be a complete MV-algebra and $(M, \mathcal{M})$ be a fuzzy measurable space. Then*

$$\int_X^{\to} A \, d\nu' = \neg \int_X^{\otimes} A \, d\mu, \tag{34}$$

$$\int_X^{\otimes} A \, d\mu' = \neg \int_X^{\to} A \, d\nu \tag{35}$$

hold for any fuzzy measure μ and complementary fuzzy measure ν, where $\nu' = \neg\mu$ and $\mu' = \neg\nu$.

5 Conclusions

In this contribution, new types of fuzzy integrals, that are quite useful for modeling fuzzy quantifiers (of the type $\langle 1 \rangle$), are introduced and some of their properties are studied. The definitions of fuzzy quantifiers using these types of fuzzy integrals and some of their semantical properties could be found in [2].

Acknowledgment

This paper has been supported by the Grant IAA108270901 of the GA AV ČR.

References

[1] V. Novák. A formal theory of intermediate quantifiers. *Fuzzy Sets and Systems*, 159(10):1229 1246, 2008.

[2] A. Dvořák and M. Holčapek. Basic properties of **L**-fuzzy quantifiers of the type $\langle 1 \rangle$ determined by fuzzy measures. In *IFSA/EUSFLAT 2009*, Lisboa, 2009.

[3] R. Mesiar. Fuzzy measures and integrals. *Fuzzy Sets and Systems*, 156:365–370, 2005.

[4] M. Ying. Linguistic quantifiers modeled by Sugeno integrals. *Artificial Intelligence*, 170:581–606, 2006.

[5] W. San-min and Z. Bin. Prenex normal form in linguistic quantifiers modeled by Sugeno integrals. *Fuzzy Sets and Systems*, 159:1719–1723, 2008.

[6] I. Glöckner. *Fuzzy Quantifiers: A Computational Theory*. Springer-Verlag, Berlin, 2006.

[7] M. Holčapek. Monadic **L**-fuzzy quantifiers of the type $\langle 1^n, 1 \rangle$. *Fuzzy Sets and Systems*, 159:1811–1835, 2008.

[8] M. Sugeno. *Theory of Fuzzy Integrals and its Applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[9] R. Bělohlávek. *Fuzzy Relational Systems: Foundations and Principles*. Kluwer Academic Publisher, New York, 2002.

[10] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, Dordrecht, 1998.

[11] Z. Wang and G.J. Klir. *Fuzzy measure theory*. Plenum Press, New York, 1992.

[12] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy Measures and Integrals. Theory and Applications*. Studies in Fuzziness and Soft Computing. Physica Verlag, Heidelberg, 2000.

[13] E.P. Klement, R. Mesiar, and E. Pap. A universal integral based on measures of level sets. *IEEE Transactions on Fuzzy Systems*, to appear.

[14] L. Mišík and J. Tóth. On asymptotic behaviour of universal fuzzy measure. *Kybernetika*, 42(3):379–388, 2006.

[15] S. Peters and D. Westerståhl. *Quantifiers in Language and Logic*. Oxford University Press, New York, 2006.

ISBN: 978-989-95079-6-8

A link between the 2-additive Choquet Integral and Belief functions

Brice Mayag[1,2] Michel Grabisch[1] Christophe Labreuche[2]

1. University of Paris I, Centre d'Economie de la Sorbonne
106-112 Boulevard de l'Hôpital, 75013 Paris, France
2. Thales Research & Technology
Campus de Polytechnique, 1 avenue Augustin Fresnel
91767 Palaiseau Cedex- France
{brice.mayag, christophe.labreuche}@thalesgroup.com, michel.grabisch@univ-paris1.fr

Abstract— In the context of decision under uncertainty, we characterize the 2-additive Choquet integral on the set of fictitious acts called binary alternatives or binary actions. This characterization is based on a fundamental property called MOPI which permits us to relate belief functions and the 2-additive Choquet integral.

Keywords— Capacity, Möbius transform, Choquet integral, k-monotone function, Belief function, Decision under uncertainty

1 Introduction

Decision under uncertainty is a part of decision making where each act has several possible consequences, depending on the state of nature whose probability of occurrence is unknown. As shown by the well-known Ellsberg's paradox [1, 2], the use of the expected utility model [3] in decision under uncertainty is limited. Therefore some non-additive models like Choquet expected utility [4] have been proposed in order to overcome the limitations of the expected utility model.

The Choquet integral is defined w.r.t. a capacity (or non-additive monotonic measure, or fuzzy measure), and can be thought of as a generalization of the expected value, the capacity playing the role of a probability measure. In this paper we focus on the 2-additive Choquet integral [5, 6], a particular Choquet integral where interaction between two states of nature can be represented, but not more complex interaction. This model is in practice already sufficiently flexible. In many situations, it is important for the Decision-Maker (DM) to construct a preference relation over the set of all acts X. Because it is not an easy task (the cardinality of X may be very large), we ask him to give, using pairwise comparisons, an ordinal information (a preferential information containing only a strict preference and an indifference relations) on a particular reference subset $\mathcal{B} \subseteq X$. The set $\mathcal{B}$ we use is the set of binary acts or binary actions. A binary action is a fictitious act which takes only two values denoted $\mathbf{1}$ and $\mathbf{0}$ belonging to the set of consequences, such that $\mathbf{1}$ is strictly preferred to $\mathbf{0}$. We present necessary and sufficient conditions on the ordinal information for the existence of a 2-additive capacity such that the Choquet integral w.r.t. this capacity represents the preference of the decision maker. We introduce the new fundamental property MOPI, a kind of monotonicity coming from the definition of a 2-additive capacity, in order to have this characterization. We found through our MOPI property the following link between the 2-additive Choquet integral and belief functions (Shafer [7]): *Any ordinal information representable by a belief function is representable by a Choquet integral w.r.t. a 2-additive capacity*.

Because a belief function is a capacity, we show another characterization of the representation of any ordinal information by a belief function. The new fundamental property defined in this case is called the 2-MOPI property. This property and the MOPI property in the previous paragraph are related by the following statement: *if the 2-MOPI property is satisfied then the MOPI property is satisfied*.

The article is organized as follows. The next section presents the basic concepts we need. Section 3 concerns a representation of ordinal information by the 2-additive Choquet integral. In the last section, after some results on the case of the k-monotone functions, we study the representation of an ordinal information by a belief function.

2 Preliminaries

Let us denote by $N = \{1, \ldots, n\}$ the set of n states of nature and by 2^N the set of all subsets of N. The set of possible consequences (also called "outcomes") is denoted by C. An act x is identified to an element of $X = C^n$ with $x = (x_1, ..., x_n)$. We introduce the following convenient notation: for two acts $x, y \in X$ and a subset $A \subseteq N$, the compound act $z = (x_A, y_{N-A})$ is defined by $z_i := x_i$ if $i \in A$, and $z_i := y_i$ otherwise. For all i, j in N, the element $i \vee j$ denotes one of the elements i, j.

We want to construct a preference relation over X, but this is not easy because X may contain infinitely many acts. In practice [8] one can only ask to the DM pairwise comparisons of acts on a finite subset X' of X. Hence we get a preference relation $\succsim_{X'}$ on X'. The question is then: how to construct $\succsim_X$ from $\succsim_{X'}$? To this end, people usually suppose that $\succsim_X$ is representable by an overall utility function:

$$x \succsim_X y \Leftrightarrow F(U(x)) \geq F(U(y)) \qquad (1)$$

where $U(x) = (u(x_1), \ldots, u(x_n))$, $u : C \to \mathbb{R}$ is called a utility function, and $F : \mathbb{R}^n \to \mathbb{R}$ is an aggregation

ISBN: 978-989-95079-6-8

function. Usually, we consider a family of aggregation functions characterized by a parameter vector θ (e.g., a probability distribution over the states of nature). The parameter vector θ can be deduced from the knowledge of $\succsim_{X'}$, that is, we determine the possible values of θ for which (1) is fulfilled over X'. We study the case where F is the Choquet integral, the parameters are the 2-additive capacity and X' is the set of binary actions. The aim of this paper is to give necessary and sufficient conditions on $\succsim_{X'}$ to be represented by a 2-additive Choquet integral. The model obtained in X' will be automatically extended to X.

2.1 The 2-additive Choquet Integral

The 2-additive Choquet integral [6] is a particular case of the Choquet integral [9, 10]. This integral generalizes the arithmetic mean and takes into account interactions between the states of nature. A 2-additive Choquet integral is based on a 2-additive capacity defined below and its Möbius transform [11, 12]:

Definition 2.1.

1. A *capacity* on N is a set function $\mu : 2^N \to [0,1]$ such that:
   (a) $\mu(\emptyset) = 0$
   (b) $\mu(N) = 1$
   (c) $\forall A, B \in 2^N$, $[A \subseteq B \Rightarrow \mu(A) \leq \mu(B)]$ (monotonicity).
2. The *Möbius transform* of a capacity μ on N is a function $m : 2^N \to \mathbb{R}$ defined by:

$$m(T) := \sum_{K \subseteq T} (-1)^{|T \setminus K|} \mu(K), \forall T \in 2^N \qquad (2)$$

When m is given, it is possible to recover the original μ by the following expression:

$$\mu(T) := \sum_{K \subseteq T} m(K), \forall T \in 2^N \qquad (3)$$

Definition 2.2. A capacity μ on N is said to be *2-additive* if

- For all subset T of N such that $|T| > 2$, $m(T) = 0$;
- There exists a subset B of N such that $|B| = 2$ and $m(B) \neq 0$.

Notations We simplify our notations by using for a capacity μ and its Möbius transform m: $\mu_i := \mu(\{i\})$, $\mu_{ij} := \mu(\{i,j\})$, $m_i := m(\{i\})$, $m_{ij} := m(\{i,j\})$, for all $i, j \in N$, $i \neq j$. Whenever we use i and j together, it always means that they are different.

The following important Lemma shows that a 2-additive capacity is entirely determined by the value of the capacity on the singletons $\{i\}$ and pairs $\{i,j\}$ of 2^N:

Lemma 1.

1. *Let μ be a 2-additive capacity on N. We have*

$$\mu(K) = \sum_{\{i,j\} \subseteq K} \mu_{ij} - (|K|-2) \sum_{i \in K} \mu_i, \ \forall K \subseteq N, \ |K| \geq 2. \qquad (4)$$

2. *If the coefficients μ_i and μ_{ij} are given for all $i, j \in N$, then the necessary and sufficient conditions that μ is a 2-additive capacity are:*

$$\sum_{\{i,j\} \subseteq N} \mu_{ij} - (n-2) \sum_{i \in N} \mu_i = 1 \qquad (5)$$

$$\mu_i \geq 0, \ \forall i \in N \qquad (6)$$

$$\sum_{i \in A \setminus \{k\}} (\mu_{ik} - \mu_i) \geq (|A| - 2)\mu_k, \ \forall A \subseteq N, |A| \geq 2, \forall k \in A. \qquad (7)$$

Proof. See [6] □

For an act $x := (x_1, ..., x_n) \in X$, the expression of the Choquet integral w.r.t a capacity μ is given by:

$$C_\mu((u(x_1), ..., u(x_n))) := u(x_{\tau(1)})\mu(N) + \sum_{i=2}^{n} (u(x_{\tau(i)}) - u(x_{\tau(i-1)}))\mu(\{\tau(i), ..., \tau(n)\})$$

where τ is a permutation on N such that $u(x_{\tau(1)}) \leq u(x_{\tau(2)}) \leq ... \leq u(x_{\tau(n-1)}) \leq u(x_{\tau(n)})$.

A Choquet integral with a 2-additive capacity μ is called a 2-additive Choquet integral. Given an act $x := (x_1, ..., x_n) \in X$, the 2-additive Choquet integral can be written also as follows:

$$C_\mu((u(x_1), \ldots, u(x_n))) = \sum_{i=1}^{n} v_i u(x_i) - \frac{1}{2} \sum_{\{i,j\} \subseteq N} I_{ij} |u(x_i) - u(x_j)| \qquad (8)$$

where $v_i = \sum_{K \subseteq N \setminus i} \frac{(n - |K| - 1)!|K|!}{n!} (\mu(K \cup i) - \mu(K))$ represents the importance of the state of nature i and corresponds to the Shapley value [13]; $I_{ij} = \mu_{ij} - \mu_i - \mu_j$ is the interaction index between the two states of nature i and j.

The above development suggests that the Choquet integral w.r.t. a 2-additive capacity seems to be of particular interest, and offers a good compromise between flexibility of the model and complexity. Therefore, we focus in this paper on the 2-additive model.

2.2 Binary actions

We assume in this paper that the DM is able to identify in C two consequences denoted $\mathbf{1}$ and $\mathbf{0}$ such that he strictly prefers $\mathbf{1}$ to $\mathbf{0}$. In the sequel, we call $\mathbf{0}$ the "neutral level" (even if this is not the neutral level understood in bipolar model).

We call a *binary action or binary act*, an element of the set $\mathcal{B} = \{\mathbf{0}_N, \ (\mathbf{1}_i, \mathbf{0}_{N-i}), \ (\mathbf{1}_{ij}, \mathbf{0}_{N-ij}), \ i, j \in N, i \neq j\} \subseteq X$ where

- $\mathbf{0}_N = (\mathbf{1}_\emptyset, \mathbf{0}_N) =: a_0$ is an act which has a consequence $\mathbf{0}$ on all states of nature.

ISBN: 978-989-95079-6-8

- $(\mathbf{1}_i, \mathbf{0}_{N-i}) =: a_i$ is an act which has a consequence $\mathbf{1}$ on state of nature i and a consequence $\mathbf{0}$ on the other states of nature.

- $(\mathbf{1}_{ij}, \mathbf{0}_{N-ij}) =: a_{ij}$ is an act which has a consequence $\mathbf{1}$ on states of nature i and j and a consequence $\mathbf{0}$ on the other states of nature.

By convention we set $u(\mathbf{0}) = 0$ and $u(\mathbf{1}) = 1$. The above convention have the following consequences:

Remark 1.

1. The Choquet integral satisfies the following property [14, 10]: if μ is a capacity then

$$C_\mu(U(\mathbf{1}_A, \mathbf{0}_{N-A})) = \mu(A),\ \forall A \subseteq N. \qquad (9)$$

2. Let μ be a 2-additive capacity. We have

$C_\mu(U(a_0)) = 0$;

$C_\mu(U(a_i)) = \mu_i = v_i - \frac{1}{2} \sum_{k\in N,\ k\neq i} I_{ik}$;

$C_\mu(U(a_{ij})) = \mu_{ij} = v_i + v_j - \frac{1}{2} \sum_{k\in N,\ k\notin\{i,j\}} (I_{ik} + I_{jk})$

Generally the DM knows how to compare some acts using his knowledge of the problem, his experience, etc. These acts form a set of reference acts and allows to determine the parameters of a model (utility functions, subjective probabilities, weights,...) in the decision process (see [8] for more details). As shown by the previous Remark 1 and Lemma 1, it should be sufficient to get some preferential information from the DM only on binary acts. To entirely determine the 2-additive capacity this information is expressed by the following relations:
$P = \{(x, y) \in \mathcal{B} \times \mathcal{B} :$ DM strictly prefers x to $y\}$, $I = \{(x, y) \in \mathcal{B} \times \mathcal{B} :$ DM is indifferent between x and $y\}$.

Definition 2.3. The *ordinal information on* $\mathcal{B}$ is the structure $\{P,\ I\}$.

Now we will suppose P nonempty for any ordinal information $\{P,\ I\}$ ("non triviality axiom"). Before we end this section, let us introduce another relation M which completes the ordinal information $\{P,\ I\}$ given by the DM and models the natural relations of monotonicity between binary actions. For $(x, y) \in \{(a_i, a_0), i \in N\} \cup \{(a_{ij}, a_i), i, j \in N, i \neq j\}$,

$$x\ M\ y \text{ if not}(x\ (P \cup I)\ y).$$

The relation M models the monotonicity conditions $\mu(\{i\}) \geq 0$ and $\mu(\{i, j\}) \geq \mu(\{i\})$ for a capacity μ.

Example 1. *If we consider*
$N = \{1, 2, 3\}$, $\mathcal{B} = \{a_o, a_1, a_2, a_3, a_{12}, a_{13}, a_{23}\}$, $P = \{(a_{13}, a_3), (a_2, a_3), (a_{23}, 0)\}$, $I = \{(a_{12}, a_1)\}$, *then the relation* M *is* $M = \{(a_{12}, a_0), (a_{13}, a_0), (a_1, a_0), (a_2, a_0), (a_3, a_0), (a_{12}, a_2), (a_{13}, a_1), (a_{23}, a_2), (a_{23}, a_3)\}$.

3 The representation of the ordinal information by the Choquet integral

An ordinal information $\{P, I\}$ is said to be *representable by a 2-additive Choquet integral* if there exists a 2-additive capacity μ such that:

1. $\forall x, y \in \mathcal{B},\ x\ P\ y \Rightarrow C_\mu(U(x)) > C_\mu(U(y))$
2. $\forall x, y \in \mathcal{B},\ x\ I\ y \Rightarrow C_\mu(U(x)) = C_\mu(U(y))$.

Given an ordinal information $\{P, I\}$, we look for the necessary and sufficient conditions on $\mathcal{B}$ for which $\{P, I\}$ is representable by a 2-additive Choquet integral. To do it, we need to define first the notion of strict cycle of the relation $(P \cup I \cup M)$.

3.1 Cycle of $(P \cup I \cup M)$

For a binary relation $\mathcal{R}$ on $\mathcal{B}$ and x, y elements of $\mathcal{B}$, $\{x_1, x_2, \cdots, x_p\} \subseteq \mathcal{B}$ is a *path* of $\mathcal{R}$ from x to y if $x = x_1\ \mathcal{R}\ x_2\ \mathcal{R} \cdots \mathcal{R}\ x_{p-1}\ \mathcal{R}\ x_p = y$. A path of $\mathcal{R}$ from x to x is called a *cycle* of $\mathcal{R}$.

- A path $\{x_1, x_2, ..., x_p\}$ of $(P \cup I \cup M)$ is said to be a *strict path* from x to y if there exists i in $\{1, ..., p-1\}$ such that $x_i\ P\ x_{i+1}$. In this case, we will write $x\ TC_P\ y$.

- A cycle $(x_1, x_2, ..., x_p)$ of $(P \cup I \cup M)$ is a *nonstrict cycle* if it is not strict.

- We note $x \sim y$ if there exists a nonstrict cycle of $(P \cup I \cup M)$ containing x and y.

Contrarily to the strict cycle which is a classic concept used in graph theory [15, 16], we need to define a new fundamental property called *MOPI*.

3.2 MOPI property and theorem of Characterization

Before defining the property *MOPI*, let us discover this new condition through a simple example:

Example 2. *Suppose that the DM says :* $a_{12}\ I\ a_3$, $a_{13}\ I\ a_2$ *and* $a_1\ P\ a_0$. *Using the relation* M, *we have* $a_{12}\ M\ a_2\ I\ a_{13}\ M\ a_3\ I\ a_{12}$. *So,* $(a_{12}, a_2, a_{13}, a_3, a_{12})$ *forms a nonstrict cycle of* $(P \cup I \cup M)$. *If* $\{P, I\}$ *is representable by a 2-additive Choquet integral* C_μ, *this implies* $\mu_{12} = \mu_{13} = \mu_2 = \mu_3$ *and* $\mu_1 > 0$. *However, we get a contradiction with the monotonicity constraint* $\mu_{12} + \mu_{13} \geq \mu_1 + \mu_2 + \mu_3$ *of a 2-additive capacity with the subset* $A = \{1, 2, 3\}$, $k = 1$ *(see Equation* (7) *in Lemma 1).*

This type of inconsistency is defined by:

Definition 3.1 (MOPI property)**.** Let $i, j, k \in N$, i fixed.

1. We call *Monotonicity of Preferential Information in* $\{i, j, k\}$ *w.r.t.* i the following property (denoted by **(**$\{i, j, k\}$**,***i***)-MOPI**):

$$\left.\begin{array}{l} a_{ij} \sim a_{i\vee j} \\ a_{ik} \sim a_{i\vee k} \\ i \vee j \in \{i, j\} \\ i \vee k \in \{i, k\} \\ i \vee j \neq i \vee k \end{array}\right\} \Rightarrow \begin{array}{l} [not(a_l\ TC_P\ a_0) \\ l \in \{i, j, k\} \setminus \{i \vee k, i \vee j\}] \end{array}$$

ISBN: 978-989-95079-6-8

If the property **($\{i,j,k\}$,i)-MOPI** is satisfied then the element $a_l, l \in \{i,j,k\} \setminus \{i \vee k, i \vee j\}$ is called the *neutral binary action of* $\{i,j,k\}$ *w.r.t.* i.

2. We say that $\{i,j,k\}$ satisfies the property *Monotonicity of Preferential Information* (MOPI) if $\forall l \in \{i,j,k\}$, ($\{i,j,k\}$,l)-MOPI is satisfied.

Example 3. *Let* $N = \{1,2,3,4\}$ *and* $i = 1$ *fixed. The property* $(\{1,2,3\},1)$*- MOPI reads as follows:*

- $\left\{ \begin{array}{l} a_{12} \sim a_2 \\ a_{13} \sim a_1 \end{array} \right. \Rightarrow not(a_3\ TC_P\ a_0)$
- $\left\{ \begin{array}{l} a_{12} \sim a_1 \\ a_{13} \sim a_3 \end{array} \right. \Rightarrow not(a_2\ TC_P\ a_0)$
- $\left\{ \begin{array}{l} a_{12} \sim a_2 \\ a_{13} \sim a_3 \end{array} \right. \Rightarrow not(a_1\ TC_P\ a_0)$

The MOPI condition given in this paper is equivalent to the MOPI property presented in [5]. We give below our theorem of characterization of consistent ordinal information $\{P,I\}$ representable by a 2-additive Choquet integral:

Theorem 1. *An ordinal information* $\{P,I\}$ *is representable by a 2-additive Choquet integral on* $\mathcal{B}$ *if and only if the following conditions are satisfied:*

1. $(P \cup I \cup M)$ *contains no strict cycle;*
2. *Any subset* K *of* N *such that* $|K| = 3$ *satisfies the MOPI property.*

Is it possible to represent an ordinal information by another operator instead of the 2-additive Choquet integral? If the answer is yes, can we give a similar characterization like in Theorem 1? In the next section, we will show that it is possible by using for instance belief functions.

4 The representation of ordinal information by belief functions

4.1 General definitions

Beliefs functions are one of the fundamental concepts used in the theory of evidence of Shafer [7]. They are defined by the belief function mass m as follows:

Definition 4.1. A function $m : 2^N \rightarrow [0,1]$ is called a *mass distribution* or a *basic belief assignment* if m satisfies the following two properties:

1. $m(\emptyset) = 0$;
2. $\sum_{A \subseteq N} m(A) = 1$.

The quantity $m(A)$ expresses the total amount of belief that supports the proposition: " the actual state of nature is in A", and does not support any more specific subset of N because of lack of information.

Based on this concept, we define the belief function Bel by:

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad \forall A \subseteq N.$$

Remark 2.

- Bel is a capacity;
- The sets A such that $m(A) > 0$ are called the focal elements;
- If all focal elements are singletons then a mass distribution can be considered as a probability distribution;
- The mass distribution m corresponds to the Möbius transform of Bel. So we have $\forall T \in 2^N$,

$$m(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B).$$

Thus, we can have a definition of the representation of ordinal information by a belief function which is similar to the same representation by a Choquet integral (see Section 3).

Definition 4.2. An ordinal information $\{P,I\}$ is said to be *representable by a belief function* if there exists a belief function Bel such that

1. $\forall x,y \in \mathcal{B},\ x\ P\ y \Rightarrow C_{Bel}(U(x)) > C_{Bel}(U(y))$
2. $\forall x,y \in \mathcal{B},\ x\ I\ y \Rightarrow C_{Bel}(U(x)) = C_{Bel}(U(y))$.

By using Definition 2.2, a 2-additive belief function has a mass distribution m characterized by:

1. $\exists i,j \in N$ tel que $m(\{i,j\}) \neq 0$;
2. $\forall K \in 2^N$ tel que $|K| \geq 3$, $m(K) = 0$.

Theorem 2 below provides a relation between a k-monotone function [6, 12] and a belief function, and a relation between k-monotone functions and the previous MOPI property.

4.2 k-monotone functions and belief functions

Given an integer $k \geq 2$, a function $\mu : 2^N \rightarrow [0,1]$ is k-monotone (shorthand for: monotone of order k) if for each family $\{A_1, A_2, \ldots, A_k\} \subseteq 2^N$, we have

$$\mu(\bigcup_{i=1}^{k} A_i) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \mu(\bigcap_{i \in I} A_i). \qquad (10)$$

A simpler characterization of k-monotone functions by their Möbius inversion is given by the following proposition:

Proposition 1. *Let* $\mu : 2^N \rightarrow [0,1]$ *and* m *its Möbius transform.* μ *is* k*-monotone (*k *integer,* $k \geq 2$*) if and only if*

$$\sum_{A \subseteq L \subseteq B} m(L) \geq 0 \quad \forall A,B \subseteq N, A \subseteq B \text{ and } 2 \leq |A| \leq k. \qquad (11)$$

ISBN: 978-989-95079-6-8

Proof. See [12] □

It is well-known that $\mu : 2^N \to [0,1]$ is a belief function if and only if μ is a k-monotone capacity for all $k \geq 2$. The following result gives another sufficient condition to obtain a belief function from a k-monotone and 2-additive capacity, and relates belief function with the MOPI condition translated in terms of capacity.

Theorem 2.
Let $\mu : 2^N \to [0,1]$ be a function and k be an integer such that $k \geq 2$.

1. *If μ is monotone, k-monotone and 2-additive then μ is a belief function (precisely a 2-additive belief function);*
2. *If μ is monotone and k-monotone then μ satisfies the following property: for all $i, j, k \in N$, i fixed*

$$\left.\begin{array}{l}\mu_{ij} = \mu_{i\vee j}\\ \mu_{ik} = \mu_{i\vee k}\\ i\vee j \in \{i,j\}\\ i\vee k \in \{i,k\}\\ i\vee j \neq i\vee k\end{array}\right\} \Rightarrow \begin{array}{l}[\mu_l = 0,\\ l \in \{i,j,k\}\setminus\{i\vee k, i\vee j\}]\end{array}$$

We end the paper by a characterization of ordinal information by belief functions.

4.3 A link between Belief functions and the 2-additive Choquet integral

In this section, we give through the MOPI property (see Section 3) a link between beliefs functions and the 2-additive Choquet integral.

Proposition 2. *Let $\{P,\ I\}$ be an ordinal information on $\mathcal{B}$.*
If there exist $i, j, k \in N$, i fixed such that the property $(\{i,j,k\}, i)$-MOPI is violated, then there is no belief function Bel which represents $\{P,\ I\}$.

Corollary 1. *Every ordinal information $\{P,\ I\}$ on $\mathcal{B}$ representable by a belief function $Bel : 2^N \to [0,1]$ is representable by a 2-additive Choquet integral.*

The inverse of Corollary 1 is false. If we suppose $P = \{(a_2, a_0)\}$, $I = \{(a_{12}, a_1)\}$ and μ a 2-additive capacity, we will have $\{P,\ I\}$ representable by a 2-additive Choquet integral and $I_{12} = m_{12} = \mu_{12} - \mu_1 - \mu_2 < 0$. So no belief function can represent $\{P,\ I\}$ in this case. Then it is interesting to look for the class of 2-additive capacities which are belief functions. In order to characterize them, we introduce a new fundamental property called 2-MOPI property:

Definition 4.3. An ordinal information $\{P,\ I\}$ satisfies the 2-*MOPI property* if

$$\forall i, j \in N, i \neq j, [a_{ij} \sim a_i \Rightarrow not(a_j\ TC_P\ a_0)]. \qquad (12)$$

The relation between the 2-MOPI property and the MOPI property is given by the following proposition:

Proposition 3. *Let $\{P,\ I\}$ an ordinal information on $\mathcal{B}$.*

$$\{P,\ I\} \text{ satisfies the 2-MOPI property}$$
$$\Downarrow$$
$$\forall i, j, k \in N,\ \{i,j,k\} \text{ satisfies the MOPI property}$$

Now we have the main result of this section:

Theorem 3.
$\{P,\ I\}$ is representable by a 2-additive belief function if and only if the two following conditions are satisfied:

1. *$(P \cup I \cup M)$ contains no strict cycle;*
2. *$\{P,\ I\}$ satisfies the 2-MOPI property.*

4.4 Interpretation of 2-MOPI and MOPI properties

We try to give an interpretation in terms of decision behavior of the two main conditions introduced in this paper. We assume here for clarity that consequence **1** is a good consequence for the DM, while consequence **0** is neither bad nor good (statu quo).

Facing a situation where for two states of nature i and j the DM is indifferent between the two acts a_{ij} and a_i, the 2-MOPI property says that act a_j is equivalent to act a_0 (statu quo for every state of nature). Hence in such a situation, the DM thinks that state of nature j is unlikely to occur. This is a strong condition, since it suffices that one such state i exists to infer the "nullity" of state j. This condition can be related to the notion of null set in generalized measure theory (see, e.g., [17]): a set $A \subseteq N$ is said to be *null* for capacity μ if $\mu(B \cup A) = \mu(B)$, $\forall B \subseteq N \setminus A$. Taking $A = \{j\}$ and $B = \{i\}$ gives our condition 2-MOPI. Observe that for the nullity condition, $\{j\}$ would be null if for *all* subsets B not containing j we would have $\mu(B \cup j) = \mu(B)$, but the 2-MOPI condition asks to find only one singleton satisfying this equality.

The MOPI property is a weakening of the above one, and can be interpreted in a similar way. Let us consider now three states of nature i, j and k. The MOPI condition can be translated as follows (see Example **??**, with $i = 1$, $j = 2$, and $k = 3$). Suppose that a_{ij} and a_j are indifferent. As above, this would suggest that i is unlikely to occur for the DM, but this is *relatively to the occurence of* j, or put differently, i is much less likely than j. Suppose in addition that a_{ik} is indifferent to a_i. Again, this suggests that k much less likely to occur than i. Since i is much less likely than j, the conclusion is that k is very unlikely to occur, hence a_k is indifferent to a_0. This explains the first case in the MOPI condition. The second case (indifference between a_{ik} and a_k, and between a_{ij} and a_i) works exactly the same way. The third case says that a_{ij} and a_j are indifferent (i is much less likely than j) as well as a_{ik} and a_k (i is much less likely than k). Since i is much less likely than both j and k, the conclusion is that i is very unlikely, so that a_i is indifferent with a_0.

For $N = \{1,2\}$, the 2-MOPI property can be also viewed as uncertainty aversion[1] (see [18]). Indeed,

[1] Uncertainty aversion, as presented in [18], is defined as follows: For three acts x, y, z, if y and z are comonotonic then:

$$x \sim y \Rightarrow x + z \gtrsim y + z.$$

Comonotonicity between two acts y, z means that there are no $i, j \in N$ such that $u(y_i) > u(y_j)$ and $u(z_i) < u(z_j)$.

ISBN: 978-989-95079-6-8

given the three acts $a_{12} = (\mathbf{1}, \mathbf{1})$, $a_1 = (\mathbf{1}, \mathbf{0})$, $z = (-\mathbf{1}, \mathbf{0})$ and using the property of uncertainty aversion, we have:

$$(\mathbf{1}, \mathbf{0}) \sim (\mathbf{1}, \mathbf{1}) \Rightarrow (\mathbf{0}, \mathbf{0}) \gtrsim (\mathbf{0}, \mathbf{1})$$

which corresponds to the 2-MOPI property in this case. However, this interpretation does not work any more for the MOPI condition.

References

[1] D. Ellsberg. Risk, ambiguity, and the Savage axioms. *Quart. J. Econom.*, 75:643–669, 1961.

[2] A. Chateauneuf. Ellsberg paradox intuition and Choquet expected utility. In *Mathematical models for handling partial knowledge in artificial intelligence*. Plenum Press, 1995.

[3] J. Von Neuman and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[4] D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587, 1989.

[5] B. Mayag, M. Grabisch, and C. Labreuche. A characterization of the 2-additive Choquet integral. In *Proceedings of IPMU'08 (CD)*, Malaga, Spain, June 2008.

[6] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.

[7] G. Shafer. *A Mathematical Theory of Evidence*. Princeton Univ. Press, 1976.

[8] Th. Marchant. Towards a theory of MCDM: Stepping away from social choice theory. *Mathematical Social Sciences*, 45(3):343–363, 2003.

[9] M. Grabisch and Ch. Labreuche. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *4OR*, 6:1–44, 2008.

[10] M. Grabisch and C. Labreuche. Fuzzy measures and integrals in MCDA. In J. Figueira, S. Greco, and M. Ehrgott, editors, *Multiple Criteria Decision Analysis*, pages 563–608. Kluwer Academic Publishers, 2004.

[11] M. Grabisch. The Möbius transform on symmetric ordered structures and its application to capacities on finite sets. *Discrete Mathematics*, 287 (1-3):17–34, 2004.

[12] A. Chateauneuf and J.Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.

[13] L.S. Shapley. A value for n-person games. *Contributions to the Theory of games*, pages 307–317, 1953.

[14] J.L. Marichal. An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. *IEEE Tr. on Fuzzy systems*, 8(6):800–807, 2000.

[15] M. Gondran and M. Minoux. *Graphes et algorithmes*. Eyrolles, 1986.

[16] M. Pirlot and Ph. Vincke. *Semiorders*. Kluwer academic publishers, 1997.

[17] E. Pap. *Null-Additive Set Functions*. Kluwer Academic, 1995.

[18] A. Chateauneuf. Modeling attitudes towards uncertainty and risk through the use of Choquet integral. *Annals of Operations Research*, 52:3–20, 1994.

ISBN: 978-989-95079-6-8

Agent-based Fuzzy Constraint-directed Negotiation Mechanism for Planning and Scheduling in Supply Chain

Bo-Ruei Kao,[1] K. Robert Lai,[1] Menq-Wen Lin[2]

1.Department of Computer Science & Engineering, Yuan Ze University,
Chung-Li, Taiwan
2.Department of Information Management, Ching Yun University,
Chung-Li, Taiwan
Email: s909402@mail.yzu.edu.tw, krlai@cs.yzu.edu.tw, linmw@cyu.edu.tw

Abstract— *This paper presents an agent-based fuzzy constraint-directed negotiation mechanism for planning and scheduling in supply chain. The supply chain scheduling problem is modeled as a set of fuzzy constraint satisfaction problems (FCSPs), interlinked together by inter-agent constraints. For converging each distinct firm's interests, the conflicts among the set of FCSPs are resolved through negotiation by iteratively exchanging offers/counters with limited sharing of their perspectives and preferences. During the negotiation, proposing offers/counter-offers takes not only firm's self-interest and preferences but also opponents' perspectives into consideration. By sharing perspectives between agents to gradually uncovering the intent of opponents, consensus can be obtained and the quality of consensus can be guaranteed at satisfactory level. Experimental results suggest that the proposed approach obtains a superior solution for supply chain scheduling than other negotiation models in a fully distributed processing.*

Keywords— Supply Chain Planning and Scheduling, Fuzzy Constraint-directed Negotiation, Fuzzy Constraint Satisfaction Problem, Multi-Agent System.

1 Introduction

Facing with global economic downturn and intense competition, firms have to rapidly respond to the variations in market situation. Business entities (suppliers, manufacturers, distributors, retailers and customers) are urged to integrate their operations into multi-layer supply chain for product/service provision [1]. However, supply chain usually handles multiple projects concurrently with shared components, facilities, and capacities governed by distinct firms. Collaborating with supply chain partners in planning and scheduling in a cost effective manner is the key to survive in today's fierce market competition.

Since the supply chain environment in nature is distributed, autonomous, and heterogeneous, agent-based approaches which are characterized by decentralization of computation and information processing are particularly attractive for supply chain modeling and problem solving [2, 3]. To ensure the global performance with limited interaction among agents, a mediator or a third party agent is used in some systems for coordination. Negotiation processes can be facilitated by a mediator which helps the parties understand their needs, suggests possible agreements and/or supports them in the implementation of the agreement. But, agents may have to share sensitive strategic information that should not be revealed to competitors or even to a third party agent, to the mediator for coordination.

So, instead of coordinating with a third party agent, coordination/negotiation in a fully distributed environment for privacy concern has been addressed in some agent-based approaches. Those papers address the dynamic nature of supply chain and emphasize the flexibility and responsiveness. Two popular and fully distributed negotiation models, contract net protocol (CNP) and market-based protocol, have been proposed for these purposes [4]. The CNP can rapidly produce a feasible solution to overcome the frequently changes in supply chain. With simplified negotiation protocol, however, agents with very limited interaction and information sharing can only make their decisions independently and optimize their local objectives in a myopic way. Accordingly, more sophisticated negotiation mechanism is needed and thus bringing us to a market-based approach. Market-based approaches employ a bargaining or auction process characterized by iterative bidding among agents. During the bidding process, agents will adjust bids according to the direction of surplus and deficit of demand. The bids which imply the degree of competition can be used to resolve the conflicts caused by contention and improve the system performance. But the bid indicates only the demand that is in high or low contention, not where the contention can be resolved. Thus, the process might oscillate and not achieve the convergence. It also affects the quality of solution.

Accordingly, facilitating the convergence and guaranteeing the system performance in supply chain are the critical challenges and they are highly influenced by the level of information sharing [5, 6, 7, 8] as well. To cope with these problems, this paper proposes an agent-based fuzzy constraint-directed negotiation mechanism (AFCN) in a fully distributed environment.

In AFCN, a supply chain problem is modeled as a set of fuzzy constraint satisfaction problems (FCSPs), interlinked together by inter-agent constraints. Each FCSP represents the firm's perspectives and issues in supply chain. Constraints which are characterized by non directional, declarative, and intuitive properties are closely related to the real-world problem descriptions. Besides, the subjective, imprecision and qualitative knowledge (e.g., human cognition, preferences, or even opponent's perspectives) which are frequently used for business decision making in supply chain can easily be coped

ISBN: 978-989-95079-6-8

by fuzzy constraints with the levels of consistency. Fuzzy constraints also can be used to rank the solutions by specifying the possibilities prescribing to what extent the solutions are suitable. Additionally, it even provides a measure of similarity between solutions and opponents' counteroffers and a basis for the selection of multi-objective decision from a set of feasible alternatives. Thus, AFCN can be a practical and effective methodology for the supply chain problem modeling.

To coordinate supply chain parties, the conflicts among set of FCSPs are resolved through AFCN negotiation protocol by iteratively exchanging offers/counters with limited sharing of their perspectives. For converging each distinct firm's interests in supply chain, proposing an offer/counteroffer takes not only firm's self-interest and preferences but also opponents' perspectives into consideration. In a proposed offer/counteroffer, it indicates not only the region of acceptable solutions and preference degrees but the possibility of conflicts in the region. For each FCSP, incremental propagation eliminates the redundant decision values and infeasible combination of solutions. According to the ranking of solutions by fuzzy constraints, the set of feasible solutions can be further restricted in a preferred region with an acceptable/controlled threshold. It supports agent to quick response the changes in the environment and promises the proposed offers/counteroffers to be focused within the interest/attention area.

By sharing limited information among agents to incrementally uncovering the intention of opponents, consensus can be obtained and the quality of solution can be guaranteed at a satisfaction level as well. Moreover, the framework of AFCN also provides flexibility to incorporate negotiation strategies, such as self-interested, cooperative, and win-win, for various global performance measures.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical basis for modeling supply chain planning and scheduling problem as a distributed fuzzy constraint satisfaction problem (DFCSP). Section 3 then presents an agent-based fuzzy constraint-directed negotiation mechanism in detail. Then, Section 4 provides the experimental results to show the effectiveness of the proposed approach followed by conclusions in Section 5.

2 Modeling Supply Chain Planning and Scheduling as DFCSP

During the process of production/service provision from acquiring material to deliver to the end customers, it passes through the stages of raw material supply, intermediate supply, manufacturing, distribution, and retail. This paper focuses on the stage of manufacturing in supply chain. The supply chain project scheduling problem, as shown in Figure 1, is to schedule multiple projects in a network of manufacturers and suppliers (or contractors). Each project involves a set of tasks or operations with complex precedence relationships in which the task can be performed by a set of alternative suppliers. Suppliers differ from each other in terms of resource capacity, processing times, and costs of performing tasks. The process of project planning and scheduling specifies how the suppliers are selected from alternatives and tasks are scheduled. The problem is multi–objective in nature. For example, manufacturers wish to optimize the scheduling performance as well as to minimize the operation cost. On the other hand, the suppliers are more inclined to maximize the utilization and profit.

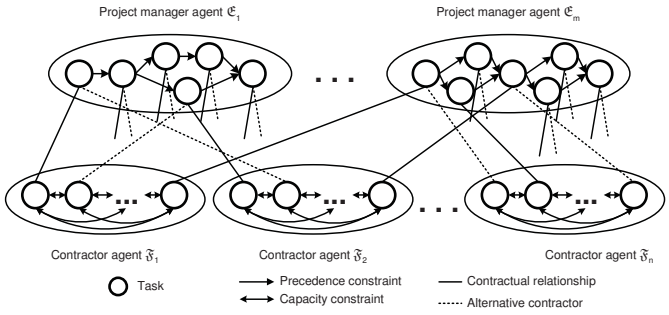


Figure 1: Supply chain project scheduling with m project agents and n contractor agents.

The problem can be represented as a triple $(\mathfrak{E}, \mathfrak{F}, \mathfrak{I})$, where $\mathfrak{E}$ is a set of m project manager agents (PAs) for the manufacturers, $\mathfrak{F}$ is a set of n contractor agents (CAs) for the suppliers, and $\mathfrak{I}$ is a set of inter-agent constraints between two class of agents. Each PA consists of a chain of tasks which are specified further by precedence constraints, including the processing time on the set of alterative suppliers for each task, due date, arrival (release) date, and tardiness cost. Each CA consists of capacity constraints and processing cost. Thus, the problem can be modeled as a distributed constraint satisfaction problem (DFCSP) in that coming up to a mutually acceptable solution between two classes of agents is the same as uncovering a consistent solution satisfying all the constraints in a distributed fuzzy constraint network specifying the fuzzy relationships inside each agent and between agents. Adapted from [9], a distributed fuzzy constraint network (DFCN) can be defined as below.

Definition 1
A distributed fuzzy constraint network $(\mathcal{U}, \mathbf{X}, \mathbf{C})$ *for a supply chain project scheduling problem* $(\mathfrak{E}, \mathfrak{F}, \mathfrak{I})$ *can be defined as a set of* $m+n$ *fuzzy constraint networks* $\{\mathfrak{N}_1, \ldots, \mathfrak{N}_{m+n}\}$, $\mathfrak{N}_l$ *representing an agent* $l \in \{\mathfrak{E}, \mathfrak{F}\}$, *where*

- $\mathcal{U}_l$ *is a universe of discourse for an FCN* $\mathfrak{N}_l$*;*
- $\mathbf{X}_l$ *is a tuple of non-recurring objects of agent* l*;*
- $\mathbf{C}_l$ *is a set of fuzzy constraints which involves a set of internal fuzzy constraints existing among objects in* $\mathbf{X}_l$*, and a set of external fuzzy constraints* $\mathfrak{I}_l$ *between agent* l *and opponent agents;*
- $\mathfrak{N}_l$ *is connected to other FCNs by a set of external fuzzy constraints* $\mathfrak{I}_l$*;*
- $\mathcal{U}$ *is a universe of discourse;*
- $\mathbf{X} = \left(\cup_{l=1}^{m+n} \mathbf{X}_l\right)$ *is a tuple of all non-recurring objects;*
- $\mathbf{C} = \left(\cup_{l=1}^{m+n} \mathbf{C}_l\right)$ *is a set of all fuzzy constraints.*

In Definition 1, a set of $\mathbf{X}_l$ of agent l is corresponded to its beliefs, its knowledge of the environment (tasks, resources, etc.), and any other attitudes (desires, intentions, opponents' response, etc.). A set of fuzzy constraints $\mathbf{C}_i$ for project agent $\mathfrak{E}_i$ is corresponding to a set of restrictions (e.g. precedence

ISBN: 978-989-95079-6-8

constraints), objectives (e.g. flowtime, operating cost) and inter-agent constraints between project agent $\mathfrak{E}_i$ and related contractor agents; $\mathbf{C}_k$ for contractor agent $\mathfrak{F}_k$ is corresponding to a set of restrictions (e.g. capacity constraints), objectives (e.g. utilization) and inter-agent constraints between contractor agent $\mathfrak{F}_k$ and related project agents.

By Definition 1, a consistent solution of a DFCSP for scheduling is an instantiation of all time allocation of the tasks such that all the constraints of the agents are satisfied at a degree of membership function. No agent knows about other agents' feasible solutions and possible agreements a priori. An agent negotiation mechanism is to explore potential agreements and then to move the negotiation toward a globally beneficial solution.

3 Agent-based fuzzy constraint-directed negotiation mechanism

Fuzzy constraint-directed approach has been demonstrated in [10, 11, 12, 9, 13] as an effective framework for agent negotiation. During the negotiation, each agent plans the solution by solving its own FCSP, and exchanges the proposals among the agents for resolving the inconsistency of time allocation of activities. As a proposal cannot be accepted by its opponents, a counter-proposal will be made according to the negotiation strategies which consider alternative solutions at the same constraint satisfaction level or offer a solution with a lower constraint satisfaction level. Exchanging proposals and counter-proposals will continue until termination conditions (e.g., achieving a consensus or a failure) are met.

In the proposed negotiation mechanism, the process of proposing a offer/counter-offer in an agent is regarded as the inference process for solving its own FCSP. The process includes the following steps: opponent responsive state evaluation, internal state update, behavioral state determination, set of feasible solutions generation, prospective solution selection and offer/counter-offer generation. The process of proposing an offer $\mathbf{A}^*_{i,j}$ with the FCN $\mathfrak{N}_{i,j} = (\mathcal{U}_{i,j}, \mathbf{X}_{i,j}, \mathbf{C}_{i,j})$ for the task j in project manager agent $\mathfrak{E}_i$ in a negotiation round is described as follows.

While receiving the counter-offer $\mathbf{B}_{i,j}$ over negotiation issues $\mathbf{I}_{i,j} \in \mathbf{X}_{i,j}$ from opponent agents, the opponent responsive state $\sigma_{i,j}$, which indicates the difference between current self-interest and opponents' perspectives, is obtained by

$$\sigma_{i,j} = 1 - ((\widehat{D}_{i,j} - D_{i,j})/(\widehat{D}_{i,j})), \tag{1}$$

where $D_{i,j}$ is the distance between the offer $\mathbf{A}'_{i,j}$ generated in last round and the latest counter-offer $\mathbf{B}_{i,j}$ and $\widehat{D}_{i,j}$ is the basis of distance for normalizing which is obtained at the first negotiation round.

The distance D between an offer $\mathbf{A}$ and a counter-offer $\mathbf{B}$ over set of issues $\mathbf{I}$ is obtained from

$$D = \frac{1}{N_I}\sqrt{\sum\nolimits_{k=1}^{N_I} \mathcal{L}(A^k, B^k)^2}, \tag{2}$$

where $\mathcal{L}$ is a distance measure for two fuzzy set, A^k and B^k is the fuzzy set for the offer $\mathbf{A}$ and counter-offer over issue $I^k \in \mathbf{I}$, respectively.

As none of the counter-offers can be accepted by the project manager agent $\mathfrak{E}_i$, agent $\mathfrak{E}_i$ has to decide the behavioral state to proposal alternative solutions in the current preferred region or offer a solution with a lower constraint satisfaction level. The internal states of agent and the opponent responsive state are involving to determine the behavioral state. In this manner, the internal states include the satisfaction level $\rho_{i,j}$ for task j and degree of tightness in solution space $\delta_{i,j}$. The aggregated satisfaction value $\rho_{i,j}$ for task j indicates the satisfaction level for the current prospective solution $\mathbf{S}'_{i,j}$ and is obtained by

$$\rho_{i,j} = \Psi(\mathbf{S}'_{i,j}) = \frac{1}{N_G}\sum\nolimits_{k=1}^{N_G} \mu_{\mathbf{C}_k}(\mathbf{S}^k_{i,j}), \tag{3}$$

where $\mathbf{S}'_{i,j}$ is the prospective solution of task j obtained in last negotiation round, $\Psi(\mathbf{S}'_{i,j})$ denotes the aggregated satisfaction value for the prospective solution, $\mu_{\mathbf{C}_k}(\mathbf{S}^k_{i,j})$ is the satisfaction degree of the fuzzy constraint $\mu_{\mathbf{C}_k}$ over goal $\mathbf{G}$.

The tightness of solution space $\delta_{i,j}$ indicates the remaining feasible solution space between the aggregated satisfaction value $\rho_{i,j}$ and the acceptable threshold $\epsilon'_{i,j}$ and is obtained by

$$\delta_{i,j} = 1 - (\rho'_{i,j} - \epsilon'_{i,j}), \tag{4}$$

where $\rho'_{i,j}$ is the aggregated satisfaction value, and $\epsilon'_{i,j}$ is the threshold of aggregated satisfaction determined in last negotiation round.

According to the internal state and the opponent responsive state, the level cut τ for the desire of concession $\mathbf{V}$ obtained by

$$\tau = (\mu_\rho(\rho'_{i,j}) \wedge \mu_\delta(\delta'_{i,j}) \wedge \mu_\sigma(\sigma_{i,j}))^{W_c}, \tag{5}$$

where $\mu_\rho(\rho'_{i,j})$, $\mu_\delta(\delta'_{i,j})$, and $\mu_\sigma(\sigma_{i,j})$ denote the desire of concession according to the degree of tightness, aggregated satisfaction and degree of difference, respectively; W_c denotes the weight associated with the desire of concession.

Then, the behavioral state regarded as the threshold of aggregated satisfaction $\epsilon_{i,j}$ is obtained from

$$\epsilon_{i,j} = \epsilon'_{i,j} - \Delta\epsilon = \epsilon'_{i,j} - \mathcal{D}(\mathbf{V}_\tau), \tag{6}$$

where $\epsilon'_{i,j}$ is the threshold of aggregated satisfaction obtained from last negotiation round, and $\Delta\epsilon$ is the concession value transformed from the fuzzy set of the desire of concession $\mathbf{V}$ with τ level cut. $\mathcal{D}$ is the defuzzification method.

To evaluate and generate the offer and counter-offer, an agent has to plan its prospective solution first. Each agent can only plan the prospective solution from its individual area of interests limited by the threshold of aggregated satisfaction $\epsilon_{i,j}$. That is, a set of feasible solution $\mathfrak{P}_{i,j}$ for task j in agent $\mathfrak{E}_i$ can be defined as

$$\mathfrak{P}_{i,j} = \{\mathbf{S}_{i,j} \mid (\mathbf{S}_{i,j} \in_\alpha \mathbf{\Pi}_{i,j}) \wedge (\Psi(\mathbf{S}_{i,j}) \geq \epsilon_{i,j})\}, \tag{7}$$

where ${}_\alpha\mathbf{\Pi}_{i,j}$ is the solution set with α level cut, $\Psi_{i,j}(\cdot)$ is the aggregated satisfaction value, and $\epsilon_{i,j}$ is the threshold of aggregation satisfaction degree of objectives.

Given the counter-offer $\mathbf{B}_{i,j}$ and the feasible solution set $\mathfrak{P}_{i,j}$, the prospective solution $\mathbf{S}^*_{i,j}$ is generated by solution selection method

$$\mathbf{S}^*_{i,j} = \arg_{\mathbf{S}}(\max_{\mathbf{S}\in\mathfrak{P}_{i,j}} \mathcal{H}(\mathbf{S}, \mathbf{B}_{i,j})), \tag{8}$$

where $\mathcal{H}$ is a utility function to evaluate the appropriateness between the feasible schedule $\mathbf{S}$ and the counter-offer set $\mathbf{B}$ in

ISBN: 978-989-95079-6-8

which the utility function $\mathcal{H}$ can be defined by

$$\mathcal{H}(\mathbf{S},\mathbf{B}) = \frac{1}{N_I}\sqrt{\sum_{k=1}^{N_I}(\mathcal{P}(S_k)^{w_p} \wedge \mathcal{S}(S_k,B_k)^{w_s})^2}, \quad (9)$$

where $\mathcal{P}$ is a satisfaction function over the issue I_k, and $\mathcal{S}$ is a similarity function (a distance measure) between the solution $\mathbf{S}$ and the counter-offer $\mathbf{B}$; w_p, and w_q denote the weight associated with the satisfaction and the similarity of the solution of task, respectively.

For the task j, given the feasible solution set $\mathfrak{P}_{i,j}$ and the prospective solution $\mathbf{S}^*_{i,j}$ which involves a set of values $\{\mathbf{S}^1_{i,j},\mathbf{S}^2_{i,j},...,\mathbf{S}^{N_I}_{i,j}\}$ for N_I issues, the offer $\mathbf{A}^*_{i,j}$ for task j is a tuple of fuzzy set $\{A^1_{i,j},A^2_{i,j},...,A^{N_I}_{i,j}\}$ in which each fuzzy set $A^k_{i,j}$ for issue $I^k_{i,j} \in \mathbf{I}_{i,j}$ is the marginal particularized possibility distribution

$$A^k_{i,j} = Proj_{X^k_{i,j}}(\mathfrak{P}_{i,j} \cap \overline{\Pi}_{X^1_{i,j}} \cap ... \cap \overline{\Pi}_{X^{k-1}_{i,j}} \cap \overline{\Pi}_{X^{k+1}_{i,j}} \cap ... \cap \overline{\Pi}_{X^{N_I}_{i,j}}), \quad (10)$$

where $\overline{\Pi}_{X^k_{i,j}}$ is the cylindrical extension of $\Pi_{X^k_{i,j}}$ in the space $(\mathbf{X}^1_{i,j},...\mathbf{X}^{N_I}_{i,j})$, $\Pi_{X^k_{i,j}} = \mathbf{S}^k_{i,j}$, and $\mathbf{X}^k_{i,j}$ is the object of issue $I^k_{i,j}$ for task j.

Meanwhile, when project manager agent $\mathfrak{E}_i$ receives a counter-offer $B_{i,j} = \{B^1_{i,j},B^2_{i,j},...,B^{N_I}_{i,j}\}$ for the task j, the opponent agents' preferred solution $\widehat{\mathfrak{P}}_{i,j}$ can be obtained by

$$\widehat{\mathfrak{P}}_{i,j} = (\overline{B}^1_{i,j} \cap ... \cap \overline{B}^{N_I}_{i,j}), \quad (11)$$

where $\overline{B}^k_{i,j}$ is the cylindrical extension of $B^k_{i,j}$ in the space $\mathbf{X}_l$. Each element $\widehat{\mathfrak{P}}$ represents a solution preferred by opponent agents and the membership degree of each element represents the acceptability of solution for all opponent agents.

The project manager agent $\mathfrak{E}_i$ will accept the counter-offer $B_{i,j}$ proposed by its opponent as an agreement if

$$(\mathfrak{P}_{i,j} \cap \widehat{\mathfrak{P}}_{i,j}) \neq \{\}, \quad (12)$$

The negotiation process for task allocation will be terminated when the project manager agent reach an agreement with one of contractor agents or the project manager agent or all of their contractor agents withdraws the negotiation. A project manager agent will terminate the negotiation when each of tasks in project is assigned to a contractor agent or the project manager agent withdraws the negotiation while failing to reach agreement. And the supply chain planning/scheduling process will be terminated when all of project manager agents are termination.

4 Experiments

To demonstrate the utility of the proposed model for project planning and scheduling problem in supply chain, the experiments are meant to compare AFCN with centralized heuristic (CTR) [14], conventional CNP (CNP) [15], modified CNP (MCNP) [4], market-based auction CNP (MA-CNP) (i.e., extended from CNP) and market-based auction MCNP (MA-MCNP) (extend from MCNP) (i.e., extended from MCNP) in terms of the number of project managers over makespan, total operating cost, and computational time.

The experiment was implemented on a Pentium M PC with 0.8 GHz running windows XP and 256 MB RAM. In the experiment, each project has a linear sequence of tasks, and each task is then specified by its precedence constraints, required resources, processing time and tardiness cost. Each supplier is specified by its unique capacity and processing cost. The total operating cost is the sum of the total processing cost and tardiness cost of each project. For simplicity, the number of tasks per project is identical at five, alternative suppliers for a task are identical at three, and 50 instances are generated for each number of project managers. Figures 2 to 4 show the performance comparisons over makespan, total operating cost, and computational time, respectively.

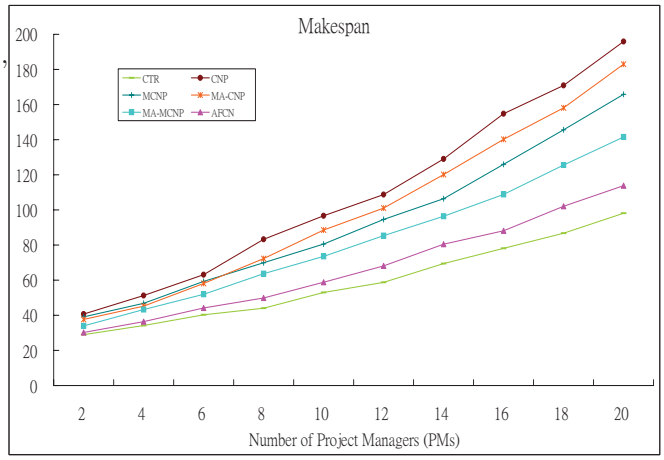


Figure 2: Makespan of various negotiation mechanisms over a number of project managers.

The conventional CNP supports one PA to assign a task at a time and another PA can start only after the PA finishes selecting CAs for all its tasks. In this form of negotiation model, a large number of slacks will be generated since that the conflicts among PAs cannot be coordinated. Each PA's performance highly depends on the order of contracting.

Instead of one PA assigning a task, MCNP supports multiple PAs to contract tasks to multiple CAs simultaneously. In the manner, PAs propose the whole time window of task and price to CAs. CAs resolve the conflicts among tasks according to the urgency of task be completed and the price of the task. It reduces the possibility of violating the due date of the PAs and decreases the tardiness when a project is delayed.

Since the simplicity of negotiation protocol, the single-shot negotiation models (CNP and MCNP) have the best time efficiency in Figure 4. With very limited information sharing and myopic decision-making, however, the global performances can easily get trapped at local optima and are highly unstable. It also can be observed in Figures 2 and 3, these models worse results than other approaches over makespan and total operating cost.

For reflecting the service preference in PA and the task contention in CA, MA-CNP and MA-MCNP incorporate the market-based auction mechanism [16] which is characterized

ISBN: 978-989-95079-6-8

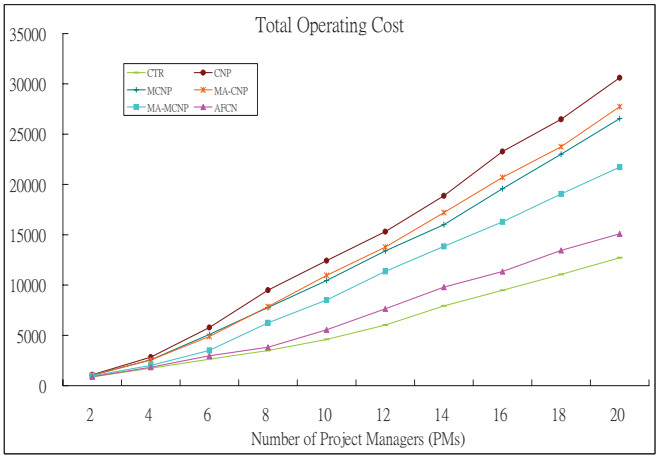


Figure 3: Operating cost of various negotiation mechanisms over a number of project managers.

by iterative bidding among agents. Agents will adjust the prices of bid according to the direction of surplus and deficit of demand. Task allocation processing in CA can be adjusted iteratively according to time boundary and the updating price of tasks. In this way, MA-MCNP with iterative bidding have better makespan and total operating cost than the single-shot negotiation models. However, due to the characteristic of protocol in MA-CNP, MA-CNP even worse than MCNP over makespan and total operational cost when the number of project managers is larger than 8 in Figure 2 and 3, respectively.

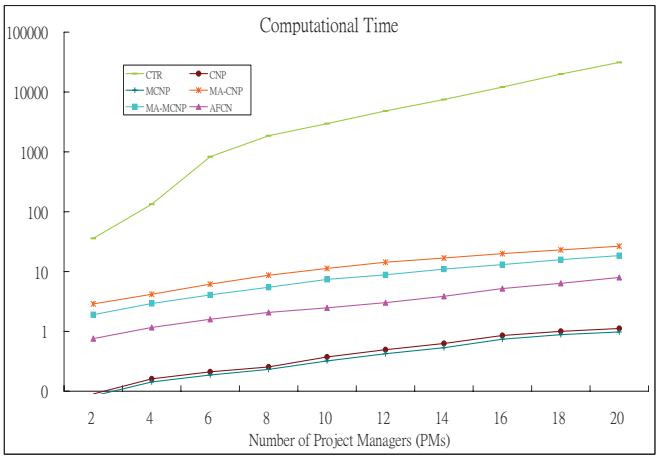


Figure 4: Computational time of various negotiation mechanisms over a number of project managers.

However, the price of task proposed by CA indicates the contention degree of the desired time slots of task, but not indicates where the contention can be resolved. PA might blindly increase price but the schedule of task can not be improved (overpricing). Besides, sharing full of task's time boundary in PA is not practical in reality.

In AFCN, PA generates the offer not only considers how much suppliers can provide this service, but how pressing the task be finished. Instead of the bid information in the market-based mechanism, the negotiation information in the AFCN varies not only depending on the particular combination of task demands/resource contracts, but the objectives and preferences of agent. Both of PA and CA proposes offers/counter-offers which involve the current feasible time slots of task with preference and price. The time slots with preference in offer/counter-offer indicate not the current time boundary of task, but the time window for conflict-free. According to time slots with preference from PA, CA can arrange the task allocation efficiency and can response the conflict in counter-offer directly. Meanwhile, CA generates the counter-offers considering not only the tasks of PAs are satisfied but the profit obtained in the transaction. Based on the counter-offers from CA, PA can avoid requiring the high contention area in CA or can abort lower quality service of suppliers. Thus, PA can bid the desired time slots to preferred suppliers with controllable budget.

In Figure 2 and 3, it can be observed that the AFCN has superior performance than other approaches in makespan and total operating cost. Meanwhile, AFCN performs better in time efficiency than that of market based mechanism. Additionally, as the problem size is increasing, the performance improved by the AFCN grows more significantly.

5 Conclusions

We have presented an agent-based fuzzy constraint-directed negotiation protocol for project planning and scheduling in supply chain. Constraint modeling gives a more direct fusion to the real-world problem descriptions and the impreciseness of knowledge in supply chain can easily be represented by fuzzy constraints with the levels of consistency as well. Experimental results suggest that the AFCN indeed can provide a practical and efficient framework for supply chain planning and scheduling.

References

[1] B. M. Beamon. Supply chain design and analysis: Models and methods. *International Journal of Production Economics*, 55(3):281–294, 1998.

[2] F. R. Lin, H. C. Kuo, and S. M. Lin. The enhancement of solving the distributed constraint satisfaction problem for cooperative supply chains using multi-agent systems. *Decision Support Systems*, 45(4):795–810, 2008.

[3] T. Calosso, M. Cantamessa, and M. Gualano. Negotiation support for make-to-order operations in business-to-business electronic commerce. *Robotics and Computer-Integrated Manufacturing*, 20(5):405–416, 2004. Cited By (since 1996): 7.

[4] J. S. K. Lau, G. Q. Huang, K. L. Mak, and L. Liang. Agent-based modeling of supply chains for distributed scheduling. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 36(5):847–861, 2006.

[5] Y. Aviv. The effect of collaborative forecasting on supply chain performance. *Management Science*, 47(10):1326–1343, 2001.

ISBN: 978-989-95079-6-8

[6] P. J. Byrne and C. Heavey. The impact of information sharing and forecasting in capacitated industrial supply chains: A case study. *International Journal of Production Economics*, 103(1):420–437, 2006.

[7] G. P. Cachon and M. Fisher. Supply chain inventory management and the value of shared information. *Management Science*, 46(8):1032–1048, 2000.

[8] L. Yan and T. G. Woo. Information sharing in a supply chain with dynamic consumer demand pattern. In *Proceedings of the Hawaii International Conference on System Sciences*, volume 37, pages 1143–1152, 2004.

[9] K. R. Lai. *Fuzzy Constraint Processing*. PhD thesis, NCSU, Raleigh, N. C., 1992.

[10] K. Y. Cai. Robustness of fuzzy reasoning and δ-equalities of fuzzy sets. *IEEE Transactions on Fuzzy Systems*, 9(5):738–750, 2001.

[11] D. Dubois, H. Fargier, and H. Prade. Propagation and satisfaction of flexible constraints. In *Fuzzy Sets, Neural Networks, and Soft Computing*, pages 166–187. Van Nostrand Reinhold Company, New York, 1994.

[12] D. Dubois, H. Fargier, and P. Fortemps. Fuzzy scheduling: Modelling flexible constraints vs. coping with incomplete knowledge. *European Journal of Operational Research*, 147(2):231–252, June 2003.

[13] J. Moura-Pires and H. Prade. Specifying fuzzy constraints interactions without using aggregation operators. In *Proceedings of the Ninth IEEE International Conference on Fuzzy Systems*, volume 1, pages 228–233, 2000.

[14] C. G. Li, J. B. Yu, and Liao. X. F. Fuzzy tabu search for solving the assignment problem. In *Communications, Circuits and Systems and West Sino Expositions, IEEE 2002 International Conference on*, volume 2, pages 1151–1155, June-1 July 2002.

[15] R. G. Smith. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, C-29(12):1104–1113, 1980.

[16] M.P. Wellman, W.E. Walsh, P.R. Wurman, and J.K. MacKie-Mason. Auction protocols for decentralised scheduling. *Games and Economic Behavior*, 14(5):429–439, 2001.

ISBN: 978-989-95079-6-8

Learning and Classification of Events in Monitored Environments

J. Albusac[1] J.J Castro-Schez[2] L.M Lopez-Lopez[2] D. Vallejo[2] L. Jimenez-Linares[2]

1.E.U.P.A, University of Castilla-La Mancha
Almaden, Spain
2.Escuela Superior de Informatica, University of Castilla-La Mancha
Ciudad Real, Spain
Email: {javieralonso.albusac, josejesus.castro, luis.jimenez, lorenzomanuel.lopez, david.vallejo}@uclm.es

Abstract—
This paper presents a prototype system to automatically carry out surveillance tasks in monitored environments. This system consists in a supervised machine learning algorithm that generates a set of highly interpretable rules in order to classify events as normal or anomalous from 2D images without needing to build a 3D model of the environment. Each security camera has an associated knowledge base which is updated when the environmental conditions change. To deal with uncertainty and vagueness inherent in video surveillance, we make use of Fuzzy Logic. The process of building the knowledge base and how to apply the generated sets of fuzzy rules is described in depth for a virtual environment.

Keywords— Automated Video Surveillance, Visual Information Analysis, Machine Learning, and Fuzzy Logic.

1 Introduction

The problem of intelligent surveillance deals with the perception, interpretation, and identification of activities and situations that occur in monitored environments [1]. A typical surveillance scenario consists of a set of CCTV cameras deployed on different places in order to perform explicit surveillance on behalf of people and information storage to carry out forensic analysis if needed [2]. When a security guard watches an environment by means of video cameras, he is able to detect if something is going wrong. In other words, he perceives video events and classifies them as normal or anomalous. If he detects an anomalous behaviour or event, he makes the appropriate decisions to solve the problem as soon as possible. However, the system depends on the human component to classify events and to continuously pay attention to the video stream, which is very tiring and, therefore, error-prone [3]. This is our motivation to face the design of a system able to classify behaviours, that is, to propose a method for automating this task in different domains and scenarios.

Events in real environments can be classified as simple or composite. Really, a composite event is a sequence of simple events which are temporally related [4, 5]. However, the analysis of object characteristics and spatial properties may be enough to understand simple events, i.e, without needing a temporal analysis. In this context, there are many elements which can be learned and identified from data sensors. For instance, the object physics characteristics to determine its class or object type [6], trajectories followed by moving objects [7], entry/exit areas [8], allowed proximity relationships between objects and areas, allowed speed for each type of object, etc.

On the other hand, it is necessary to study how to deal with the uncertainty in real-time to carry out surveillance. Normally, an artificial surveillance system cannot totally ensure what is happening from data sensors in most cases. Uncertainty, imprecision, and vagueness are frequently present when these systems try to solve real world problems. For this reason, one of the main problems of this type of systems is the high number of false alarms due to incorrect interpretations.

Several authors have addressed these issues in the literature. Foresti et al. [9] proposed a visual-based surveillance system for real-time event detection and classification, which is based on adaptive high order neural trees. This system carries out object classification, object tracking, and event recognition for understanding normal, suspicious, and dangerous events in parking lots. Buxton and Gong [10] also provided solutions to the problem of event recognition. They proposed techniques based on Bayesian networks for interpreting traffic situations in dynamic scenes. Remagnino et al. [11] described events in a surveillance scenario by using a Bayesian classifier instead of the Hidden Markov Model. The different techniques proposed in this work were employed to model the common event behaviours in car park. On the other hand, fuzzy logic [12, 13] may provide another interesting approach for dealing with the same problem [14]. In [15], a prototype fuzzy system for describing human activity in natural language is described. This description is made by analysing the possible relations among objects in a monitored environment. To do that, the linguistic labels and the fuzzy rules are defined by an expert to classify people behaviour.

The crafting of detailed scene models may provide an effective means of interpreting situations and detecting anomalous behaviours in a static scene, but it is not an appropriate approach in dynamic scenes where the environmental conditions change over time [7]. This is why in this work we use a fuzzy-based machine learning algorithm to generate a set of highly interpretable set of rules to classify real-time events. To do that, the system analyses spatial data from 2D images obtained by cameras in outdoor/indoor environments, and it generates sets of highly interpretable fuzzy rules. The system analysis situations in a similar way as human beings do, that is, it obtains conclusions without needing precise data. For instance, a person does not need to know the exact speed or the absolute position of an object to determine if it is moving fast on the zone in which it is located.

The remainder of this paper is organised as follows. Section 2 describes the problem statement and overviews the processes required to build the knowledge base of a camera and how to apply this knowledge to classify events. Section 3 de-

ISBN: 978-989-95079-6-8

scribes in depth the machine learning algorithm employed to build the different sets of fuzzy rules that compose the knowledge base. Section 4 studies how to apply the system in a well-defined environment. Finally, conclusions are presented in Section 5.

2 Problem Statement

As previously described, one of the main objectives of advanced surveillance systems is to interpret events in an environment from data sensors. Besides, events not only must be identified, but the system must be also able to classify events as normal or abnormal in order to make adequate decisions. This task covers several research areas, we attack two of them: i) how to build the surveillance knowledge base in a concrete scenario and ii) how to apply this knowledge to classify events as normal or abnormal.

On the other hand, a simple event can be defined as a concrete action that occurs at a time instant, and an anomalous simple event may be enough to activate an alarm in a monitored environment. The method proposed in this work determines the normality of a scene according to this type of events. Examples of anomalous simple events are as follows: a vehicle moving too fast in a concrete time instant or an object located in a forbidden zone. We characterise our surveillance domain by means of the following assumptions:

- Video stream is obtained from cameras placed on certain locations of the environment.
- Each camera has an own knowledge base.
- Every camera is fixed, that is, vertical or horizontal movements and zoom imply a new generation of the surveillance knowledge base.
- Video data can be imprecise.

Each video camera has its own knowledge base, which is composed of three set of fuzzy rules, and each set is generated by an independent training set:

1. Set of rules to determine the object's class. This information is critical when classifying events as normal or abnormal because it defines how objects should normally behave.
2. A second set of rules to determine the object's speed taking into account the object's class and its motion vectors obtained in the segmentation process.
3. Finally, a set of rules to infer whether a situation is normal or anomalous according to the object's class, speed, and the areas in which the object could be located.

To generate the first set of fuzzy rules, we employ the size and the position of the objects in the environment from 2D images (frames), and we take into account the camera view point. Thus, the system may learn, for instance, that vehicles are large or very large size objects when they are close to the camera, and small or medium size objects when they are far from the camera. The segmentation algorithm [16] used in this work determines the set of macroblocks in each frame for every object that appears in the scene. A macroblock is the basic unit in a MPEG stream and it is an area of 16 by 16 pixels in which the motion vectors are stored. The displacement between two macroblocks in different frames gives the motion vector and it specifies a distance and a direction. Each object is represented by means of an ellipse that involves the set of macroblocks. As it will be seen further (Section 4), the parameters of the ellipse are used to determine the location and the size of an object.

On the other hand, in order to generate the rules used to determine the object's speed, the position (centre of the ellipse), the displacement of the object between consecutive frames (distance between the centres of the ellipses), and the object class are used as input variables, i.e, each sample of the training set takes a value for every variable. Movements done by people and vehicles may imply different speed. A medium displacement for one person may involve high speed and for a car may involve low or medium speed. The main goal of this stage is to learn to interpret the object speed depending on their sizes and displacements by taking into account the camera point of view. Finally, the last set of fuzzy rules is generated from a training set where each sample represents the situation of an object in a concrete time instant, i.e, its class, speed, and the spatial relations with the zones or areas of the environment.

Once the rules have been generated and the knowledge base has been built, it can be used to classify simple events. The system analyses frames of a MPEG video stream, evaluating where are the objects located in the environment (2D position) and what are their size. From this information, the system obtains the class of each moving object. In this point, the surveillance system possibly knows where are the objects in each moment and what are their classes. Next, the system uses this information and the second set of rules to determine the speed of each object. Finally, the system determines the areas in which each object could be located and the information previously obtained (class and speed) to classify the simple events as normal or anomalous. In other words, the classification process can be defined as the chaining of three sets of fuzzy if-then rules in which the knowledge generated by a set is employed by the next one.

3 Suggested Machine Learning Algorithm

The machine learning algorithm described in this section is based on the work proposed by Castro *et al.* [17]. This method learns a set of maximal structure fuzzy if-then rules from a set of training data instances. This choice has been motivated by the following reasons: (a) The use of fuzzy logic provides a well-defined mathematical framework to deal with the uncertainty and vagueness inherently contained in the visual information used as input data. (b) This algorithm represents the inferred knowledge by means of a set of fuzzy if-then rules expressed in terms of linguistic variables. As a result, we obtain rules that are easily comprehensible by security guards, who are watching the monitors. Thus, an user can easily analyse why an alarm was activated. (c) As a supervised learning approach, once the set of rules has been inferred from the training data set, the algorithm does not require a significant amount of time to classify new data instances.

Basically, Castro's algorithm works in two phases. In the first one, each example of a training set is converted into a

ISBN: 978-989-95079-6-8

specific rule which describes how to act in a concrete situation, that is, how this example must be classified. Therefore, there will be so many particular rules as examples. In a second phase, the algorithm generalises these particular rules in order to act in a wide range of possibilities, that is to say, many examples of a class can be correctly classified by using one generalised rule. To do that, an amplification process extends the particular rules covering situations of the space of possible situations, which have not been covered before.

Formally, the original algorithm starts out the training phase from a set of data instances $\Theta = \{e_1, \ldots, e_m\}$. Each $e_i = ((x_{i1}, \ldots, x_{in}), y_j)$ is a data example which is conveniently described by means of the set of input variables $V = \{v_1, \ldots, v_n\}$ and one output variable y_j, which is called the *class* of e_i. The elements $(x_{i1}, \ldots, x_{in})$ are the concrete values that the example e_i takes for each variable $v_k \in V$. Besides, there is a domain definition DDV_i for each variable in V, being $DDV = \{DDV_1, \ldots, DDV_n\}$ the set of all domains. Each $DDV_i = \{L_1, \ldots, L_p\}$ is composed of a set of linguistic labels which correspond to the fuzzy sets represented by means of trapezoidal functions as specified in expression (1).

$$\prod(u; a, b, c, d) = \begin{cases} 0 & u < a \\ \frac{(u-a)}{(b-a)} & a \leq u < b \\ 1 & b \leq u \leq c \\ \frac{(d-u)}{(d-c)} & c < u \leq d \\ 0 & u > d \end{cases} \quad (1)$$

In this way, each DDV_i verifies the following properties:

1. $\forall L_x \in DDV_i, height(L_x) = 1$
2. $\forall L_x, L_y \in DDV_i, nucleus(L_x) \cap nucleus(L_y) = \emptyset$
3. $\forall x \in X_i, \sum_{j=1}^{|DDV_i|} \mu_{L_j}(x) = 1$, being X_i the domain where v_i is defined.

Besides, the inferred knowledge is represented as a set of fuzzy if-then rules with the following structure:

$$\textbf{if } v_0 \textit{ is } ZD_0 \wedge \ldots \wedge v_n \textit{ is } ZD_n \textbf{ then } y_j \quad (2)$$

where $v_i \in V$ and $ZD_i \subseteq DDV_i$ is a subset of the linguistic labels defined in DDV_i for the variable v_i.

The original algorithm converts each $e_i \in \Theta$ into the fuzzy domain, according to the values $(x_{i1}, \ldots, x_{in})$ that e_i takes for each $v_i \in V$ and their corresponding $DDV_i \in DDV$. For this purpose, each $L_k \in DDV_i$ has associated a function $\mu_{L_k} : X_j \rightarrow [0, 1]$, being X_j the domain where the variable v_j takes its values (i.e $\Re, \aleph$, an interval $[a, b]$, a finite set A, etc). Thus, every $e_i \in \Theta$ is converted into $e'_i = ((L_{1x}, \ldots, L_{nz}), y_j)$ such that L_{jk} is the label which matches best, according to $L_{jk} = max\{\mu_{L_{jk}}(x_j)\}$ of the DDV_j. Each e'_i is an initial rule which belongs to the set of *initial rules*. Second, if some rule of the initial set does not subsume in any rule of the final set, the algorithm proceeds to amplify the rule. A rule R_i can be amplified to $R_{i'}$: **if** v_0 *is* $ZD_{i'0} \wedge \ldots \wedge v_n$ *is* $ZD_{i'n}$ **then** y_p if there is no rule R_j : **if** v_0 *is* $ZD_{j0} \wedge \ldots \wedge v_{jn}$ *is* ZD_n **then** y_q in the set of initial rules that verify $ZD_{jk} \subseteq ZD_{i'k}$ and $y_p \neq y_q$. In other words, an amplification is possible whenever there is no counterexample that conflicts with the amplified rule. This makes possible for this method to generate a set of rules that are as general as possible. A description in detail of the original algorithm by Castro *et al.* can be found in [17].

However, this algorithm may cause troubles when applied in the surveillance context due to the production of rules that are over-generalised. With over-generalised we refer to the tendency of the final rules to cover spaces for which there are no counterexamples. For example, let us suppose three initial rules for classifying a moving object according to its size by taking into account the distance from the camera:

- R_i: **if** $size$ is {SMALL} $\wedge$ $distance$ is {MEDIUM} **then** y_1 is people
- R_j: **if** $size$ is {MEDSMALL} $\wedge$ $distance$ is {MEDIUM} then y_1 is motorbike
- R_k: **if** $size$ is {MEDBIG} $\wedge$ $distance$ is {MEDIUM} then y_1 is car

all of them contained in the set of initial rules. Let DDV_{size} = {VSMALL, SMALL, MEDSMALL, MEDBIG, BIG, VBIG}, if we try to amplify the rule R_i to achieve $R_{i'}$, according to the variable **size**, it could result in:

- $R_{i'}$: **if** $size$ is {VSMALL, SMALL, BIG, VBIG} $\wedge$ $distance$ is {MEDIUM} **then** y_1 is people

Note that $R_{i'}$ means that if an object is located at a *medium* distance from the camera and its size is *very small*, *small*, *big*, or *very big*, it certainly belongs to the class *people*. However, R_j and R_k mean that if an object is located at a *medium* distance from the camera and its size is *medium small* or *medium big*, the object is a *motorbike* or a *car*, respectively. Obviously, $R_{i'}$ lacks of any sense and it can cause problems because of the misclassification of objects. This happens because when amplifying R_i to $R_{i'}$, the labels $VSMALL$, BIG, and $VBIG$ are added to $R_{i'}$ as there are no counterexamples in the set of rules that conflict with it. Moreover, this type of counterexamples might never exist in the training set because it is possible that cars and motorbikes have never a big or very big size, when they are located at a medium distance from the camera.

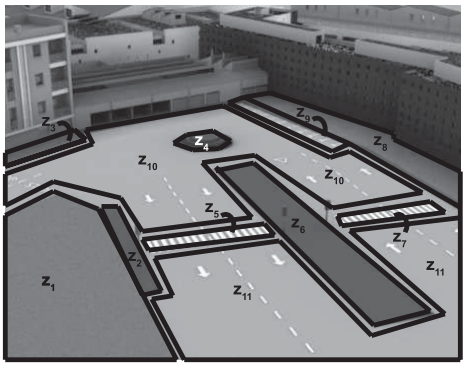
In order to solve this issue, we have modified the original algorithm to restrict the amplification of rules. We are interested in amplifying a rule only if the last label added to the rule in the last amplification step and the label being considered to be added are not too far from each other. Thus, we need to use a measure of separability s to estimate the dissimilarity between linguistic labels L_i and L_j (with $L_i < L_j$), which are expressed in terms of fuzzy sets:

$$s(L_i, L_j) = \frac{(b_{Lj} - c_{Li}) + (a_{Lj} - d_{Li})}{2} \quad (3)$$

being $(a_{Li}, b_{Li}, c_{Li}, d_{Li})$ and $(a_{Lj}, b_{Lj}, c_{Lj}, d_{Lj})$ the trapezoids that define the fuzzy sets L_i and L_j respectively, and verifying that $d_{Li} < a_{Lj}$. If $d_{Li} = a_{Lj}$ or $d_{Li} > a_{Lj}$, then $s(L_i, L_j) = 0$, as the area of the fuzzy set $s(L_i, L_j)$ is not significant enough.

The separability threshold is empirically calculated by observing how the algorithm selects the different values of percentage of the maximum separability between the fuzzy sets

ISBN: 978-989-95079-6-8



Zone Classification	
z_1	Garden
z_2	Sidewalk
z_3	Garden
z_4	Traffic Circle
z_5	Pedestrian Crossing
z_6	Sidewalk
z_7	Pedestrian Crossing
z_8	Garden
z_9	Parking
z_{10}	Road
z_{11}	Road

Figure 1: Definition and classification of areas for the analysed scene.

of a given linguistic variable. If we apply the expression (3) with L_i being the first label and L_j the last label of a linguistic variable, we obtain the maximum possible distance value for that variable. Then, it is possible to choose the 25% or the 10% of such separability as the threshold for amplifying a rule. Note that when the maximum allowed separability is 100%, the new version of the algorithm works as the original one.

Next, the steps of the modified algorithm are described in detail.

1. Convert each training example into an initial fuzzy rule (translation into the fuzzy domain). Each element $e_i \in \Theta$ is translated into a fuzzy rule in which the value of each input variable is represented by means of a linguistic label. This step creates the set of *initial* rules.
2. Take a rule from the set of initial rules.
3. Try to subsume the taken rule in some rule of the set of definitive rules. If that happens so, ignore the taken rule and go back to step 2.
4. If the taken rule does not subsume in any rule of the definitive set, try to amplify it. For each variable:
   (a) For each unconsidered label:
      i. Try to amplify the rule. If it is not possible, go to step 4.a; otherwise, proceed to step 4.a.ii. One rule can be amplified only if:
         A. There is no R_j : **if** v_0 *is* ZD_{j0} $\wedge$ ... $\wedge$ v_n *is* ZD_{jn} **then** y_q in the set of initial rules that $ZD_{jk} \subseteq ZD_{i'k}$ and $y_p \neq y_q$ (we maintain the constraint of the original algorithm).
         B. The separability between the last label added to the rule and the label being considered for amplifying the rule does not exceed a separability threshold.
      ii. Amplify the rule and include it in the set of definitive rules.
5. If there are still unconsidered rules in the initial set of rules, go to step 2. Otherwise, **END**.

4 Applying the Suggested Algorithm to an Example

The proposed machine learning algorithm has been tested on the virtual environment shown in Figure 2 due to the following reasons: (a) the difficulty of obtaining a wide range of real video scenes with anomalous situations, (b) the possibility of generating all abnormal situations as needed without attempting on people security, and (c) the freedom of changing camera position with no configuration cost.

Figure 1 shows the scenario used in this work, which represents a typical urban environment composed of buildings, roads, gardens, traffic signals, and so on. There are also pedestrian areas in which vehicles are not allowed to drive. In the same way, there are only vehicle areas in which pedestrians should not walk.

In our particular application, there are three training sets: Θ_1, Θ_2, and Θ_3, one for each phase of the learning process. Each training set is composed of a set of examples where every one of them e_i is made up from a set of features extracted from the 2D images captured by a video camera, such as the position of a moving object in a frame or the size of the ellipse that contains it. The algorithm described in Section 3 is performed for every training set Θ_i, in order to obtain a set of fuzzy rules. On the other hand, there are three output variables: the class of the moving object (y_1), its speed (y_2), and whether its behaviour is normal or anomalous (y_3).

4.1 Learning Process

As described before, the first step consists in learning the set of rules to determine the object class. Determining the object class is crucial to know if its behaviour in a monitored environment is normal. Each one of these classes has a set of norms or rules. If an object meets the norms associated to its class, then its behaviour will be considered as normal. To do that, we first need to build a training set in which each sample is defined in the following way: $e_i = ((x_{i1}, \ldots, x_{in}), y_j)$, being $(x_{i1}, \ldots, x_{in})$ the variable values used to learn and y_j the output variable. The variables employed in this first phase are $((X_{pos}, Y_{pos}, R_h, R_v)\, y_1)$. X_{pos} represents the moving object horizontal position and Y_{pos} represents the moving object vertical position. The values of X_{pos} and Y_{pos} comes from the coordinate x, y of the ellipse central point that encloses the object. On the other hand, R_h determines the horizontal size measured in pixels of the ellipse that encloses the object and R_v the vertical size measured also in pixels. Finally, y_1 is the output variable, which can take the following values: $\{pedestrian, vehicle\}$.

Each sample refers to a moving object detected in the segmentation process, which has been marked with an ellipse. Precisely, the ellipse parameters (central point coordinates, horizontal radius, and vertical radius) are used to generate the rules in this step. The goal is to determine whether a certain object is a pedestrian or a vehicle depending on its size and position from the camera point of view. Distinguishing be-

ISBN: 978-989-95079-6-8

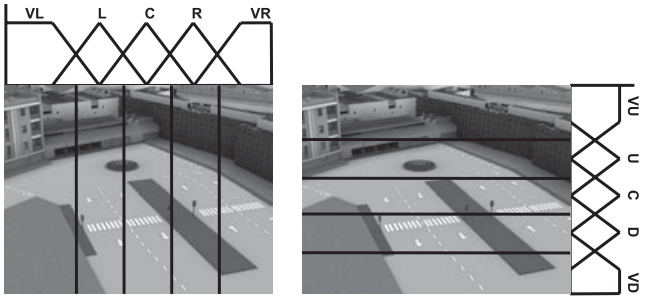


Figure 2: Vertical-horizontal scene division from an image captured by a surveillance camera.

tween people and vehicles depending on size and position is feasible because the difference between them is notable independently of the area in which they are located. On the other hand, to determine the zone in which an object is placed, a vertical-horizontal division as shown in Figure 2 is carried out. In this environment, the far zones from camera position, where objects may move, are located in $X_{pos} = \{\text{centre, right}\}$ and $Y_{pos} = \{\text{up}\}$. However, the close zones from camera position are located in $X_{pos} = \{\text{very left, left, centre}\}$ and $Y_{pos} = \{\text{very down, down}\}$.

An example of a fuzzy rule generated in this stage is as follows:

R : if $Xpos$ is {very right} and $Ypos$ is not {very up, up} and Rh is {very small, small} and Rv is {very small, small} then y_1 is pedestrian

Once applied the first stage, the next step consists in determining the object speed depending on its position and movement (measured in pixels) between the current and previous frame. We consider that two frames per second for the studied environment is enough for detecting anomalous situations. For every object class and zone there is a range of allowed speed values, which should be met by every object. Therefore, the speed study is interesting for the analysis of behaviours in monitored environments. In order to learn the second set of rules, we use input examples with the following form: $((X_{pos}, Y_{pos}, y_1, Mov)\ y_2)$, being Mov the movement of the object measured in pixels between two consecutive frames, and y_2 the output variable which can take the possible following values: $\{slow, normal, fast\}$.

An example of a rule generated in this step is the following:

R : if $Xpos$ is {very left, left} and $Ypos$ is {very down, down} and y_1 is {pedestrian} and Mov is not {medium, low, very low} then y_2 is slow

Finally, the last step in the general learning process consists in generating a set of rules that allows the surveillance system to detect if an object behaves normally or anomalously depending on the object class, its speed and the zones in which it could be located. As in previous stages, a training set is built for the rule generation. Every input sample is represented in the following way: $((y_1, y_2, Garden, Sidewalk, Trafficcircle, Parking, PedestrianCrossing, Road)\ Y_3)$, where every kind of zone represents an input variable whose value sets the intersection grade between the object and the type of zone. Finally, y_3 is the output variable whose values can be $\{normal\ situation, anomalous\ situation\}$.

An example of two rules generated in this final step is:

R : if y_1 is {pedestrian} and y_2 is {slow, normal} and $garden$ is {very high} and $road$ is {out, very low, low} then y_3 is normal situation

R : if y_1 is {pedestrian, vehicle} and y_2 is {fast} and $road$ is {low, medium, high} then y_3 is anomalous situation

4.2 Classification process

Once the learning process is done, the set of rules is available to determine, in each frame, the normality of an object behaviour. To carry out the classification process, first it is necessary to obtain a set of low-level objects data. Next, we expose the description of such data and how to obtain it.

For each frame, the system performs a segmentation process to detect every moving object, which is then bounded within a ellipse. The parameters of the ellipse are employed to determine the position and the size of the bounded object (X_{pos},Y_{pos},R_h,R_v). If an object is moving, the position of its bounding ellipse in the current frame changes respecting to the previous frame. Thus, the movement of an object is calculated through the distance between the origin of the previous and the current ellipses.

To complete the low level data acquisition of the objects, the system performs a proximity analysis of every tracked object to the areas of the environment which are defined according to the camera point of view. The goal of this analysis is to learn in which areas can be located a concrete object class. Thus, an expert is required to define such areas through the definition of their corresponding bounding polygons and the class which every area belongs to. Figure 1 shows the definition and classification of areas performed for the analysed scene.

The next step is to study the intersection between the objects bounding ellipses and the defined areas to learn in which extent an objects is on an area. For this purpose, the system calculates the amount of points in common in an object's bounding ellipse and the polygon defining an area. To do that, we use the following algorithm [18] represented in the programming language C:

```
int pnpoly(int nvert, float *vertx, float *verty,
           float testx, float testy)
{
  int i, j, c = 0;
  for (i = 0, j = nvert-1; i < nvert; j = i++) {
    if ( ((verty[i]>testy) != (verty[j]>testy)) &&
         (testx < (vertx[j]-vertx[i]) * (testy-verty[i])
         / (verty[j]-verty[i]) + vertx[i]) )

         c = !c;
  }
  return c;
}
```

where *nvert* is the number of vertices in the polygon, *vertx* and *verty* the arrays containing the x- and y-coordinates of the polygon vertices and, finally, *testx* and *testy* are the x- and y-coordinate of the test point (the algorithm is performed for each point of the ellipse). The amount of points in common is converted into the fuzzy domain and is classified as *out*, *very low*, *low*, *medium*, *high*, or *very high*.

The classification process proceeds in each frame and it begins once the system knows the low level data of every object in the scene. The system makes use of the three sets of fuzzy rules to determine the object's class and speed, and if its behaviour is normal or, on the contrary, anomalous.

ISBN: 978-989-95079-6-8

In order to evaluate the learning method proposed in this paper, we have run ten tests for every step of the general process. Each training set had a different number of examples, and the 80% of them were used for learning and the rest were used in the classification process. Table 1 resumes the results obtained in the tests, where the average of the generated rules, examples correctly classified, classification errors, and percentage of successful are shown for every step.

Table 1: Results obtained in the tests

Step	Gen. Rules	Correct	Wrong	% Correct
step 1	29	43	2	96.28
step 2	22	79	3	96.34
step 3	50	120	5	96

5 Conclusions

Surveillance systems are being used in a wide range of environments which require more and more sophisticated solutions. In fact, a notable investment has been made during the last years in order to provide surveillance services, both in public and private environments, which increase the efficiency of such systems and manage high-level information to allow users to adequately make decisions and manage crisis situations. In this paper, we have presented a possible solution to classify and describe simple events related to spatial properties in a monitored environment having into account the camera perspective. Moreover, we have also determined whether these events are normal or not.

We have proposed a machine learning algorithm to acquire the necessary knowledge from examples of situations and a method that uses this algorithm. The proposed method includes three calls to the algorithm as has been explained in Section 2.

Our work starts from the fuzzy machine learning algorithm proposed by Castro et al. [17]. A key innovation has been added to this algorithm with the aim of adapting it to the new needs of our problem. This was primarily motivated due to that several generated rules with different consequences could be applied to the same premises, generally scenarios that are not present in the training data. This is due to the generalisation of each particular rule into a definitive rule. The modified algorithm prunes the over-generalisation of a definitive rule during the amplification process. As has been described in Section 3, the over-generalisation is not desirable for us and it will be controlled by means of a separability measurement. This measurement allows one definitive rule to capture the premises that are not present in the training data but close to those evidences in the data that justify that rule. Thanks to this separability measurement the results obtained are improved.

Finally, we want to remark that the separability threshold is empirically calculated by observing how the algorithm selects the different values of percentage of the maximum separability between the fuzzy sets of a given linguistic variable. For this reason, one of our lines of future research consists in designing a new algorithm to decide the best separability value to optimise the results.

Acknowledgment

This work has been founded by the Regional Goverment of Castilla-La Mancha under Research Projects PII2I09-0052-3440 and PII1C09-0137-6488.

References

[1] W. Wang and S. Maybank. A survey on visual surveillance of object motion and behaviors. *Systems, Man and Cybernetics, Part C, IEEE Transactions on*, 34(3):334–352, 2004.

[2] M. Valera and S.A. Velastin. Intelligent distributed surveillance systems: a review. In *Vision, Image and Signal Processing, IEE Proceedings-*, volume 152, pages 192–204, 2005.

[3] G.J.D. Smith. Behind the screens: Examining constructions of deviance and informal practices among cctv control room operators in the uk. *Surveillance and Society*, 2(2/3):376–95, 2004.

[4] J.F. Allen. An interval-based representation of temporal knowledge. In *Proc. 7th International Joint Conference on Artificial Intelligence, Vancouver, Canada*, pages 221–226, 1981.

[5] J. Allen and G. Ferguson. Actions and events in interval temporal logic. In *Journal of Logic and Computation*, volume 4(5), pages 531–579, 1994.

[6] F. Fusier, V. Valentin, F. Bremond, M. Thonnat, M. Borg, D. Thirde, and J. Ferryman. Video understanding for complex activity recognition. *Machine Vision and Applications*, 18(3):167–188, 2007.

[7] H.M. Dee and D.C. Hogg. Navigational strategies and surveillance. In *Proceedings of the IEEE International Workshop on Visual Surveillance*, pages 73–81, 2005.

[8] D. Makris and T. Ellis. Path detection in video surveillance. *Image and Vision Computing*, 20(12):895–903, 2002.

[9] G.L. Foresti, C. Micheloni, and L. Snidaro. Event classification for automatic visual-based surveillance of parking lots. *Proc. of the 17th International Confrence on Pattern Recognition*, 3:314–317, 2004.

[10] H. Buxton and S. Gong. Visual surveillance in a dynamic and uncertain world. *Artificial Intelligence*, 78(1-2):431–459, 1995.

[11] P. Remagnino and GA Jones. Classifying Surveillance Events from Attributes and Behaviour. In *the Proceeding of the British Machine Vision Conference*, pages 10–13.

[12] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[13] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, Jan-Feb 1985.

[14] P. Angelov and X. Zhou. Evolving Fuzzy-Rule-Based Classifiers From Data Streams. *IEEE Transactions on Fuzzy Systems*, 16(6):1462–1475, 2008.

[15] H. Stern, U. Kartoun, and A. Shmilovici. A prototype Fuzzy System for Surveillance Picture Understanding. *IASTED International Conference Visualization, Imaging, and Image Processing (VIIP 2001), Marbella, Spain*, pages 624–629, 2001.

[16] L. Rodriguez-Benitez, J. Moreno-Garcia, J.J. Castro-Schez, J. Albusac, and L. Jimenez-Linares. Automatic objects behaviour recognition from compressed video domain. *Image and Vision Computing*, 27(6):648 – 657, 2009.

[17] J.L. Castro, J.J. Castro-Schez, and J.M. Zurita. Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems. *Fuzzy Sets and Systems*, 101(3):331–342, 1999.

[18] M. Shimrat. Algorithm 112: Position of point relative to polygon. *Communications of the ACM*, 5(8):434, 1962.

ISBN: 978-989-95079-6-8

A Hybrid Method for Soccer Video Events Retrieval Using Fuzzy Systems

Ehsan Lotfi[1], M.-R. Akbarzadeh-T[2], H.R. Pourreza[3], Mona Yaghoubi[1] and F. Dehghan[1]

[1]Department of Computer Engineering, Islamic Azad University,
Mashhad Branch, Iran.
[2]Departments of Electrical Engineering and Computer
Engineering, Ferdowsi University of Mashhad, Iran and IEEE Senior Member.
[3]Departments of Computer Engineering, Ferdowsi University of Mashhad.
esilotf@gmail.com, Akbarzadeh@ieee.org, hpourreza@um.ac.ir, mona.yaghoobi@gmail.com

Abstract— *Most information retrieval systems make indirect use of human knowledge in their retrieval process. The new method we present here aims to efficiently use human knowledge directly in combination with support vector machines for clustering. As illustrated in this paper, this approach is particularly applicable to concept retrieval from soccer-related videos. The first phase consists of extracting suitable features from video shots. Then, using a fuzzy rule base containing the experiences of experts, shots that do not include significant events are removed. Finally, the last phase uses SVM to classify results coming from the fuzzy system. The results of the classification phase are accompanied by a textual description and enables retrieval through text based query. Experimental result show good classification and satisfying retrieval process.*

Keywords— Event Retrieval, Semantic Indexing, Hybrid Method, Fuzzy System, SVM, Soccer Video.

1 Introduction

Videos are used for different purposes such as education, mass media, entertainment, and surveillance systems. In each of these applications a video conveys a significant message to the audience. For example, in a football match video events such as penalties, corners, goals, and the very concept of victory or loss are conveyed when fans watch the video. The meanings contained in a video are quite obvious for men, but the same does not go for computers. A computer cannot understand the scoring of a goal in a football match video. This fact is usually referred to as the semantic gap between man and computers. Imagine if computers were capable of detecting and understanding the meaning of videos then we could use them to summarize sport videos automatically and retrieving meanings from a multimedia database. On the other hand the dramatic increase of multimedia data volumes in recent years will eventually force us to use some sort of retrieval system so that the indexing and retrieval of information is handled automatically. As noted earlier there is a semantic gap between the user and retrieval systems. The user queries the system with his/her ideas and the system presents it findings for such ideas as an answer. Such queries are done in different manners: 1. Query by keywords[11] 2. Sketch-based queries 3. Example-based queries[6] 4. Semantic-based queries[1]. Methods 2 and 3 are usually used in Content-Based Information Retrieval (CBIR). Recent studies have focused on Context instead of Content resulting in the presentation of a range of Semantic-Based Video Retrieval methods which are discussed in the published works section. The system presented here is for assessing significant events included in a database containing shots of football videos. Queries with the user are based on text. For instance, the user can query the system with items such as this: shots of goals scored by the yellow team. The user interface is designed in such a way that the user can ask for the summary of any given video. Published works include several major methods and techniques for retrieval of information-which are common in some features- as summarized here: 1. Context Based Information Retrieval (CBIR); In such methods the media is modelled by features such as colour, texture, etc. and special relationships between objects and movements. In CBIR suitable features are extracted and related to high level concepts and meanings, without the system knowing such meanings. In fact the system displays and models the video contents in such a way that is efficient for content retrieval. For example there is a statistical model of motion in [3]. After that there is a training session on dynamic contents and then the data are recognized and classified. The result is a database of different video shots (football, basketball, meeting, and highways) each divided into minor events. The provided structure classifies shots in various groups based on similarity of their contents. Also there is fuzzy presentation of video contents in [12]. In this paper suitable features of a multidimensional fuzzy histogram are made for each video frame. One feature will be suitable for studying the similarities between frames. Consecutive frames are classified based on the fuzzy histogram. In fact the paper divides the videos into shots by this method, and then it chooses a keyframe for each of these frames. Then the shot related to the keyframe is provided as the answer.

2. Semantic Based Retrieval; One of the most essential phases of designing a Semantic-based retrieval system is defining a meaningful list of concept-oriented meanings based on human's available knowledge. Thus each single

ISBN: 978-989-95079-6-8

concept in the list should be accompanied by correct and true descriptions on the video collection. A new perspective introduced in recent years includes the use of Semantic Description of Object Motion for the retrieval process, which has been studied in [1],[2],[8],[7], and [6]. The method offered in [1] is as follows: First objects included in video frames are extracted and their trajectories are tracked. Once the trajectories are extracted they are classified for the training of object motion models. Finally, a meaningful description is added to these motion models. It should be added that such a meaningful description was added manually.

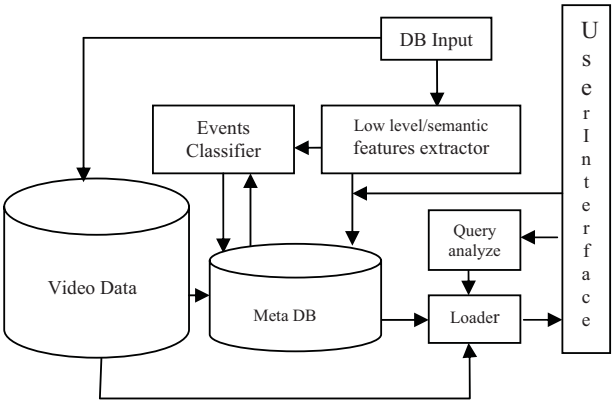


Fig. 1 The overall proposed structure

2 Proposed Structure

Most information retrieval systems make indirect use of human knowledge in their retrieval process(for example in the query by example, all of instance are a priori knowledge and in Semantic Based retrieval a meaningful description as priori knowledge should be added to the visual models manually.) The new method we are presenting in here uses human knowledge directly and in a very efficient way by a fuzzy rule base. The presented structure allows the system to process based on soccer video shots available in the database. The first phase is devoted to extracting shots from each video and making a list of features extracted from each shot. Then a fuzzy system is used to eliminate shots including insignificant events. Finally shots are classified and associated with predefined classed using a SVM. Then shots related to the class associated with the user query are provided as an answer to that query. The user may make queries on different events and concepts such as occurrence of penalties, corners or goals or team attacks throughout the DB. The overall structure is provided in Figure 1. As can been seen each video is processed before entering the DB and its low level features and conceptual properties are extracted, and once the events are classified all collected information is stored along with all data provided by the user (such as names of teams, time and place of the match, etc.) inside the Meta DB, while the video itself is stored in another location. Indeed the assumed database is of Multimedia linked meta database type. Figure 2 demonstrates feature extraction, Shot filtering using fuzzy system and event classification stages.

2.1 Extracting Low-Level and Semantic-Based Features

Used low-level features includes: Gi percent of pixels including grass of the field in keyframe of shot i for which we used the method provided in [5]. Once the grass region was recognized we divided it into 3 sections according to the Fig.3 and calculated Gi(E2) , percent of pixels covering the section E2 of the grass field, for keyframe. These features(Gi,Gi(E2)) may be used for Shot boundary detection and Shot classification and may be used as input to the fuzzy system not only in the Shot Rejection but also in the next phase of Event Classification along with low-level features.

The first step of proposed structure is devoted to extracting shots from each video (Fig. 2) and making a list of features extracted from each shot. Then a fuzzy system is used to eliminate shots including insignificant events. Finally shots are classified and associated with predefined classed using SVM The results of the classification phase are accompanied by a textual description and enables retrieval through text based query. The proposed method is based on the idea of using a fuzzy system for shot rejection and textual description for Soccer Events retrieval. In this section first the general structure is studied and we proceed to describe the extraction of low level features. Afterwards we go to high level features and then development of visual models. The fuzzy system including fuzzy rule base is presented in the section 2.2.

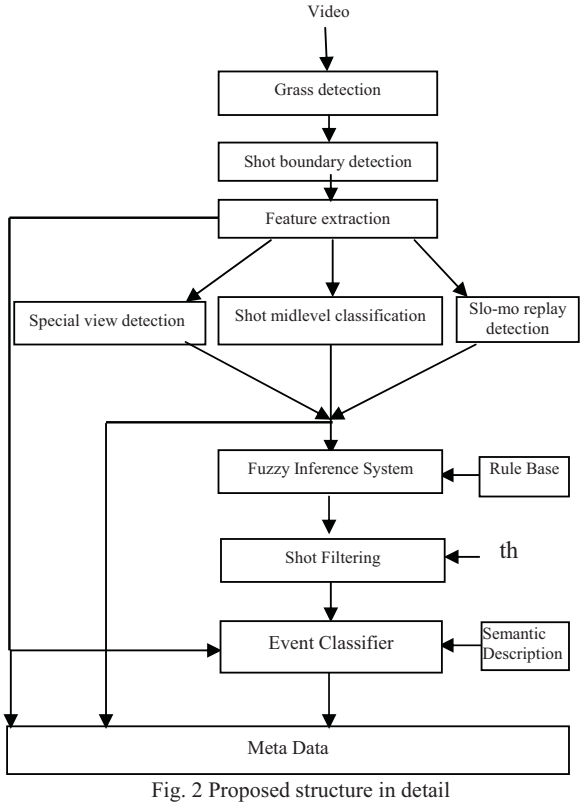


Fig. 2 Proposed structure in detail

ISBN: 978-989-95079-6-8

Fig. 3 Segmenting the area including the grass field

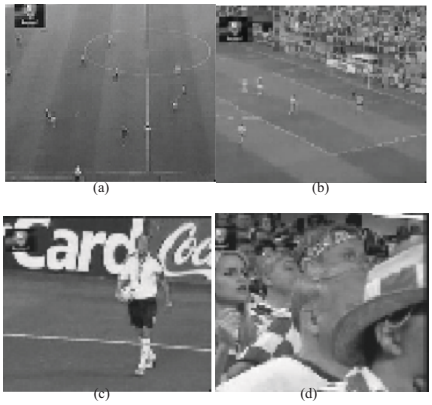


Fig.4 Types of views: a) Far-center, b) Far-side, c) Medium view, d) Out of field

Used semantic based features includes: 1. Far views from midfield, Fig. 4.a. 2. Far views from field sides, Fig. 4.b. 3. Medium views from inside the field 4.c, 4. View of the area outside of the field / closed view, showing the spectators or the upper body of a player, Fig. 4.d 5. Slow-Motion Replay Detection[9]. we have proposed a structural classification method for extracting this views in [5].

2.2 Using a fuzzy system to eliminate shots containing no significant events

Data retrievable from each shot by the above features as input of fuzzy system up to this phase are included in Table 1 along with abbreviations used for each one of them. The goal of designing a fuzzy system is to use the collected data in defining a degree of significance for each shot. This fuzzy system shows how important are the contents of that shot in terms of events occurring in it. For example, the system is expected to give a higher degree of significance to a shot containing a goal scoring than the one containing images of fans and no other important event. The rulebase and 3 triangle membership functions (low, Medium, High) are presented in Figure 5,6. The inference method used here is Mamdani product, for example for one execution of fuzzy inference result see Fig 7. Once each shot is given a degree of significance by the fuzzy system, a thresholding phase is run to eliminate the fuzzy system outputs for some shots. In this phase all shots with a significance degree below th(our system parameter) are eliminated from the classification. by defining a threshold(th) on the output of the fuzzy system we can separate the useful and none-useful shots for the purpose of classification. The threshold must be between 0 and 1. The lowest amount of the threshold is zero. By defining the threshold as 0 the all of shots in the database will be seen in the output of rejection phase as well. the higher the degree of significance the more sensitive the system vice versa.

Table 1 The input/output of fuzzy system with their abbreviations

Shot Degree	Out	Sd
Percent of Far-center view in a shot	In	Fc
Percent of Far-side view in a shot	In	Fs
Percent of Medium view	In	Mv
Percent of Out of field	In	Of
Percent of Frame includes Slow-Motion	In	Sm

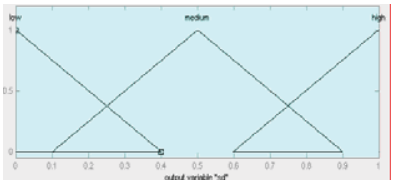


Fig. 5 triangle membership functions (low, Medium, High) for most In/Out variables

1. If (fc is high) then (sd is low) (1)
2. If (fs is high) then (sd is very high) (1)
3. If (fs is medium) then (sd is high) (1)
4. If (mv is high) then (sd is high) (1)
5. If (of is high) then (sd is low) (1)
6. If (cu is high) then (sd is low) (1)
7. If (cu is medium) then (sd is medium) (1)
8. If (fs is low) then (sd is medium) (1)
9. If (sm is high) then (sd is high) (1)
10. If (sm is medium) then (sd is high) (1)

Fig. 6 Fuzzy rule base for none-useful shot rejection

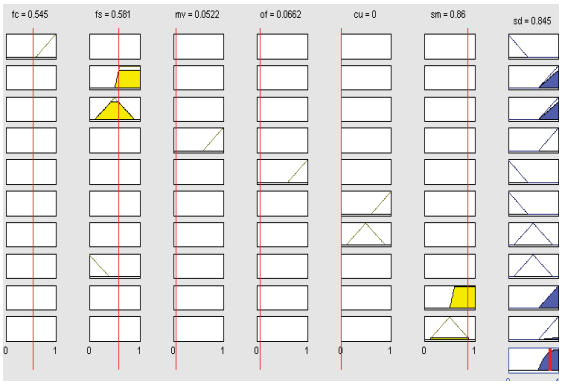


Fig. 7 if(fc=0.545 & fs=0.581 & mv=0.052 & of=0.0682 & cu=0 & sm=0.86) then Shot Degree = 0.845

ISBN: 978-989-95079-6-8

3 Event Retrieval

Up to the previous level we rejected most video shots imported into the DB with high probability of being irrelevant to significant events. So classification of shots in this phase may be done based on the events included in them. We have listed the set of features that may be used to classify the events , features include: G, percent of pixels covering the grass field in shot keyframes, G2, percent of pixels in section E2 on the grass field in shot keyframes and the Sd (the output of fuzzy system) and Fs, Percent of Far-side view in a shot. After normalizing the above 4 features the value of each of them for each shot will range between 0 and 1, so the feature vector for each shot is shown by Fv (I,V) and expressed in the following form:

Fv(i,V)={G,G2,Fs,Sd}
Where: i is the shot number from video V.

We used SVM to classify the events. In this phase events must be associated with 5 predefined classes. All classes have their special keywords which are presented here:

EventClass = { Goal, Penalty, Kroner, Free kick, other}

After the learning phase the videos stored in the DB for making up the Meta Data are processed and then the events are classified. Afterwards the above keywords are associated with the created classes manually and then each new event automatically inherits its keyword and is recorded in the Meta Data for that event. A table is formed in the Meta Data for each single video processed (Figure 8). During implementation the name given to each table of the Meta Data is taken to be the same as the file name recorded in the database, so relations with the raw video are maintained. Also we can save the information entered into the system for each video in Title of main video file to be used in the retrieval process. Such information may include the name of competing teams, time or place of the match or other details.

Shot Index	Start of shot	End of Shot	Type of Shot	Slow-Motion	Class of Event
. . .	. .	...	. .	... ...	

Fig 8 A table is formed in the Meta Data for each single video processed (The field related to Type of Shot shows the view that has been most frequent in the shot.)

The query model we used for this project was the same proposed in [2]. The base expression used in the model included the following:

select video from search_space [where condition];
select shot from search_space [where condition];

Expression SELECT specifies the SEARCH TARGET which may be Videos or Shots. The search space comes after FROM which may be all videos available in the DB or a list of chosen videos from any given video specified with a "where" condition which may only constitute one restriction or a combination of and/or phrases used to create conditions for time dependencies, which we ignore here.The provided system will be able to answer questions such as these:

Q1: "Find all the goal shots from all the soccer videos"
Q2: "Find all the corner kick shots from all the soccer videos where the goal accurse by England team"
Remember that the prerequisite to doing the condition contained in the second question is that the required information are added to the Meta Data by the system manager while the video is being entered into the database. To find the answer it suffices to extract the defined terms from the query expression and search for them in the Meta Data.

4 Evaluation

We have used more than 5.5 hours of football Videos in the database includes: 1 match of the World Cup 2006, 2 matches from the UEFA Champions League 2005, 1 match from the FA Premier League 2004, and more than 5 clips from Euro 2004(Table 2). The file format of the input films was non-compressed avi and the size of output films is 88×79.

Table 2. Names and the length (min:sec) of the clips in the database

Euro001(46:40),Euro002(47:00),Euro003(47:30), Euro004(46:00), Euro005(46:45), FIFA-div01f(12:38), FIFA-div02f(18:23),FIFA-div03f(15:38),FIFA-div04f(17:40),FIFA-div05f(18:22),Eng1(42:7) ,Eng2(20:51)

In the training phase we used 119 shots to learning for SVM and the rate of correct class in training data is 87%. As noted earlier the output of fuzzy system is a degree for each shot and the system rejects no useful shots using a threshold(th) as system input parameter. The input margin of the system for evaluation is 0.2 and then inputs include 0.2, 0.4, 0.6, 0.8 and 1. The results of changing the input amount of the system and the accuracy of the events classification are shown in tables 3,4,5,6. In this evaluation scenario we used 187 shots includes 5 Goals, 5 Penalties, 6 Free kicks, 9 Corners and 162 shots containing no significant events. Table 3 where the input th is 0.2 then system rejected 23

ISBN: 978-989-95079-6-8

shots before classification, table 3 shows the result of classification after rejection phase with input 0.2, Table 4 where the input th is 0.4 then system rejected 86 shots. Table 5 where th is 0.6 and the system rejected 114 shots and Table 6 where th is 0.8 and the system rejected 172 shots.
When th is 0.6, 4 shots containing corner and free kicks are rejected incorrectly(see the '# of rejected' row of table 5) and When th is 0.8, 10 shots containing very important events are rejected incorrectly (see the '# of rejected row' of table 6).
In the tables accuracy formula is:
Accuracy =100* Correct/(Total + False)
The Tables indicate that the higher the threshold, the higher the number of rejected shots will be. Also, the higher the threshold, the lower the number of incorrect shot detection (see 'False' row of tables) will be. See Table 6 where th = 0.8 and False detection for all events is zero but 1 Goal, 5 Free kicks and 4 corner are rejected incorrectly.

Table 3. The result of proposed method with th = 0.2

	Goal	Penalties	Free kick	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	0	0	0	0	23	23
Correct	5	5	6	9	86	111
False	13	9	17	14	0	53
Accuracy	27%	36%	26%	39%	-	-

Table 4. The result of proposed method with th = 0.4

	Goal	Penalties	Free kicks	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	0	0	0	0	86	86
Correct	5	5	5	8	51	74
False	6	0	13	6	1	27
Accuracy	45%	100%	32%	53%	-	-

Table 5. The result of proposed method with th = 0.6

	Goal	Penalties	Free kicks	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	0	0	2	2	110	114
Correct	5	5	4	7	42	63
False	2	0	5	3	0	10
Accuracy	71%	100%	36%	58%	-	-

Table 6. The result of proposed method with th = 0.8

	Goal	Penalties	Free kicks	Corner	other	sum
Total	5	5	6	9	162	187
# of rejected	1	0	5	4	162	172
Correct	4	5	1	5	0	15
False	0	0	0	0	0	0
Accuracy	80%	100%	16%	55%	-	-

5 Conclusions

The proposed method uses human expert knowledge for soccer video events retrieval by fuzzy systems. After feature extraction, each shot is given a degree of significance by the fuzzy system, thereby significantly reducing time complexity of video processing. By defining a threshold on the output of the fuzzy system, the useful and none-useful shots are separated for the purpose of event classification. Support vector machines are then applied for final classification of video information.

References

[1] Weiming Hu, Dan Xie, Zhouyu Fu, Wenrong Zeng, and Steve Maybank, "Semantic-Based Surveillance Video Retrieval", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 16, NO. 4, APRIL 2007.

[2] Cees G. M. Snoek, Bouke Huurnink, Laura Hollink,"Adding Semantics to Detectors for Video Retrieval",IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 9, NO. 5, AUGUST 2007.

[3] Gwenaëlle Piriou, Patrick Bouthemy, and Jian-Feng Yao,"Recognition of Dynamic Video Contents With Global Probabilistic Models of Visual Motion' , IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 15, NO. 11, NOVEMBER 2006.

[4] Cees G.M. Snoek , "Multimedia Event based Video Indexing using Time Intervals," IEEE TRANSACTIONS ON MULTIMEDIA, 2004.

[5] Ehsan Lotfi, H.R Pourreza, "Event Detection and Automatic Summarization in Soccer Video ," 4th Iranian Conference on Image Processing and Machine Vision, Mashhad, Iran, Feb 2007.

[6] Z. Q. Liu, L. T. Bruton, J. C. Bezdek, J. M. Keller, S. Dance, N. R. Bartley, and C. Zhang, "Dynamic image sequence analysis using fuzzy measures," IEEE Trans. Syst., Man, Cybern., vol. 31, no. 4, pp. 557–571, Aug. 2001.

[7] Haiping Sun, Joo-Hwee Lim, Qi Tian, Mohan S. Kankanhalli, "SEMANTIC LABELING OF SOCCER VIDEO", Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on MultimediaVolume 3, Issue , 15-18 Dec. 2003.

[8] B. Bell and L. F. Pau, "Context knowledge and search control issues in object-oriented prolog-based image understanding," Pattern Recognit. Lett., vol. 13, no. 4, pp. 279–290, 1992

[9] Ahmet Eklin,A.murat Tekalp, and Rajiv Mehrotra,"Automatic Soccer Video Analysis and Summarization", IEEE Transactions on Image processing Volume 12, Issue 7, July 2003.

[10] N.-S. Chang and K.-S. Fu, "Query-by-Pictorial-Example," IEEE Transactions on Software Engineering, vol. 6, no. 6, pp. 5 19-524, November 1980.

[11] Yu-Lin Kang ,"SOCCER VIDEO EVENT DETECTION WITH VISUAL KEYWORDS", ICICS-CM 200315-18 December 2003 Singapore.

[12] Anastasios D. Doulamis*, Nikolaos D. Doulamis, Stefanos D. Kollias "A fuzzy video content representation for video summarization and content-based retrieval" , ELAIVIER, Signal Processing 80 (2000) 1049.

[13] " Inference and retrieval of soccer event", SUN Xing-hua, YANG Jing-yu, Journal of Communication and Computer, Volume 4, No.3 (Serial No.28) , Mar. 2007.

ISBN: 978-989-95079-6-8

Pointwise defined CRI-based aggregation distributive operators are trivial

Siegfried Gottwald[1] Philipp Hüsam[2]

1. Universität Leipzig, Abt. Logik und Wissenschaftstheorie, Institut für Philosophie
Beethovenstr. 15, 04107 Leipzig, Germany
2. Universität Leipzig, Institut für Mathematik
Augustusplatz 9, 04109 Leipzig, Germany
Email: gottwald@uni-leipzig.de, philipp@huesam.de

***Abstract*— *In a rather general, essentially aggregation operator based discussion of the traditional fuzzy control strategies known as FATI and FITA strategies, a way to reduce these strategies to one another has been to define pairs of aggregation distributive aggregation operators. In this paper it is shown that for some often used special cases this reduction condition allows only the set theoretic union as aggregation operator.*

***Keywords*—** fuzzy control, FATI strategy, FITA strategy, aggregation operators, aggregation distributivity

1 Introduction

The standard paradigm behind standard approaches toward fuzzy control is that one supposes to have given, as an incomplete and fuzzy description of a control function Φ from an input space $\mathbb{X}$ to an output space $\mathbb{Y}$, a family

$$\mathcal{D} = (\langle A_i, B_i\rangle)_{1\leq i\leq n} \tag{1}$$

of (fuzzy) input-output data pairs to characterize this function Φ.

In the usual approaches such a family of input-output data pairs is provided by a finite list

$$\text{IF } x \text{ is } A_i \text{ THEN } y \text{ is } B_i, \tag{2}$$

(with $i = 1, \ldots, n$) of linguistic control rules, also called fuzzy IF-THEN rules.

The basic examples of fuzzy control approaches are Zadeh's original approach via fuzzy relations and the compositional rule of inference (CRI), as prototypically realized by the Mamdani–Assilian approach in [7], and the Holmblad–Ostergaard approach toward fuzzy control of a cement kiln as explained in [5].

Derived from these two approaches there is the well known distinction between FATI and FITA strategies to evaluate systems of linguistic control rules w.r.t. arbitrary fuzzy inputs from $\mathcal{F}(\mathbb{X})$.

2 Preliminaries

The core idea of a FITA strategy is that it is a strategy which **F**irst **I**nfers (by reference to the single rules) and **T**hen **A**ggregates starting from the actual input information A. Contrary to that, a FATI strategy is a strategy which **F**irst **A**ggregates (the information in all the rules into one fuzzy relation) and **T**hen **I**nfers starting from the actual input information A.

Both these strategies use the set theoretic union as their aggregation operator. Furthermore, both of them refer to the CRI as their core tool of inference.

In general, however, the interpolation operators may depend more generally upon some inference operator(s) as well as upon some aggregation operator.

2.1 Aggregation operations and fuzzy control strategies

By an *inference operator* we mean here simply a mapping from the class of fuzzy subsets of the input space to the class of fuzzy subsets of the output space.[1]

And an *aggregation operator* $\mathbf{A}$, as explained e.g. in [1, 2], is a family $(f^n)_{n\in\mathbb{N}}$ of ("aggregation") operations, each f^n an n-ary one, over some partially ordered set $\mathbf{M}$, with ordering $\leqslant$, with a bottom element $\mathbf{0}$ and a top element $\mathbf{1}$, such that each operation f^n is non-decreasing, maps the bottom to the bottom: $f^n(\mathbf{0}, \ldots, \mathbf{0}) = \mathbf{0}$, and the top to the top: $f^n(\mathbf{1}, \ldots, \mathbf{1}) = \mathbf{1}$.

Such an aggregation operator $\mathbf{A} = (f^n)_{n\in\mathbb{N}}$ is a *commutative* one iff each operation f^n is commutative. And $\mathbf{A}$ is an *associative* aggregation operator iff

$$\begin{aligned} &f^n(a_1, \ldots, a_n) = \\ &\quad f^r(f^{k_1}(a_1, \ldots, a_{k_1}), \ldots, f^{k_r}(a_{m+1}, \ldots, a_n)) \end{aligned} \tag{3}$$

for $n = \sum_{i=1}^{r} k_i$ and $m = \sum_{i=1}^{r-1} k_i$.

Our aggregation operators further on are supposed to be commutative as well as associative ones.[2]

As in [3, 4], we now consider operators Ψ of FATI-type operators Ξ of FITA-type and which have the abstract forms

$$\Psi_{\mathcal{D}}(A) = \mathbf{A}(\theta_{\langle A_1,B_1\rangle}(A), \ldots, \theta_{\langle A_n,B_n\rangle}(A)), \tag{4}$$

$$\Xi_{\mathcal{D}}(A) = \widehat{\mathbf{A}}(\theta_{\langle A_1,B_1\rangle}, \ldots, \theta_{\langle A_n,B_n\rangle})(A). \tag{5}$$

Here we assume that each one of the "local" inference operators θ_i is determined by the single input-output pair $\langle A_i, B_i\rangle$. This restriction is in general sufficient. For the present purpose we assume that our inference operators are CRI-based, i.e. we assume that $\theta_{\langle A_1,B_1\rangle}(A)$ has

[1] This terminology has its historical roots in the fuzzy control community. There is no relationship at all with the logical notion of inference intended and supposed here; but–of course–also not ruled out.

[2] It seems that this is a rather restrictive choice from a theoretical point of view. However, in all the usual cases these restrictions are satisfied.

ISBN: 978-989-95079-6-8

the form

$$\theta_{\langle A_i,B_i\rangle}(A) = R_i'' A \tag{6}$$

for some fuzzy relation R. In this case we call the inference operator $\theta_{\langle A_1,B_1\rangle}$ CRI-*based*.

Furthermore $\mathbf{A}$ has to be an aggregation operator for fuzzy subsets of the universe of discourse $\mathbb{X}$, and $\widehat{\mathbf{A}}$ has to be an aggregation operator for inference operators.

2.2 Stability conditions

If $\Theta_{\mathcal{D}}$ is a fuzzy inference operator of one of the types (4), (5), then the interpolation property one likes to have realized is that one has

$$\Theta_{\mathcal{D}}(A_i) = B_i \tag{7}$$

for all the data pairs $\langle A_i, B_i\rangle$. In the particular case that the operator $\Theta_{\mathcal{D}}$ is detrmined by the CRI-methodology, this is just the usual problem to solve a system (7) of fuzzy relation equations.

In the present generalized context the property (7) has been called the $\mathcal{D}$-*stability* of the fuzzy inference operator $\Theta_{\mathcal{D}}$.

To find $\mathcal{D}$-stability conditions on this abstract level seems to be rather difficult in general. However, the restriction to fuzzy inference operators of FITA-type makes things easier.

To explain some of the known results it is necessary to have a closer look at the aggregation operator $\mathbf{A} = (f^n)_{n\in\mathbb{N}}$ involved in (4) which operates on $\mathbb{F}(\mathbb{Y})$, of course with the inclusion relation for fuzzy sets as partial ordering.

Definition 1 *Having $B, C \in \mathbb{F}(\mathbb{Y})$ we say that C is* $\mathbf{A}$-negligible w.r.t. *B iff $f^2(B, C) = f^1(B)$ holds true.*

The core idea here is that in any aggregation by $\mathbf{A}$ the presence of the fuzzy set B among the aggregated fuzzy sets makes any presence of C superfluous.

For examples and further interesting properties of aggregation operators the interested reader may consult [3, 4].

Now we are in a position to state one of the results from [3, 4] to give an impression of what becomes of interest in the present context.

Proposition 1 *Consider a fuzzy inference operator $\Psi_{\mathcal{D}}$ of FITA-type (4). It is sufficient for the $\mathcal{D}$-stability of $\Psi_{\mathcal{D}}$, i.e. to have*

$$\Psi_{\mathcal{D}}(A_k) = B_k \qquad \text{for all } k = 1, \ldots, n \tag{8}$$

that one always has

$$\theta_{\langle A_k,B_k\rangle}(A_k) = B_k \tag{9}$$

and additionally that for each $i \neq k$ the fuzzy set

$$\theta_{\langle A_k,B_k\rangle}(A_i) \quad \text{is } \mathbf{A}\text{-negligible w.r.t.} \quad \theta_{\langle A_k,B_k\rangle}(A_k)\,. \tag{10}$$

This result has two quite interesting specializations which themselves generalize well known results about fuzzy relation equations. The interested reader may consult [3, 4].

To extend such considerations from inference operators (4) of the FITA type to those ones of the FATI type (5) let us consider the following notion.

Definition 2 *Suppose that $\widehat{\mathbf{A}}$ is an aggregation operator for inference operators, and that $\mathbf{A}$ is an aggregation operator for fuzzy sets. Then $(\widehat{\mathbf{A}}, \mathbf{A})$ is an* application distributive *pair of aggregation operators iff*

$$\widehat{\mathbf{A}}(\theta_1, \ldots, \theta_n)(X) = \mathbf{A}(\theta_1(X), \ldots, \theta_n(X)) \tag{11}$$

holds true for arbitrary inference operators $\theta_1, \ldots, \theta_n$ and fuzzy sets X.

Using this notion it is easy to see that one has on the left hand side of (11) a FATI type inference operator, and on the right hand side an associated FITA type inference operator. So one is able to give a reduction of the FATI case to the FITA case, assuming that such application distributive pairs of aggregation operators exist.

Proposition 2 *Suppose that $(\widehat{\mathbf{A}}, \mathbf{A})$ is an application distributive pair of aggregation operators. Then a fuzzy inference operator $\Xi_{\mathcal{D}}$ of FATI-type is $\mathcal{D}$-stable iff its associated fuzzy inference operator $\Psi_{\mathcal{D}}$ of FITA-type is $\mathcal{D}$-stable.*

3 Application distributivity

Based upon the notion of application distributive pair of aggregation operators the property of $\mathcal{D}$-stability can be transferred back and forth between two inference operators of FATI-type and of FITA-type if they are based upon a pair of application distributive aggregation operators.

What has not been discussed previously was the existence and the uniqueness of such pairs. Here are some results concerning these problems.

The uniqueness problem has a simple solution.

Proposition 3 *If $(\widehat{\mathbf{A}}, \mathbf{A})$ is an application distributive pair of aggregation operators then $\widehat{\mathbf{A}}$ is uniquely determined by $\mathbf{A}$, and conversely also $\mathbf{A}$ is uniquely determined by $\widehat{\mathbf{A}}$.*

And for the existence problem we have a nice reduction to the two-argument case.

Theorem 1 *Suppose that $\mathbf{A}$ is a commutative and associative aggregation operator. For the case that there exists an aggregation operator $\widehat{\mathbf{A}}$ such that $(\widehat{\mathbf{A}}, \mathbf{A})$ form an application distributive pair of aggregation operators it is necessary and sufficient that there exists some operation G for fuzzy inference operators satisfying*

$$\mathbf{A}(\theta_1(X), \theta_2(X)) = G(\theta_1, \theta_2)(X) \tag{12}$$

for all fuzzy inference operators θ_1, θ_2 and all fuzzy sets X.

However, there is an important restriction concerning the existence of such pairs of application distributive aggregation operators, at least for the interesting particular case that the application operation is determined by the compositional rule of inference (CRI). And this means simply that the inference operations θ_i are determined via suitable fuzzy relations R_i.

ISBN: 978-989-95079-6-8

Definition 3 *An aggregation operator* $\mathbf{A} = (f^n)_{n\in\mathbb{N}}$ *for fuzzy subsets of a universe of discourse* $\mathbb{X}$ *is* pointwise defined *iff for each* $n \in \mathbb{N}$ *there exists a function* $g_n : [0,1]^n \to [0,1]$ *such that for all* $A_1, \ldots, A_n \in \mathcal{F}(\mathbb{X})$ *and all* $x \in \mathbb{X}$ *there hold*

$$f^n(A_1, \ldots, A_n)(x) = g_n(A_1(x), \ldots, A_n(x)) \,. \tag{13}$$

And an aggregation operator $\widehat{\mathbf{A}}$ *for inference operators is* pointwise defined *iff it can be reduced to a pointwise defined aggregation operator for fuzzy relations.*

From the isotonicity behavior of the aggregation operator $\mathbf{A}$ it follows that also these characterizing functions g_n are isotonic, and similarly in the case of $\widehat{\mathbf{A}}$.

The restrictive result, first proved in [6], now reads as follows.

Theorem 2 *Suppose that all inference operators are CRI-based. Then the pair* $(\bigcup, \bigcup)$ *is the only application distributive pair among the commutative, associative, and pointwise defined aggregation operators.*

Proof: Obviously the considerations can restricted to the binary case $n = 2$. So let us start in this CRI-based case with an aggregation operator $\widehat{\mathbf{A}}$, which has to give a fuzzy relation $\widehat{\mathbf{A}}(R_1, R_2)$ for any two inference operators θ_1, θ_2 determined by the fuzzy relations R_1, R_2, respectively. Because $\widehat{\mathbf{A}}$ has to be pointwise defined, according to Definition 3 there has to be a function $\widehat{g} : [0,1] \times [0,1] \to [0,1]$ such that one has for the membership degrees of the corresponding fuzzy relations

$$\widehat{\mathbf{A}}(R_1, R_2)(x, y) = \widehat{g}(R_1(x, y), R_2(x, y)) \,. \tag{14}$$

In a similar way, again by Definition 3, a corresponding aggregation operator $\mathbf{A}$ has to be determined by a function $g : [0,1] \times [0,1] \to [0,1]$. Assuming that these aggregation operators $(\mathbf{A}, \widehat{\mathbf{A}})$ form an aggregation distributive pair, gives for arbitrary fuzzy inputs A the condition

$$\begin{aligned}
&\widehat{\mathbf{A}}(R_1, R_2)''(A)(y) \\
&= \bigvee_x T\Big(A(x), \widehat{\mathbf{A}}(R_1, R_2)(x, y)\Big) \\
&= \bigvee_x T\Big(A(x), \widehat{g}(R_1(x, y), R_2(x, y))\Big) \qquad (15) \\
&= g\Big(\bigvee_x T(A(x), R_1(x, y)), \bigvee_x T(A(x), R_2(x, y))\Big) \\
&= \mathbf{A}(R_1'' A, R_2'' A)(y) \,,
\end{aligned}$$

which has to be satisfied for arbitrary fuzzy sets A and fuzzy relations R_1, R_2. Of course, T here is the t-norm involved in the CRI application process.

So let be always $A(x) = 1$ and furthermore $R_1(x, y) = a, R_2(x, y) = b$ for some $a, b \in [0,1]$. Now routine calculations yield $\widehat{g}(a, b) = g(a, b)$, which means equality of the functions $\widehat{g} = g$ which determine the aggregation operators $\widehat{\mathbf{A}}, \mathbf{A}$, respectively.

So application distributivity of the pair $(\mathbf{A}, \widehat{\mathbf{A}})$ becomes a condition which has to be satisfied by the characterizing function g, and this condition reads

$$\begin{aligned}
&\bigvee_x T\Big(A(x), g(R_1(x, y), R_2(x, y))\Big) \; = \qquad (16) \\
&\qquad g\Big(\bigvee_x T(A(x), R_1(x, y)), \bigvee_x T(A(x), R_2(x, y))\Big) \,.
\end{aligned}$$

To continue our discussion and to finish the proof of Theorem 2, we insert two lemmata.

Lemma 1 *Suppose that* $g : [0,1] \times [0,1] \to [0,1]$ *determines a commutative and associative pointwise defined (binary) aggregation operator. Then the condition*

$$\sup_{i\in I} g(a_i, b_i) = g(\sup_{i\in I} a_i, \sup_{i\in I} b_i) \tag{17}$$

is equivalent to the fact that g *is left continuous and satisfies*

$$g(a, b) = g(\max\{a, b\}, \max\{a, b\}) \,. \tag{18}$$

It is easy to see that (17) implies the left continuity of g. So assume (17) and that (18) is not generally satisfied. Then there are $a_0, b_0 \in [0,1]$ such that

$$g(a_0, b_0) \neq g(\max\{a_0, b_0\}, \max\{a_0, b_0\}) \tag{19}$$

and additionally, w.l.o.g., also $b_0 \leq a_0$. This last condition forces even its strengthening $b_0 < a_0$, and together with the isotonicity of g yields $g(a_0, b_0) < g(a_0, a_0)$. But this now means

$$\begin{aligned}
&\max\{g(a_0, b_0), g(b_0, a_0)\} = g(a_0, b_0) \\
&\quad < \quad g(a_0, a_0) = g(\max\{a_0, b_0\}, \max\{a_0, b_0\}) \,, \qquad (20)
\end{aligned}$$

contradicting (17). So the $(\Rightarrow)$-part of the lemma is proved.

If otherwise g is left continuous and satisfies (18), one has

$$g(\sup_{i\in I} a_i, \sup_{i\in I} b_i) \geq \sup_{i\in I} g(a_i, b_i) \tag{21}$$

by the isotonicity of g. But $g(a_j, b_j) \leq \sup_{i\in I} g(a_i, b_i)$ and (18) yield

$$g(c_j, c_j) \leq \sup_{i\in I} g(a_i, b_i) \tag{22}$$

for $c_j = \max\{a_j, b_j\}$, and by the left continuity and isotonicity of g this gives

$$g(\sup_{i\in I} c_i, c_j) \leq \sup_{i\in I} g(c_i, c_i) \leq \sup_{i\in I} g(a_i, b_i), , \tag{23}$$

and thus also

$$g(\sup_{i\in I} a_i, \sup_{i\in I} b_i) \leq g(\sup_{i\in I} c_i, \sup_{i\in I} c_i) \leq \sup_{i\in I} g(a_i, b_i) \,. \tag{24}$$

All together gives (17) and hence the $(\Leftarrow)$-part of the lemma.

The condition (17) obviously means that the corresponding aggregation operator commutes with the supremum.

ISBN: 978-989-95079-6-8

Lemma 2 *Suppose that* $g : [0,1] \times [0,1] \to [0,1]$ *determines a commutative and associative pointwise defined (binary) aggregation operator. Then condition (17) is satisfied iff there exists a left continuous isotonic function* $h : [0,1] \to [0,1]$ *satisfying* $h(0) = 0$, $h(1) = 1$ *and*

$$g(a,b) = \max\{h(a), h(b)\}\,. \tag{25}$$

For the $(\Rightarrow)$-part consider the function $h(x) = g(0,x)$. It is left continuous as well as isotonic, satisfies $h(0) = 0$ and $h(1) = 1$, and one has by Lemma 1

$$\begin{aligned} g(a,b) &= g(\max\{a,b\}, \max\{a,b\}) \\ &= g(\max\{0, \max\{a,b\}\}, \max\{0, \max\{a,b\}\}) \\ &= g(0, \max\{a,b\}) \;=\; \max\{h(a), h(b)\}\,. \end{aligned} \tag{26}$$

The $(\Leftarrow)$-part follows from routine calculations.

So we come back to the proof of Theorem 2. What we did not discuss up to now is that for having an aggregation distributive pair of pointwise defined operators the characterizing function g has to be distributive relative to the t-norm T which determines the CRI-application. So we need to have always satisfied

$$T(a, g(b,c)) = g(T(a,b), T(a,c))\,, \tag{27}$$

which means, via Lemmata 1 and 2, to have always satisfied

$$\begin{aligned} &\max\{T(a,h(b)), T(a,h(c))\} \\ &\quad = \quad \max\{h(T(a,b)), h(T(a,c))\}\,, \end{aligned} \tag{28}$$

which forces that one always has to have

$$T(a,h(b)) = h(T(a,b))\,. \tag{29}$$

And this yields $h = \mathsf{id}$: because otherwise there would exist some c with $h(c) \neq c$ and

$$T(c,h(1)) = T(c,1) = c \neq h(c) = h(T(c,1))\,. \tag{30}$$

So Theorem 2 is finally proved.

4 Conclusion

The type of approach explained in Section 2.2 works actually well only in the FITA case. This was the starting point for the considerations on aggregation distributive operator pairs. They give a transfer possibility to the FATI case.

However, as the main result, i.e. Theorem 2 of the present paper shows, these transfer possibilities are quite restricted under some conditions which have, up to now, been considered rather natural ones.

For the authors understanding this result points into two directions. (i) It may be appropriate to try to find other ways and to discuss the FATI case differently. And this way may become essentially different from the reduction strategy toward fuzzy relation equations which stands behind the generalization in [3, 4]. (ii) It may be suitable to move into the realm of aggregation operators which are no longer pointwise defined, and it may also be suitable to leave the world of the CRI-based approaches.

Particularly point (ii) here seems promising, even having in mind that the actual standard cases all fall into the class of pointwise defined aggregation operators. Further investigations into this topic are necessary.

References

[1] Tomasa Calvo, G. Mayor, and Radko Mesiar, editors. *Aggregation Operators: New Trends and Applications*. Physica-Verlag, Heidelberg, 2002.

[2] Didier Dubois and Henri Prade. On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems*, 142:143–161, 2004.

[3] Siegfried Gottwald. On a generalization of fuzzy relation equations. In *Proc. 11th Internat. Conf. IPMU 2006, Paris*, volume 2, pages 2572–2577, Paris, 2006. Edition EDK.

[4] Siegfried Gottwald. An abstract approach toward the evaluation of fuzzy rule systems. In Paul P. Wang, Da Ruan, and Etienne E. Kerre, editors, *Fuzzy Logic. A Spectrum of Theoretical & Practical Issues*, volume 215 of *Studies in Fuzziness and Soft Computing*, pages 281–297. Springer-Verlag, Berlin, 2007.

[5] L.P. Holmblad and J.J. Ostergaard. Control of a cement kiln by fuzzy logic. In Madan M. Gupta and Elie Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 389–399. North-Holland Publ. Comp., Amsterdam, 1982.

[6] Philipp Hüsam. FATI versus FITA. Ein abstrakter Vergleich gängiger Strategien bei "fuzzy control"-Systemen. Diplomarbeit, Mathematisches Institut, Universität Leipzig, 2008.

[7] Abe Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *Internat. J. Man-Machine Studies*, 7:1–13, 1975.

ISBN: 978-989-95079-6-8

A possibilistic approach to bottleneck combinatorial optimization problems with Ill-known weights

Adam Kasperski[1] Paweł Zieliński[2]

1.Institute of Industrial Engineering and Management, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
2.Institute of Mathematics and Computer Science, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
Email: adam.kasperski@pwr.wroc.pl, pawel.zielinski@pwr.wroc.pl

Abstract— *In this paper a general bottleneck combinatorial optimization problem with uncertain element weights modeled by fuzzy intervals is considered. A rigorous possibilistic formalization of the problem and solution concepts in this setting that lead to finding robust solutions under fuzzy weights are given. Some algorithms for finding a solution according to the introduced concepts and evaluating optimality of solutions and elements are provided.*

Keywords— Bottleneck combinatorial optimization, Interval, Fuzzy interval, Fuzzy optimization and design, Possibility theory.

1 Introduction

A *combinatorial optimization problem* consists in finding an object composed of elements of a given ground set E. In a deterministic case, every element of E has a precise weight. In *bottleneck combinatorial optimization problems*, we wish to find an object that minimizes the weight of its heaviest element. Such formulation encompasses a large variety of classical combinatorial optimization problems, for instance: the bottleneck path [1], the bottleneck assignment [2], the bottleneck spanning tree [3] (or a more general the bottleneck matroid base problem [4]) etc. All these problems are efficiently solvable when the weights of all elements are precisely known. Unfortunately, in real world it is not easy to estimate element weights exactly. In many cases, the exact values of weights are not known in advance and the imprecision must be taken into account.

In this paper, we wish to investigate a general bottleneck combinatorial optimization problem with uncertain element weights modeled by fuzzy intervals. The membership function $\mu_{\widetilde{W}}$ of a fuzzy interval $\widetilde{W}$ is a possibility distribution describing, for each value w of the element weight, the extent to which it is a possible value. Equivalently, it means that the value of this weight belongs to a λ-cut interval $\widetilde{W}^{\lambda} = \{t : \mu_{\widetilde{W}}(t) \geq \lambda\}$ with confidence (or degree of necessity) $1 - \lambda$. Now to each feasible solution or element a *degree of possible optimality* and a *degree of necessary optimality* can be assigned. The notion of the necessary optimality of a solution may be weaken by assigning a *degree of necessary soft optimality*. Moreover, all the degrees of optimality of a solution (an element) can be derived from a *fuzzy deviation*, that is a possibility distribution representing the set of plausible values of deviations of the solution (the element) from optimum. In order to choose a "robust solution" under fuzzy weights, we adopt two criteria. The first one consists in choosing a solution of the maximum degree of necessary optimality, called a *best necessarily optimal solution*. Such a solution, if exists, seems to be an ideal choice. The second criterion is weaker than the first one and consists in choosing a solution of the maximum degree of necessary soft optimality, called a *best necessarily soft optimal solution*. This criterion has been originally proposed in [5, 6] for the linear programming problem with a fuzzy objective function. A best necessarily soft optimal solution seems to be a compromise choice and it minimizes a "distance" to the necessary optimality.

In this paper, we provide some methods for the optimality evaluation and for choosing a solution under fuzzy weights. In Section 3, we investigate the interval case, that is the problems in which the element weights are specified as closed intervals. It turns out, that it is possible to construct some polynomial algorithms for such problems if only their deterministic counterparts are polynomially solvable. In consequence, the interval bottleneck problems are easier to solve than the interval problems with a linear sum objective discussed in [7, 8]. We obtain polynomial algorithms for such problems as the bottleneck shortest path, the bottleneck assignment and the bottleneck matroid base. In Section 4, we show that the optimality evaluation and the problem of choosing a solution under fuzzy weights can be reduced to examining a family of interval problems. In particular, we show that a best necessarily soft optimal solution can be computed in polynomial time for a wide class of problems.

2 Preliminaries

In this section, we recall a formulation of a general bottleneck combinatorial problem with precise weights.

Let $E = \{e_1, \ldots, e_n\}$ be a finite ground set and let $\Phi \subseteq 2^E$ be a set of subsets of E called the set of the *feasible solutions*. A nonnegative real weight w_e is given for every element $e \in E$. A *bottleneck combinatorial optimization problem* $\mathcal{BP}$ consists in finding a feasible solution X that minimizes the weight of its heaviest element, namely:

$$\mathcal{BP}: \quad \min_{X \in \Phi} \max_{e \in X} w_e. \tag{1}$$

A solution that minimizes (1) is called an *optimal solution*. We call an element $e \in E$ *optimal* if it is a part of an optimal solution. If a solution (element) is not optimal a natural question arises how far from optimality this solution (element) is. To answer this question one can introduce the concept of a deviation. *A deviation of solution* $X \in \Phi$ and *a deviation of*

ISBN: 978-989-95079-6-8

element $f \in E$ are defined in the following way:

$$\delta_X = \max_{e\in X} w_e - \min_{Y\in\Phi} \max_{e\in Y} w_e,$$
$$\delta_f = \min_{Y\in\Phi_f} \max_{e\in Y} w_e - \min_{Y\in\Phi} \max_{e\in Y} w_e,$$

where Φ_f is the set of all feasible solutions that contain element f. It is clear that solution X (element f) is optimal if and only if $\delta_X = 0$ ($\delta_f = 0$).

In this paper we will assume that problem $\mathcal{BP}$ is polynomially solvable. Some polynomial algorithms for the bottleneck path, the bottleneck assignment, the bottleneck spanning tree, and the bottleneck matroid base problem can be found for instance in [1, 2, 3, 4]. For all these problems the deviations δ_X and δ_f can be computed in polynomial time.

3 Interval-valued bottleneck combinatorial optimization problems

In this section we consider an *interval* version of problem (1), in which the weights of the elements are ill-known and they are modeled by closed intervals $W_e = [\underline{w}_e, \overline{w}_e]$, $e \in E$. Assigning some interval to an element weight means that the actual weight of this element will take some value within the interval but it is not possible at present to predict which one. Every precise instantiation of the element weights is called a *scenario* and we denote it by S. Thus every scenario is a vector $S = (w_e)_{e\in E}$, $w_e \in W_e$ that expresses a realization of the weights. We denote by Γ the set of all the scenarios, i.e. $\Gamma = \times_{e\in E}[\underline{w}_e, \overline{w}_e]$ and we use $w_e(S)$ to denote the weight of element $e \in E$ in a fixed scenario $S \in \Gamma$. Among the scenarios of Γ, we distinguish the *extreme* ones, which belong to $\times_{e\in E}\{\underline{w}_e, \overline{w}_e\}$. Let $A \subseteq E$ be a fixed subset of elements. In scenario S_A^+ all elements $e \in A$ have weights $\overline{w}_e$ and all the remaining elements have weights $\underline{w}_e$. Similarly, in scenario S_A^- all elements $e \in A$ have weights $\underline{w}_e$ and all the remaining elements have weights $\overline{w}_e$. For a given solution $X \in \Phi$, we define its weight under a fixed scenario $S \in \Gamma$ as $F(X, S) = \max_{e\in X} w_e(S)$. We will denote by $F^*(S)$ the value of the weight of an optimal solution under scenario $S \in \Gamma$, that is $F^*(S) = \min_{X\in\Phi} F(X, S)$.

The optimality in the interval-valued problem $\mathcal{BP}$ can be characterized as follows: a given solution $X \in \Phi$ (element $f \in E$) is *possibly optimal* if and only if it is optimal in some scenario $S \in \Gamma$. A given solution $X \in \Phi$ (element $f \in E$) is *necessarily optimal* if and only if it is optimal in all scenarios $S \in \Gamma$. Similarly to the deterministic case, we can express the possible and necessary optimality in terms of the deviation. Let $\delta_X(S) = F(X, S) - F^*(S)$ and $\delta_f(S) = \min_{Y\in\Phi_f} F(Y, S) - F^*(S)$ denote the deviations of solution X and element f under a fixed scenario $S \in \Gamma$, respectively. Consider the following optimization problems:

$$\underline{\delta}_X = \min_{S\in\Gamma} \delta_X(S), \quad \overline{\delta}_X = \max_{S\in\Gamma} \delta_X(S)$$
$$\underline{\delta}_f = \min_{S\in\Gamma} \delta_f(S), \quad \overline{\delta}_f = \max_{S\in\Gamma} \delta_f(S).$$

The solutions to the above problems determine the so called *deviation interval* $\Delta_X = [\underline{\delta}_X, \overline{\delta}_X]$ containing all possible values of deviations for solution X. Similarly $\Delta_f = [\underline{\delta}_f, \overline{\delta}_f]$ is a deviation interval for element f. It is worth pointing that in literature (see e.g. [9]) the quantity $\overline{\delta}_X$ is called the *maximal regret* or *robust deviation* and it expresses the maximal possible deviation from optimum. Obviously, we can easily deduce the optimality of a solution from the deviation interval. Namely, X is possibly (resp. necessarily) optimal if and only if $\underline{\delta}_X = 0$ (resp. $\overline{\delta}_X = 0$). Similarly, an element f is possibly (resp. necessarily) optimal if and only if $\underline{\delta}_f = 0$ (resp. $\overline{\delta}_f = 0$).

3.1 Asserting possible and necessary optimality

In this section we establish some sufficient and necessary conditions for possible and necessary optimality of solutions and elements in the interval-valued problem. We start by proving the following proposition:

Proposition 1. *Let X be a given feasible solution. Then*

$$\underline{\delta}_X = \max\{0, \max_{e\in X} \underline{w}_e - F^*(S_E^+)\}, \tag{2}$$
$$\overline{\delta}_X = \max_{e\in X} \max\{0, \overline{w}_e - F^*(S_{\{e\}}^+)\}. \tag{3}$$

Proof. Equality (3) has been proved in [10]. We now give a proof sketch of (2). It is easy to verify that

$$\underline{\delta}_X \geq \max\{0, \max_{e\in X} \underline{w}_e - F^*(S_E^+)\}. \tag{4}$$

It remains to show that the inequality $\leq$ also holds in (4). Let Y be an optimal solution under S_E^+ and let $g = \arg\max_{e\in Y} \overline{w}_e$. We consider two cases. (i) $\max_{e\in X} \underline{w}_e > \overline{w}_g$. Denote $h = \arg\max_{e\in X} \underline{w}_e$. Consider scenario S such that $w_e^S = \min\{\underline{w}_h, \overline{w}_e\}$ for all $e \in X$ and $w_e^S = \overline{w}_e$ for all $e \in E \setminus X$. Since $\underline{w}_h \geq \underline{w}_e$ for all $e \in X$, $S \in \Gamma$. It is easy to check that $F(X, S) = \underline{w}_h$ and $F^*(S) = F(S_E^+)$. Hence $\underline{\delta}_X \leq \delta_X(S) = \max_{e\in X} \underline{w}_e - F^*(S_E^+) \leq \max\{0, \max_{e\in X} \underline{w}_e - F^*(S_E^+)\}$, which together with (4) yield (2). (ii) $\max_{e\in X} \underline{w}_e \leq \overline{w}_g$. Consider scenario S such that under this scenario all elements $e \in E \setminus X$ have weights $\overline{w}_e$ and all the elements $e \in X$ have weights $\min\{\overline{w}_e, \overline{w}_g\}$. Since $\underline{w}_e \leq \overline{w}_g$ for all $e \in X$, $S \in \Gamma$. One can easily verify that X is optimal under S, which means that $\underline{\delta}_X = 0 \leq \max\{0, \max_{e\in X} \underline{w}_e - F^*(S_E^+)\}$. This, together with (4), give (2). □

Proposition 1 allows us to compute the solution deviation interval Δ_X and leads to the following two corollaries:

Corollary 1. *A solution $X \in \Phi$ is possibly optimal if and only if $F(X, S_E^-) = \max_{e\in X} \underline{w}_e \leq F^*(S_E^+)$.*

Corollary 2. *A solution $X \in \Phi$ is necessarily optimal if and only if $\max_{e\in X} \max\{0, \overline{w}_e - F^*(S_{\{e\}}^+)\} = 0$.*

Making use of Corollaries 1 and 2, we can efficiently evaluate the optimality of a given solution X if only the underlying bottleneck deterministic problem $\mathcal{BP}$ is efficiently solvable. In order to evaluate the possible optimality of X, it suffices to compute the value of $F^*(S_E^+)$ and check if $F(X, S_E^-) \leq F^*(S_E^+)$. This can be done in $O(|X| + f(|E|))$ time, where $f(|E|)$ is the running time of an algorithm for the deterministic $\mathcal{BP}$ problem. Note that evaluating the possible optimality of a next solution, say $X' \in \Phi$, requires only $O(|X'|)$ time. Evaluating the necessary of optimality of X is a little more complex, since it requires computing the difference $\overline{w}_e - F^*(S_{\{e\}}^+)$ for each $e \in X$ and, consequently, the

ISBN: 978-989-95079-6-8

overall running time of evaluating the necessary optimality is $O(|X|f(|E|))$.

Let us consider the problem of computing an element deviation interval. The following proposition is true and its proof is similar to the proof of Proposition 1:

Proposition 2. *Let f be a specified element. Then*

$$\underline{\delta}_f = \max\{0, \min_{X\in\Phi_f} F(X, S_X^-) - F^*(S_E^+)\}.$$

From Proposition 2 we immediately get the following corollary:

Corollary 3. *An element $f \in E$ is possibly optimal if and only if* $\min_{X\in\Phi_f} F(X, S_X^-) \leq F^*(S_E^+)$.

Corollary 3 shows a significant difference between the problems with the bottleneck objective and the problems with the linear sum objective discussed for instance in [7, 8]. For the latter problems, deciding whether $\underline{\delta}_f = 0$ for a given element $f \in E$ may be NP-hard even if a deterministic counterpart is polynomially solvable [8]. For the bottleneck problems the situation is much better. From Proposition 2, it follows that if problem $\mathcal{BP}$ is solvable in $f(|E|)$ time, then the bound $\underline{\delta}_f$ for a given element f can be determined in $O(f(|E|))$ time.

We are unable here to provide a general formula for computing the upper bound of an element deviation $\overline{\delta}_f$. Also, the complexity status of the problem of checking whether f is necessarily optimal is unknown. This is an interesting subject of further research. In the next section we show how to compute efficiently quantities $\underline{\delta}_f$ and $\overline{\delta}_f$ when $\mathcal{BP}$ has a matroidal structure.

3.1.1 Matroidal problems

Let us recall the notion of a matroid (see e.g. [11]). A *matroid* is a system $(E, \mathcal{I})$, where E is a ground set and $\mathcal{I}$ is a set of subsets of E closed under inclusions (if $A \subseteq B$ and $B \in \mathcal{I}$ then $A \in \mathcal{I}$) and fulfilling the so-called *growth* property (if $A, B \in \mathcal{I}$ and $|A| < |B|$ then there is $e \in B \setminus A$ such that $A \cup \{e\} \in \mathcal{I}$). The maximal (under inclusion) elements of $\mathcal{I}$ are called *bases*. We will denote the set of all bases by $\mathcal{B}$. The minimal (under inclusion) sets not in $\mathcal{I}$ are called *circuits*. Matroids have the following property, which will be used in this section. Namely, if $B \in \mathcal{B}$ is a base and f is an element such that $f \notin B$, then $B \cup \{f\}$ contains the unique circuit C. Furthermore, for every $e \in C$, set $(B \cup \{f\}) \setminus \{e\}$ is a base.

In a *matroidal problem* the set of feasible solutions Φ consists of all bases of a given matroid, $\Phi = \mathcal{B}$. An example of a matroidal problem is bottleneck spanning tree, where Φ consists of all bases of a *graphic matroid* [11]. Due to a matroidal structure of the interval problem $\mathcal{BP}$, we can simplify the computation of $\underline{\delta}_f$ and provide a method of computing $\overline{\delta}_f$. We first show how to evaluate the possible and necessary optimality of a given element f.

Proposition 3. *An element $f \in E$ is possibly optimal if and only if $\underline{w}_f \leq F^*(S_E^+)$.*

Proof. Using Corollary 3, it is sufficient to proof that $\underline{w}_f \leq F^*(S_E^+)$ if and only if there is a base $B \in \mathcal{B}_f$ such that $F(B, S_E^-) \leq F^*(S_E^+)$, where $\mathcal{B}_f$ stands for the set of all bases that contain element f.

($\Leftarrow$) Obvious.

($\Rightarrow$) Let B^* be an optimal base in scenario S_E^+. If $f \in B^*$, then $B^* \in \mathcal{B}_f$, $F(B^*, S_E^-) \leq F(B^*, S_E^+) = F^*(S_E^+)$ and we are done. Otherwise, $B^* \cup \{f\}$ contains an unique circuit C. Set $B' = (B^* \setminus \{e\}) \cup \{f\}$, $e \in C \setminus \{f\}$, is a base and $B' \in \mathcal{B}_f$. Since $F^*(S_E^+) \geq \underline{w}_f$, we have $F^*(S_E^+) = \max_{e\in B^*} \overline{w}_e \geq \max_{e\in B'} \underline{w}_e$. Thus $F^*(S_E^+) \geq F(B', S_E^-)$ and the proposition follows. □

Proposition 4. *An element $f \in E$ is necessarily optimal if and only if $\overline{w}_f \leq F^*(S_{\{f\}}^+)$.*

Proof. ($\Rightarrow$) If f is necessarily optimal, then it is optimal under all scenarios, in particular, under $S_{\{f\}}^+$. Thus there is a base B containing f such that $F^*(S_{\{f\}}^+) = F(B, S_{\{f\}}^+)$, which yields $F^*(S_{\{f\}}^+) \geq \overline{w}_f$.

($\Leftarrow$) Consider any scenario $S \in \Gamma$. We first show that inequality $w_f(S) \leq F^*(S)$ holds for S. Define scenario S' where $w_f(S') = \overline{w}_f$ and the remaining elements, $e \neq f$, have weights $w_e(S') = w_e(S)$. Using the assumption we get $\overline{w}_f \leq F^*(S_{\{f\}}^+) \leq F^*(S')$. After decreasing the weight of f to $w_f(S)$, we get inequality $w_f(S) \leq F^*(S)$. Let B^* be an optimal base under S. If $f \in B^*$ then f is optimal under S. Otherwise, $B^* \cup \{f\}$ contains an unique circuit C. Set $B' = (B^* \setminus \{e\}) \cup \{f\}$, $e \in C \setminus \{f\}$, is a base such that $B' \in \mathcal{B}_f$. Since $w_f(S) \leq F^*(S)$, $F(B', S) \leq F(B^*, S) = F^*(S)$. Hence B' is an optimal base and f is optimal under S. □

Using a similar reasoning to that in Propositions 3 and 4 one can prove the following proposition, which allows us to determine the deviation interval $\Delta_f = [\underline{\delta}_f, \overline{\delta}_f]$.

Proposition 5. *Let X be a given feasible solution. Then*

$$\underline{\delta}_f = \max\{0, \underline{w}_f - F^*(S_E^+)\}, \overline{\delta}_f = \max\{0, \overline{w}_f - F^*(S_{\{f\}}^+)\}.$$

Propositions 3 and 4 allow us to evaluate efficiently the possible (resp. necessary) optimality of a specified element $f \in E$. It suffices to replace the interval weights with their exact values determined according to S_E^+ (resp. $S_{\{f\}}^+$), compute the optimal value of $F^*(S_E^+)$ (resp. $F(S_{\{f\}}^+)$) in the resulting deterministic problem and compare it with $\underline{w}_f$ (resp. $\overline{w}_f$). This can be done in $O(|E| \log^*(|E|))$ time [4], where $\log^* |E|$ is the iterated logarithm of $|E|$. Note also that using Proposition 3 we can detect all possibly optimal elements in $O(|E| \log^*(|E|))$ because we need to execute an algorithm for the deterministic problem only once. On the other hand, Proposition 4 does not allow us to detect all necessary optimal elements without extra effort. Computing the deviation interval Δ_f costs the same time as evaluating the possible and necessary optimality of f (see Proposition 5).

3.2 Choosing a robust solution

An important task in the interval-valued problem is to choose a *robust solution*, that is the one which performs reasonably well under any possible scenario. A necessarily optimal solution ($\overline{\delta}_X = 0$) is an ideal choice because it is optimal regardless of weight realizations. On the other hand, the possible optimality of a chosen solution ($\underline{\delta}_X = 0$) is the minimum requirement that should be satisfied. A possibly optimal solution

ISBN: 978-989-95079-6-8

always exists. But a necessarily optimal solution rarely exists. Hence the possible optimality is too weak criterion while the necessary optimality seems to be too strong. A solution that minimizes the maximal regret $\overline{\delta}_X$ seems to be a compromise choice. It minimizes the maximal possible deviation from optimum. In literature [9] such a solution is called an *optimal minmax regret solution*.

Proposition 6. *Every optimal minmax regret solution X is possibly optimal and is composed of possibly optimal elements.*

Proof. We use a proof by contraposition. Assume that X is not possibly optimal. From Corollary 1, we have $\max_{e\in X} \underline{w}_e > F^*(S_E^+)$. Let Y^* be an optimal solution in scenario S_E^+. Define $g = \arg\max_{e\in Y^*} \overline{w}_e$ and $h = \arg\max_{e\in X} \underline{w}_e$. Thus $w_h(S) > \overline{w}_g$ for all scenarios $S \in \Gamma$. But $\overline{w}_g \geq w_e(S)$ for all $e \in Y^*$ in every scenario S, which yields $w_h(S) > w_e(S)$ for $e \in Y^*$. From this we conclude that $F(X,S) > F(Y^*,S)$ for all $S \in \Gamma$, which implies $\overline{\delta}_X > \overline{\delta}_Y$. In consequence, X cannot be an optimal minmax regret solution. It is obvious that every possibly optimal solution is composed of possibly optimal elements. □

Therefore, any optimal minmax regret solution X is possibly optimal and it minimizes a distance to the necessary optimality. In other words its deviation interval is of the form $\Delta_X = [0, \overline{\delta}_X]$ where $\overline{\delta}_X$ is the smallest among all $X \in \Phi$. Fortunately, the problem of determining an optimal minmax regret solution can be efficiently solved if the underlying bottleneck deterministic problem $\mathcal{BP}$ is polynomially solvable [10]. Using Equality (3) we get

$$\min_{X\in\Phi} \overline{\delta}_X = \min_{X\in\Phi} \max_{e\in X} \hat{w}_e, \tag{5}$$

where weights $\hat{w}_e = \max\{0, \overline{w}_e - F^*(S^+_{\{e\}})\}$, $e \in E$, are deterministic. So, the minmax regret problem can be reduced to a deterministic problem $\mathcal{BP}$. It can be shown [10] that the minmax regret problem can be solved in $O(|E|+|X^*|f(|E|))$ time, where X^* is such that $F(X^*, S_E^-) = F^*(S_E^-)$ and $f(|E|)$ is the running time of an algorithm for $\mathcal{BP}$.

4 Fuzzy-valued bottleneck combinatorial optimization problems

In this section, we study problem $\mathcal{BP}$ with uncertain element weights, where the uncertainty is modeled by fuzzy intervals. We give a rigorous possibilistic interpretation, solution concepts and some solution algorithms in this setting.

4.1 A possibilistic formalization of the problem

We now give a formalization of problem $\mathcal{BP}$, in which the weights of elements of E are uncertain and they are modeled by fuzzy intervals $\widetilde{W}_e$, $e \in E$, in the setting of *possibility theory* [12]. Let us recall that a *fuzzy interval* $\widetilde{W}_e$ is a fuzzy set in the space of real numbers $\mathbb{R}$, whose membership function $\mu_{\widetilde{W}_e} : \mathbb{R} \to [0,1]$ is normal, quasiconcave and upper semicontinuous on $\mathbb{R}$. We will additionally assume that the support of $\widetilde{W}_e$ is bounded. A membership function of $\widetilde{W}_e$ is regarded as a *possibility distribution* for the values of the unknown weight $\mathfrak{w}_e$. The possibility degree of the assignment $\mathfrak{w}_e = v$ is $\Pi(\mathfrak{w}_e = v) = \pi_{\mathfrak{w}_e}(v) = \mu_{\widetilde{W}_e}(v)$. Let $S = (v_e)_{e\in E}$ be a scenario that represents a state of the world in which $\mathfrak{w}_e = v_e$, for all $e \in E$. Assuming that the weights are unrelated to one another, the fuzzy intervals associated with the weights induce the following possibility distribution over all scenarios in $\mathbb{R}^n$ (see [13]): $\pi(S) = \min_{e\in E} \Pi(\mathfrak{w}_e = v_e) = \min_{e\in E} \mu_{\widetilde{W}_e}(v_e)$ The degrees of possibility and necessity that a solution $X \in \Phi$ is optimal are defined as follows:

$$\begin{aligned} \Pi(X \text{ is optimal}) &= \sup_{\{S:\, X \text{ is optimal in } S\}} \pi(S), \\ \mathrm{N}(X \text{ is optimal}) &= \inf_{\{S:\, X \text{ is not optimal in } S\}} (1 - \pi(S)). \end{aligned} \tag{6}$$

As in the deterministic and interval cases we can characterize the optimality of a solution using the concept of deviation. In the fuzzy problem a solution deviation belongs to a fuzzy interval $\widetilde{\Delta}_X$, whose membership function $\mu_{\widetilde{\Delta}_X}$ is a possibility distribution for δ_X defined as follows:

$$\mu_{\widetilde{\Delta}_X}(v) = \Pi(\delta_X = v) = \sup_{\{S:\, \delta_X(S)=v\}} \pi(S). \tag{7}$$

Since the statement "X is optimal in S" is equivalent to condition $\delta_X(S) = 0$, it is easily seen that (7) generalizes the degrees of optimality (6), that is

$$\begin{aligned} &\Pi(X \text{ is optimal}) = \Pi(\delta_X = 0) = \mu_{\widetilde{\Delta}_X}(0), \\ &\mathrm{N}(X \text{ is optimal}) = \mathrm{N}(\delta_X = 0) = 1 - \sup_{v>0} \mu_{\widetilde{\Delta}_X}(v). \end{aligned} \tag{8}$$

Replacing X with f in (6)-(8) we get the same formulae for element f.

In practice the requirement $\delta_X = 0$ may be very strong and the degree of necessary optimality of every solution $X \in \Phi$ may be very small or even equal to 0. Suppose that a decision maker knows his/her preferences about δ_X and expresses it by a *fuzzy goal* $\widetilde{G}$. The membership function of the fuzzy goal $\mu_{\widetilde{G}}$ is a nonincreasing mapping from $[0,\infty]$ into $[0,1]$ such that $\mu_{\widetilde{G}}(0) = 1$. The value of $\mu_{\widetilde{G}}(\delta_X)$ expresses the degree to which deviation δ_X satisfies the decision maker. We can now replace the strong requirement "X is optimal" ($\delta_X = 0$) with weaker "X is soft optimal" ($\delta_X \in \tilde{G}$). Notice that $\delta_X \in \tilde{G}$ is a fuzzy event and the necessity that it holds can be computed as follows [5, 6]:

$$\mathrm{N}(X \text{ is soft optimal}) = \inf_S \max\{1 - \pi(S), \mu_{\widetilde{G}}(\delta_X(S))\}. \tag{9}$$

One can check that $\mathrm{N}(X \text{ is soft optimal}) = \alpha$ means that for all scenarios S such that $\pi(S) > 1 - \alpha$ it holds $\mu_{\widetilde{G}}(\delta_X(S)) \geq \alpha$ or equivalently $\delta_X(S) \in \widetilde{G}^\alpha = [0, \mu^{-1}_{\widetilde{G}}(\alpha)]$, which represents the suboptimality of X. Function $\mu^{-1}_{\widetilde{G}} : [0,1] \to \mathbb{R} \cup \{+\infty\}$ is a pseudo-inverse of $\mu_{\widetilde{G}}$ that is $\mu^{-1}_{\widetilde{G}}(\alpha) = \sup\{v : \mu_{\widetilde{G}}(v) \geq \alpha\}$.

4.2 Computing the optimality degrees

Every fuzzy interval $\widetilde{U}$ can be decomposed into its λ-cuts, i.e. sets $\widetilde{U}^\lambda = \{u : \mu_{\widetilde{U}}(u) \geq \lambda\}$, $\lambda \in (0,1]$. We will assume that $\widetilde{U}^0$ is the smallest closed set containing the support of $\widetilde{U}$. It can be easily verified (see e.g. [12]) that if $\widetilde{U}$ is a fuzzy interval with a bounded support, then $\widetilde{U}^\lambda$ is a closed interval for every $\lambda \in [0,1]$. The membership function $\mu_{\widetilde{U}}$ can be retrieved from the family of λ-cuts in the following way:

$$\mu_{\widetilde{U}}(u) = \sup\{\lambda \in [0,1] : u \in \widetilde{U}^\lambda\}, \tag{10}$$

ISBN: 978-989-95079-6-8

and $\mu_{\widetilde{U}}(u) = 0$ if $v \notin \widetilde{U}^0$.

Let us denote by $\mathcal{BP}^\lambda$, $\lambda \in [0,1]$, the interval-valued problem $\mathcal{BP}$ with element weights $\widetilde{W}_e^\lambda = [\underline{w}_e^\lambda, \overline{w}_e^\lambda]$, $e \in E$. Note that the scenario set in $\mathcal{BP}^\lambda$ is composed of all scenarios S such that $\pi(S) \geq \lambda$. Now $\widetilde{\Delta}_X^\lambda = [\underline{\delta}_X^\lambda, \overline{\delta}_X^\lambda]$ is the interval of possible values of the deviation of solution X in problem $\mathcal{BP}^\lambda$. It contains all values of the deviation of X, whose possibility of occurrence is not less than λ. From (8) and (10), it follows easily that

$$\Pi(X \text{ is optimal}) = \sup\{\lambda \in [0,1] : \underline{\delta}_X^\lambda = 0\} \tag{11}$$

and $\Pi(X \text{ is optimal}) = 0$ if $\underline{\delta}_X^0 > 0$. A similar reasoning leads to the following equality:

$$\mathrm{N}(X \text{ is soft optimal}) = 1 - \inf\{\lambda : \overline{\delta}_X^\lambda \leq \mu_{\widetilde{G}}^{-1}(1-\lambda)\}. \tag{12}$$

Replacing expression $\mu_{\widetilde{G}}^{-1}(1-\lambda)$ with 0 in (12) we get a formula for computing the degree of necessary optimality of X. Exactly the same considerations apply to elements. It is enough to replace X with f in formulae (11)–(12).

Equations (11) and (12) form a theoretical basis for calculating the values of the optimality degrees. The function $\underline{\delta}_X^\lambda$ is nondecreasing, so in order to compute the degree of possible optimality of solution X we can apply a binary search technique on $\lambda \in [0,1]$. Condition $\underline{\delta}_X^\lambda = 0$ means that X is possibly optimal in $\mathcal{BP}^\lambda$ and it can be checked efficiently by using Corollary 1. Similarly, $\overline{\delta}_X^\lambda$ is nonincreasing and $\mu_{\widetilde{G}}^{-1}(1-\lambda)$ is nondecreasing function of λ. Hence the degree of necessary soft optimality (and necessary optimality) can also be computed by using a binary search. We obtain $\overline{\delta}_X$ using Corollary 2. The calculations can be done in $\mathcal{O}(I(|E|)\log \epsilon^{-1})$ time, where $\epsilon > 0$ is a given precision and $I(|E|)$ is time required to find the bound (the lower or the upper) of deviation interval $\widetilde{\Delta}_X^\lambda$ in the corresponding interval problem $\mathcal{BP}^\lambda$. Of course the same reasoning can be repeated for elements.

Fuzzy deviations $\widetilde{\Delta}_X$ and $\widetilde{\Delta}_f$ can also be determined via the use of λ-cuts. That is, we compute interval $\widetilde{\Delta}_X^\lambda = [\underline{\delta}_X^\lambda, \overline{\delta}_X^\lambda]$ (resp. $\widetilde{\Delta}_f^\lambda$) in the interval valued $\mathcal{BP}^\lambda$ for suitably chosen λ-cuts. Then the fuzzy quantity $\widetilde{\Delta}_X$ (resp. $\widetilde{\Delta}_f$) is reconstructed from its λ-cuts by using equality (10). This method gives an approximation of $\widetilde{\Delta}_X$ (resp. $\widetilde{\Delta}_f$). Its overall running time is $\mathcal{O}(rI(|E|))$ time, r is the number of chosen λ-cuts and $I(|E|)$ is time required to determine $\widetilde{\Delta}_X^\lambda$ (resp. $\widetilde{\Delta}_f^\lambda$) in the interval problem $\mathcal{BP}^\lambda$.

From the above it follows that the running time of the proposed methods heavily relies on the interval case, because they can be reduced to the optimality analysis in a sequence of the interval-valued problems $\mathcal{BP}^\lambda$. Thus the main difficulty lies in the interval case. Fortunately, the optimality analysis in the interval case can be efficiently done, see Corollaries 1,2 and Proposition 1 for a solution and Corollary 3 and Proposition 2 for an element. Unfortunately, we cannot give an efficient method for asserting necessary optimality and determining the upper bound on possible values of the deviation of an element in a general problem $\mathcal{BP}^\lambda$. However, we have proposed such method when problem $\mathcal{BP}^\lambda$ has a matroidal structure (see Section 3.1.1).

4.3 Choosing a robust solution

We now propose some concepts of choosing a robust solution in the fuzzy-valued problem $\mathcal{BP}$. An ideal choice is a solution with the highest degree of certainty that it is optimal under all possible scenarios, i.e. an optimal solution to the following problem:

$$\max_{X \in \Phi} \mathrm{N}(X \text{ is optimal}) = \max_{X \in \Phi} \mathrm{N}(\delta_X = 0). \tag{13}$$

A solution of (13) is called a *best necessarily optimal solution*. However, this concept has a drawback, since the criterion used in (13) is very strong. Namely, a solution X such that $\mathrm{N}(X \text{ is optimal}) > 0$ may not exist or even if it exists, its necessary optimality degree may be very small. To overcome this drawback, we can replace the degree of necessary optimality with the degree of soft necessary optimality, which leads to the following problem:

$$\max_{X \in \Phi} \mathrm{N}(X \text{ is soft optimal}). \tag{14}$$

A solution to (14) is called a *best necessarily soft optimal solution*. Let us recall that this solution concept was first applied to fuzzy linear programming in [5, 6]. We can see from (12) that problem (14) is equivalent to the following one:

$$\begin{array}{ll} \min & \lambda \\ s.t. & \overline{\delta}_X^\lambda \leq \mu_{\widetilde{G}}^{-1}(1-\lambda), \\ & \lambda \in [0,1], \\ & X \in \Phi. \end{array} \tag{15}$$

If problem (15) is infeasible then $\mathrm{N}(X \text{ is soft optimal}) = 0$ for all solutions $X \in \Phi$. If λ^* is the optimal objective value of (15) and X is a best necessarily soft optimal solution, then $\mathrm{N}(X \text{ is soft optimal}) = 1 - \lambda^*$. Since $\overline{\delta}_X^\lambda$ is nonincreasing and $\mu_{\widetilde{G}}^{-1}(1-\lambda)$ is nondecreasing function of λ, problem (15) can also be solved by the binary search technique on $\lambda \in [0,1]$. Note that $\overline{\delta}_X^\lambda$ is the maximal regret of X in $\mathcal{BP}^\lambda$ (see Section 3.2). Thus to find the optimal λ^*, we seek solution that minimizes $\overline{\delta}_X^\lambda$ for a fixed λ, i.e. an optimal minmax regret solution in $\mathcal{BP}^\lambda$ (see problem (5)). An optimal solution X^* to problem (15), determined by finding an optimal minmax regret solution at λ-cut, has a nice property. Namely, it is an optimal minmax regret solution in $\mathcal{BP}^{\lambda^*}$ and, by Proposition 6, it is also possibly optimal one in $\mathcal{BP}^{\lambda^*}$. Thus $\mathrm{N}(X^* \text{ is soft optimal}) \leq \Pi(X^* \text{ is optimal})$. Consequently,

$$\begin{array}{rcl} \mathrm{N}(X^* \text{ is optimal}) & \leq & \mathrm{N}(X^* \text{ is soft optimal}) \\ & \leq & \Pi(X^* \text{ is optimal}). \end{array}$$

From this, we conclude that the concept of a necessary soft optimality is a natural extension of the minmax regret approach to the fuzzy case. The running time of the method based on the binary search is $\mathcal{O}(I(|E|)\log \epsilon^{-1})$ time, where $\epsilon > 0$ is a given precision and $I(|E|)$ is time required for seeking an optimal minmax regret solution in $\mathcal{BP}^\lambda$. It is polynomial for the bottleneck path, the bottleneck assignment, the bottleneck spanning tree and the bottleneck matroid base problems.

ISBN: 978-989-95079-6-8

4.4 A parametric approach

We now present a parametric approach to finding a best necessary optimal solution. A similar method can be applied to compute fuzzy deviations $\widetilde{\Delta}_X$ and $\widetilde{\Delta}_f$. It is easy to check that the following parametric problem may be associated with problem (15) (see also problems (15) and (5))

$$\overline{\delta}^{\lambda} = \min_{X\in\Phi} \overline{\delta}^{\lambda}_X = \min_{X\in\Phi} \max_{e\in X}\{\hat{w}^{\lambda}_e\}, \quad \lambda \in [0,1], \tag{16}$$

where $\hat{w}^{\lambda}_e = \max\{0, \overline{w}_e - F^*(S^{+\lambda}_{\{e\}})\}$, $e \in E$, are parametric weights (functions of a parameter λ), $S^{+\lambda}_{\{e\}}$ is the scenario in which we fix the weight of element e to $\overline{w}^{\lambda}_e$ and the weights of elements $g \in E \setminus \{e\}$ to $\underline{w}^{\lambda}_g$. In order to solve (16), we need to compute parametric weights $\hat{w}^{\lambda}_e = \max\{0, \overline{w}_e - F^*(S^{+\lambda}_{\{e\}})\}$, $e \in E$. Determining these weights requires computing functions $F^*(S^{+\lambda}_{\{e\}})$ for each $e \in E$. This can be done by applying known parametric techniques. Observe that $F^*(S^{+\lambda}_{\{e\}})$ is a function of parameter $\lambda \in [0,1]$. We wish to compute sequences $0 = \lambda_0 < \lambda_1 < ... < \lambda_k = 1$ and $X_0, \ldots, X_{k-1}$ such that X_i is an optimal solution for $\lambda \in [\lambda_i, \lambda_{i+1}]$, $i = 0, \ldots, k-1$. Having these sequences it is easy to describe function $F^*(S^{+\lambda}_{\{e\}})$ and, in consequence, $\hat{w}^{\lambda}_e$ for $\lambda \in [0,1]$. It turns out that if elements weights are linear functions of λ for each $e \in E$, then for some particular bottleneck problems their parametric counterparts can be efficiently solved (see e.g. [14]). Now, we are ready to solve (16) with parametric weights $\hat{w}^{\lambda}_e$. Again, we compute sequences $0 = \lambda_0 < \lambda_1 < ... < \lambda_k = 1$ and optimal solutions (optimal minmax regret solutions) $X_0, \ldots, X_{k-1}$ and the upper bounds for their deviations $\overline{\delta}^{\lambda}_{X_i}$, $\lambda \in [\lambda_i, \lambda_{i+1}]$, $i = 0, \ldots, k-1$. Values $\overline{\delta}^{\lambda}_{X_i}$, $\lambda \in [\lambda_i, \lambda_{i+1}]$, $i = 0, \ldots, k-1$, provide an analytical description of function $\overline{\delta}^{\lambda}$ (deviations of optimal minmax regret solutions) for $\lambda \in [0,1]$. To find a best necessary optimal solution, we need solve the following problem

$$\lambda^* = \arg \min_{\lambda\in[0,1]} \{\overline{\delta}^{\lambda}, \mu^{-1}_{\widetilde{G}}(1-\lambda)\},$$

which is equivalent to finding an intersection of $\overline{\delta}^{\lambda}$ and $\mu^{-1}_{\widetilde{G}}(1-\lambda)$. An optimal minmax regret solution that corresponds to the optimal value λ^* is a best necessary optimal solution.

5 Conclusions

In this paper, we have studied a general bottleneck combinatorial optimization problem with uncertain element weights modeled by fuzzy intervals, whose membership functions are regarded as possibility distributions for the values of the unknown weights. We have described, in this setting, the notions of possible and necessary optimality of a solution and an element and the necessary soft optimality of a solution. These notions are natural generalizations of the ones introduced in the interval-valued case. In order to choose a robust solution, we compute a best necessary soft optimal solution. This concept of choosing a solution is also a generalization, to the fuzzy case, of the known from literature minmax regret criterion [9]. Thus, we have shown in the paper that there exists a link between interval and possibilistic uncertainty. Hence, we have discussed first the interval-valued case and then we have extended the notions and the methods introduced for the interval-valued problem to the fuzzy-valued one. Indeed, the optimality evaluation and choosing a robust solution in the fuzzy problem boil down to solving a number of interval problems $\mathcal{BP}^{\lambda}$. Both problems can be solved in polynomial time if the corresponding interval counterparts are polynomially solvable, that holds true for a wide class of classical bottleneck combinatorial problems.

Acknowledgment

The first author was partially supported by Polish Committee for Scientific Research, grant N N111 1464 33.

References

[1] Abraham P. Punnen. A linear time algorithm for the maximum capacity path problem. *European Journal of Operational Research*, 53:402–404, 1991.

[2] Harold N. Gabow and Robert E. Tarjan. Algorithms for two bottleneck optimization problems. *Journal of Algorithms*, 9:411–417, 1988.

[3] Paolo M. Camerini. The minimax spanning tree problem and some extensions. *Information Processing Letters*, 7:10–14, 1978.

[4] Abraham P. Punnen. A fast algorithm for a class of bottleneck problems. *Computing*, 56:397–401, 1996.

[5] Masahiro Inuiguchi and Masatoshi Sakawa. Robust optimization under softness in a fuzzy linear programming problem. *International Journal of Approximate Reasonning*, 18:21–34, 1998.

[6] Masahiro Inuiguchi. On possibilistic/fuzzy optimization. In *Foundations of Fuzzy Logic and Soft Computing, IFSA 2007*, Lecture Notes in Artificial Intelligence 4529, pages 351–360. Springer-Verlag, 2007.

[7] Adam Kasperski and Paweł Zieliński. On possibilistic combinatorial optimization problems. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 324–329, Hong Kong, 2008.

[8] Adam Kasperski and Paweł Zieliński. Minmax regret approach and optimality evaluation in combinatorial optimization problems with interval and fuzzy weights. *European Journal of Operational Research*, to appear.

[9] Panos Kouvelis and Gang Yu. *Robust Discrete Optimization and its applications*. Kluwer Academic Publishers, 1997.

[10] Igor Averbakh. Minmax regret solutions for minimax optimization problems with uncertainty. *Operations Research. Letters*, 27:57–65, 2000.

[11] C. Papadimitriou and K. Steiglitz. *Combinatorial Optimization*. Dover Publications, INC, Mineola, New York, 1998.

[12] Didier Dubois and Henri Prade. *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, New York, 1988.

[13] Didier Dubois, Hélène Fargier, and Vincent Galvagnon. On latest starting times and floats in activity networks with ill-known durations. *European Journal of Operational Research*, 147:266–280, 2003.

[14] Timothy M. Chan. Finding the shortest bottleneck edge in a parametric minimum spanning tree. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 232—-240, 2005.

ISBN: 978-989-95079-6-8

Web Usage Mining: users' navigational patterns extraction from web logs using Ant-based Clustering Method

Kobra Etminani [1] Mohammad-R. Akbarzadeh-T. [2] Noorali Raeeji Yanehsari [3]

1 Dept. of Computer Engineering, Ferdowsi University of Mashhad, Iran
2 Depts of Electrical Engineering and Computer Engineering, Ferdowsi University of Mashhad, Iran
3 Iran Khodro, Khorasan, Iran
Email: etminani@wali.um.ac.ir, akbarzadeh@ieee.org, raeeji@ikkco.ir

Abstract— *Web Usage Mining is the process of applying data mining techniques to the discovery of usage patterns from data extracted from Web Log files. It mines the secondary data (web logs) derived from the users' interaction with the web pages during certain period of Web sessions. Web usage mining consists of three phases, namely preprocessing, pattern discovery, and pattern analysis. In this paper, web logs of our university web server logs (http://www.um.ac.ir/) are pre-processed. Then, ant-based clustering is applied to pre-processed logs to extract frequent patterns for pattern discovery and then it is displayed in an interpretable format. Result of this paper would be useful for our university web site owner.*

Keywords— web usage mining, ant-based clustering, frequent pattern extraction, web mining.

1 Introduction

Web usage information mining could help to engage new customers, maintain current customers, track customers who are leaving web site, and so on [1]. Usage information can be extracted to increase web server efficiency by pre-fetching and caching strategies [2].Based on several researches done in the area of web mining, we can broadly classify it into three domains: web content mining, web structure mining, and web usage mining.

Web content mining is the process of extracting knowledge from web documents such as text and multimedia. Knowledge extraction from the structure of web and hyperlink references is called web structure mining. Web usage mining is the process of knowledge exploitation from the secondary data [3]. By secondary data, we mean browser logs, user profiles, web server access logs, registration data, user sessions or transactions, cookies, user queries, mouse clicks and any other data that is the result of interaction with the Web.

In the following section we give an overview over the related work. Section 3 explains the problem and ant-based clustering in more detail. Section 4 goes into detail how we implement the proposed method i.e. the experimental procedure of the proposed method and results are shown. We concluded our work in Section 5.

2 Related Work

There are several methods for pattern extraction from the secondary data (web logs) such as Masseglia et al. [2] [7], Spiliopoulou et al. [8], Bonchi et al. [9], Hay et al. [10], Zhu et al. [11], Nakagawa and Mobasher [12]. In [2], Masseglia et al. invented a new method named PSP (Prefix-tree for Sequential Patterns) that follows the basic principles of GSP (Generalized Sequential Patterns algorithm) in [13]. The tree structure in PSP is similar to the *prefix-tree* used in [14]. At the *k*th step, the tree has a depth of *k*. Each branch from the root to a leaf stands for a candidate sequence.

In [8], Spiliopoulou et al. proposed the WUM (Web utilization miner) tool that determines patterns which are noticeable from the statistical view. So, it emphasizes the frequency (minimum support) of the patterns.

Hay et al. in [10] applied the notion of time embedded in the navigations to cluster user sessions. They used an alignment algorithm to compute the distance between sessions.

Zhu et al. in [11] considered navigating between web sites as a Markov chain and mainly discussed about Markov model problems.

In [12], Nakagawa and Mobasher show that depending on whether the propositions are based on frequent itemsets or frequent sequences, the features of the site have an impact on the quality of the refinement shown to users.

In this paper, we propose a new method for extracting patterns from web logs based on ant clustering algorithm. We apply ant-based clustering for pattern discovery, other similar methods applied ant colony clustering to segregate visitors [15]. Some methods applied Markov models for modeling user web navigation behavior. But the proposed method has the similarity and speed of ant-based clustering algorithm rather than other clustering algorithms.

3 Problem description

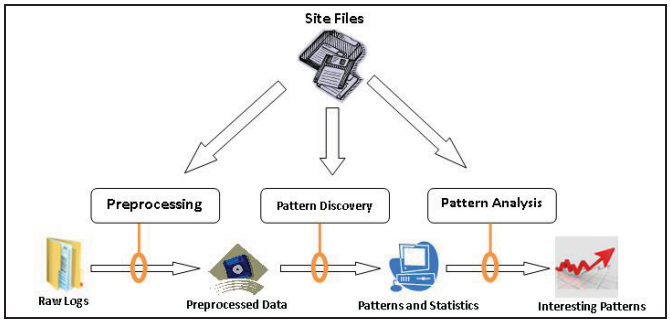


Figure 1: Usage Mining Process.

As shown in Figure 1 from [6], there are three main tasks for performing Web Usage Mining: Pre-processing, Pattern

ISBN: 978-989-95079-6-8

Discovery and Pattern Analysis. This section presents an overview of the tasks for each step.

At least two log file formats exists: Common Log File format (CLF) and Extended Log File format ([16] for more details). Our university log file consists of these fields: Date, Time, client IP address, Method, URI stem, Protocol status, Bytes sent, Protocol version, Host, User Agent and Referrer.

3.1 Pre-processing

As said in [6], pre-processing *"consists of converting the usage, content, and structure information contained in the various available data sources into the data abstractions necessary for pattern discovery"*.

This step can break into at least four substeps: Data Cleaning, User Identification, Session Identification and Formatting.

Unneeded data will be deleted from raw data in web log files in the data cleaning step. When a user requests a page, the request is added to the Log File; but if this page contains images, javascripts, flash animations, video, etc., they are added to the Log file as well. Most of the time, these are not needed for pattern discovery and should be omitted from log files.

There are some methods for User Identification, the second phase: detecting cookies, Identd, through IP address and user name. W3C [17] define cookies as *"data sent by the server to the client, data locally stored in cookies and is sent to the server with each request"*. [18] and [19] use cookies to identify users. But they have two main problems: the users can lock the use of cookies, so server cannot store information locally in the user machine; other problem is the user can delete the cookies.

Identd can be used for user identification. It is a protocol defined in RFC 1413 [20]. It allows us to identify connected users with the unique TCP connection. The problem with Identd is that users should configure with this protocol.

Another way for user detection is through user names added in the log file in field authuser. But this field can be empty (default value dash(-)) according to server/user command.

At last we can identify users through their IP address registered in each record in log file. We used this method although it has several problems: different users can be registered with same IP address; same user can be registered with different IP addresses. But with the help of session identification we can identify a user with IP address and be sure of solving the first problem. The second problem is not important because in this paper, a specific user is not wanted. We want to recognize general user's pattern.

For Session Identification in third phase, first we should define session timeout. Different timeouts considered are between 25-30 minutes .The thirty minute timeout is based on the results of [21]. We assume 30 minutes session timeout for the experimental procedure.

And in the last phase of pre-processing step, we should display pre-processed data in a suitable format, for the second step, pattern discovery.

3.2 Pattern Discovery

As stated in [6], pattern discovery *"draws upon methods and algorithms developed from several fields such as statistics, data mining, machine learning and pattern recognition"*. Several methods and techniques have already been developed for this step as summarized below:

- *Statistical Analysis* such as frequency analysis, mean, median, etc.
- *Clustering* of users help to discover groups of users with similar navigation patterns (provide personalized Web content).
- *Classification* is the technique to map a data item into one of several predefined classes.
- *Association Rules* discover correlations among pages accessed together by a client.
- *Sequential Patterns* extract frequently occurring inter-session patterns such that the presence of a set of items s followed by another item in time order.
- *Dependency Modeling* determines if there are any significant dependencies among the variables in the Web.

We choose clustering to discover users' navigational patterns. Our clustering method is based on Ant-based Clustering algorithm explained in Section 3.4.

3.3 Pattern Analysis

Pattern Analysis is the final stage of WUM (Web Usage Mining), which involves the validation and interpretation of the mined pattern.

- *Validation*: to eliminate the irrelevant rules or patterns and to extract the interesting rules or patterns from the output of the pattern discovery process.
- *Interpretation*: the output of mining algorithms is mainly in mathematic form and not suitable for direct human interpretations.

3.4 Ant-based Clustering

Deneubourg et al. in [22] proposed ant-based clustering and sorting. In the case of ant-based clustering and sorting, two related types of natural ant behaviors are modeled. When clustering, ants gather items to form heaps. And when sorting, ants discriminate between different kinds of items and spatially arrange them according to their properties [23].

Lumer and Faieta in [24] proposed ant-based data clustering algorithm (shown in Figure 2), which resembles the ant behavior described in [22].

As shown in Figure 2, the agents (ants) and data are randomly initialized on a toroidal grid. By moving agents, data is sorted according to its neighbors.

The picking and dropping probabilities, given a grid position and a particular data item *i*, are computed using the density functions:

$$p_{pick}(i) = \left(\frac{k^{+}}{k^{+} + f(i)}\right)^{2} \quad (1)$$

$$p_{drop}(i) = \begin{cases} 2f(i) & if\ f(i) < k^{-} \\ 1 & otherwise, \end{cases} \quad (2)$$

ISBN: 978-989-95079-6-8

where $k+$ and $k-$ are constants, and $f(i)$ is a neighborhood function:

$$f(i) = \max(0, \frac{1}{\sigma^2} \sum_{j \in L} (1 - \frac{d(i,j)}{\alpha})) \quad (3)$$

where, $d(i, j) \in [0, 1]$ is a measure of the dissimilarity between data points i and j , $\alpha \in [0, 1]$ is a data-dependent scaling parameter, and σ^2 is the size of the local neighborhood L.

Handl & Meyer in [25] proposed an extension of this algorithm where the parameter α is adaptively updated during the execution of the algorithm.

We applied Handl & Meyer's Ant-based clustering algorithm for detecting user's patterns.

```
(1)  Procedure Lumer and Faieta
(2)     randomly scatter data items on the toroidal grid
(3)     randomly place agents on the toroidal grid
(4)     for t = 1 to max_iterations
(5)        j = random agent
(6)        move agent j randomly by stepsize grid cells
(7)        l = does agent j carry a data item?
(8)        e = is agent j's grid position occupied by a data
           item?
(9)        if (i = TRUE) and (e = FALSE) then
(10)          i = data item carried by agent j
(11)          drop = (random() ≤ Pdrop(i)) // see
              equations (2) and (3)
(12)          if drop = TRUE then
(13)             Let agent j drop data item i at its current
                 position
(14)          end if
(15)       end if
(16)       if (l = FALSE) and (e = TRUE) then
(17)          i = data item at agent j's grid position
(18)          pick = (random() ≤Ppick(i)) // see equations
              (1) and (3)
(19)          if pick = TRUE then
(20)             let agent j pick up data item i
(21)          end if
(22)       end if
(23)    end for
(24) end procedure
```

Figure 2: Lumer & Faieta 's ant-based clustering algorithm [24].

4 Experimental Procedure and Results

4.1 Experimental Procedure

We used our university (http://www.um.ac.ir/) web server logs of two weeks for the experimental procedure.

For the first step, i.e. pre-processing, we wrote a c++ program compiled using gcc without any optimization flags. As mentioned in section 3.1, this step contains four phases. First, we omit unneeded records from log file. The log file consists of image files (gif, jpg, bmp, jpeg …) and other unneeded resources like javascripts and errors. For user identification, we use IP address and session timeout of 30 minutes. So, a user with an IP address has 30 minutes to navigate in the web site. After a user's navigational sequence is extracted, it is displayed in a suitable format for the second step, Pattern Discovery. We then classify the URLs of the web site into 28 groups and assign a number to each group, as shown in Table 1. Then, each user's requested URL is substituted with its corresponding number. The output of this step is a file that consists of records, each record representing a navigational sequence of users in numbers.

For the second step, Patten Discovery, we used Ant-based Clustering algorithm based on [25]. Julia Handl's written source code is used in Java and changes are made according to our application. Each user's navigational sequence is defined as an array with 28 elements. The element i is 1 if the related user had seen one of the pages in group i; otherwise it is 0.

Dissimilarity of two sequences $s1$ at point i and $s2$ at point j in the grid is computed through the following equation:

$$d(i,j) = \frac{\sqrt{\sum_{k=1}^{28} (s1_k - s2_k)^2}}{28} \quad (4)$$

$d(i,j)$ becomes 1 when two sequences do not have any similar elements, and becomes 0 when they are exactly the same.

Table 1: Classification of URLs.

1	Web site content	2	Newsletter
3	Miscellaneous	4	Black board
5	Web services	6	Search
7	News	8	Help
9	FAQ	10	Societies
11	Publications	12	Abstract
13	Photos	14	Web links
15	RSS	16	People
17	Staff	18	Student
19	Faculty	20	Professor
21	User	22	Guest
23	Research	24	About university
25	Education	26	Download
27	English homepage	28	homepages

The output of this program is a grid that contains numbers: >-1 and =-1 indicates if there is/is not a data item, respectively. So, clusters should be extracted according to these numbers and size of the local neighborhood, σ^2. Figure 3 shows the positions of the data points in different phases of running this algorithm. Figure 3.a shows the distribution of data points at the first step of program execution. As the execution continues, Figures 3.b, 3.c and 3.d shows the results after 200, 400 and 600 iterations, respectively. Clusters are created through moving of the ants.

We examined several numbers of ants for the clustering step. The experiments shows that the smaller the number of ants, the slower will be the rate of convergence, but also better results in clustering the data items. On the other hand, the larger the number of ants, the faster will be the speed of convergence, but also weaker results in clustering. So, empirically, we choose number of ants to be %20 of the number of data items.

At last for the third step, Pattern Analysis, the results are shown in an interpretable way. The output of the second step

ISBN: 978-989-95079-6-8

is the clusters of users' navigational sequence. Each cluster may include lots of members, so we should represent a cluster with one pattern and display it in a suitable format for users.

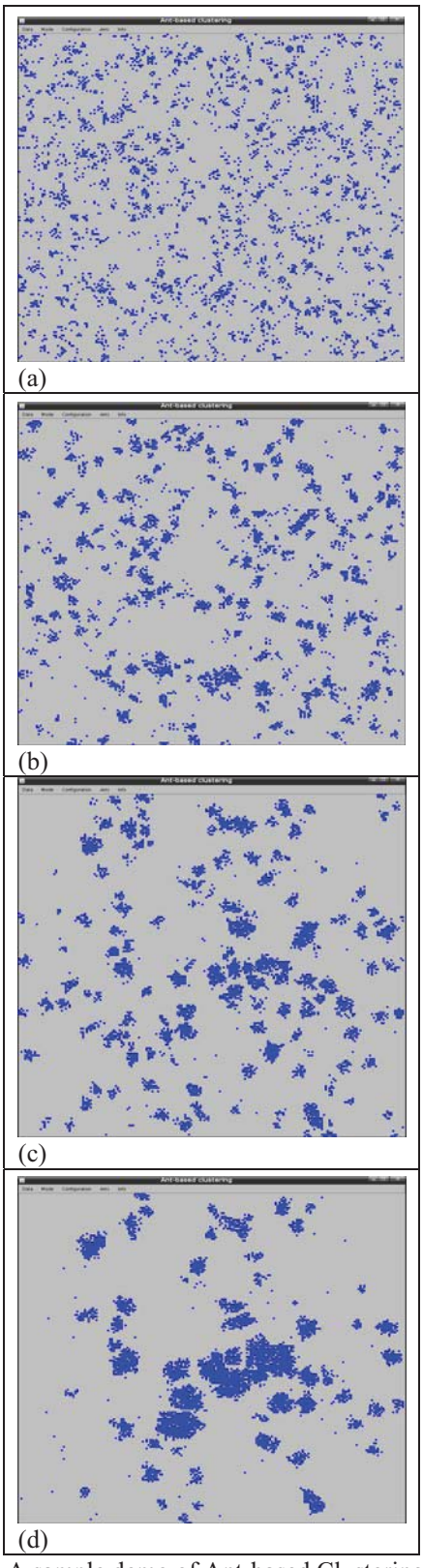


Figure 3: A sample demo of Ant-based Clustering of users' sequences.

The clustering algorithm results in clusters of similar sequences, which is a key element for sequence alignment. The alignment of sequences leads to a weighted sequence (as defined in [26]), represented as follows: $SA = \{I1 : n1, I2 : n2, \ldots, Ir, nr\} : m$. In this representation, m stands for the total number of sequences involved in the alignment. Ip ($1 \leq p \leq r$) is an itemset represented as ($xi1 : mi1 , \ldots xit : mit$), where mit is the number of sequences containing the item xi at the npth position in the aligned sequences. Finally, np is the number of occurrences of itemset Ip in the alignment. Figure 4 shows an example of the alignment process. The approximate sequential pattern can be obtained by specifying k: the number of occurrences of an item in order for it to be displayed. For instance, with the sequence $SA14$ from Figure 4 and $k = 3$, the filtered aligned sequence will be: {2, 4, 8}

To show the results, the aligned patterns that consist of numbers is re-substituted by the groups in Table 1.

S1={	2	4	5	8}
S2={	2		5	8}
SA12={	(2:2):2	(4:1):1	(5:2):2	(8:2):2}
SA12={	(2:2):2	(4:1):1	(5:2):2	(8:2):2}
S3={	2	4		8}
SA13={	(2:3):3	(4:2):2	(5:2):2	(8:3):3}
SA13={	(2:3):3	(4:2):2	(5:2):2	(8:3):3}
S4={	2	4		8}
SA14={	(2:4):4	(4:3):3	(5:2):2	(8:4):4}

Figure 4: Alignment processing example.

4.2 Results

The above mentioned procedure is applied to our university (http://www.um.ac.ir/) web server logs. All the experiments were performed on an Intel Core 2 Duo 2.5GHz PC running Linux (Mint).

The log files are collected in two different weeks in 2008: first week of June and the middle week of August each have 300 MB. Two different weeks are selected and the result of applying the proposed method on each week is compared. Applying ant-based clustering on each week, on average, took 300 seconds for 800 iterations. Average hourly web traffic for each group in these two weeks can be seen in Figures 5 and 6, respectively. The extracted behaviors from these two weeks are listed in Table 2 and Table 3, respectively. As shown in Table 2 and Table 3, behaviors of users in these two weeks were similar in most of the cases. One should notice that these results show just most frequent patterns of users' navigational behaviors, not all of the users' behaviors. Patterns of single navigations are not listed above, i.e. patterns that contain only one navigation.

Comparing Table 2 and Table 3 with Figure 5 and Figure 6, one could understand that pattern <Web site content → News > is the most popular one. <Professor → Homepages> is frequent, too. On the other hand, groups that do not exist in the patterns, like FAQ and Societies, have the least access in hours of a day, too.

Comparing our method to other methods, it has some advantages:

- It is simple. One only needs to define a suitable dissimilarity function for the clustering step and do

ISBN: 978-989-95079-6-8

not need to involve in complex mathematical relations.

- It does not depend on pattern length. Other similar methods are limited in terms of the length of the extracted patterns. The proposed method is able to extract patterns of any length.

Different patterns can be extracted depending on the occurrence of that page group in a cluster. So, even page groups that are less accessed can be extracted.

These kinds of results are useful for web site owners. They can put their advertisements in these sequences, because these are the most frequent ones. They are useful for page pre-fetching, too.

5 Conclusions

In this paper, we have proposed a new method to extract navigational patterns from web logs. Ant-based clustering has been used for this purpose. It needs a neighborhood function to be defined for. After the clustering is completed, alignment processing has been applied to the extracted sequences in each cluster and extract the representative for each cluster.

We apply the following procedure on our university web server logs for two different weeks (http://www.um.ac.ir/) and the results are satisfactory and true according to the hourly web traffic. These kinds of results are suitable for web site owners, for example, to put their advertisements there or to even change the structure of the web site according to users' navigational behavior.

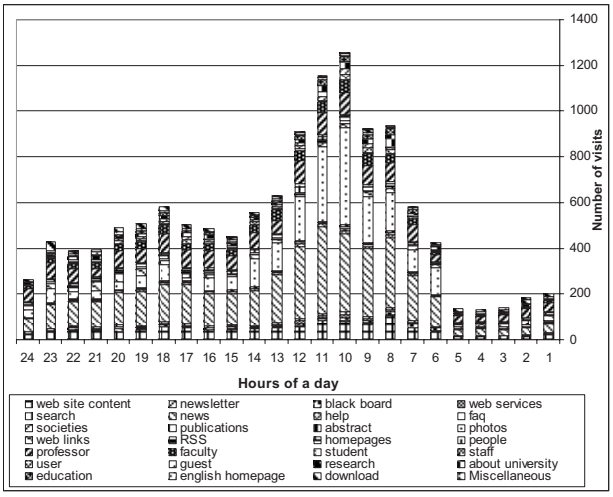


Figure 5: Hourly web traffic of the first week.

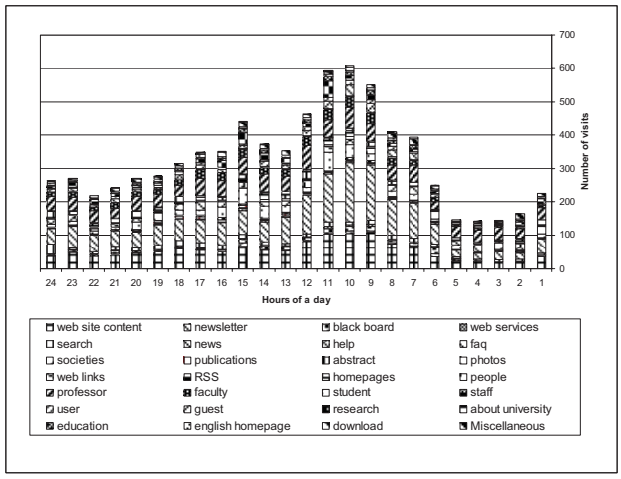


Figure 6: Hourly web traffic of the second week.

Table 2: Extracted behaviors from the first week.

Web site content	→	News						
Web services	→	Search						
Web site content	→	News	→	Faculty	→	Education		
News	→	Faculty	→	Education				
News	→	Professor						
Web site content	→	News	→	Professor	→	Homepages		
Web site content	→	News	→	Help	→	About university	→	Education
Web site content	→	News	→	Search	→	Photos	→	Web links → Professor → About university → Downloads
Web site content	→	Black board						
English homepage	→	Guest						
Professor	→	Homepages						
Professor	→	Faculty						
News	→	Faculty						

Table 3: Extracted behaviors from the second week.

Web site content	→	News						
Faculty	→	English Homepage						
Web site content	→	News	→	Users				
Web site content	→	Users						
Web site content	→	Search	→	News	→	Web links	→	Professor → About university → Download
Web site content	→	Faculty						
News	→	Professor	→	Homepages				
Faculty	→	Education						
Web site content	→	Photos	→	Web links	→	Professor	→	Homepages → English homepage → Download
News	→	Professor						
News	→	Download						

ISBN: 978-989-95079-6-8

References

[1] A. Abraham. Natural Computation for Business Intelligence from Web Usage Mining, *Proceeding of Seventh International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNAC2005),* pp. 3-11, 2005.

[2] F. Masseglia, P. Poncelet, and R. Cicchetti, An Efficient Algorithm for Web Usage Mining, *Networking and Information Systems Journal (NIS)*, 2(5-6), pp. 571-603, 1999.

[3] R. Cooley, Web Usage Mining: Discovery and Application of Interesting patterns from Web Data, Ph. D. Thesis, University of Minnesota, Department of Computer Science, 2000.

[4] P. Pirolli, J. Pitkow, and R. Rao, Silk From a Sow‘s Ear: Extracting Usable Structures from the Web, *Proceeding on Human Factors in Computing Systems (CHI’96)*, ACM Press, pp. 118-125, 1996.

[5] M. Spiliopoulou, and L.C. Faulstich, WUM: A Web Utilization Miner, *Proceeding of EDBT Workshop on the Web and Data Bases (WebDB’98)*, Springer Verlag, pp. 109-115, 1999.

[6] J. Srivastava, R. Cooley, M. Deshpande, and P.N. Tan, Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, *SIGKDD Explorations*, 1(2), pp. 12-23, 2000.

[7] F. Masseglia , P. Poncelet, M. Teisseire, A. Marascu, Web usage mining: extracting unexpected periods from web logs, *Data Min Knowl Disc*, 16, pp.39–65, 2008.

[8] M. Spiliopoulou , L.C. Faulstich , K. Winkler , A data miner analyzing the navigational behavior of web users, *Proceeding of the workshop on machine learning in user modeling of the ACAI’99 international conference Creta*, Greece, 1999.

[9] F. Bonchi , F. Giannotti , C. Gozzi , G. Manco, M. Nanni , D. Pedreschi, C. Renso , S. Ruggieri, Web log data warehousing and mining for intelligent web caching , *Data Knowl Eng*, 39(2), pp. 165–189, 2001.

[10] B. Hay , G. Wets, K. Vanhoof , Mining navigation patterns using a sequence alignment method , *Knowl Inf Syst*, 6(2), pp.150–163, 2004.

[11] Zhu, J., Hong, J., Hughes, J.G. 2002 Using Markov chains for link prediction in adaptive web sites. *Proceeding of soft-ware: first international conference on computing in an imperfect world*, Belfast, UK, pp. 60–73, 2002.

[12] M. Nakagawa, B. Mobasher, Impact of site characteristics on recommendation models based on association rules and sequential patterns. *Proceeding of the IJCAI’03 workshop on intelligent techniques for web personalization*, Mexico, 2003.

[13] R. Srikant, R. Agrawal, Mining sequential patterns: generalizations and performance improvements. *Proceeding of the 5th international conference on extending database technology (EDBT’96)*, pp. 3–17, France, 1996.

[14] A. Mueller, Fast sequential and parallel algorithms for association rules mining: a comparison, Technical report CS-TR-3515, Department of Computer Science, University of Maryland-College Park, 1995.

[15] A. Abraham, V. Ramos, Web Usage Mining Using Artificial Ant Colony Clustering and Genetic Programming, *Congress on Evolutionary Computation (CEC)*, IEEE 2003.

[16] W3C extended log file format. Available at ,http://www.w3.org/TR/WD-logfile

[17] WCA. Web characterization terminology & definitions. Available at ,http://www.w3.org/1999/05/WCA-terms/ .

[18] M. Eirinaki, M.Vazirgiannis, Web Mining for Web Personalization, Athens University of Economics and Business, 2003.

[19] J. Huysmans, B. Baesens , J. Vanthienen , Web Usage Mining: A Practical Study, Katholieke Universiteit Leuven, Dept. of Applied Economic Sciences, 2003.

[20] RFC 1413. Identification Protocol. Available at ,http://www.rfceditor.org/rfc/rfc1413.txt .

[21] L. Catledge, J. Pitkow, Characterizing browsing behaviors on the World Wide Web, *Computer Networks and ISDN Systems*, 27(6), 1995.

[22] J. Deneubourg -L., S. Goss, N. Franks, A. Sendova-Franks, C. Detrain, L. Chrétien, The dynamics of collective sorting: robot-like ants and ant-like robots. *Proceeding of the first international conference on simulation of adaptive behavior*, pp. 356–365, MIT Press, 1991.

[23] J. Handl, B. Meyer, Ant-based and Swarm-based clustering, *Swarm Intelligence*, 1, pp. 95–113, 2007.

[24] E. Lumer, B. Faieta, Diversity and adaptation in populations of clustering ants. *Proceeding of the third international conference on simulation of adaptive behaviour*, pp. 501–508, MIT Press, 1994.

[25] J. Handl, B. Meyer, Improved ant-based clustering and sorting in a document retrieval interface. *Proceeding of the Seventh International Conference on Parallel Problem Solving from Nature*, Vol. 2439 of Lecture Notes in Computer Science, pp. 913-923, Germany: Springer-Verlag, 2002.

[26] H. Kum, J. Pei, W. Wang, D. Duncan, ApproxMAP: approximate mining of consensus sequential patterns, *Proceeding of SIAM International Conference on data mining*, CA, 2003.

ISBN: 978-989-95079-6-8

Multiplicative and implicative importance weighted averaging aggregation operators with accurate andness direction

Henrik Legind Larsen

Department of Electronic Systems
Aalborg University, Denmark
Email: legind@aaue.dk

Abstract— *Importance weighted averaging aggregation operators play a key role in utilizations of electronic data and information resources for retrieving, fusing, and extracting information and knowledge, as needed for decision making. Two central issues in the choice of such operators are the kind of importance weighting and the andness (or conjunction degree) of the operator. We present and discuss two main kinds of importance weighting, namely multiplicative and implicative, each yielding a particular kind of operator for a particular kind of aggregation problems. Importance weighting generalizations of each kind are proposed for each of two classes of averaging operators, namely the Power Means and the Ordered Weighted Averaging operators, each in a De Morgan dual form for increased efficacy. For each class is proposed a function for a rather accurate direct control of the andness. Operators of the same kind appear to behave rather similarly at the same andness, independent of the class of averaging operators generalized.*

Keywords— Aggregation operators, averaging operators, andness, orness, importance weighting, implicative importance.

1 Introduction

With the increasing amount of electronically accessible data in databases, document bases, and data streams, and the need of efficient utilization of such data, aggregation operators have attained new interest. This is due to the central role of such operators in the key reasoning tasks, such as information fusion and pattern recognition, applied in utilizations of many kinds, *e.g.*, information retrieval, information extraction, object recognition, and knowledge discovery. Of particular interest are the aggregations operators between AND (conjunction) and OR (disjunction), *i.e.*, the averaging operators, or generalized conjunction/disjunction functions (GCD) [1], especially the importance weighted generalizations of such operators.

An importance weighted averaging aggregation operator aggregates a number of arguments, each qualified by an importance weight, into a single score. We distinguish between two kinds of importance weighting, namely multiplicative and implicative. Each kind provides an importance weighting generalization of the unweighted averaging operator, in the sense that the latter is retained by the case where all arguments have the same importance weight. We propose such generalizations of two classes of averaging operators, namely the Power Means (PM) and the Ordered Weighted Averaging (OWA) operators [2], each in a De Morgan dual form for increased efficacy, and each with an accurate direct control of the andness, *i.e.*, the degree the operator aggregates like an AND rather than an OR. The generalizations of the PM include the AIWA operators [3].

A few words on the notation applied in the following. $\bigodot_i a_i$, where $\bigodot$ is an operator, is an abbreviation for $\bigodot_{i=1}^{n}(a_i)$; a bold letter like $\mathbf{x}$ denotes the vector $(x_1, \ldots, x_n)$; $\overline{x}$ denotes the standard negation $\overline{x} = 1 - x$; $\overline{\mathbf{x}}$ denotes $(\overline{x_1}, \ldots, \overline{x_n})$. Some specific letters are used all over with the same meaning: I denotes the real unit interval $[0, 1]$; n is the dimension, *i.e.*, number of arguments aggregated by the averaging operator considered; ρ denotes the targeted andness of an averaging operator; $\mathbf{a}$ denotes a vector of arguments $(a_1, \ldots, a_n) \in I^n$; $\mathbf{v}$ denotes an importance weighting vector $(v_1, \ldots, v_n) \in I^n$ that is max-normalized, *i.e.*, $\max_i v_i = 1$; $\mathbf{w}$ denotes an importance weighting vector $(w_1, \ldots, w_n) \in I^n$ that is sum-normalized, *i.e.*, $\sum_i w_i = 1$; $h : I^n \times I^n \to I : (\mathbf{v}, \mathbf{a}) \mapsto h(\mathbf{v}, \mathbf{a})$ is an importance weighted averaging operator, such that v_i is the importance of a_i; for a particular operator h, its name acronym and parameters are attached on the form $h^{\text{acronym}}_{\text{parameters}}$.

Section 2 introduces some basic concepts applied in the following. Sections 3 and 4 introduce, for each class of averaging operators, the PM based and the OWA based, the basis of the class, its multiplicative and implicative importance weighting generalizations, and a function for accurate direct control of the andness of its operators. Sections 5 and 6 discuss, for each kind of importance weighting, the multiplicative and the implicative, the kind of applications they apply to and some key issues of relevance for their application. Section 7 compares and discusses the effect of the two kinds of importance weighting on the common problem of ordering a set of options by their score. Section 8 concludes.

2 Basic concepts

2.1 The andness of an averaging operator

The andness may be viewed as the degree of universal quantification over the arguments, with AND (andness $= 1$) representing *all* and OR (andness $= 0$) representing *at least one*. In this view, the andness is the degree to which all arguments, rather than at least one argument, must support the result of the aggregation. In importance weighted averaging operators, each argument is considered to the degree it is important. The dual interpretation applies to the orness, defined by: orness $= 1 -$ andness.

Since, AND and OR are evaluated by the operators min and max, respectively, the andness of an averaging operator h may be defined as distance between h and the max, relative to the distance between the min and the max, with operators h, max, and min evaluated as the mean of these operators over the ar-

ISBN: 978-989-95079-6-8

gument space, as proposed by Dujmović [4]:

$$\text{andness}(h) = \frac{E(\max_i(\mathbf{x})) - E(h(\mathbf{x}))}{E(\max_i(\mathbf{x})) - E(\min_i(\mathbf{x}))} \qquad (1)$$

where $E(f(\mathbf{x}))$ is the mean of $f(\mathbf{x})$ over the argument space I^n, as defined by $E(f(\mathbf{x})) = \int_{I^n} f(\mathbf{x})d\mathbf{x}$. The terms *andness* and *orness* were coined by Yager for a measure of the OWA operators [2] that is fully consistent with the measure defined by (1).

The andness is, by OWA operators, obtained by positional weights (OWA weights), such that the OWA aggregate is the sum of the products of the i'th OWA weight and the i'th largest argument, whereas it by PM operators is obtained by a power function, yielding emphasis on the smaller or larger valued arguments to the degree the PM represents, respectively, AND and OR (see, *e.g.*, [1]).

2.2 The two kinds of importance weighting

A well-known example of an importance weighted averaging operator is the Weighted Arithmetic Mean (WAM) defined by $h^{\text{WAM}}(\mathbf{v}, \mathbf{a}) = \sum_i (w_i, a_i)$ where, for all i, $w_i = \frac{v_i}{\sum_i v_i}$, for which $\text{andness}(h^{\text{WAM}}) = \text{orness}(h^{\text{WAM}}) = \frac{1}{2}$. The unweighted case, as obtained by $v_i = 1$ for all i, is the Arithmetic Mean (AM): $h^{\text{WAM}}((1, 1, \ldots, 1), \mathbf{a}) = \frac{1}{n}\sum_i a_i = h^{\text{AM}}(\mathbf{a})$. The WAM appears to be the only importance weighted averaging operator that represents both kinds of importance weighting, namely the multiplicative and the implicative. For other degrees and andness than $\frac{1}{2}$, the two kinds of importance weighting provide, as we shall see, different behaviors that apply to different kinds of aggregation problems.

While Multiplicative importance Weighted Averaging (for short, MWA) operators are weighted means, Implicative importance Weighted Averaging (for short, IWA) operators are logic operators for pattern matching inference. In general, the andness of an IWA operator "penalizes" (decreases) the score (the aggregate) to the degree that there are criteria that are important, but not satisfied, while the orness "rewards" (increases) the score to the degree that there are criteria that are important and satisfied.

By the importance weighting, the arguments are, essentially, transformed by the importance weights in a way giving the desired effect, depending on the kind of importance weighting and the andness, in the aggregation. In principle, an argument a with the importance weight w is, by the two kinds of weighting, transformed as follows: by MWA, to the product wa, and by IWA at andness ρ to $\rho(w \Rightarrow a) + \overline{\rho}(\overline{w} \Rightarrow \overline{a})$ where $\Rightarrow$ is a fuzzy implication operator. It is the multiplication in the MWA case and the implication in the IWA case that give name to the two kinds of importance weighting.

As the fuzzy implication $\Rightarrow$ for IWA operators, we choose the Reichenbach implication, $\Rightarrow_{\text{R}}$, as defined by $(v \Rightarrow_{\text{R}} a) = \overline{v\overline{a}} = 1 - v(1-a)$. The reasons for this choice are, first, that it allows us easily to obtain the WAM at $\rho = \frac{1}{2}$, and, second, that the choice of the fuzzy implication appears not to be critical for the behavior of such operators. The latter was tested empirically; the paper doesn't leave space for presenting details from the test.

A key difference between between MWA and IWA operators is that in MWA operators, as opposed to IWA operators, the effect of the importance weights decreases from full effect to no effect as the andness goes from $\frac{1}{2}$ to one of its extremes, 0 or 1. For an appropriate choice of the fuzzy implication operator, the IWA operators represent, like the MWA operators, the WAM at andness $\frac{1}{2}$. These behaviors of the two kind of importance weighted averaging operators are illustrated by Figure 1. We notice, as the figure indicates, that in cases where

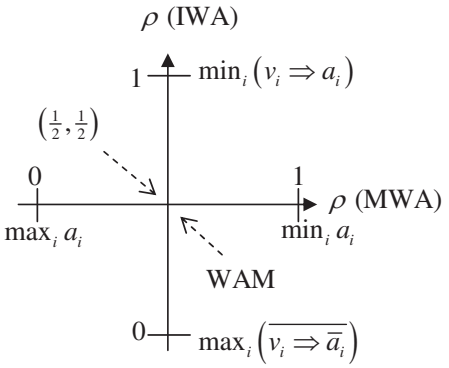


Figure 1: Illustration of MWA and IWA operators

we are completely uncertain about the proper kind of importance weighting and andness for an application, WAM may be a good choice as a starting point.

2.3 The two classes of operators generalized

In the following, we shall in particular consider importance weighted generalizations of two classes of operators, namely the Power Means (PM) and the Ordered Weighted Averaging (OWA) operators, each in a De Morgan dual form for increased efficacy. For each class, we propose two andness-directing functions, $\psi(\rho)$ and $\psi(\rho, n)$, each yielding a value ($\in [1, \infty[$) of the andness controlling parameter of the class. While $\psi(\rho)$ directly yields an andness of the operator that, in general, is somewhat close to the target ρ, $\psi(\rho, n)$ yields the target rather accurately, due to it also being dependent of n, as applied in the definition by (1).

For each class, we provide a common expression of the two kinds of importance weighting generalization, such that the choice of a kind is controlled by a parameter $\gamma \in \{0, 1\}$, yielding the multiplicative generalization by $\gamma = 0$ and the implicative generalization by $\gamma = 1$. The operators are in their common generalized form referred to as, respectively, Andness-directed importance Weighted averaging (AWA) operators and andness-directed importance Weighted OWA operators (W-OWA). Figure 2 gives an overview over the operator name acronyms applied.

Basis	*Without* importance weighting	*With* importance weighting		
		Common class	**M**ultipl. $\gamma = 0$ **MWA**	**I**mpl. $\gamma = 1$ **IWA**
Power Means	AA	AWA	A**M**WA	A**I**WA
OWA	OWA	W-OWA	**M**W-OWA	**I**W-OWA

Figure 2: Operator and operator class name acronyms applied

ISBN: 978-989-95079-6-8

3 AWA operators

3.1 Basis in the Power Means

The AWA (Andness-directed importance Weighted Averaging) operators are based on the Power Means (PM) that (in their unweighted form) are defined by $h^{\text{PM}}_\alpha(\mathbf{a}) = (\frac{1}{n}\sum_i a_i^\alpha)^{\frac{1}{\alpha}}$, $\alpha \in \mathbb{R}$. When α goes from $-\infty$ to $+\infty$, the andness of h^{PM}_α goes from 1 down to 0. It has an asymmetric behavior around andness $= \frac{1}{2}$, as obtained at $\alpha = 1$; in particular, it is "mandatory" for $\alpha \leq 0$, corresponding to andness $\tilde{>}\frac{2}{3}$, in the sense that $h^{\text{PM}}_\alpha(\mathbf{a}) = 0$, if $\min_i(a_i) = 0$, which normally is an undesired property. [1] However, by the De Morgan dual variant PM′, defined by (i) $h^{\text{PM}}_\alpha(\mathbf{a})$, $\alpha \geq 1$, for andness $\leq \frac{1}{2}$, and (ii) $1 - h^{\text{PM}}_\alpha(\overline{\mathbf{a}})$, $\alpha \geq 1$, for andness $\geq \frac{1}{2}$, we obtain a symmetric behavior around andness $= \frac{1}{2}$ and avoid the "mandatory" property. For both (i) and (ii), the arithmetic mean is obtained at $\alpha = 1$; when α goes to $+\infty$, the andness of $h^{\text{PM}'}_\alpha$ goes in case (i) to 0 and in case (ii) to 1.

The AWA operators comprise the multiplicative and implicative importance weighted generalizations of PM′, namely the AMWA (Andness-directed Multiplicative importance Weighted Averaging) operators and the AIWA (Andness-directed Implicative importance Weighted Averaging) operators. The AIWA operators were introduced in [3] where we analyzed their properties and showed that they implement implicative importance weighting by the Riechenbach implication. The AMWA operators is the multiplicative weighting variant of the AIWA operators.

3.2 Definition of AWA operators

The AWA operators are defined by the following semi-recursive expression, where the case of $\rho > \frac{1}{2}$ is defined by its duality to the case of $\rho \leq \frac{1}{2}$:

$$h^{\text{AWA}}_\rho(\mathbf{v},\mathbf{a}) = \begin{cases} \left(\sum_i (y_i a_i^\alpha)\right)^{\frac{1}{\alpha}} & \rho \leq \frac{1}{2} \\ 1 - h^{\text{AWA}}_{\overline{\rho}}(\mathbf{v},\overline{\mathbf{a}}) & \rho > \frac{1}{2} \end{cases} \tag{2}$$

where $\alpha = \psi^{\text{AA}}(\rho) = \overline{\rho}/\rho$, and for all i:

$$y_i = \frac{v_i^{\alpha^\gamma}}{\sum_i v_i^{\alpha^\gamma}} \tag{3}$$

with $\gamma \in \{0,1\}$, $\gamma = 0$ (by which $y_i = \frac{v_i}{\sum_i v_i} = w_i$) yielding the AMWA operators, and $\gamma = 1$ (by which $y_i = \frac{v_i^\alpha}{\sum_i v_i^\alpha}$) yielding the AIWA operators.

By $\mathbf{v} = (1,1,\ldots,1)$ we obtain the common unweighted case of AMWA and AIWA that we shall refer to as Andness-directed Averaging (AA) operators; thus, these operators are, in the case of $\rho \leq \frac{1}{2}$, defined by: $h^{\text{AA}}_\rho(\mathbf{a}) = h^{\text{AWA}}_\rho((1,1,\ldots,1),\mathbf{a}) = (\frac{1}{n}\sum_i a_i^\alpha)^{\frac{1}{\alpha}}$.

3.3 Accurate control of AWA andness

Above, the parameter α is, for a targeted andness ρ, defined by $\alpha = \psi^{\text{AA}}(\rho) = \overline{\rho}/\rho$ in the case $\rho \leq \frac{1}{2}$ (and, by duality, as implemented by the semi-recursive call, $\alpha = \psi^{\text{AA}}(\overline{\rho}) = \rho/\overline{\rho}$ in the case $\rho > \frac{1}{2}$). While by $\psi^{\text{AA}}(\rho)$, andness(h^{AWA}) is equal to ρ, if $\psi^{\text{AA}}(\rho) \in \{0, \frac{1}{2}, 1\}$ and, otherwise, somewhat close to ρ, a much more precise approximation can be obtained by also considering the dimension n, namely through replacing $\psi^{\text{AA}}(\rho)$ by $\psi^{\text{AA}}(\rho, n)$ as defined by:

$$\psi^{\text{AA}}(\rho,n) = (-\log_2 \rho)^{1+(\log_2 n)^{1/3}} \tag{4}$$

An empirical test of the error $\|\rho - \text{andness}(h^{\text{AWA}})\|$, showed that by $\psi^{\text{AA}}(\rho)$ the average error over $\rho \in I$ increases from about 0.03 to about 0.09 as n increases from 2 to 100, while it by $\psi^{\text{AA}}(\rho, n)$ remains around 0.005.

3.4 Decomposability of AWA operators

A property of particular interest for applications of AWA operators is their decomposability, allowing us to update the aggregate if the value of an argument has changed or a new argument has to be considered, without re-computing the whole aggregate. [3] Thus, considering the common case by AWA (2) for $\rho \leq \frac{1}{2}$, let $h^{\text{AWA}}_\rho(\mathbf{v},\mathbf{a}) = c$ to be updated with a new argument a_{n+1} with the importance weight v_{n+1}. Let $\cup$ denote the concatenation of vectors, such that $\mathbf{x} \cup (x_{n+1}) = (x_1,\ldots,x_n) \cup (x_{n+1}) = (x_1,\ldots,x_{n+1})$. Then, the updated aggregate is computed as the AWA aggregate of c and the new argument, as defined by: $h^{\text{AWA}}_\rho(\mathbf{v} \cup (v_{n+1}), (\mathbf{a} \cup (a_{n+1})) = h^{\text{AWA}}_\rho((s, v_{n+1}), (c, a_{n+1})) = \left(\frac{s^{\alpha^\gamma} c^\alpha + v_{n+1}^{\alpha^\gamma} a_{n+1}^\alpha}{s^{\alpha^\gamma} + v_{n+1}^{\alpha^\gamma}}\right)^{1/\alpha}$, with $s = \left(\sum_{i=1}^n v_i^{\alpha^\gamma}\right)^{1/\alpha^\gamma}$.

4 W-OWA operators

4.1 Basis in the OWA operators

The properties of OWA operators [2] are controlled by a vector of sum-normalized position weights $\mathbf{u}$ (called OWA weights) such that the OWA aggregate of an argument vector $\mathbf{a}$ is defined by $h^{\text{OWA}}(\mathbf{a}) = \sum_i (u_i, a_{(i)})$ where $(\cdot)$ is an index permutation such that $a_{(1)} \geq \cdots \geq a_{(n)}$. Two key properties of OWA operators are their andness, defined by andness$(\mathbf{u}) = \frac{1}{n-1}\sum_i ((i-1)u_i)$, and their normalized dispersion (or entropy), defined by $\text{ndisp}(\mathbf{u}) = -\frac{1}{\ln n}\sum_i (u_i \ln u_i)$. This andness measure is consistent with (1).

In [5, 6], Yager proposed an importance weighted generalization of quantifier guided OWA operators that essentially is a multiplicative importance weighting generalization, which we here shall refer to as MW-OWA operators. They are defined by $h^{\text{MW-OWA}}_f(\mathbf{v},\mathbf{a}) = \sum_i (u_i, a_{(i)})$, where $u_i = f(s_i) - f(s_{i-1})$, $i = 1,\ldots,n$, with $s_0 = 0$ and $s_i = \frac{\sum_{k=1}^i v_{(k)}}{\sum_{k=1}^n v_k}$ for $i > 0$, $f : I \to I$ is a regular increasing monotone quantifier (*i.e.*, $f(0) = 0$, $f(1) = 1$, and $f(x) \geq f(y)$ if $x > y$), such as $f(x) = x^\beta$, $\beta \geq 0$, and $(\cdot)$ as before is a permutation of the index such that $a_{(1)} \geq \cdots \geq a_{(n)}$. It is easily seen that when β goes from 0 to $+\infty$, andness$(h^{\text{MW-OWA}})$ goes from 0 to 1, with $h^{\text{MW-OWA}}_f$ representing the max (with andness $= 0$) at $\beta = 0$, the WAM (with andness $= \frac{1}{2}$) at $\beta = 1$, and the min (with andness $= 1$) at $\beta \to +\infty$.

In the case $v_i = 1$, $i = 1,\ldots,n$, we get $s_i = \frac{i}{n}$, and, hence, $\mathbf{u}$ representing the underlying OWA weights (for the OWA operator without importance weighting) that by $f(x) = x^\beta$ are somewhat close to the MEOWA weights [7, 8], *i.e.*, the unique OWA weighting vector with the maximum dispersion at the given andness and dimension n. However, the behavior is not symmetric around andness $= \frac{1}{2}$; in fact, while the dispersion

[1] The mandatory threshold of andness $\tilde{>}\frac{2}{3}$ enforced by the PM is, from a practical point of view, rather arbitrary.

ISBN: 978-989-95079-6-8

of the OWA weights for andness $> \frac{1}{2}$ (*i.e.*, $\beta > 1$) is rather close to the maximum, it is much less for andness $< \frac{1}{2}$ (*i.e.*, $\beta < 1$). For instance, for $n = 5$ the maximum normalized dispersion at andness $= 0.8$ and andness $= 0.2$ is 0.755 in both cases, while the OWA weights with $f(x) = x^\beta$ have the normalized dispersion 0.744 in the first case, but only 0.717 in the second case. Therefore, an improvement with a near optimal dispersion for all degrees of andness, is obtained by defining $h_f^{\text{MW-OWA}}$ as above for andness $\geq \frac{1}{2}$, and by its De Morgan duality for andness $< \frac{1}{2}$, namely (with a change of notation to express the andness ρ) $h_\rho^{\text{MW-OWA}}(\mathbf{v}, \mathbf{a}) = 1 - h_{\overline{\rho}}^{\text{MW-OWA}}(\mathbf{v}, \overline{\mathbf{a}})$. By this improvement, the normalized dispersion in the second case (andness $= 0.2$) above is increased to 0.744, as in the first case, *i.e.*, in both cases only 0.011 lower than the maximum.

4.2 Definition of the W-OWA operators

In the following, we present an extension of the duality based MW-OWA operators to a common class, W-OWA, comprising both the multiplicative importance weighting generalization, MW-OWA, and an implicative importance weighting generalization, IW-OWA, with parametric control of the kind of generalization. The W-OWA operators are defined by:

$$h_\rho^{\text{W-OWA}}(\mathbf{v}, \mathbf{a}) = \begin{cases} \left(\sum_i (u_i b_{(i)})\right) & \rho \geq \frac{1}{2} \\ 1 - h_{\overline{\rho}}^{\text{W-OWA}}(\mathbf{v}, \overline{\mathbf{a}}) & \rho < \frac{1}{2} \end{cases} \tag{5}$$

where, for all i, b_i is defined by:

$$b_i = (v_i^{\gamma(2\rho-1)} \Rightarrow_{\text{R}} a_i) = 1 - v_i^{\gamma(2\rho-1)}(1 - a_i) \tag{6}$$

with $\gamma \in \{0, 1\}$, $\gamma = 0$ (by which $b_i = a_i$) yielding the class of MW-OWA operators, and $\gamma = 1$ (by which $b_i = (v_i^{(2\rho-1)} \Rightarrow_{\text{R}} a_i)$) yielding the class of IW-OWA operators; $u_i = f(s_i) - f(s_{i-1}), i = 1, \ldots, n$, with $s_0 = 0$ and

$$s_i = \frac{\sum_{k=1}^{i} v_{(k)}}{\sum_{k=1}^{n} v_k} \tag{7}$$

for $i > 0$, and $f(x) = x^\beta$, with $\beta = \psi^{\text{W-OWA}}(\rho) = \rho/\overline{\rho}$; finally, $(\cdot)$ is an index permutation such that $b_{(1)} \geq \cdots \geq b_{(n)}$. We notice that the Reichenbach implication applied above may be replaced by any fuzzy implication.

By $\mathbf{v} = (1, 1, \ldots, 1)$ we obtain the common case of MW-OWA and IW-OWA, without importance weighting, namely the duality based, quantifier guided OWA operators that, in the case of $\rho \geq \frac{1}{2}$, are defined by: $h_\rho^{\text{OWA}}(\mathbf{a}) = h_\rho^{\text{W-OWA}}((1, 1, \ldots, 1), \mathbf{a}) = \sum_i (u_i a_{(i)})$ where $u_i = f(\frac{i}{n}) - f(\frac{i-1}{n})$, with $f(x) = x^\beta$.

4.3 Accurate control of W-OWA andness

Above, the parameter β is, for a targeted andness ρ, defined by $\beta = \psi^{\text{W-OWA}}(\rho) = \rho/\overline{\rho}$ in the case $\rho \geq \frac{1}{2}$ (and, by duality, as implemented by the semi-recursive call, $\beta = \psi^{\text{W-OWA}}(\overline{\rho}) = \overline{\rho}/\rho$ in the case $\rho < \frac{1}{2}$). While by $\psi^{\text{W-OWA}}(\rho)$, andness$(h^{\text{W-OWA}})$ is equal to ρ, if $\rho \in \{0, \frac{1}{2}, 1\}$ or $n \to +\infty$, and, otherwise, close to ρ, a more precise approximation can be obtained by also considering the dimension n, namely through replacing $\psi^{\text{W-OWA}}(\rho)$ by $\psi^{\text{W-OWA}}(\rho, n)$ as defined by:

$$\psi^{\text{W-OWA}}(\rho, n) = \frac{0.5 + n\rho}{0.5 + n\overline{\rho}} \tag{8}$$

For instance, by $\rho = 0.8$ (and, through the duality, by $\rho = 0.2$), we get $\beta = \psi^{\text{W-OWA}}(0.8) = \frac{0.8}{1-0.8} = 4$. For $n = 5$, this value of β gives an andness of about of about 0.86, i.e., 0.06 more than the targeted andness, while we by $\beta = \psi^{\text{W-OWA}}(0.8, 5) = \frac{0.5+5\cdot 0.8}{0.5+5(1-0.8)} = 3.0$ obtain the desired andness of 0.8. As n increases from 2 to 100, the average over $\rho \in I$ of the error $\|\rho - \text{andness}(h_\rho^{\text{W-OWA}})\|$ decreases from about 0.07 to about 0.002 by $\psi^{\text{W-OWA}}(\rho)$, and from about 0.0003 to about 0.00001 by $\psi^{\text{W-OWA}}(\rho, n)$.

5 On the multiplicative importance weighting

MWA operators are essentially weighted means that, qua means, are symmetric, monotonic increasing, continuous, and idempotent. They produce means in the interval $[\min_i a_i, \max_i a_i]$, and are monotonically increasing with the orness. If an MWA operator has the andness $= \frac{1}{2}$, it represents the WAM.

5.1 Applications of MWA operators

A common application is for *estimation of a utility variable*. In this case, each argument represents the estimation by an expert (or the measure by some source), while its importance weight represents the decision maker's confidence in the experts's ability to estimate the correct value. The andness and the orness represent decision maker's risk attitude, namely, the degree of, respectively, pessimism and optimism. The aggregate is in this case an estimate of the utility variable, considering the decision maker's risk attitude and, for each expert, the estimate by the expert and the decision masker's confidence in the expert.

Another application is for *selection between (alternative) options*, where the option with the highest score (weighted mean) is winning. In this case, each argument represents the degree to which the option considered has a particular property of interest, while its importance weight represents the importance of having the property to a high degree. The andness represents the degree to which all properties of interest must be present to a high degree. The outcome of the averaging is a ranking of options, possible with a threshold distinguishing acceptable options from unacceptable options.

5.2 Discontinuity property of MWA operators

While MWA and IWA operators both represent the WAM at andness $= \frac{1}{2}$, a key difference is that for MWA, as opposed to IWA, the effect of the importance weights decreases as the andness converges to one of its extremes, 1 or 0. In these cases, MWA evaluates to the smallest argument with a positive weight, yielding a discontinuity, since a small change in an importance weight, from, say, 0.01 to 0, may give a drastic change in the aggregate.

For instance, let h_ρ^{MWA} be an MWA operator (like AMWA or MW-OWA) at andness ρ, and assume $\mathbf{a} = (0.9, 0.1)$. Then $h_1^{\text{MWA}}((1, 0.01), \mathbf{a}) = 0.1$, while $h_1^{\text{MWA}}((1, 0), \mathbf{a}) = 0.9$. Similarly, $h_0^{\text{MWA}}((0.01, 1), \mathbf{a}) = 0.9$, while $h_0^{\text{MWA}}((0, 1), \mathbf{a}) = 0.1$. In both cases, the small decrease from 0.01 to 0 of an importance weight, caused a drastic change (of size 0.8) in the aggregate.

This behavior is an effect of MWA operators as weighted means. It models the attitude of the decision maker, in the cases of extreme pessimism and extreme optimism, to select

ISBN: 978-989-95079-6-8

the most pessimistic, respectively, most optimistic, estimate (argument) by any source with a positive confidence (importance weight). Such behavior is not acceptable in multicriteria aggregation for object recognition; for instance, it is not acceptable that the unimportant second criterion determines the overall satisfaction as by aggregate $h_1^{\text{MWA}}((1, 0.01), \mathbf{a}) = 0.1$. IWA operators avoid this by only considering criteria to the degree they are important; thus $h_1^{\text{IWA}}((1, 0.01), \mathbf{a}) = h_1^{\text{IWA}}((1, 0), \mathbf{a}) = 0.9$.

5.3 Other MWA operators

The PM, on which the AMWA and AIWA operators are based, are in the family of the quasi-arithmetic means (QM) [9, 10]. The weighted QM (WQM) forms the family of MWA operators defined by $h^{\text{WQM}}(\mathbf{v}, \mathbf{a}) = \phi^{-1}(\sum_i (w_i \phi(a_i)))$ where ϕ is a continuous strictly monotonic function, and ϕ^{-1} is the inverse of ϕ. The Weighted PM (WPM), $h^{\text{WPM}}(\mathbf{v}, \mathbf{a}) = (\sum_i (w_i a_i^\alpha))^{1/\alpha}$, $\alpha \in \mathbb{R}$, belong to this family from which they are derived by $\phi_\alpha(x) = x^\alpha$ (with $\phi_\alpha^{-1}(x) = x^{1/\alpha}$).

By $\phi_\alpha(x) = e^{\alpha x}$ (with $\phi_\alpha^{-1}(x) = \frac{\ln x}{\alpha}$), we obtain the Weighted Exponential Means (WEM), $h_\alpha^{\text{WEM}}(\mathbf{v}, \mathbf{a}) = \frac{1}{\alpha} \ln(\sum_i (w_i e^{\alpha a_i}))$, that converges to the WAM for $\alpha \to 0$, and to $\min_i a_i$ and $\max_i a_i$ for α going to, respectively, $-\infty$ and $+\infty$. A nice property of the WEM is its symmetric behavior around andness $= \frac{1}{2}$ as obtained for $\alpha \to 0$. Unlike the WPM, the WEM do not impose the mandatory property.

6 On the implicative importance weighting

6.1 Applications of IWA operators

Implicative importance weighting is applied by IWA operators for pattern matching inference in multicriteria recognition or classification problems, where an observed object (*e.g.*a document or some physical object) is compared to a goal concept (e.g., a query or a class). The goal concept is intensionally characterized by a set of criteria that, in general, are importance weighted. Each criterion expresses a constraint on the values of an object attribute and may, in general, be represented by a fuzzy subset. The importance weight of a criterion expresses the importance of satisfying the criterion in recognizing or classifying an object as an instance of the goal concept.

This scheme applies for querying of the two main kinds, namely object querying and concept querying. [11, 12] In *object querying*, *e.g.*, document retrieval and database querying, the goal concept forms a query. When posed to an object base, the query determines a fuzzy subset of the set of objects, representing the query's fuzzy extension in that set. An object's degree of membership in the extension can be seen as the degree to which the object is an instance of the goal concept and is also referred to as the object's *score* in the goal concept. In *concept querying*, *e.g.*, object recognition and classification, the characterization of the observed object forms a query. When posed to a concept base, *i.e.*, a base of possible goal concepts or classes, the query determines in this case a fuzzy subset of the set of concepts. The answer may in both cases be presented by a ranked list of the objects or concepts (depending on the query type) for which the score is above a given threshold.

6.2 IWA weighting and reasoning scheme

Let C be a goal concept, characterized by the criteria $C_1, \ldots, C_n$ with the importance weights $v_1, \ldots, v_n$, with v_i being the importance of satisfying C_i, and let $a_i = C_i(x) = \mu_{C_i}(x)$ be the degree to which the object x (or, actually, the constrained attribute of x) satisfies C_i. Then, the inference implemented by IWA operators has a Modus Ponens form that in the case of andness = 1 is expressed by:

$$\frac{\begin{array}{l} a_1 = C_1(x), \ldots, a_n = C_n(x) \\ (v_1 \Rightarrow C_1) \wedge \cdots \wedge (v_n \Rightarrow C_n) \to C \end{array}}{C(x) = \wedge_i (v_i \Rightarrow C_i) \ (= h_{\rho=1}^{\text{IWA}}(\mathbf{v}, \mathbf{a}))} \tag{9}$$

Equation 9 expresses that, for an AND aggregation, the goal concept is satisfied by an object to the degree that the requirement *'for all criteria, the criterion is satisfied if it is important'* is met. This may be expressed by $\forall i(\text{important}(\text{C}_\text{i}) \Rightarrow \text{satisfied}(\text{C}_\text{i}))$ that, by $v_i = \text{important}(\text{C}_\text{i})$ and $a_i = \text{satisfied}(\text{C}_\text{i})$, may be written $\forall i(v_i \Rightarrow a_i)$, or, by the standard evaluation of $\forall$ by the min operator, $\min_i(v_i \Rightarrow a_i)$. By other equivalent expressions of $\forall i(v_i \Rightarrow a_i)$, the requirement can be expressed by other words, for instance, by $\neg\exists i(v_i \wedge \neg a_i)$, *'there does not exist a criterion that is important but not satisfied'*.

The transformation of an argument a by an implicative importance weight v for an averaging aggregation at andness ρ is modeled by a function $g_\rho : I^2 \to I : (v, a) \mapsto g_\rho(v, a)$. Thus, for a fuzzy implication $\Rightarrow$, $g_1(v, a) = (v \Rightarrow a)$ is a transformation for an AND aggregation, and $g_0(v, a) = \overline{g_{\overline{0}}(v, \overline{a})} = \overline{v \Rightarrow \overline{a}}$ is its De Morgan dual transformation for an OR aggregation. An important property of g_1 and g_0 is that they for $v = 0$ evaluate to the neutral element for the aggregation, namely, respectively, 1 and 0. In general, for $\rho \in I$, we define g_ρ as the andness-orness weighted sum of g_1 and g_0:

$$g_\rho(v, a) = \rho g_1(v, a) + \overline{\rho} g_0(v, a) \tag{10}$$

which in the case of the Reichenbach implication $\Rightarrow_\text{R}$, with $g_1^\text{R}(v, a) = (v \Rightarrow_\text{R} a) = \overline{v\overline{a}}$ $(= 1 - v(1 - a))$ and $g_0^\text{R}(v, a) = \overline{v \Rightarrow_\text{R} \overline{a}} = va$, evaluates to $g_\rho^\text{R}(v, a) = \rho - v(\rho - a)$. In [13], we presented an (implicative) importance weighting generalization of OWA operators and showed that transformation by g_ρ^R satisfies the requirements for such an importance weighting.

6.3 Other IWA operators

In [3], we introduced another approach to implicative importance weighting generalization of OWA operators, in that case the MEOWA operators [7, 8], namely the IW-MEOWA operators that, by the current notation, are defined by:

$$h_\rho^{\text{IW-MEOWA}}(\mathbf{v}, \mathbf{a}) = \frac{h_\rho^{\text{IW-MEOWA}}(\mathbf{u}, \mathbf{v}, \mathbf{a}) - l}{u - l} \tag{11}$$

where $h_\rho^{\text{IW-MEOWA}}(\mathbf{u}, \mathbf{v}, \mathbf{a}) = \sum_i (u_i b_{(i)})$, $\mathbf{u}$ is the MEOWA weighting vector for the given ρ and n, $b_i = g_\rho^\text{R}(v, a) = \rho - v(\rho - a)$, and $(\cdot)$ is an index permutation, such that $b_{(1)} \geq \cdots \geq b_{(n)}$. Finally, l and u are the lower and upper bounds for $h_\rho^{\text{IW-MEOWA}}(\mathbf{v}, \mathbf{a})$, namely, repectively, $h_\rho^{\text{IW-MEOWA}}(\mathbf{u}, \mathbf{v}, (0, \ldots, 0))$ and $h_\rho^{\text{IW-MEOWA}}(\mathbf{u}, \mathbf{v}, (1, \ldots, 1))$. It was shown that (11) provides a linear transformation of $h_\rho^{\text{IW-MEOWA}}(\mathbf{u}, \mathbf{v}, \mathbf{a})$, such that $h_\rho^{\text{IW-MEOWA}}(\mathbf{v}, \mathbf{a})$ represents the WAM at andness $= \rho = \frac{1}{2}$.

ISBN: 978-989-95079-6-8

7 Comparison of MWA and IWA aggregates

While operators of the same kind, MWA or IWA, tend to behave similarly—in the sense that they provide the same ordering of a given set of options at the same andness and the same importance weights—this is, as we may expect, not the case for operators of difference kinds. This is illustrated by the following examples, where the andness of the MWA and W-OWA classes are targeted by, respectively, $\psi^{\text{AA}}(\rho, n)$ and $\psi^{\text{W-OWA}}(\rho, n)$, as denoted by the aggregation operator symbol; thus, $h^{\text{AMWA}}_{\frac{2}{3},2}$ denotes that h^{AMWA} is applied with $\psi^{\text{AA}}(\frac{2}{3}, 2)$.

Assume that the decision problem is characterized by andness $= \frac{2}{3}$, $n = 2$, $\mathbf{v} = (0.4, 1)$ that is sum-normalized to $\mathbf{w} = (0.286, 0.714)$, and consider the two options represented by the argument vectors $\mathbf{a}_1 = (0.1, 0.7)$ and $\mathbf{a}_2 = (0.9, 0.4)$.

By *MWA aggregation*, the second option, as represented by $\mathbf{a}_2$, get the highest score and is therefore ranked higher than the first option, as the following evaluations, all at andness $= \frac{2}{3}$, show: $h^{\text{AMWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_1) = 0.420 < h^{\text{AMWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_2) = 0.474$ and $h^{\text{MW-OWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_1) = 0.454 < h^{\text{MW-OWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_2) = 0.470$. This is also the case for MWA operators of other classes; thus, for the Weighted Geometric Mean (WGM), $h^{\text{WGM}}(\mathbf{v}, \mathbf{a}) = \prod_i a_i^{w_i}$, that at $n = 2$ has andness $= \frac{2}{3}$, we get $h^{\text{WGM}}(\mathbf{v}, \mathbf{a}_1) = 0.401 < h^{\text{WGM}}(\mathbf{v}, \mathbf{a}_2) = 0.504$; similarly for the Weighted Exponential Means (WEM), $h^{\text{WEM}}_{\alpha}(\mathbf{v}, \mathbf{a}) = \frac{1}{\alpha} \ln \left(\sum_i (w_i e^{\alpha a_i}) \right)$, that at $(\alpha, n) = (-3, 2)$ has andness $= \frac{2}{3}$, we get $h^{\text{WEM}}_{-3}(\mathbf{v}, \mathbf{a}_1) = 0.402 < h^{\text{WEM}}_{-3}(\mathbf{v}, \mathbf{a}_2) = 0.484$.

By *IWA aggregation*, the first option, as represented by $\mathbf{a}_1$, get the highest score and is therefore ranked higher than the second option: $h^{\text{AIWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_1) = 0.579 > h^{\text{AIWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_2) = 0.422$ and $h^{\text{IW-OWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_1) = 0.551 > h^{\text{IW-OWA}}_{\frac{2}{3},2}(\mathbf{v}, \mathbf{a}_2) = 0.473$; similarly, for the IW-MEOWA operator (11), we get $h^{\text{IW-MEOWA}}_{\frac{2}{3}}(\mathbf{v}, \mathbf{a}_1) = 0.530 > h^{\text{IW-MEOWA}}_{\frac{2}{3}}(\mathbf{v}, \mathbf{a}_2) = 0.515$.

This behavior may be explained as follows. By IWA aggregation, the first criterion (argument) is ignored to some degree, due to its rather low importance (0.4); the higher satisfaction of the second, highly important criterion by the first option is the sufficient to give this option the highest rank. By MWA aggregation, the very low value of the first argument in the first option contributes to give this option the lowest rank, despite the rather low importance of the argument.

These observations are supported by the experiment in [3], where the comparison of the IWA operators AIWA and IW-MEOWA for a small data set and a set of implicatively importance weighted queries, indicated a quite similar behavior of the two operators.

8 Conclusion

Importance weighting of two kinds, namely multiplicative and implicative, were proposed as generalizations of two classes of averaging aggregation operators, namely the Power Means (PM) and the Ordered Weighted Averaging (OWA) operators. Each class is applied in a De Morgan dual version, yielding symmetric behavior on both side of andness $= \frac{1}{2}$, and, at the same time, avoiding the mandatory property of the PM and obtaining a near maximum dispersion of OWA at all degrees of andness. The two generalizations of a class are represented by a common expression, where the kind of importance weighting is controlled through a binary parameter.

For each class, were proposed an andness-directing function that allows us to obtain a targeted andness for an operator rather accurately through also considering the number of arguments.

For each kind of importance weighting, we presented and discussed how it works in the operator under different degrees of andness, and the kind of application problems they apply to. Operators of different classes, but with the same kind of importance weighting, appear to have rather similar behavior at the same andness.

For a given choice of the kind of importance weighting, computational issues may affect the choice of operator class. While the PM based operators (AWA) in particular require much computationally rather heavy power functions, OWA based operators (W-OWA) in particular require an ordering (sorting) of the arguments. If reevaluation of a set of options by a modified query occurs frequently, AWA operators have an advantage through their decomposability property.

References

[1] J.J. Dujmović and H.L. Larsen. Generalized conjunction/disjunction. *Int. J. Approx. Reason*, 46(3):423–446, 2007.

[2] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Fuzzy Syst.*, 18(1):183–190, 1988.

[3] H. Legind Larsen. Efficient andness-directed importance weighted averaging operators. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(Suppl.):67–82, 2003.

[4] J.J. Dujmović. Weighted conjunctive and disjunctive means and their application in systems evaluation. *Journal of the University of Belgrade, EE Dept., Series Mathematics and Physics*, (483):147–158, 1974.

[5] R. R. Yager. Multicriteria decision making using fuzzy quantifiers. In *IEEE Conference on Computational Intelligence for Financial Engineering (CIFEr)*, pages 42–47, New York, 1995. Introduces Yager's multiplicative importance weighting of OWA operators.

[6] R.R. Yager. A hierarchical document retrieval language. *Information Retrieval*, 3:357–377, 2000.

[7] D. Filev and R.R. Yager. Analytic properties of Maximum Entropy OWA operators. *Information Sciences*, 85:11–27, 1995.

[8] M. Carbonel, M. Mas, and G. Mayor. On a class of monotonic extended OWA operators. In *Proceeding of FUZZ-IEEE'1997*, pages 1695–1700, Barcelona, 1997.

[9] M. Nagumo. Über eine klasse der mittelwerte. *Japanese Journal of Mathematics*, 7:71–79, 1930.

[10] A.N. Kolmogoroff. Sur la notion de la moyenne. *Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Natur. Sez.*, 12(6):388–391, 1930.

[11] H.L. Larsen and R.R. Yager. The use of fuzzy relational thesauri for classificatory problem solving in information retrieval and expert systems. *IEEE Trans. Syst. , Man, and Cybern.*, SMC-23(1):31–41, 1993.

[12] H.L. Larsen and R.R. Yager. A framework for fuzzy recognition technology. *IEEE Trans. Syst. , Man, and Cybern.*, SMC-30(1):65–76, 2000.

[13] H.L. Larsen. Importance weighted OWA aggregation of multicriteria queries. In *Proceeding of the North American Fuzzy Information Processing Society conference (NAFIPS'99)*, pages 740–744, New York, 10-12 June 1999, 1999.

ISBN: 978-989-95079-6-8

Training a Personal Alert System for Research Information Recommendation

Germán Hurtado Martín [1,2] Chris Cornelis[2] Helga Naessens[1]

1.Dept. of Applied Engineering Sciences, University College Ghent, Ghent, Belgium
2.Dept. of Applied Mathematics and Computer Science, Ghent University, Ghent, Belgium
Email: german.hurtadomartin@{hogent.be,ugent.be}, chris.cornelis@ugent.be, helga.naessens@hogent.be

Abstract— *Information Systems, and in particular Current Research Information Systems (CRISs), are usually quite difficult to query when looking for specific information, due to the huge amounts of data they contain. To solve this problem, we propose to use a personal search agent that uses fuzzy and rough sets to inform the user about newly available information. Additionally, in order to automate the operation of our solution and to provide it with sufficient information, a document classification module is developed and tested. This module also generates fuzzy relations between research domains that are used by the agent during the mapping process.*

Keywords— Automatic classification, Current Research Information Systems, Fuzzy-rough sets, Personal search agents

1 Introduction

Current Research Information Systems (CRISs) are information systems, operating at regional, national, or international level[1], that store and provide access to data on current research carried out by organizations or researchers. These data typically include information on people, projects, organizations, funding possibilities, facilities, etc. CRISs are usually not easy to query, due to several reasons. On the one hand, the data volume is huge. This makes them a very interesting source of information, but can also be overwhelming for the user when he tries to find some specific information in the system. On the other hand, this information is frequently updated, so users are often not aware of new information that could be relevant to them. As a consequence, they are often unable to express their information needs by means of a conventional query. In the last years, personal search agents seem to be the solution to those problems [1]: they gather potentially useful information for users to subsequently present it to them in the form of recommendations. This idea has been received well and can be found not only in research-related applications like CRISs[2] or article recommenders [2, 3], but also in a significant amount of systems from different domains that share the aforementioned problems [4, 5].

However, and despite the fact that these agents can indeed be helpful for the users, they are subject to limitations. Since what they generally do is just to look for exact matches between the user's interests (given as keywords) and the rest of the information in the system, the user will often miss out on useful information as the documents use different terms to refer to the same, or similar, concepts. Furthermore, most CRISs also face other information defects such as missing, ambiguous, or imprecise information [6].

Concepts from fuzzy [7] and rough [8] set theory, upon which the present approach is based, appear as a solution to the problems mentioned above by allowing for a more flexible matching process. Fuzzy sets allow to express partial relationships, which describe reality in a more faithful way than a binary classification, while the rough component provides mechanisms for query expansion: in this way, a user profile and a document may still be matched when they refer to different, but related, keywords, resulting in a higher recall (a higher number of retrieved relevant documents).

To put these ideas into practice, a Personal Alert System (PAS) is currently under development. This system contains profiles of researchers and their interests, activities, papers, etc., as well as documents with information regarding research projects, conference announcements, or funding possibilities. Both these information sources are mapped to a common ontology, which is currently the three-level IWETO[3] taxonomy of the Flemish government. The main goal of the system is to alert users whenever a document can be matched to their research interests by the search agent. A basic prototype has been implemented, using fuzzy-rough algorithms, and in which the user can also influence term relations through a simple feedback process.

While a conceptual version of PAS was described in [9], this work goes further and describes its concrete implementation, paying special attention to the development and evaluation of an automatic classification module which allows the system to classify new documents according to the IWETO taxonomy, as well as to acquire new information with a view to the matching process.

The remainder of this paper is organized as follows: in Section 2, PAS is presented, giving a brief overview of its architecture, and focusing on how the information is represented in the system. In Section 3, the automatic document classification mechanism is described in detail, while in Section 4 we show how the mapper works. Finally, the paper is concluded in Section 5, where we address some issues for future work as well.

2 General design of the system

As said in the introduction, PAS contains information about researchers, projects, funding possibilities, etc., and it will try

[1]Some examples can be found at http://cris.csrees.usda.gov/ (USA), http://www.ris.is/ (Iceland), or http://sicris.izum.si/ (Slovenia). More information can be found at http://www.eurocris.org/, the professional association of CRIS experts and developer of the standard for these systems.

[2]The possibility of using an agent to retrieve funding possibilities is given at EraCareers: http://ec.europa.eu/euraxess/

[3]IWETO, Inventaris Wetenschappelijk en Technologisch Onderzoek Vlaanderen (now FRIS); http://www.iweto.be

ISBN: 978-989-95079-6-8

to match researchers with research information by using the "intelligence" that fuzzy and rough set theories provide. This section first gives a brief, general view on the system, and then focuses on the representation of the information.

2.1 Architecture of the system

Fig. 1 below shows the different modules of the system and how they interact with each other. In brief, the system is organized as follows. Both users (researchers) and documents (information about projects, etc.) are represented by means of profiles. Those profiles, described in more detail in the next subsection, contain several fields, the field storing keywords being the most important one. These keywords represent the researchers' interests (in the cases of the users), or the fields the documents are related to (in the case of the projects), and correspond to the IWETO taxonomy, where they are distributed in three hierarchical levels: a first level with 5 main nodes, a second level with 29 nodes, and a third level containing 359 nodes (in this last level, a node may contain more than one research field, when the fields are closely related). The left-hand side of Fig. 2 shows a sample of the taxonomy. Once a document comes into the system, it is automatically classified (this process will be shown in detail in Section 3), to subsequently be sent to the mapper. The mapper is the central and, thanks to its fuzzy-rough algorithms, intelligent part of the system. It decides whether or not a given document could be interesting for a given user. After this process, and if the mapper has decided so, a notification (about that given document) is sent to the user. Finally, if he desires so, the notified user can give his feedback about his satisfaction degree with the received alert: the system will then use this information to adjust the relations described in Section 2.2.

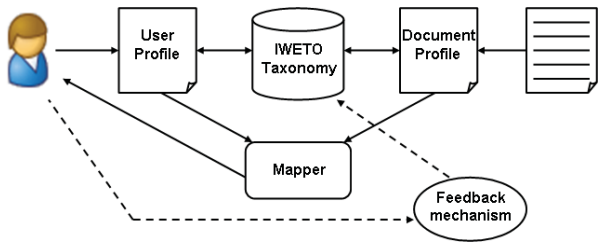


Figure 1: General architecture of PAS

For additional information about the architecture and a more detailed description of the different parts of the system, we refer to [9].

2.2 Representation of the information

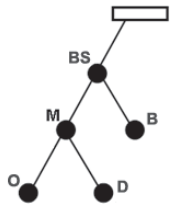
A key issue in PAS is the representation of the information: not only how profiles are represented but also how the IWETO taxonomy is stored.

Profiles are based on IWETO keywords, and their representation is common for both users and documents. Leaving aside the fields inherent to users or documents (i.e. "name" for users, "title" or "description" for documents), both user and document profiles contain a field dedicated to store the keywords that will represent them in the system. The structure used for that is a set where the pairs $<$*interest terms*, *degree of interest*$>$ are stored (*degree of relationship*, in the case of the documents). For example, the list $\{$*(AI,1)*, *(T,0.7)*, *(P,0.2)*$\}$ would correspond to a user very interested in *Artificial Intelligence*, quite interested in *Translation*, and slightly interested in *Physics*. The degree of interest/relation lies always between 0 (no interest/relation) and 1 (strong interest/relation). Every profile can therefore be seen as a fuzzy set in the whole collection of keywords X (since the keywords can come from any of the levels of the taxonomy, its hierarchical structure does not play any role at this point).

In the current implementation, the information contained in the profile keywords has different origins. While documents are automatically classified by using the techniques explained in Section 3, users must select their interests from a list of keywords. Users also select the different interest degrees from a list where these are linguistically represented (to subsequently map them onto a numerical value automatically).

Apart from user-keyword and document-keyword relations, we also consider relationships between keywords. A square matrix is used to represent the different relations between keywords, expressed as a degree ranging between 0 (no relation) and 1 (strong relationship). The right-hand side of Fig. 2 gives an example of this representation. A way of assigning degrees to term pairs is based on their co-occurrence during the training process; the more two terms co-occur in training documents, the higher their degree of relationship (see (10) in Section 4 for more detail).

These degrees make it possible to define a fuzzy relation R reflecting how closely two keywords k_1 and k_2 are related. It is important to remark that R is not a fuzzy tolerance relation, since it is not symmetric. It is reflexive, because a term is perfectly related to itself, so $R(k_1, k_1) = 1$, but $R(k_1, k_2)$ and $R(k_2, k_1)$ are not always equal. This is necessary to depict reality in a more faithful way. For instance, and referring to the example shown in Fig. 2, it is logical to think that there is a stronger relation from *Ophtalmology* towards *Medicine* than vice versa: a document related to *Ophtalmology* will always be related to *Medicine*, but a document related to *Medicine* will not be necessarily relevant to *Ophtalmology*.



	BS	M	B	O	D
BS	1	0.4	0.2	0.1	0.1
M	1	1	0.3	0.3	0.2
B	1	0.3	1	0.1	0
O	1	1	0.1	1	0.1
D	1	1	0	0.1	1

O: Ophtalmology *M: Medicine* *BS: Biomedical Sciences*
D: Dermatology *B: Biophysics*

Figure 2: Example of how the classification is stored

The information about the links between keywords (research fields) contained in the keyword relationship matrix can be used by the system, along with the interest research fields in the profiles, to compute a fuzzy-rough upper approximation for every profile. This process will be explained in more detail in Section 4.

3 Automatic document classification

Clearly, information is the core of any information system, CRISs included. Therefore, it is important for the system to

ISBN: 978-989-95079-6-8

be able to acquire new information easily. Since PAS aims to be as human-independent as possible, it is equipped with an automatic document classification module for this purpose. A representation of the process that documents undergo in this module is shown in Fig. 3 below. To build the module, some of the ideas in [10], such as the usage of keyword vectors for documents and classification, have been used.

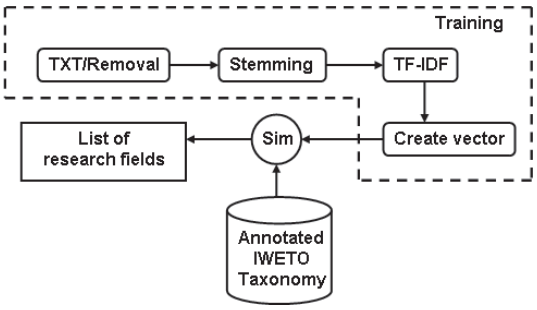


Figure 3: Automatic classification process

3.1 Training

First of all, the system needs to be trained, which is done by means of already classified documents. Currently, these documents come from the FRIS database[4]. Basically, each of them consists of an abstract about a given project, the keywords of that project, and the research fields under which the project was classified in the FRIS database. The documents must first be converted to a format readable by the system, namely txt. Afterwards, every document d is processed by a stemming algorithm. To this purpose, PAS currently uses a purpose-built stemmer, although the list of stopwords that this stemmer uses is an extended version of that which can be found at the site of the implementation of the Porter stemmer [11] in the Snowball language for stemming algorithms[5]. After that, the TF-IDF algorithm [12] is applied to calculate a weight w_i for each term in the document. Those terms with a weight w_i greater than a given threshold γ ($\gamma \geq 0$) are put into a vector $\vec{d}$ that represents the document. This vector $\vec{d}$ is further normalized so that it has unit length.

An important aspect at this point are the terms for which a TF-IDF value is computed. The first implementations of the classifier worked with two-word terms (bigrams), since such terms offer more information and are less ambiguous; for example "information system" is more meaningful than "information" and "system". However, they are also more difficult to match between documents, and some combinations are not so fortunate. So since early tests with one-word terms (unigrams) showed a sensible improvement, the classifier currently works with unigrams. Nevertheless, for some terms composed of more than 1 word, no splitting is applied. This happens for terms extracted via pattern recognition, for example those preceded by the string *"Keywords:"* in a document. In this cases, the terms are added, untouched, to the vector $\vec{d}$, with weight 1.0. Additionally, if the term is not a unigram, another copy of it is added to the rest of the text in the document to be split and processed as described above.

When the training documents have been reduced to their term vector form, the system is ready to start its learning process. It is here where the hierarchy of the IWETO thesaurus is used. First, and based on the information regarding the nodes (research fields) under which the documents have been classified, every vector $\vec{d}$ is assigned to its corresponding node or nodes n. A vector $\vec{n}$ for every node n is then computed. These vectors $\vec{n}$ are normalized vectors containing the terms that are most relevant to node n, or in other words, the terms (and respective weights) contained in the term vectors $\vec{d}$ assigned to n. More formally, let $S(n)$ be the set of subconcepts under concept n (children of node n). Also, let $\{d_1^n, d_2^n, \cdots, d_{k_n}^n\}$ be the individual training documents classified under concept n. $Docs(n)$, the set of all the documents classified under concept n augmented with the documents classified under all its children is defined as:

$$Docs(n) = \left[\bigcup_{n' \in S(n)} Docs(n') \right] \cup \{d_1^n, d_2^n, \cdots, d_{k_n}^n\} \quad (1)$$

The term vector $\vec{n}$ is then computed as:

$$\vec{n} = \left[\sum_{d \in Docs(n)} \vec{d} \right] / |Docs(n)| \quad (2)$$

This term vector $\vec{n}$ is finally normalized into a unit vector.

3.2 Automatic classification

The classification process starts in a similar way to that of the training process, including the application of the stemming and TF-IDF algorithms to create their representative vectors $\vec{d}$. Each test document is then compared to every node in the hierarchical classification. This is done by comparing their representative term vectors $\vec{d}$ and $\vec{n}$ by means of cosine similarity. Those nodes n for which the value of the comparison $sim(\vec{d}, \vec{n})$ is greater than zero are added to a priority queue along with their sim values, ordered with respect to these values. The cosine similarity measure sim for normalized vectors is defined as their inner product, i.e.,

$$sim(\vec{d}, \vec{n}) = \frac{\vec{d} \cdot \vec{n}}{|\vec{d}||\vec{n}|} = \vec{d} \cdot \vec{n} \quad (3)$$

Since the values $sim(\vec{d}, \vec{n})$ reflect how related the document d is to the class represented by node n, they are normalized so that d and n_1, the class with which d has the strongest relation, are related in a degree 1. Finally, the nodes in the N first places of the priority queue (i.e. the nodes with the N greatest values for sim, $N \geq 0$) are retrieved, along with their computed and normalized sim values, as long as they are greater or equal than a given threshold δ, $\delta \geq 0$. Thus, δ controls that only those classes to which d is sufficiently strongly related are taken into account.

The document is then classified under those research fields. This means that a profile is created for the document, with its representative keywords the research fields in which it has been classified, and with the respective sim values giving an idea of how related the document is to them.

[4]FRIS, Flanders Research Information Space; http://www.ewi-vlaanderen.be/fris/

[5]M. Porter, *Snowball: a language for stemming algorithms*; http://snowball.tartarus.org/index.php

ISBN: 978-989-95079-6-8

3.3 Classification results

The last part of this section is dedicated to show some results obtained with the proposed classifier. At the same time, some design decisions are discussed.

First of all, an evaluation measure is needed. Evaluating hierarchical classification is no trivial problem, a problem that gets yet more complicated by the fact of the classification being multi-label, i.e., documents can be labelled with different keywords. In this work we will use an adaptation to fuzzy sets of the measure hF_β proposed by Kiritchenko et al. in [13]. Specifically:

$$hF_\beta = \frac{(\beta^2+1)\cdot hP \cdot hR}{(\beta^2 \cdot hP + hR)}, \beta \in [0, +\infty) \quad (4)$$

where hP and hR are hierarchical precision and hierarchical recall, respectively. These measures are a variation of the traditional precision and recall evaluation measures which also take into account the hierarchical structure of the classes. In particular, rather than comparing the classes a document belongs to and the ones that were predicted for it, we also consider the ancestors of these classes. Formally:

$$\hat{C}_i = \{\cup_{c_k \in C_i} Anc(c_k)\} \qquad \hat{C}'_i = \{\cup_{c_l \in C'_i} Anc(c_l)\} \quad (5)$$

where C_i and C'_i are the actual and the predicted class sets, respectively, and $Anc(n)$ is the set of ancestors of n, n included. Therefore, $\hat{C}_i$ and $\hat{C}'_i$ are the actual and predicted class sets extended by adding them the ancestors of the classes they contain. Note that $\hat{C}'_i$ is treated as a fuzzy set, with its membership values equal to the predicted sim values, and that $\hat{C}_i$ is actually a crisp set, since a document d is related to all the classes under which it is actually classified in a degree of 1. Measures hP and hR are then defined as:

$$hP = \frac{\sum_i |\hat{C}_i \cap \hat{C}'_i|}{\sum_i |\hat{C}'_i|} \qquad hR = \frac{\sum_i |\hat{C}_i \cap \hat{C}'_i|}{\sum_i |\hat{C}_i|} \quad (6)$$

Since we are dealing with fuzzy sets, corresponding operations are used, defining the intersection of $\hat{C}_i$ and $\hat{C}'_i$ by:

$$(\hat{C}_i \cap \hat{C}'_i)(x) = \min(\hat{C}_i(x), \hat{C}'_i(x)) \quad (7)$$

for a keyword x, and defining the cardinality of $\hat{C}'_i$ as:

$$|\hat{C}'_i| = \sum_{x \in X} \hat{C}'_i(x) \quad (8)$$

By this definition, wrongly predicted classes are penalized less strongly when they belong to the same branch as one of the actual classes. The value of β can be chosen; in our case we will use $\beta = 1$, so both hierarchical precision and hierarchical recall have the same weight in hF.

All the results shown throughout the rest of the section have been calculated with this evaluation measure. On the other hand, the tests to obtain those results were carried out with a dataset formed by 9438 IWETO project descriptions, on which 10-fold cross validation was performed.

The results are subject to a number of parameter settings. The parameters with more impact in the results are the ones related with the term vector $\vec{d}$ used to represent a given document d, and the way the system chooses the classes under which the document d will be classified. Those problems are related with the parameter γ mentioned in Section 3.1, and the parameters N and δ mentioned in Section 3.2, respectively. All these parameters are discussed here and some results about their tuning are presented in Table 1.

Parameter γ is necessary to restrict the number of terms in every term vector $\vec{d}$, mainly because of memory usage and system performance reasons, since terms with very low TF-IDF weights would be probably irrelevant anyway if they were included in the term vector. Due to the short length of the documents we used, parameter γ is not very essential for our purposes. Therefore, the value used for our tests was $\gamma = 0$. Nevertheless, as mentioned above, its use is recommended when dealing with longer documents (or documents with an unknown length).

More relevant in our tests was the number of classes in which the system classifies the document: too many or too few classes probably signify a big difference with respect to the actual classification (resulting in too low precision and/or recall), and will also have an important impact on the user's satisfaction (if a document is classified in too many categories, it will be considered for the matching process even when it is not strongly related to a given category; if it is classified in just one category, it probably will not reach its whole target group of users). Parameters N and δ control this.

As said in Section 3.2, N sets the number of research fields in which a document will be classified, i.e., the number of classes that will be retrieved from the first places of the priority queue. On the other hand, a threshold δ is also necessary, in order to avoid cases of weak relationships: N is the number of classes with the highest value for the similarity measure sim, but that does not necessarily mean that all N values are high. Threshold δ tries to solve this. Since it is applied to normalized results, δ must be a value between 0 and 1, and since strong relationships are desired, the value cannot be too low. The results of the different tests are shown in Table 1.

Table 1: Results for tests with different values for N and δ.

δ	N	6	5	4	3	2	1
0.3	hF_1	0.502	0.512	0.523	**0.53**	0.527	0.478
	hP	0.419	0.442	0.472	0.508	0.555	0.623
	hR	0.625	0.609	0.587	0.555	0.501	0.388
0.4	hF_1	0.505	0.514	0.524	**0.53**	0.526	0.478
	hP	0.426	0.447	0.475	0.509	0.556	0.623
	hR	0.62	0.605	0.584	0.553	0.5	0.388
0.5	hF_1	0.509	0.517	0.525	**0.53**	0.526	0.478
	hP	0.439	0.457	0.481	0.513	0.557	0.623
	hR	0.607	0.595	0.577	0.548	0.497	0.388
0.6	hF_1	0.514	0.519	0.525	0.528	0.524	0.478
	hP	0.458	0.473	0.493	0.52	0.56	0.623
	hR	0.584	0.575	0.562	0.538	0.493	0.388
0.7	hF_1	0.516	0.519	0.522	0.524	0.52	0.478
	hP	0.487	0.496	0.511	0.531	0.564	0.623
	hR	0.549	0.546	0.535	0.518	0.483	0.388
0.8	hF_1	0.514	0.514	0.515	0.516	0.514	0.478
	hP	0.528	0.531	0.537	0.55	0.574	0.623
	hR	0.5	0.499	0.495	0.486	0.465	0.388
0.9	hF_1	0.503	0.503	0.503	0.503	0.501	0.478
	hP	0.577	0.577	0.578	0.581	0.591	0.623
	hR	0.445	0.445	0.444	0.443	0.435	0.388

The average number of classes to which a given document from the dataset belongs is 3. Due to this fact, $N = 3$ offers the best values for the evaluation measure hF_1. It can also be

ISBN: 978-989-95079-6-8

seen that $N = 3$ offers the best equilibrium between hP and hR. Moreover, note that the precision hP is inversely proportional to N: the fewer classes d is classified in, the higher the precision. This is logical since a low value for N means that the document is classified in those classes which offered the highest values for the similarity measure sim, classes that are then probably among those under which the document was actually classified. But that small number of classes has a drawback: the recall is of course lower. In the same way, a higher value of N results in higher values of hR: the document is classified under a lot of concepts and therefore it is more probable that they cover all the classes actually linked with the document. Of course, this higher number of classes makes the classifier less precise.

Also δ is important, mainly for higher values of N: these high values imply taking more classes into account when classifying a document, but δ imposes a threshold that leaves out all those "highest values" that are not actually high enough to be considered for the classification. In that way, δ allows higher values of N to be used in order to avoid leaving out potentially adequate classes, but without the risk of considering classes that probably are not so related to the document.

Note that the combination of $N = 1$ and δ is useless, as reflected by the results: since the results of sim are normalized, the highest value in the priority queue (and therefore the only one that will be retrieved when $N = 1$) will be equal to 1.

Although a high precision is desirable, it is necessary to keep in mind that the system aims to alert the researchers about potentially interesting documents. Therefore, the documents are preferably classified under more than one research field (as long as those fields are relatively strongly related to the document). That means that a balance between precision and recall ($\beta = 1$) is recommended in this case. As the combination $\delta = 0.5$, $N = 3$ offers the best results ($\delta = 0.4$ and $\delta = 0.3$ obtain similar values for hF but a lower precision), they are the values currently used by the system.

There are nonetheless some remarks to be made. In this case, the hierarchical classification problem gets complicated by the fact that the documents in the dataset have been manually classified by different people. This means that in some cases the document has been indexed under all the possible research fields while in other cases it has been vaguely classified under a parent concept. As previously explained, this can cause big differences between the set containing the research fields under which the documents were actually classified and that predicted by the system.

4 Mapper

Since the mapper is the most important part of the system, this section is used to explain in detail how it works. The mapper determines whether a document is interesting enough to notify a user about it. In this process, as discussed in the introduction, the system should be able to identify interesting documents even when their keywords do not exactly match those in the user's profile, but are semantically related to them. To achieve this added intelligence, some ideas from fuzzy-rough set theory are used.

Specifically, fuzzy-rough query expansion is applied. In particular, the system currently uses an adaptation of the approach described in [14]: to assess how well a document profile D matches a user profile U (both represented as fuzzy sets in the set X of keywords) the algorithm first uses a fuzzy relation R to generate their respective upper approximations $R{\uparrow}D$ and $R{\uparrow}U$, where the upper approximation of a fuzzy set A in X under a fuzzy relation R is defined as:

$$(R{\uparrow}A)(y) = \sup_{x \in X} \min(R(x, y), A(x)), \forall y \in X \qquad (9)$$

A keyword belongs to this upper approximation to the extent that it is related, by means of the fuzzy relation R, to at least one of the keywords in A. In other words, (9) defines the set of objects possibly belonging to A to a certain degree. In the current system, the relatedness of two terms is based on their co-occurrence during the training process, as explained in Section 2.2. To be precise, the relation R between two terms x and y can be defined as:

$$R(x, y) = \frac{|Docs(x) \cap Docs(y)|}{|Docs(x)|} \qquad (10)$$

The similarity between D and U is then computed, using α-cuts[6], as

$$Sim_\alpha(U, D) = 1 - \frac{|B_{u\alpha}|}{|(R{\uparrow}D)_\alpha|} \qquad (11)$$

where

$$B_u = R{\uparrow}D \ominus [R{\uparrow}U \cap R{\uparrow}D] \qquad (12)$$

with $\ominus$ the difference of A and B, defined by $(A \ominus B)(x) = \max(0, A(x) - B(x))$.

If $|(R{\uparrow}U)_\alpha| = 0$, the similarity is defined to be 0. Currently, the system uses a fixed value $\alpha = 0.5$, which turned out to yield the best performance experimentally in preliminary tests. It is important to point out that Sim is asymmetric and that therefore the order of U and D is relevant. This keeps the focus of the comparison on the document (B_u is the set of terms in the upper approximation of D that are not shared with the upper approximation of U) and allows the document to reach more users, as long as it is of relevance for them.

The obtained similarity between user and document is then compared to a user-set notification threshold (entered linguistically and then mapped onto a numerical value): if the similarity is greater than or equal to the threshold, user U will be notified by the system about document D. In this way, the user can tune the sensitivity of the mapper: if he only wants to receive alerts about documents that are definitely relevant, he can impose a higher threshold.

5 Conclusion and future work

We have shown how fuzzy relations are used to represent gradual relationships between research fields, and how these relations can be generated from a given dataset. It has also been shown how to construct user and document profiles as fuzzy sets, and how these sets can be matched by applying ideas from fuzzy-rough set theory. In addition, we have also proposed and explained a method to automatically classify documents, in order to keep the system easily updated. However, and despite the promising performance offered by the current

[6] The α-cut of A is defined as $A_\alpha = \{x \in X | A(x) \geq \alpha\}$, where $\alpha \in [0, 1]$

ISBN: 978-989-95079-6-8

implementation of the system, there is still a lot of work to be done.

Since the main goal of the project is to match researchers with documents that can be potentially useful for them, the main effort will be put into the mapping algorithms. New algorithms will be developed and tested to replace the one currently used. This current algorithm, though very interesting, is quite basic, so we think that there is a lot of room for improvement. For example, different similarity measures to compare the upper approximation can be investigated, and we plan to set up an experiment with a group of real test users.

The performance of the fuzzy-rough algorithms is of course also linked to the availability of sufficient document profiles and the quality of the classifier. However, the current version of this module of the system is quite simple, and further improvements could be very interesting and fruitful. For instance, more complex techniques from language technology could be used to refine the keyword extraction. Also, the fuzzy weights of the relationships between concepts can play a more important role, in the classification process as well as in the evaluation of results. The possibility of semi-automatically enriching the IWETO thesaurus with a fourth level, for the sake of precision, will be studied too. Besides, an alternative representation of the information, with the profiles as copies of the classification (which would allow the user to assign his own weights to the relations), will be considered.

Other modules will be added as well. In order to achieve a greater autonomy of the system, the feedback mechanism will be automated, and different techniques will be used to monitor the user's behavior while using the system, so that it can automatically update his list of interests and preferences, adding information that can be useful for the mapper. Also, the classifier could be used to extract additional information from the researcher's publications, when available. All this information can also be used, along with that provided by the feedback mechanism, to refine and adjust the relationships between concepts from the thesaurus. The user will also be free to explore the system on his own: a browsing utility and a search engine will be added with that finality. This engine can also use the same query expansion techniques that are used for personalized notification.

Acknowledgment

The authors thank the Flanders Research Information Space program for its cooperation. Also, Chris Cornelis thanks the Research Foundation-Flanders for supporting his research.

References

[1] P. Maes. Agents that reduce work and information overload. *Communications of the ACM*, 37(7):31–40, 1994.

[2] T. Bogers and A. van den Bosch. Recommending scientific articles using citeulike. In *Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys'08)*, pages 287–290, 2008.

[3] S.M. McNee, I. Albert, D. Cosley, P. Gopalkrishnan, S.K. Lam, A.M. Rashid, J.A. Konstan, and J. Riedl. On the recommending of citations for research papers. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work (CSCW'02)*, pages 116–125, 2002.

[4] M. Montaner, B. López, and J.L. de la Rosa. A taxonomy of recommender agents on the internet. *Artificial Intelligence Review*, 19(4):285–330, 2003.

[5] G. Adomavicius and A. Tuzhilin. Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[6] A. Motro and P. Smets. *Uncertainty management in information systems: from needs to solutions*. Springer, New York, NY, 1997.

[7] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[8] Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, 11(5):341–356, 1982.

[9] G. Hurtado Martín, C. Cornelis, and H. Naessens. Personalizing information retrieval in CRISs with fuzzy sets and rough sets. In *Proceedings of the 9th International Conference on Current Research Information Systems (CRIS2008)*, pages 51–59, 2008.

[10] A. Sieg, B. Mobasher, and R. Burke. Learning ontology-based user profiles: a semantic approach to personalized web search. *IEEE Intelligent Informatics Bulletin*, 8(1), 2007.

[11] M. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.

[12] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, NY, 1983.

[13] S. Kiritchenko, S. Matwin, R. Nock, and A. Fazel Famili. Learning and evaluation in the presence of class hierarchies: application to text categorization. In *Canadian Conference on AI*, pages 395–406, 2006.

[14] P. Srinivasan, M.E. Ruiz, D.H. Kraft, and J. Chen. Vocabulary mining for information retrieval: rough sets and fuzzy sets. *Information Processing and Management*, 37(1):15–38, 2001.

ISBN: 978-989-95079-6-8

On Reduced Semantics for Fuzzy Predicate Logics

Pilar Dellunde[1,2]

1.Universitat Autònoma de Barcelona
2.Artificial Intelligence Research Institute IIIA - CSIC
08193 Bellaterra, Catalonia
Email: pilar.dellunde@uab.cat

Abstract— Our work is a contribution to the model-theoretic study of equality-free fuzzy predicate logics. We present a *reduced semantics* and we prove a completeness theorem of the logics with respect to this semantics. The main concepts being studied are the *Leibniz congruence* and the *relative relation*. On the one hand, the Leibniz congruence of a model identifies the elements that are indistinguishable using equality-free atomic formulas and parameters from the model, a reduced structure is the quotient of a model modulo this congruence. On the other hand, the relative relation between two structures plays the same role that the isomorphism relation plays in classical predicate languages with equality.

Keywords— Equality-free Language, Fuzzy Predicate Logic, Model Theory, Reduced Structure, Relative Relation.

1 Introduction

This work is a contribution to the model-theoretic study of equality-free fuzzy predicate logics. Model theory is the branch of mathematical logic that studies the construction and classification of structures. Construction means building structures or families of structures, which have some feature that interest us. In our case we devote our investigation to the class of reduced structures, that help us to shed light to the characteristic role played by equality in predicate fuzzy logics.

Classifying a class of structures means grouping the structures into subclasses in a useful way, and then proving that every structure in the collection does belong in just one of the subclasses. The most basic classification in classical model theory is given by the relations of elementary equivalence and isomorphism. Our purpose in the present article is to investigate and characterize the structure-preserving maps between structures in a fuzzy setting. In classical predicate logics with equality, homomorphisms are structure-preserving, but if they are not isomorphisms, they don't necessarily preserve all the formulas of the language. On the contrary, in equality-free fuzzy predicate logics, σ-homomorphisms preserve all the formulas, but unlike isomorphisms, the relation between structures of one being the σ-homomorphic image of another is not an equivalence relation in the class of structures.

The main concepts studied in this work are the *Leibniz congruence* and the *relative relation*. The notion of Leibniz congruence arises in a very natural way. It is said that two elements of a model are related by this congruence when they satisfy exactly the same equality-free atomic formulas with parameters in the model. This congruence always exists, and it turns out to be the greatest congruence of the model. This idea has its origin in the Principle of the identity of the indiscernibles of G. W. Leibniz.

Given a model $(\mathbf{M}, \mathbf{B})$ the quotient structure modulo the Leibniz congruence $\Omega(\mathbf{M}, \mathbf{B})$ is denoted by $(\mathbf{M}, \mathbf{B})^r$ and is called its *reduction*. When we make the quotient modulo the Leibniz congruence, we identify the elements that are indistinguishable using equality-free atomic formulas and parameters in the model, thus the Leibniz congruence of the reduction of a model is always the identity. An structure with the property that its Leibniz congruence is the identity is said to be a *reduced* structure. The importance of reduced structures in equality-free logic comes from the fact that the reduction of a model is a σ- homomorphic image of the model and therefore, the model and its reduction satisfy exactly the same equality-free sentences.

The other main concept analysed is the relative relation. It is said that two structures are *relatives* when they have isomorphic reductions. Along this work we will give different characterizations of this relation. Our aim is to point out that it plays the same role in equality-free logic that the isomorphism relation plays in logic with equality. The actual interest of the Leibniz congruence and the relative relation comes from the work of W. Blok and D. Pigozzi. They introduced the concept of relative relation for the special case of logical matrices in [1], and in [2] they made an extensive use of what they named the Leibniz congruence.

Different definitions have been introduced so far for basic model-theoretic operations on structures. For instance, the notion of *elementary submodel, morphism* and *congruence of a fuzzy model* of [3], *elementary embeddings and submodels* of [4], *fuzzy submodel, elementary fuzzy submodel* and *isomorphism* of structures of first-order fuzzy logic with graded syntax of [5], *complete morphism and congruence* in languages with a similarity predicate of [6] and the notion of σ-embedding of [7]. Being our starting point all these works, in the Preliminaries section we introduce the notions of homomorphism and congruence of a model, trying both to encompass the most commonly used definitions in the literature and to extend the corresponding notions of classical predicate logics.

In section 3 we introduce the notion of reduced structure and some basic model-theoretic properties of this kind of structures. In section 4 we characterize when two structures are relative and we prove that the relative relation is the transitivization of the relation of being a σ-homomorphic image. Finally, section 5 is devoted to future work.

ISBN: 978-989-95079-6-8

2 Preliminaries

Our study of the model theory of fuzzy predicate logics is focused on the basic fuzzy predicate logic MTL∀ and stronger t-norm based predicate calculi, the so-called *core fuzzy logics*. We start by introducing the notion of core fuzzy logic in the propositional case.

Definition 1 *A propositional logic L is a* core fuzzy logic *iff L satisfies:*

1. *For all formulas ϕ, φ, α, $\varphi \equiv \phi \vdash \alpha(\varphi) \equiv \alpha(\phi)$.*
2. (LDT) Local Deduction Theorem: *for each theory and formulas ϕ, φ:*

   $$T, \varphi \vdash \phi \text{ iff for some natural number } n,\ T \vdash \varphi^n \to \phi.$$

3. *L expands* MTL.

For a thorough treatment of core fuzzy logics we refer to [4], [8] and [7]. A *predicate language* Γ is a triple (P,F,A) where P is a non-empty set of predicate symbols, F is a set of function symbols and A is a function assigning to each predicate and function symbol a natural number called the *arity of the symbol*. Functions f for which $\mathrm{A}(f) = 0$ are called *object constants*. Formulas of the predicate language Γ are built up from the symbols in (P,F,A) together with logical symbols $(\forall, \exists, \&, \to, \overline{0}, \overline{1})$, variables and punctuation. Throughout the paper we consider the equality symbol as a binary predicate symbol not as a logical symbol, we work in equality-free fuzzy predicate logics. That is, the equality symbol is not necessarily present in all the languages and its interpretation is not fixed.

Let L be a fixed propositional core fuzzy logic and $\mathbf{B}$ an L-algebra, we introduce now the semantics for the fuzzy predicate logic $L\forall$. A $\mathbf{B}$-structure for predicate language Γ is a tuple $\mathbf{M} = (M, (P_{\mathbf{M}})_{P\in\Gamma}, (F_{\mathbf{M}})_{F\in\Gamma}, (c_{\mathbf{M}})_{c\in\Gamma})$ where M is a non-empty set and

1. For each n-ary predicate $P \in \Gamma$, $P_{\mathbf{M}}$ is a $\mathbf{B}$-fuzzy relation $P_{\mathbf{M}} : M^n \to \mathbf{B}$.
2. For each n-ary function symbol $F \in \Gamma$, $F_{\mathbf{M}} : M^n \to M$.
3. For each constant symbol $c \in \Gamma$, $c_{\mathbf{M}} \in M$.

Let $\mathbf{M}$ be a $\mathbf{B}$-structure, an $\mathbf{M}$-*evaluation* of the variables is a mapping v which assigns to each variable an element from M. By $\phi(x_1, \ldots, x_k)$ we mean that all the free variables of ϕ are among $x_1, \ldots, x_k$. If v is an evaluation such that for each $0 < i \leq n$, $v(x_i) = d_i$, and λ is either a Γ-term or a Γ-formula, we abbreviate by $\|\lambda(d_1, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}$ the expression $\|\lambda(x_1, \ldots, x_n)\|^{\mathbf{B}}_{\mathbf{M},v}$. Let ϕ be a Γ-sentence, given a $\mathbf{B}$-*structure* $\mathbf{M}$, it is said that $\mathbf{M}$ is a *model* of ϕ iff $\|\phi\|^{\mathbf{B}}_{\mathbf{M}} = 1$.

From now on, we say that $(\mathbf{M}, \mathbf{B})$ is a Γ-structure instead of saying that $\mathbf{M}$ is a $\mathbf{B}$-structure in the language Γ. We say that a structure is *safe*, if a truth value is defined for each formula and evaluation. We assume that all our structures are safe. It is denoted by $(\mathbf{M}, \mathbf{B}) \equiv (\mathbf{N}, \mathbf{A})$ when these two structures are elementarily equivalent. In this section we have presented only a few definitions and notation, a detailed introduction to the syntax and semantics of fuzzy predicate logics can be found in [9].

Definition 2 *Let $(\mathbf{M}_1, \mathbf{B}_1)$ be a Γ_1-structure and $(\mathbf{M}_2, \mathbf{B}_2)$ be a Γ_2-structure with $\Gamma_1 \subseteq \Gamma_2$. We say that the pair (f, g) is a* homomorphism *of $(\mathbf{M}_1, \mathbf{B}_1)$ into $(\mathbf{M}_2, \mathbf{B}_2)$ iff*

1. *$g : \mathbf{B}_1 \to \mathbf{B}_2$ is a L-algebra homomorphism of $\mathbf{B}_1$ into $\mathbf{B}_2$.*
2. *$f : M_1 \to M_2$ is a mapping of M_1 into M_2.*
3. *For each constant symbol $c \in \Gamma_1$, $f(c_{\mathbf{M}_1}) = c_{\mathbf{M}_2}$.*
4. *For each n-ary function symbol $F \in \Gamma_1$ and elements $d_1, \ldots, d_n \in M_1$,*

   $$f(F_{\mathbf{M}_1}(d_1, \ldots, d_n)) = F_{\mathbf{M}_2}(f(d_1), \ldots, f(d_n))$$

5. *For each n-ary predicate $P \in \Gamma_1$ and elements $d_1, \ldots, d_n \in M_1$,*

   $$g(P_{\mathbf{M}_1}(d_1, \ldots, d_n)) = P_{\mathbf{M}_2}(f(d_1), \ldots, f(d_n))$$

We say that (f, g) is a σ-homomorphism *if g preserves the existing infima and suprema (that is, if I is a non-empty set and $\sup_{i\in I} a_i$ and $\sup_{i\in I} g(a_i)$ exist, then $g(\sup_{i\in I} a_i) = \sup_{i\in I} g(a_i)$ and analogously for the infima).*

It is denoted by $(\mathbf{M}, \mathbf{B}) \cong (\mathbf{N}, \mathbf{A})$ when these two structures are isomorphic (that is, there is a homomorphism (f, g) from $(\mathbf{M}, \mathbf{B})$ into $(\mathbf{N}, \mathbf{A})$ with f and g onto and one-to-one). It is easy to check, by induction on the complexity of the formulas, that homomorphisms preserve quantifier-free formulas. Note that, by definition, homomorphisms are not always σ-complete, as are in [3] or [6], and unlike [6] homomorphisms are crisp on the algebraic reduct of the first-order structure. If (f, g) is a σ-homomorphism of $(\mathbf{M}_1, \mathbf{B}_1)$ into $(\mathbf{M}_2, \mathbf{B}_2)$ such that f is onto, then for each formula $\phi(x_1, \ldots, x_n) \in \Gamma_1$ and elements $d_1, \ldots, d_n \in M_1$,

$$g(\|\phi(d_1, \ldots, d_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1}) = \|\phi(f(d_1), \ldots, f(d_n))\|^{\mathbf{B}_2}_{\mathbf{M}_2} \quad (1)$$

The proof can be found in [3] (Propositions 6.1 and 6.2). We will refer to homomorphisms satisfying condition (1) as *elementary homomorphisms*.

Definition 3 *A* congruence *on a Γ-structure $(\mathbf{M}, \mathbf{B})$ is a pair (E, θ) where:*

1. *θ is an L-congruence on the algebra $\mathbf{B}$.*
2. *E is an equivalence relation $E \subseteq M \times M$ such that:*

   - *For each n-ary function symbol $F \in \Gamma$ and elements $d_1, \ldots, d_n, e_1, \ldots, e_n \in M$, if for each $0 < i \leq n$, $(d_i, e_i) \in E$, then*

     $$(F_{\mathbf{M}}(d_1, \ldots, d_n), F_{\mathbf{M}}(e_1, \ldots, e_n)) \in E$$

   - *For each n-ary predicate $P \in \Gamma$ and elements $d_1, \ldots, d_n, e_1, \ldots, e_n \in M$, if for each $0 < i \leq n$, $(d_i, e_i) \in E$, then*

     $$(P_{\mathbf{M}}(d_1, \ldots, d_n), P_{\mathbf{M}}(e_1, \ldots, e_n)) \in \theta$$

Now, given a congruence (E, θ) on $(\mathbf{M}, \mathbf{B})$ we define the *quotient structure* $(\mathbf{M}/E, \mathbf{B}/\theta)$ by:

ISBN: 978-989-95079-6-8

- For each constant symbol $c \in \Gamma$, $c_{(\mathbf{M}/E,\mathbf{B}/\theta)} = [c_{(\mathbf{M},\mathbf{B})}]_E$.
- For each n-ary function symbol $F \in \Gamma$ and elements $d_1, \ldots, d_n \in M$,

$$F_{\mathbf{M}/E}([d_1]_E, \ldots, [d_n]_E) = [F_{\mathbf{M}}(d_1, \ldots, d_n)]_E$$

- For each n-ary predicate $P \in \Gamma$ and elements $d_1, \ldots, d_n \in M$,

$$P_{\mathbf{M}/E}([d_1]_E, \ldots, [d_n]_E) = [P_{\mathbf{M}}(d_1, \ldots, d_n)]_\theta$$

where, given an element $d \in M$ and $b \in B$, $[d]_E$ and $[b]_\theta$ denote respectively the equivalence classes of d modulo E and of b modulo θ. We will say that a (E, θ) is an *elementary* congruence (σ-congruence, respectively) if its canonical mapping (f_E, g_θ) is an elementary homomorphism (σ-homomorphism, respectively).

3 Reduced Structures

In this section we introduce the notions of Leibniz congruence and of reduced structure and we establish some basic model-theoretic properties of this kind of models, giving some examples of first-order theories with reduced models. The study of reduced structures and Leibniz congruences for classical predicate logics was done in [10]. In the context of fuzzy predicate logics, X. Caicedo introduced this notion in [11] for the particular case of models of first-order Rational Pavelka's logic in a language with the $\approx$ symbol. At the end of this section we study similarities on reduced structures.

Definition 4 *Let $(\mathbf{M}, \mathbf{B})$ be a Γ-structure and θ an L-congruence on $\mathbf{B}$. We define the relation $\Omega(\mathbf{M}, \mathbf{B}, \theta) \subseteq M \times M$ as follows: for every $d, e \in M$, $(d, e) \in \Omega(\mathbf{M}, \mathbf{B}, \theta)$ iff for every atomic formula, $\phi(y, x_1, \ldots, x_n) \in \Gamma$ and elements $d_1, \ldots, d_n \in M$,*

$$(\|\phi(d, d_1, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}, \|\phi(e, d_1, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$$

Fixed an L-congruence θ on the L-algebra $\mathbf{B}$, next lemma shows that $\Omega(\mathbf{M}, \mathbf{B}, \theta)$ is the greatest E such that (E, θ) is a congruence on the model $(\mathbf{M}, \mathbf{B})$.

Lemma 5 *Let $(\mathbf{M}, \mathbf{B})$ be a Γ-structure and θ an L-congruence on $\mathbf{B}$, then*

1. *$(\Omega(\mathbf{M}, \mathbf{B}, \theta), \theta)$ is a congruence on $(\mathbf{M}, \mathbf{B})$.*
2. *For every (E, θ) congruence on $(\mathbf{M}, \mathbf{B})$, $E \subseteq \Omega(\mathbf{M}, \mathbf{B}, \theta)$.*

Proof: 1. Since θ is an equivalence relation, by definition, $\Omega(\mathbf{M}, \mathbf{B}, \theta)$ is also an equivalence relation. For each n-ary predicate $P \in \Gamma$ and elements $d_1, \ldots, d_n, e_1, \ldots, e_n \in M$, if for each $0 < i \leq n$, $(d_i, e_i) \in \Omega(\mathbf{M}, \mathbf{B}, \theta)$, then by using the definition of $\Omega(\mathbf{M}, \mathbf{B}, \theta)$, for every $0 < i \leq n$, we have the following chain:

$$(\|P(d_1, d_2, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}, \|P(e_1, d_2, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$$

$$(\|P(e_1, d_2, d_3, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}, \|P(e_1, e_2, d_3, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$$

$$\vdots$$

$$(\|P(e_1, e_2, \ldots, e_{n-1}, d_n)\|^{\mathbf{B}}_{\mathbf{M}}, \|P(e_1, \ldots, e_n)\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$$

Assume now that $F \in \Gamma$ is an n-ary function symbol and $d_1, \ldots, d_n, e_1, \ldots, e_n \in M$ such that for each $0 < i \leq n$, $(d_i, e_i) \in \Omega(\mathbf{M}, \mathbf{B}, \theta)$. Let $\overline{k} = k_1, \ldots, k_s \in M$, $\phi(y, x_1, \ldots, x_s) \in \Gamma$ an atomic formula and ϕ' the formula obtained from ϕ by substitution of the variable y for the term $F(z_1, \ldots, z_n)$ (where $z_1, \ldots, z_n$ are new variables not occurring in ϕ). By definition of $\Omega(\mathbf{M}, \mathbf{B}, \theta)$, since ϕ' is also atomic, we can build a chain similar to the one defined in the predicate case and then, we obtain $(\|\phi'(d_1, \ldots, d_n, \overline{k})\|^{\mathbf{B}}_{\mathbf{M}}, \|\phi'(e_1, \ldots, e_n, \overline{k})\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$, consequently,

$$(\|\phi(F_{\mathbf{M}}(d_1, \ldots, d_n), \overline{k})\|^{\mathbf{B}}_{\mathbf{M}}, \|\phi(F_{\mathbf{M}}(e_1, \ldots, e_n), \overline{k})\|^{\mathbf{B}}_{\mathbf{M}}) \in \theta$$

and then $(F_{\mathbf{M}}(d_1, \ldots, d_n), F_{\mathbf{M}}(e_1, \ldots, e_n)) \in \Omega(\mathbf{M}, \mathbf{B}, \theta)$.

2. By definition of $\Omega(\mathbf{M}, \mathbf{B}, \theta)$, because (E, θ) is a congruence. □

Definition 6 *A Γ-structure $(\mathbf{M}, \mathbf{B})$ is* reduced *iff $\Omega(\mathbf{M}, \mathbf{B}, Id_{\mathbf{B}})$ is the identity relation.*

From now on we denote $(\Omega(\mathbf{M}, \mathbf{B}, Id_{\mathbf{B}}), Id_{\mathbf{B}})$ simply by $\Omega(\mathbf{M}, \mathbf{B})$ and we call it the *Leibniz congruence* of $(\mathbf{M}, \mathbf{B})$. Since the identity map clearly preserves infima and suprema, $\Omega(\mathbf{M}, \mathbf{B})$ is always a σ-congruence, therefore for every $(d, e) \in \Omega(\mathbf{M}, \mathbf{B})$, every formula, $\phi(y, x_1, \ldots, x_n) \in \Gamma$ and elements $d_1, \ldots, d_n \in M$,

$$\|\phi(d, d_1, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}} = \|\phi(e, d_1, \ldots, d_n)\|^{\mathbf{B}}_{\mathbf{M}}$$

We will denote by $(\mathbf{M}, \mathbf{B})^r$ the quotient structure modulo the Leibniz congruence $\Omega(\mathbf{M}, \mathbf{B})$ and call it the *reduction* of $(\mathbf{M}, \mathbf{B})$.

Lemma 7 *For every Γ-structure $(\mathbf{M}, \mathbf{B})$ we have:*

1. *$(\mathbf{M}, \mathbf{B}) \equiv (\mathbf{M}, \mathbf{B})^r$.*
2. *$(\mathbf{M}, \mathbf{B})^r$ is a reduced structure.*
3. *$((\mathbf{M}, \mathbf{B})^r)^r \cong (\mathbf{M}, \mathbf{B})^r$.*
4. *If there exists a σ-congruence (E, θ) on $(\mathbf{M}, \mathbf{B})$, then*

$$(\mathbf{M}, \mathbf{B}) \equiv (\mathbf{M}/\Omega(\mathbf{M}, \mathbf{B}, \theta), \mathbf{B}/\theta)$$

Proof: 1. holds because $(\mathbf{M}, \mathbf{B})^r$ is a σ-homomorphic image of $(\mathbf{M}, \mathbf{B})$, 2. and 3. by definition of the Leibniz congruence and of quotient structure. To prove 4. use the fact that for every σ-congruence (E, θ), $(\Omega(\mathbf{M}, \mathbf{B}, \theta), \theta)$ is also a σ-congruence. □

Corollary 8 [Completeness Theorem] *Let $L\forall$ be a fuzzy predicate logic and $\mathbb{K}$ a class of structures such that $L\forall$ is $\mathbb{K}$-complete (strong or finite strong $\mathbb{K}$-complete, respectively), then $L\forall$ is $\mathbb{K}^r$-complete (strong or finite strong $\mathbb{K}^r$-complete, respectively), where $\mathbb{K}^r$ is the class of reductions of the structures in $\mathbb{K}$.*

ISBN: 978-989-95079-6-8

Reduced structures are of common use in computer science and in mathematics. The Rado Graph (infinite random graph) and fuzzy linear orders are examples of reduced structures. However, we can see a more developed example of this process of reduction in the well-known case of similarities. In Section 5 of [9], P. Hájek studies similarities and applies the obtained results to the analysis of fuzzy control in Chapter 7 of of [9]. For a reference about model-theoretic properties of algebras with fuzzy equalities see [8] and [6]. Now (and only for the rest of this section) we assume that our predicate language Γ contains a binary predicate symbol $\approx$. Similarity is understood as fuzzified equality (for a reference see [12] or [13]). Given a core fuzzy logic L, let our axiomatic system for $L\forall$ contain also the following axioms:

1. (Reflexivity) $\forall x\ x \approx x$
2. (Symmetry) $\forall x \forall y (x \approx y \rightarrow y \approx x)$
3. (Transitivity) $\forall x \forall y \forall z ((x \approx y \& y \approx z) \rightarrow x \approx z)$
4. For each n-ary function symbol $F \in \Gamma$, $\forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall y_n ((x_1 \approx y_1 \& \ldots \& x_n \approx y_n) \rightarrow (F(x_1, \ldots, x_n) \approx F(y_1, \ldots, y_n))$
5. For each n-ary predicate $P \in \Gamma$, $\forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall y_n ((x_1 \approx y_1 \& \ldots \& x_n \approx y_n) \rightarrow (P(x_1, \ldots, x_n) \leftrightarrow P(y_1, \ldots, y_n))$

Axioms 1-3 are called *Similarity Axioms* (Sim) and axioms 4-5 *Congruence Axioms* (Cong).

Definition 9 *A Γ-structure $(\mathbf{M}, \mathbf{B})$ has the* equality property (EQP) *if the following condition holds: for every $d, e \in M$, $\|d \approx e\|^{\mathbf{B}}_{\mathbf{M}} = 1$ iff $d = e$.*

Lemma 10 *Given a set Σ of Γ-sentences, $\Sigma \cup Sim \cup Cong$ is satisfiable iff Σ has a model $(\mathbf{M}, \mathbf{B})$ that has* EQP.

Proof: Let $(\mathbf{M}, \mathbf{B})$ be a model of $\Sigma \cup Sim \cup Cong$. If we define $E = \{(a, b) \in M \times M : \|a \approx b\|^{\mathbf{B}}_{\mathbf{M}} = 1\}$, then $(E, Id_{\mathbf{B}})$ is a congruence. Thus, in the quotient structure $(\mathbf{M}/E, \mathbf{B})$, $\|x \approx y\|^{\mathbf{B}}_{\mathbf{M}/E} = 1$ iff $x = y$. □

In [11] X. Caicedo called *reduced structure* to a model of first-order Rational Pavelka's logic with the property EQP. Next lemma shows that Caicedo's notion coincides with ours when we consider axiomatic systems including axioms $Sim \cup Cong$.

Lemma 11 *$(\mathbf{M}, \mathbf{B})$ is a reduced structure iff $(\mathbf{M}, \mathbf{B})$ has property* EQP.

Proof: Assume that $(\mathbf{M}, \mathbf{B})$ is a reduced structure, by definition, the Leibniz congruence $\Omega(\mathbf{M}, \mathbf{B}, Id_{\mathbf{B}})$ is the identity on M. Then, if we define $(E, Id_{\mathbf{B}})$ as in the previous proof, by Lemma 5, $E \subseteq \Omega(\mathbf{M}, \mathbf{B}, Id_{\mathbf{B}})$, consequently E is also the identity and thus $(\mathbf{M}, \mathbf{B})$ has the EQP. Conversely, assume that $(\mathbf{M}, \mathbf{B})$ has the EQP. If $d, e \in M$ and $(d, e) \in \Omega(\mathbf{M}, \mathbf{B}, Id_{\mathbf{B}})$, since $x \approx y$ is an atomic formula and $\|d \approx d\|^{\mathbf{B}}_{\mathbf{M}} = 1$, by definition of the Leibniz congruence, $\|d \approx e\|^{\mathbf{B}}_{\mathbf{M}} = 1$. Then, by EQP, we have that $d = e$, therefore $(\mathbf{M}, \mathbf{B})$ is a reduced structure. □

Note that the interpretation of the $\approx$ symbol in a reduced structure is not necessarily crisp. Adding a new axiom it is possible to obtain crisp interpretations: Crispness Axiom (Crisp) $\forall x \forall y (x \approx y \vee \neg(x \approx y))$ (for the details about this axiom see Chapter 5 of [8]).

Corollary 12 *Let T be a set of Γ-sentences containing axioms $Sim \cup Cong \cup Crisp$. Then, for every formula $\phi(x_1, \ldots, x_n) \in \Gamma$, the following holds:* $T \vdash \forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall y_n ((x_1 \approx y_1 \& \ldots \& x_n \approx y_n) \rightarrow (\phi(x_1, \ldots, x_n) \leftrightarrow \phi(y_1, \ldots, y_n))$

Proof: Since axiom Crisp holds, the interpretation of the $\approx$ symbol in a reduced structure is the identity. Therefore, since the logic is complete with respect to its reduced models, we obtain the desired result. □

4 The relative relation

We now present the notion of relative relation, a relation between structures that will play in fuzzy predicate languages the same role that the isomorphism relation plays in classical predicate languages with equality. This notion was introduced by G. Zubieta in [14], but only for relational structures, and independently by W. Blok and D. Pigozzi in [1], for the special case of logical matrices. A characterization of the relative relation for classical first-order logics can be found in [10].

Definition 13 *Let $(\mathbf{M}_1, \mathbf{B}_1)$ and $(\mathbf{M}_2, \mathbf{B}_2)$ be two Γ-structures, we say that the pair (R, T) is a* relative relation *between $(\mathbf{M}_1, \mathbf{B}_1)$ and $(\mathbf{M}_2, \mathbf{B}_2)$ iff*

1. *$T \subseteq B_1 \times B_2$ is a relation such that $dom(T) = B_1$, $rg(T) = B_2$ and*
   (a) *for every connective $\delta \in L$, if $(a_i, b_i) \in T$, then*
   $$(\delta_{\mathbf{B}_1}(a_1, \ldots, a_n), \delta_{\mathbf{B}_2}(b_1, \ldots, b_n)) \in T$$
   (b) *for every $a \in B_1$, if $b, b' \in rg(a)$, then $dom(b) = dom(b')$.*
   (c) *for every $b \in B_2$, if $a, a' \in dom(b)$, then $rg(a) = rg(a')$.*
2. *$R \subseteq M_1 \times M_2$ is a relation such that $dom(R) = M_1$, $rg(R) = M_2$ and*
   (a) *For each constant symbol $c \in \Gamma$, $(c_{\mathbf{M}_1}, c_{\mathbf{M}_2}) \in R$.*
   (b) *For each n-ary function symbol $F \in \Gamma$, if for every $0 < i \leq n$, $(a_i, b_i) \in R$,*
   $$(F_{M_1}(a_1, \ldots, a_n), F_{M_2}(b_1, \ldots, b_n)) \in R$$
   (c) *For each n-ary predicate $P \in \Gamma$, if for every $0 < i \leq n$, $(d_i, e_i) \in R$,*
   $$(\|P(d_1, \ldots, d_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1}, \|P(e_1, \ldots, e_n))\|^{\mathbf{B}_2}_{\mathbf{M}_2}) \in T$$

where for every $a \in B_1$, $rg(a) = \{b \in B_2 : (a, b) \in T\}$ and for every $b \in B_2$, $dom(b) = \{a \in B_1 : (a, b) \in T\}$. We denote by $(R, T) : (\mathbf{M}_1, \mathbf{B}_1) \sim (\mathbf{M}_2, \mathbf{B}_2)$ when (R, T) is a relative relation (or simply by $(\mathbf{M}_1, \mathbf{B}_1) \sim (\mathbf{M}_2, \mathbf{B}_2)$ when there is a relative relation between them).

ISBN: 978-989-95079-6-8

Theorem 14 *Let* $(\mathbf{M}_1, \mathbf{B}_1)$ *and* $(\mathbf{M}_2, \mathbf{B}_2)$ *be two* Γ*-structures. The following are equivalent:*

1. *There is a relative relation* $(R, T) : (\mathbf{M}_1, \mathbf{B}_1) \sim (\mathbf{M}_2, \mathbf{B}_2)$
2. *There are congruences* (E_1, θ_1) *and* (E_2, θ_2) *such that*

$$(\mathbf{M}_1/E_1, \mathbf{B}_1/\theta_1) \cong (\mathbf{M}_2/E_2, \mathbf{B}_2/\theta_2)$$

Proof: 1. $\Rightarrow$ 2. Assume that (R, T) is a relative relation. We define $\theta_1 = \{(a, a') \in B_1 \times B_1 : rg(a) = rg(a')\}$, $\theta_2 = \{(b, b') \in B_2 \times B_2 : dom(b) = dom(b')\}$, $E_1 = \Omega(\mathbf{M_1}, \mathbf{B_1}, \theta_1)$ and $E_2 = \Omega(\mathbf{M_2}, \mathbf{B_2}, \theta_2)$.

It is clear by definition that θ_1 and θ_2 are equivalence relations. Now we show that they are L-congruences, we prove that, for every connective $\delta \in L$, for every $0 < i \leq n$ and $a_i, a_i' \in B_1$, if $rg(a_i) = rg(a_i')$, then $rg(\delta_{\mathbf{B}_1}(a_1, \ldots, a_n)) = rg(\delta_{\mathbf{B}_1}(a_1', \ldots, a_n'))$. Let us assume that for every $0 < i \leq n$, $rg(a_i) = rg(a_i')$. Since $dom(T) = B_1$, for every $0 < i \leq n$, we choose $b_i \in B_2$ such that $(a_i, b_i) \in T$. Thus, by assumption, since $rg(a_i) = rg(a_i')$, we have also that $(a_i', b_i) \in T$. By condition 1.(a) of the definition of relative relation,

$$(\delta_{\mathbf{B}_1}(a_1, \ldots, a_n), \delta_{\mathbf{B}_2}(b_1, \ldots, b_n)) \in T$$

and $(\delta_{\mathbf{B}_1}(a_1', \ldots, a_n'), \delta_{\mathbf{B}_2}(b_1, \ldots, b_n)) \in T$, finally, by condition 1.(c) of the definition of relative relation, we have that

$$rg(\delta_{\mathbf{B}_1}(a_1, \ldots, a_n)) = rg(\delta_{\mathbf{B}_1}(a_1', \ldots, a_n')).$$

In order to show that θ_2 is a congruence, we can follow an analogous proof, using condition 1.(b) of the definition of relative relation, instead of 1.(c). Now we define a mapping $g : \mathbf{B}_1/\theta_1 \to \mathbf{B}_2/\theta_2$. First we fix enumerations (possibly with repetitions) $(a_i : i \in I)$ and $(b_i : i \in I)$ of $\mathbf{B}_1$ and $\mathbf{B}_2$ respectively, with the property that, for every $i \in I$, $(a_i, b_i) \in T$. And then let, for every $i \in I$, $g([a_i]_{\theta_1}) = [b_i]_{\theta_2}$. By using the definition of relative relation it is easy to check that g is well-defined, and it is an L-isomorphism.

Now, since $dom(R) = M_1$ and $rg(R) = M_2$, we can fix enumerations (possibly with repetitions) $(d_j : j \in J)$ and $(e_j : j \in J)$ of M_1 and M_2 respectively, with the property that, for every $j \in J$, $(d_j, e_j) \in R$. And then let, for every $j \in J$,

$$f([d_j]_{\Omega(\mathbf{M_1}, \mathbf{B_1}, \theta_1)}) = [e_j]_{\Omega(\mathbf{M_2}, \mathbf{B_2}, \theta_2)}$$

Let us see that f is well defined. Let $(d_j, d_j') \in \Omega(\mathbf{M_1}, \mathbf{B_1}, \theta_1)$, we show that $(e_j, e_j') \in \Omega(\mathbf{M_2}, \mathbf{B_2}, \theta_2)$. Let $\phi(y, x_1, \ldots, x_n)$ be an atomic formula, and a sequence of elements $k_1, \ldots, k_n \in M_2$. Since $rg(R) = M_2$, we can choose $l_1, \ldots, l_n \in M_1$ such that for every $0 < i \leq n$, $(l_i, k_i) \in R$. Remark that, since $(d_j, e_j) \in R$ and for every $0 < i \leq n$, $(l_i, k_i) \in R$, by conditions 2.(a) and 2.(b) of the definition of relative relation, we have for every Γ-term t, $(t_{M_1}(d_j, l_1, \ldots, l_n), t_{M_2}(e_j, k_1, \ldots, k_n)) \in R$. Assume that the formula $\phi(y, x_1, \ldots, x_n)$ is of the form $P(t_1, \ldots, t_s)(y, x_1, \ldots, x_n)$, where P is a s-ary predicate symbol and $t_1, \ldots, t_s$ are Γ-terms. Since $(d_j, d_j') \in \Omega(\mathbf{M_1}, \mathbf{B_1}, \theta_1)$, $(\|P(t_1, \ldots, t_s)(d_j, l_1, \ldots, l_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1}$, $\|P(t_1, \ldots, t_s)(d_j', l_1, \ldots, l_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1}) \in \theta_1$, thus, by 2.(c) of the definition of relative relation,
$\|P(t_1, \ldots, t_s)(e_j, k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2}$ and
$\|P(t_1, \ldots, t_s)(e_j', k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2}$
$\in rg(\|P(t_1, \ldots, t_s)(d_j, l_1, \ldots, l_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1})$

and by 1.(b), $dom(\|P(t_1, \ldots, t_s)(e_j, k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2}) = dom(\|P(t_1, \ldots, t_s)(e_j', k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2})$ and we obtain the desired result:

$$(\|\phi(e_j, k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2}, \|\phi(e_j', k_1, \ldots, k_n)\|^{\mathbf{B}_2}_{\mathbf{M}_2}) \in \theta_2.$$

In an analogous way it is easy to check that f is one-to-one and that (f, g) is an isomorphism.

2. $\Rightarrow$ 1. Let (f, g) be an isomorphism. Define (R, T) in the following way: for every $b_1 \in B_1, b_2 \in B_2$,$(a, b) \in T$ iff $g([a]_{\theta_1}) = [b]_{\theta_2}$ and for every $d \in M_1, e \in M_2$, $(d, e) \in R$ iff $f([d]_{E_1}) = [e]_{E_2}$. Using the fact that (f, g) is an isomorphism, it is easy to check that $(R, T) : (\mathbf{M}_1, \mathbf{B}_1) \sim (\mathbf{M}_2, \mathbf{B}_2)$ is a relative relation. □

Remark that our approach differs from [6] because the relative relation is not a measure of the degree of similarity between structures. We left for future work the study of the relationship between these two notions.

Definition 15 (R, T) *is an* elementary *relative relation if for every formula* $\phi(x_1, \ldots, x_n) \in \Gamma$ *and for every* $0 < i \leq n$*, if* $(a_i, b_i) \in R$*, then*

$$(\|\phi(a_1, \ldots, a_n)\|^{\mathbf{B}_1}_{\mathbf{M}_1}, \|\phi(b_1, \ldots, b_n))\|^{\mathbf{B}_2}_{\mathbf{M}_2}) \in T$$

By induction on the complexity of the formulas it is straightforward to show that, in case (E_1, θ_1) and (E_2, θ_2) are elementary congruences in Theorem 14, then (R, T) is also an elementary relative relation and $(\mathbf{M}_1, \mathbf{B}_1) \equiv (\mathbf{M}_2, \mathbf{B}_2)$. The following corollary show us that, when we study structures over the same algebra, we can improve Theorem 14.

Corollary 16 *Let* $(\mathbf{M}_1, \mathbf{B})$ *and* $(\mathbf{M}_2, \mathbf{B})$ *be two* Γ*-structures. The following are equivalent:*

1. *There is a relative relation* $(R, Id_{\mathbf{B}}) : (\mathbf{M}_1, \mathbf{B}) \sim (\mathbf{M}_2, \mathbf{B})$
2. $(\mathbf{M}_1, \mathbf{B})^r \cong (\mathbf{M}_2, \mathbf{B})^r$

Proof: By the proof of Theorem 14. □

Let us see now an example of two fuzzy equivalence relations that are relatives but there is no homomorphism from one onto the other. Let $(\mathbf{M}_1, \mathbf{B})$ and $(\mathbf{M}_2, \mathbf{B})$ be defined as follows: the domains of the structures are $M_1 = \{d_1, d_2, e_1, e_2\}$ and $M_2 = \{d_1', d_2', d_3', e_1'\}$ respectively. The fuzzy equivalence relation E_1 is defined by: for every $i, j \in \{1, 2\}$, $E_1(d_i, d_j) = 1 = E_1(e_i, e_j)$ and $E_1(d_i, e_j) = r$, where $r \neq 1$ is a fixed element of $\mathbf{B}$. And the fuzzy equivalence relation E_2 is defined by: for every $i, j \in \{1, 2, 3\}$, $E_2(d_i', d_j') = 1 = E_2(e_1', e_1')$ and $E_2(d_i', e_1') = r$. It is easy

ISBN: 978-989-95079-6-8

to check that $(R, Id_{\mathbf{B}}) : (\mathbf{M}_1, \mathbf{B}) \sim (\mathbf{M}_2, \mathbf{B})$, where R is the relation $R = \{(d_i, d'_j), (e_i, e'_j) : i, j \in \{1, 2, 3\}\}$.

The relation of being either a homomorphic image or a homomorphic counter-image is not in general transitive. Its transitivization is precisely the relative relation, as the following propositions show.

Lemma 17 *Given an L-algebra* $\mathbf{B}$*,* $\sim$ *is an equivalence relation in the class of* $\mathbf{B}$*-structures.*

Proof: By Corollary 16, because $\cong$ is an equivalence relation on the class of reduced structures. □

Notation: Given Γ-structures $(\mathbf{N}, \mathbf{B})$ and $(\mathbf{O}, \mathbf{B})$, we denote by $(\mathbf{N}, \mathbf{B}) \in \mathbf{H}(\mathbf{O}, \mathbf{B})$ the fact that there exists a mapping f from O onto N such that $(f, Id_{\mathbf{B}})$ is a homomorphism.

Proposition 18 *Let* $(\mathbf{M}, \mathbf{B})$ *and* $(\mathbf{N}, \mathbf{B})$ *be two* Γ*-structures. The following are equivalent:*

1. $(\mathbf{M}, \mathbf{B}) \sim (\mathbf{N}, \mathbf{B})$.
2. *There is a natural number* n *and* Γ*-structures* $(\mathbf{O}_1, \mathbf{B}), \ldots, (\mathbf{O}_n, \mathbf{B})$ *such that* $(\mathbf{M}, \mathbf{B}) = (\mathbf{O}_1, \mathbf{B})$*,* $(\mathbf{N}, \mathbf{B}) = (\mathbf{O}_n, \mathbf{B})$ *and for every* $0 < i < n$*, either* $(\mathbf{O}_{n+1}, \mathbf{B}) \in \mathrm{H}(\mathbf{O}_n, \mathbf{B})$ *or* $(\mathbf{O}_n, \mathbf{B}) \in \mathrm{H}(\mathbf{O}_{n+1}, \mathbf{B})$.
3. *There is a* Γ*-structure* $(\mathbf{O}, \mathbf{B})$ *such that* $(\mathbf{M}, \mathbf{B}), (\mathbf{N}, \mathbf{B}) \in \mathrm{H}(\mathbf{O}, \mathbf{B})$.
4. *There is a* Γ*-structure* $(\mathbf{O}, \mathbf{B})$ *such that* $(\mathbf{O}, \mathbf{B}) \in \mathrm{H}(\mathbf{M}, \mathbf{B})$ *and* $(\mathbf{O}, \mathbf{B}) \in \mathrm{H}(\mathbf{N}, \mathbf{B})$.

Proof: 4. $\Rightarrow$ 2. and 3. $\Rightarrow$ 2. are clear. 1. $\Rightarrow$ 4. By Corollary 16. 2. $\Rightarrow$ 1. By the definition of relative relation, given two Γ-structures $(\mathbf{O}, \mathbf{B})$ and $(\mathbf{D}, \mathbf{B})$, if $(\mathbf{O}, \mathbf{B}) \in \mathbf{H}(\mathbf{D}, \mathbf{B})$, then $(\mathbf{O}, \mathbf{B}) \sim (\mathbf{D}, \mathbf{B})$ therefore we can apply the transitive property of the relative relation (Lemma 17).

1. $\Rightarrow$ 3. Assume that there is a relative relation $(R, Id_{\mathbf{B}}) : (\mathbf{M}, \mathbf{B}) \sim (\mathbf{N}, \mathbf{B})$. Since $dom(R) = M$ and $rg(R) = N$, we can fix enumerations (possibly with repetitions) $(d_j : j \in J)$ and $(e_j : j \in J)$ of M and N respectively, with the property that, for every $j \in J$, $(d_j, e_j) \in R$. Now we define a structure $(\mathbf{O}, \mathbf{B})$ and homomorphisms $(f^M, Id_{\mathbf{B}})$ and $(f^N, Id_{\mathbf{B}})$ from $(\mathbf{O}, \mathbf{B})$ onto $(\mathbf{M}, \mathbf{B})$ and $(\mathbf{N}, \mathbf{B})$ respectively.

The algebraic reduct of $(\mathbf{O}, \mathbf{B})$ is the algebra Ter_J of Γ-terms generated by the set of variables $V_J = \{v_j : j \in J\}$. We define the function $f_0^M : V_J \to M$ as follows: for every $j \in J$, $f_0^M(v_j) = d_j$. Then we extend f_0^M in the usual way, to a homomorphism f^M from Ter_J onto M. Finally we define the interpretation of the predicate symbols in $(\mathbf{O}, \mathbf{B})$: for each n-ary predicate $P \in \Gamma$ and elements $t_1, \ldots, t_n \in O$, $P_O(t_1, \ldots, t_n) = P_M(f^M(t_1), \ldots, f^M(t_n))$. So defined $(f^M, Id_{\mathbf{B}})$ is clearly a homomorphism onto $(\mathbf{M}, \mathbf{B})$. Now we define f_0^N by: for every $j \in J$, $f_0^N(v_j) = e_j$ and we extend f_0^N as before, to a homomorphism f^N of the terms algebra onto N. Since for every $j \in J$, $(d_j, e_j) \in R$, for every Γ-term t,

$$(t_M(d_{j_1}, \ldots, d_{j_n}), t_N(e_{j_1}, \ldots, e_{j_n})) \in R$$

and using this fact it is easy to check that $(f^N, Id_{\mathbf{B}})$ is also a homomorphism onto $(\mathbf{N}, \mathbf{B})$. □

5 Conclusions

Work in progress includes the development of usual tools of model theory such as the method of diagrams or ultraproducts in order to work in fuzzy predicate logic. The use of a reduced semantics will allow us to show when one structure is either embeddable or elementarily embeddable in another in terms of extensions of the usual diagrams with special sentences. By using relative relations we could define new operations among structures, such as *ultrafilter-products*, more suitable for working with equality-free languages. Future work will be devoted also to provide different characterizations of the relation of elementary equivalence and some strenghtenings of this notion.

Acknowledgment

Research partially supported by the Generalitat de Catalunya, under the grant 2005-SGR-00093 and the Spanish project "Agreement Technologies" (CONSOLIDER CSD2007-0022, INGENIO 2010).

References

[1] W. J. Blok and D. Pigozzi. Protoalgebraic logics. 45:337–369, 1986.

[2] W. J. Blok and D. Pigozzi. *Algebraizable logics*, volume 396 of *Mem. Amer. Math. Soc.* A.M.S., Providence, January 1989.

[3] A. Di Nola and G. Gerla. Fuzzy models of first order languages. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 32:331–340, 1986.

[4] P. Hájek and P. Cintula. On theories and models in fuzzy predicate logics. *The Journal of Symbolic Logic*, 71(3):863–880, 2006.

[5] V. Novák et al. *Mathematical principles of fuzzy logic*. The Kluwer International Series in Engineering and Computer Science, 517. Kluwer Academic Publishers, Boston, MA, 1999.

[6] R. Belohlávek. *Fuzzy Relational Systems: Foundations and Principles.* Springer, 2002.

[7] P. Cintula et al. Distinguished algebraic semantics for t-norm based fuzzy logics: methods and algebraic equivalencies. *Annals of Pure and Applied Logic (article in press)*, 2009. Submitted.

[8] P. Cintula and P. Hájek. Triangular norm based predicate fuzzy logics. In *Fuzzy Sets and Systems ((article in press)*, 2009. Forthcoming.

[9] P. Hájek. *Metamathematics of fuzzy logic*, volume 4 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 1998.

[10] P. Dellunde. *Contributions to the model theory of equality-free logic*. Ph. D. Dissertation, University of Barcelona, 1996.

[11] X. Caicedo. A characterization of first order rational pavelka's logic. Preprint, slides of the talk in the Conference *Logic, Algebra and Truth Degrees*, Siena http://www.mat.unisi.it/ latd2008/slides/Caicedo.pdf (2008).

[12] E. H. Ruspini. On the semantics of fuzzy logic. *International Journal of Approximate Reasoning*, 5(1):45–88, 1991.

[13] L. A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

[14] G. Zubieta. Clases aritméticas definidas sin igualdad. *Boletín de la Sociedad Matemática Mejicana*, 2:45–53, 1957.

ISBN: 978-989-95079-6-8

Fuzzy Clustering for Finding Fuzzy Partitions of Many-Valued Attribute Domains in a Concept Analysis Perspective

Yassine Djouadi[1] Basma Alouane[2] Henri Prade[1]

1. IRIT, University Paul Sabatier, Toulouse
118 Route de Narbonne, 31062 Toulouse, Cedex 9, France
2. LIFAB Laboratory, University of Boumerdes
Avenue de l'Independance, Boumerdes, Algeria
Email: djouadi@irit.fr, B_alouane@umbb.dz, prade@irit.fr

Abstract— *Although an overall knowledge discovery process consists of a distinct pre-processing stage followed by the data mining step, it seems that existing formal concept analysis (FCA) and association rules mining (ARM) approaches, dealing with many-valued contexts, mainly focus on the data mining stage. An "intelligent" pre-processing of input contexts is often absent in existing FCA/ARM approaches, leading to an unavoidable information loss. Usually, many-valued attribute domains need to be first fuzzily partitioned. However, it is unrealistic that the most appropriate fuzzy partitions can be provided by domain experts. In this paper, an unsupervised learning stage, based on Fuzzy C-Means algorithm, is proposed in order to get fuzzy partitions that are faithful to data for quantitative attribute domains, and consequently for avoiding the loss of valuable association rules due to the use of empirical fuzzy partitions. More precisely, the paper reports an experiment where it is shown that some rules are no longer found because their support or confidence is too low when using such empirical partitions. Experimental results show that the learned fuzzy partition outperforms human expert fuzzy partitions. More generally, the paper provide discussions about the handling of many-valued attributes in both fuzzy FCA and fuzzy ARM.*

Keywords— Many-valued formal contexts, fuzzy partitions, fuzzy C-means, association rules.

1 Introduction

Association rules mining (ARM for short) [1] is one of the most widely used data mining technique. This model of knowledge represents the patterns of co-occurrence of items in a collection of transactions. Typically, input data set is in the form of sets of items called transactions (e.g. an item may be things we buy in a market). Thus an example of an association rule is an expression of the form *"beer"* $\wedge$ *"sausage"* $\Rightarrow$ *"mustard"* which represents the fact that purchasing beer and sausage implies purchasing mustard with some degree. It is important to point out that, in the classical setting, considered input data sets (called *contexts* in the rest of this paper) are binary (i.e. are expressed in terms of binary attributes).

Formal concept analysis [2] (FCA for short) consists also of learning some knowledge representation model in an unsupervised way. In the classical setting of this theory, the considered input data set is called *formal context* (also called *context* in this paper). It consists of a binary relation between a set of objects and a set of properties. This relation is usually represented as a table with rows corresponding to objects, columns corresponding to properties (or conversely), and table entries containing 1's or 0's depending on whether an object has or not the corresponding property.

It appears that, in their classical setting, both FCA and ARM are concerned with crisp binary contexts. However real-world data usually contain heterogenous kinds of data (binary, categorical, quantitative, etc...). In order to generalize binary settings to many-valued contexts (i.e. with quantitative or categorical attributes) it is usually suggested to partition quantitative attribute domain into many fuzzy intervals.

Few approaches propose a complete pre-processing stage that is intended to get an optimal fuzzy partition. In [3], authors propose a genetic algorithm-based clustering method that adjusts the centroids of the clusters, which are to be handled as midpoints of the triangular fuzzy partitions. Another approach, also based on genetic algorithms, is proposed in [4]. Based on a linguistic representation model, authors propose to perform a genetic lateral tuning of the membership functions. Authors consider also that each attribute has a predefined number of linguistic terms associated to it. For example, *Low*, *Middle* and *High* are the fuzzy linguistic terms covering the domain of the attribute *Age*.

Concerning these approaches, one may notice that they consider fuzzy partitions restricted to triangular membership functions. On the other hand, the number of clusters has to be empirically fixed for each attribute. Then, the fuzzy partitioning may be considered as ad-hoc to some extent and is far to be information lossless. In other words, the distribution of the data into the context is not taken into account for the assessment of discovered patterns. For this purpose, we present an approach that consists of learning the fuzzy partitions by means of the well-known Fuzzy C-Means algorithm. We also consider cluster validity measures in order to get an optimal multi-prototype context representation.

The paper is organized as follows. Section 2 gives some basic notions of FCA and ARM frameworks, while both crisp and fuzzy partitioning methods are reviewed. The next section highlights the information loss problem with empirical or expert-based fuzzy partitions. In the fourth section, data faithful fuzzy partitions are learned using Fuzzy C-Means algorithm. Experimental results on the *adult* database (UCI Machine Learning Repository) show that the proposed approach avoids information loss and outperforms in terms of support and confidence measures, expert-based partitions. Finally, we conclude and give future perspectives.

ISBN: 978-989-95079-6-8

2 Basic Notions

This section introduces main notions and terminology for both FCA and ARM paradigms and then, presents quantitative context representation and its inherent sharp boundary problem.

2.1 Association Rules Mining

Formally, let $\mathcal{X}$ be a set of objects (called also transactions) $\left(\mathcal{X} = \{x_1, x_2, ..., x_{|\mathcal{X}|}\}\right)$ and let $\mathcal{Y}$ be a set of items $\left(\mathcal{Y} = \{y_1, y_2, ..., y_{|\mathcal{Y}|}\}\right)$. A context $\mathbb{K}$ is denoted as $\mathbb{K} = (\mathcal{X}, \mathcal{Y}, \mathcal{R})$ where $\mathcal{R}$ is a binary relation ($x\mathcal{R}y$ indicates that item y satisfies the transaction x). An association rule is of the form $A \Rightarrow B$ where $A, B \subseteq \mathcal{Y}$, $A \cap B = \emptyset$ and, $A, B \neq \emptyset$. The intended meaning of this rule is that the presence of all of the items of A in a transaction implies the presence of all of the items of B in the same transaction. Association rules are assigned support and confidence measures for a given context. The support defines the percentage of transactions that satisfy (contain) both A and B. Whereas the confidence expresses the conditional probability that B is satisfied given that A is satisfied. These measures denoted respectively `Sup` and `Conf` are defined as:

$$\texttt{Sup}\,(A \Rightarrow B) = \frac{|\{x \in \mathcal{X} \mid \forall y \in A \cup B,\, x\mathcal{R}y\}|}{|\mathcal{X}|} \quad (1)$$

$$\texttt{Conf}(A \Rightarrow B) = \frac{\texttt{Sup}(A \Rightarrow B)}{\texttt{Sup}(A)} \quad (2)$$

Hence, an itemset A is said frequent, if its support is greater than a fixed threshold ξ (i.e. $\texttt{Sup}(A) \geqslant \xi$). A large variety of representational and computational approaches for association rules mining have been published in the literature in the case of binary items. Among them, frequent-based [5] and closed-based [6] approaches. While the former generate the lattice of all frequent itemsets, the latter generate a more condensed representation through the so-called closed itemsets lattice. For instance, based on the notion of closed itemsets, and minimal implication rules base (e.g. Duquenne-Guigues base, Luxemburger base) [7], association rules may be easily inferred from a closed itemsets lattice structure without enumerating them [8]. This kind of approach is closely related to formal concept analysis since the closed itemsets lattice is isomorphic to the concepts lattice which is described below.

2.2 Formal Concept Analysis

Formal concept analysis theory, proposed by [2], deals with a particular kind of analysis of data based on a formal context. Let us consider a similar notation as used above. A set $\mathcal{X}$ of objects, a set $\mathcal{Y}$ of attributes and a formal context $\mathbb{K} = (\mathcal{X}, \mathcal{Y}, \mathcal{R})$ with $\mathcal{R}$ a binary relation ($x\mathcal{R}y$ means that object x verifies attribute y). Let A (resp. B) be a subset of $\mathcal{X}$ (resp. $\mathcal{Y}$), two mappings $\uparrow$ and $\downarrow$ are symmetrically defined:

$$A\uparrow = \{y \mid \forall x \in A,\, x\mathcal{R}y\}$$
$$B\downarrow = \{x \mid \forall y \in B,\, x\mathcal{R}y\}$$

From a formal context, one can construct pairs (A, B) such that $A\uparrow = B$ and $B\downarrow = A$ known as formal concepts, where A and B are called respectively the extent and the intent of the corresponding concept. The set $\mathfrak{B}(\mathbb{K})$ of all formal concepts in the formal context $\mathbb{K}$ is equipped with a partial order (denoted $\preceq$) defined as:

$$(A_1, B_1) \preceq (A_2, B_2) \text{ iff } A_1 \subseteq A_2 \text{ (or, equivalently, } B_1 \supseteq B_2)$$

$(\mathfrak{B}(\mathbb{K}), \preceq)$ forms a complete lattice, called the *concept lattice* of $\mathbb{K}$ (for more details on the lattice structure see [9]).

It appears that ARM and more generally FCA both require, in their classical setting, the same context representation which is a binary context.

2.3 Quantitative Context Representation

In order to generalize binary context settings to many-valued contexts, Srikant et al. [10] suggested to partition a quantitative attribute domain into intervals. In their approach, these authors proposed to map a quantitative or categorical attribute value into a Boolean attribute value in the following way. Let y be an attribute whose domain is quantitative (the principle remains the same with a categorical attribute). Let $Dom\,(y)$ denote the domain of y (according to the relational paradigm, $Dom\,(y)$ is necessarily a finite set). For a given object x we denote by $x[y]$ the value v of y for the object x (i.e. $x[y] = v$).

If the set $Dom\,(y)$ consists of few values, the mapping is straightforward. Conceptually, the authors propose to have instead of the attribute y, as many new attributes as the cardinality of the set $Dom\,(y)$. The new attributes are of the form "$y.v_k$" (where $v_k \in Dom\,(y)$). The values of the new binary formal context are given as:

$$x[y.v_k] = \begin{cases} 1 & \text{if } x[y] = v_k \\ 0 & \text{otherwise.} \end{cases}$$

If the set $Dom\,(y)$ is large, it is partitioned into different intervals $I_1, I_2, ..., I_p$. Attributes of the new binary context are of the form "$y.I_k$" and are given as:

$$x[y.I_k] = \begin{cases} 1 & \text{if } x[y] \in I_k \\ 0 & \text{otherwise.} \end{cases}$$

Note that in the two above cases only particular subsets (singletons or definite intervals) are allowed.

An example of a many-valued context is illustrated in Table 1, while Table 2 gives the corresponding binary mapping.

Table 1: An example of a many-valued context.

Obj\Attr	Age	Marital Status	Native Country	Work Class
Alice	23	single	USA	never-worked
Boris	24	married	USA	private
Cyril	31	married	USA	fed-gov
David	60	divorced	Haiti	self-emp

Table 2: An example of a many-valued context mapping.

Obj\Attr	Age.[20,29]	Age.[30,39]	Age.[40,49]	...
Alice	1	0	0	...
Boris	1	0	0	...
Cyril	0	1	0	...
David	0	0	0	...

ISBN: 978-989-95079-6-8

2.4 Crisp vs. Fuzzy Context Representation

In [10], the authors point out a dilemma between support and confidence measures caused by crisp partitions. On the one hand, if the number of intervals for a quantitative attribute is important, the support for any single interval may be low. Hence, without using larger intervals, some rules involving this attribute may not be found because they lack minimal support. On the other hand, if the intervals are too large, some rules may not have minimal confidence. Indeed, when interval becomes larger, the support of the antecedent A of a rule $A \Rightarrow B$ becomes higher. Consequently, the confidence of the rule i.e. the ratio $\texttt{Supp}\left(A \Rightarrow B\right) / \texttt{Supp}\left(A\right)$ decreases and the rule may not have minimal confidence.

Besides, in [11, 12] other authors point out some undesirable threshold effects caused by crisp partitions. Such effects are well-known, for instance, from histograms in statistics: A slight variation of the boundary points of the intervals can have a noticeable effect on the histogram induced by a number of observations. Likewise, the variation of a partition can strongly influence the evaluation of association rules.

Considering the above dilemma and the undesirable threshold effects, fuzzy partitions (in fuzzy intervals) have been widely used instead of crisp partitions (crisp intervals). This has led to fuzzy association rules [13, 14, 15] and to fuzzy concepts analysis [16, 17]. For example, Table 3 illustrates a fuzzy representation for the many-valued context given in Table 1, when choosing the fuzzy partition $\mathcal{F}_1$ (see Fig. 1).

Table 3: Fuzzy representation of a many-valued context.

Obj\Attr	Age.Very Young	Age.Young	Age. Middle Age	...
Alice	0.4	0.6	0	...
Boris	0.2	0.8	0	...
Cyril	0	1	0	...
David	0	0	0	...

3 Information Loss in Quantitative Context

There are two distinct tendencies to transform a many-valued context into a fuzzy partitioned one. The former is based on human expertise, or is even given empirically. The latter uses dedicated heuristics (e.g. [16]). These partitioning methods induce important information loss as it will be showed in this section. Let us before, recall the definition of a fuzzy partition.

Definition1.
A fuzzy partition of a set X on a universe U is a family $\mathcal{F}\left(X\right) = \left(F_1, ..., F_n\right)$ such that:

1. $\forall i = 1..n \quad F_i \neq \emptyset \wedge F_i \neq U$
2. $\forall u \in U, \quad \sum_i F_i\left(u\right) = 1$

3.1 Empirical Fuzzy Partitions

Classical fuzzy set theory has a powerful tool to manage granularity, namely linguistic variables, introduced by Zadeh [18]. Fuzzy partitions are generally based on those linguistic variables. However the complete specification of a linguistic variable may obviously differ from a user to another. For example, in a survey of the literature [11, 12, 14, 15] related to fuzzy association rules discovery, one may find different fuzzy representations of the attribute *Age*. These partitions $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$, $\mathcal{F}_4$ are respectively showed in Fig. 1, Fig. 2, Fig. 3, Fig. 4.

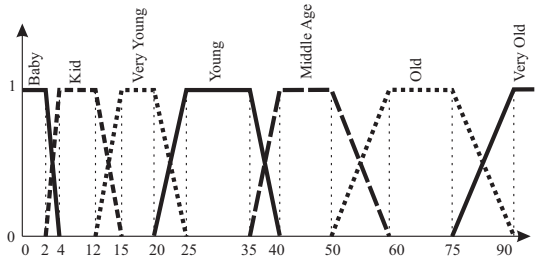


Figure 1: Fuzzy partition $\mathcal{F}_1$ given in [14].

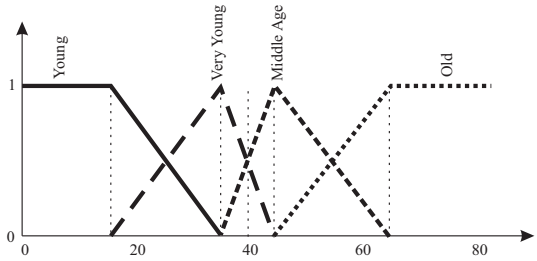


Figure 2: Fuzzy partition $\mathcal{F}_2$ given in [15].

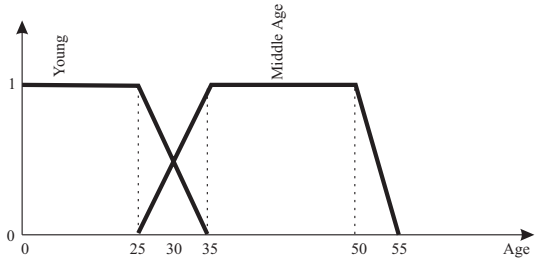


Figure 3: Fuzzy partition $\mathcal{F}_3$ given in [11].

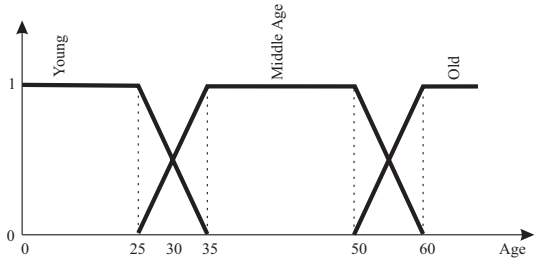


Figure 4: Fuzzy partition $\mathcal{F}_4$ given in [12].

In order to highlight the information loss that may arise with a given fuzzy context compared to another one, we induce all the fuzzy association rules w.r.t a fixed support threshold ξ. The *Adult* benchmark database (UCI Machine Learning Repository) available at http://www.ics.uciedu/mlearn/MLRepositery is taken as a target context for our experiments. Note that, we have restricted the *Adult* context up to six categorical attributes (i.e. *Marital_Status*, *Native_Country*, *Work_Class*, *Salary_Class*, *Education*, *Occupation*) and one quantitative attribute (i.e. *Age*). According to

ISBN: 978-989-95079-6-8

the fuzzy partitions represented respectively in Fig. 1, Fig. 2, Fig. 3, Fig. 4 (namely $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$, $\mathcal{F}_4$) we apply the fuzzy association rules mining algorithm that is fully described in [13] for each of the four fuzzy *Adult* contexts.

Note that, the algorithm in [13] uses a co-occurrence semantics [19] for fuzzy rules. However the information loss phenomenon could be observed as well with gradual or certainty semantics [19] .

Thus among the fuzzy rules induced by each partition $\mathcal{F}_1$-$\mathcal{F}_4$, we have chosen the following association rules to illustrate information loss:
$R1$:(Age.young)⇒(MaritalStatus.single)
$R2$:(Age.middle)⇒(WorkClass.private)
$R3$:(Age.young)⇒(WorkClass.private)
$R4$:(Age.middle)∧(WorkClass.private)⇒(NativeCountry.USA)

By analyzing the results, illustrated in Table 4, it may be remarked that:

1. Different support and confidence measures are found. This confers an uncertainty to the final decision process.

2. More important is a clear information loss materialized by the non discovery of many rules. For example, the rule $R1$ does not exist in the fuzzy context $\mathcal{F}_1$ even though this fuzzy context contains the fuzzy label *young*.

Besides, defining such fuzzy partitions may not be intuitive for humans. This is especially true for attributes where there does not exist a more or less standard linguistic vocabulary which fuzzily partition the attribute domain.

Table 4: Information loss for expert fuzzy partitioning.

	$R1$ Sup/Conf	$R2$ Sup/Conf	$R3$ Sup/Conf	$R4$ Sup/Conf
$\mathcal{F}_1$	$\not\models$	0.26/0.72	$\not\models$	0.13/0.90
$\mathcal{F}_2$	0.08/0.77	0.17/0.68	0.31/0.75	0.08/0.88
$\mathcal{F}_3$	0.05/0.98	$\not\models$	$\not\models$	0.20/0.92
$\mathcal{F}_4$	0.13/0.59	0.47/0.77	$\not\models$	0.30/0.94

where the mark $\not\models$ indicates that the rule could not be discovered since it lacks support.

3.2 Heuristic-Based Contexts Mapping

The second family of approaches deals with a particular generalization of many-valued contexts based on some heuristics depending on the nature of the context. In [16] a "single-pivot" representation is proposed. The so-called pivot is taken as the maximum value in the attribute domain. The equivalent fuzzy context is obtained by dividing the quantitative values by the maximal value of the domain, thus achieving the estimation of a fuzzy proximity to the maximum value. For example, the equivalent fuzzy representation of the attribute *Age* given in Table 1 will be: {(Alice,0.38), (Boris,0.4), (Cyril,0.51), (David,1)}. One may remark that some non uniform data distribution (e.g. with few values near the pivot and almost all values far from it) will lead to an unavoidable information loss for any knowledge discovery process.

As a summary, it appears that empirical, or heuristic-based fuzzy partitioning for data mining applications may be not well-suited. Indeed, in [20] two important stages (i.e. *Processing* and *Transformation*) are intended to get a well-suited representation for the next data mining stage. This is why, we propose a fuzzy learning stage.

4 Learning Fuzzy Partitions

The Fuzzy C-Means algorithm [21, 22] is one of the strongest and widely used fuzzy clustering method. It is based on an iterative optimization of a fuzzy objective function. This function characterizes the partition and is given as:

$$J_m(X,U,V) = \sum_{j=1}^{n}\sum_{k=1}^{c} \mu_{jk}^m \left\|x_j - v_k\right\|^2 \tag{3}$$

where $X = \{x_1, ..., x_n\}$ represents the domain to partition, $V = \{v_1, ..., v_c\}$ represents the set of cluster centers and $U = (\mu_{ki})_{n\times c}$ is the fuzzy partition matrix composed of the membership degree of each item x_j in each cluster k. The exponent m is a parameter called *fuzzifier*.

The distance measure usually used is the Euclidian distance between a datum and a prototype (cluster center). Each cluster V_k is determined as:

$$V_k = \frac{\sum_{j=1}^{n}(\mu_{jk})^m x_j}{\sum_{j=1}^{n}(\mu_{jk})^m} \tag{4}$$

Whereas the elements μ_{jk} of the fuzzy partition matrix are given as:

$$\begin{cases} \mu_{jk} = 0 & \text{if } J_j \neq \emptyset \wedge k \notin J_j \\ \sum_{k\in J_j}\mu_{jk} = 1 & \text{if } J_j \neq \emptyset \wedge k \in J_j \\ \mu_{jk} = \dfrac{1}{\sum_{l=1}^{c}\left(\frac{\|x_j - v_k\|^2}{\|x_j - v_l\|^2}\right)^{\frac{1}{m-1}}} & \text{if } J_j = \emptyset \end{cases} \tag{5}$$

with $J_j = \{k \mid 1 \leqslant k \leqslant c, \|x_j - v_k\| = 0\}$

4.1 Algorithm

The proposed algorithm integrates cluster validity measures [23] in order to obtain an optimal fuzzy partition. It takes as input a minimal (resp. maximal) number of fuzzy partitions allowed, denoted $Cmin$ (resp. $Cmax$). It takes also a fixed threshold tolerance denoted θ. The vector V contains the cluster centers and it is randomly initialized. The algorithm is given as follows:

Begin.
1: $Cmin := 2$; $Cmax := 10$;
2: $\theta := 0.001$;
3: **For** l:=$Cmin$ To $Cmax$ **Do**
4: Initialize $V^{(l)}$ randomly;
5: Compute the partition matrix $U^{(l)}$ with expression (5);
6: Compute the objective function J with (3);
7: Update the cluster centers using (4);
8: Compute the new partition matrix with (5);
9: Compute the new objective function J_{new};
10: If $(J - J_{new}) < \theta$ goto 3 else goto 7;
11: **End For**
12: Determine l s.t. $Vpc(l) = \min\left(Vpc(k)\right)_{k=Cmin,Cmax}$
13: Return $V^{(l)}$, $U^{(l)}$

ISBN: 978-989-95079-6-8

End.

Definition2.
The equivalent fuzzy representation of a quantitative attribute y corresponds to the pair $(\mathcal{F}_y, U_y)$ where:

$$\mathcal{F}_y = \{y.V_1, .., y.V_k, .., y.V_c\}$$
$$U_y\,(x_i, y.V_k) = \mu_{ik}$$

This definition means also that we have chosen to label each fuzzy cluster k by its prototype (cluster center) V_k. Considering that membership degrees to these fuzzy clusters are determined using a distance measure (i.e. Euclidian distance), the semantics attached to the discovered fuzzy association rules $Y.V_k(\overrightarrow{y}) \Rightarrow Z.V_l(\overrightarrow{z})$ is of the type: *the presence of $\overrightarrow{y}$'s near to the prototype $Y.V_k$, imply the presence of $\overrightarrow{z}$'s near to the prototype $Z.V_l$.*

4.2 Cluster validity measures

Our algorithm is implemented considering two options:
i) the end-user may specify the number c of partitions, and even the center of classes (i.e. V_1, ..., V_c),
ii) the learning algorithm determines itself the optimal number of classes using the *Vpc* validity measure as done in the above algorithm.

The partition coefficient (*Vpc*) [23] measure has been chosen since this measure is based on the subsethood degrees μ_{jk}. It is given as:

$$Vpc = \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{c}\mu_{jk}^2 \qquad (6)$$

5 Experimental Results

All experiments were performed on a 3.0 Ghz Pentium-IV PC with 512 MB main memory running on Microsoft Windows XP. Algorithms were implemented in C++.

The fuzzy association rules mining algorithm is an A-PRIORI like implementation. Cardinalities of fuzzy sets are computed using a $\Sigma-$count operator. While conjunction in expressions of support and confidence, resp. (1) and (2), is performed using the *min* operator. We have also considered weighted $\alpha-cuts$ for computing support and confidence measures which avoids accumulation of small (irrelevant) cardinalities. The reader is referred to [13] for complete details.

Experiments were performed on the *Adult* database (see section §3.1.). Many objects in the *Adult* context contain missing (null) values. Such objects have been deleted from the initial context which reduces the size of the considered context to 32561 objects. The probability distribution for the *Age* attribute is represented in Fig. 5.

Our fuzzy partition learning process for the *Age* attribute yields two fuzzy clusters and induces the partition:

$$\mathcal{F}_{age} = \{age.25, age.43\}$$

where 25 and 43 correspond to the prototypes (cluster centers). The complete partition is showed in Fig. 6.

According to the partition $\mathcal{F}_{age}$, we show below some new fuzzy association rules that are now discovered.

(Age.25)∧(Workclass.*Private*)⇒(NativeCountry.*USA*)
[support=0.32;confidence=0.89]

(Age.25)∧(Workclass.*Private*)∧(NativeCountry.*USA*)⇒
(Salary_Class.*LeastThan50K$*) [support=0.31;confidence=1]

(Age.43)⇒(Workclass.*Private*)
[support=0.24;confidence=0.69]

(Age.25)⇒(Workclass.*Private*)
[support=0.35;confidence=0.94]

(Age.43)∧(Workclass.*Private*)⇒(NativeCountry.*USA*)
[support=0.32;confidence=0.89]

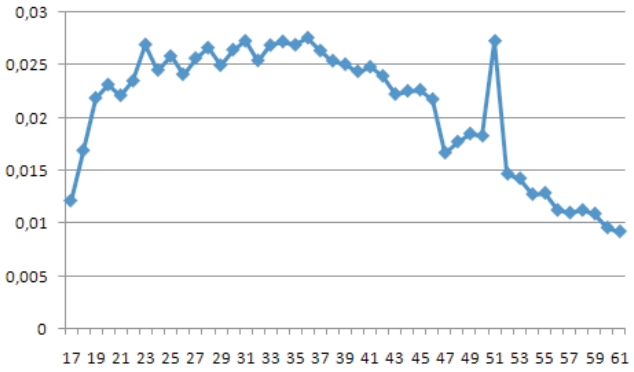


Figure 5: Probability distribution for *Age* attribute in *Adult* context.

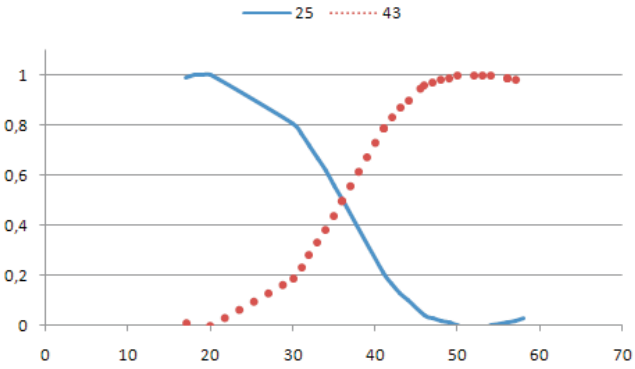


Figure 6: The learned fuzzy partition for *Age* attribute in *Adult* context.

It has been checked that, related to the prototypes, almost all rules are now discovered. For instance, the fourth above rule is a counterpart of the rule $R3$ ((Age.young)⇒(WorkClass.private)), which was not found using the three partitions $\mathcal{F}_1$, $\mathcal{F}_3$, $\mathcal{F}_4$.

6 Conclusion and Perspectives

This paper deals with the transformation stage in a knowledge discovery process. An automatic approach is proposed for the mapping of a many-valued context into a fuzzy context. This mapping is achieved using a fuzzy clustering method. The difficulties of humans for determining fuzzy partitions and fuzzy grades are avoided, which makes things easier for end-users in real life applications. The Fuzzy C-Means algorithm has been

ISBN: 978-989-95079-6-8

chosen as a strong means to get optimized fuzzy partitions. Besides experimental results show that the proposed approach avoids information loss, which departs from the results of almost all existing approaches.

As a first perspective, we intend to enlarge the fuzzy partitions setting by allowing a hierarchical clusters representation. Similar ideas were already explored at least in two different ways. Wolff in [24] proposes a so-called conceptual scale theory which handles many-valued (quantitative) attributes in a fuzzy setting without being restricted to particular subsets (singletons or definite intervals). Bosc et al. in [25] have outlined a knowledge extraction method based on the idea on the fuzzy summarization of data, which may also be used in order to cluster data in a hierarchical-like way.

As pointed out in [26] fuzzy association rules may be also of interest for their increased expressivity. That's why, we intend secondly to investigate other semantics than co-occurrence, namely the one of gradual and certainty rules.

Lastly, it is worth of interest to enlarge fuzzy contexts to incomplete information fuzzy contexts (i.e. with unknown values) [27],[28].

Acknowledgment

The first author is grateful to the A.U.F. (Agence Universitaire de la Francophonie at Brussels-Belgium) who has supported this work under a mobility grant.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining associations rules between sets of items in large databases. *Proc. of the ACM SIGMOD Conf on Management of Data. Washington D.C.*, pages 207–216, 1993.

[2] Rudolf Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. *Rival. I. (Ed): Ordered Sets. Reidel, Dordrecht. Boston*, pages 445–470, 1982.

[3] M. Kaya and R. Alhajj. Genetic algorithm based framework for mining fuzzy association rules. *Fuzzy Sets and Systems*, 152(3):587–601, 2005.

[4] J. Alcala-Fdez, R. Alcala, M.J. Gacto, and F. Herrera. Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms. *Fuzzy Sets and Systems*, 160(7):905–921, 2009.

[5] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. *Proc* 20^{th} *International Conf Very Large Databases. Santiago, Chile*, pages 478–499, 1994.

[6] M.J. Zaki and H.J. Hsiao. Efficient algorithms for mining closed itemsets and their lattice structure. *IEEE Transactions on Knowledge and Data Enginnering*, 17(4):462–478, 2005.

[7] V. Duquenne and J.L. Guigues. Famille minimale d'implications informatives resultant d'un tableau de donnes binaires. *Mathmatiques et sciences humaines*, 24(95):5–18, 1986.

[8] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Efficient mining of association rules using closed itemsets lattices. *Information Systems*, 24:25–46, 1999.

[9] B. Ganter and R. Wille. *Formal Concept Analysis*. Springer Mathematical Foundations, edn, 1999.

[10] R. Srikant and R. Agrawal. Mining quantitative association rules in large relational tables. *Proc. of the ACM SIGMOD Conf on Management of Data. Montreal Canada*, pages 1–12.

[11] D. Dubois, H. Prade, and T. Sudkamp. On the representation, measurement and discovery of fuzzy associations. *IEEE Transactions on Fuzzy Systems*, 13(2):250–262, 2005.

[12] T. Sudkamp. Examples, counterexamples, and measuring fuzzy associations. *Fuzzy Sets and Systems*, 149:57–71, 2005.

[13] Y. Djouadi., S. Redaoui, and K. Amroun. Mining association rules under imprecision and vagueness: towards a possibilistic approach. *FUZZ'IEEE 07, International Conference on Fuzzy Systems. London, UK*, pages 722–727, July 2007.

[14] M. Delgado, N. Marin, D. Sanchez, and M.A. Vila. Fuzzy association rules : general model and applications. *IEEE Transactions on Fuzzy Systems*, 11(2):214–255, 2005.

[15] W.H. Au and K.C.C. Chan. Mining fuzzy association rules in a bank-account databases. *IEEE Transactions on Fuzzy Systems*, 11(2):238–248, 2003.

[16] N. Messai, M.D. Devignes, A. Napoli, and M.S. Tabbone. Many-valued concept lattices for conceptual clustering and information retrieval. *ECAI'08,* 18^{th} *European Conference on Artificial Intelligence. Patras, Greece*, pages 722–727, July.

[17] R. Belohlavek and V. Vychodil. What is a fuzzy concept lattice. *CLA 05, Proc. Olomounc. Czech Republic*, pages 34–45, 2005.

[18] L.A. Zadeh. The concept of a linguistic and its application to approximate reasoning. *Information Science*, 8:199–249, 1975.

[19] M. Serrurier, D. Dubois, H. Prade, and T. Sudkamp. Learning fuzzy rules with their implication operators. *Data and Knowledge Engineering*, 60:71–89, 2007.

[20] U. Fayyad, G.P. Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *Artificial Intelligence Magazine*, pages 37–54, 1996.

[21] J.C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters. *Journal of Cybernetics*, 3:35–57, 1973.

[22] J.C. Bezdek, J. Keller, R. Krishnapuram, and N.R. Pal. *Fuzzy models and algorithms for pattern recognition and image processing*. Kluwer Academic Publisher, 1999.

[23] J.C. Bezdek and N.R. Pal. On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems*, 3(3):370–379, 1995.

[24] K.E. Wolff. Concepts in fuzzy scaling theory: order and granularity. *Fuzzy Sets and Systems*, 132:63–75, 2002.

[25] P. Bosc, D. Dubois, O. Pivert, H.Prade, and M. De Calmes. Fuzzy summarization of data using fuzzy cardinalities. *Proc. of the 9th Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU 2002), Annecy, France*, pages 1553–1559, 2002.

[26] M. Serrurier and H. Prade. Improving expressivity of inductive logic programming by learning different kinds of fuzzy rules. *Soft Computing*, 11(5):459–466, 2007.

[27] P. Burmeister and R. Holzer. Treating incomplete knowledge in formal concepts analysis. *Lecture Notes in Computer Science 3626, Bernhard Ganter (Eds.)*, pages 114–126, 2005.

[28] Didier Dubois, Florence Dupin de Saint Cyr Bannay, and Henri Prade. A possibilty-theoretic view of formal concept analysis. *Fundamenta Informaticae*, 75(1-4):195–213, 2007.

ISBN: 978-989-95079-6-8

On possibilistic sequencing problems with fuzzy parameters

Adam Kasperski[1] Paweł Zieliński[2]

1.Institute of Industrial Engineering and Management, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370, Wrocław, Poland
2.Institute of Mathematics and Computer Science, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370, Wrocław, Poland
Email: adam.kasperski@pwr.wroc.pl, pawel.zielinski@pwr.wroc.pl

Abstract— *In this paper a wide class of sequencing problems with imprecise parameters is discussed. The imprecision is modeled by using closed intervals and fuzzy intervals, whose membership functions are regarded as possibility distributions for the values of unknown parameters. A possibilistic interpretation of fuzzy problems is provided, some solution concepts are proposed and some algorithms for computing the solutions are designed.*

Keywords— Sequencing, min-max regret, possibility theory, fuzzy interval, fuzzy optimization and design

1 Introduction

In a sequencing problem we wish to find a feasible order of elements, called jobs, to achieve some goal. This goal typically depends on job completion times and may also depend on some other job parameters such as due dates or weights. There are a lot of deterministic sequencing problems with different computational properties and a comprehensive description of them can be found for instance in [1]. Unfortunately, most of sequencing problems turned out to be NP-hard but there are also some important problems for which efficient polynomial algorithms exist.

Sequencing problems involve many parameters whose exact values are often unknown. For instance, a job processing time, which is a crucial parameter in all sequencing problems, is rarely precisely known. In order to model such a situation a stochastic approach can be applied (see e.g. [2]), where for every unknown parameter a probability distribution is specified and, typically, the expected cost of a solution is minimized. The stochastic approach has several drawbacks. Namely, it may be hard or expensive to estimate the probability distributions for the parameters. Also, minimizing the expected performance may be not reasonable if the obtained solution is used only once.

In practice, decision makers are often interested in minimizing the cost of a decision in the worst case, that is under the worst realization of parameters that may occur. In recent years a *robust* approach do discrete optimization has attracted a considerable attention. In this approach we specify a *scenario set* containing all possible realizations of parameters which may occur. No probability distribution over the scenario set is given. One of methods of determining the scenario set consists of specifying an interval of possible values for every unknown parameter. A scenario set is then the Cartesian product of all the uncertainty intervals. A natural criterion for choosing a solution under imprecision is the *maximal regret*, which expresses the maximum distance of a solution from optimality over all scenarios [3]. A deeper discussion on using the maximal regret criterion in decision making under uncertainty can be found in [4]. The class of min-max regret sequencing problem with interval parameters has been discussed in a number of papers, for instance in [5, 6, 7, 8, 9, 10, 11, 12]. We briefly describe the known results in this area later in this paper.

In recent years, theory of fuzzy sets has been applied to model the imprecision in optimization, in particular in sequencing problems. A good review of different concepts in fuzzy optimization can be found in [13]. In papers [14, 15, 16] some single machine sequencing problems with fuzzy processing times, fuzzy due dates and fuzzy precedence constraints have been discussed. In these papers a fuzzy due date expresses a degree of satisfaction with job completion time and a sequence is computed, which maximizes the minimum satisfaction or the sum of satisfactions over all jobs. In [17] an alternative approach has been proposed, where the possibility or necessity of job delays is minimized. Also, in [18] an optimality evaluation of sequences under fuzzy parameters has been investigated. We recall and extend this approach here.

In this paper we wish to propose a new approach to sequencing problems with fuzzy parameters. We generalize the min-max regret problems with interval parameters to the fuzzy case by extending the interval uncertainty representation to the fuzzy interval one. The fuzzy parameters induce a possibility distribution over the scenario set, which becomes then reacher in information than the classical scenario set. We provide a possibilistic interpretation of the obtained fuzzy problem and describe a solution concept, which is an adaptation of the elegant solution method proposed in [19] for fuzzy linear programming. This solution concept has been recently adopted in [20] for a class of combinatorial optimization problems. Apart from showing a general framework, we also point out some difficulties which arise when one try to solve a particular problem. Contrary to the class of combinatorial optimization problems described in [20], the sequencing problems are typically more complex to solve and there are very few general properties that are valid for all problems. However, as for the problems described in [20], the main computational difficulties are in the classical interval case and the algorithms known for the interval uncertainty representation can be generalized to the fuzzy case.

This paper is organized as follows. In Section 2 we recall the definition of the classical deterministic sequencing problem. In Section 3 we present the min-max regret approach to sequencing problems with interval parameters - we provide a

ISBN: 978-989-95079-6-8

general formulation and we recall some known results in this area. Finally, in Section 4 we introduce the class of sequencing problems with fuzzy parameters. We give a link between the fuzzy problems and the min-max regret ones. We construct some general methods of solving a fuzzy problem and we illustrate them using a sample problem.

2 Deterministic sequencing problems

We are given a set of *jobs* $J = \{J_1, \ldots, J_n\}$, which may be partially ordered by some *precedence constraints* of the form $i \rightarrow j$, where $i, j \in J$. For the simplicity of notations we will identify every job $J_i \in J$ with its index $i \in \{1, \ldots, n\}$. A solution is a *sequence* (permutation) $\sigma = (\sigma(1), \ldots, \sigma(n))$ of J and it represents an order in which the jobs are processed. A sequence σ is *feasible* if $i \rightarrow j$, $i \neq j$, implies that job j appears after i in σ. We will denote by $\boldsymbol{\mathcal{S}}$ the set of all feasible sequences. We use $C_i(\sigma)$ to denote the *completion time* of job i in sequence σ. In a *single machine* case, a processing time p_i is given for every job $i \in J$ and if $i = \sigma(k)$, then $C_i(\sigma) = \sum_{j=1}^{k} p_{\sigma(j)}$. If every job must be processed on $m > 1$ machines, first on machine 1 next on machine 2 and so on, then we get the *permutation flow shop problem*. In this case p_{ij} is a processing time of job i on machine j and $C_i(\sigma)$ is the time when job i is finished on the m-th machine (see Figure 1).

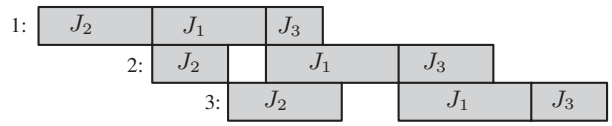


Figure 1: A permutation flow show problem with $m = 3$, $n = 3$ and the schedule that corresponds to the sequence $(2, 1, 3)$.

For every job $i \in J$, there is a cost function $f_i : R \rightarrow R$, which measures the cost of completing i at time t. The value of this function may depend on some other parameters associated with job i such as *due date* d_i or *weight* w_i. Finally, $F(\sigma)$ denotes a cost of sequence σ. There are two general types of the cost function $F(\sigma)$, namely a *bottleneck objective* $F(\sigma) = \max_{i=1,\ldots,n} f_i(C_i(\sigma))$ and a *sum objective* $F(\sigma) = \sum_{i=1}^{n} f_i(C_i(\sigma))$.

In a deterministic sequencing problem we seek a feasible sequence of the minimal cost, that is we solve the following optimization problem:

$$\min_{\sigma \in \boldsymbol{\mathcal{S}}} F(\sigma). \tag{1}$$

Our analysis in the next sections of this paper will be based on the concept of a *deviation*. In the deterministic case the deviation of a sequence σ expresses its distance to optimum and is defined as follows:

$$\delta_\sigma = F(\sigma) - \min_{\rho \in \boldsymbol{\mathcal{S}}} F(\rho). \tag{2}$$

Obviously, σ is optimal if and only if $\delta_\sigma = 0$. If the deviation is positive, then it indicates a distance of σ from optimality.

Sequencing problems are usually described by using a convenient Graham's notation. So, every sequencing problem can be denoted by a triple $\alpha|\beta|\gamma$, where α is the machine environment ($\alpha = 1$ for the single machine case), β specifies the job characteristic and γ describes the objective function. The following examples of sequencing problems are well known and will be explored later in this paper:

- $1|prec|L_{max}$. In this problem we seek a sequence, which minimizes the maximum *lateness*. So, the bottleneck objective is $\max_{i \in J}\{C_i(\sigma) - d_i\}$. We can generalize this problem by minimizing the maximal weighted lateness or modify it to minimize the maximal weighted *tardiness* (see [1]).

- $1||\sum C_i$. In this problem there are no precedence constraints between jobs and we seek a sequence for which the sum of completion times of all jobs, i.e. the *total flow time* $\sum_{i \in J} C_i(\sigma)$, is minimal.

- $1|p_i = 1|\sum w_i U_i$. In this problem there are no precedence constraints between jobs and all jobs have unit processing times. A job is *late* in σ if $C_i(\sigma) > d_i$. In this case we write $U_i(\sigma) = 1$ and $U_i(\sigma) = 0$ otherwise. The cost of σ is $\sum_{i \in J} w_i U_i(\sigma)$, so it expresses the weighted number of late jobs.

- $Fm||C_{max}$. This is the permutation flow shop problem with $m > 1$ machines. There are no precedence constraints between jobs and the bottleneck cost of σ is the *makespan* $\max_{i \in J} C_i(\sigma)$, that is the completion time of the last job on the last machine. The problem is polynomially solvable only if $m = 2$ and becomes NP-hard for $m \geq 3$ (see [1]).

The above examples illustrate a large variety of basic sequencing models. As we will see in the next section, they have quite different computational properties under imprecision and the state of the art in this area is still far from being complete.

3 Minmax regret sequencing problems

In practice, the exact values of parameters in a sequencing problem such as processing times, due dates or weights may be not precisely known. Assume that we have l parameters and the value of a parameter ξ_i, $i = 1, \ldots, l$, may fall within a closed interval $[\underline{\xi}_i, \overline{\xi}_i]$ independently of the values of the other parameters. A parameter ξ_i is *precise* if $\underline{\xi}_i = \overline{\xi}_i$. Every vector $S = (s_1, \ldots, s_l) \in R^l$ such that $s_i \in [\underline{\xi}_i, \overline{\xi}_i]$ is called a *scenario* and it expresses a possible state of the world, where $\xi_i = s_i$ for $i = 1, \ldots, l$. A scenario is called *extreme* if all parameters take the lower or upper bounds in their uncertainty intervals. We use Γ to denote the set of all possible scenarios. Hence Γ is the Cartesian product of all the uncertainty intervals. Now the cost of a sequence σ depends on scenario $S \in \Gamma$ and we will denote it as $F(\sigma, S)$. We will also denote by $F^*(S)$ the cost of an optimal sequence under scenario S. In order to obtain the value of $F^*(S)$ we need to solve problem (1) under the fixed realization of parameters S. It is clear that deviation of σ also depends on scenario S and we will denote it as $\delta_\sigma(S) = F(\sigma, S) - F^*(S)$.

Now the optimality of a sequence σ can be characterized by a *deviation interval* $[\underline{\delta}_\sigma, \overline{\delta}_\sigma]$, where $\underline{\delta}_\sigma = \min_{S \in \Gamma} \delta_\sigma(S)$ is the minimal deviation and $\overline{\delta}_\sigma = \max_{S \in \Gamma} \delta_\sigma(S)$ is the maximal deviation over all scenarios. The quantity $\overline{\delta}_\sigma$ is called in

ISBN: 978-989-95079-6-8

literature the *maximal regret* of σ and it expresses the largest distance of σ from optimum over the scenario set Γ. A scenario S_σ, for which the deviation of σ attains maximum, is called a *worst case scenario* for σ. So, under the interval uncertainty representation, we only know that $\delta_\sigma \in [\underline{\delta}_\sigma, \overline{\delta}_\sigma]$ and we can give the following characterization of a feasible sequence: a sequence σ is *possibly optimal* if $\underline{\delta}_\sigma = 0$ and it is *necessarily optimal* if $\overline{\delta}_\sigma = 0$. Notice that a sequence is possibly optimal if and only if it is optimal under some scenario $S \in \Gamma$ and it is necessarily optimal if and only if it is optimal for all scenarios $S \in \Gamma$.

Now the question arises which sequence of $\boldsymbol{\mathcal{S}}$ should be chosen. A necessarily optimal one is a natural choice. However, it rarely exists in most practical situations, because the necessary optimality is very strong criterion. In a more reasonable approach we can minimize the the maximal regret, that is the quantity $\overline{\delta}_\sigma$, which may be viewed as a distance to the necessary optimality. As a result we get the following *min-max regret sequencing problem*:

$$\min_{\sigma \in \boldsymbol{\mathcal{S}}} \overline{\delta}_\sigma. \tag{3}$$

We call an optimal solution to (3) a *min-max regret sequence*. In the next section we briefly review some known facts on problem (3).

3.1 Complexity of min-max regret sequencing problems

The first problem that arises while analyzing a particular min-max regret sequencing problem is the computation of the maximal regret for a given sequence σ, that is the quantity $\overline{\delta}_\sigma$. Unfortunately, contrary to the class of problems discussed in [20], there is no general method of performing this task. For the problems considered in [20], it is possible to find two extreme scenarios that maximize and minimize the deviation and, consequently, the computation of the maximal regret has the same complexity as the deterministic problem. For sequencing problems the situation is much more complex. First of all, for some problems there may be no extreme scenario that maximizes (minimizes) the deviation [21]. Furthermore, computing the maximal regret may be much more time consuming than solving a deterministic problem. For instance, in the min-max regret $1||\sum C_i$ problem with interval processing times, computing $\overline{\delta}_\sigma$ requires solving an assignment problem while an optimal sequence under a given scenario can be computed in $O(n \log n)$ time. An extreme case has been described in [9], where a permutation flow shop problem with m machines and with interval processing times and with only 2 jobs has been discussed. Since the solution set contains only two sequences, the total computational effort is focused on computing the maximal regret of a given sequence.

It is not surprising that min-max regret sequencing problems are typically hard to solve. There are only several problems that are known to be polynomially solvable. A polynomial algorithm for $1|prec|L_{max}$ with interval processing times and interval due dates has been constructed in [6]. A polynomial algorithm for $1|prec|\max w_i T_i$ with interval weights and precise processing times and precise due dates has been proposed in [8]. Apart from this two problems, only some very special cases are known to be polynomially solvable, for instance $Fm||C_{max}$ with interval processing times and only two jobs [9] and $1|p_i = 1|\sum w_i U_i$ with interval weights and a precise common due date ($d_1 = d_2 = \cdots = d_n$) [21]. This latter problem is equivalent to the min-max regret version of the selecting items problem, which is known to be polynomially solvable [21, 22].

Among the known negative results, the most important one has been obtained in [10], where it has been shown that the min-max regret $1||\sum C_i$ problem with interval processing times is NP-hard. This problem is also known to be approximable within 2 [12] and can be solved by using a mixed integer programming formulation proposed in [11]. However, the complexity status of a number of basic problems is still unknown. We do not know whether $F2||C_{max}$ with interval processing times is NP-hard. We can, however, compute the maximal regret of a given sequence and solve the problem by using a branch and bound algorithm [7, 3]. Also the problem $1|p_i = 1|\sum w_i U_i$ with interval weights and arbitrary precise due dates and the problem $1|prec|\max w_i L_i$ with interval processing times, interval due dates and precise weights are open. The former problem can be solved by a mixed integer programming formulation [21]. There are also a large number of different sequencing problems whose min-max regret versions have never been investigated and they should be the subjects of further research.

4 Fuzzy sequencing problems

In this section we propose an extension of the min-max regret approach to sequencing problems. We will apply fuzzy intervals to model the imprecise parameters and we will use possibility theory to define a solution concept. A comprehensive description of possibility theory can be found in [23].

4.1 Basic notions on possibility theory

A *fuzzy interval* $\tilde{A}$ is a fuzzy set in the space of reals whose membership function $\mu_{\tilde{A}}$ is normal, quasi concave and upper semicontinuous. It is typically assumed that the support of a fuzzy interval is compact. The main property of a fuzzy interval is the fact that all its λ-cuts, that is the sets $\tilde{A}^\lambda = \{x : \mu_{\tilde{A}}(x) \geq \lambda\}$, $\lambda \in (0, 1]$, are closed intervals. We will assume that $\tilde{A}^0$ is the smallest closed set containing the support of $\tilde{A}$. So, every fuzzy interval $\tilde{A}$ can be represented as a family of closed intervals $\tilde{A}^\lambda = [\underline{a}^\lambda, \overline{a}^\lambda]$, parametrized by the value of $\lambda \in [0, 1]$. In many practical applications, the class of *trapezoidal fuzzy intervals* is used. A trapezoidal fuzzy interval, denoted by a quadruple $(\underline{a}, \overline{a}, \alpha, \beta)$, can be represented as the family $[\underline{a} - \alpha(1 - \lambda), \overline{a} + \beta(1 - \lambda)]$ for $\lambda \in [0, 1]$. Notice that this representation contains classical intervals ($\alpha = \beta = 0$) and real numbers (additionally $\underline{a} = \overline{a}$) as special cases.

In this paper we adopt a possibilistic interpretation of a fuzzy interval [23]. Assume that for an imprecise real quantity ξ a fuzzy interval with membership function μ_ξ is given. This membership function expresses a *possibility distribution* for the values of ξ, namely $\Pi(\xi = x) = \mu_\xi(x)$ is the possibility of the event that ξ will take the value of x. It is easily seen that the closed interval $[\underline{\xi}^\lambda, \overline{\xi}^\lambda]$, $\lambda \in [0, 1]$, contains all values of ξ, whose possibility of occurrence is not less than λ. In particular, the interval $[\underline{\xi}^0, \overline{\xi}^0]$ should contain all possible values of ξ, while the interval $[\underline{\xi}^1, \overline{\xi}^1]$ should contain the most plausible ones. Some methods of obtaining the possibility distribution of an unknown quantity can be found in [23].

ISBN: 978-989-95079-6-8

Let $\tilde{G}$ be a fuzzy set in the space of reals with membership function $\mu_{\tilde{G}}$. Then $\xi \in \tilde{G}$ is a *fuzzy event* and the necessity that $\xi \in \tilde{G}$ holds is defined in the following way:

$$\begin{aligned} \mathrm{N}(\xi \in \tilde{G}) &= 1 - \Pi(\xi \notin \tilde{G}) = \\ &= 1 - \sup_{x \in R} \min\{\mu_{\xi}(x), 1 - \mu_{\tilde{G}}(x)\} \end{aligned} \tag{4}$$

where $1 - \mu_{\tilde{G}}(x)$ is the membership function of the complement of the fuzzy set $\tilde{G}$. It is not difficult to see that if $\tilde{G} = (0, \overline{g}, 0, \beta) = (\overline{g}, \beta)$, then the following equality is true:

$$\mathrm{N}(\xi \in \tilde{G}) = 1 - \inf_{\lambda \in [0,1]} \{\overline{\xi}^{\lambda} \leq \overline{g}^{1-\lambda}\} \tag{5}$$

and $\mathrm{N}(\xi \in \tilde{G}) = 0$ if $\overline{\xi}^{1} > \overline{g}^{0}$. Equality (5) is illustrated in Figure 2.

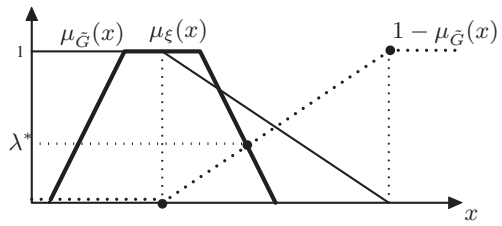


Figure 2: $\mathrm{N}(\xi \in \tilde{G}) = 1 - \lambda^*$.

4.2 Possibilistic sequencing problem

Assume that for every unknown parameter ξ_i, $i = 1, \ldots, l$, in a sequencing problem a fuzzy interval with membership function μ_{ξ_i} is specified. According to the interpretation given in the previous section μ_{ξ_i} is a possibility distribution for the values of ξ_i. Under the assumption that all parameters are unrelated, there is a possibility distribution over all scenarios $S = (s_1, \ldots, s_l) \in R^l$ defined as follows:

$$\begin{aligned} &\pi(S) = \Pi\left(\bigwedge_{i=1}^{l} [\xi_i = s_i]\right) = \min_{i=1,\ldots,l} \Pi(\xi_i = s_i) = \\ &= \min_{i=1,\ldots,l} \mu_{\xi_i}(s_i). \end{aligned} \tag{6}$$

So, the value of $\pi(S)$ express the possibility of the event that scenario S will occur. Notice that we generalize in this way the scenario set $\Gamma \subset R^l$ described in Section 3. Indeed, for the interval uncertainty representation $\pi(S) = 1$ if $S \in \Gamma$ and $\pi(S) = 0$ otherwise. In the fuzzy case $\pi(S)$ may take any value in the interval $[0, 1]$ so fuzzy intervals allow us to model the imprecision in a more sophisticated manner.

In the interval case the value of deviation δ_{σ} falls within a closed interval. Analogously, in the fuzzy case it falls within a fuzzy interval with membership function $\mu_{\delta_{\sigma}}$. Of course, $\mu_{\delta_{\sigma}}$ is a possibility distribution for the values of δ_{σ} and, according to possibility theory, it is defined as follows:

$$\mu_{\delta_{\sigma}}(x) = \Pi(\delta_{\sigma} = x) = \sup_{\{S:\ \delta_{\sigma}(S) = x\}} \pi(S). \tag{7}$$

Recall that the statement "'σ is optimal"' is equivalent to the assertion $\delta_{\sigma} = 0$. Consequently, we can define the degrees of possible and necessary optimality of a given sequence as follows:

$$\Pi(\sigma \text{ is optimal}) = \Pi(\delta_{\sigma} = 0) = \mu_{\delta_{\sigma}}(0), \tag{8}$$

$$\mathrm{N}(\sigma \text{ is optimal}) = \mathrm{N}(\delta_{\sigma} = 0) = 1 - \sup_{x > 0} \mu_{\delta_{\sigma}}(x). \tag{9}$$

As in the interval case, a question arises which sequence should be chosen. In order to perform this task we adopt a concept first applied to fuzzy linear programming in [19]. Assume that a decision maker knows his/her preference about the sequence deviation and expresses it by using a fuzzy interval $\tilde{G} = (\overline{g}, \beta)$. So, the values of deviation in $[0, \overline{g}]$ are fully accepted, the values in $[\overline{g} + \beta, \infty)$ are not at all accepted and the degree of acceptance decreases from 1 to 0 in $[\overline{g}, \overline{g} + \beta]$. Our aim is to compute a feasible sequence $\sigma \in \boldsymbol{\mathcal{S}}$, for which the necessity of the event $\delta_{\sigma} \in \tilde{G}$ is maximal, namely we wish to solve the following optimization problem:

$$\max_{\sigma \in \boldsymbol{\mathcal{S}}} \mathrm{N}(\delta_{\sigma} \in \tilde{G}). \tag{10}$$

This solution concept has been first proposed for linear programming problem with fuzzy objective function coefficients [19]. Observe that (10) can also be expressed as $\min_{\sigma \in \boldsymbol{\mathcal{S}}} \Pi(\delta_{\sigma} \in \tilde{G}^d)$, where $\tilde{G}^d$ is the complement of the fuzzy goal $\tilde{G}$ with membership function $1 - \mu_{\tilde{G}}(x)$. If we fix $\tilde{G} = (0, 0)$ in (10), then we get a special case, in which we seek a feasible sequence that maximizes the degree of necessary optimality:

$$\max_{\sigma \in \boldsymbol{\mathcal{S}}} \mathrm{N}(\delta_{\sigma} = 0) = \max_{\sigma \in \boldsymbol{\mathcal{S}}} \mathrm{N}(\sigma \text{ is optimal}). \tag{11}$$

We focus now on a method of solving the problem (10). According to equality (5), the problem (10) is equivalent to the following optimization problem:

$$\begin{aligned} &\min \lambda \\ &\overline{\delta}_{\sigma}^{\lambda} \leq \overline{g}^{1-\lambda} \\ &\sigma \in \boldsymbol{\mathcal{S}} \\ &\lambda \in [0, 1] \end{aligned} \tag{12}$$

If λ^* is the optimal objective value and σ^* is an optimal solution to (12), then $\mathrm{N}(\delta_{\sigma^*} \in \tilde{G}) = 1 - \lambda^*$. If (12) is infeasible, then $\mathrm{N}(\delta_{\sigma} \in \tilde{G}) = 0$ for all feasible sequences σ. Of course, if we replace expression $\overline{g}^{1-\lambda}$ with 0 in (12), then we get an equivalent formulation of the problem (11).

Let us focus now on the quantity $\overline{\delta}_{\sigma}^{\lambda}$. The closed interval $[\underline{\delta}_{\sigma}^{\lambda}, \overline{\delta}_{\sigma}^{\lambda}]$, $\lambda \in [0, 1]$, contains all values of deviation δ_{σ}, whose possibility of occurrence is not less than λ. So

$$\overline{\delta}_{\sigma}^{\lambda} = \sup_{\{S\,:\,\pi(S) \geq \lambda\}} \{F(\sigma, S) - F^*(S)\} \tag{13}$$

is the greatest deviation of σ, whose possibility of occurrence is not less than λ. From the definition of $\pi(S)$ it is easy to see that $\{S:\ \pi(S) \geq \lambda\} = [\underline{\xi}_1^{\lambda}, \overline{\xi}_1^{\lambda}] \times \cdots \times [\underline{\xi}_l^{\lambda}, \overline{\xi}_l^{\lambda}] = \Gamma^{\lambda}$. Consequently, the quantity $\overline{\delta}_{\sigma}^{\lambda}$ is the maximal regret of σ in the min-max regret version of the problem with scenario set Γ^{λ}. In particular, the condition $\overline{\delta}_{\sigma}^{\lambda} \leq 0$ means that σ is necessarily optimal under Γ^{λ}.

ISBN: 978-989-95079-6-8

Let us point out that the fuzzy problem is not simpler than the corresponding min-max regret one. Indeed, the interval uncertainty representation is a special case of the fuzzy one. If we additionally fix $\tilde{G} = (0, M)$ for a sufficiently large M, then the fuzzy problem is equivalent to the min-max regret one.

Observe that $\overline{\delta}_\sigma^\lambda$ is a nonincreasing and $\overline{g}^{1-\lambda}$ is a nondecreasing function of λ. In consequence, the problem (12) can be solved be a standard binary search technique if we only can decide somehow whether there is a feasible sequence $\sigma \in \boldsymbol{\mathcal{S}}$ fulfilling inequality $\overline{\delta}_\sigma^\lambda \leq \overline{g}^{1-\lambda}$ for a fixed $\lambda \in [0, 1]$. This task is easy if we have an algorithm for the corresponding min-max regret sequencing problem with interval parameters. Solving this problem for scenario set Γ^λ we get a min-max regret sequence σ^*. Then $\overline{\delta}_\sigma^\lambda \leq \overline{g}^{1-\lambda}$ for some $\sigma \in \boldsymbol{\mathcal{S}}$ if and only if $\overline{\delta}_{\sigma^*}^\lambda \leq \overline{g}^{1-\lambda}$. Notice that in the problem (11) it is enough to detect a necessarily optimal sequence for scenario set Γ^λ, which may be computationally easier than solving the min-max regret problem. The binary search algorithm is shown in Figure 3.

1: Find a min-max regret sequence σ under Γ^1
2: **if** $\overline{\delta}_\sigma^1 > \overline{g}^0$ **then return** $\emptyset$
3: $\lambda_1 \leftarrow 0.5$, $k \leftarrow 1$, $\lambda_2 \leftarrow 0$
4: **while** $|\lambda_1 - \lambda_2| \geq \epsilon$ **do**
5: $\quad\lambda_2 \leftarrow \lambda_1$
6: $\quad$Find a min-max regret sequence ρ under Γ^{λ_1}
7: $\quad$**if** $\overline{\delta}_\rho^{\lambda_1} \leq \overline{g}^{1-\lambda_1}$ **then**
8: $\quad\quad\lambda_1 \leftarrow \lambda_1 - 1/2^{k+1}$, $\sigma \leftarrow \rho$
9: $\quad$**else**
10: $\quad\quad\lambda_1 \leftarrow \lambda_1 + 1/2^{k+1}$
11: $\quad$**end if**
12: $\quad k \leftarrow k + 1$
13: **end while**
14: **return** σ

Figure 3: Solving the problem (10) with a given precision $\epsilon \in (0, 1)$. Algorithm returns $\emptyset$ if $\mathrm{N}(\delta_\sigma \in \tilde{G}) = 0$ for all $\sigma \in \boldsymbol{\mathcal{S}}$.

It is easy to check that if the min-max regret problem with interval parameters can be solved in $f(n)$ time, then the corresponding fuzzy problem is solvable in $O(f(n) \log \epsilon^{-1})$ time. W thus can see that the problem with fuzzy parameters boils down to solving a small number of min-max regret sequencing problems. So, every exact algorithm for the min-max regret problem can easily be adopted for the more general fuzzy case. In [6] a polynomial algorithm for the min-max regret $1|prec|L_{max}$ problem with interval processing times and interval due dates has been constructed. Consequently, the fuzzy $1|prec|L_{max}$ problem is solvable efficiently as well because we can use this algorithm as a subroutine in the binary search.

The binary search algorithm is the most general method of solving the fuzzy problem which, however, requires a given precision of calculations. The formulation (12) sometimes allows us to design an exact algorithm based on a mixed integer programming (MIP) formulation. The obtained MIP model can be then solved by using some standard packages such as CPLEX or GLPK. We will show an example in the next section.

4.3 The fuzzy problem $1||\sum C_i$

Assume that for every job $i \in J$, a fuzzy processing time with possibility distribution μ_{p_i} is given. A scenario in this problem is a particular realization of the processing times. The cost of a sequence σ under S is the total flow time, that is the sum of job completion times. Under a given processing times scenario S, an optimal sequence can be obtained in $O(n \log n)$ time by ordering the jobs with respect to nondecreasing processing times [1].

We first show that the problem of computing the most necessarily optimal sequence, i.e. the problem (11), is efficiently solvable. Let us fix $\lambda \in [0, 1]$ and consider the min-max regret $1||\sum C_i$ problem with interval processing times specified by scenario set $\Gamma^\lambda = [\underline{p}_i^\lambda, \overline{p}_i^\lambda] \times \cdots \times [\underline{p}_n^\lambda, \overline{p}_n^\lambda]$. Consider a scenario $S \in \Gamma^\lambda$ such that under this scenario the processing times are $\frac{1}{2}(\underline{p}_i^\lambda + \overline{p}_i^\lambda)$ for $i \in J$. We can compute in $O(n \log n)$ time and optimal sequence ρ under S. It turns out that $\overline{\delta}_\rho^\lambda \leq 2\overline{\delta}_\sigma^\lambda$ for all sequences σ (see [20]). In consequence, if there is a necessarily optimal sequence σ such that $\overline{\delta}_\sigma^\lambda = 0$, then ρ must also be necessarily optimal. We thus have an efficient method of detecting a necessarily optimal sequence and the problem (11) can be solved in $O(n \log n \log \epsilon^{-1})$ time by using the binary search shown in Figure 3.

We now focus on the more general problem (10). Unfortunately, this problem is NP-hard because the corresponding min-max regret problem with interval processing times is NP-hard [10]. Using the formulation (12) we will show that it is possible to design a mixed integer linear programming model to solve the fuzzy problem. Let $\mathcal{A}$ be the set of all binary vectors (x_{ij}), $i = 1, \ldots, n$, $j = 1 \ldots, n$, fulfilling the so called *assignment constraints*, that is $\sum_{i=1}^n x_{ij} = 1$ for all $j = 1, \ldots, n$ and $\sum_{j=1}^n x_{ij} = 1$ for all $i = 1, \ldots, n$. A vector $(x_{ij}) \in \mathcal{A}$ represents a sequence σ in which $x_{ij} = 1$ if job $i \in J$ occupies position j in σ. Obviously, there is one to one correspondence between the sequences of the set of jobs J and the vectors in $\mathcal{A}$ (recall that there are no precedence constraints in J). If $(x_{ij}) \in \mathcal{A}$ corresponds to sequence σ, then the maximal regret of σ under scenario set Γ^λ can be computed in the following way [3, 11]:

$$\overline{\delta}_\sigma^\lambda = \max_{(z_{ij}) \in \mathcal{A}} \sum_{i=1}^n \sum_{j=1}^n c_{ij}^\lambda z_{ij}, \tag{14}$$

where

$$c_{ij}^\lambda = \overline{p}_i^\lambda \sum_{k=1}^{j} (j-k) x_{ik} + \underline{p}_i^\lambda \sum_{k=j+1}^{n} (j-k) x_{ik}. \tag{15}$$

Observe that (14) is an assignment problem. We can construct the dual to (14) and it is well known that this dual has the same optimal objective function value as (14). So, it holds:

$$\begin{aligned} \overline{\delta}_\sigma^\lambda = \quad & \min \sum_{i=1}^n \alpha_i + \sum_{i=1}^n \beta_i \\ & \alpha_i + \beta_j \geq c_{ij}^\lambda \text{ for } i, j = 1, \ldots, n \end{aligned} \tag{16}$$

where α_i and β_i, $i = 1, \ldots, n$, are unrestricted dual variables associated with the assignment constraints. Now, using formulation (12), we can represent the fuzzy problem in the fol-

ISBN: 978-989-95079-6-8

lowing way:

$$\begin{array}{l} \min \lambda \\ \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \beta_i \leq \overline{g}^{1-\lambda} \\ \alpha_i + \beta_j \geq c_{ij}^{\lambda} \text{ for } i,j = 1,\ldots,n \\ (x_{ij}) \in \mathcal{A} \\ \lambda \in [0,1] \end{array} \tag{17}$$

where c_{ij}^{λ} are given as (15). If trapezoidal fuzzy intervals $(\underline{p}_i, \overline{p}_i, \alpha_i, \beta_i)$, $i = 1,\ldots,n$, are used to model the processing times, then we can substitute $\overline{p}_i^{\lambda} = \overline{p}_i + (1-\lambda)\beta_i$ and $\underline{p}_i^{\lambda} = \underline{p}_i - (1-\lambda)\alpha_i$ in (15). The resulting model will be still not linear because some expressions of the form λx_{ij} will appear. However, we can easily linearize the model by substituting $t_{ij} = \lambda x_{ij}$ and adding additional constraints $t_{ij} - x_{ij} \leq 0$, $\lambda - t_{ij} + x_{ij} \leq 0$, $-\lambda + t_{ij} \leq 0$, $t_{ij} \geq 0$ for all $i,j = 1,\ldots,n$. Hence the resulting model is a mixed integer linear one and can be solved by using an available software. We refer the reader to [11] for some techniques that allow us to refine the formulation and speed up the computations.

5 Conclusions

In this paper we have adopted a general approach to sequencing problems with fuzzy parameters. We have applied possibility theory to model the imprecise parameters and to define a solution concept. We have extended the approach used in [19] and [20] to another class of problems. As in many fuzzy problems, the main computational difficultly is in the classical interval case. Namely, every exact algorithm that computes a min-max regret sequence under interval data can be easily adopted to solve the fuzzy problem. One can use a binary search method or try to design a mixed integer programming model.

Unfortunately, the min-max regret sequencing problems with interval data are mostly hard to solve. There are only few problems that are known to be polynomially solvable and the complexity of some important problems, such as the permutation flow shop on two machines, remains open. Contrary to the class of problems discussed in [20], there is also lack of general properties of the min-max regret sequencing problems. Specifically, there is no general method of computing the maximal regret of a given sequence and there are no general relationships between the optimality evaluation and the min-max regret sequences. In other words, every sequencing problem possesses its own properties and the area of research in this field is still not fully explored.

Acknowledgment

The second author was partially supported by Polish Committee for Scientific Research, grant N N111 1464 33.

References

[1] P. Brucker. *Scheduling Algorithms, 5th edition*. Springer Verlag, Hedielberg, 2007.

[2] X. Wu and X. Zhou. Stochastic scheduling to minimize expected maximum lateness. *European Journal of Operational Research*, 190(1):103–115, 2008.

[3] P. Kouvelis and G. Yu. *Robust discrete optimization and its applications*. Kluwer Academic Publishers, Boston, 1997.

[4] R. D. Luce and H. Raiffa. *Games and decisions. Introduction and critical survey*. Dover Publications, INC., New York, 1957.

[5] R.L. Daniels and P. Kouvelis. Robust scheduling to hedge against processing time uncertainty in single-stage production. *Management Science*, 41(2):363–376, 1995.

[6] A. Kasperski. Minimizing maximal regret in the single machine sequencing problem with maximum lateness criterion. *Oper. Res. Let.*, 33(4):431–436, 2005.

[7] P. Kouvelis, R.L. Daniels, and G. Vairaktarakis. Robust scheduling of a two-machine flow shop with uncertain processing times. *IIE Transactions*, 32:421–432, 2000.

[8] I. Averbakh. Minmax regret solutions for minimax optimization problems with uncertainty. *Oper. Res. Let.*, 27(2):57–65, 2001.

[9] I. Averbakh. The minmax regret permutation flow-shop problem with two jobs. *European Journal of Operational Research*, 169:761–766, 2006.

[10] V. Lebedev and I. Averbakh. Complexity of minimizing the total flow time with interval data and minmax regret criterion. *Discrete Applied Mathematics*, 154(15):2167–2177, 2006.

[11] R. Montemanni. A mixed integer programming formulation for the total flow time single machine robust scheduling problem with interval data. *Journal of Mathematical Modelling and Algorithms*, 6(2):287–296, 2007.

[12] A. Kasperski and P. Zieliński. A 2-approximation algorithm for interval data minmax regret sequencing problems with the total flow time criterion. *Oper. Res. Let.*, 36:343–344, 2008.

[13] M. Inuiguchi M. and J. Ramik. Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem. *Fuzzy Sets and Systems*, 111(11):3–28, 2000.

[14] M. Vlach. Single machine scheduling under fuzziness. In R. Słowiński and M. Hapke, editors, *Scheduling under Fuzziness*, volume 37 of *Studies in Fuzziness and Soft Computing*, pages 223–245. Physica–Verlag, 2000.

[15] H. Ishii, M. Tada, and T. Masuda. Scheduling problem with fuzzy due dates. *Fuzzy Sets and Systems*, 46:339–347, 1992.

[16] S. Han, H. Ishii, and S. Fuji. One machine scheduling problem with fuzzy due dates. *European Journal of Operational Research*, 79:1–12, 1994.

[17] A. Kasperski. A possibilistic approach to sequencing problems with fuzzy parameters. *Fuzzy Sets and Systems*, 150(1):77–86, 2005.

[18] S. Chanas and A. Kasperski. Possible and necessary optimality of solutions in the single machine scheduling problem with fuzzy parameters. *Fuzzy Sets and Systems*, 142:359–371, 2004.

[19] M. Inuiguchi and M. Sakawa. Robust optimization under softness in a fuzzy linear programming problem. *International Journal of Approximate Reasonning*, 18:21–34, 1998.

[20] A. Kasperski and P. Zieliński. On possibilistic combinatorial optimization problems. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 324–329, Hong Kong, 2008.

[21] A. Kasperski. *Discrete optimization with interval data. Minmax regret and fuzzy approach*, volume 228 of *Studies in Fuzziness and Soft Computing*. Springer, Berlin Heidelberg, 2008.

[22] E. Conde. An improved algorithm for selecting p items with uncertain returns according to the minmax regret criterion. *Mathematical Programming*, 100:345–353, 2004.

[23] D. Dubois and H. Prade. *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, New York, 1988.

ISBN: 978-989-95079-6-8

The Bandler-Kohout Subproduct as a Suitable Inference Mechanism

Martin Štěpnička[1] Balasubramaniam Jayaram[2]

1. Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava 701 33, Czech Republic
2. Department of Mathematics, Indian Institute of Technology Madras
Chennai 600 036, India
Email: martin.stepnicka@osu.cz, jbala@iitm.ac.in,

***Abstract*— *The compositional rule of inference (CRI) introduced by Zadeh is widely used in approximate reasoning schemes using fuzzy sets. In this work we show that the Bandler-Kohout subproduct does possess all the important properties such as equivalent and reasonable conditions for their solvability, their interpolative properties and the preservation of the indistinguishability that may be inherent in the input fuzzy sets, as cited in favor of using CRI.*

***Keywords*—** Bandler-Kohout subproduct, compositional rule of inference, correctness and continuity of inference, fuzzy relation equations.

1 Introduction

Fuzzy systems are one of the best known applications of fuzzy logic. These are usually based on a set of fuzzy rules. Systems using fuzzy rules have been applied in a wide variety of applications, viz., automatic control, decision making, risk analysis, etc.

If X is a classical set, we denote the set of all fuzzy sets on X by $\mathcal{F}(X)$. Given two non-empty classical sets X, Y, a fuzzy rule is usually given in the following form:

$$\textbf{IF } x \text{ is } \mathcal{A} \textbf{ THEN } y \text{ is } \mathcal{B} \,, \tag{1}$$

where $\mathcal{A}, \mathcal{B}$ are membership predicates represented by fuzzy sets $A \in \mathcal{F}(X)$ and $B \in \mathcal{F}(Y)$, respectively. Given a fuzzy observation $A' \in \mathcal{F}(X)$ a corresponding output fuzzy set $B' \in \mathcal{F}(Y)$ is deduced using an inference mechanism.

1.1 Fuzzy Relational Inference

Fuzzy inference systems can be broadly classified as those inference mechanisms that are fuzzy relation based, i.e., those that interpret a given fuzzy rule base as a fuzzy relation, and those that are not. Similarity Based Reasoning (SBR) (see, for instance, TÜRKSEN [1]) and Inverse Truth-functional Modification of BALDWIN [2] are two of the representative examples of the latter type of inference mechanisms that do not use fuzzy relations and which are well established in the literature.

In this work, we focus only on those inference mechanisms that are fuzzy relation based. However, it should be mentioned that, under certain conditions, an equivalent fuzzy relation based description of some of these inference mechanisms can be given (see [3, 4]).

1.2 Compositional Rule of Inference

The earliest inference scheme, proposed by Zadeh himself and well established in the literature, is the Compositional Rule of Inference (CRI) [5]. In CRI, a fuzzy rule is interpreted by a fuzzy relation $R \in \mathcal{F}(X \times Y)$. Given a fuzzy observation $A' \in \mathcal{F}(X)$ an output fuzzy set $B' \in \mathcal{F}(Y)$ is deduced by the CRI based on the fuzzy relation R. From the fuzzy relational point of view the CRI is defined as follows:

$$B'(y) = \bigvee_{x \in X} (A'(x) * R(x, y)), \quad y \in Y \,, \tag{2}$$

where $*$ is a fuzzy conjunction, typically a t-norm (see KLEMENT et al. [6] for more details). We use the following notation to indicate the CRI scheme:

$$B' = A' \circ R \,. \tag{3}$$

1.3 Popularity of CRI

The CRI mechanism is most often and widely used for the following reasons:

(i) An important issue in the applicability of CRI scheme is to determine an appropriate fuzzy relation R based on the given fuzzy rules so as to obtain meaningful conclusions. This pertains to the solvability of the fuzzy relation equations. A necessary and sufficient condition for the solvability of fuzzy relation equations has been well established for long, see DI NOLA et al [7]. Moreover, recently, PERFILIEVA and LEHMKE [8] showed its solvability under conditions that are more favourable and prevalent in many applicational contexts.

(ii) In PERFILIEVA et al. [8, 9] the authors have dealt with the continuity of a fuzzy function adjoint to the CRI mechanism and a fuzzy relation interpreting fuzzy rules (8). The authors, defining continuity suitably, have shown that it is equivalent to the correctness of the model under consideration.

(iii) KLAWONN and CASTRO [10] have proven two important and interesting results about the CRI scheme using fuzzy rules and the indistinguishability inherent to the fuzzy sets considered. They showed the robustness of the CRI mechanism in scenarios where there can be slight discrepancies between the intended input and the actual input. They have also shown that the indistinguishability induced by the fuzzy set representing the linguistic expression in the premise of the rule cannot be overcome.

However, it should be noted that there are still some drawbacks that are usually cited about CRI. Recently, JAYARAM

ISBN: 978-989-95079-6-8

[11] has proposed a modified form of CRI called the Hierarchical CRI that addresses many of these drawbacks. The above properties will be dealt with in more detail in Sec. 3.

1.4 Motivation for this work: Bandler-Kohout Subproduct

Generally, the CRI - represented by $\circ$ - may be replaced by any appropriate image of a fuzzy set under the fuzzy relation R denoted by @, which fulfills some required properties. Besides the most often used CRI inference scheme, another scheme of inference is the Bandler-Kohout subproduct $\lhd$ (BK-Subproduct, for short) proposed by BANDLER and KOHOUT [12]. Considering the fuzzy rule as given in (1), for a given input $A' \in \mathcal{F}(X)$ the inference obtained using the BK-Subproduct is given as follows:

$$B'(y) = \bigwedge_{x \in X} (A'(x) \to R(x,y)), \quad y \in Y \ , \tag{4}$$

where $\to$ is a fuzzy implication (see Sec. 2 for more details), $B' \in \mathcal{F}(Y)$ and R is the fuzzy relation that interprets the fuzzy rule (1).

Later on, PEDRYCZ [13] has shown certain solvability conditions for the BK-Subproduct inference as given in (4). Buoyed by this important result on the BK-Subproduct, this paper is an attempt at studying the suitability of the BK-Subproduct as an inference mechanism, by showing that the BK-Subproduct does have all the above desirable properties possessed by CRI.

1.5 Organisation of the work

In Section 2 we recall the preliminaries required for the rest of the paper. Also the structure and inference in the Compositional Rule of Inference (CRI) mechanism is discussed. Following this, in Section 3 we discuss some of the important properties cited in favor of using CRI,e.g., equivalent and reasonable conditions for their solvability, their interpolative properties and the preservation of the indistinguishability that may be inherent in the input fuzzy sets. Section 4, which contains our main work, shows that all the above mentioned properties are valid even in the case of the BK-Subproduct.

2 Preliminaries

2.1 Fuzzy Inference Mechanisms

An inference mechanism acting upon fuzzy rules is usually based on fuzzy logic connectives. In this work, we restrict ourselves to inference schemes based on fuzzy logic connectives. Let us fix a complete residuated lattice $\mathcal{L} = ([0,1], \wedge, \vee, *, \to, 0, 1)$ (see, for example, NOVÁK et al. [14]) as the basic algebraic structure for the whole paper.

Since the set $[0,1]$ is totally ordered, $\mathcal{L}$ becomes an MTL-algebra and hence, in our context, $*$ becomes a left-continuous t-norm and $\to$ is the residual implication *uniquely* given as follows:

$$a \to b = \bigvee \{t \in \mathcal{L} | a * t \leq b\} \ , \tag{5}$$

which is also a fuzzy implication, for more details we refer the readers to [6, 14, 15].

From the two operations $\wedge, \to$ of $\mathcal{L}$ we can derive yet another operation known as the *biresiduum* which is defined as follows:

$$a \leftrightarrow b = (a \to b) \wedge (b \to a), \qquad a, b \in \mathcal{L} \ . \tag{6}$$

Finally, by an extension of an algebraic operation on $\mathcal{L}$ to an operation between fuzzy sets we mean the following:

$$(C \star D)(u,v) = C(u) \star D(v), \qquad u \in U, v \in V \ , \tag{7}$$

where $\star \in \{\wedge, \vee, *, \to, \leftrightarrow\}$ and where C, D are arbitrary fuzzy sets on arbitrary universes U, V, respectively.

2.2 Structure and Inference in CRI

The CRI inference mechanism given in (2) is for a fuzzy relation $R \in \mathcal{F}(X \times Y)$ that interprets a single fuzzy rule. However, usually multiple fuzzy rules are a necessity. Let us denote the multiple fuzzy rules as follows:

$$\textbf{IF } x \text{ is } \mathcal{A}_i \quad \textbf{THEN} \quad y \text{ is } \mathcal{B}_i \ , \tag{8}$$

where $\mathcal{A}_i, \mathcal{B}_i$ are membership predicates represented by fuzzy sets $A_i \in \mathcal{F}(X)$ and $B_i \in \mathcal{F}(Y)$, respectively, for $i = 1, \ldots, n$.

Then a fuzzy relation interpreting the whole *fuzzy rule base* (altern. *linguistic description*), composed of the n fuzzy rules (8), has to be constructed. basically, there are two main approaches to the construction of such a fuzzy relation.

The following fuzzy relation $\check{R} \in \mathcal{F}(X \times Y)$

$$\check{R}(x,y) = \bigvee_{i=1}^{n} (A_i(x) * B_i(y)) \ , \tag{9}$$

is used in most of the real world applications. This is mainly due to the successful applications of this, say Cartesian product approach, published by Mamdani and Assilian in [16], which was followed by a huge number of researchers and practitioners, see e.g. [17, 18].

Alternatively, to keep the conditional IF-THEN form of the fuzzy rules (8) fuzzy relation $\hat{R} \in \mathcal{F}(X \times Y)$ given as follows

$$\hat{R}(x,y) = \bigwedge_{i=1}^{n} (A_i(x) \to B_i(y)) \ , \tag{10}$$

can be chosen to interpret the fuzzy rule base. It deals with a mathematically correct extension of a classical implication.

To stress the difference between both the approaches, let us recall the work of DUBOIS et al. [19], where the authors state: "*In the view given by* (10)*, each piece of information (fuzzy rule) is viewed as a constraint. This view naturally leads to a conjunctive way of merging the individual pieces of information since the more information, the more constraints and the less possible values to satisfy them*." While the same authors describe the second approach proposed by Mamdani and Assilian as follows: "*It seems that fuzzy rules modelled by* $\check{R}$ *are not viewed as constraints but are considered as pieces of data. Then the maximum in (9) expresses accumulation of data*".

In other words, instead of interpreting fuzzy rules as logical implications, the approach given using $\check{R}$ builds up an input-output relation from smaller units, and those units are examples of fuzzy input-output pairs. And then Mamdani-Assilian approach is a method that interpolates between known input-output pairs [20]. For an extensive study of different fuzzy rules we refer to [21, 22].

ISBN: 978-989-95079-6-8

It should be stressed that both approaches have sound logical foundations but from different viewpoints, see e.g. [23, 24, 14]. However, only the approach using $\check{R}$ was widely used in applications although the implicational approach using $\hat{R}$ is probably as useful as the Mamdani-Assilian one, see [25]. Nevertheless, as we show in Section 4, the implicational approach using $\hat{R}$ does have an important role to play in the case of BK-Subproducts (see Theorem 4.7).

3 Some desirable properties of CRI

Fuzzy rules may be viewed as a partial mapping from $\mathcal{F}(X)$ to $\mathcal{F}(Y)$ assigning $B_i \in \mathcal{F}(Y)$ to $A_i \in \mathcal{F}(X)$ for every $i = 1, \ldots, n$. The inference process then may be viewed as an extension of this partial mapping to a total one [26]. For better understanding, let us adopt the notation from [8] and consider the following structure

$$S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, @) \ ,$$

where $@ : \mathcal{F}(X) \times \mathcal{F}(X \times Y) \to \mathcal{F}(Y)$ is an image of a fuzzy set under a fuzzy relation. For instance, $@$ could be one of $\circ$ or $\triangleleft$. Now, by the choice of the fuzzy relation R interpreting the fuzzy rule base and by the choice of $@$, we define a fuzzy function $f_R(A) : \mathcal{F}(X) \to \mathcal{F}(Y)$ such that $f_R(A) = A@R$, for arbitrary $A \in \mathcal{F}(X)$.

3.1 Solvability of systems of fuzzy relation equations with the $\circ$ image

One of the fundamental properties of such a mapping is its interpolativity, i.e., $f_R(A_i) = B_i$. In this case, we say that R is a *correct model* of given fuzzy rules in the given structure S [8].

This leads us to deal with a system of fuzzy relation equations [27] where generally the system of equations

$$A_i@R = B_i, \quad i = 1, \ldots, n \ , \tag{11}$$

is solved with respect to the known $A_i \in \mathcal{F}(X), B_i \in \mathcal{F}(Y)$ and unknown $R \in \mathcal{F}(X \times Y)$. If R is a solution to (11) then the adjoint fuzzy function fulfills $f_R(A_i) = B_i$.

In the case of CRI, the above system of equations, viz., (11), reduces to the following:

$$A_i \circ R = B_i, \quad i = 1, \ldots, n \ . \tag{12}$$

Let us recall some main results which may be found, e.g., in [7, 27, 28, 29].

Theorem 3.1 *System (12) is solvable if and only if $\hat{R}$ is a solution of the system and moreover, $\hat{R}$ is the greatest solution of (12).*

On the one hand, Theorem 3.1 states the necessary and sufficient condition of the solvability of system (12) and it determines the solution. Moreover, it ensures that the given solution is the greatest one. On the other hand, we still do not know, when $\hat{R}$ is the solution, i.e., how to ensure the solvability.

Theorem 3.2 [29] *Let A_i for $i = 1, \ldots, n$ be normal. Then $\check{R}$ is a solution of (12) if and only if the following condition*

$$\bigvee_{x \in X} (A_i(x) * A_j(x)) \leq \bigwedge_{y \in Y} (B_i(y) \leftrightarrow B_j(y)) \ , \tag{13}$$

holds for arbitrary $i, j \in \{1, \ldots, n\}$.

Theorem 3.2 specifies a sufficient condition under which the system is solvable and moreover, it ensures that not only $\hat{R}$ but also even $\check{R}$ is a solution of system (12).

It is worth mentioning that condition (13) appearing in Theorem 3.2 is from the practical point of view not very convenient. However, if the antecedent fuzzy sets form the so called $*$-semi-partition it forces fulfilling the discussed condition in advance, see DE BAETS and MESIAR [30] for more details.

Another sufficient condition for solvability of the systems with a high practical importance was published in [31].

Theorem 3.3 (Theorem 7 in [31]) *Let A_i for $i = 1, \ldots, n$ be normal and fulfill the Ruspini condition*

$$\sum_{i=1}^{n} A_i(x) = 1, \quad x \in X. \tag{14}$$

Then the system (12) is solvable.

3.2 Continuity of a model of fuzzy rules

In [8, 9] the authors have dealt with the continuity of a fuzzy function adjoint to the CRI mechanism and a fuzzy relation interpreting fuzzy rules (8). They have defined continuity suitably and have shown that it is equivalent to the correctness of the model under consideration.

Although the original definition in [8] of a continuous model was given for the particular inference mechanism CRI, i.e., for $@ \equiv \circ$, the particular image plays absolutely no role in the proof of Theorem 3.6 explaining the nature of the definiton and hence can be generalized for an arbitrary image of a fuzzy set under a fuzzy relation.

Definition 3.4 A fuzzy relation $R \in \mathcal{F}(X \times Y)$ is said to be a *continuous model of fuzzy rules* (8) *in a structure* $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, @)$ if for each $i \in I$ and for each $A \in \mathcal{F}(X)$ the following inequality holds:

$$\bigwedge_{y \in Y} (B_i(y) \leftrightarrow (A@R)(y)) \geq \bigwedge_{x \in X} (A_i(x) \leftrightarrow A(x)) \ . \tag{15}$$

Inequality (15) can be rewritten in terms of the adjoint fuzzy function f_R as follows:

$$\bigwedge_{y \in Y} (B_i(y) \leftrightarrow (f_R(A))(y)) \geq \bigwedge_{x \in X} (A_i(x) \leftrightarrow A(x)). \tag{16}$$

Remark 3.5 Let us explain why formula (15) expresses the continuity. The closeness between fuzzy sets is measured by the biresiduation operation $\leftrightarrow$, i.e. it is a dual information to the metric one. Let us consider a continuous Archimedean t-norm $*$ with an additive generator $g : [0, 1] \to [0, +\infty]$. Then the biresiduum may be written in the form

$$a \leftrightarrow b = g^{-1}(|g(a) - g(b)|) \tag{17}$$

where $g^{-1} : [0, \infty] \to [0, 1]$ is the inverse function and where in the case of $g(0) = \infty$ we define $g(0) - g(0) = 0$. Now, for an arbitrary non-empty universe X it is possible to define a metric D_g on $\mathcal{F}(X)$ generated by g as follows:

$$D_g(A, B) = \bigvee_{x \in X} |g(A(x)) - g(B(x))|. \tag{18}$$

The following theorem justifies the use of the notion of continuity in Definition 3.4. For more details, see PERFILIEVA and LEHMKE [8].

ISBN: 978-989-95079-6-8

Theorem 3.6 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, @)$ be a structure for fuzzy rules* (8) *such that $\mathcal{L}$ be a residuated lattice on $[0,1]$ with a continuous Archimedean t-norm $*$ having a continuous additive generator g. A fuzzy relation $R \in \mathcal{F}(X \times Y)$ is a continuous model of the fuzzy rules in the given structure S if and only if*

$$D_g(B_i, (A@R)) \leq D_g(A_i, A), \quad i = 1, \ldots, n \tag{19}$$

for each fuzzy set $A \in \mathcal{F}(X)$.

The following lemma was crucial for further results published in [8].

Lemma 3.7 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \circ)$ be a structure for fuzzy rules* (8) *and let $R \in \mathcal{F}(X \times Y)$. Then for any $A \in \mathcal{F}(X)$ and all $i = 1, \ldots, n$ and $y \in Y$ it is true that*

$$B_i(y) \leftrightarrow (A \circ R)(y) \geq \delta_{R,i}(y) * \bigwedge_{x \in X} (A_i(x) \leftrightarrow A(x)) , \tag{20}$$

where $\delta_{R,i}(y) = B_i(y) \leftrightarrow (A_i \circ R)(y)$.

Finally, we recall here the main result in PERFILIEVA et al. [8, 9] concerning the relationship of the above mentioned continuity and the interpolativity for the CRI.

Theorem 3.8 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \circ)$ be a structure for fuzzy rules* (8)*. A fuzzy relation $R \in \mathcal{F}(X \times Y)$ is a correct model of fuzzy rules* (8) *in the given structure S if and only if it is a continuous model of these rules in S.*

3.3 CRI and the Indistinguishability of Premises

Let X be a classical set and let $\sim$ be an equivalence relation defined on X, i.e., $\sim$ is reflexive, symmetric and transitive. Immediately, $\sim$ partitions X into equivalence classes. It is well-known then that an $M \subseteq X$ belongs to this partition if, and only if, whenever $x \in M$ and $x \sim y$ for some $y \in X$ then $y \in M$. In a sense, the elements of M are indistinguishable and can be represented mathematically as follows:

$$x \in M \text{ and } x \sim y \text{ implies } y \in M .$$

A similar relation between fuzzy equivalence relations and fuzzy sets on X was introduced by KLAWONN and CASTRO [10]. Once again the operation $*$ comes from the residuated lattice $\mathcal{L}$.

Definition 3.9 A fuzzy subset E of the Cartesian product X^2 is called a *fuzzy equivalence* relation on X if the following properties are satisfied for all $x, y, z \in X$:

$$\begin{aligned} &\text{(Reflexivity)} && E(x,x) = 1, && \text{(ER)} \\ &\text{(Symmetry)} && E(x,y) = E(y,x), && \text{(ES)} \\ &\text{(Transitivity)} && E(x,z) \geq E(x,y) * E(y,z) . && \text{(ET)} \end{aligned}$$

Definition 3.10 A fuzzy set $\mu \in \mathcal{F}(X)$ is called *extensional* with respect to a fuzzy equivalence relation E on X if

$$\mu(x) * E(x,y) \leq \mu(y) , \qquad x, y \in X . \tag{21}$$

If a fuzzy set μ is not extensional with respect to the considered fuzzy equivalence relation E, instead one considers the smallest fuzzy set that is extensional with respect to E and contains μ.

Definition 3.11 Let $\mu \in \mathcal{F}(X)$ and let E be a fuzzy equivalence relation on X. The fuzzy set

$$\widetilde{\mu}(x) = \bigwedge \{\nu \mid \mu \leq \nu \text{ and } \nu \text{ is extensional with respect to } E\} , \tag{22}$$

is called the *extensional hull* of μ. Note that by $\mu \leq \nu$ we mean that for all $x \in X$, $\mu(x) \leq \nu(x)$, i.e., we mean ordering in the sense of inclusion, not in the sense of ordering fuzzy quantities.

Proposition 3.12 [10] *Let $\mu \in \mathcal{F}(X)$ and let E be a fuzzy equivalence relation on X.*

(i) $\widetilde{\mu}(x) = \bigvee \{\mu(y) * E(x,y) \mid y \in X\}$,

(ii) $\widetilde{\mu}$ *is extensional with respect to E,*

(iii) $\widetilde{\widetilde{\mu}} = \widetilde{\mu}$.

KLAWONN and CASTRO [10] have proven the following two important and interesting results about the CRI scheme using fuzzy rules and the indistinguishability inherent to the fuzzy sets considered.

Theorem 3.13 *Let $S = (X, Y, \{A, B\}, \mathcal{L}, \circ)$ be a structure for fuzzy rule* (1)*. Let E be a fuzzy equivalence relation on X with respect to which A is extensional. Let $A' \in \mathcal{F}(X)$ be any fuzzy set, then*

$$A' \circ \hat{R} = \widetilde{A'} \circ \hat{R},$$
$$A' \circ \check{R} = \widetilde{A'} \circ \check{R}.$$

The following interpretation of the above result is given in [10]: *The output obtained from CRI for a given fuzzy rule and an input fuzzy set A' does not change if we substitute A' by its extensional hull $\widetilde{A'}$. The indistinguishability inherent in the fuzzy set A cannot be avoided even if the input fuzzy set A' stands for a crisp value*. Further, *a fuzzified input does not change the outcome of a rule as long as the fuzzy set obtained by the fuzzification is contained in the extensional hull of the original crisp input value.* They finally conclude that *it does not make sense to measure more exactly than the indistinguishability admits.*

In other words, this shows the robustness of the inference in scenarios where there can be slight discrepancies between the intended input and the actual input.

It is immediate now, as already observed in [10], that *the indistinguishability induced by the fuzzy set representing the linguistic expression in the premise of the rule cannot be overcome.*

Let us stress that the proof of Theorem 3.13 may be easily modified in order to capture the situation with n fuzzy rules. It can be observed that only the Mamdani-Assilian approach $\check{R}$ generally works in the combination with the CRI.

Theorem 3.14 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \circ)$ be a structure for fuzzy rules (8). Let E be a fuzzy equivalence relation on X with respect to which A_i is extensional for arbitrary $i = 1, \ldots, n$. Let $A' \in \mathcal{F}(X)$ be any fuzzy set, then*

$$A' \circ \check{R} = \widetilde{A'} \circ \check{R}.$$

ISBN: 978-989-95079-6-8

4 BK-Subproduct and some desirable properties

4.1 Solvability of systems of fuzzy relation equations with the $\triangleleft$ image

In the case of BK-Subproduct, the system of equations (11), reduces to the following:

$$A_i \triangleleft R = B_i, \quad i = 1, \ldots, n \, . \tag{23}$$

Concerning system (23), let us recall the following two basic theorems .

Theorem 4.1 [13] *System (23) is solvable if and only if $\check{R}$ is a solution of the system and moreover, $\check{R}$ is the least solution of system (23).*

Theorem 4.2 [32] *Let A_i for $i = 1, \ldots, n$ be normal. Then $\hat{R}$ is a solution of (23) if and only if the condition* (13) *holds for arbitrary $i, j \in \{1, \ldots, n\}$.*

Again, condition (13) to which Theorem 4.2 refers to, is not very convenient from a practical point of view and may be a priori fulfilled by using antecedent fuzzy sets which form the $*$-semi-partition.

Fortunately, that the sufficient condition for solvability of the systems with a high practical importance stated in Theorem 3.3 is valid even for system (23).

Theorem 4.3 (Theorem 7 in [31]) *Let A_i for $i = 1, \ldots, n$ be normal and fulfill the Ruspini condition (14). Then the system (23) is solvable.*

4.2 Continuity of BK-Subproduct

The BK-Subproduct, unlike the CRI, was not motivated by approximate reasoning, see BANDLER and KOHOUT [12]. However, as mentioned in GOTTWALD [33], it is not necessary to insist on purely logical foundations for an inference mechanism, and it can simply be a mapping from $\mathcal{F}(X)$ to $\mathcal{F}(Y)$ fulfilling certain properties. Let us also recall that it was PEDRYCZ [13] who firstly proposed the BK-Subproduct as an inference scheme.

We may again state an analogous lemma to Lemma 3.7 also for the case of the BK-Subproduct.

Lemma 4.4 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \triangleleft)$ be a structure for fuzzy rules* (8) *and let $R \in \mathcal{F}(X \times Y)$. Then for any $A \in \mathcal{F}(X)$ and all $i = 1, \ldots, n$ and $y \in Y$ it is true that*

$$B_i(y) \leftrightarrow (A \triangleleft R)(y) \geq \delta_{R,i}(y) * \bigwedge_{x \in X} (A_i(x) \leftrightarrow A(x)) \, , \tag{24}$$

where $\delta_{R,i}(y) = B_i(y) \leftrightarrow (A_i \triangleleft R)(y)$.

Due to Lemma 4.4, we may state the following analogous theorem to Theorem 3.8 which again shows that the BK subproduct as an inference mechanism carries the same property as the CRI.

Theorem 4.5 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \triangleleft)$ be a structure for fuzzy rules* (8)*. A fuzzy relation $R \in \mathcal{F}(X \times Y)$ is a correct model of fuzzy rules* (8) *in the given structure S if and only if it is a continuous model of these rules in S.*

4.3 BK-Subproduct and the Indistinguishability of Premises

In this subsection, we show the robustness of the BK-Subproduct inference mechanism along similar lines as KLAWONN and CASTRO [10]. Once again the employed operations come from the residuated lattice $\mathcal{L}$. Firstly note that if a fuzzy set $\mu \in \mathcal{F}(X)$ is extensional with respect to a fuzzy equivalence relation E on X then

$$E(x, y) \to \mu(y) \geq \mu(x), \; x, y \in X \, . \tag{25}$$

Proposition 4.6 *Let $\mu \in \mathcal{F}(X)$ and E a fuzzy equivalence relation on X. Then*

$$\widetilde{\mu}(x) = \bigwedge \{E(x, y) \to \mu(y) \mid y \in X\} \, . \tag{26}$$

Now, we present the main result analogous to Theorem 3.13.

Theorem 4.7 *Let $S = (X, Y, \{A, B\}, \mathcal{L}, \triangleleft)$ be a structure for fuzzy rule* (1)*. Let E be a fuzzy equivalence relation on X with respect to which A is extensional. Let $A' \in \mathcal{F}(X)$ be any fuzzy set, then*

$$A' \triangleleft \hat{R} = \widetilde{A'} \triangleleft \hat{R},$$
$$A' \triangleleft \check{R} = \widetilde{A'} \triangleleft \check{R}.$$

The above result, as already noted in the case of CRI, shows the robustness of the BK-Subproduct inference in scenarios where there can be slight discrepancies between the intended input and the actual input and reinforces the fact that even in the case of BK-Subproduct the indistinguishability induced by the fuzzy set representing the linguistic expression in the premise of the rule cannot be overcome.

And again as in case of the CRI, we may generalize the result concerning the indistinguishability of the premises for an arbitrary finite number of rules. Note that in the case of the BK-Subproduct the $\hat{R}$ is plays the main role.

Theorem 4.8 *Let $S = (X, Y, \{A_i, B_i\}_{i=1,\ldots,n}, \mathcal{L}, \triangleleft)$ be a structure for fuzzy rules (8). Let E be a fuzzy equivalence relation on X with respect to which each A_i is extensional, for arbitrary $i = 1, \ldots, n$. Let $A' \in \mathcal{F}(X)$ be any fuzzy set, then*

$$A' \triangleleft \hat{R} = \widetilde{A'} \triangleleft \hat{R}.$$

5 Conclusions

In this work, after recalling some of the properties that are usually cited in favor of using the Compositional Rule of Inference (CRI) introduced by ZADEH [5], viz., equivalent and reasonable conditions for their solvability, their interpolative properties and the preservation of the indistinguishability that may be inherent in the input fuzzy sets, we have shown that the Bandler-Kohout subproduct introduced in [12] does possess all the above properties and hence is equally suitable for consideration when reasoning with a system of fuzzy rules. Towards this end some new but equivalent results on indistinguishability operations has also been presented.

Acknowledgment

This work has been supported by projects MSM6198898701 of the MŠMT ČR and by IAA108270902 of the GA AV ČR.

ISBN: 978-989-95079-6-8

References

[1] I. B. Turksen and Z. Zhong. An approximate reasoning schema based on similarity measures and interval valued fuzzy sets. *IEEE Trans. Syst. Man Cybern.*, 18:1049–1056, 1988.

[2] J. F. Baldwin. A new approach to approximate reasoning using fuzzy logic. *Fuzzy Sets and Systems*, 2:309–325, 1979.

[3] V. Cross and T. Sudkamp. Patterns of fuzzy rule-based inference. *Int. J. of Approximate Reasoning*, 11:235–255, 1994.

[4] B. Bouchon-Meunier, R. Mesiar, C. Marsala, and M. Rifqi. Compositional rule of inference as an analogical scheme. *Fuzzy Sets and Systems*, 138:56–65, 2003.

[5] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Systems, Man, and Cybernetics*, SMC-3:28–44, 1973.

[6] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer, Dordrecht, 2000.

[7] Di Nola A., S. Sessa, W. Pedrycz, and E. Sanchez. *Fuzzy Relation Equations and Their Applications to Knowledge Engineering*. Kluwer, Boston, 1989.

[8] I. Perfilieva and S. Lehmke. Correct models of fuzzy if-then rules are continuous. *Fuzzy Sets and Systems*, 157:3188–3197, 2006.

[9] I. Perfilieva and V. Novák. System of fuzzy relation equations as a continuous model of IF-THEN rules. *Information Sciences*, 177:3218–3227, 2007.

[10] F. Klawonn and J. L. Castro. Similarity in fuzzy reasoning. *Mathware & Soft Computing*, 2:197–228, 1995.

[11] B. Jayaram. On the law of importation $(x \wedge y) \rightarrow z \equiv (x \rightarrow (y \rightarrow z))$ in fuzzy logic. *IEEE Trans. Fuzzy Syst.*, 16:130–144, 2008.

[12] W. Bandler and L. J. Kohout. Semantics of implication operators and fuzzy relational products. *Int. J. Man-Mach. Stud.*, 12:89–116, 1979.

[13] W. Pedrycz. Applications of fuzzy relational equations for methods of reasoning in presence of fuzzy data. *Fuzzy Sets and Systems*, 16:163–175, 1985.

[14] V. Novák, I. Perfilieva, and J. Močkoř. *Mathematical Principles of Fuzzy Logic*. Kluwer, Boston, 1999.

[15] M. Baczyński and B. Jayaram. *Fuzzy Implications (Studies in Fuzziness and Soft Computing series)*. Springer-Verlag, Heidelberg, Germany, 2008.

[16] E.H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. of Man-Machine Studies*, 7:1–13, 1975.

[17] W. J. M. Kickert and E. H. Mamdani. Analysis of a fuzzy logic controller. *Fuzzy Sets and Systems*, 1:29–44, 1978.

[18] P. M. Larsen. Industrial applications of fuzzy logic control. *Int. J. Man-Mach. Stud.*, 12:3–10, 1980.

[19] D. Dubois, H. Prade, and L. Ughetto. Checking the coherence and redundancy of fuzzy knowledge bases. *IEEE Trans. Fuzzy Systems*, 5(6):398–417, 1997.

[20] F. Klawonn and V. Novák. The relation between inference and interpolation in the framework of fuzzy systems. *Fuzzy Sets and Systems*, 81(3):331–354, 1996.

[21] D. Dubois and H. Prade. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84:169–185, 1996.

[22] V. Novák and S. Lehmke. Logical structure of fuzzy IF-THEN rules. *Fuzzy Sets and Systems*, 157:2003–2029, 2006.

[23] L. Godo and P. Hájek. A note on fuzzy inference as deduction. In D. Dubois, E. P. Klement, and H. Prade, editors, *Fuzzy Sets, Logics and Reasoning about Knowledge (Applied Logic series Vol.15)*, pages 237–241. Kluwer Academic Publishers, Dordrecht, Netherlands, 1999.

[24] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer, Dordrecht, 1998.

[25] U. Bodenhofer, M. Daňková, M. Štěpnička, and V. Novák. A plea for the usefulness of the deductive interpretation of fuzzy rules in engineering applications. In *Proc. of FUZZ-IEEE'07*, pages 1567–1572, London, 2007.

[26] T. Vetterlein and M. Štěpnička. Completing fuzzy if-then rule bases by means of smoothing splines. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Syst.*, 14:235–244, 2006.

[27] E. Sanchez. Resolution of composite fuzzy relation equations. *Inform. and Control*, 30:38–48, 1976.

[28] B. De Baets. Analytical solution methods for fuzzy relation equations. In D. Dubois and H. Prade, editors, *The Handbook of Fuzzy Set Series Vol. 1*, volume 15. 2000.

[29] F. Klawonn. Fuzzy points, fuzzy relations and fuzzy functions. In V. Novák and I. Perfilieva, editors, *Discovering the World with Fuzzy Logic*, pages 431–453. Springer, Berlin, 2000.

[30] B. De Baets and R. Mesiar. T-partitions. *Fuzzy Sets and Systems*, 97:211–223, 1998.

[31] M. Štěpnička, B. De Baets, and L. Nosková. On additive and multiplicative fuzzy models. In M. Štěpnička, V. Novák, and U. Bodenhofer, editors, *New Dimensions in Fuzzy Logic and Related Technologies, Int. Conference EUSFLAT 2007, Ostrava, Proceedings*, volume 2, pages 95–102, Ostrava, 2007.

[32] L. Nosková. Systems of fuzzy relation equation with inf$\rightarrow$ composition: solvability and solutions. *J. of Electrical Eng.*, 12(s):69–72, 2005.

[33] S. Gottwald. An abstract approach toward the evaluation of fuzzy rule systems. In P.P. Wang, D. Ruan, and E.E. Kerre, editors, Fuzzy Logic. *A Spectrum of Theoretical & Practical Issues (Studies in Fuzziness and Soft Computing 215)*, pages 281–297. Springer-Verlag, 2007.

ISBN: 978-989-95079-6-8

Early Start Can Inhibit Learning: Towards A New Explanation

Olga Kosheleva

Department of Teacher Education
University in the Texas at El Paso
El Paso, TX 79968, USA
Email: olgak@utep.edu

Abstract— *The age at which we teach different topics change. If it turns out that students do not learn, say, reading by the time they should, a natural idea is to start teaching them earlier. Several decades ago, reading and writing started in the first grade, now they start at kindergarten and even earlier. At first glance, the earlier we start, the better the students will learn. Yes, they may play less, their childhood may be not as careless as it used to be, but at least they will learn better. In practice, however, this does not always work this way: early start often inhibits learning.*

In this paper, we propose a simple model that explains why this can happen, and how we can avoid such situations and enhance the student learning. In particular, we explain why, contrary to the traditional concrete-first way of teaching abstracts concepts of school mathematics, abstract-first approach is often beneficial.

Keywords— Abstract-first, early start, education, optimal education, uncertainty

1 Formulation of the Problem

1.1 Early Start: A Seemingly Natural Idea

The age at which we teach different topics change. If it turns out that students do not learn, say, reading by the time they should, a natural idea is to start teaching them earlier.

Several decades ago, reading and writing started in the first grade, now they start at kindergarten and even earlier. At first glance, the earlier we start, the better the students will learn.

1.2 Early Start: Known Side Effects

With the early start, children may play less, their childhood may be not as careless as it used to be – but a usual expectation is that with an early start, children will learn better.

1.3 Early Start: Serious Problems

In practice, however, early start does not always help: often, early start inhibits learning.

For example, according to [18],

- human infants who started learning to turn their heads to specific sounds at the age of 31 days mastered this task, on average, at the age of 71 days, while
- infants who started learning this task at birth mastered this task, on average, at the age of 128 days.

This phenomenon is not limited to human infants: according to [8], an early start in training rhesus monkeys to discriminate objects decreased their peak performance level.

Numerous examples when an early start inhibits learning are presented and discussed in [2, 3, 4, 6, 19].

1.4 Natural Questions

The empirical fact that an early start often inhibits learning leads to the following natural question: how do take this phenomenon into account when enhancing student learning?

To be able to take this phenomenon into account in the learning process, we must be able to understand this phenomenon – and ideally, understand on the quantitative level.

1.5 These Questions Are Still Largely Open

In [2, 3, 4, 6], an attempt is made to understand why early start can inhibit learning. However,

- the existing understanding is still mostly on the qualitative level, and
- even on this level, the proposed explanations are still not fully satisfactory; see, e.g., [19].

1.6 More General Questions

The above questions about the efficiency of the early start can be viewed as a particular case of more general questions: what is the best order of presenting the material, the order that leads to the best possible learning?

1.7 These More General Questions Are Often Very Important

Many empirical studies have shown that a change in the order in which different parts of the material are presented often drastically changes the learning efficiency; see, e.g., [5, 9, 10, 13, 14, 16, 21, 22].

This is not only about using common sense: sometimes, the empirical results are counter-intuitive. For example:

- it is usually assumed that most students learn mathematical concepts better if
  - they are first presented with concrete examples of these concepts, and
  - they only learn abstract ideas later on;
- however, it turns out that empirically, the abstract-first approach for presenting the material often enhances learning; see, e.g., [9, 10, 14, 21].

1.8 What We Do in This Paper

In this paper, we propose a simple model that attempts:

- to explain the negative effect of early start and,

ISBN: 978-989-95079-6-8

- more generally, to explain the reasons why a change in presentation order can drastically change the efficiency of learning.

We then show how this explanation can be used to avoid inhibition of learning – and to enhance the student learning.

2 Towards A Theoretical Explanation

2.1 Learning: A Natural Geometric Representation

To facilitate reasoning about learning, let us start with a simple geometric representation of learning.

The process of learning means that we change the state of a student:

- from a state in which the student did not know the material (or does not have the required skill)
- to a state in which the student has (some) knowledge of the required material (or has the required skill).

Let s_0 denote the original state of a student, and let S denote the set of all the states corresponding to the required knowledge or skill.

- We start with a state which is not in the set S ($s_0 \notin S$), and
- we end up in a state s which is in the set S.

On the set of all possible states, it is natural to define a metric $d(s, s')$ as the difficulty (time, effort, etc.) needed to go from state s to state s'. Our objective is to help the students learn in the easiest (fastest, etc.) way. In terms of the metric d, this means that we want to go from the original state $s_0 \notin S$ to the state $s \in S$ for which the effort $d(s_0, s)$ is the smallest possible.

In geometric terms, the smallest possible effort means the shortest possible distance. Thus, our objective is to find the state $s \in S$ which is the closest to s_0. Such closest state is called the *projection* of the original state s_0 on the set S.

2.2 Learning Complex Material: Geometric Interpretation

The above geometric description of learning as a transition from the original state s_0 to its projection on the desired set S describes learning *as a whole*. Our objective is to find out which order of presenting information is the best. Thus, our objective is to analyze the *process* of learning, i.e., learning as a multi-stage phenomenon. For this analysis, we must explicitly take into account that the material to be learned consists of several pieces.

Let S_i, $1 \le i \le n$, denote the set of states in which a student has learned the i-th part of the material. Our ultimate objective is to make sure that the student learns all the parts of the material. In terms of states, learning the i-th part of the material means belonging to the set S_i. Thus, in terms of states, our objective means that the student should end up in a state which belongs to all the sets $S_1, \ldots, S_n$ – i.e., in other words, in a state which belongs to the intersection

$$S \stackrel{\text{def}}{=} S_1 \cap \ldots \cap S_n \tag{1}$$

of the corresponding sets S_i.

In these terms, if we present the material in the order S_1, S_2, …, S_n, this means that:

- we first project the original state s_0 onto the set S_1, resulting is a state $s_1 \in S_1$ which is the closest to s_0;
- then, we project the state s_1 onto the set S_2, resulting is a state $s_2 \in S_2$ which is the closest to s_1;
- …
- at the last stage of the cycle, we project the state s_{n-1} onto the set S_n, resulting is a state $s_n \in S_n$ which is the closest to s_{n-1}.

In some cases, we end up learning all the material – i.e., in a state $s_n \in S_1 \cap \ldots \cap S_n$. However, often, by the time the students have learned S_n, they have somewhat forgotten the material that they learned in the beginning. So, it is necessary to repeat this material again (and again). Thus, starting from the state s_n, we again sequentially project onto the sets S_1, S_2, etc.

2.3 The Above Geometric Interpretation Makes Computational Sense

The above "sequential projections" algorithm is actually actively used in many applications; see, e.g., [7, 12, 20]. In the case when all the sets S_i are convex, the resulting Projections on Convex Sets (POCS) method actually guarantees (under certain reasonable conditions) that the corresponding projections converge to a point from the intersection $S_1 \cap \ldots \cap S_n$ – i.e., in our terms, that the students will eventually learn all parts of the necessary material.

In the more general non-convex case, the convergence is not always guaranteed – but the method is still efficiently used, and often converges.

2.4 The Simplest Case: Two-Part Knowledge

Let us start with the simplest case when knowledge consists of two parts. In this simplest case, there are only two options:

- The first option is that:
  - we begin by studying S_1;
  - then, we study S_2,
  - then, if needed, we study S_1 again, etc.
- The second option is that:
  - we begin by studying S_2;
  - then, we study S_1,
  - then, if needed, we study S_2 again, etc.

We want to get from the original state s_0 to the state $\widetilde{s} \in S_1 \cap S_2$ which is the closest to s_0. The effectiveness of learning is determined by how close we get to the desired set $S = S_2 \cap S_2$ in a given number of iterations.

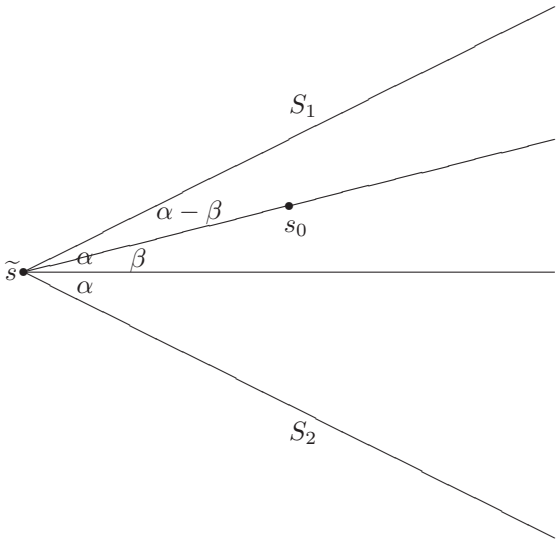
In the case of two-part knowledge, it is natural to conclude that the amount of this knowledge is reasonably small – otherwise, we would have divided into a larger number of easier-to-learn pieces.

In geometric terms, this means that the original state s_0 is close to the desired intersection set $S_1 \cap S_2$, i.e., that the distance $d_0 \stackrel{\text{def}}{=} d(s_0, \widetilde{s})$ is reasonably small.

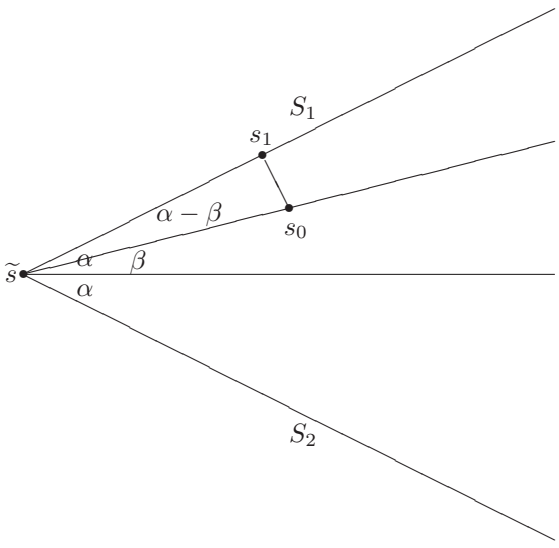
ISBN: 978-989-95079-6-8

Since all the states are close to each other, in the vicinity of the state $\widetilde{s}$, we can therefore expand the formulas describing the borders of the sets S_i into Taylor series and keep only terms which are linear in the (coordinates of the) difference $s - \widetilde{s}$. Thus, it is reasonable to assume that the border of each of the two sets S_i is described by a linear equation – and is hence a (hyper-)plane: a line in 2-D space, a plane in 3-D space, etc.

As a result, we arrive at the following configuration. Let 2α denote the angle between the borders of the sets S_1 and S_2, so that the angles between each of these borders and the midline is exactly α. Let β denote the angle between the direction from $\widetilde{s}$ to s_0 and the midline. In this case, the angle between the border of S_1 and the midline is equal to $\alpha - \beta$. So, we arrive at the following configuration:



In the first option, we first project s_0 onto the set S_1. As a result, we get the following configuration:



Here, the projection line $s_0 s_1$ is orthogonal to the border of S_1. From the right triangle $\triangle \widetilde{s} s_0 s_1$, we therefore conclude that the distance $d_1 \stackrel{\text{def}}{=} d(\widetilde{s}, s_1)$ from the projection point s_1 to the desired point $\widetilde{s}$ is equal to

$$d_1 = d_0 \cdot \cos(\alpha - \beta). \quad (2)$$

On the next step, we project the point s_1 from S_1 onto the line S_2 which is located at the angle 2α from S_1. Thus, for the projection result s_2, we will have

$$d_2 = d(s_2, \widetilde{s}) = d_1 \cdot \cos(2\alpha) = d_0 \cdot \cos(\alpha - \beta) \cdot \cos(2\alpha). \quad (3)$$

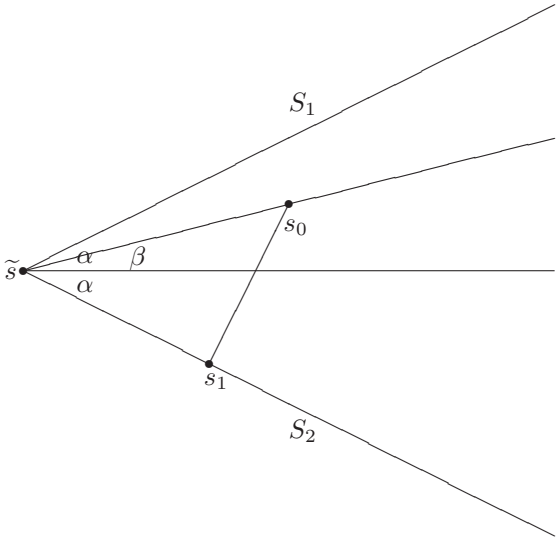
After this, we may again project onto S_2, then again project onto S_1, etc. For each of these projections, the angle is equal to 2α, so after each of them, the distance from the desired point $\widetilde{s}$ is multiplied the same factor $\cos(2\alpha)$.

As a result, after k projection steps, we get a point s_k at a distance

$$d_k = d(s_k, \widetilde{s}) = d_0 \cdot \cos(\alpha - \beta) \cdot \cos^{k-1}(2\alpha) \quad (4)$$

from the desired state $\widetilde{s}$.

In the second option, we start with teaching S_2, i.e., if we first project the state s_0 into the set S_2. In this option, we get the following configuration:



Here, we have

$$d_1 = d_0 \cdot \cos(\alpha + \beta), \quad (5)$$

$$d_2 = d(s_2, \widetilde{s}) = d_1 \cdot \cos(2\alpha) = d_0 \cdot \cos(\alpha + \beta) \cdot \cos(2\alpha), \quad (6)$$

$$\ldots$$

$$d_k = d(s_k, \widetilde{s}) = d_0 \cdot \cos(\alpha + \beta) \cdot \cos^{k-1}(2\alpha). \quad (7)$$

Since, in general, $\cos(\alpha - \beta) \neq \cos(\alpha + \beta)$, we can see that a change in the presentation order can indeed drastically change the success of the learning procedure.

2.5 Conclusion: Dependence Explained

Thus, our simple geometric model explains why the effectiveness of learning depends on the order in which the material is presented.

3 Specific Recommendations: Case of Two-Part Knowledge

3.1 Analysis

Let us extract more specific recommendations from our model. According to the above formulas, starting with S_1

ISBN: 978-989-95079-6-8

leads to a more effective learning than starting with S_2 if and only if

$$d_0 \cdot \cos(\alpha - \beta) \cdot \cos^{k-1}(2\alpha) < d_0 \cdot \cos(\alpha + \beta) \cdot \cos^{k-1}(2\alpha), \quad (8)$$

i.e., equivalently, if and only if

$$\cos(\alpha - \beta) < \cos(\alpha + \beta). \quad (9)$$

Since for the angles $x \in [0, \pi]$, the cosine $\cos(x)$ is a decreasing function, we conclude that projection of S_1 is better if and only if

$$\alpha - \beta > \alpha + \beta. \quad (10)$$

Thus, we arrive at the following recommendation:

3.2 Recommendation

To make learning more efficient, we should start with studying the material which is further away from the current state of knowledge. In other words, we should start with a material that we know the least.

This ties in nicely with a natural commonsense recommendation that to perfect oneself, one should concentrate on one's deficiencies.

This recommendation is also in a very good accordance:

- with the seemingly counter-intuitive conclusion from [9, 10, 14, 21], that studying more difficult (abstract) ideas first enhances learning, and
- with the human infant studies [18] according to which a concentration on teaching, to human infants, skills that they can easily learn is detrimental in the long run.

4 General Case

4.1 Analysis of the Problem

What happens in the general case, when instead of only two knowledge components, we have a large number of different components? In the beginning, it still makes sense to project to the set S_{i_1} which is the farthest from the original state s_0.

After this original projection, in the general case, we still have a choice. We can project to any set S_{i_2}, $i_2 \neq i_1$, in which case the current distance d_1 to the desired state is multiplied by the cosine $\cos(\alpha_{i_1 i_2})$ of the angle between the corresponding sets S_{i_1} and S_{i_2}. After k steps, we get the original distance multiplied by the product of the corresponding cosines.

Our objective is to find the best order, i.e., the sequence S_{i_1}, S_{i_2}, ..., S_{i_n} that covers all n sets $S_1, \ldots, S_n$ and for which the corresponding product

$$\cos(\alpha_{i_1 i_2}) \cdot \cos(\alpha_{i_1 i_2}) \cdot \ldots \cdot \cos(\alpha_{i_n i_1}) \quad (11)$$

attains the smallest possible value.

Usually, it is easier to deal with the sums than with the products. To transform the product into a sum, we can use the fact that:

- minimizing the product is equivalent to minimizing its logarithm, and
- the logarithm of the product is equal to the sum of the logarithms.

Thus, minimizing the product (11) is equivalent to minimizing the sum

$$D(i_1, i_2) + \ldots + D(i_n, i_1), \quad (12)$$

where

$$D(i, j) \stackrel{\text{def}}{=} \log(\cos(\alpha_{ij})). \quad (13)$$

In other words, we arrive at the following problem:

- we have n objects with known distances

$$D(i, j), \ \ 1 \leq i, j \leq n;$$

- we must find a way to traverse all the objects and come back in such a way that the overall traveled distance is the smallest possible.

This is a well-known problem called a *traveling salesman* problem.

It is known that

- in general, this problem is NP-hard (see, e.g., [17]), and
- in many cases, there exist reasonable algorithms for solving this problem; see, e.g., [1].

4.2 Recommendations and the Need for Expert (Fuzzy) Knowledge

Based on the above analysis, we can make the following recommendations:

- To find the optimal order of presenting the material, we must solve the corresponding instance of the traveling salesman problem, with the distances determined by the formula (13).
- Since in general, the traveling salesman problem is computationally difficult (NP-hard), to efficiently solve this problem, we must use expert knowledge – the knowledge for which fuzzy technique have been invented; see, e.g., [11, 15].

Acknowledgment

The author is thankful to Mourat Tchoshanov for valuable discussions, and to the anonymous referees for useful suggestions.

References

[1] D. L. Applegate, R. M. Bixby, V. Chvátal, and W. J. Cook, *The Traveling Salesman Problem*, Princeton University Press, Princeton, New Jersey, 2006.

[2] D. F. Bjorklund, *Why Youth is Not Wasted on the Young: Immaturity in Human Development*, Blackwell Publishing, 2007.

[3] D. F. Bjorklund and B. L. Green, The adaptive nature of cognitive immaturity, *American Psychologist*, 47:46–54, 1992.

[4] D. F. Bjorklund and A. Pellegrini, *The Origins of Human Nature: Evolutionary Developmental Psychology*, American Psychological Association, Washington, DC, 2002.

ISBN: 978-989-95079-6-8

[5] V. Davydov, *Types of Generalizations in Instruction: Logical and Psychological Problems in the Structuring of Schhol Curricular*, NCTM, Reston, Virginia, 1990.

[6] B. J. Ellis and D. F. Bjorklund (eds.), *Origins of the Social Mind: Evolutionary Psychology and Child Development*, Guilford Press, New York, 2005.

[7] L. G. Gubin, B. T. Polyak, and E. V. Raik, The method of projections for finding the comon point of convex sets, *USSR Computational Mathematics and Mathematical Physics*, 7:1–24, 1967.

[8] H. F. Harlow, The development of learning in the rhesus monkey, *American Scientist*, 47:459–479, 1959.

[9] J. A. Kaminski, V. M. Sloutsky, and A. F. Heckler, Do children need concrete instantiations to learn an abstract concept?, In: R. Sun and N. Miyake (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society, Vancouver, BC, July 26–29, 2006*, Lawrence Erlbaum, Mahwah, NJ, 2006, pp. 411–416.

[10] J. A. Kaminski, V. M. Sloutsky, and A. F. Heckler, The advantage of abstract examples in learning math, *Science*, 320:454–455, April 25, 2008.

[11] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, Upper Saddle River, New Jersey, 1995.

[12] E. J. Kontoghiorghes, *Handbook of Parallel Computing and Statistics*, CRC Press, Boca Raton, Florida, 2006.

[13] L. M. Lesser and M. A. Tchoshanov, The effect of representation and representational sequence on students' understanding, In: G. M. Lloyd, M. Wilson, J. M. L. Wilkins, and S. L. Behm (eds.), *Proceedings of the 27th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education PME-NA'2005*, Roanoke, Virginia, October 20–23, 2005.

[14] L. M. Lesser and M. A. Tchoshanov, Selecting representations, *Texas Mathematics Teacher*, 53(2):20–26, Fall 2006.

[15] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic*, CRC Press, Boca Raton, Florida, 2005.

[16] S. Paper and M. A. Tchoshanov, The role of representation(s) in developing mathematical understanding, *Theory into Practice*, 40(2):118–127, 2001.

[17] C. H. Papadimitriou, *Computational Complexity*, Addison Wesley, 1994.

[18] H. Papousek and M. Papousek, Mothering and the cognitive head-start: Psychobiological considerations. In H. R. Schaffer (Ed.), *Studies in Mother-Infant Interaction*, Academic Press, London, 1977.

[19] E. Remmel, The benefits of a long childhood, *American Scientist*, 96(3):250–251, May-June 2008.

[20] H. Stark and Y. Yang, *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets and Optics*, Wiley, New York, 1988.

[21] M. A. Tchoshanov, *Visual Mathematics*, ABAK Publ., Kazan, 1997 (in Russian).

[22] J. Van Patten, C. Chao, and C. Reigeluth, A review of strategies for sequencing and synthesizing instruction, *Review of Educational Research*, 56(4):437–471, 1986.

ISBN: 978-989-95079-6-8

Intensity assessment of pedestrian collisions in city of Mashhad based on fuzzy probabilities

Farzaneh Tatari and Mohammad-R. Akbarzadeh-T.

Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
Email: fa_tatari@yahoo.com, akbarzadeh@ieee.org

Abstract- *Fuzzy probabilities have been widely used in the areas of risk assessment and decision making. Here, we propose a system based on fuzzy probabilities for assessing the intensity of pedestrian collisions and levels of injury. The proposed approach uses possibility probability distributions for demonstrating fuzzy probabilities. A composition of possibility probability distributions (fuzzy probabilities) of collision intensities coming from different environmental variables yields the final fuzzy probability of intensity with higher reliability. The proposed approach is then applied to an actual database of pedestrian collisions from the City of Mashhad for over a period of five years.*

Keywords_ *Collision Intensity, Collision Variables, Fuzzy Probability, Possibility Probability Distribution.*

1 Introduction

Walking is a healthy and natural mean of transposition; it is therefore desirable that it also be safe. However, with the modern society and its rising crowd of transportation vehicles, this aspect of transposition has been sidelined and pedestrian collisions have caused significant social and financial losses every year. The safety of pedestrians during daily transposition is influenced by various variables. An appropriate paradigm for handling the effect of these variables can help insurance companies as well as city planners to better evaluate the safety of pedestrians in different situations and to plan the city environment for their improved safety. Unfortunately, recognizing all of the effective variables in a collision and their precise amount of effectiveness is non-trivial. The available information is highly imprecise and can lead to uncertain conclusions. With respect to the existing uncertainty in the collected information, the first source of data is known as "soft" data [17] that can be represented in linguistic form. This data is from expert opinions based on their background knowledge about the collision's intensity by considering related variables. Due to the linguistic form of this information and imprecision of data in this application, soft computing based methods, and in particular fuzzy-logic based analysis, offer a promising solution paradigm [18] to handle this type of uncertainty. In addition to soft data, the statistical historical data of pedestrian collisions are another valid data source which can be helpful in this case. The statistical data of pedestrian collisions may be incomplete when the number of observed data is not considerably large. Hence the statistical data are also accompanied by uncertainty which is caused by sparsity and insufficiency of data. We have found fuzzy probability framework as a suitable approach which enables us to enhance the reliability of our assessments by employing the both databases of soft data and statistical data at the same time. Fuzzy probabilities were first introduced by Zadeh (1984). Fuzzy probability theory [2] is a fuzzy approach to probability theory and is a generalized form of probability theory. In fuzzy probability, probability theory is complemented with an extra dimension of uncertainty provided by fuzzy set theory [16]. Generally we can divide the applications of fuzzy probabilities to two different main areas. The first is the area of reliability and risk assessment. In this area fuzzy probability has been widely applied in fuzzy fault trees [3,4,8] to assess the fault risk, reliability assessment for pressure piping [20], risk assessment of natural hazards [7] and reliability enhancement by combining expert opinions [14]. The second area is the area of the decision making. In this area, the fuzzy probabilities have been employed for decision making in perception-based theory [19], optimal decision fusion [11,12], inference by aggregation [9], information retrieval [6] and inventory control [1].

At any collision, there are two states for the pedestrian who has been hit by a car, in one state he would be just injured and in the other state he would die. Here, we employ fuzzy probabilities to handle the imprecision in expert opinions as the first dataset to assess the probability of injury and death for the pedestrians of Mashhad. We use fuzzy numbers to show these fuzzy probabilities. The second database is the statistical numeric data of pedestrian's collisions in Mashhad [10]. We combine these two insufficient databases to provide a more reliable result by the use of possibility probability distribution (PPD) which can model fuzzy probabilities in a suitable way.

2 Possibility probability distribution

The idea of representing assessments with PPD has been originally proposed by Haung [5]. This idea was improved by Karimi and Hullermeier [7] which was a new concept of fuzzy probability expressed by PPD. As we mentioned, on one hand, the statistical data in our type of application are typically insufficient, on the other hand background knowledge is often accompanied by imprecision. Therefore we apply the PPD to combine these two databases based on Bayesian approach. In the PPD approach, the prior knowledge (soft data) which is called prior distribution can be specified by a fuzzy number. $\overline{\theta}^{(i)}$ is an estimated probability by an expert for the state of death or injury in any collision where i is the variable that influences this probability and $p_i(\theta)$ is the prior distribution modeled in the form of symmetric, triangular fuzzy number, whose center is given by $\overline{\theta}^{(i)}$ (θ is the variable representing any probability in unit interval of [0,1]). As stated in [7] the support of the fuzzy number is of the form

ISBN: 978-989-95079-6-8

$$[\bar{\theta}^{(i)} - \bar{\theta}^{(i)}(1-\bar{\theta}^{(i)}).s \ , \bar{\theta}^{(i)} + \bar{\theta}^{(i)}(1-\bar{\theta}^{(i)}).s] \quad (1)$$

$$0 \le s \le \min(1/\bar{\theta}^{(i)}, 1/(1-\bar{\theta}^{(i)})) \quad (2)$$

the so-called uncertainty factor, is a constant that specifies the reliability of the point estimation. For combining prior knowledge with empirical data which is the available statistical numeric data, we first divide the unit interval of probabilities into m points, $\theta_j, j = 1,...,m$, and then calculate the posterior probability for each point θ_j as follows [7].

$$p'_i(\theta_j) = \binom{n}{n_i} \theta_j^{n_i} (1-\theta_j)^{n-n_i} . p_i(\theta_j) \approx$$

$$\frac{1}{c}(\gamma_j . q_j + (1-\gamma_j).q_{j+1}),$$

$$q_j = \theta_j^{n_i} (1-\theta_j)^{n-n_i} \, p_i(\theta_j) \quad (3)$$

where n_i is the number of observations that satisfies the features of the ith variable among n observations. $\gamma_j \in [0,1)$ is a unique value such that

$$\theta_j = \gamma_j \theta_j + (1-\gamma_j).q_{j+1} \quad (4)$$

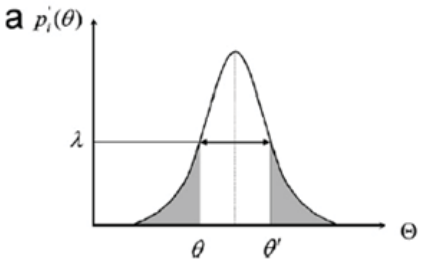
and c is a normalizing constant that guarantees

$$\int_0^1 p_i'(\theta) d\theta = 1 \quad (5)$$

Now is the time of transforming probability distribution to possibility distribution. Suppose a probability measure P on a set X is obtained via some statistical experiment. This probability function is a very rich piece of information, if the number of statistical experiments supporting it, is high enough. According to widely accepted consistence principle, the possibility measures should dominate P in the sense that $P(A) \le \pi(A)$ for all events $A \subseteq X$ and the maximally specific possibility distribution that exists must be unique [7]. As illustrated in [7] the most specific possibility distribution that approximates P from above (in the sense that $P \le \pi$) can be derived quite easily: for obtaining a possibility degree for any θ, $\pi_i(\theta)$, first $\lambda = p_i(\theta)$ is computed. Next, the second boundary point θ' that satisfies $p_i(\theta') = \lambda$ is found. Finally, the possibility degree $\pi_i(\theta)$ can be obtained as:

$$\pi_i(\theta) = \pi_i(\theta') = 1 - \left| \int_\theta^{\theta'} p'_i(x) dx \right|. \quad (6)$$

This possibility is equal to the area shaded grey in Fig. 1.a.



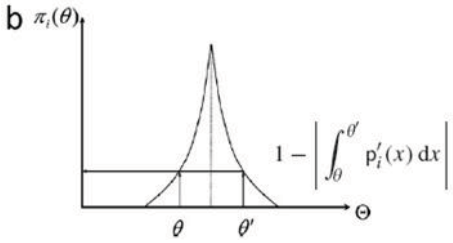


Figure 1: a. probability distribution to be transformed into a possibility distribution; b. transformed possibility distribution

3 Combining experts opinions and statistical data

We collected expert opinions (based on their own background knowledge and their reviews of traffic statics of death and injury in middle east countries [10]) about the percentage of collisions intensity (injury and death) by considering just one certain significant variable. The results of expert opinions are shown in Tables 1~5.

Table 1: Significant variable: hit organ

Injury probability if	Death probability if
Head and neck is hit, is approximately 0.09	Head and neck is hit, is approximately 0.55
Hand and arm is hit, is approximately 0.01	Hand and arm is hit, is approximately 0.001
Chest and stomach is hit, is approximately 0.2	Chest and stomach is hit, is approximately 0.2
Legs and pelvis is hit, is approximately 0.3	Legs and pelvis is hit, is approximately 0.03
More than one organ are hit, is approximately 0.5	More than one organ are hit, is approximately 0.3

Table 2: Significant variable: agility (crossing ability)

Injury probability for a pedestrian	Death probability for a pedestrian
Able to cross is 0.3	Able to cross is 0.4
Unable to cross is 0.6	Unable to cross is 0.6

Table 3: Significant variable: vehicle front shape

Injury probability if vehicle	Death probability if vehicle
Has bonnet is 0.7	Has bonnet is 0.4
Does not have bonnet is 0.2	Does not have bonnet is 0.2

Table 4: Significant variable: collision location

Injury probability if location is	Death probability if location is
In the city is 0.6	In the city is 0.3
Out of city is 0.04	Out of city is 0.05
In the suburb is 0.2	In the suburb is 0.15

Table 5: Significant variable: type of road

Injury probability if the road is a	Death probability if the road is a
Highway is 0.2	Highway is 0.6
Boulevard is 0.5	Boulevard is 0.6
Two_way road is 0.3	Two_way road is 0.2
incidental way is 0.1	incidental way is 0.05

With respect to the resulting experts ideas we obtained the fuzzy numbers for the proposed probability based on the approach illustrated in Section 2 where s is chosen as

$$s = 1/2\,(\min(1/\bar{\theta}^{(i)}, 1/(1-\bar{\theta}^{(i)}))) \quad (7)$$

It should be mentioned that we normalized $P(\bar{\theta}^{(i)})$ of all

ISBN: 978-989-95079-6-8

the prior probability distributions by factor N, to convert the prior probability distributions to fuzzy numbers.

$$N = \begin{cases} 2/(\overline{\theta}^{(i)}) & \overline{\theta}^{(i)} \le 0.5 \\ 2/(1-\overline{\theta}^{(i)}) & \overline{\theta}^{(i)} \ge 0.5 \end{cases} \quad (8)$$

As an example, through Figs. 2~3 we have shown the result of fuzzy probabilities of injury and death based on hit organ variable.

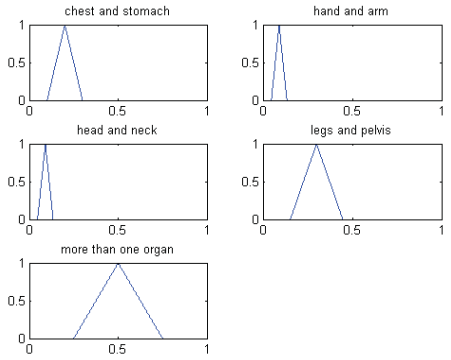


Figure 2: Injury fuzzy probability considering hit organ

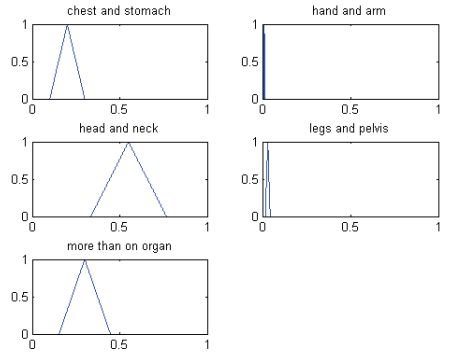


Figure 3: Death fuzzy probability considering hit organ

In the next step we apply (3) to obtain the posterior distribution based on statistical data which was gathered during one month in Shahidkamyab and Emamreza hospitals. In our database among 131 victims 99 of them had just injured and 32 of them had died. While injury and death are two independent states, we have considered the injuries which ends to fatality in the death statistical society. Therefore n=32 for dead pedestrians statistical society and n=99 for injured statistical society, these numbers are distributed through any variable partitions based on the collected data which is available in Tables 6 ~ 10. Among 18 factors which consist of human, vehicle, road and environmental factors, we could find 5 variables (hit organ, agility (crossing ability), vehicle front shape, location of collision and type of road) that had significant relations with collision intensity by the 5 percentage of error using SPSS.

Table 6: Distribution of statistical database on Significant variable of hit organ

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Head and neck = 10	Head and neck = 17
Hand and arm = 0	Hand and arm = 4
Chest and stomach = 5	Chest and stomach = 2
Legs and pelvis = 33	Legs and pelvis = 1
More than one point = 51	More than one point = 8

Table 7: Distribution of statistical database on Significant variable of agility (crossing ability)

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Able = 97	Able = 24
Unable =2	Unable = 8

Table 8: Distribution of statistical database on Significant variable of vehicle of front shape

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Has bonnet = 86	Has bonnet = 23
Does not have bonnet = 13	Does not have bonnet = 9

Table 9: Distribution of statistical database on Significant variable of location of the collision

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
In the city =60	In the city =14
Out of city =3	Out of city =6
Suburb =36	Suburb =12

Table 10: Distribution of statistical database on Significant variable vehicle of type of road

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Highway = 4	Highway = 7
Boulevard = 51	Boulevard = 10
Two way road = 34	Two way road = 11
Incidental way = 10	Incidental way = 4

The resulting posterior distributions of injury and death for the hit organ variable are as below:

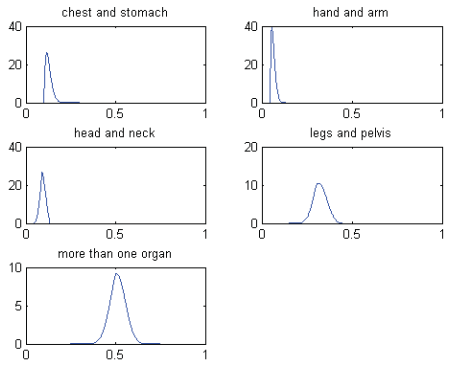


Figure 4: Posterior distribution of injury considering hit organ

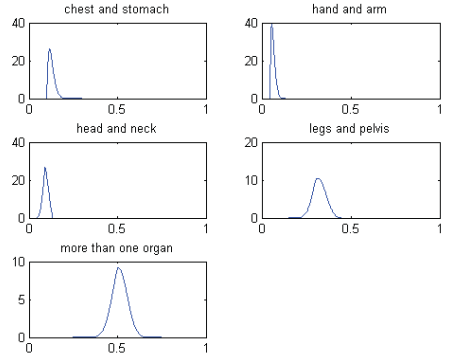


Figure 5: Posterior distribution of death considering hit organ

ISBN: 978-989-95079-6-8

Based on the procedure mentioned for transforming probability distribution to possibility distribution in Section 2, we obtained the possibility distributions of probability distributions. By comparing the prior fuzzy probabilities and possibility-probability distributions (Figs. 6~15), It can be easily interpreted that by combining the information of two databases less ambiguous, i.e. more certain and reliable results, are obtained in PPDs. It is also obvious that the PPD (fuzzy probability) can easily handle the existing uncertainty in the results.

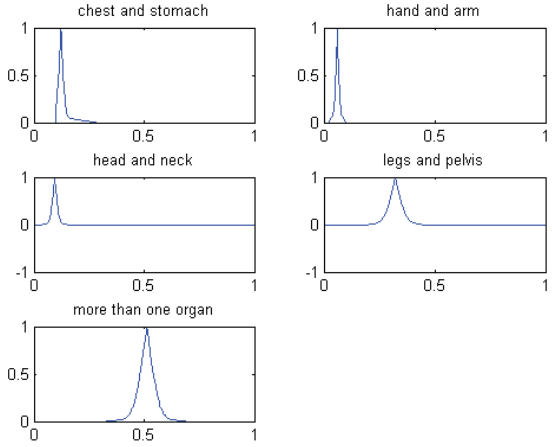


Figure 6: Possibility-probability distribution of injury considering hit organ

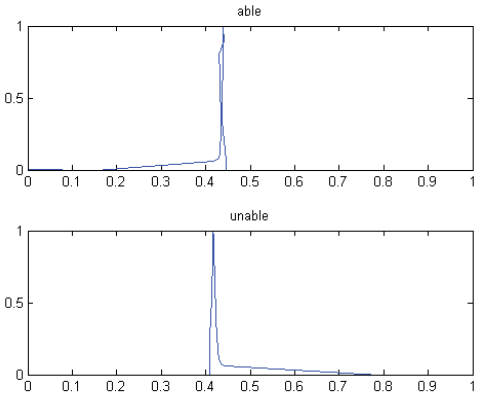


Figure 7: Possibility-probability distribution of injury knowing agility (crossing ability)

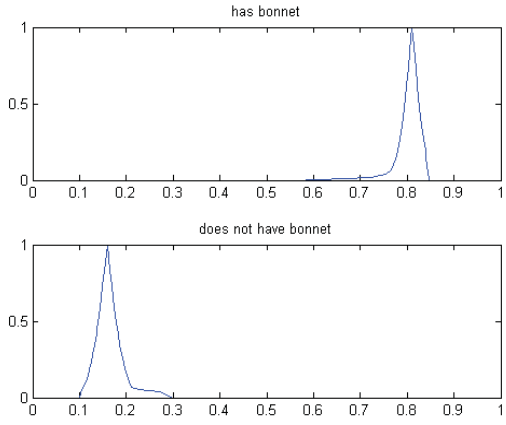


Figure 8: Possibility_probability distribution of injury considering vehicle front shape

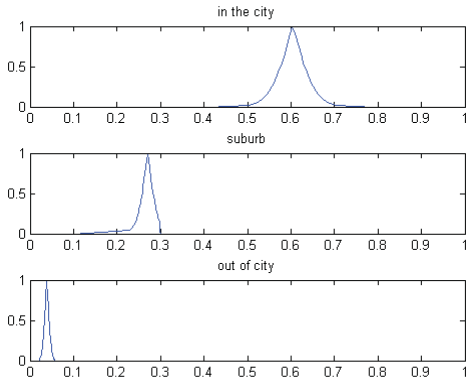


Figure 9: Possibility_probability distribution of injury considering location of collision

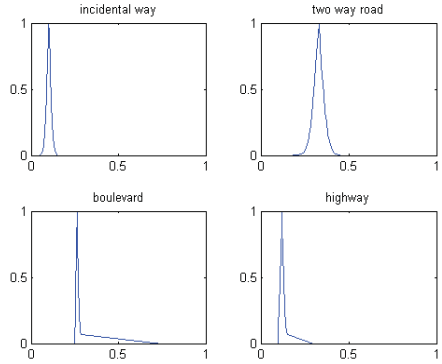


Figure 10: Possibility_probability distribution of injury considering type of road

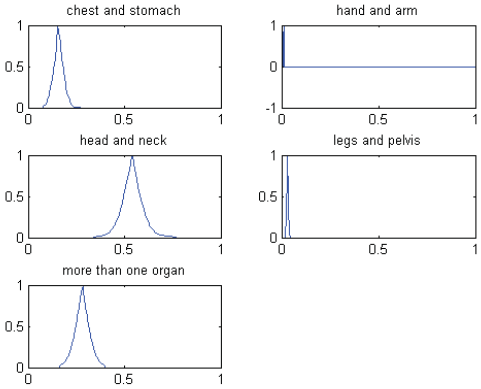


Figure 11: Possibility_probability distribution of death considering hit organ

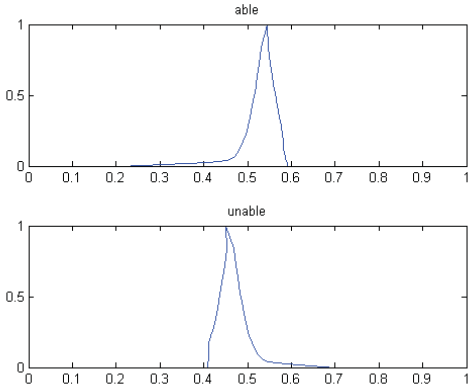


Figure 12: Possibility_probability distribution of death considering agility (crossing ability)

ISBN: 978-989-95079-6-8

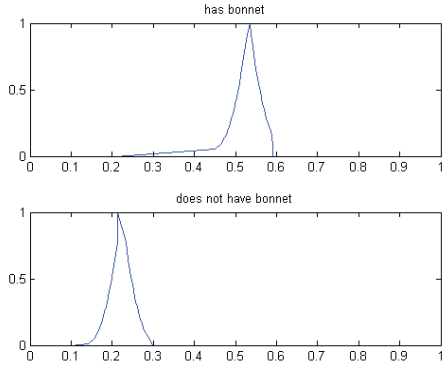


Figure 13: Possibility_probability distribution of death considering vehicle front shape

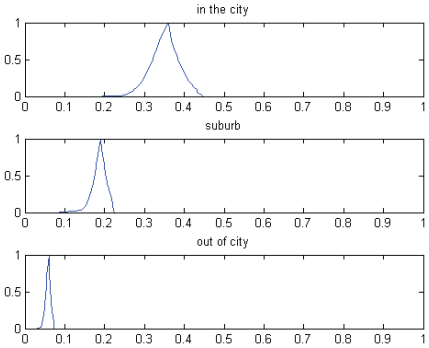


Figure 14: Possibility_probability distribution of death considering location of collision

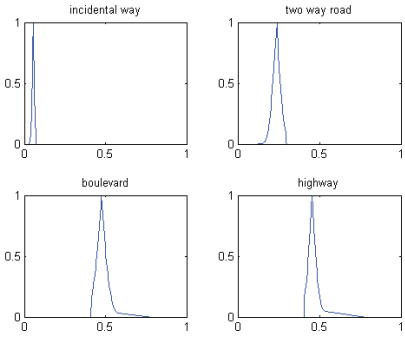


Figure 15: Possibility-probability distribution of death considering type of road

4 Collision intensity assessment

As it was explained before there are various variables which influence the intensity of collisions. If we assess the collision intensity based on one or few known variables the uncertainty of the assessment is so higher in comparison with the time that the assessment is based on more known variables. Therefore we can assess the fuzzy probability of injury or death of the pedestrians based on the combination of all significant variables PPDs. It should be mentioned that the employed method for combining fuzzy probabilities is based on fuzzy arithmetic [2] not Bayesian approach.

Assume that $\overline{A}$ and $\overline{B}$ are fuzzy probabilities of X. By having $\overline{A}$ and $\overline{B}$ which have been calculated based on different variables of X , we want to obtain $\overline{C}$ as a more reliable fuzzy probability of X by considering both variables, as the following

$$P[X(a)] \text{ is } \overline{A} \text{ and } P[X(b)] \text{ is } \overline{B} \equiv P[X(a,b)] \text{ is } \overline{C}.$$

Where a and b are related variables of X which effect assessing the fuzzy probability of X. One of the most common ways to aggregate fuzzy probabilities such as $\overline{A}$ and $\overline{B}$ is to obtain the α _cuts of $\overline{C}$ for any $\alpha \in [0,1]$ as below

$$\overline{A}[\alpha] = [a_1(\alpha), a_2(\alpha)], \overline{B}[\alpha] = [b_1(\alpha), b_2(\alpha)] \Rightarrow$$

$$\overline{C}(\alpha) = [\min(a_1(\alpha), b_1(\alpha)), \max(a_2(\alpha), b_2(\alpha))] \quad (9)$$

We can extend this procedure to more than two fuzzy probabilities with different variables to obtain a more reliable fuzzy probability of X, $\overline{C}$. It is also notable to know that any α _cut in any fuzzy probability is a sub interval of the unit interval of probabilities. However the mentioned method may exhibit shortcomings, first the max and min operations are noninteractive so they tend to lead to the results that are getting close to 1 and 0 respectively. Second, the result of combination at any level of α is entirely dependent upon the extreme bounds. To overcome these shortcomings we propose to calculate the α _cuts of $\overline{C}$ based on fuzzy arithmetic [2].

$$\overline{A}[\alpha] = [a_1(\alpha), a_2(\alpha)], \overline{B}[\alpha] = [b_1(\alpha), b_2(\alpha)] \Rightarrow$$

$$\overline{C}(\alpha) = [a_1(\alpha) \times b_1(\alpha) \ , \ a_2(\alpha) \times b_2(\alpha)] \quad (10)$$

The only drawback of fuzzy arithmetic method is that it results in low fuzzy probabilities which do not conform the reality (statistical data) but we can solve this problem by normalizing the obtained fuzzy probabilities of all the existing states. This normalization is in the way that for each $\alpha -$ cut , $\alpha \in [0,1]$, the summation of $\alpha -$ cut boundary points (probabilities) must be equal to one. In the specific application that we focus on it, we have two states of "X= death, Y= injury" in any collision, therefore we must normalize the final fuzzy probabilities in the way that

$$P'_{X1\alpha} = \frac{P_{X1\alpha}}{(P_{X1\alpha} + P_{Y2\alpha})}, \quad P'_{Y2\alpha} = \frac{P_{Y2\alpha}}{(P_{X1\alpha} + P_{Y2\alpha})},$$

$$P'_{X2\alpha} = \frac{P_{X2\alpha}}{(P_{X2\alpha} + P_{Y1\alpha})}, \quad P'_{Y1\alpha} = \frac{P_{Y1\alpha}}{(P_{X2\alpha} + P_{Y1\alpha})}$$

$$\text{then } P'_{X1\alpha} + P'_{Y2\alpha} = 1 \text{ and } P'_{X2\alpha} + P'_{Y1\alpha} = 1 \ \forall \alpha \in [0,1] \quad (11)$$

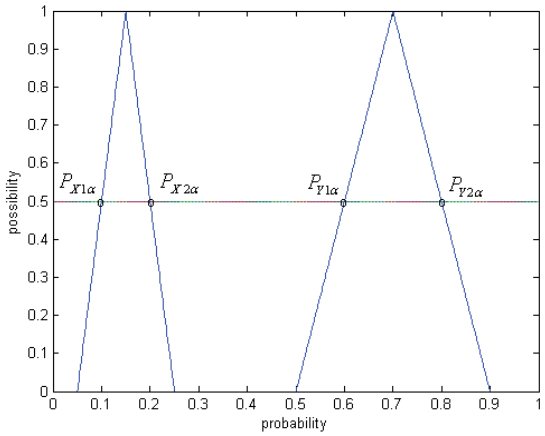


Figure 16: The probabilities which must be normalized at any α level

ISBN: 978-989-95079-6-8

We continue by an example where we used the fuzzy arithmetic procedure to obtain a more reliable fuzzy probability of injury and death based on the obtained results. In the following example for combining the fuzzy probabilities of injury and death which were assessed based on 5 significant variables, we use seven α _cuts (α = 0, 0.1, 0.3, 0.5, 0.7, 0.8, 1) of all fuzzy probabilities (PPDs) to do the combination.

Example: A pedestrian who does not have the agility (crossing ability) is crashed by a mini-bus while crossing a two-way road in city of Mashhad. His chest and stomach are hit by the vehicle. (a) What is the fuzzy probability of his injury? (b) What is the fuzzy probability of his death?

Hit organ: chest and stomach, agility (crossing ability): unable to cross, Vehicle front shape: does not have bonnet, Collision location: in the city, Type of road: two-way road. By combining the PPDs of the related variables we can obtain the final fuzzy probability of injury as below:

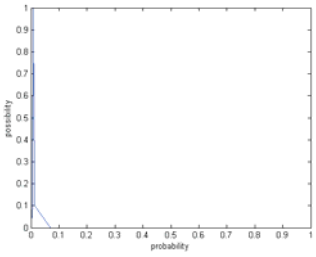


Figure 17: Fuzzy probability of injury before normalization

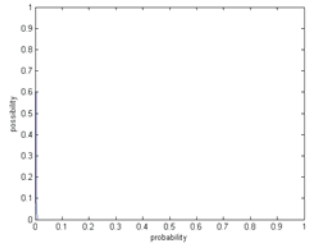


Figure 18: Fuzzy probability of death before normalization

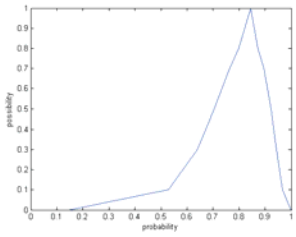


Figure 19: Normalized fuzzy probability of injury

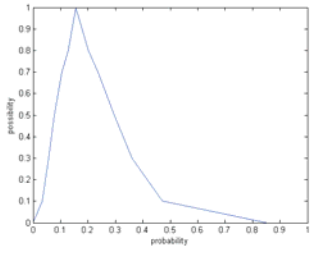


Figure 20: Normalized fuzzy probability of death

The normalized fuzzy probabilities are closer to statistical data and reality and the fuzzy probability of death and injury are complement of each other. The obtained result is more reliable while it has considered 5 significant variables at the same time instead of one.

5 Conclusion

Each year, too many pedestrians die all over the world with significant social and financial losses. The intensity of these collisions is of great importance for city planners and insurance companies alike. Here, we propose the use of fuzzy set theory to complement the probability theory with an additional dimension of uncertainty which leads to fuzzy probability characterized in terms of possibility probability distribution. We assessed the intensity of pedestrian collisions in Mashhad based on the variables that had significant relation with the intensity by use of fuzzy probability. The results indicate that by considering more significant variables in our assessment, we can decrease the amount of uncertainty and reach a more reliable result.

References

[1] J.J.Buckley et al., Fuzzy Mathematics in Economics and Engineering, Physica-Verlag, Heidelberg, Germany, 2002.
[2] J.J.Buckley, Fuzzy probabilities, new approach and applications, Springer, 2005.
[3] James P. Dunyak and Donald Wunsch, Fuzzy Probability for System Reliability, Proceedings of the 37th IEEE Conference on Decision & Control Tampa, 1998.
[4] James P. Dunyak et al., A Theory of Independent Fuzzy Probability for System Reliability, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 7, NO. 2, 1999.
[5] C. Huang, Fuzzy risk assessment of urban natural hazards, Fuzzy Sets and Systems 83 , 271–282, 1995.
[6] Z. Huang and T. D. Gedeon, Information retrieval estimation via fuzzy probability, World Automation Congress (WAC), 2006.
[7] I. Karimi and E. Hüllermeier, Risk assessment system of natural hazards: A new approach based on fuzzy probability, Fuzzy Sets and Systems 158, 987 – 999, 2007.
[8] R. Kenarangui, Event-Tree Analysis by Fuzzy Probability, IEEE TRANSACTIONS ON RELIABILITY, VOL. 40, NO. 1, 1991.
[9] A. L. Ralescu , D. A. Ralescu, Yoko Yamakata , Inference by aggregation of evidence with applications to fuzzy probabilities, Information Sciences 177, 378–387, 2007.
[10] M. Rezaei, estimation of collisions intensity for pedestrians in Mashhad using logistic regression, pp.92_107, 2004.
[11] G. Wang , S. Mao and Y. He, Optimal Decision Fusion With Fuzzy Priori Probabilities, Proceedings of IBP, 1998.
[12] G.Wang, S.Mao and Y. He, Distributed detection fusion with fuzzy priori probabilities and fuzzy cost function, IEEE INTERNATIONAL RADAR CONFERENCE, 2000.
[13] Ronald R. Yager, Decision Making with Fuzzy Probability Assessments, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 7, NO. 4, 1999.
[14] D. Yu and W. S. Park, Combination and evaluation of expert opinions characterized in terms of fuzzy probabilities, Annals of Nuclear Energy 27, 713-726, 2000.
[15] L. A. Zadeh, Fuzzy probabilities, Information processing and management 20(3), pp. 363_372, 198.
[16] L.A. Zadeh, Fuzzy sets , Information and control, 8(3), pp. 338_353, 1965.
[17] L.A. Zadeh, Posibility theory and soft data analysis, In: Cobb, L. and R. M. Thrall, eds., mathematical frontiers of the social and policy sciences. Westview press, Boulder, Colorado, pp.69_129, 1981.
[18] L.A. Zadeh, Soft computing and fuzzy logic, IEEE Softw are, 11 (6), pp.48_56, 1994.
[19] L. A. Zadeh, Toward a perception_based theory of probabilistic reasoning with imprecise probabilities, Journal of statistical planning and inference 105, 233-264, 2002.
[20] J. Zhou, Reliability assessment method for pressure piping containing circumferential defects based on fuzzy probability, International Journal of Pressure Vessels and Piping 82, 669–678, 2005.

ISBN: 978-989-95079-6-8

Number-free Mathematics Based on T-norm Fuzzy Logic

Libor Běhounek

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic
Email: behounek@cs.cas.cz

Abstract—*The paper presents a form of rendering classical mathematical notions by formal theories over suitable t-norm fuzzy logics in such a way that references to real numbers are eliminated from definitions and theorems, being removed to the standard semantics of fuzzy logic. Several examples demonstrate how this move conceptually simplifies the theory in exchange for non-classical reasoning, facilitates certain generalizations, and puts the concepts into a different perspective. The formal framework employed for the number-free formalization of mathematical concepts is that of higher-order fuzzy logic, also known as Fuzzy Class Theory.*

Keywords—Fuzzy Class Theory, fuzzy mathematics, fuzzy set, real-valued function, similarity-based limit, t-norm fuzzy logic.

1 Introduction

In the standard semantics of t-norm fuzzy logics [1], truth values are represented by real numbers from the unit interval $[0, 1]$; truth functions of n-ary propositional connectives are certain real-valued functions on $[0, 1]^n$; and the quantifiers $\exists, \forall$ are interpreted as the suprema and infima of sets of truth values. Read inversely, the apparatus of t-norm fuzzy logics expresses, by means of their standard semantics, certain first- and second-order constructions over real numbers. The axioms and rules of t-norm logics are designed to capture basic properties of such constructions and ensure the soundness of formally derivable theorems expressible in the language.

A suitable formal theory in first-order fuzzy logic thus can, by means of its standard semantics, express facts about classical mathematical notions and make them derivable by logical deductions in fuzzy logic. Since real numbers occur in the standard semantics of such a theory, they need no longer be explicitly mentioned by the theory itself. A notion of classical mathematics then becomes represented as the standard model of another notion of formal fuzzy mathematics that makes no explicit reference to real numbers: the reals are only implicitly present behind the logical axioms that govern reasoning in formal fuzzy mathematics.

This way of eliminating real numbers from the theory in favor of reasoning by means of t-norm fuzzy logic will here be called the *number-free* approach.

Number-free formalization of mathematical concepts is not new and has implicitly been around since the beginning of the theory of fuzzy sets: in fact, the notion of fuzzy set itself can be understood as a number-free rendering of the notion of real-valued function (see Section 2). However, it was only after the advancement of t-norm fuzzy logics, mostly in the past decade, that number-free notions could be treated rigorously in formal theories over first-order fuzzy logics. An early example of the number-free treatment of a classical notion is the formalization of finitely additive probability as a modality *Probable* in Łukasiewicz logic (see Section 3). Nevertheless, number-free rendering of more advanced mathematical notions requires more complex concepts of formal fuzzy mathematics—esp. higher-order set-constructions and a formal theory of fuzzy relations. The latter prerequisites have only recently been developed in the framework of higher-order fuzzy logic [2, 3, 4], which made it possible to apply the number-free approach systematically to various classical mathematical notions.

The profit we gain under the number-free approach in exchange for having to use non-classical rules of reasoning is, in the first place, conceptual simplification (roughly speaking, we get 'a set for a function'). Secondly, the number-free rendering often reveals a new perspective upon the notion, exposing the gradual quality of the classical construct, treating its gradualness as a primitive rather than derivative feature, and telling a different story about the concepts than is told by classical mathematics. Thirdly, many theorems of classical mathematics are under this approach detected as provable by simple (often, propositional) logical derivations in a suitable fuzzy logic, instead of complex classical proofs involving arithmetic, infima, functions, etc. Furthermore, adopting a non-standard semantics (e.g., using Chang's MV-algebra instead of the standard real numbers) or a different interpretation of the logical symbols involved (e.g., taking another t-norm for conjunction) yields an effortless generalization that might be harder to find (and motivate) in the classical language of crisp mathematics. Finally, the many-valuedness of all formulae in fuzzy logic makes it possible to consider another kind of graded generalization, by admitting partial satisfaction of the axioms for the represented notion (e.g., a metric to degree .99, cf. [5, 6]).

The particular formal framework in which number-free formalization of classical mathematical concepts is carried out in this paper is that of higher-order fuzzy logic, also known as Fuzzy Class Theory (FCT), over a suitable propositional fuzzy logic at least as strong as $\mathrm{MTL}_\triangle$. A working knowledge of FCT will be assumed throughout; for an introduction to the theory and more information see [2, 5]. For reference, the definitions used in the present paper are repeated below in the Appendix.

The aim of this paper is only to introduce the number-free approach as a distinct paradigm of formalization, rather than to develop particular number-free theories in depth. Therefore it only gives definitions of and a few observations on several number-free notions and discusses the merits of such a formalization; a detailed investigation of number-free theories is left for future work. A slightly more eloquent version of this paper has been made available as a research report [7].

ISBN: 978-989-95079-6-8

2 Real-valued functions

The very first notion of fuzzy mathematics, namely Zadeh's notion of fuzzy set [8], can be regarded as a number-free representation of classical $[0,1]$-valued functions by non-classical (namely, fuzzy) sets. Even though the formal apparatus of first-order fuzzy logic, which makes it possible to cast fuzzy sets as a primitive notion instead of representing them by classical real-valued functions, was developed years later, the tendency of regarding fuzzy sets and relations as a number-free rendition of real-valued functions has partly been present since the very beginning of the fuzzy set theory, as witnessed by the vocabulary and notation employed. E.g., the function $x \mapsto \min(A(x), B(x))$ is in the traditional fuzzy set theory denoted by $A \cap B$ and called the *intersection* of A and B: that is, the functions $A, B\colon X \to [0,1]$ are regarded as (non-classical) *sets* rather than real-valued functions (as the intersection of real-valued functions is a different thing). Similarly, such notions as fuzzy relational composition or the image under a fuzzy relation would make little sense if the n-ary functions involved were not regarded as (a kind of non-classical) relations. The terminological shift towards the number-free discourse is expressed by the very term "fuzzy set" and by its informal motivation of an unsharp collection of elements.

A certain part of the talk about *real-valued functions* and their properties was thus replaced by a talk about *sets* and *relations* that behave non-classically (e.g., do not follow the rule of excluded middle). This move eliminated references to numbers at least from the wording of some theorems, giving them compact forms and new conceptual meanings.[1] Clearly, the original number-free notion of fuzzy set has proved immensely fruitful even in its semi-formal form of traditional fuzzy mathematics. The formal apparatus of logic-based fuzzy mathematics has provided means for accomplishing the long-present idea and developing a fully fledged number-free approach to fuzzy sets and fuzzy relations.[2]

3 Finitely additive probability measures

Another number-free representation, already based on formal fuzzy logic, was the axiomatization of finitely additive probability measures as models of a fuzzy modality *Probable* over propositional Łukasiewicz logic by Hájek, Godo, and Esteva [10]. Later it was elaborated in a series of papers by Flaminio, Marchioni, Montagna, and the authors of [10]. We shall briefly recapitulate the original axiomatization (adapted from [11]) as another illustration of the number-free approach.

Consider a classical probability space $(\Omega, \mathcal{B}, \pi)$, where Ω is a set of elementary events, $\mathcal{B}$ a Boolean algebra of subsets of Ω, and π a finitely additive probability measure on $\mathcal{B}$, i.e., a function $\pi\colon \mathcal{B} \to [0,1]$ satisfying the following conditions:

$$\pi(\Omega) = 1 \tag{1}$$

$$\text{If } A \subseteq B, \text{ then } \pi(A) \leq \pi(B) \tag{2}$$

$$\text{If } A \cap B = \emptyset, \text{ then } \pi(A \cup B) = \pi(A) + \pi(B) \tag{3}$$

A number-free representation of π draws on the fact that a $[0,1]$-valued function on a Boolean algebra can be understood as a standard model of a fuzzy modality P over an algebra of crisp propositions. The above conditions on π can be transformed into the axioms for P, which (due to the additivity) are expressible in Łukasiewicz logic:[3]

Definition 3.1. The axioms and rules of the logic $\mathrm{FP}(\mathbf{Ł})$ are those of Łukasiewicz propositional logic plus the following axioms and rules, for non-modal φ, ψ:

$$\varphi \vee \neg\varphi \tag{4}$$

$$\text{From } \varphi \text{ infer } P\varphi \tag{5}$$

$$P\varphi, \text{ for all Boolean tautologies } \varphi \tag{6}$$

$$P(\varphi \to \psi) \to (P\varphi \to P\psi) \tag{7}$$

$$P(\neg\varphi) \leftrightarrow \neg P\varphi \tag{8}$$

$$P(\varphi \vee \psi) \leftrightarrow ((P\varphi \to P(\varphi \wedge \psi)) \to P\psi) \tag{9}$$

The axioms and rules of $\mathrm{FP}(\mathbf{Ł})$ ensure the following representation theorem (adapted from [10]):

Theorem 3.2. *Any probability space $(\Omega, \mathcal{B}, \pi)$ is a standard model of $\mathrm{FP}(\mathbf{Ł})$. Vice versa, all standard models of $\mathrm{FP}(\mathbf{Ł})$ are probability spaces.*

The representation theorem shows that the number-free theory faithfully captures the original notion of finitely additive probability measure. Moreover, by the completeness theorem of $\mathrm{FP}(\mathbf{Ł})$ w.r.t. probability spaces, proved in [11], all valid laws of finitely additive probability that are expressible in the language of $\mathrm{FP}(\mathbf{Ł})$ can in $\mathrm{FP}(\mathbf{Ł})$ be also (number-freely) proved.

In a given probability space $(\Omega, \mathcal{B}, \pi)$, i.e., a standard model of $\mathrm{FP}(\mathbf{Ł})$ with $\|P\| = \pi$, the truth value of $P\varphi$ is the probability of the event φ: $\|P\varphi\| = \pi(\|\varphi\|)$; the formula $P\varphi$ can therefore be understood as "φ is probable". Numerical calculations with probabilities are thus in $\mathrm{FP}(\mathbf{Ł})$ replaced by *logical derivations* with the modality "is probable". The key difference is that the latter represent inference *salva probabilitate* (i.e., salvo *probabilitatis gradu,* in the sense of P): e.g., it can be observed that numberless probability is transmitted by modus ponens, as $P\varphi \,\&\, P(\varphi \to \psi) \to P\psi$, i.e., "if φ is probable and $\varphi \to \psi$ is probable, then ψ is probable", is a theorem of $\mathrm{FP}(\mathbf{Ł})$.

The number-free approach to probability facilitates several kinds of generalization. First, generalizations to models over non-standard algebras for Łukasiewicz logic: thus we can have, e.g., probability valued in Chang's MV-algebra, or in non-standard reals (as in [12]). Second, a generalization to measures with only partially satisfied additivity (by a many-valued interpretation of the axioms, see [6]). And third, a generalization to the probability of fuzzy events, where one discards the axiom $\varphi \vee \neg\varphi$ for events and adapts the finite additivity axiom to work well with fuzzy events (e.g., as $P(\varphi \oplus \psi) \leftrightarrow ((P\varphi \to P(\varphi \,\&\, \psi)) \to P\psi)$).[4]

[1] The elimination of numbers also from proofs would have required a consistent use of first-order fuzzy logic. This approach was not embraced in the early works on fuzzy set theory, even though particular first-order fuzzy logics already existed, e.g., [9].

[2] Cf. [2, 3], where the formal theory of fuzzy sets and relations is developed without making any reference to real numbers in definitions, theorems, or proofs (only in explanatory semantic examples).

[3] Ł with rational truth constants is used in [10], but the truth constants are inessential for our account. The language is two-layered, admitting only non-modal formulae and propositional combinations of non-nested modal formulae.

[4] This approach has been taken in [13], though only over finitely-valued events, as the authors strove for the completeness of the logic; in [12] this was generalized to infinitely-valued events, with completeness w.r.t. non-standard reals.

ISBN: 978-989-95079-6-8

4 Distribution functions

Classical distribution functions present a special way how to define a probabilistic measure on Borel sets, i.e., on the σ-algebra $\mathcal{B}$ of subsets of the real line generated by all intervals $(-\infty, x]$. A function $f\colon \mathbb{R} \to [0,1]$ defines a measure on $\mathcal{B}$ with $\mu(-\infty, x] = f(x)$ and $\mu(\mathbb{R}) = 1$ iff it satisfies the following conditions, which can thus be taken as the axioms for distribution functions:

1. *Monotony:* if $x \leq y$ then $f(x) \leq f(y)$, for all $x, y \in \mathbb{R}$
2. *Margin conditions:* $\lim_{x\to-\infty} f(x) = 0$, $\lim_{x\to+\infty} f(x) = 1$
3. *Right-continuity:* $\lim_{x\to x_0^+} f(x) = f(x_0)$ for all $x_0 \in \mathbb{R}$

Let us translate these conditions into the number-free language.[5] The function $f\colon \mathbb{R} \to [0,1]$ represents a standard fuzzy set of reals, i.e., in FCT over any expansion of MTL_Δ, the standard model of a predicate A on reals.[6] By the standard semantics of MTL_Δ, the above conditions translate into the following axioms on the predicate A:[7]

$$(\forall xy)(x \leq y \to (Ax \to Ay)) \tag{10}$$

$$\neg(\forall x)Ax,\ (\exists x)Ax \tag{11}$$

$$(\forall x_0)[(\forall x > x_0)Ax \to Ax_0] \tag{12}$$

In FCT, these conditions express, respectively, the *upperness* of A in $\mathbb{R}$, the *null plinth* and *full height* of A, and the *left-closedness* of the fuzzy upper class A. Consequently, number-free distribution functions are left-closed upper sets in $\mathbb{R}$ with full height and null plinth, i.e., (weakly bounded) *fuzzy Dedekind cuts on* $\mathbb{R}$. The number-free rendering of distribution functions as fuzzy Dedekind cuts corresponds to the known fact that distribution functions represent Hutton fuzzy reals (cf., e.g., [15]). A use of fuzzy Dedekind cuts for the development of a logic-based theory of fuzzy intervals (or fuzzy numbers) is hinted at in [16].

5 Continuous functions on $\mathbb{R}$

In Section 4 we abused the presence of monotony for the number-free rendering of right-continuity. Yet, if we aim at a graded theory where monotony can be satisfied to partial degrees, we need a different number-free characterization of continuity that does not rely on monotony. We shall work with functions $\mathbb{R} \to [0,1]$ only, even though various generalizations are easy to obtain.

For a number-free rendering of right-continuity, we shall use the following classical characterization. A function $f\colon \mathbb{R} \to [0,1]$ is right-continuous in x_0 iff

$$\limsup_{x\to x_0^+} f(x) = \liminf_{x\to x_0^+} f(x) = f(x_0), \tag{13}$$

where

$$\limsup_{x\to x_0^+} f(x) = \inf_{x_1>x_0} \sup_{x_1>x>x_0} f(x) \tag{14}$$

$$\liminf_{x\to x_0^+} f(x) = \sup_{x_1>x_0} \inf_{x_1>x>x_0} f(x). \tag{15}$$

This translates into the following number-free definitions in FCT over MTL_Δ:[8]

$$\mathrm{LimSup}^+(A, x_0) \equiv_{\mathrm{df}} (\forall x_1 > x_0)(\exists x \in (x_0, x_1))Ax \tag{16}$$

$$\mathrm{LimInf}^+(A, x_0) \equiv_{\mathrm{df}} (\exists x_1 > x_0)(\forall x \in (x_0, x_1))Ax \tag{17}$$

$$\Delta\,\mathrm{Cont}^+(A, x_0) \equiv_{\mathrm{df}} (Ax_0 = \mathrm{LimSup}^+(A, x_0))\ \&\ (Ax_0 = \mathrm{LimInf}^+(A, x_0)) \tag{18}$$

The left-sided predicates LimSup^- and LimInf^- are defined dually (with $<$ for $>$), and the both-sided ones as

$$\mathrm{LimSup}(A, x_0) \equiv_{\mathrm{df}} \mathrm{LimSup}^-(A, x_0) \vee \mathrm{LimSup}^+(A, x_0) \tag{19}$$

$$\mathrm{LimInf}(A, x_0) \equiv_{\mathrm{df}} \mathrm{LimInf}^-(A, x_0) \wedge \mathrm{LimInf}^+(A, x_0). \tag{20}$$

These definitions reconstruct the classical notions in a number-free way in the framework of FCT; the representation theorems follow directly from the above considerations and the standard semantics of MTL_Δ. The following observation shows that many properties of the classical notions can be reconstructed in FCT as well.

Observation 5.1. By shifts of crisp relativized quantifiers valid in first-order MTL_Δ, the following theorems are easily provable in FCT:[9]

1. $\mathrm{LimInf}(A, x_0) \leq \mathrm{LimSup}(A, x_0)$
2. $A \subseteq B \to (\mathrm{LimSup}(A, x_0) \to \mathrm{LimSup}(B, x_0))$ and analogously for LimInf.
3. $\mathrm{LimInf}(\smallsetminus A, x_0) \leq \neg\,\mathrm{LimSup}(A, x_0)$
   $\mathrm{LimSup}(\smallsetminus A, x_0) \geq \neg\,\mathrm{LimInf}(A, x_0)$
   (Equality holds in logics with involutive negation, but not generally in MTL_Δ.)
4. $\mathrm{LimInf}(A \sqcap B, x_0) = \mathrm{LimInf}(A, x_0) \wedge \mathrm{LimInf}(B, x_0)$
   $\mathrm{LimInf}(A \sqcup B, x_0) = \mathrm{LimInf}(A, x_0) \vee \mathrm{LimInf}(B, x_0)$
   and analogously for LimSup.

Since the definitions reconstruct classical notions, we have retained the classical terminology and notation referring to limits and continuity, even though these regard membership functions (i.e., semantic *models* of fuzzy classes), rather than fuzzy classes themselves. In FCT, the fuzzy predicate $\mathrm{LimSup}^+(A, x_0)$ actually expresses the condition that x_0 is a *right-limit point* of the fuzzy class A, and $\mathrm{LimInf}^+(A, x_0)$ that x_0 is an *interior point* of $A \cup \{x_0\}$ in the left half-open interval topology,[10] as these are the properties expressed by the defining formulae if all sets involved are crisp.

[5] Of course, the 'translation' is by no means unique: we always select one which is sufficiently straightforward and which results in *meaningful* concepts of fuzzy mathematics.

[6] Recall that $\mathbb{R}$ as well as other crisp mathematical structures are available in FCT by means of the Δ-interpretation, see [2, §7] and [14, §4]. In this and the next section, we shall understand all first-order quantifications relativized to $\mathbb{R}$, unless specified otherwise.

[7] We use the fact that due to the monotony assumed, the margin conditions reduce to $\inf_x f(x) = 0$ and $\sup_x f(x) = 1$, and the right-continuity to $f(x_0) \geq \inf_{x>x_0} f(x)$. The representation theorem is then immediate by the standard semantics of MTL_Δ. Observe that as the axioms are required to degree 1, they are (due to the crispness of $\leq$) independent of the particular left-continuous t-norm used.

[8] Again it can be observed that the definitions are independent of a particular t-norm and are the same in all expansions of MTL_Δ.

[9] The theorems are stated for both-sided limits only, but hold equally well for one-sided limits.

[10] I.e., the topology with the open base of all half-open intervals $[a, b)$, also known as the lower-limit topology or the Sorgenfrey line.

ISBN: 978-989-95079-6-8

Consequently, the formulae $(\forall x_0 \in A)\,\mathrm{LimInf}(A, x_0)$ and $(\forall x_0 \in A)\,\mathrm{LimSup}(A, x_0)$ express the notions of *openness* resp. *closedness* of A in a fuzzy interval topology on $\mathbb{R}$. The study of this fuzzy topology and its relationship to the fuzzy interval topologies of [17, 18] is left for future work.

Like the definitions of LimInf and LimSup, the theorems of Observation 5.1 have double meanings, too. On the one hand they can be understood as number-free reconstructions and graded generalizations of the classical theorems on (membership) functions. On the other hand, they can be interpreted as fuzzy-mathematical theorems on (fuzzy) sets, under the above fuzzy interval topology on reals. In particular, 1. says that an interior point of a fuzzy set of reals is also its limit point; 2. that a limit point of a fuzzy set is also a limit point of a larger fuzzy set (and dually for interior points); 3. that an interior point of the complement of a fuzzy set A is not a limit point of A (and vice versa); and 4. that x_0 is a limit point of $A \sqcap B$ exactly to the degree it is a limit point of A *and* ($\wedge$) a limit point of B (and dually for $\sqcup$). (The theorems are graded, thus 'is' represents fuzzy implication $\rightarrow$.)

6 Operations on $\mathbb{R}$

In the previous examples, only the codomain of real-valued functions of reals was rendered numberless. Obviously, the domain $\mathbb{R}$ (or more conveniently, $\mathbb{R} \cup \{\pm\infty\}$) can be re-scaled into $[0, 1]$ and regarded as the standard set of truth degrees as well. The functions $\mathbb{R}^n \rightarrow \mathbb{R}$ then become n-ary functions from truth values to truth values, i.e., truth functions of fuzzy propositional connectives.

An apparatus for internalizing truth values and logical connectives in FCT was developed in [4, §3]. As shown there, the truth values can be internalized as the elements of the crisp class $\mathrm{L} = \mathrm{Ker}\,\mathrm{Pow}\{a\}$, i.e., subclasses of a fixed crisp singleton. The class L of internalized truth values is ordered by crisp inclusion $\subseteq^{\Delta}$, and the correspondence between internal and semantical truth values is given as follows:[11] $\alpha \in \mathrm{L}$ represents the semantic truth value of $\emptyset \in \alpha$, and the semantic truth value of φ is represented by the class $\overline{\varphi} =_{\mathrm{df}} \{a \mid \varphi\}$; the correspondences $\varphi \leftrightarrow (a \in \overline{\varphi})$ and $\overline{\varphi} \subseteq \overline{\psi} \leftrightarrow (\varphi \rightarrow \psi)$ then hold.

Logical connectives are then internalized by crisp functions $\mathbf{c}\colon \mathrm{L}^n \rightarrow \mathrm{L}$ (which can be called *internal, inner,* or *formal* connectives). In particular, definable connectives c of the logic are represented by the corresponding class operations $\overline{c} = \{x \in \mathrm{L} \mid c(x \in X_1, \ldots, x \in X_n)\}$ on L (e.g., & by $\cap$; $\vee$ by $\sqcup$; etc.). Since n-ary internal connectives are crisp functions valued in L, they can as well be regarded as fuzzy subsets of L^n, i.e., n-ary fuzzy relations on L. Usual fuzzy class operations and graded predicates, e.g., the graded inclusion

$$\mathbf{c} \subseteq \mathbf{d} \equiv_{\mathrm{df}} (\forall x_1 \ldots x_n)(\mathbf{c}x_1 \ldots x_n \rightarrow \mathbf{d}x_1 \ldots x_n), \quad (21)$$

then apply to them and make their theory graded.

The number-free theory of functions $\mathbb{R}^n \rightarrow \mathbb{R}$ is thus in fact the fuzzy-logical theory of internal connectives, i.e., of fuzzy relations on internal truth values. An elaboration of the theory of unary and binary internal connectives has been sketched in [19, 20]. These preliminary papers focus on the defining properties of *t-norms* (i.e., monotony, commutativity, associativity, and the unit) and the relation of domination between internal connectives, making them graded by reinterpretation of their defining formulae in fuzzy logic (cf. [2, §7], [5, §2.3], or [6, §4]) and studying their graded properties. A full paper on the topic (by the authors of [20]) is currently under construction.

7 Metrics

Recall that a *pseudometric*[12] on a set X is a function $d\colon X^2 \rightarrow [0, +\infty]$ such that

$$d(x, x) = 0 \quad (22)$$
$$d(x, y) = d(y, x) \quad (23)$$
$$d(x, z) \leq d(x, y) + d(y, z). \quad (24)$$

The numberless reduction will first need to normalize the range of pseudometrics from $[0, +\infty]$ to $[0, 1]$, e.g., by setting

$$c(x, y) = 2^{-d(x,y)}. \quad (25)$$

The defining conditions on pseudometric then become the following equivalent conditions on c:[13]

$$c(x, x) = 1 \quad (26)$$
$$c(x, y) = c(y, x) \quad (27)$$
$$c(x, z) \geq c(x, y) \cdot c(y, z) \quad (28)$$

These conditions are nothing else but the defining conditions of *fuzzy equivalences,* also known as *similarity relations* [21], in the standard semantics of product fuzzy logic [1]. We can thus equate number-free pseudometrics with *product similarities,* i.e., standard models of the following axioms in product fuzzy logic:

$$Cxx \quad (29)$$
$$Cxy \rightarrow Cyx \quad (30)$$
$$Cxy \,\&\, Cyz \rightarrow Cxz \quad (31)$$

The definition of a *metric* strengthens the first condition to $d(x, y) = 0$ iff $x = y$, which is equivalent to c being a *fuzzy equality* (also called *separated* similarity), i.e., $c(x, y) = 1$ iff $x = y$, thus replacing the first axiom by $Cxy \leftrightarrow x = y$.

Using a different left-continuous t-norm represents a different way of combining the distances $d(x, y)$ and $d(y, z)$ in the triangle inequality: e.g., with the minimum t-norm, c represents a *(pseudo)ultrametrics* under the same transformation (25), while with the Łukasiewicz t-norm, c represents a *bounded* pseudometric d under a different transformation $c(x, y) = (1 - d(x, y))/d_{\max}$, where $d_{\max} < +\infty$ is an upper bound on the distances. (The obvious both-way representation theorems are left to the reader.) Since furthermore many theorems on number-free metrics hold generally over MTL_{Δ}, it is quite natural to generalize the notion of number-free metrics to any similarity, not only the product one.

Various notions based on such (generalized) number-free metrics can be defined and their properties investigated in the framework of FCT (over MTL_{Δ} or stronger). Only a few

[11] However, see [4, Rem. 3.3] for certain metamathematical qualifications regarding this correspondence.

[12] For simplicity, we shall work with *extended* pseudometrics, allowing the value $+\infty$.

[13] Notice that since the function 2^{-x} reverses the order, the fuzzy relation $c\colon X^2 \rightarrow [0, 1]$ expresses *closeness* rather than distance.

ISBN: 978-989-95079-6-8

observations on number-free limits are given here as a concluding illustration of the numberless approach.

Fix a metric d rendered in the numberless way by a closeness predicate C under the transformation (25). The *limit* $\lim_{n\to\infty} x_n$ of a sequence $\{x_n\}_{n\in\mathbb{N}}$ (abbreviated $\vec{x}$) under C can be defined as follows:[14]

$$\Delta\mathrm{Lim}_C(\vec{x},x) \equiv_{\mathrm{df}} \Delta(\exists n_0)(\forall n > n_0)Cxx_n \qquad (32)$$

Theorem 7.1. *Standard models over product logic validate* $\Delta\mathrm{Lim}_C(\vec{x},x)$ *iff* $x = \lim_{n\to\infty} x_n$ *under* d.

Proof: $\lim x_n = x$ under d iff $\limsup d(x,x_n) = 0$, iff $\liminf 2^{-d(x,x_n)} = 1$, iff $\sup_{n_0} \inf_{n>n_0} c(x,x_n) = 1$, which is the semantics of $\Delta\mathrm{Lim}_C(\vec{x},x)$. □

As noted above, the meaning of $\Delta\mathrm{Lim}_C$ is natural not only in product logic, but also in other t-norm logics, esp. if the relation C is interpreted as *indistinguishability* rather than mere closeness:[15] then (32) expresses the condition that *from some index on, x_n is indistinguishable from x*. Similarly, Theorem 7.2 below expresses the fact that all limits of $\vec{x}$ are indistinguishable (to the degree the indistinguishability relation is symmetric and transitive).

Discarding the Δ in (32) yields a graded number-free notion of limit:[16]

$$\mathrm{Lim}_C(\vec{x},x) \equiv_{\mathrm{df}} (\exists n_0)(\forall n > n_0)Cxx_n \qquad (33)$$

Interestingly, $\mathrm{Lim}_C(\vec{x},x)$ coincides with G. Soylu's notion of *similarity-based fuzzy limit* [23]. Even without employing explicitly the formalism of t-norm fuzzy logic, Soylu was able to prove graded theorems such as [23, Prop. 3.5],

$$\mathrm{Lim}_C(\vec{x},x)\ \&\ \mathrm{Lim}_C(\vec{y},y) \to \mathrm{Lim}_C(\vec{x}+\vec{y},x+y). \qquad (34)$$

With the apparatus of FCT, the gradedness of Soylu's results can be extended even further by not requiring the full satisfaction of the defining properties of the similarity C (this conforms to the standard methodology of constructing graded theories [6, §7]). An example of such graded results is the following theorem on the fuzzy uniqueness of the limit:

Theorem 7.2. FCT *over* MTL_Δ *proves:*

$$\mathrm{Sym}\,C\ \&\ \mathrm{Trans}\,C\ \&\ \mathrm{Lim}_C(\vec{x},x)\ \&\ \mathrm{Lim}_C(\vec{x},x') \to Cxx' \qquad (35)$$

Proof. By $\mathrm{Trans}\,C$ we obtain $Cxx_n\ \&\ Cx_nx' \to Cxx'$; thus $Cx_nx\ \&\ Cx_nx' \to Cxx'$ by $\mathrm{Sym}\,C$, whence

$$((n > n_0) \to Cx_nx)\ \&\ ((n > n_0) \to Cx_nx') \to Cxx' \qquad (36)$$

follows propositionally. By generalization on n and distribution of the quantifier,

$$(\forall n > n_0)Cx_nx\ \&\ (\forall n > n_0)Cx_nx' \to Cxx' \qquad (37)$$

is obtained (as in the consequent the quantification is void). Generalization on n_0 and the shift of the quantifier to the antecedent (as $\exists$) then yields the required formula. □

[14] The Δ in $\Delta\mathrm{Lim}_C$ refers to the Δ in the defining formula, which will later be dropped.

[15] Cf. Menger's probabilistic indistinguishability relations [22].

[16] Observe that Lim_C is a Σ_2-formula: compare it with the classical Π_3-definition and the Π_1-definition in non-standard analysis.

A more detailed investigation of convergence based on fuzzy indistinguishability in the formal framework of FCT exceeds the scope of the present paper, and is therefore left for future research.

Appendix: Fuzzy Class Theory

Fuzzy Class Theory (FCT) is a formal theory over a given first-order fuzzy logic aiming at axiomatic approximation of Zadeh's fuzzy sets of all orders over a fixed crisp domain. It can be characterized as Henkin-style higher-order fuzzy logic, or fuzzified Russell-style simple type theory. FCT can be regarded as a foundational theory for fuzzy mathematics [24], since other axiomatic mathematical theories over fuzzy logic can be formalized within its framework. For more details on FCT see [2, 5]; the relevant definitions of [5] are briefly repeated here for reference.

The reader's familiarity with the logic MTL_Δ and its main extensions is assumed; for details on these logics see [1, 25]. Here we only recapitulate its standard real-valued semantics, which is crucial for number-free mathematics:

$\&$	...	a left-continuous t-norm $*$
$\to$	...	the residuum $\Rightarrow$ of $*$, defined as $x \Rightarrow y =_{\mathrm{df}} \sup\{z \mid z * x \le y\}$
$\wedge, \vee$	...	$\min, \max$
$\neg$	...	$x \Rightarrow 0$
$\leftrightarrow$	...	the bi-residuum: $\min(x \Rightarrow y, y \Rightarrow x)$
Δ	...	$\Delta x = 1 - \mathrm{sgn}(1-x)$
$\forall, \exists$	...	$\inf, \sup$

Łukasiewicz logic further specifies $x * y = (x + y - 1) \vee 0$, product logic sets $x * y = x \cdot y$, and Gödel logic sets $x * y = x \wedge y$.

Fuzzy Class Theory FCT is a formal theory over a given multi-sorted first-order fuzzy logic $\boldsymbol{L}$ (at least as strong as MTL_Δ), with sorts of variables for: atomic objects (lowercase letters $x, y, \ldots$), fuzzy classes of atomic objects (uppercase letters $A, B, \ldots$), fuzzy classes of fuzzy classes of atomic objects (calligraphic letters $\mathcal{A}, \mathcal{B}, \ldots$), etc., in general for fuzzy classes of the n-th order $(X^{(n)}, Y^{(n)}, \ldots)$.

Besides the crisp identity predicate $=$, the language of FCT contains:

- The membership predicate $\in$ between objects of successive sorts
- The class terms $\{x \mid \varphi\}$ of order $n+1$, for any variable x of any order n and any formula φ
- The symbols $\langle x_1, \ldots, x_k\rangle$ for k-tuples of individuals $x_1, \ldots, x_k$ of any order

In formulae of FCT we employ usual abbreviations and defined notions known from classical mathematics or traditional fuzzy mathematics, including those listed in Table 1, for all orders of fuzzy classes.

FCT has the following axioms, for all formulae φ and variables of all orders:

- The logical axioms of multi-sorted first-order logic $\boldsymbol{L}$
- The axioms of crisp identity: $x = x$; $x = y\ \&\ \varphi(x) \to \varphi(y)$; and $\langle x_1, \ldots, x_k\rangle = \langle y_1, \ldots, y_k\rangle \to x_i = y_i$

ISBN: 978-989-95079-6-8

Table 1: Abbreviations and defined notions of FCT

$\varphi = \psi$	$\equiv_{\mathrm{df}}$	$\Delta(\varphi \leftrightarrow \psi)$
$\varphi \leq \psi$	$\equiv_{\mathrm{df}}$	$\Delta(\varphi \rightarrow \psi)$
Ax	$\equiv_{\mathrm{df}}$	$x \in A$
$x_1 \ldots x_k$	$=_{\mathrm{df}}$	$\langle x_1, \ldots, x_k \rangle$
$(\forall x \in A)\varphi$	$\equiv_{\mathrm{df}}$	$(\forall x)(x \in A \rightarrow \varphi)$
$(\exists x \in A)\varphi$	$\equiv_{\mathrm{df}}$	$(\exists x)(x \in A \,\&\, \varphi)$
$\{x \in A \mid \varphi\}$	$=_{\mathrm{df}}$	$\{x \mid x \in A \,\&\, \varphi\}$
$\emptyset$	$=_{\mathrm{df}}$	$\{x \mid 0\}$
$\mathrm{Ker}\, A$	$=_{\mathrm{df}}$	$\{x \mid \Delta Ax\}$
$\smallsetminus A$	$=_{\mathrm{df}}$	$\{x \mid \neg Ax\}$
$A \cap B$	$=_{\mathrm{df}}$	$\{x \mid Ax \,\&\, Bx\}$
$A \sqcap B$	$=_{\mathrm{df}}$	$\{x \mid Ax \wedge Bx\}$
$A \sqcup B$	$=_{\mathrm{df}}$	$\{x \mid Ax \vee Bx\}$
$\mathrm{Pow}\, A$	$=_{\mathrm{df}}$	$\{X \mid X \subseteq A\}$
$\mathrm{Hgt}\, A$	$\equiv_{\mathrm{df}}$	$(\exists x)Ax$
$\mathrm{Plt}\, A$	$\equiv_{\mathrm{df}}$	$(\forall x)Ax$
$\mathrm{Crisp}\, A$	$\equiv_{\mathrm{df}}$	$(\forall x)\Delta(Ax \vee \neg Ax)$
$\mathrm{Sym}\, R$	$\equiv_{\mathrm{df}}$	$(\forall xy)(Rxy \rightarrow Ryx)$
$\mathrm{Trans}\, R$	$\equiv_{\mathrm{df}}$	$(\forall xyz)(Rxy \,\&\, Ryz \rightarrow Rxz)$
$\mathrm{Fnc}\, R$	$\equiv_{\mathrm{df}}$	$(\forall xyy')(Rxy \,\&\, Rxy' \rightarrow y = y')$
$A \subseteq B$	$\equiv_{\mathrm{df}}$	$(\forall x)(Ax \rightarrow Bx)$
$A \subseteq^{\Delta} B$	$\equiv_{\mathrm{df}}$	$(\forall x)(Ax \leq Bx)$

- The comprehension axioms: $y \in \{x \mid \varphi(x)\} \leftrightarrow \varphi(y)$
- The extensionality axioms: $(\forall x)(Ax = Bx) \rightarrow A = B$

Models of FCT are systems of fuzzy sets (and fuzzy relations) of all orders over a crisp universe of discourse, with truth degrees taking values in an $\boldsymbol{L}$-chain $\mathbf{L}$ (e.g., the interval $[0, 1]$ equipped with a left-continuous t-norm); thus all theorems on fuzzy classes provable in FCT are true statements about $\mathbf{L}$-valued fuzzy sets. Note, however, that the theorems of FCT are derived from its axioms by the rules of the fuzzy logic $\boldsymbol{L}$ rather than classical Boolean logic. For details on proving theorems in FCT see [5] or [26].

Acknowledgment

This work was supported by grant No. A100300503 of GA AV ČR and Institutional Research Plan AV0Z10300504.

References

[1] Petr Hájek. *Metamathematics of Fuzzy Logic*, volume 4 of *Trends in Logic*. Kluwer, Dordercht, 1998.

[2] Libor Běhounek and Petr Cintula. Fuzzy class theory. *Fuzzy Sets and Systems*, 154(1):34–55, 2005.

[3] Libor Běhounek, Ulrich Bodenhofer, and Petr Cintula. Relations in Fuzzy Class Theory: Initial steps. *Fuzzy Sets and Systems*, 159:1729–1772, 2008.

[4] Libor Běhounek and Martina Daňková. Relational compositions in Fuzzy Class Theory. To appear in *Fuzzy Sets and Systems*, doi:10.1016/j.fss.2008.06.013.

[5] Libor Běhounek and Petr Cintula. Fuzzy Class Theory: A primer v1.0. Technical Report V–939, Institute of Computer Science, Academy of Sciences of the Czech Republic, 2006.

[6] Libor Běhounek and Petr Cintula. Features of mathematical theories in formal fuzzy logic. *Lecture Notes in Artificial Intelligence*, 4529:523–532, 2007. (Proc. IFSA Congr. 2007.)

[7] Libor Běhounek. Number-free reductions in logic-based fuzzy mathematics. Technical Report V–1032, Institute of Computer Science, Academy of Sciences of the Czech Republic, 2008.

[8] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[9] Louise Schmir Hay. Axiomatization of the infinite-valued predicate calculus. *Journal of Symbolic Logic*, 28:77–86, 1963.

[10] Petr Hájek, Lluís Godo, and Francesc Esteva. Fuzzy logic and probability. *Proc. 11th UAI*, pages 237–244, Montreal, 1995.

[11] Lluís Godo, Francesc Esteva, and Petr Hájek. Reasoning about probability using fuzzy logic. *Neural Network World*, 10:811–823, 2000. (Spec. issue SOFSEM 2000.)

[12] Tommaso Flaminio. Strong non-standard completeness for fuzzy logics. *Soft Computing*, 12:321–333, 2008.

[13] Tommaso Flaminio and Lluís Godo. A logic for reasoning about the probability of fuzzy events. *Fuzzy Sets and Systems*, 158:625–638, 2006.

[14] Petr Cintula and Rostislav Horčík. Fuzzy Class Theory: Some advanced topics. *Proc. 5th EUSFLAT Conf.*, volume I, pages 137–143. University of Ostrava, 2007.

[15] Ulrich Höhle. Fuzzy real numbers as Dedekind cuts with respect to a multiple-valued logic. *Fuzzy Sets and Systems*, 24(3):263–278, 1987.

[16] Libor Běhounek. Towards a formal theory of fuzzy Dedekind reals. *Proc. Joint 4th Conf. of EUSFLAT and 11th LFA*, pages 946–954, Barcelona, 2005.

[17] Libor Běhounek and Tomáš Kroupa. Topology in Fuzzy Class Theory: Basic notions. *Lecture Notes in Artificial Intelligence*, 4529:513–522, 2007. (Proc. IFSA Congr. 2007.)

[18] Libor Běhounek and Tomáš Kroupa. Interior-based topology in Fuzzy Class Theory. *Proc. 5th EUSFLAT Conf.*, volume I, pages 145–151. University of Ostrava, 2007.

[19] Libor Běhounek, Ulrich Bodenhofer, Petr Cintula, and Susanne Saminger-Platz. Graded dominance. *Abstracts of the 29th Linz Seminar on Fuzzy Set Theory*, pages 11–14, Linz, 2008.

[20] Libor Běhounek, Ulrich Bodenhofer, Petr Cintula, Peter Sarkoci, and Susanne Saminger-Platz. Graded properties of binary fuzzy connectives. Technical Report V–1042, Institute of Computer Science, Academy of Sciences of the Czech Republic, 2008.

[21] Lotfi A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

[22] Karl Menger. Probabilistic theories of relations. *Proceedings of the National Academy of Sciences*, 37:178–180, 1951.

[23] Gültekin Soylu. Similarity-based fuzzy limits. *Fuzzy Sets and Systems*, 159:3380–3387, 2008.

[24] Libor Běhounek and Petr Cintula. From fuzzy logic to fuzzy mathematics: A methodological manifesto. *Fuzzy Sets and Systems*, 157(5):642–646, 2006.

[25] Francesc Esteva and Lluís Godo. Monoidal t-norm based logic: Towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124(3):271–288, 2001.

[26] Libor Běhounek and Petr Cintula. Fuzzy class theory as foundations for fuzzy mathematics. *Proc. 11th IFSA World Congress*, volume 2, pages 1233–1238, Beijing, 2005. Tsinghua University Press & Springer.

ISBN: 978-989-95079-6-8

Uncertainty Can Decrease Privacy: An Observation

Karen Villaverde[1] Olga Kosheleva[2]

1. Department of Computer Science
New Mexico State University
Las Cruces, NM 88003, USA
2. Department of Teacher Education
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
Email: kvillave@cs.nmsu.edu, olgak@utep.edu

Abstract— *In many application areas, e.g., in medical applications, it is important to be able to get statistical information about the data without disclosing individual cases. One of the methods to preserve privacy in such statistical databases is to introduce uncertainty, i.e., to replace the exact values with intervals of possible values. In this paper, we show that while the introduction of uncertainty can often enhance privacy, sometimes, the opposite effect occurs: adding uncertainty can decrease privacy.*

Keywords— Fuzzy uncertainty, interval uncertainty, privacy, statistical databases

1 Statistical Databases – Need to Preserve Privacy

1.1 Need for Collecting Data

In many practical situations, it is very useful to collect large amounts of data.

For example, from the data that we collect during a census, we can extract a lot of information about health, mortality, employment in different regions – for different age ranges, and for people from different genders and of different ethnic groups. By analyzing these statistics, we can reveal troubling spots and allocate (usually limited) resources so that the help goes first to social groups that need it most.

Similarly, by gathering data about people's health in a large medical database, we can extract a lot of useful information on how the geographic location, age, and gender affect a person's health. Then, we can improve the public health by appropriate appropriate public health measures to different portions of the population.

Finally, a large database of purchases can help find out what people are looking for, make shopping easier for customers and at the same time, decrease the stores' expenses related to storing unnecessary items.

1.2 Need for Privacy

Privacy is an important issue in the statistical analysis of human-related data. For example, to check whether in a certain geographic area, there is a gender-based discrimination, one can use the census data to check, e.g., whether for all people from this area who have the same level of education, there is a correlation between salary and gender. One can think of numerous possible questions of this type related to different sociological, political, medical, economic, and other questions. From this viewpoint, it is desirable to give researchers *ability to perform* whatever *statistical analysis* of this data that is reasonable for their specific research.

On the other hand, we may not want to give the researchers direct access to the raw census data, because a large part of the census data is *confidential*. For example, for most people (those who work in the private sector) salary information is confidential. Suppose that a corporation is deciding where to built a new plant and has not yet decided between two possible areas. This corporation would benefit from knowing the average salary of people of needed education level in these two areas, because this information would help them estimate how much it will cost to bring local people on board. However, since salary information is confidential, the company should not be able to know the exact salaries of different potential workers.

The need for privacy is also extremely important for *medical* experiments, where we should be able to make statistical conclusions about, e.g., the efficiency of a new medicine without disclosing any potentially embarrassing details from the individual medical records.

Such databases in which the outside users cannot access individual records but can solicit statistical information are often called *statistical databases*.

1.3 Maintaining Privacy is Not Easy

Maintaining privacy in statistical databases is not easy. Clerks who set up policies on access to statistical databases sometimes erroneously assume that once the records are made anonymous, we have achieved perfect privacy. Alas, the situation is not so easy: even when we keep all the records anonymous, we can still extract confidential information by asking appropriate questions.

Many examples of such extraction can be found in a book by D. Denning [1]. For example, suppose that we are interested in the salary of Dr. X who works for a local company. Dr. X's mailing address can be usually taken from the phone book; from the company's webpage, we can often get his photo and thus find out his race and approximate age. Then, to determine Dr. X's salary, all we need is to ask what is the average salary of all people with a Ph.D. of certain age brackets who live in a small geographical area around his actual home address – if the area is small enough, then Dr. X will be the

ISBN: 978-989-95079-6-8

only person falling under all these categories.

Even if we only allow statistical information about salaries $s_1, \ldots, s_q$ when there are at least a certain amount n_0 people within a requested range, we will still be able to reconstruct the exact salaries of all these people. Indeed, for example, we can ask for the number and average salary of all the people who live on Robinson street at houses 1 through 1001, and then we can ask the same question about all the people who live in houses from 1 to 1002. By comparing the two numbers, we get the average salary of the family living at 1002 Robinson – in other words, we gain the private information that we tried to protect.

In general, we can ask for the average

$$\frac{s_1 + \ldots + s_q}{q},$$

and for several moments of salary (variance, third moment, etc):

- if we know the values v_j of at least q different functions $f_j(s_1, \ldots, s_q)$ of s_i,
- then we can, in general, reconstruct all these values from the corresponding system of q equations with q unknowns:

$$f_1(s_1 \ldots, s_q) = v_1,$$

$$\ldots,$$

$$f_q(s_1, \ldots, s_q) = v_q.$$

At first glance, moments are natural and legitimate statistical characteristics, so researchers would be able to request these characteristics. On the other hand, we do not want the researchers to be able to extract the exact up-to-cent salaries of all the people leaving in a certain geographical area.

What restriction should we impose on possible statistical queries that would guarantee privacy but restrict research in the least possible way?

These are anecdotal examples of poorly designed privacy and security, but, as we have mentioned, the problem is indeed difficult: Many seemingly well-designed privacy schemes later turn out to have unexpected privacy and security problems, and it is known that the problem of finding a privacy-preserving scheme is, in general, NP-hard [1].

Different aspects of the problem of privacy in statistical databases, different proposed solutions to this problem, and their drawbacks, are described in [1, 11, 12] (see also references therein).

2 Known Fact: Uncertainty Can Enhance Privacy

A reasonable way to avoid privacy violations is to store ranges (intervals) of values instead of the actual values. For example, instead of keeping the exact age, we only record whether the age is between 0 and 10, 10 and 20, 20 and 30, etc.

In this case, no matter what statistics we allow, the worst that can happen is that the corresponding ranges will be disclosed. However, in this situation, we do not disclose the original exact values – since these values are not stored in the database in the first place; see, e.g., [5, 6, 7].

3 New Observation: Uncertainty Can Decrease Privacy

3.1 What We Do in This Section

The successful use of uncertainty to enhance privacy in statistical databases may lead to an impression that uncertainty *always* enhances privacy. In this paper, we show that uncertainty can actually *decrease* privacy in a statistical database.

3.2 Simplest Possible Case: 1-D Databases

We will show that the privacy decrease phenomenon occurs in the simplest 1-dimensional case, when each record in the statistical database consists of a single value x_i. In this case, the database consists of n values $x_1, \ldots, x_n$.

3.3 1-D Case: Choice of Statistical Characteristics

Among the most important statistical characteristics are the mean values of certain quantities.

In the 1-D case, each record contains the value x_i of a single quantity x. So, in the 1-dimensional case, we are interested in estimating the mean values of different characteristics $u(x)$, i.e., in estimating the values

$$E[u] \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} u(x_i). \tag{1}$$

3.4 Smooth Characteristics

Among different statistical characteristics, an important class is formed by characteristics for which the corresponding function $u(x)$ is smooth.

Usually, most values x_i corresponding to different individuals do not differ much. In other words, most values x_i belong to a small-size region. In this region, we can expand the function $u(x)$ in Taylor series and ignore higher-order terms. In the first approximation, we thus approximate the function $u(x)$ by a linear expression

$$u(x) \approx u_0 + u_1 \cdot x. \tag{2}$$

In the next approximation, we approximate an arbitrary smooth function $u(x)$ by a quadratic expression:

$$u(x) \approx u_0 + u_1 \cdot x + u_2 \cdot x^2. \tag{3}$$

In this paper, we will show that the phenomenon of decreasing privacy occurs already for such quadratic characteristics.

3.5 For Exactly Known Values, the Computation of Quadratic Characteristics Preserves Privacy

In applications, we may be interested in the mean values of different statistical characteristics. If we restrict ourselves to quadratic characteristics, this means that we must be able, given an arbitrary quadratic characteristic (3), to estimate the average value (1) of this characteristic.

Substituting the general expression (3) into the formula (1), we can conclude that the desired average value has the form

$$E[u] = u_0 + u_1 \cdot E[x] + u_2 \cdot E[x^2], \tag{4}$$

where

$$E[x] \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \tag{5}$$

ISBN: 978-989-95079-6-8

and

$$E[x^2] \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2. \tag{6}$$

Thus, in effect, even when we know the mean value $E[u]$ for all possible quadratic characteristics, all we know about the values x_i are the two combinations: the mean $E[x]$ and the second moment $E[x^2]$.

Based on the knowledge of these two values, we cannot reconstruct n different values $x_1, \ldots, x_n$ and thus, privacy is preserved.

3.6 Case of Interval Uncertainty

Let us now consider what happens if instead of the exact value of x_i, we only know the *interval* $[\underline{x}_i, \overline{x}_i]$ of possible values of x_i.

Often, these intervals come from the fact that we know an approximate value $\widetilde{x}_i$, and we know the upper bound Δ_i on the approximation accuracy $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$: $|\Delta x_i| \le \Delta_i$. In this case, the only information that we have about the actual (unknown) value x_i is that x_i belongs to the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$. So, we get an interval $[\underline{x}_i, \overline{x}_i]$ with $\underline{x}_i = \widetilde{x}_i - \Delta_i$ and $\overline{x}_i = \widetilde{x}_i + \Delta_i$.

In general, every interval $[\underline{x}_i, \overline{x}_i]$ can be represented in the form $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$: it is sufficient to take:

- as $\widetilde{x}_i$, the midpoint of the interval, i.e., the value

$$\widetilde{x}_i = \frac{\underline{x}_i + \overline{x}_i}{2}; \tag{7}$$

  and

- as Δ_i, the radius (half-width) of the interval, i.e., the value

$$\Delta_i = \frac{\overline{x}_i - \underline{x}_i}{2}. \tag{8}$$

3.7 Estimating Statistical Characteristics under Interval Uncertainty: Formulation of the Problem

Suppose that we are interested in the mean value $E[u]$ of a statistical characteristic $u(x)$. In the ideal case when we know the exact values $x_1, \ldots, x_n$, this mean value is simply equal to the arithmetic average of the corresponding values $u(x_i)$.

Under interval uncertainty, we can have different values $x_i \in [\underline{x}_i, \overline{x}_i]$. In general, different values x_i can lead to different values of the mean. We are interested in computing the *range* $[\underline{u}, \overline{u}]$ of possible values of this mean, i.e., the interval

$$[\underline{u}, \overline{u}] \stackrel{\text{def}}{=} \left\{ \frac{1}{n} \cdot \sum_{i=1}^{n} u(x_i) : x_i \in [\underline{x}_i, \overline{x}_i] \right\}. \tag{9}$$

3.8 Estimating Quadratic Characteristics under Interval Uncertainty: Algorithm

How can we compute the range of the mean

$$E[u] = \frac{1}{n} \cdot \sum_{i=1}^{n} u(x_i)?$$

Each of the variables x_i can take any value from the corresponding interval $[\underline{x}_i, \overline{x}_i]$. Thus,

- the mean $E[u]$ takes the *largest* value if and only if each term $u(x_i)$ takes the *largest* possible value on the interval $[\underline{x}_i, \overline{x}_i]$; we will denote this largest value by $\overline{u}_i$;
- similarly, the mean $E[u]$ takes the *smallest* value if and only if each term $u(x_i)$ takes the *smallest* possible value on the interval $[\underline{x}_i, \overline{x}_i]$; we will denote this smallest value by $\underline{u}_i$.

For a given quadratic function $u(x_i)$ on a given interval $[\underline{x}_i, \overline{x}_i]$, computing its smallest value $\underline{u}_i$ and its largest value $\overline{u}_i$ is an easy computational task. Indeed, according to calculus, if a function attains its minimum or maximum at some point inside an interval, then its derivative is 0 at this point. Thus, to find the minimum and the maximum of a given function $u(x_i)$ on a given interval $[\underline{x}_i, \overline{x}_i]$, it is sufficient to compute the values of this function

- at the endpoints $\underline{x}_i$ and $\overline{x}_i$ of this interval and
- at the point(s) (if any) where the derivative $u'(x)$ of the function $u(x)$ is equal to 0.

Then,

- the smallest of these values is the desired minimum $\underline{u}_i$, and
- the largest of these values is the desired maximum $\overline{u}_i$.

For a quadratic function $u(x) = u_0 + u_1 \cdot x + u_2 \cdot x^2$, the derivative is a linear function $u'(x) = u_1 + 2 \cdot u_2 \cdot x$, so computing the point where the derivative is 0 is a straightforward task: $u_1 + 2 \cdot u_2 \cdot x = 0$ implies $x = -u_1/(2u_2)$.

Once we compute the values $\underline{u}_i$ and $\overline{u}_i$, we can now find the desired range $[\underline{u}, \overline{u}]$ as

$$\underline{u} = \frac{1}{n} \cdot \sum_{i=1}^{n} \underline{u}_i, \ \ \overline{u} = \frac{1}{n} \cdot \sum_{i=1}^{n} \overline{u}_i. \tag{10}$$

3.9 Under Interval Uncertainty, Privacy Is No Longer Preserved: Result

Let us show that in this case, privacy is no longer preserved. To be more precise, we assume that we are given a 1-D database, i.e., a collection of intervals $[\underline{x}_1, \overline{x}_1], \ldots, [\underline{x}_n, \overline{x}_n]$. We do not have an explicit access to intervals from this collection, but for every quadratic function $u(x)$, we can generate the range of the mean value $E[u]$ over these intervals.

We will show that in almost all cases, based on these ranges, we can actually reconstruct all the original intervals $[\underline{x}_i, \overline{x}_i]$. In other words, adding uncertainty leads to a loss of privacy.

Comment. By "almost all" cases, we mean all cases in which all n midpoints $\widetilde{x}_i$ are different. Situations when two midpoints coincide are indeed degenerate, since a minor modification of the original data leads to $\widetilde{x}_i \neq \widetilde{x}_j$.

3.10 Under Interval Uncertainty, Privacy Is No Longer Preserved: Proof

In our proof, for every real number a, we consider the quadratic function $u(x) = (x - a)^2$. For this function, the derivative is equal to 0 at the minimum point $x = a$. Thus, this function attains its largest value on the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ at one of the endpoints $\widetilde{x}_i - \Delta_i$ or $\widetilde{x}_i + \Delta_i$. One can easily check that:

ISBN: 978-989-95079-6-8

- when $a \le \widetilde{x}_i$, then the largest possible value $\overline{u}_i$ of $u(x)$ on the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ is attained when $x_i = \overline{x}_i = \widetilde{x}_i + \Delta_i$ and is equal to $\overline{u}_i = (\overline{x}_i - a)^2$;
- when $a \ge \widetilde{x}_i$, then the largest possible value $\overline{u}_i$ of $u(x)$ on the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ is attained when $x_i = \underline{x}_i = \widetilde{x}_i - \Delta_i$ and is equal to $\overline{u}_i = (\underline{x}_i - a)^2$.

Let us use this fact to describe the dependence of $\overline{u}$ on the parameter a.

When $a \ne \widetilde{x}_i$, the value $\overline{u}$ is the average of n smooth expressions.

At each point $a = \widetilde{x}_i$, all the terms $\overline{u}_j$ in the sum $\overline{u}$ are smooth except for the term $\overline{u}_i$ that turns from $(\overline{x}_i - a)^2$ to $(\underline{x}_i - a)^2$. The derivative of $\overline{u}_i$ with respect to a changes from $2 \cdot (a - \overline{x}_i)$ to $2 \cdot (a - \underline{x}_i)$, i.e., increases by

$$2 \cdot (a - \underline{x}_i) - 2 \cdot (a - \overline{x}_i) = 2 \cdot (\overline{x}_i - \underline{x}_i) = 4 \cdot \Delta_i. \quad (11)$$

Since all the other components $\overline{u}_j$ are smooth at $a = \widetilde{x}_i$, at $a = \widetilde{x}_i$, the derivative of the average $\overline{u}(a)$ increases by $\frac{4}{n} \cdot \Delta_i$.

Thus, once we know the value $\overline{u}$ for all a,

- we can find the values $\widetilde{x}_i$ as the values at which the derivative is discontinuous; and
- we can find each value Δ_i as $n/4$ times the increase of the derivative at the corresponding point $\widetilde{x}_i$.

The statement is proven.

3.11 Case of Fuzzy Uncertainty

In many practical situations, the estimates $\widetilde{x}_i$ come from experts. Experts often describe the inaccuracy of their estimates in terms of imprecise words from natural language, such as "approximately 0.1", etc. A natural way to formalize such words is to use special techniques developed for formalizing this type of estimates – specifically, the technique of fuzzy logic; see, e.g., [4, 10].

In this technique, for each possible value of $x_i \in [\underline{x}_i, \overline{x}_i]$, we describe the degree $\mu_i(x_i)$ to which this value is possible. For each degree of certainty α, we can determine the set of values of x_i that are possible with at least this degree of certainty – the *α-cut* $\mathbf{x}_i(\alpha) = \{x \,|\, \mu_i(x) \ge \alpha\}$ of the original fuzzy set. Vice versa, if we know α-cuts for every α, then, for each object x, we can determine the degree of possibility that x belongs to the original fuzzy set [2, 4, 8, 9, 10]. A fuzzy set can be thus viewed as a nested family of its (interval) α-cuts.

3.12 From the Computational Viewpoint, Fuzzy Uncertainty Can Be Reduced to the Interval One

Once we know how to propagate interval uncertainty, i.e., how to compute the range

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=} \{f(x_1, \ldots, x_n) : x_i \in \mathbf{x}_i\} \quad (12)$$

of a given function $f(x_1, \ldots, x_n)$ over given intervals $\mathbf{x}_i$, then, to propagate the fuzzy uncertainty, we can consider, for each α, the fuzzy set y with the α-cuts

$$\mathbf{y}(\alpha) = f(\mathbf{x}_1(\alpha), \ldots, \mathbf{x}_n(\alpha)); \quad (13)$$

see, e.g., [2, 4, 8, 9, 10]. This is equivalent to using Zadeh's extension principle.

So, from the computational viewpoint, the problem of propagating fuzzy uncertainty can be reduced to several interval propagation problems.

For example, the fuzzy value of $E[u]$ can be described as follows: for each α, the corresponding α-cut is equal to the range of $E[u]$ when for each i, all values x_i belong to the corresponding α-cuts: $x_i \in \mathbf{x}_i(\alpha)$.

3.13 Under Fuzzy Uncertainty, Privacy Is Also Not Preserved: Result

Let us assume that we are given a 1-D database, i.e., a collection of fuzzy numbers $\mu_1(x_1), \ldots, \mu_n(x_n)$. We do not have an explicit access to fuzzy numbers from this collection, but for every quadratic function $u(x)$, we can generate the fuzzy number $\mu(u)$ corresponding to the mean $E[u]$.

We will show that in almost all cases (i.e., when the midpoints are all different), based on the resulting fuzzy numbers $\mu(u)$, we can actually reconstruct all the original fuzzy numbers $\mu_i(x_i)$.

In other words, adding fuzzy uncertainty also leads to a loss of privacy.

3.14 Under Fuzzy Uncertainty, Privacy Is Also Not Preserved: Proof

For each α, the α-cuts of the resulting fuzzy number $\mu(u)$ is the range of $E[u]$ when $x_i \in \mathbf{x}_i(\alpha)$.

We already know that from the ranges computed for all possible quadratic functions $u(x)$, we can reconstruct all the intervals. Thus, for every α, we can reconstruct all the α-cuts of all the membership functions $\mu_i(x_i)$.

Since we can thus reconstruct all α-cuts for all α, hence, we can uniquely reconstruct all the membership functions.

The statement is proven.

Acknowledgment

The authors are thankful to the anonymous referees for useful suggestions.

References

[1] D. Denning, *Cryptography and Data Security*, Addison-Wesley, Reading, MA, 1982.

[2] D. Dubois and H. Prade, Operations on fuzzy numbers, *International Journal of Systems Science*, 1978, Vol. 9, pp. 613–626.

[3] P. C. Fishburn, *Nonlinear preference and utility theory*, The John Hopkins Press, Baltimore, MD, 1988.

[4] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, Upper Saddle River, New Jersey, 1995.

[5] V. Kreinovich and L. Longpré, Computational complexity and feasibility of data processing and interval computations, with extension to cases when we have partial information about probabilities, In: V. Brattka, M. Schroeder, K. Weihrauch, and N. Zhong, editors, *Proc. Conf. on Computability and Complexity in Analysis CCA'2003*, Cincinnati, Ohio, USA, Aug. 28–30, 2003, pp. 19–54.

[6] V. Kreinovich, L. Longpré, S. A. Starks, G. Xiang, J. Beck, R. Kandathi, A. Nayak, S. Ferson, and J. Hajagos, "Interval Versions of Statistical Techniques, with Applications to Environmental Analysis, Bioinformatics, and Privacy in Statistical Databases", *Journal of Computational and Applied Mathematics*, 2007, Vol. 199, No. 2, pp. 418–423.

[7] L. Longpré, G. Xiang, V. Kreinovich, and E. Freudenthal, "Interval Approach to Preserving Privacy in Statistical Databases: Related Challenges and Algorithms of Computational Statistics", In: V. Gorodetsky, I. Kotenko, and V. A. Skormin (eds.), *Proceedings of the International Conference "Mathematical Methods, Models and Architectures for Computer Networks Security"*, MMM-ACNS-07, St. Petersburg, Russia, September 13–15, 2007, Springer Lecture Notes in Computer Science, 2007, Vol. CCIS-1, pp. 346–361.

[8] R. E. Moore and W. Lodwick, Interval Analysis and Fuzzy Set Theory, *Fuzzy Sets and Systems*, 2003, Vol. 135, No. 1, pp. 5–9.

[9] H. T. Nguyen and V. Kreinovich, Nested Intervals and Sets: Concepts, Relations to Fuzzy Sets, and Applications, In: R. B. Kearfott and V. Kreinovich, eds., *Applications of Interval Computations*, Kluwer, Dordrecht, 1996, pp. 245–290.

[10] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic*, CRC Press, Boca Raton, Florida, 2005.

[11] L. Sweeney, "Weaving technology and policy together to maintain confidentiality", *Journal of Law, Medicine and Ethics*, 1997, Vol. 25, pp. 98–110.

[12] L. Willenborg and T. De Waal, *Statistical disclosure control in practice*, Springer Verlag, New York, 1996.

ISBN: 978-989-95079-6-8

Decentralized Adaptive Fuzzy-Neural Control of an Anaerobic Digestion Bioprocess Plant

Ieroham S. Baruch[1] Rosalba Galvan-Guerra[1]

1 Department of Automatic Control, CINVESTAV-IPN, Mexico D.F., 07360 Mexico
Email: {baruch, rgalvan}@ctrl.cinvestav.mx

***Abstract*—** *The paper proposed to use recurrent Fuzzy-Neural Multi-Model (FNMM) identifier for decentralized identification of a distributed parameter anaerobic wastewater treatment digestion bioprocess, carried out in a fixed bed and a recirculation tank. The distributed parameter analytical model of the digestion bioprocess is used as a plant data generator. It is reduced to a lumped system using the orthogonal collocation method, applied in three collocation points (plus the recirculation tank), which are used as centers of the membership functions of the fuzzyfied plant output variables with respect to the space variable. The local and global weight parameters and states of the proposed FNMM identifier are used to design hierarchical FNMM direct and indirect controllers. The comparative graphical simulation results of the digestion system direct and indirect control, obtained via learning, exhibited a good convergence, and precise reference tracking. The comparative numerical results, giving the final means squared error of control of each output variable showed that the indirect adaptive decentralized fuzzy-neural control outperformed the direct one, and the it outperformed the linearized proportional optimal control too.*

***Keywords*—** Decentralized control, direct adaptive control, indirect adaptive control, distributed parameter digestion bioprocess system, recurrent neural network model, hierarchical fuzzy neural identification and control.

1 Introduction

The Distributed Parameter Systems (DPS) are distinguished by the fact that the states, controls, and outputs may depend on spatial position, [1]. Thus the natural form of the system description is by Partial Differential Equations (PDE), integral equations, or transcendental transfer functions, [1]. On the other side, in the last two decades, a new identification and control tools like Neural Networks (NN), used for biotechnological plants, [2], rose fame. Among several possible network architectures the ones most widely used are the Feedforward NN (FFNN) and the Recurrent NN (RNN), [3]. The main NN property namely the ability to approximate complex non-linear relationships without prior knowledge of the model structure makes them a very attractive alternative to the classical modeling and control techniques. This property has been proved for both types of NNs by the universal approximation theorem [3]. The preference given to NN identification with respect to the classical methods of process identification is clearly demonstrated in the solution of the "bias-variance dilemma" [3]. The FFNN and the RNN have been applied for DPS identification and control too. In [4], a RNN is used for system identification and process prediction of a DPS dynamics - an adsorption column for wastewater treatment of water contaminated with toxic chemicals. In [5, 6], a spectral-approximation-based intelligent modeling approach is proposed for the distributed thermal processing of the snap curing oven DPS that is used in semiconductor packaging industry. In [7], it is presented a new methodology for the identification of DPS, based on NN architectures, motivated by standard numerical discretization techniques used for the solution of PDE. In [8], an attempt is made to use the philosophy of the NN adaptive-critic design to the optimal control of distributed parameter systems. In [9] the concept of proper orthogonal decomposition is used for the model reduction of DPS to form a reduced order lumped parameter problem. In [10], measurement data of an industrial process are generated by solving the PDE numerically using the finite differences method. Both centralized and decentralized NN models are introduced and constructed based on this data. The models are implemented on FFNN using Backpropagation (BP) and Levenberg-Marquardt learning algorithms. In [11, 12, 13, 14], Baruch et al. defined direct and indirect Fuzzy Neural Multi-Model (FNMM) control system, based on Takagi-Sugeno (T-S) fuzzy rules, [15], containing in its consequent parts computational procedure of a Backpropagation Learning (BP) of a Recurrent Trainable NN (RTNN), [14]. In the present paper, the direct and indirect FNMM control system are modified and used for decentralized identification and control of a digestion anaerobic DPS of wastewater treatment, [16]. The anaerobic bioprocess plant model, used as an input/output plant data generator, is described by PDE/ODE, and simplified using the orthogonal collocation technique in three collocation points and a recirculation tank, [16], [17].

2 Analytical Model of the Anaerobic Digestion Bioprocess

The anaerobic digestion systems conformed by a fixed bed reactor and a recirculation tank is depicted on Fig. 1. It contained a fixed bed bioreactor and a recirculation tank. The anaerobic digestion process is modeled using PDE, [16]:

$$\frac{\partial X_1}{\partial t} = (\mu_1 - \varepsilon D) X_1 , \quad \mu_1 = \mu_{1\max} \frac{S_1}{K_{s_1}^{'} X_1 + S_1}, \tag{1}$$

$$\frac{\partial X_2}{\partial t} = (\mu_2 - \varepsilon D) X_2 , \quad \mu_2 = \mu_{2s} \frac{S_2}{K_{s_2}^{'} X_2 + S_2 + \frac{S_2^2}{K_{I_2}}}, \tag{2}$$

Where: X_1 is concentration of acidogenic bacteria; X_2-

ISBN: 978-989-95079-6-8

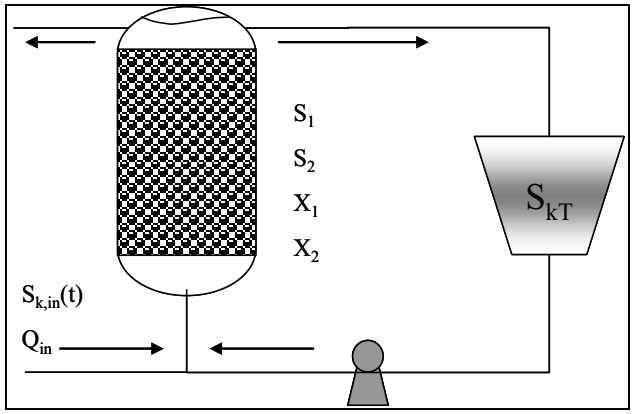


Figure 1: Block-Diagram of Anaerobic Digestion Bioreactor

concentration of methanogenic bacteria; S_1 - chemical oxygen demand; S_2 - volatile fatty acids; μ_1, μ_2- Monod functions, representing the acidogenesis and methanogenesis growth rate; D - dilution rate; ε - bacteria fraction in the liquid phase. The next equations of the bioprocess model are:

$$\frac{\partial S_1}{\partial t} = \frac{E_z}{H^2}\frac{\partial^2 S_1}{\partial z^2} - D\frac{\partial S_1}{\partial t} - k_1\mu_1 X_1\,, \tag{3}$$

$$\frac{\partial S_2}{\partial t} = \frac{E_z}{H^2}\frac{\partial^2 S_2}{\partial z^2} - D\frac{\partial S_2}{\partial t} + k_2\mu_1 X_1\,, \tag{4}$$

$$S_1(0,t) = \frac{S_{1,in}(t) + RS_{1T}}{R+1},\quad S_2(0,t) = \frac{S_{2,in}(t) + RS_{2T}}{R+1},\quad R = \frac{Q_T}{DV_{eff}} \tag{5}$$

$$\frac{\partial S_1}{\partial z}(1,t) = 0\,,\quad \frac{\partial S_2}{\partial z}(1,t) = 0\,. \tag{6}$$

$$\frac{dS_{1T}}{dt} = \frac{Q_T}{V_T}\left(S_1(1,t) - S_{1T}\right),\quad \frac{dS_{2T}}{dt} = \frac{Q_T}{V_T}\left(S_2(1,t) - S_{2T}\right). \tag{7}$$

Where: S_{1T} , S_{2T} are concentrations of the chemical oxygen demand and the volatile fatty acids in the recirculation tank, respectively; H - fixed bed length; Q_T - recycle flow rate; V_T - volume of the recirculation tank. The physical meaning of the other constants and initial values of the variables of the process model are given in [16]. For practical purpose, the full PDE process model, [16], could be reduced to an ODE system using an early lumping technique and the Orthogonal Collocation Method (OCM), [16], [17]. The precision of the OCM approximation of the PDE model depended on the number of measurement (collocation) points, but the approximation is always exact in that points. If the number of points is very high and the point positions are chosen inappropriately, the ODE model could loose identifiability. Furthermore the ODE plant model here is used as a plant data generator for neural identification and control of PDE system and the number of point not need to be too high. So to fulfill this objective we need a reduced order model having only three points, (0.25H, 0.5 H, 0.75H), but generating 14 measured variables. The reference set points generated for all that variables keep the form but differ in amplification due to its position. The plant ODE system model, obtained by OCM is taken from [16] and used as a plant process input/output data generator so to obtain identification and control simulation results.

3 Description of the Direct and Indirect Fuzzy-Neural Multi-Model Control System

3.1 Direct Adaptive FNMM Control System Design

The block-diagrams of the complete control system and its identification and control parts are schematically depicted in Fig. 2, Fig. 3 and Fig. 4.

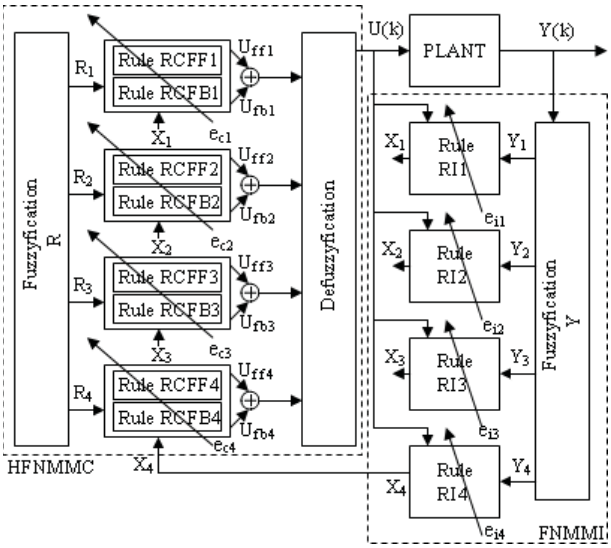


Figure .2: Block-Diagram of the FNMM Control System.

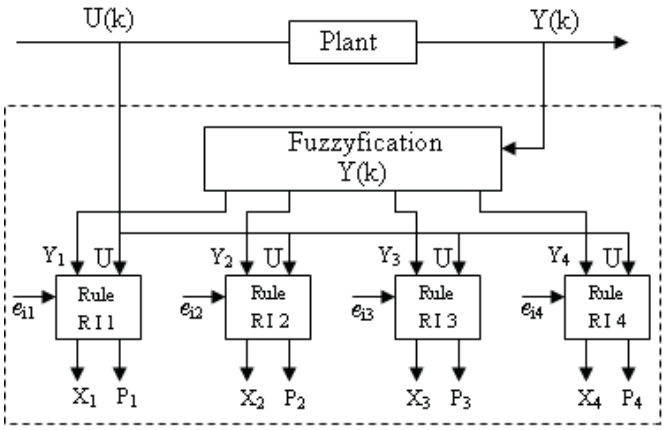


Figure 3: Detailed block-diagram of the FNMM identifier

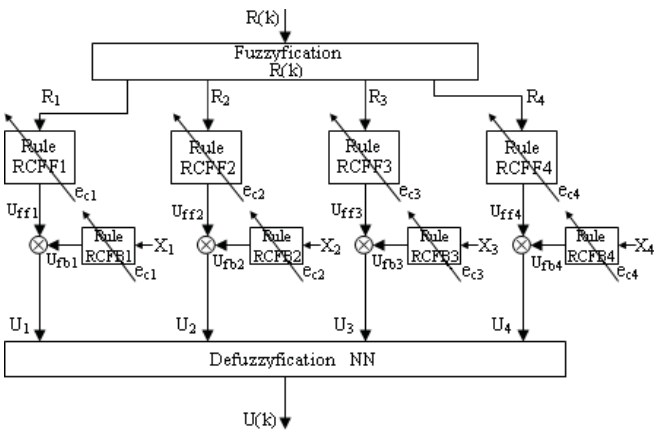


Figure 4: Detailed block-diagram of the HFNMM controller

ISBN: 978-989-95079-6-8

The structure of the entire control system, [12, 14], contained Fuzzyfier, Fuzzy Rule-Based Inference System (FRBIS), containing four identification, four feedback control and four feedforward control T-S rules (RI_i, $RCfb_i$, $RCff_i$) and a Defuzzyfier. The plant output variable and its correspondent reference variable depended on space and time, and they are fuzzyfied on space. The membership functions of the fixed-bed output variables are triangular/trapezoidal ones and that belonging to the output variables of the recirculation tank are singletons. The centers of the membership functions are the respective collocation points of the plant. The main objective of the Fuzzy-Neural Multi-Model Identifier (FNMMI), containing four rules, is to issue states for the direct adaptive Fuzzy-Neural Multi-Model Feedback Controller (FNMMFBC) when the FNMMI outputs follows the outputs of the plant in the four measurement (collocation) points with minimum Means Squared Error (MSE) of approximation. The direct fuzzy neural controller has also a direct adaptive Fuzzy-Neural Multi-Model Feedforward Controller (FNMMFFC). The objective of the direct adaptive FNMM controller, containing four Feedback (FB) and four Feedforward (FF) T-S control rules is to reduce the error of control. The upper hierarchical level of the FNMM control system is one layer perceptron which represented the defuzzyfier, [12, 14]. The hierarchical FNMM controller has two levels – Lower Level of Control (LLC), and Upper Level of Control (ULC). It is composed of three parts: 1) Fuzzyfication, where the normalized reference vector signal contained reference components of four measurement points; 2) Lower Level Inference Engine, which contains twelve T-S fuzzy rules (four rules for identification and eight rules for control- four in the feedback part and four in the feedforward part), operating in the corresponding measurement points; 3) Upper Hierarchical Level of neural defuzzification.

The detailed block-diagram of the FNMMI, given on Fig. 3, contained a space plant output fuzzyfier and four identification T-S fuzzy rules, labeled as RI_i, which consequent parts are RTNN learning procedures, [14]. The identification T-S fuzzy rules have the form:

$$RI_i\text{: } \textbf{\textit{If}}\ \mathrm{x(k)}\ \textbf{\textit{is}}\ \mathrm{A_i}\ \textbf{\textit{and}}\ \mathrm{u(k)}\ \textbf{\textit{is}}\ \mathrm{B_i}\ \textbf{\textit{then}}\ \mathrm{Y_i = \Pi_i\ (L,M,N_i,Y_{di},U,X_i,A_i,B_i,C_i,E_i),\ i=1,2,..,4} \tag{8}$$

The detailed block-diagram of the FNMMC, given on Fig. 4, contained a spaced plant reference fuzzyfier and eight control T-S fuzzy rules (four FB and four FF), which consequent parts are also RTNN learning procedures, [12, 14], using the state information, issued by the corresponding identification rules. The consequent part of each feedforward control rule (the consequent learning procedure) has the M, L, N_i RTNN model dimensions, R_i, Y_{di}, E_{ci} inputs and U_{ffi}, outputs used to form the total control. The T-S fuzzy rule has the form:

$$RCFF_i\text{: } \textbf{\textit{If}}\ \mathrm{R(k)}\ \textbf{\textit{is}}\ B_i\ \textbf{\textit{then}}\ \mathrm{Uff_i = \Pi_i\ (M, L, N_i, R_i, Y_{di}, X_i, J_i, B_i, C_i, E_{ci}),\ i =1,2,..,4} \tag{9}$$

The consequent part of each feedback control rule (the consequent learning procedure) has the M, L, N_i RTNN model dimensions, Y_{di}, X_i, E_{ci} inputs and U_{fbi}, outputs used to form the total control. The T-S fuzzy rule has the form:

$$RCFB_i\text{: } \textbf{\textit{If}}\ \mathrm{Y_{di}}\ \textbf{\textit{is}}\ A_i\ \textbf{\textit{then}}\ \mathrm{Ufb_i = \Pi_i\ (M, L, N_i, Y_{di}, X_i, X_{ci}, J_i, B_i, C_i, E_{ci}),\ i =1,2,..,4} \tag{10}$$

The total control corresponding to each of the four measurement points is a sum of its corresponding feedforward and feedback parts:

$$U_i\,(k) = -U_{ffi}\,(k) + U_{fbi}\,(k) \tag{11}$$

The defuzzyfication learning procedure, which correspond to the single layer perceptron learning, performed a weighted sum of the control variables Ui, [15]. It is described by:

$$\mathrm{U = \Pi\ (M, L, N, Y_d, Uo, X, A, B, C, E)} \tag{12}$$

3.1 Indirect Adaptive FNMM Control System Design

The block-diagram of this control system is schematically depicted in Fig.5. The structure of the entire control system, [14, 15], contained Fuzzyfier, Fuzzy Rule-Based Inference System (FRBIS), containing four identification and four control T-S rules (RI_i, RC_i), and a Defuzzyfier. The plant output variable and its correspondent reference variable depended on space and time, and they are fuzzyfied on space. The membership functions and the Identifier (FNMMI) are the same used in the direct controller. The objective of the indirect adaptive FNMM controller is equivalent to that of the direct controller. The hierarchical FNMM controller has two levels – Lower Level of Control (LLC), and Upper Level of Control (ULC). It is composed of three parts: 1) Fuzzyfication, where the normalized reference vector signal contained reference components of four measurement points; 2) Lower Level Inference Engine, which contains eight T-S fuzzy rules (four rules for identification and four rules for control), operating in the corresponding measurement points; 3) Upper Hierarchical Level of neural defuzzification. The detailed block-diagram of the FNMMC is given on Fig. 6. It contained a spaced plant reference fuzzyfier and four sliding mode control T-S fuzzy rules, which consequent parts are SMC procedures, [13, 14], using the state, and parameter information, issued by the corresponding identification rules. The control T-S fuzzy rules have the form:

$$RC_i\text{: } \textbf{\textit{If}}\ \mathrm{R(k)}\ \textbf{\textit{is}}\ \mathrm{C_i}\ \textbf{\textit{then}}\ \mathrm{U_i = \Pi_i\ (M, L, N_i, R_i, Y_{di}, X_i, A_i, B_i, C_i, E_{ci}),\ i =1, 2,\ldots, 4} \tag{13}$$

The defuzzyfication of the control variable is a learning procedure, which correspond to the single layer perceptron learning. It is described by:

$$\mathrm{U = \Pi\ (M, L, N, Y_d, Uo, X, A, B, C, E)} \tag{14}$$

The T-S rule and the defuzzification of the plant output of the fixed bed with respect to the space variable z ($\lambda i,z$ is the correspondent membership function), [14, 15], are given by:

$$RO_i\text{: } \textbf{\textit{If}}\ \mathrm{Y_{i,t}}\ \textbf{\textit{is}}\ \mathrm{A_i}\ \textbf{\textit{then}}\ \mathrm{Y_{i,t} = a_i^T Y_t + b_i,\ i=1,2,3} \tag{15}$$

$$\mathrm{Y_z=[\Sigma_i\ \gamma_{i,z}\ a_i^T]\ Y_t + \Sigma_i\ \gamma_{i,z}\ b_i\ ;\ \gamma_{i,z} = \lambda_{i,z} / (\Sigma_j\ \lambda_{j,z})} \tag{16}$$

The indirect adaptive neural control algorithm, which is the consequent part of the local fuzzy control rule RCi (13) is viewed as a Sliding Mode Control (SMC), [13, 14], using the

ISBN: 978-989-95079-6-8

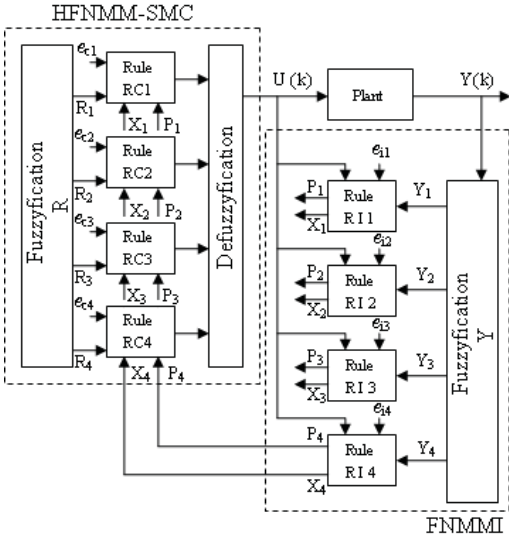


Figure 5: Block-diagram of the FNMM Control system.

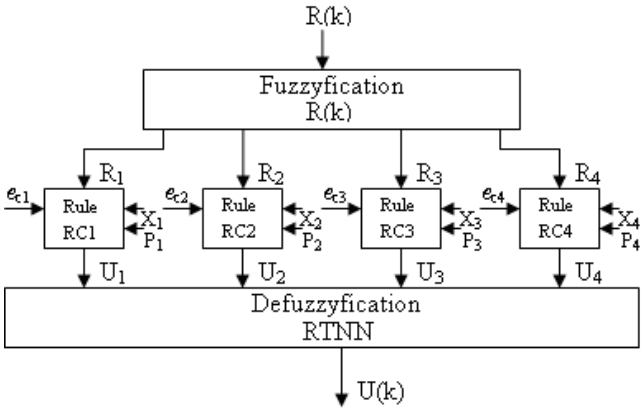


Figure 6: Detailed block-diagram of the HFNMM controller

parameters and states issued by the correspondent identification local fuzzy rule RIi (8). The equivalent control capable to lead the system to the sliding surface in the case when L>M is given by the equations:

$$U_{eq}(k) = (CB)^{+}\left[-CAX(k) + R(k+1) + \sum_{i=1}^{P}\gamma_i E(k-i+1)\right] + Of \qquad (17)$$
$$(CB)^{+} = [(CB)^{T}(CB)]^{-1}(CB)^{T}$$

Here the added offset Of is a learnable M-dimensional constant vector which is learnt using a simple delta rule (see [3]), where the error of the plant input is obtained backpropagating the output error through the adjoint RTNN model (see [18, 14]). The SMC avoiding chattering is taken using a saturation function inside a bounded control level Uo, taking into account plant uncertainties. So the SMC takes the form:

$$U(k) = \begin{cases} U_{eq}(k), & if \ \|U_{eq}(k)\| < U_0 \\ -U_0 U_{eq}(k)/\|U_{eq}(k)\|, & if \ \|U_{eq}(k)\| \geq U_0 \end{cases} \qquad (18)$$

The proposed SMC cope with the characteristics of the wide class of plant model reduction neural control with reference model, and represents an indirect adaptive neural control, given by [11, 13, 14].

3.1 Description of the RTNN and its Learning Algorithm

The RTNN topology, including thresholds in both layers, and its learning algorithm, which appeared in the consequent part of the identification fuzzy rule as a learning procedure, are described in vector-matrix form as:

$$X(k+1) = AX(k) + BU(k);\ B = [B_1 ; B_0];\ U^T = [U_1 ; U_2] \quad (19)$$
$$Z_1(k) = G[X(k)] \quad (20)$$
$$V(k) = CZ(k);\ C = [C_1 ; C_0];\ Z^T = [Z_1 ; Z_2] \quad (21)$$
$$Y(k) = F[V(k)];\ A = \text{block-diag}\ (Ai),\ |Ai| < 1 \quad (22)$$
$$W(k+1) = W(k) + \eta\,\Delta W(k) + \alpha\,\Delta W_{ij}(k-1) \quad (23)$$
$$E(k) = T(k) - Y(k);\ E_1(k) = F'[Y(k)]\,E(k) \quad (24)$$
$$\Delta C(k) = E_1(k)\,Z^T(k);\ F'[Y(k)] = [1 - Y^2(k)] \quad (25)$$
$$E_3(k) = G'[Z(k)]\,E_2(k);\ E_2(k) = C^T(k)\,E_1(k) \quad (26)$$
$$\Delta B(k) = E_3(k)\,U^T(k);\ G'[Z(k)] = [1 - Z^2(k)] \quad (27)$$
$$\Delta A(k) = E_3(k)\,X^T(k);\ Vec(\Delta A(k)) = E_3(k) \square X(k) \quad (28)$$

Where: X, Y, U are vectors of state, output, and augmented input with dimensions N, L, (M+1), respectively, Z is an (L+1) –dimensional input of the feedforward output layer, where Z_1 and U_1 are the (Nx1) output and (Mx1) input of the hidden layer; the constant scalar threshold entries are $Z_2 = -1$, $U_2 = -1$, respectively; V is a (Lx1) pre-synaptic activity of the output layer; T is the (Lx1) plant output vector, considered as a RNN reference; A is (NxN) block-diagonal weight matrix; B and C are [Nx(M+1)] and [Lx(N+1)]-augmented weight matrices; B_0 and C_0 are (Nx1) and (Lx1) threshold weights of the hidden and output layers; F[.], G[.] are vector-valued tanh(.)-activation functions with corresponding dimensions; F'[.], G'[.] are the derivatives of these tanh(.) functions; W is a general weight, denoting each weight matrix (C, A, B) in the RTNN model, to be updated; ΔW (ΔC, ΔA, ΔB), is the weight correction of W; η, α are learning rate parameters; ΔC is an weight correction of the learned matrix C; ΔB, ΔA are weight corrections of the learned matrices B, A; the diagonal of the matrix A is denoted by Vec (.) where (28) represents its learning as an element-by-element vector product; E, E_1, E_2, E_3, are error vectors with appropriate dimensions. The stability of the RTNN model is assured by the activation functions [-1, 1] bounds and by the local stability weight bound condition, given by (22). The learning procedure having forward and backward steps (19)-(28), could be denoted by Y(k) = Π (L, M, N, Y_d, U, X, A, B, C, E). The defuzzyfication learning procedure, corresponding to the single layer perceptron, denoted by (14) is a simple delta rule (see [3]), where the error of the plant controlled input is obtained backpropagating the output error through the adjoint RTNN model, [18].

4 Simulation Results

The decentralized FNMM identifier used a set of four T-S fuzzy rules containing in its consequent part RTNN learning procedures (see Fig. 3). The topology of the first three RTNNs is (2-6-4) (2 inputs, 6 neurons in the hidden layer, 4 outputs) and the last one has topology (2-6-2) corresponding to the fixed bed plant behavior in each collocation point and the recirculation tank. The RTNNs identified the following fixed bed variables: X_1 (acidogenic bacteria), X_2

ISBN: 978-989-95079-6-8

(methanogenic bacteria), S_1 (chemical oxygen demand) and S_2 (volatile fatty acids), in the following collocation points, z=0.25H, z=0.5H, z=0.75H, and the following variables in the recirculation tank: S_{1T} (chemical oxygen demand) and S_{2T} (volatile fatty acids). The graphical simulation results of RTNNs learning are obtained on-line during 100 days with a step of 0.1 day (To=0.1 sec.; Nt=1000 iterations). The learning rate parameters of RTNN have small values which are different for the different measurement point variables. The Fig. 7, Fig.8, Fig. 9, Fig. 10 showed similar three dimensional and two dimensional graphical simulation results of the direct and indirect decentralized FNMM control of X_1, X_2, where the outputs of the plant are compared with the reference signals at the collocation points. The reference signals for all variables are proportional train of pulses with uniform duration and random amplitude. The Means Squared Error (MSE%) of identification, direct, indirect and optimal control for each output signal and each measurement point are given on Table 1, 2, 3, 4, respectively. The comparison showed a slight preference of the indirect FNMM-SMC control over the direct FNMM control, due to the better adaptation of the first low.

Table 1: MSE% of the decentralized FNMM approximation of the bioprocess output variables

Coll. point	X_1	X_2	S_1/S_{1T}	S_2/S_{2T}
z=0.25H	1.2524e-8	6.5791e-8	2.9615e-5	4.3302e-4
z=0.5H	5.0180e-9	2.9067e-8	1.1840e-8	2.7851e-6
z=0.75H	1.0487e-9	2.7977e-9	9.3562e-5	2.8941e-4
Recir Tank	-	-	8.6967e-7	2.0205e-6

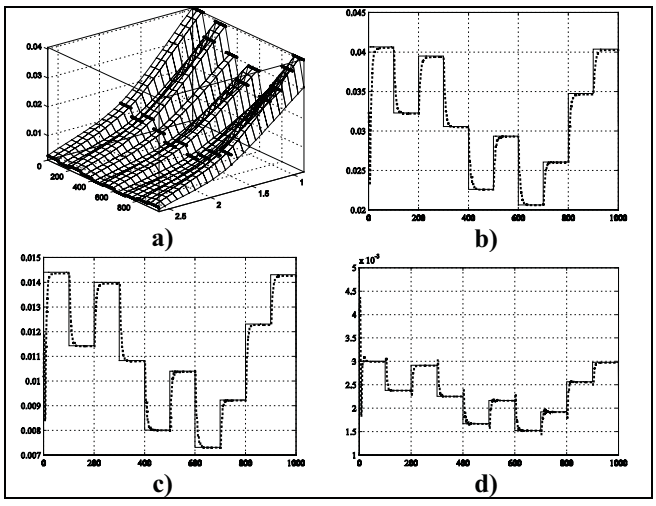


Figure 7: Results of the direct decentralized FNMM control of X_1 (acidogenic bacteria in the fixed bed) (dotted line-plant output, continuous-reference); a) 3d view of X_1 ; b) Ref vs X_1 in z=0.25H; c) Ref vs X_1 in z=0.5H; d) Ref vs X_1 in z=0.75H.

Table 2: MSE% of the direct decentralized FNMM control of the bioprocess output variables

Coll. point	X_1	X_2	S_1/S_{1T}	S_2/S_{2T}
z=0.25H	1.2524e-8	6.5791e-8	2.9615e-5	4.3302e-4
z=0.5H	5.0180e-9	2.9067e-8	1.1840e-8	2.7851e-6
z=0.75H	1.0487e-9	2.7977e-9	9.3562e-5	2.8941e-4
Recir Tank	-	-	8.6967e-5	2.0205e-4

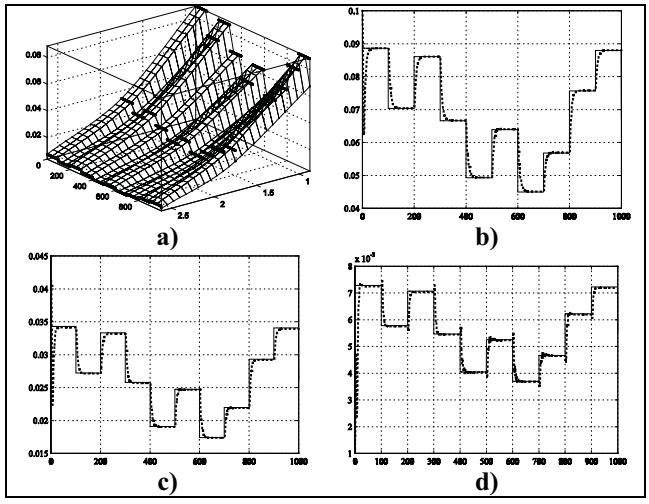


Figure 8: Results of the direct decentralized FNMM control of X_2 (methanogenic bacteria in the fixed bed) (dotted line-plant output, continuous-reference); a) 3d view of X_2 ; b) Ref vs X_2 in z=0.25H; c) Ref vs X_2 in z=0.5H; d) Ref vs X_2 in z=0.75H.

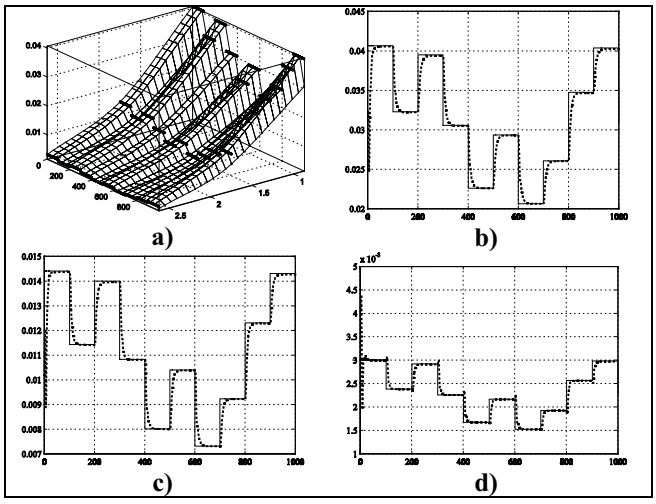


Figure 9: Results of the decentralized FNMM-SMC of X_1 (acidogenic bacteria in the fixed bed) (dotted line-plant output, continuous-reference); a) 3d view of X_1; b) SMC of X_1 in z=0.25H; c) SMC of X_1 in z=0.5H; d) SMC of X_1 in z=0.75H.

Table 3: MSE% of the decentralized FNMM–SMC of the bioprocess output variables

Coll. point	X_1	X_2	S_1/S_{1T}	S_2/S_{2T}
z=0.25H	1.7494E-8	3.0157E-9	3.9538E-6	7.9391E-9
z=0.5H	2.2131E-9	1.9669E-8	4.8951E-7	2.1116E-6
z=0.75H	1.0415E-10	1.3238E-9	2.3548E-8	1.3095E-7
Recir Tank			8.4352E-8	5.8734E-9

Table 4: MSE% of the proportional optimal control of the bioprocess output variables

Coll. point	X_1	X_2	S_1/S_{1T}	S_2/S_{2T}
z=0.25H	5.3057E-8	1.7632E-7	1.1978E-5	2.1078E-5
z=0.5H	6.6925E-9	4.2626E-8	1.4922E-6	4.4276E-6
z=0.75H	3.0440E-10	2.0501E-9	6.8737E-8	2.0178E-7
Recir Tank			2.7323E-7	6.0146E-7

The comparison also showed a slight preference of the FNMM-SMC control over the linearized optimal control, due to the adaptation of the FNMM-SMC control low.

ISBN: 978-989-95079-6-8

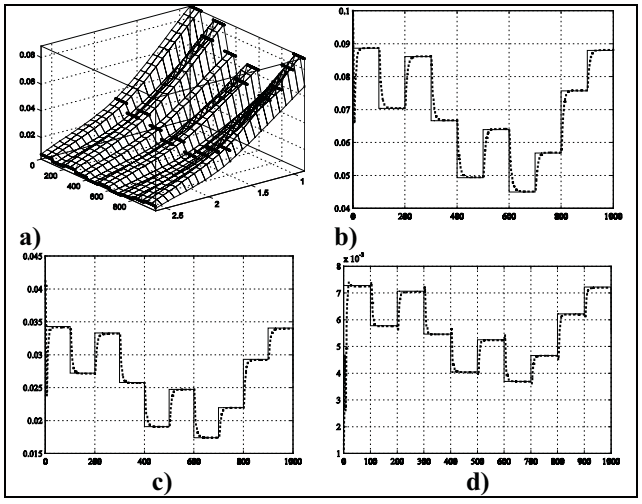


Figure 10: Results of the decentralized FNMM-SMC of X_2 (methanogenic bacteria in the fixed bed) (dotted line-plant output, continuous-reference); a) 3d view of X_2; b) SMC of X_2 in z=0.25H; c) SMC of X_2 in z=0.5H; d) SMC of X_2 in z=0.75H.

5 Conclusions

The paper performed decentralized recurrent fuzzy-neural identification, direct and indirect control of an anaerobic digestion wastewater treatment bioprocess, composed by a fixed bed and a recirculation tank represented a DPS. The simplification of the PDE process model by ODE is realized using the orthogonal collocation method in three collocation points (plus the recirculation tank) represented centers of membership functions of the space fuzzyfied output variables. The obtained from the FNMMI state and parameter information is used by a HFNMM direct and indirect control. All graphical control results exhibited good convergence and precise reference tracking. All obtained comparative numerical identification and control results (final MSE%) exhibited a high precision and showed that the indirect decentralized FNMM control is the better one outperforming the linearized optimal control and the direct decentralized fuzzy-neural control.

Acknowledgment

The Ph.D. student Rosalba Galvan-Guerra is thankful to CONACYT, Mexico for the scholarship received during her studies at the Department of Automatic Control, CINVESTAV-IPN, Mexico City, Mexico.

References

[1] W.H. Ray, *Advanced Process Control*, Butterworths, Boston, London, Singapore, Sydney, Toronto, Wellington, 133-242, 1989.

[2] J.D. Boskovic, and K. S. Narendra, Comparison of Linear, Nonlinear and Neural-Network-Based Adaptive Controllers for a Class of Fed-Batch Fermentation Processes, *Automatica,* 31: 817-840, 1995.

[3] S. Haykin, *Neural Networks, a Comprehensive Foundation*, Second Edition, Prentice-Hall, Upper Saddle River, New Jersey 07458, Section 2.13: 84-89; Section 4.13: 208-213, 1999.

[4] A. Bulsari, and S. Palosari, Application of Neural Networks for System Identification of an Adsorption Column, *Neural Computing and Applications*, 1: 160-165, 1993.

[5] H. Deng, and H.X. Li, Hybrid Intelligence Based Modeling for Nonlinear Distributed Parameter Process with Applications to the Curing Process, *IEEE Transactions on SMC*, 4: 3506-3511, 2003.

[6] H. Deng, and H.X. Li, Spectral-Approximation-Based Intelligent Modeling for Distributed Thermal Processes, *IEEE Transactions on Control Systems Technology*, 13: 686-700, 2005.

[7] R. Gonzalez-Garcia, R. Rico-Martinez, and I. Kevrekidis, Identification of Distributed Parameter Systems: A Neural Net Based Approach, *Computers and Chemical Engineering*, 22 (4-supl. 1): 965-968, 1998.

[8] R. Padhi, S. Balakrishnan, and T. Randolph, Adaptive Critic based Optimal Neuro Control Synthesis for Distributed Parameter Systems, *Automatica*, 37: 1223-1234, 2001.

[9] R. Padhi, and S. Balakrishnan, Proper Orthogonal Decomposition Based Optimal Neurocontrol Synthesis of a Chemical Reactor Process Using Approximate Dynamic Programming, *Neural Networks*, 16: 719-728, 2003.

[10] S. Pietil, and H.N. Koivo, Centralized and Decentralized Neural Network Models for Distributed Parameter Systems, *In: Proc. of the Symposium on Control, Optimization and Supervision, In: Proc. of CESA'96, IMACS Multiconference on Computational Engineering in Systems Applications*, Lille, France, 1043-1048, 1996.

[11] I. Baruch, J.M. Flores, and R. Garrido, A Fuzzy-Neural Recurrent Multi-Model for Systems Identification and Control, *In: Proc. of the ECC'01*, Porto, Portugal, 3540-3545, Sept. 4-7, 2001.

[12] I. Baruch, J.L. Olivares-Guzman, C.R. Mariaca-Gaspar, and R. Galvan-Guerra, A Fuzzy-Neural Hierarchical Multi-Model for Systems Identification and Direct Adaptive Control, *In: P. Melin, O.Castillo, E.G. Ramirez,, J. Kacprzyk, W. Pedrycz (eds.): Analysis and Design of Intelligent Systems Using Soft Computing Techniques,* ASC Vol. 41, Springer-Verlag Berlin Heidelberg, 163-172, 2007.

[13] I. Baruch, J.L. Olivares-Guzman, C.R. Mariaca-Gaspar, and R. Galvan-Guerra, A Sliding Mode Control Using Fuzzy-Neural Hierarchical Multi-Model Identifier, *In: O.Castillo, P. Melin, O.M. Ross, R.S. Cruz, W. Pedrycz, J. Kacprzyk (eds.): Theoretical Advances and Applications of Fuzzy Logic and Soft Computing,* ASC Vol. 42, Springer-Verlag Berlin Heidelberg, 762-771, 2007.

[14] I. Baruch, R. Beltran-Lopez, J.L. Olivares-Guzman, and J.M. Flores, A Fuzzy-Neural Multi-Model for Nonlinear Systems Identification and Control, *Fuzzy-Sets and Systems*, 159: 2650-2667, 2008.

[15] T.Takagi, and M. Sugeno, Fuzzy Identification of Systems and its Applications to Modeling and Control, *IEEE Trans. on SMC*, 15: 116-132, 1985.

[16] F. Aguilar-Garnica, V. Alcaraz-Gonzalez, and V. Gonzalez-Alvarez, Interval Observer Design for an Anaerobic Digestion Process Described by a Distributed Parameter Model, *In: Proc. of the 2-nd Intern. Meeting on Environm. Biotech. and Engng. (2IMEBE),* Mexico City, paper 117: 1-16, 2006.

[17] B. Bialecki, and G. Fairweather, Orthogonal Spline Collocation Methods for Partial Differential Equations, *Journal of Computational and Applied Mathematics*, 128: 55-82, 2001.

[18] E. Wan, and F. Beaufays, Diagrammatic Method for Deriving and Relating Temporal Neural Networks Algorithms, *Neural Computations*, 8: 182-201, 1996.

ISBN: 978-989-95079-6-8

Enforcing Local Properties in Online Learning First Order TS-fuzzy Systems by Incremental Regularization

Nils Rosemann Werner Brockmann Bastian Neumann
University of Osnabrück
Albrechtstraße 28
D–49069 Osnabrück

Abstract— *Embedded systems deseminate more and more. Because their complexity increases and their design time has to be reduced, they have to be increasingly equipped with self-tuning properties. One form is self-adaption of the system behavior, which can potentially lead the system into safety critical states. In order to avoid this and to speed up the self-tuning process, we apply a specific form of regularization, incremental regularization. The SILKE approach has been developed as an incremental regularization scheme for a special class of online learning Takagi-Sugeno fuzzy systems. Its aim is to control the process of self-tuning by guiding the online learning process towards local meta-level characteristics such as a smooth system behavior without outliers. This ability has been investigated experimentally and formally for zero order systems before. This paper now analyzes the regularization ability of the SILKE approach to enforce local smoothness in first order TS-fuzzy systems in order to enlarge the methodological basis for more complex applications.*

Keywords— First order Takagi Sugeno fuzzy systems, incremental learning, self-optimization, incremental regularization, fuzzy control

1 Introduction

1.1 Background

Due to the growing dissemination of embedded technical systems and their higher integration into critical environments, the requirements for the next generation of those technical systems will change. The systems will have to adapt intelligently to different environments, different aims and changes within the environment in order to open up new application areas. This greatly increases the complexity of such new systems. At the same time, it will become necessary to decrease the design time. We tackle these problems within the context of Organic Computing, There one tries to face these challenges by introducing self-x properties into the systems, e.g., self-organization and self-optimization [1]. This means that technical systems are equipped with a much higher flexibility. They have to adapt their behavior based on changing demands and prior experience.

But although such a technical system adapts its behavior at run time, safety and trustworthiness are a must. This is especially problematic as the higher integration into natural, i.e., not specially prepared, environments comes with many sources of uncertainty. Thus, any form of adaption based on uncertain interaction with the environment will be uncertain itself. Accordingly, a self-tuning system has to be both flexible enough to react dynamically to new conditions, and stable enough to protect already learned knowledge from transient disturbances, noise and other uncertainties (*stability-plasticity-dilemma*). And, what is more, at the same time it has to be guaranteed that the system does not enter critical states by learning. This means that the self-tuning process has to be controlled, but without limiting its flexibility.

Within this paper, we address issues of self-tuning the control behavior of an embedded system, i.e., a form of self-adaptive control. This online changing of the behavior is done by *incremental online learning*. It has a strong influence on the stability of the system: In addition to the feedback loop between the control system and the environment into which it is embedded, online learning creates another, indirect feedback loop [2]. Whatever is learned in a certain situation changes the system's (future) behavior, and thus also what is learned in the future. Because of this, small deviations of (initial) conditions or the self-optimization process as such might have arbitrarily high impact and can potentially lead to a chaotic system behavior.

Different candidate function approximators have been used in adaptive control by online learning, each of them fullfilling the different requirements to a varying degree. An overview over the different families of function approximators is given in [3]. Accordingly, one has to choose an appropriate approximator to represent the control behavior in a way which suits the application specific needs, e.g., concerning expressiveness, complexity, interpretability and, what is more in our context, controllability of the learning process itself.

For adaptive nonlinear modeling and control, Takagi-Sugeno systems are a favoured trade-off [4, 5]. We hence build upon a specific variant. Then, in order to employ incremental online learning for adaptive control in uncertain environments, one has to address the stability-flexibility-dilemma. In the next section it is argued that this can be done by an appropriate form of regularization.

1.2 Related Work

In data-driven function approximation, regularization techniques have been successfully applied in many different areas, especially in nonlinear system identification [6] and nonlinear adaptive control [7]. Typical for these applications is that the distribution of data points is dense in one region of the input space and sparse in another. For practical reasons it might be impossible to obtain enough data in a sparse region, e.g., because it is related to critical states of a controlled process. Hence regularization techniques have found widespread use in data-driven fields. The reason is that in many cases data-driven problems are ill-posed because there is no unique solution [8]. This is usually due to a misbalance between the degrees of freedom of the function approximator and the amount of data available.

The idea of regularization is to incorporate additional in-

ISBN: 978-989-95079-6-8

formation about the solution into the problem statement, i.e., to reformulate it [9]. Usually this is some kind of structural or meta-information. A well known example is to assign a penalty to non-zero parameter values [10]. Another example is to assign a penalty to solutions with a high overall curvature [11]. They are called ridge regression or Tikhonov regularization. Despite missing data, sparse regions and large input spaces, regularization is able to improve the generalization performance of the identified models.

Basically, these approaches require all data points to be accessible at once in order to find a globally optimal solution, i.e., they are offline approaches. But for many online applications, there is a change of requirements while new data points are generated (continuously), for example because the system which is to be modeled changes or is modelled from scratch.

Recent approaches for these kind of applications have so far neglected the problem of requiring regularization [4]. In order to tackle the above challenges online, these approaches learn incrementally, i.e., by looking at every data point only once. But without access to the whole dataset, it is not *a priori* clear how to apply the conventional regularization techniques. Other approaches use regularization techniques for ill-posed matrix inversions [12] which is suitable for certain modeling applications.

Because of our background of adaptive nonlinear control by direct adaptation [2], we are developing schemes for *incremental regularization*, i.e., which perform special forms of regularization online, for appropriate function approximators. The control tasks which we address typically have only a low number of input dimensions (less than 10). In the case of more input dimensions, it is usually possible and advisable to decompose the overall problem into subproblems. Thus, the first one of the schemes has been developed for special Takagi-Sugeno fuzzy systems which are especially suited for adaptive nonlinear modeling and control [13, 5]. They employ an all-coverage approach for generating the rule base which allows for a high interpretability albeit limiting the maximal possible number of input dimensions. These fuzzy systems and the incremental regularization, called SILKE approach, will be reviewed in section 2. Up to now, we have developed incremental regularization schemes for zero order Takagi-Sugeno fuzzy systems [14]. They have been investigated experimentally and formally [15, 16]. The purpose of this paper is to extend the SILKE approach to first order Takagi-Sugeno fuzzy systems. In section 3 it is then shown that each incremental application of our approach to a first order Takagi-Sugeno system increases its smoothness. Because of this, the SILKE approach can be seen as a kind of incremental Tikhonov regularization which is also applicable to first order TS-fuzzy systems.

2 Controlled Self-optimization by the SILKE Approach

2.1 Basic Concept

The basis of the SILKE approach is a special class of Takagi-Sugeno fuzzy systems (hereinafter called sTS fuzzy systems) with polynomial functions as rule conclusions. They use linear B-spline input membership functions, such that each input dimension is split by a partition of unity. This allows for an intuitive handling by the system designer and an implementation which requires a very low computational effort. The

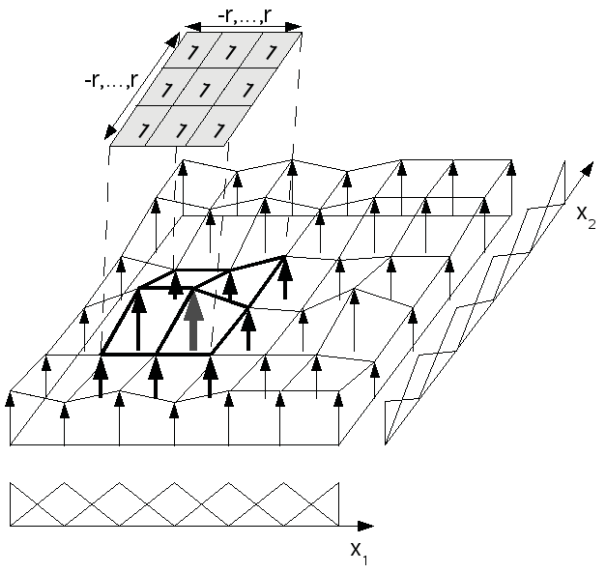


Figure 1: Schematic correspondence of a SILKE-template (shaded in grey) to the conclusions of a part of an sTS-fuzzy rule base in a two-dimensional system. The membership functions determine the lattice of rule conclusions. The thick arrow indicates the active rule under consideration. The numbers within the template are the coefficients of the averaging template mask [15].

rule inference is done by the *sum-prod* technique. These sTS-systems are ideally suited for application to online learning within embedded systems [3].

The idea of the SILKE approach is then to enhance learning in sTS-systems by a correction step which introduces an additional feedback loop at a higher, functional level in order to rate the learning process itself online [17]. This is done by implementing so called *SILKE templates* which determine whether the rule base of the sTS-system complies with predefined meta-level properties. If this *monitoring* detects too strong deviations, then the rule base should be corrected, which closes the feedback loop at a higher level. It should be noted that such a template acts only on those rules which apply to the current system state and which form a local neighborhood (*active rules*), i.e., a template acts only locally and incrementally on each new data point.

Many different meta-level properties can be expressed, e.g., steadiness, monotonicity, local linearity or gradient ratios of the underlying function. One of the most important examples of meta-properties in the field of adaptive control is *smoothness* of the approximated function, i.e., the learned knowledge, because smoothness means that small changes of the inputs do not cause strong changes at the output. So given a process which is smooth, then one wants a rule base which is also smooth. If online learning changes the rule base in a way that it is smooth, then the process of learning harmonizes with its environment. Hence, robustness of learning is increased when faced with learning stimuli which would otherwise violate the smoothness property and smoothness is enforced, for example the effect of outliers is reduced.

In the next section, it is formally shown that this smoothness property can be enforced for first order sTS-systems by

ISBN: 978-989-95079-6-8

using the SILKE approach. For this purpose, the formalism of the SILKE approach is presented in its general form. Afterwards, a measure for local smoothness in zero and first order sTS-systems is developed. This measure is then used to investigate the effect of the SILKE approach in a simplified, one-dimensional scenario for clarity reasons.

2.2 Formalization

Let n be the number of inputs of a first order sTS-fuzzy system. By compilation of an sTS rule base, it becomes normalized, so that one can imagine it as a lattice [17]. This is illustrated by the schematic in Fig. 1. It shows a two-dimensional lattice of a rule base with two input variables, x_1 and x_2. Both inputs are partitioned by triangular membership functions. For every combination of two membership functions, one from each dimension, there is a single rule. Due to normalization, the rules thus form a lattice given by these membership functions. The membership functions can be numbered sequentially in each input dimension. As there is a single rule for each combination of membership functions, one can assign a vector of the individual membership function indixes to each rule. For example, in a two-dimensional rule base, the index vector $(2, 1)$ is assigned to the rule which has the second membership function of the first input dimension and the first membership function of the second input dimension as its antecedents.

The rule conclusions (which are affine functions) can be viewed as being placed upon the nodes of this lattice. It is useful here to identify the affine conclusions by vectors of their parameters, i.e., a function $g : \mathbb{R}^n \to \mathbb{R}$ with

$$g(\mathbf{x}) = c_0 + c_1 x_1 + \ldots + c_n x_n, \quad \mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}, \tag{1}$$

can be identified by the vector of the parameters

$$\mathbf{c} = (c_0, \ldots, c_n)^T \in \mathbb{R}^{n+1}. \tag{2}$$

With this, the rule base of a first-order sTS fuzzy system can be seen as a map $o : \mathbb{N}^n \to \mathbb{R}^{n+1}$ from the space $\mathbb{N}^n$ of rule index vectors to the space $\mathbb{R}^{n+1}$ of parameter vectors. A single node $p \in \mathbb{N}^n$ of the lattice corresponds to a rule, and $o(p)$ to the parameters of an affine function, i.e., the rule conclusion.

A so called SILKE template T is than a map

$$T : \mathbb{N}^n \to \mathbb{R}^{n+1}. \tag{3}$$

The value $\|T(p)\|$ is called *violation degree* of the rule p with respect to T. It indicates how much a rule p violates the meta-level property which is given by T with respect to the neighboring rules. $\|T(p)\| = 0$ means a complete compliance with this property. The higher $\|T(p)\|$, the further rule p is violating the property.

The core of the SILKE approach is then to correct the given rule p according to

$$o(p) \leftarrow (1 - \alpha) \cdot o(p) + \alpha \cdot (o(p) - T(p)). \tag{4}$$

Obviously, if $\|T(p)\| = 0$, the rule is not changed. The *adjustment rate* $\alpha \in [0; 1]$ determines how much the meta-level property is enforced, i.e., $\alpha = 0$ results in no correction of the rule base, whereas $\alpha = 1$ makes it fully compliant with the meta-level property in one step. As said before, such a correction step is done directly after each learning step. Every rule which has been changed by learning is subject to the SILKE template(s). This means that $\alpha = 1$ ignores all changes of the rules by learning. Instead, the meta-level property is fully enforced. Thus, high values of α limit the influence of learning severely but increase the stability.

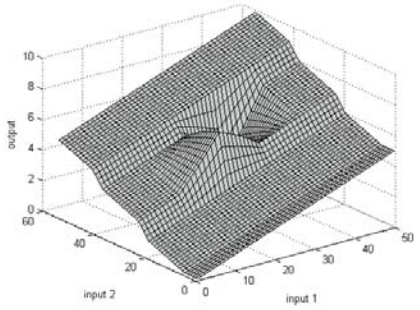


Figure 2: Original graph of a sample two-dimensional sTS-system which is smoothed by the SILKE approach, see Fig. 3 and Fig. 4.

In the following, only SILKE templates are considered which can be represented by a convolution of the rule base:

$$T(p) = o(p) - \frac{1}{(2r+1)^n} \sum_{u \in U} o(p-u)m(u). \tag{5}$$

Here, $r \in \mathbb{N}^+$ is a radius. All neighbor rules with a L_1-distance to the given rule p of less than or equal to r are considered by the template (For illustration, in Fig. 1 an exemplary template is shown with radius $r = 1$). So the sum indices u are in

$$U = [-r, r]^n. \tag{6}$$

The map $m : U \to \mathbb{R}$ is called a *mask* and determines the effect of the SILKE template, i. e., the represented meta-level property, by specific coefficients. Because of this, the mask is the most important part of the SILKE approach concerning design. The designer has to use the right coefficients in the mask in order to represent the desired meta-level properties. The mask thus gives the SILKE approach the necessary flexibility to express a wide range of local meta-level properties. The example in Fig. 1 shows the mask for an averaging template which is given by $m(u) = 1 \forall u \in U$. This mask can be used to guide zero order sTS systems towards a higher smoothness [15]. Another example is the one-dimensional averaging template which only use one of the input dimensions. It is given by $m(u) = 1 \forall u \in \{0\}^{a-1} \times [-r, r] \times \{0\}^{n-a}$ with a denoting the dimension to be used.

3 Smoothness in First Order sTS Fuzzy Systems

In order to extend the formal investigations of the SILKE approach to first order sTS fuzzy systems, they are investigated concerning their smoothness properties. In this section, based on the definition of a measure of smoothness for the individual rules, it is proven that the application of an averaging SILKE

ISBN: 978-989-95079-6-8

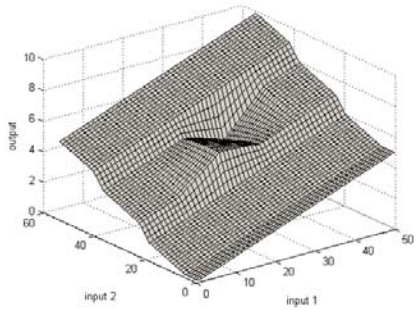


Figure 3: Graph of the sample two-dimensional sTS-system from Fig. 2 after the first application of an averaging template ($\alpha = 0.5$).

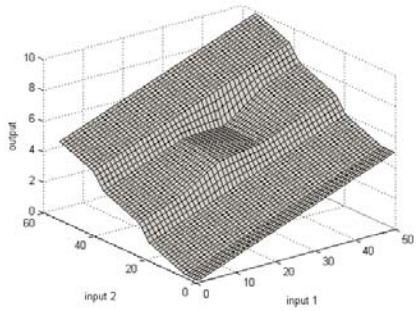


Figure 4: Graph of the sample two-dimensional sTS-system from Fig. 2 after applying an averaging template for two times ($\alpha = 0.5$).

template to a rule increases its smoothness. This can be seen as a form of incremental regularization.

For clarity reasons, we first consider a one dimensional first order sTS-fuzzy system f with triangular membership functions which sum up to 1. Then the input space is covered completely by a set of neighboring membership functions. Due to the normalization of the rule base, there is a single rule for each membership function. Thus, the rules can be ordered according to the ordinality of the membership functions so that they can be addressed by an index $i \in \mathbb{N}$, such that the indices i and $i+1$ address neighboring rules. The rule conclusion of rule number i is given by

$$f_i(x) = c_{0,i} + c_{1,i} \cdot x \tag{7}$$

with x as the input to the fuzzy system.

In the following, let $p_i \in \mathbb{R}$ denote the core of the membership function of rule i. As the membership functions are triangular, it is convenient here to distinguish between the left and the right part, $\mu_{i,l}(x)$ and $\mu_{i,r}(x)$, respectively, of each of them:

$$\begin{aligned} \mu_{i,l}(x) &= \frac{x - p_{i-1}}{p_i - p_{i-1}}, \quad p_{i-1} \le x < p_i \\ \mu_{i,r}(x) &= \frac{p_{i+1} - x}{p_{i+1} - p_i}, \quad p_i \le x < p_{i+1}. \end{aligned} \tag{8}$$

We are now considering the smoothness of the sTS fuzzy system in the environment of p_i. The functional behavior in this environment is determined by the neighboring rules, i.e., the rules $i-1$, i and $i+1$. To analyze the smoothness, we first calculate the first derivative of the sTS fuzzy system on the left and on the right of p_i. Between p_{i-1} and p_i, the output of the fuzzy system is then given by

$$f(x) = \mu_{i-1,r}(x) f_{i-1}(x) + \mu_{i,l}(x) f_i(x) \tag{9}$$

because of the linear interpolation due to sum-prod inference. Similarly, between p_i and p_{i+1}, the output is

$$f(x) = \mu_{i,r}(x) f_i(x) + \mu_{i+1,l}(x) f_{i+1}(x). \tag{10}$$

With these preparations, one can calculate the first derivative f' of f between the cores. For $p_{i-1} < x < p_i$ one gets

$$\begin{aligned} f'(x) &= \mu'_{i-1,r}(x) f_{i-1}(x) + \mu_{i-1,r}(x) f'_{i-1}(x) \\ &\quad + \mu'_{i,l}(x) f_i(x) + \mu_{i,l}(x) f'_i(x) \end{aligned} \tag{11}$$

and for $p_i < x < p_{i+1}$

$$\begin{aligned} f'(x) &= \mu'_{i,r}(x) f_i(x) + \mu_{i,r}(x) f'_i(x) + \\ &\quad \mu'_{i+1,l}(x) f_{i+1}(x) + \mu_{i+1,l}(x) f'_{i+1}(x). \end{aligned} \tag{12}$$

We can now determine the difference between the right and the left limit of $f'(x)$ with $x \to p_i$ and simplify it as the membership functions are either 0 or 1 at p_i:

$$\begin{aligned} \Delta_i &= \lim_{x \searrow p_i} f'(x) - \lim_{x \nearrow p_i} f'(x) \\ &= (\mu'_{i,r}(p_i) - \mu'_{i,l}(p_i)) f_i(p_i) \\ &\quad + \mu_{i+1,l}(p_i) f_{i+1}(p_i) - \mu'_{i-1,r}(p_i) f_{i-1}(p_i) \\ &= \left(\frac{-1}{p_{i+1} - p_i} + \frac{-1}{p_i - p_{i-1}} \right) (c_{0,i} + c_{1,i} p_i) \\ &\quad + \frac{1}{p_{i+1} - p_i} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad + \frac{1}{p_i - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i). \end{aligned} \tag{13}$$

ISBN: 978-989-95079-6-8

This expression means that in general one obtains a different result for the first derivative when approaching p_i from left than from right. To put it otherwise, Δ_i is a measure of the violation of the smoothness property at rule i.

Now suppose that our sTS fuzzy system contains a rule i with $\Delta_i \neq 0$. We will now prove that we can use a special SILKE template to reduce the violation of the smoothness property. This special SILKE template is given by the radius $r = 1$ and the mask

$$\begin{aligned} m(-1) &= \frac{3(p_{i+1} - p_i)}{p_{i+1} - p_{i-1}} \\ m(0) &= 0 \\ m(1) &= \frac{3(p_i - p_{i-1})}{p_{i+1} - p_{i-1}}. \end{aligned} \tag{14}$$

The application of this mask according to (4) and (5) changes $c_{0,i}$ and $c_{1,i}$ to $\hat{c}_{0,i}$ and $\hat{c}_{1,i}$, respectively:

$$\begin{aligned} \hat{c}_{a,i} &= c_{a,i} + \alpha \left(\frac{p_{i+1} - p_i}{p_{i+1} - p_{i-1}} c_{a,i-1} \right. \\ &\quad \left. + \frac{p_i - p_{i-1}}{p_{i+1} - p_{i-1}} c_{a,i+1} - c_{a,i} \right), \; a \in \{1, 2\}. \end{aligned} \tag{15}$$

This yields an updated value $\hat{\Delta}_i$ for the violation of the smoothness property. By comparing the updated value $\hat{\Delta}_i$ with the old value Δ_i, we can prove that the SILKE template (14) has the desired effect:

$$\begin{aligned} \hat{\Delta}_i &= \left(\frac{-1}{p_{i+1} - p_i} + \frac{-1}{p_i - p_{i-1}} \right) (\hat{c}_{0,i} + \hat{c}_{1,i} p_i) \\ &\quad + \frac{1}{p_{i+1} - p_i} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad + \frac{1}{p_i - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i) \\ &= \Delta_i + \alpha \left(\frac{-1}{p_{i+1} - p_i} + \frac{-1}{p_i - p_{i-1}} \right) \\ &\quad \cdot \left(\frac{p_{i+1} - p_i}{p_{i+1} - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i) \right. \\ &\quad + \frac{p_i - p_{i-1}}{p_{i+1} - p_{i-1}} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad \left. + \frac{-1}{1} (c_{0,i} + c_{1,i} p_i) \right) \\ &= \Delta_i + \alpha \left(\left(\frac{1}{p_{i+1} - p_i} + \frac{1}{p_i - p_{i-1}} \right) (c_{0,1} + c_{1,i} p_i) \right. \\ &\quad + \frac{-1}{p_{i+1} - p_i} (c_{0,i+1} + c_{1,i+1} p_i) \\ &\quad \left. + \frac{-1}{p_i - p_{i-1}} (c_{0,i-1} + c_{1,i-1} p_i) \right) \\ &= \Delta_i - \alpha \Delta_i \\ &= (1 - \alpha) \Delta_i. \end{aligned} \tag{16}$$

From this, it can be concluded that

$$|\hat{\Delta}_i| \leq |\Delta_i|. \tag{17}$$

The effect of this parameter update on a sample two-dimensional sTS-system is illustrated in Fig. 2 to 4. The figures demonstrate that each application of (4) on the central rule (in this case with $\alpha = 0.5$) increases the smoothness.

4 Discussion

The above investigations have shown that first order sTS-fuzzy systems with triangular membership functions have discontinuities of their first derivative precisely at the cores of the input membership functions, i.e., at the nodes of the normalized rule base. These discontinuities correspond to violating the *local smoothness* of the system behavior. It was shown for the one dimensional case that the SILKE approach reduces the discontinuity locally by a special averaging template, i.e., it increases the local smoothness. The effect depends on the adjustment rate α. For $\alpha = 0$, the rule base is not changed at all. But for higher values of α, the discontinuity of the first derivative is more and more reduced. Thus, the learning process is guided towards the meta-property of global smoothness expressed by the template function. Up to now, the formal proof is for the one dimensional case, although the SILKE algorithm is of course applicable to arbitrary dimensionality. We have shown by examplary investigations that the smoothing property also holds for higher dimensions. The extension of the formal proof is currently in progress.

In contrast to prior work on regularization, the SILKE approach has the advantage of working incrementally. It is thus well suited for incremental online learning which is required for self-adapting systems or building models online. During online operation, it is applied periodically and affects all currently active rules of the fuzzy system. This way, SILKE templates get a global influence although they work locally, which results in a low computational effort and good real-time capabilities. But what is more, sufficient plasticity of learning is still given due to locality, although the learning process is guided towards global meta-level properties.

Incremental online learning in a closed-loop setting is potentially chaotic. But with the SILKE approach, the influence of learning on the behavior of an embedded system in its environment can be guided by correcting the learned knowledge locally towards higher smoothness or towards compliance with other local meta-level properties when other templates are used. Depending on the adjustment rate, this yields a system which is very flexible in the face of learning stimuli that comply the meta-properties, but which is also very robust against learning stimuli which violate these meta-properties. So for this case, self-adaptive systems can fight outliers and overcome the stability-plasticity-dilemma to a certain extent. Based on this it is clear that the focus of this work is not to achieve the highest possible modeling quality, but to stabilize adaptive control systems.

At system level, the presented approach is an effective and efficient method for controlling incremental online learning without limiting the required flexibility. Such a system can cope with changes online and in a reliable way. It is based on the use of rules. On the one hand this aids engineerability at design time. On the other hand, rules which have been modified by self-tuning can also be analyzed directly. By that good engineerability and high trustworthiness are achieved.

In future work, the interplay between closed-loop incremental online learning and the SILKE approach will be investigated especially for the application of other template functions to zero and first order sTS-fuzzy systems in order to achieve other meta-properties such as monotony and curvature. As before, this will be done in experiments as well as

ISBN: 978-989-95079-6-8

in formal analysis. In addition, the extension of the results to more than one dimension has to be addressed.

Acknowledgement

This project is funded by the German Research Foundation DFG in the framework of the *Organic Computing* priority research program under No. BR 1980/3-1.

References

[1] C. Müller-Schloer, C. von der Malsburg, and R. P. Würtz. Organic Computing. *Informatik-Spektrum*, 27(4):332–336, 2004.

[2] W. Brockmann. Online Machine Learning For Adaptive Control. *IEEE Int. Work. Emerging Technologies and Factory Automation, 1992*, pages 190–195, 1992.

[3] J. Farrell and M. Polycarpou. *Approximation Based Control: Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches*. Wiley Interscience, Hoboken, New Jersey, 2006.

[4] P. P. Angelov and D. P. Filev. An Approach to Online Identification of Takagi-Sugeno Fuzzy Models. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(1):484–498, 2004.

[5] T. A. Johansen and R. Murray-Smith. The operating regime approach to nonlinear modelling and control. *Multiple Model Approaches to Modelling and Control*, pages 3–72, 1997.

[6] T. A. Johansen. Robust identification of Takagi-Sugeno-Kang fuzzy models using regularization. In *Fifth IEEE Int. Conf. Fuzzy Systems*, volume 1, pages 180–193, 1996.

[7] J. Farrell and M. Polycarpou. *Adaptive Approximation Based Control*. Wiley–Interscience, 2006.

[8] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 1996.

[9] F. Bauer, S. Pereverzev, and L. Rosasco. On regularization algorithms in learning theory. *Journal of Complexity*, 23(1):52–72, 2007.

[10] A.E. Hoerl and R.W. Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 42(1):80–86, 2000.

[11] A. Tikhonov. On the stability of inverse problems. *CR (Dokl.) Acad. Sci. URSS, n. Ser.*, 39:176–179, 1943.

[12] E. Lughofer and S. Kindermann. Improving the Robustness of Data-driven Fuzzy Systems with Regularization. In *Proc. Int. Conf. Fuzzy Systems*, pages 703–709. IEEE, 2008.

[13] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, 1985.

[14] W. Brockmann and A. Horst. Stabilizing the Convergence of Online-Learning in Neuro-Fuzzy Systems by an Immune System-inspired Approach. In *Proc. IEEE Int. Conf. Fuzzy Systems, FUZZ-IEEE*, pages 351–356. IEEE, 2007.

[15] N. Rosemann and W. Brockmann. Concept for Controlled Self-optimization in Online Learning Neuro-fuzzy Systems. In J. Hertzberg, M. Beetz, and R. Englert, editors, *KI 2007: Advances in Artificial Intelligence, LNAI*, volume 4667, pages 498–501. Springer, Berlin, 2007.

[16] N. Rosemann, B. Neumann, and W. Brockmann. Formale Eigenschaften des SILKE-Ansatzes zur Kontrolle selbstoptimierender Systeme. In H.-G. Hegering and A. et al. Lehmann, editors, *Proc. Informatik 2008 Beherrschbare Systeme – dank Informatik, LNI*, volume 134, pages 755–762. Gesellschaft für Informatik, Bonn, 2008.

[17] W. Brockmann and N. Rosemann. Instantaneous Anomaly Detection in Online Learning Fuzzy-Systems. In F. Hoffmann and O. et al. Cordon, editors, *3rd Int. Workshop on Genetic and Evolving Fuzzy Systems*, pages 23–28. IEEE Press, Piscataway, 2008.

ISBN: 978-989-95079-6-8

Dynamic vs. static decision strategies in adversarial reasoning

David A. Pelta[1] Ronald R. Yager[2]
1. Models of Decision and Optimization Research Group
Department of Computer Science and A.I., University of Granada,
C/Periodista Daniel Saucedo s/n, 18071 Granada, Spain
2. Machine Intelligence Institute, Iona College
New Rochelle, NY 10801, USA
Email: dpelta@decsai.ugr.es, yager@panix.com

Abstract— Adversarial decision making is aimed at determining optimal decision strategies to deal with an adversarial and adaptive opponent. One defense against this adversary is to make decisions that are intended to confuse him, although our rewards can be diminished. It is assumed that making decisions in an uncertain environment is a hard task. However, this situation is of upmost interest in the case of adversarial reasoning as what we want is to *force the presence of uncertainty* in order to confuse the adversary in situations of repeated conflicting encounters. Using simulations, the use of dynamic vs. static decision strategies is analyzed. The main conclusions are: a) the use of the proposed dynamic strategies has sense, b) the presence of an adversary may produce a decrease of, at least, 45% with respect to the theoretical best payoff and c) the relation between this reduction and the way the uncertainty is forced should be further investigated.

Keywords— Adversarial reasoning, uncertain environment, decision strategies, simulation

1 Introduction

Adversarial decision is largely about understanding the minds and actions of one's opponent. It is relevant to a broad range of problems where the actors are actively and consciously contesting at least some of each others' objectives and actions [1]. The field is also known as decision making in the presence of adversaries or adversarial reasoning.

In its most basic form, adversarial decision making involves two participants, white and black, each of which chooses an action to respond to a given event without knowing the choice of the other. As a result of these choices, a payoff is assigned to the participants. When this scenario is repeated many times, i.e. situations of repeated conflicting encounters arise, then the situation becomes complex as the participants have the possibility to learn the others strategy. Examples of this type can be found in the military field, but also in problems of real-time strategy games, government vs government conflicts, economic adversarial domains, team sports (e.g., RoboCup), competitions (e.g., Poker), etc. [1]

Adversarial decision making is aimed at determining optimal strategies (for white) against an adversarial and adaptive opponent (black). One defense against this adversary is to make decisions that are intended to confuse him, although white's rewards can be diminished.

It is assumed that making decisions in an uncertain environment is a hard task. However, this situation is of upmost interest in the case of adversarial reasoning as what white wants is to make its behaviour as uncertain or unpredictable as possible. In other words, white wants to *force the presence of uncertainty* in order to confuse the adversary while its payoff is as less affected as possible.

In previous work [2], we proposed a model to study the balance between the level of confusion induced and the payoff obtained and we concluded that one way to produce uncertainty is through decision strategies for white that contain certain amount of randomness. Here we focus on learning strategies that white can use as a means of optimizing his payoffs in situations of repeated conflicting encounters. Essentially we are studying how white can defend against an opponent who is trying to learn their decision rules.

In this paper, we want to analyze the case where white's decision strategy is not constant along the time, but modified following certain rules. We explore two alternatives, one is to vary the number of candidates alternatives in terms of their associated payoffs and second is based on the basic concept of α-cuts, where the value of α is varied.

The contribution is organized as follows: some basic concepts on adversarial reasoning are outlined in Section 2. Then, Section 3 describes the main characteristics and components of the model used. Section 4 introduces static decision strategies for both agents and then, shows how they can be transformed into dynamic ones. In Section 5 we describe the computational experiments performed and the results obtained and finally, Section 6 is devoted to discussions and further work.

2 Adversarial Reasoning

As stated before, adversarial decision making is largely about understanding the minds and actions of one's opponent. A typical example is the threat of terrorism and other applications in Defense, but it is possible to envisage less dramatic applications in computer games where the user is the adversary and the computer characters are provided with adversarial reasoning features in order to enhance the quality, hardness and adaptivity of the game. The development of intelligent training systems is also an interesting field.

The threat of terrorism, and in particular the 9/11 event, fueled the investments and interest in the development of computational tools and techniques for adversarial reasoning. However, the field has earlier developments. For example, almost twenty years ago, P. Thagard [3] states

> In *adversarial problem solving*, one must anticipate, understand and counteract the actions of an opponent. Military strategy, business, and game playing all require an agent to construct a model of an

ISBN: 978-989-95079-6-8

opponent that includes the opponent's model of the agent.

Game theory is perceived as a natural good choice to deal with adversarial reasoning problems. For example, a brief survey of techniques where the combination of game theory with other approaches is highlighted, jointly with probabilistic risk analysis and stochastic games is presented in [4]

However, nowadays it is assumed that the field transcends the boundaries of game theory [1]. As stated in [5]: "we argue that practical adversarial reasoning calls for a broader range of disciplines: artificial intelligence planning, cognitive modeling, control theory, and machine learning in addition to game theory. An effective approach to problems of adversarial reasoning must combine contributions from disciplines that unfortunately rarely come together".

3 The model

The framework used to conduct our study is a slight modification of the previously proposed in [2].

It is based on two agents white W and black B (the adversary), a set of possible inputs or events $E = \{e_1, e_2, \ldots, e_n\}$ issued by a third agent R, and a fuzzy set of potential responses or actions $A_i = \{a_1, a_2, \ldots, a_m\}$ associated with every event. These fuzzy sets are organized as rows in a matrix P as :

$$P(n \times m) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ p_{31} & p_{32} & \cdots & p_{3m} \\ \\ p_{n1} & p_{n2} & \cdots & p_{nm} \end{pmatrix}$$

where $p_{ij} \in [0, 1]$ is the level of suitability of action j to respond to the event i. More precisely, p_{ij} is the degree of membership of action j to the fuzzy set of *suitable actions* associated with event e_i. We do not require these fuzzy sets to be normalized.

Agent W has a strategy to decide which action to take given a particular event e_k and perfect knowledge of matrix P. The aim for W is to maximize the sum of the profits or payoffs given a set of inputs. These inputs or events are issued one at a time by R and, in principle, they are independent.

The payoff of a given action is proportional to its suitability with respect to the given event. For the sake of simplicity, in this contribution we assume that the payoff is the value of suitability.

Agent B wishes to learn the actions that W is going to take given a particular input e_k so as to reduce agent W payoff. Agent B does not know matrix P. It has access to the decisions made previously by W and, if the guess matched the decision taken by W, then B obtains some reward. We may think the situation as an "imitation game", where the aim for W is to avoid being imitated.

A graphical view of the model is shown in Figure 1 while the whole procedure is described in Algorithm 1.

The payoff's calculation for W at stage j is defined as:

$$p' = p_{jk} \times F(a_g, a_k) \tag{1}$$

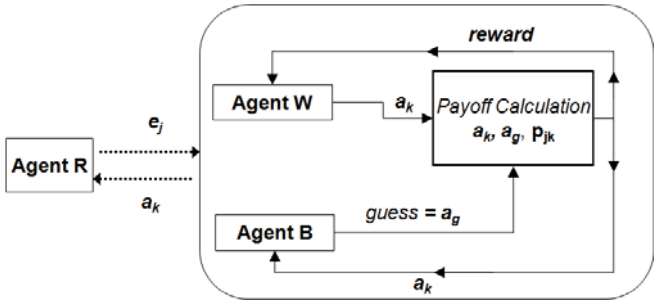


Figure 1: Graphical representation of the model. Events e_j are issued by agent R while response or actions a_k are taken by agent W.

Algorithm 1 Sequence of steps in the model.

for $j = 1$ to N **do**
  A new event e_j arises.
  Agent B "guesses" an action a_g
  Agent W determines an action a_k
  Calculate payoff for W
  Agent B records the pair e_j, a_k
end for

where F is:

$$F(a, b) = \begin{cases} 0 & \text{if a = b} \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

As stated before, the aim for W is to maximize the sum of the payoffs. In other words, a strategy for W should be aimed at defining a sequence of actions that gives the higher payoff while avoiding being correctly guessed.

The terms *event* and *action* used here should be understood in a broad sense. For example, an event may represent a particular simulation scenario while an action may represent a full plan. Also, the actions may represent a Dempster-Schaffer belief structure to reflect the ideas posed in [6].

4 Modeling the Behavior of the Agents

In this section, we provide alternatives for modeling the behavior of both agents. For simplicity, we assume that the inputs issued by agent R are equiprobable and that the number of inputs equals the number of actions (i.e, the payoff matrix is square).

4.1 Strategies for Agent B

Agent B applies a very simple frequency-based decision strategy. We define a matrix of observations O with $M \times M$ dimensions, where each O_{ij} stores the number of times that action i was observed (from W) when the event was e_j. Given an event e_j, the following decision strategy for B is used:

Proportional to the Frequency (PF): the probability of selecting an action i is proportional to O_{ij} (the observed frequency from agent W) [2].

4.2 Strategies for Agent W

Agent W knows he is being observed, so the idea is to change its behavior in order to confuse the adversary. The agent needs to take sub-optimal decisions in order to get benefits in the long term.

ISBN: 978-989-95079-6-8

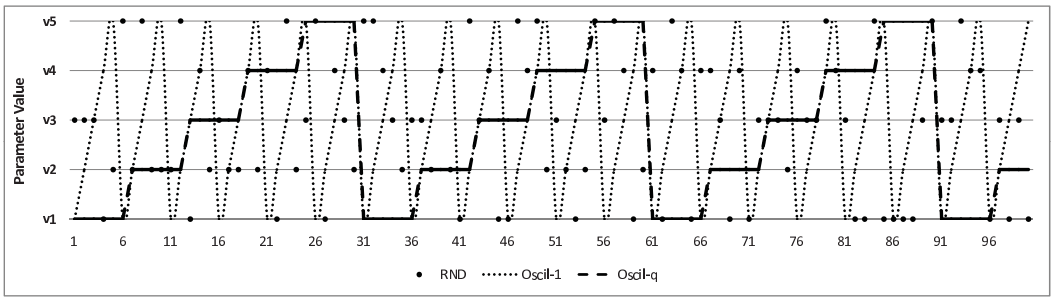


Figure 2: Adaptation schemes for the parameters in the dynamic strategies. The example shows the case where a parameter takes values from a set of possible discrete alternatives $\{v_1, v_2, v_3, v_4, v_5\}$

Given an observed input e_i and the suitability matrix P, we propose two basic strategies:

1. *Random-Among $< k >$ Best Actions (R-k-B)*: Select randomly an alternative among the k most suitable ones [2].
2. *Random-From α-cut (R-α)*: select randomly an alternative among those with a minimum level of suitability α.

We remark here that these strategies are not *mixed strategies* in the sense of game theory as we do not define a probability distribution over the set of alternatives. Once a subset of alternatives is selected, then any of them can be selected with equal probability.

It is clear that both parameters, k and α, can be used to control the amount of uncertainty in the action selection. Some specific values for those parameters lead to interesting behaviours:

- $k = 1$: always select the most suitable action.
- $k = M$: any action is considered equally suitable.
- $0 < k < M$ lead to a behaviour where suboptimal actions may have the chance of being selected.

Particular cases for α:

- $\alpha = 1$: always select the action whose level of suitability is 1. This is not a good strategy as it is easily learnable.
- $\alpha = 0$: any action is considered equally suitable.
- $0 < \alpha < 1$ lead to a behaviour where suboptimal actions may have the chance of being selected.

Although at first sight both strategies look similar, there is an important difference. The former strategy (*R-k-B*) is independent of the suitability scale: it just selects the k most suitable alternatives and then takes a decision. The later strategy is based on α-cuts, so if the value of α is high, then it may happen that the subset of alternatives of the corresponding α-cut becomes empty. In this case, an alternative would be randomly selected.

4.2.1 Dynamic Strategies

When α and k are assigned specific values, then a "static strategy" is obtained. Here, we propose a set of "dynamic strategies" where the values of the control parameter are varied along the time following the patterns shown in Fig. 2.

In order to adapt the parameter k, we propose three schemes:

1. *RND*: at each stage, k is a value from a uniform distribution in $[1, M]$.
2. *Oscil-1*: after each event, $k = k + 1$. When $k = M + 1$ then $k = 1$.
3. *Oscil-10*: the same as before but the value of k is changed after ten events.

For the adaptation of α we propose similar schemes:

1. *RND*: at each stage, α is selected randomly from the set $V = \{0.15, 0.3, 0.45, 0.6, 0.75\}$
2. *Oscil-1*: after each event, $i = i + 1$, $\alpha = V[i]$. When $i = M + 1$ then $i = 1$
3. *Oscil-10*: the same as before but the value of α is changed after ten events.

5 Experiments and Results

The aim of the experiment is to evaluate how W's payoff is affected when dynamic vs. static decision strategies are used.

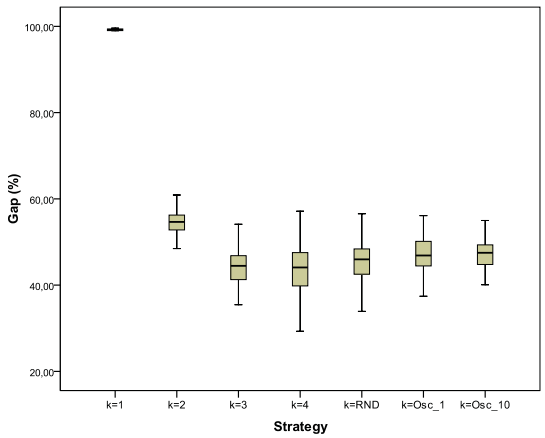
In order to do this, the following considerations are taken. We fix the number of events and alternatives to $M = 5$. Then, for every W's strategy we made 100 repetitions of the scheme shown in Algorithm 1, where $N = 500$. The matrix P is randomly generated each repetition. In this way we avoid potential biases due to particular configuration of values in P. At the end of each repetition, we record:

- Gap: the gap (as percentage) between the payoff obtained by W and the optimum (calculated as the sum of the payoffs associated with the best action for every event),
- $Guess$: the number of times that B correctly guessed the action of W (also as percentage with respect to 500 actions).

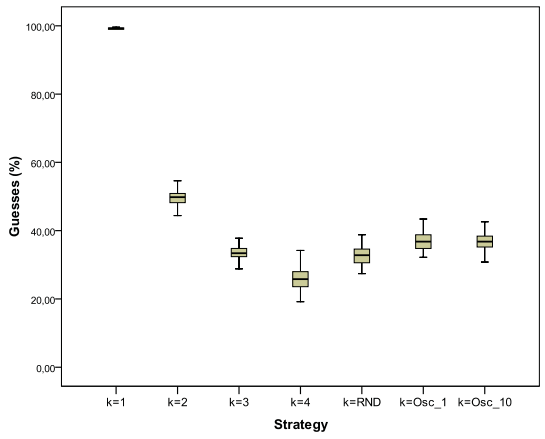
For both cases, the lower the values are, the better the performance of the strategy is.

We evaluated six dynamic strategies (three for α and three for k), four static strategies *R-k-B* using $k = 1, 2, 3, 4$ and five static strategies *R-α* using $\alpha = \{0.15, 0.3, 0.45, 0.6, 0.75\}$.

ISBN: 978-989-95079-6-8



(a)



(b)

Figure 3: Strategy *R-k-B*: average Gap (a) and $Guess$ (b) for each static and dynamic configuration

5.1 Results

In first place we will analyze the results for the *R-k-B*. Figure 3 shows boxplots for the average Gap and $Guess$ (the average number of correct guesses) when the parameter k is fixed or adapted during the repetition.

As it also occurred in [2], the value $k = 1$ is the worst alternative in terms of both measures. In this case, the best action is always selected, thus after a few events the frequency of those actions in the observations matrix kept by B are the only ones different than zero and B always choose them. As k increases, the performance is better on average but the standard deviation is increased. The proposed adaptive schemes for k do not show big differences among them. What is clear, is the reduction in the standard deviation from the $k = RND$ strategy to the one that force oscillations for k every ten events. In terms of the $Guess$ measurement, it is interesting to note that a variation of this single parameter can decrease the number of correct predictions from 50% when $k = 2$ to a lowest value of 25% when $k = 4$.

In order to assess if the differences among the strategies in the average gap have statistical significance, we performed an ANOVA test followed by a post-hoc analysis using Tamhane's test with $p < 0.05$. The results are shown in Table 1; the signs indicate that the average gap between strategies (i, j) (row,col) is different with statistical significance. A '+ sign indicates that strategy i is better, while a '- denotes that j is better. Absence of symbol indicates that the difference had no statistical significance. In this case, the best results are obtained by static strategies with parameters $k = 3$ and $k = 4$. The strategy RND shows a similar performance, however it is not so good to outperform the other dynamic strategies.

R-k-B	1	2	3	4	RND	Osc-1	Osc-10
1		-	-	-	-	-	-
2	+		-	-	-	-	-
3	+	+				+	+
4	+	+				+	+
RND	+	+					
Osc-1	+	+	-	-			
Osc-10	+	+	-	-			

Table 1: Summary of the statistical testing for *R-k-B* strategies. See text for details.

R-α	0.15	0.30	0.45	0.60	0.75	RND	Osc-1	Osc-10
0.15			+	+	+			
0.30			+	+	+			-
0.45	-	-		+	+	-	-	-
0.60	-	-	-		+	-	-	-
0.75	-	-	-	-		-	-	-
RND			+	+	+			
Osc-1			+	+	+			
Osc-10		+	+	+	+			

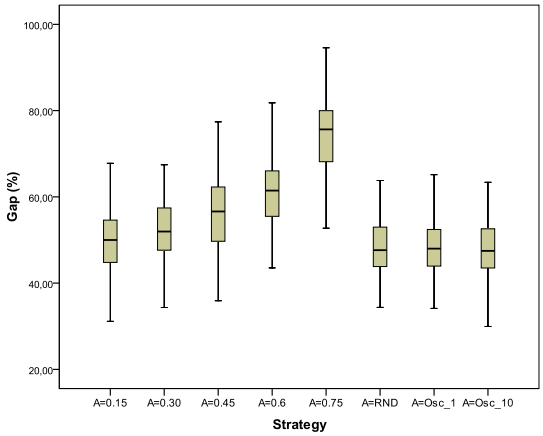
Table 2: Summary of the statistical testing for *R-*α strategies. See text for details.

A similar analysis can be done for the *R-*α strategy. Figure 4 shows boxplots for the average Gap and $Guess$ when the parameter α is fixed or adapted during the repetition. The label A in the plots stands for α.

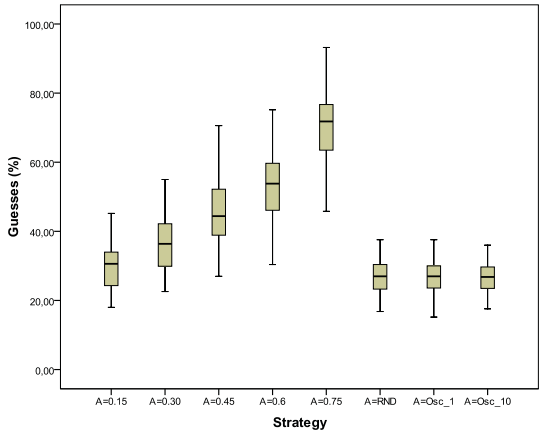
As α increases, the performance becomes worst. The reason is related with the way the matrix P is generated. It may happen that no action has a level of suitability with a degree of membership higher than α. In other words, the corresponding α-cut leads to an empty set of alternatives. In this case, a pure random decision strategy is taken. The counterpart is that now, the proposed adaptive schemes for α are clearly better than the static alternatives. Between the adaptive schemes, no clear differences appeared.

As before, in order to assess if the differences among the strategies in the average Gap have statistical significance, we performed statistical testing. The results are shown in Table 2. It can be confirmed that when using a static strategy, lower values of α like 0.15 or 0.3 led to better results than those obtained with higher ones. However, any of the dynamic strategies proposed, even RND, obtained the same average Gap. Moreover, the strategy *Osc-10* (that changed the α value every ten events) provided better performance than all the static strategies, excepting $\alpha = 0.15$.

ISBN: 978-989-95079-6-8



(a)



(b)

Figure 4: Strategy *R*-α: average Gap (a) and $Guess$ (b) for each static (A stands for α) and dynamic configuration

It is reasonable to assume that a relation exist between the number of correct guesses and the payoff obtained. If the former is high, it is clear that the payoff should be low (this is the case when $k = 1$). But what happen when the number of correct guesses is low? What are the payoff values that can be reached?. Figure 5 shows the average of each measure for every strategy (the case when $k = 1$ is omitted for visualization purposes). Within each kind of strategy (*R-k-B*, R-α), the alternatives are ordered increasingly in terms of Gap.

It is clear that just reducing the number of guesses is not enough to improve the gap. For example, *A=Osc-10* achieved a lower value of $Guess$ than $k = 3$ but a higher value of Gap. This was also already notice in [2], where a purely random strategy led to the lowest value of $Guess$. Interestingly, all the cases where the curve for $Guess$ stands below the one for Gap correspond to strategies based on α-cuts. The reason is not clear and further investigations are needed.

To conclude the analysis, Figure 6 shows several scatter plots displaying the relation between both measures. Every point correspond to a particular repetition, thus 100 point per plot are displayed. Plots on the left correspond to *R*-α strategy while those on the right to *R-k-B*. Three static values are shown per each parameter, while the bottom plot corresponds to the best dynamic strategy. The X axis is the Gap while Y axis shows the measure $Guess$.

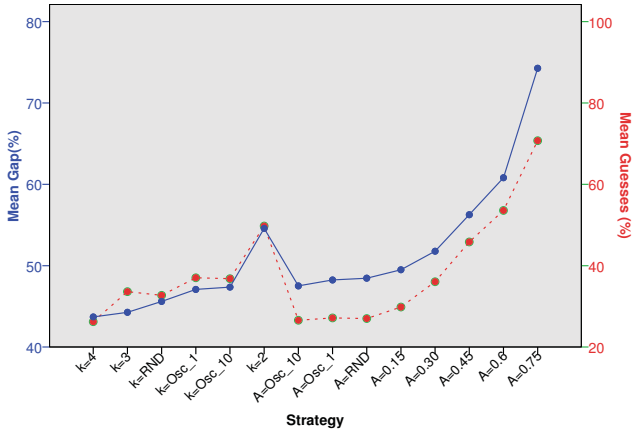


Figure 5: Average Gap (left Y axis) and $Guess$ (right Y axis) for every strategy. The dotted line is used for $Guess$.

Several elements can be observed in the Figure. Let's start with the *R*-α strategy (left). As α increases, we can observe a progressive concentration of the points towards the upper right corner of the plot. This means that the results are getting worse: more guesses, higher gap (agent W is getting much lower payoff than the potential optimum). It is noticeable the high variation the results are showing. For instance, when $\alpha = 0.45$ (second plot, in the left), one can observe simulation with guesses around 30% and gap around $30\% - 45\%$ and others with guesses higher than 60% and gap higher than 60%. When a simple dynamic strategy is used ($\alpha = RND$), then the number of correct guesses is almost always below 30%. However, the range of values that can be obtained for gap value is quite wide, going from 30% to 60%.

The plots are a bit different for the *R-k-B* strategy. As k increases, the percentage of guesses decrease. This is reasonable taking into account how the strategy works: take the k best alternatives and then, choose random. In the experiments performed, the total number of actions is 5, so, when $k = 4$, then just the worst strategy is eliminated. In other words, as k increases, agent B may conclude that W is behaving randomly. In turn, the gap value is more variable, ranging from 35% to 55% approx.

6 Discussion and Future Work

In this work we focused in the context of adversarial reasoning where a player W wants to *force the presence of uncertainty* in order to confuse the adversary B while its payoff is as less affected as possible in situations of repeated conflicting encounters. We extended a previous work where the decision strategies were fixed along the time, to consider dynamic strategies: the way the decisions are taken varies with the time.

When the decision strategy is based on α-cuts, then a dynamic variation of α led to better results than the static counterpart. Moreover, none of the adaptation schemes proposed led to worst results than those obtained by a specific configuration of the control parameter. However, in the $R - k - B$ case, the results are not so clear but still good. The best results are obtained with $k = 3, k = 4$. Using any of these values, the strategy obtained performed better than four of the other

ISBN: 978-989-95079-6-8

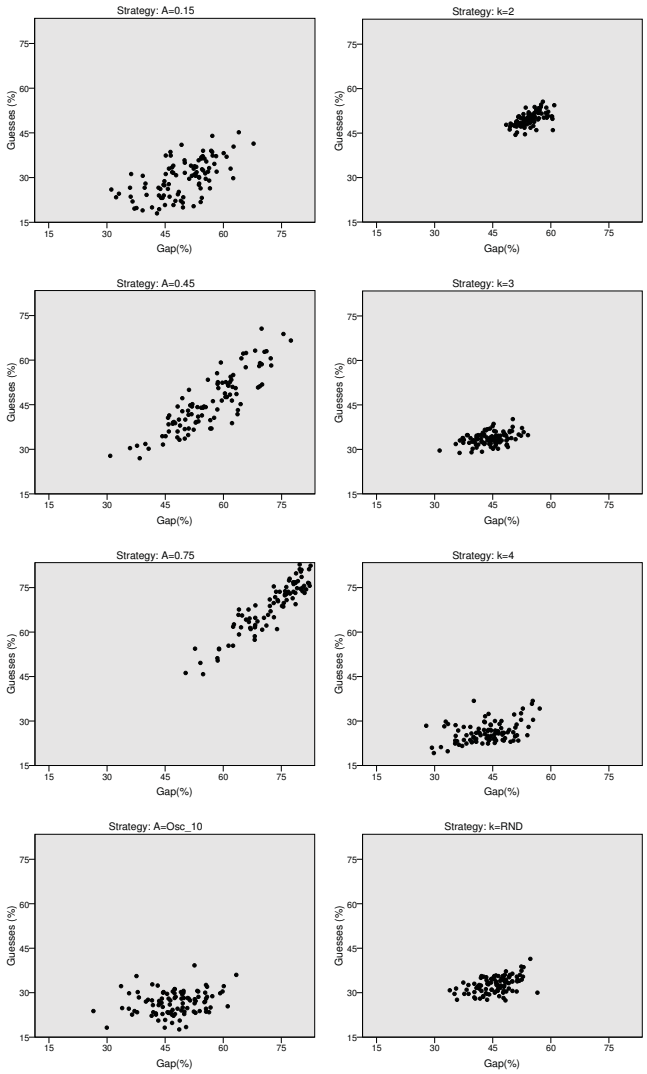


Figure 6: Every point represents the result $(Gap, Guess)$ of a repetition.

strategies available. The best dynamic alternative is RND which assigns a random value to k before applying the strategy. This strategy achieved a similar performance (on average) with respect to $k = 3, k = 4$.

The use of dynamic strategies, besides its potential performance, has another benefit which is to avoid determining a specific value for the control parameter. This is not a trivial matter as there is no simple way to infer which setting will provide the best value for a different simulation scenario.

The study performed also revealed an interesting point: the lowest average Gap value for a dynamic strategy was 45% when using $k = RND$ and this implies that W's payoff is 45% lower than the theoretical optimum one. In [2], we showed that using a purely random decision strategy, the gap is higher than 50%. The question now is: would it be possible to design a strategy with guaranteed (in a formal sense) performance?.

The way the dynamism is included in this proposal is quite simple, but several alternatives are available for producing improvements. Now, agent W is not analyzing the payoff obtained against the one expected (which is stored in matrix P). Agent W can use the fact that its payoff is being affected in order to produce an intelligent adaptation of its strategy. In this context, the use of fuzzy rules may play an important role as it would help to model adaptation strategies as: *if the reduction of payoff is high, then increase uncertainty*, or *if this event occurred a high number of times then increase uncertainty*. Other lines of research are related with the model used. Here, the way the payoff matrix is generated may affect the results obtained by the strategy based on α-cuts as the differences between the suitability of alternatives may be low. Other ways to define such matrix may be necessary. Also, the sequence of events is completely random now, but other options are available.

The relation between the number of correct guesses and the gap value is clear in some situations but in others (as shown in Fig. 6), it is possible to obtain a broad range of gap values for the same number of guesses. In this context, it seems not trivial to use the number of guesses as a predictor for the gap value. A different alternative may take into account that the presence of uncertainty is reflected in the matrix of observations O that agent B construct, so it would be interesting to look for correlations between the gap value and some measures about this matrix O. The ideas posed here are left as future work.

Finally, we would like to mention that one of the reviewers claimed that the problem posed here can be solved with the standard tools of game theory arguing that the problem can be seen as a sequence of n independent non cooperative games. However, we do not think that this is the case as, for example, the matrix of observations changes at each step, we are not dealing with mixed strategies and so on. Moreover, we claim that this kind of analysis can be applied when the events are correlated in some way, and also, when more sophisticated adaptation and learning mechanisms are considered in both agents.

Acknowledgments

This work is supported in part by projects TIN2008-01948 from the Spanish Ministry of Science and Innovation, and P07-TIC-02970 from the Andalusian Government.

References

[1] A. Kott and W. M. McEneany. *Adversarial Reasoning: Computational Approaches to Reading the Opponents Mind*. Chapman and Hall/ CRC Boca Raton, 2007.

[2] D. Pelta and R. Yager. On the conflict between inducing confusion and attaining payoff in adversarial decision making. *Information Science*, 179:33–40, 2009.

[3] P. Thagard. Adversarial problem solving: Modeling an opponent using explanatory coherence. *Cognitive Science*, 16(1):123 – 149, 1992.

[4] E. Kardes and R. Hall. Survey of literature on strategic decision making in the presence of adversaries. Report 05-006, National Center for Risk and Economic Analysis of Terrorism Events, 2005.

[5] A. Kott and M. Ownby. Tools for real-time anticipation of enemy actions in tactical ground operations. In *Proceedings of the 10th International Command and Control Research and Technology Symposium*, 2005.

[6] R. R. Yager. A knowledge-based approach to adversarial decision making. *International Journal of Intelligent Systems*, 23:1–21, 2008.

ISBN: 978-989-95079-6-8

Generalized Atanassov's Intuitionistic Fuzzy Index. Construction Method

Edurne Barrenechea Humberto Bustince Miguel Pagola Javier Fernandez Josean Sanz

Dpt. Automática y Computación, Universidad Pública de Navarra
Campus de Arrosadía, s/n
31006 Pamplona (Spain)
Email: {edurne.barrenechea,bustince,miguel.pagola,fcojavier.fernandez,joseantonio.sanz}@unavarra.es

Abstract— In this work we introduce the concept of *Generalized Atanassov's Intuitionistic Fuzzy Index*. We characterize it in terms of fuzzy implication operators and we propose a construction method with automorphisms. Finally, we study some special properties of the generalized Atanassov's intuitionistic fuzzy index.

Keywords— Atanassov's Intuitionistic Fuzzy Index, Fuzzy Implication Operator, Automorphisms.

1 Introduction

In 1983 Atanassov [1] introduced Atanassov's intuitionistic fuzzy sets ($A-IFSs$) in such a way that each element of the set has two values assigned, the membership degree and the non-membership degree. In this direction, Atanassov defined these sets also indicating that the index of intuitionism of each element obtained by subtracting the sum of membership and non-membership from one (a subtraction that should be positive and less than or equal to one), it is a measurement of the effect of working with Atanassov's intuitionistic fuzzy sets. We consider this intuitionistic index (or condition of intuitionism) a very important characteristic of $A-IFSs$ since from it we can obtain very valuable information of each element and taking on advantage of this potentiality in different applications. For example, in image processing the task of divide into disjoint parts a digital image is denoted as segmentation. The most commonly used strategy for segmenting images is global thresholding that refers to the process of dividing the pixels in an image on the basis of their intensity levels of gray. The experts have uncertainty when assigning the pixels either to the background or to the object trough the choice of the membership functions. Moreover, this choice has proven to be of uttermost importance regarding the algorithms performance. In order to overcome this problem, we consider using the Atanassov's intuitionistic fuzzy index values for representing the uncertainty of the expert in determining that the pixel belongs to the background or that it belongs to the object. From this point of view, we can consider the expert provides the degree of membership of an element to an $A-IFS$ and also the degree of intuitionism the expert has in given this membership degree (see [9]). This fact has led us to present the new concept of *Generalized Atanassov's Intuitionistic Fuzzy Index* that generalize the expression given by Atanassov. We also provide a characterization method by means of fuzzy implication operators. Moreover, we study a construction method using automorphisms that allow us to present simple expressions of said index.

Pankowska and Wygralak ([16, 17, 21]) proposed another generalization of the intuitionistic index based on strong negations and triangular norms. This approach is used to construct flexible algorithms of group decision making which involve relative scalar cardinalities defined by means of generalized sigma counts of fuzzy sets.

2 Preliminary definitions

Let U be an ordinary finite non-empty set. An *Atanassov's intuitionistic fuzzy set* ($A-IFS$) [1] in U is an expression A given by

$$A=\{(u,\mu_A(u),\nu_A(u))|u\in U\} \tag{1}$$

where

$$\begin{array}{l}\mu_A\ :\ U\longrightarrow[0,1]\\ \nu_A\ :\ U\longrightarrow[0,1]\end{array}$$

satisfy the condition $0\leq\mu_A(u)+\nu_A(u)\leq 1$ for all u in U.

The numbers $\mu_A(u)$ and $\nu_A(u)$ denote respectively the degree of membership and the degree of non-membership of the element u in set A. We will also use the notation $A(u)=(\mu_A(u),\nu_A(u))$. We will represent as $A-IFSs(U)$ the set of all the Atanassov's intuitionistic fuzzy sets in U.
Atanassov defined the *Atanassov's intuitionistic fuzzy index* of the element u in $A\in A-IFSs(U)$ as:

$$\Pi_A(u)=1-\mu_A(u)-\nu_A(u). \tag{2}$$

We know fuzzy sets are represented exclusively by the membership function degree, that is,

$$A=\{(u,\mu_A(u))|u\in U\}. \tag{3}$$

Hereinafter, fuzzy sets have associated a non-membership degree given by one minus the membership degree:

$$\begin{array}{ll}A= & \{(u,\mu_A(u),\nu_A(u))|u\in U\}=\\ & \{(u,\mu_A(u),1-\mu_A(u))|u\in U\}.\end{array} \tag{4}$$

Since $\mu_A(u)+\nu_A(u)=\mu_A(u)+1-\mu_A(u)=1$, in this sense fuzzy sets are considered as a particular case of Atanassov's intuitionistic fuzzy sets. We will represent as $FSs(U)$ the set of all the fuzzy sets in U.

We will call automorphism of the unit interval every function $\varphi:[0,1]\to[0,1]$ that is continuous and strictly increasing such that $\varphi(0)=0$ and $\varphi(1)=1$.

A function $n:[0,1]\to[0,1]$ such that $n(0)=1$ and $n(1)=0$ is called a *strong negation* whenever it is strictly decreasing, continuous and involutive. Trillas ([19, 20]) proved that $n:[0,1]\to[0,1]$ is a strong negation if and only if there exists an automorphism φ of the unit interval such that

ISBN: 978-989-95079-6-8

$n(x) = \varphi^{-1}(1 - \varphi(x))$. In this work we will only consider strong negations.

We denote as L^* the following set:

$$L^* = \{(x, y) | (x, y) \in [0, 1] \times [0, 1] \; and \; x + y \leq 1\} \quad (5)$$

and the elements $0_{L^*} = (0, 1)$ and $1_{L^*} = (1, 0)$.

For every $(x, y), (z, t) \in L^*$ the following expressions are known ([1]-[8],[12]-[14]):

- $(x, y) \leq_{L^*} (z, t) \; if \; and \; only \; if \; x \leq z \;\; and \; y \geq t$. This relation is transitive, reflexive and antisymmetric.
- $(x, y) = (z, t) \;\; if \;\; and \;\; only \;\; if \;\; (x, y) \leq_{L^*} (z, t) \;\; and \;\; (z, t) \leq_{L^*} (x, y)$.
- $(x, y) \preceq (z, t) \; if \; and \; only \; if \; x \leq z \;\; and \; y \leq t$.

  In [14] is proven $(L^*, \leq_{L^*})$ is a complete lattice.

Therefore, if $A \in FSs(U)$, then

$$A_c = \{(u, 1 - \mu_A(u), \mu_A(u)) | u \in U\}. \quad (6)$$

Next, we recall the definition of Atanassov's intuitionistic fuzzy t-norm and t-conorm and also the notion of t-representability, in such a way, Deschrijver *et al.* gave a construction method of Atanassov's intuitionistic fuzzy t-norms and t-conorms by means of t-norms and t-conorms on $[0, 1]$ (see [14]).

Definition 1 *A function* $\mathbf{T} : (L^*)^2 \to L^*$ *is said to be an Atanassov's intuitionistic fuzzy t-norm if it is commutative, associative, and increasing (in both arguments with respect to the order* $\leq_{L^*}$*), with neutral element* 1_{L^*}*. In the same way, a function* $\mathbf{S} : (L^*)^2 \to L^*$ *is said to be an Atanassov's intuitionistic fuzzy t-conorm if it is commutative, associative, increasing and with neutral element* 0_{L^*}*.*

Definition 2 *An Atanassov's intuitionistic fuzzy t-norm* $\mathbf{T}$ *is called t-representable if and only if there exist a t-norm* T *and a t-conorm* S *on* $[0, 1]$ *such that, for all* $(x, y), (z, t) \in L^*$

$$\mathbf{T}((x, y), (z, t)) = (T(x, z), S(y, t)) \in L^*. \quad (7)$$

An Atanassov's intuitionistic fuzzy t-conorm $\mathbf{S}$ *is called t-representable if and only if there exist a t-norm* T *and a t-conorm* S *on* $[0, 1]$ *such that, for all* $(x, y), (z, t) \in L^*$

$$\mathbf{S}((x, y), (z, t)) = (S(x, z), T(y, t)) \in L^*. \quad (8)$$

Example 1 *Let be* $(x, y), (z, t) \in L^*$, $T = min$ *and* $S = max$ *on* $[0, 1]$

$$(a) \quad \mathbf{min}((x, y), (z, t)) = (min(x, z), max(y, t)). \quad (9)$$

$$(b) \quad \mathbf{max}((x, y), (z, t)) = (max(x, z), min(y, t)). \quad (10)$$

Definition 3 *([10, 14]) An Atanassov's intuitionistic fuzzy negation is a function* $\mathbf{n} : L^* \to L^*$ *that is decreasing (with respect to* $\leq_{L^*}$*) such that* $\mathbf{n}(0_{L^*}) = 1_{L^*}$ *and* $\mathbf{n}(1_{L^*}) = 0_{L^*}$*. If for all* $(x, y) \in L^*$ $\mathbf{n}(\mathbf{n}((x, y))) = (x, y)$ *it is said that* $\mathbf{n}$ *is involutive.*

The characterization of the Atanassov's intuitionistic fuzzy negations and the following result are presented in [14], for interval-valued fuzzy sets are proven in [10].

Theorem 1 *A function* $\mathbf{n} : L^* \to L^*$ *is an involutive Atanassov's intuitionistic fuzzy negation if and only if there exists an involutive fuzzy negation* n *such that*

$$\mathbf{n}((x, y)) = (n(1 - y), 1 - n(x)). \quad (11)$$

Throughout this work we will restrict to involutive Atanassov's intuitionistic fuzzy negations $\mathbf{n}$ generated from a given negation n, as in Theorem 1.

3 Generalized Atanassov's Intuitionistic Fuzzy Index

In this section we propose the definition of generalized Atanassov's intuitionistic fuzzy index and we characterize such index by means of fuzzy implication operators and automorphisms.

Definition 4 *A function* $\Pi_G : L^* \to [0, 1]$ *is called a generalized Atanassov's intuitionistic fuzzy index associated with the strong negation* n*, if it satisfies the following conditions:*

(i) $\Pi_G((x, y)) = 1$ *if and only if* $x = 0$ *and* $y = 0$*;*

(ii) $\Pi_G((x, y)) = 0$ *if and only if* $x + y = 1$*;*

(iii) If $(z, t) \preceq (x, y)$*, then* $\Pi_G((x, y)) \leq \Pi_G((z, t))$*;*

(iv) $\Pi_G((x, y)) = \Pi_G(\mathbf{n}((x, y)))$ *for all* $(x, y) \in L^*$ *such that* $\mathbf{n}$ *is generated from an involutive fuzzy negation* n*, as in Theorem 1.*

Example 2

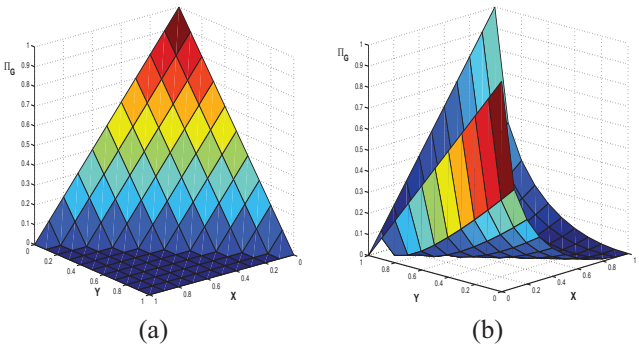
$$(a) \quad \Pi_G((x, y)) = 1 - y - x, \quad (12)$$

with $n(x) = 1 - x$ *for all* $x \in [0, 1]$*. As we can observe, this expression is equal to the expression (2) given by Atanassov.*

$$(b) \quad \Pi_G((x, y)) = ((1 - y)^{0.5} - x^{0.5})^2 \quad (13)$$

with $n(x) = (1 - x^{0.5})^2$ *for all* $x \in [0, 1]$.

Fig. 1 depicts the expressions given by Example 2.



(a) (b)

Figure 1: (a) $\Pi_G((x, y)) = 1 - y - x$ (b) $\Pi_G((x, y)) = ((1 - y)^{0.5} - x^{0.5})^2$

Next, we study the symmetry property of Π_G and we prove that only it is satisfied if we take as n the standard negation. Later, we introduce a construction method of the generalized Atanassov's intuitionistic fuzzy index using two automorphisms. In this way, it is quite simple to give different expressions of the intuitionistic index. Moreover, we analyze the case if the two automorphisms that we are taking are equal.

ISBN: 978-989-95079-6-8

Proposition 1 *Let Π_G be a generalized Atanassov's intuitionistic fuzzy index associated with the strong negation n. Then,*

$$\Pi_G((x,y)) = \Pi_G((y,x))$$

if and only if

$$n(x) = 1 - x;\ that\ is, \mathbf{n}((x,y)) = (y,x).$$

Proof. If $\Pi_G((x,y)) = \Pi_G((y,x))$ and by means of item (iv) of Definition 4 and Theorem 1, we obtain, $\Pi_G((x,y)) = \Pi_G((y,x)) = \Pi_G(\mathbf{n}((x,y))) = \Pi_G((n(1-y), 1-n(x)))$. Therefore, $n(1-y) = y$ and $n(x) = 1-x$.
On the other side, if n is the standard negation, then $\Pi_G((x,y)) = \Pi_G(\mathbf{n}((x,y))) = \Pi_G((n(1-y), 1-n(x))) = \Pi_G((y,x))$.

Proposition 2 *If φ_1, φ_2 are two automorphisms of the unit interval, then*

$$\Pi_G((x,y)) = \varphi_1^{-1}(\varphi_2(1-y) - \varphi_2(x)) \tag{14}$$

with $n(x) = \varphi_2^{-1}(1 - \varphi_2(x))$ is a generalized Atanassov's intuitionistic fuzzy index associated with the strong negation n.

Proof. We must proof the four properties of a generalized Atanassov's intuitionistic fuzzy index. The proof of properties (i) and (ii) is direct bearing in mind the bound conditions of automorphisms definition. (iii) if $(z,t) \preceq (x,y)$ then, $z \leq x$ and $t \leq y$. Automorphisms are strictly increasing therefore, $\varphi_2(z) \leq \varphi_2(x)$ and $\varphi_2(1-y) \leq \varphi_2(1-t)$, so we obtain $\Pi_G((x,y)) \leq \Pi_G((z,t))$. (iv) Bearing in mind Teorem 1 and also $n(x) = \varphi_2^{-1}(1-\varphi_2(x))$, we have $\Pi_G(\ \mathbf{n}((x,y))) = \Pi_G((n(1-y), 1-n(x))) = \varphi_1^{-1}(\varphi_2(n(x)) - \varphi_2(n(1-y))) = \varphi_1^{-1}(\varphi_2(\varphi_2^{-1}(1-\varphi_2(x))) - \varphi_2(\varphi_2^{-1}(1-\varphi_2(1-y)))) = \Pi_G((x,y))$.

Proposition 3 *In the conditions of Proposition 2:*

$$\Pi_G((0,y)) = 1 - y$$

if and only if

$$\varphi_1(x) = \varphi_2(x) \quad for\ all \quad x \in [0,1].$$

Proof. Direct.

For us, a fuzzy implication operator will be an implication in the sense of Fodor and Roubens [15], that is, a function $I : [0,1]^2 \to [0,1]$ that satisfies the following properties:

I_1. If $x \leq z$ then $I(x,y) \geq I(z,y)$ for all $y \in [0,1]$;

I_2. If $y \leq t$ then $I(x,y) \leq I(x,t)$ for all $x \in [0,1]$;

I_3. $I(0,x) = 1$ (dominance of falsity) for all $x \in [0,1]$;

I_4. $I(x,1) = 1$ for all $x \in [0,1]$.

I_5. $I(1,0) = 0$.

Depending on the application the following properties can also be demanded to a fuzzy implication operator:

I_6. $I(1,x) = x$ (neutrality of truth).

I_7. $I(x, I(y,z)) = I(y, I(x,z))$ (exchange property).

I_8. $I(x,y) = 1$ if and only if $x \leq y$.

I_9. $I(x,y) = I(n(y), n(x))$ (contraposition) with a strong negation n.

I_{10}. I is a continuous function.

In [6] is provide the following important result.

Proposition 4 *Let $I : [0,1]^2 \to [0,1]$. For all $x, y, z \in [0,1]$, the following properties hold:*

(i) *I satisfies I_1 if and only if $I(max(x,y), z) = min(I(x,z), I(y,z))$.*

(ii) *I satisfies I_1 if and only if $I(min(x,y), z) = max(I(x,z), I(y,z))$.*

(iii) *I satisfies I_2 if and only if $I(x, min(y,z)) = min(I(x,y), I(x,z))$.*

(iv) *I satisfies I_2 if and only if $I(x, max(y,z)) = max(I(x,y), I(x,z))$.*

Next, we will give a characterization of the generalized Atanassov's intuitionistic fuzzy index Π_G by means of a function $I : [0,1]^2 \to [0,1]$ that satisfies some properties of fuzzy implication operators.

Theorem 2 *Let n be a strong negation. A function $\Pi_G : L^* \to [0,1]$ is a generalized Atanassov's intuitionistic fuzzy index associated with the strong negation n*

if and only if

there exists a function $I : [0,1]^2 \to [0,1]$ satisfying $I1, I8, I9$, and $I(x,y) = 0$ if and only if $x = 1$ and $y = 0$, such that

$$\Pi_G((x,y)) = n(I(1-y, x)).$$

Proof. Sufficiency. We need to prove the four properties of Definition 4. Properties (i) and (ii) are direct. Property (iii) we must consider the result proven in [6] that if I satisfies I_1 and I_9, then I satisfies I_2. Property (iv) is direct bearing in mind that I satisfies I_9 and n is a strong negation. In order to prove necessity, let us suppose that Π_G is a generalized Atanassov's intuitionistic fuzzy index associated with the strong negation n. Define I

$$I(x,y) = \begin{cases} 1 \text{ if } x \leq y \\ n(\Pi_G((y, 1-x))) \text{ if } x > y. \end{cases} \tag{15}$$

Now it is easy to prove that I satisfies I_1, I_8, I_9 and also $I(x,y) = 0$ if and only if $x = 1$ and $y = 0$.

Example 3 *We build these generalized Atanassov's intuitionistic fuzzy indices associated with the strong negation $n(x) = 1 - x$ for all $x \in [0,1]$.*

(a) Lukasiewicz implication: $I(x,y) = min(1, 1-x+y)$. Then,

$$\Pi_G((x,y)) = n(I(1-y,x)) =$$

$$n(min(1, 1-1+y+x)) = n(y+x) = 1-y-x. \tag{16}$$

(b) Fodor implication:

$$I(x,y) = \begin{cases} 1 & \text{if } x \leq y \\ max(1-x,y) & \text{if } x > y \end{cases} \quad (17)$$

In this case we have

$$\Pi_G((x,y)) = \begin{cases} 0 & \text{if } x+y=1 \\ n(max(y,x)) & \text{if } x+y<1. \end{cases} \quad (18)$$

Corollary 1 *Let n be a strong negation. A continuous function $\Pi_G : L^* \to [0,1]$ such that $\Pi_G((x,0)) = n(x)$ for all $x \in [0,1]$ is a generalized Atanassov's intuitionistic fuzzy index associated with the function $I : [0,1]^2 \to [0,1]$ with I satisfying $I7$ and $I10$*

if and only if

there exists an automorphism φ of the unit interval such that

$$\Pi_G((x,y)) = \varphi^{-1}(\varphi(1-y) - \varphi(x)) \text{ and}$$

$$n(x) = \varphi^{-1}(1-\varphi(x))$$

Proof. It's enough to take into account Theorem 2, the relations between the properties of I studied in [6] and the following theorem proved in 1987 by Smets and Magrez (see [18]): a function $I : [0,1]^2 \to [0,1]$ verifies $I2$, $I7$, $I8$ and $I10$ if and only if there exists an automorphism φ of the unit interval such that $I(x,y) = \varphi^{-1}(min(1, 1-\varphi(x)+\varphi(y)))$.

Example 4 *if $\varphi(x) = x^2$ for all $x \in [0,1]$, then*

$$\Pi_G((x,y)) = ((1-y)^2 - x^2)^{0.5}, \quad (19)$$

with $n(x) = (1-x^2)^{0.5}$ for all $x \in [0,1]$.

Next, we provide the expressions of Π_G when it is applied to the meet operator min and the join operator max on L^*.

Theorem 3 *Under conditions of Theorem 2 the following items hold:*

$$(1)\ \Pi_G(\mathbf{max}((x,y),(z,t))) = min\{max(\Pi_G((x,y)), \Pi_G((x,t))), max(\Pi_G((z,y)), \Pi_G((z,t)))\}$$

$$(2)\ \Pi_G(\mathbf{min}((x,y),(z,t))) = max\{min(\Pi_G((x,y)), \Pi_G((x,t))), min(\Pi_G((z,y)), \Pi_G((z,t)))\}$$

Proof. (1) We must take into account the definition of meet operator min and the join operator max on L^*, also if I satisfies I_1 and I_9, then I satisfies I_2 and finally the items (i), (iv) of Proposition 4.

$\Pi_G(\mathbf{max}((x,y),(z,t))) = \Pi_G((max(x,z), min(y,t))) = n(I(1-min(y,t), max(x,z)))$. Therefore,

$$\begin{array}{ll} n(\Pi_G((max(x,z), min(y,t)))) & = \\ I(1-min(y,t), max(x,z)) & = \\ I(max(1-y,1-t), max(x,z)) & = \\ max\{I(max(1-y,1-t),x), & \\ I(max(1-y,1-t),z)\} & = \\ max\{min(I(1-y,x), I(1-t,x)), & \\ min(I(1-y,z), I(1-t,z))\} & = \\ max\{min(n(\Pi_G((x,y))), n(\Pi_G((x,t)))), & \\ min(n(\Pi_G((z,y))), n(\Pi_G((z,t))))\} & = \\ max\{n(max(\Pi_G((x,y)), \Pi_G((x,t)))), & \\ (n(max(\Pi_G((z,y)), \Pi_G((z,t))))\} & = \\ n(min\{max(\Pi_G((x,y)), \Pi_G((x,t))), & \\ max(\Pi_G((z,y)), \Pi_G((z,t)))\}). & \end{array}$$

(2) It is quite similar to the previous one but we must take into account items (ii) and (iii) of Proposition 4.

Note the result presented in Theorem 3 is not valid for any t-norm, t-conorm in L^*. The reason is that we have not defined the order relation $\leq_{L^*}$ in Definition 4, only we have defined the item (iii) If $(z,t) \preceq (x,y)$, then $\Pi_G((x,y)) \leq \Pi_G((z,t))$. That is, $\mathbf{max}((x,y),(z,t)) = (max(x,z), min(y,t)) \leq_{L^*} (S(x,z), T(y,t)) = \mathbf{S}((x,y),(z,t))$ and we obtain that $\Pi_G(\mathbf{max}((x,y),(z,t)))$ and $\Pi_G(\mathbf{S}((x,y),(z,t)))$ are imcomparable.

4 Conclusions

We define the concept of *generalized Atanassov's intuitionistic fuzzy index* and we give different construction methods. We consider this concept could be applied to image processing and we want to relate it with the concept of local contrast of a window of an image, that is, the local variations in brightness. In [11] we proposed an expression of the local contrast constructed from Atanassovs intuitionistic index by means of Atanassov's intuitionistic fuzzy S-implications. Local contrast, from our point of view, must satisfy some specific properties and we would like to study the performance of the new concept presented in this work with regard to our definition of local contrast of an image.

Acknowledgment

This paper has been partially supported by the National Science Foundation of Spain, Reference TIN2007-65981.

References

[1] K. T. Atanassov. Intuitionistic fuzzy sets. *In: V. Sgurev, ed., VII ITKR's Session, Sofia*, June 1983 (deposed in Central Science and Technical Library, Bulgarian Academy of Sciences, 1697/84, in Bulgarian).

[2] K. T. Atanassov. Intuitionistic Fuzzy Sets. *Fuzzy Sets and Systems*,20:87–96, 1986.

[3] K. Atanassov, G. Gargov. Elements of Intuitionistic Fuzzy Logic. Part I. *Fuzzy Sets and Systems*,95:39-52, 1998.

[4] P. Burillo, H. Bustince. Orderings in the referential set induced by an intuitionistic fuzzy relation. *Notes on Intuitionistic Fuzzy Sets*, 1:93-103, 1995.

ISBN: 978-989-95079-6-8

[5] H. Bustince, P. Burillo, V. Mohedano. About Intuitionistic Fuzzy Implication Operators. *JCIS 2002, Research Triangle Park*, 109–113, North Carolina, USA, 2002.

[6] H. Bustince, P. Burillo, F. Soria. Automorphisms, Negations and Implication Operators. *Fuzzy Sets and Systems*, 134:209-229, 2003.

[7] H. Bustince, E. Barrenechea, V. Mohedano. Intuitionistic Fuzzy Implication Operators. An expression and main properties. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(3):387–406, 2004.

[8] H. Bustince, V. Mohedano, E. Barrenechea, M. Pagola . A study of the Intuitionistic Fuzzy S-Implication. First Properties. *Procesos de Toma de Decisiones, modelado y Agregación de Preferencias*, 141–150, Eds.:E. Herrera-Viedma, 2005.

[9] H. Bustince, M. Pagola, P. Melo-Pinto, E. Barrenechea, P. Couto, Image Threshold Computation by modelizing Knowledge/Unknowledge by means of Atanassov's Intuitionistic Fuzzy Sets. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models. Intelligent Systems from Decision Making to Data Mining, Web Intelligence and Computer Vision*, 621–638, Eds.: H. Bustince, F. Herrera, J. montero, (Springer-Verlag), 2008.

[10] H. Bustince, E. Barrenechea, M. Pagola. Generation of interval-valued fuzzy and Atanassov's intuitionistic fuzzy connectives from fuzzy connectives and from K_α operators: Laws for conjunctions and disjunctions, amplitude. *International Journal of Intelligent Systems*, 23(6):680-714, 2008.

[11] H. Bustince, E. Barrenechea, M. Pagola, J. Fernandez, J. Olagoitia, P. Melo-Pinto, P. Couto. Contrast computing using Atanassov's intuitionistic fuzzy sets. *In Proc. 8th International FLINS Conference on Computational Intelligence in Decision and Control* , Madrid: 259–264, 2008.

[12] C. Cornelis, G. Deschrijver. The Compositional Rule of Inference in an Intuitionistic Fuzzy Logic Setting. *In Proc. 6th ESSLLI Students Session*, 2001.

[13] C. Cornelis, G. Deschrijver, E. Kerre. Classification of Intuitionistic Fuzzy Implicators: an Algebraic Approach. *In Proc. 6th Joint Conf. Information Sciences*, Research Triangle Park, USA: 105–108, 2002.

[14] G. Deschrijver, C. Cornelis, E. Kerre. On the representation of intuitionistic fuzzy t-norms and t-conorms. *IEEE Transactions on Fuzzy Systems*, 12(1):45-61, 2004.

[15] J. Fodor, M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support. Theory and Decision.* Kluwer Academic Publishers, 1994.

[16] A. Pankowska, M. Wygralak. On hesitation degrees in IF-set theory. *In: LNAI 3070,* Springer, 338–343, 2004.

[17] A. Pankowska, M. Wygralak. General If-sets with triangular norms and their applications to group decision making. *Information Sciences*, 176:2713–2754, 2006.

[18] P. Smets, P. Magrez. Implication in fuzzy logic. *International Journal of Approximate Reasoning*, 1:327-347, 1987.

[19] E. Trillas. Sobre funciones de negación en la teoría de conjuntos difusos. *Stochastica*, III(1):47-59, 1979 (in Spanish). Reprinted (English version) in: Advances of Fuzzy Logic. S. Barro et altri-Universidad de Santiago de Compostela, 31-43,1998.

[20] E. Trillas, C. Alsina, J.M. Terricabras. *Introducción a la lógica borrosa*, Ariel Matemática, 1995.

[21] M. Wygralak. Representing incomplete knowledge about fuzzy sets. *In Proc. 5th Eusflat Conf.*, Ostrava, Czech Rep.: Vol.II, 287–292, 2007.

ISBN: 978-989-95079-6-8

Time Series Analysis and Prediction Based on Fuzzy Rules and the Fuzzy Transform

Martin Štěpnička Viktor Pavliska Vilém Novák

Irina Perfilieva Lenka Vavříčková Iva Tomanová

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech Republic
Email: {martin.stepnicka,viktor.pavliska,vilem.novak,irina.perfilieva, lenka.vavrickova, iva.tomanova}@osu.cz,

Abstract— *A new methodology for analysis and forecasting of time series is proposed. It directly employs two techniques: the fuzzy transform and the perception-based logical deduction. Due to the usage of both of them and due to the innovative approach consisting in a construction of several independent models, the methodology is successfully applicable for robust long time predictions.*

Keywords— Time series, fuzzy transform, perception-based logical deduction, fuzzy partition.

1 Introduction

We propose a new methodology for forecasting time series which is based on combination of two techniques: fuzzy transform and perception-based logical deduction. The proposed methodology consists of two phases: analysis of a time series and its forecast. In both phases, the above mentioned techniques play an essential role.

The organization of the work is as follows. After preliminaries and recalling the techniques in Section 2, we continue with a detailed description of the proposed approach in Section 3. Then we demonstrate the results in Section 4.

2 Preliminaries

Let us briefly recall the main tools employed in the suggested approach, the *fuzzy transform* (F-transform) [1] and the *perception-based logical deduction* [2, 3], in particular.

2.1 Fuzzy (F)-Transform

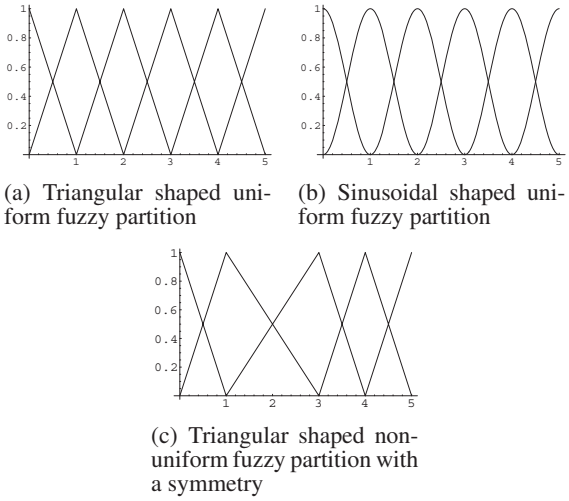
Fuzzy transform, in detailed described in [1], consists in two transformations. The first one, sometimes called the direct one, maps a continuous function defined on a fixed real interval $[a, b]$ into an n-dimensional vector. The inverse one maps the vector back to the space of continuous functions.

The F-transform is defined with respect to *basic functions* forming a fuzzy partition.

Definition 2.1 Let $c_1 < \cdots < c_n$ be fixed nodes within $[a, b]$, such that $c_1 = a, c_n = b$ and $n \geq 2$. We say that fuzzy sets $A_1, \ldots, A_n \in \mathcal{F}([a, b])$ are *basic functions* forming a fuzzy partition of $[a, b]$ if they fulfill the following conditions for $i = 1, \ldots, n$:

1. $A_i(c_i) = 1$;
2. $A_i(x) = 0$ for $x \notin (c_{i-1}, c_{i+1})$ where for the uniformity of the denotation we put $x_0 = x_1 = a$ and $x_{n+1} = x_n = b$;
3. A_i is continuous
4. A_i strictly increases on $[c_{i-1}, c_i]$ and strictly decreases on $[c_i, c_{i+1}]$;
5. for all $x \in [a, b]$

$$\sum_{i=1}^{n} A_i(x) = 1. \tag{1}$$



(a) Triangular shaped uniform fuzzy partition

(b) Sinusoidal shaped uniform fuzzy partition

(c) Triangular shaped non-uniform fuzzy partition with a symmetry

Figure 1: Graphic presentation of distinct fuzzy partitions.

Usually, the *uniform fuzzy partition* is used i.e. n equidistant nodes $c_i = c_{i-1} + h$ are fixed. Let us remark that the shape of the basic functions is not predetermined and it can be chosen based on further requirements. For some examples of fuzzy partitions fulfilling the Definition 2.1, see Fig. 1.

Definition 2.2 Let a fuzzy partition of $[a, b]$ be given by basic functions $A_1, \ldots, A_n, n \geq 2$ and let $f : [a, b] \to \mathbb{R}$ be an arbitrary continuous function. The n-tuple of real numbers $[F_1, \ldots, F_n]$ given by

$$F_i = \frac{\int_a^b f(x)A_i(x)dx}{\int_a^b A_i(x)dx}, \ i = 1, \ldots, n \tag{2}$$

is the *direct fuzzy transform (F-transform)* of f with respect to the given fuzzy partition. $F_1, \ldots, F_n$ are the *components* of the F-transform of f.

ISBN: 978-989-95079-6-8

In practice, function f is usually not given analytically but we are at least provided by some data obtained, e.g., by some measurements. In this case, Definition 2.2 can be modified in such a way that finite summations replace definite integrals in formula (2).

Let us be given n basic functions forming a fuzzy partition of $[a, b]$ and let the function f be given at $T > n$ fixed nodes $x_1, \ldots, x_T \in [a, b]$. We say that the set of nodes $\{x_1, \ldots, x_T\}$ is *sufficiently dense with respect to the fuzzy partition* if

$$(\forall i)(\exists t) \quad A_i(x_t) > 0. \tag{3}$$

Definition 2.3 Let a fuzzy partition of $[a, b]$ be given by basic functions $A_1, \ldots, A_n, n \geq 2$ and let $f : [a, b] \to \mathbb{R}$ be a function known at a set $\{x_1, \ldots, x_T\}$ of sufficiently dense nodes with respect to the given fuzzy partition. The n-tuple of real numbers $[F_1, \ldots, F_n]$ given by

$$F_i = \frac{\sum_{t=1}^{T} f(x_t)A_i(x_t)}{\sum_{t=1}^{T} A_i(x_t)}, \; i = 1, \ldots, n \tag{4}$$

is the *(discrete) direct F-transform* of f with respect to the given fuzzy partition. $F_1, \ldots, F_n$ are the *components* of the F-transform of f.

Since this paper deals with the application of the F-transform to the analysis and prediction of time series, i.e., discrete problem, there is no danger of confusion and we may easily, for the sake of simplicity, omit the word "discrete" and simply talk about the "F-transform" in the rest of the paper.

The F-transform of f with respect to $A_1, \ldots, A_n$ will be denoted by $F_n[f] = [F1, \ldots, Fn]$. It has been proved [1] that the components of the F-transform are the weighted mean values of an original function where the weights are given by the basic functions.

The original function f can be approximately reconstructed (with the help of the inversion formula) from its fuzzy transform $F_n[f]$. The function represented by the inversion formula is called the *inverse F-transform*.

Definition 2.4 Let $F_n[f]$ be the direct F-transform of f with respect to $A_1, \ldots, A_n \in \mathcal{F}([a, b])$. Then the function $f_{F,n}$ given on $[a, b]$ as follows

$$f_{F,n}(x) = \sum_{i=1}^{n} F_i A_i(x), \tag{5}$$

is called the *inverse F-transform* of f.

The inverse F-transform is a continuous function on $[a, b]$ no matter whether we have started from analytically given continuous f or we have been given only function values on a sufficiently dense set. Anyhow, for the time series problem, we consider the inverse F-transform at the same points where the original function is given.

2.2 Perception-based Logical Deduction

In this subsection, we briefly recall the main idea of the *perception-based logical deduction* (PbLD). For details, let us refer to [2, 3].

The PbLD is a method of deducing conclusions on the basis of a *linguistic description* which is a set of m fuzzy/linguistic rules:

IF x_1 is $\mathcal{A}_{11}$ **AND** $\cdots$ **AND** x_q is $\mathcal{A}_{1q}$ **THEN** z is $\mathcal{B}_1$
$\ldots$
AND
$\ldots$
IF x_1 is $\mathcal{A}_{m1}$ **AND** $\cdots$ **AND** x_q is $\mathcal{A}_{mq}$ **THEN** z is $\mathcal{B}_m$ (6)

where $\mathcal{A}_{11}, \ldots, \mathcal{A}_{nq}, \mathcal{B}_1, \mathcal{B}_m$ are specific linguistic expressions called *evaluative expressions* which have the form

$$\langle \text{ linguistic hedge } \rangle \; \{ \textit{ small, medium, big } \}.$$

Note that extension of $\mathcal{A}$ is a functional value of its intension in a point w. These extensions are specific fuzzy sets having the form depicted in Fig.2. Let us shortly mention, that formal theory of evaluative linguistic expression is a deep and firm theory and it would be out of range of the conference paper to recall it in a more detailed way to readers. Therefore, we refer to its detailed description and justification published in [4].

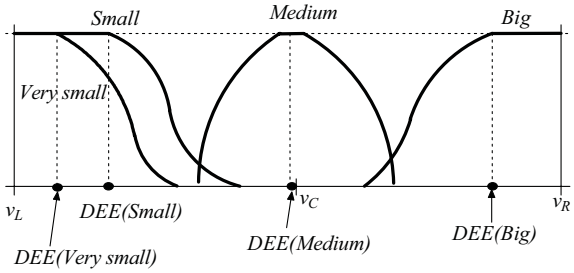


Figure 2: Fuzzy sets representing typical evaluative linguistic expressions.

In the frame of the above outlined theory, that the rules in (6) are taken as genuine conditional expressions of natural language that are interpreted accordingly, see [5], i.e. intension and extensions of them are also defined. The extension of each rule is a special fuzzy relation computed using a t-norm (usually minimum) and Łukasiewicz (residual) implication.

By *perception* we understand an evaluative expression assigned to the given value in the given context. The choice of perception is not arbitrary and besides the context, it also depends on the topic of the specified linguistic expression. The PbLD method uses a sophisticated algorithm for choosing the best fitting rule(s). The most proper crisp output is obtained using a DEE (Defuzzification of Evaluative Expressions) method [2].

2.3 Time Series

Let

$$\{x_t | t = 1, \ldots, T\} \subset \mathbb{R}, \quad T \geq 3 \tag{7}$$

be a given time series. The task is to analyze and forecast the time series, i.e., to determine values

$$\{x_t | t = T + 1, \ldots, T + k\} \subset \mathbb{R}, \quad k \geq 1. \tag{8}$$

There are two standard approaches to forecast a time series. The first one uses autoregressive and moving averages models of the so called *Box-Jenkins methodology* [6]. This approach,

ISBN: 978-989-95079-6-8

no matter how useful and successful in forecasts it may be, besides stationarity and seasonality provides us with a non-transparent and non-interpretable analysis. Therefore, it is not appropriate for an extension by means and tools directly using the fuzzy sets theory, where the interpretability and transparency play a crucial role.

Remark 2.5 A study presenting Takagi-Sugeno rules [7] in the view of the autocorrelation Box-Jenkins methodology has been published [8]. Nevertheless, the Takagi-Sugeno rules are rather regression-based than linguistic-based in comparison with (6). Analogously, distinct neuro-fuzzy approaches, which are on the border of neural networks, Takagi-Sugeno models and evolving fuzzy systems, are very often and successfully used [9, 10]. But well tuned Gaussian fuzzy sets with a centroid at node 5.6989 and a width parameter equal to 2.8893 (see [10]) constructed as a product of an employed optimization technique are undoubtably far from the interpretability of systems using models of fragments of natural language. Therefore, these approaches, no matter how effective and powerful, are closer to standard regression methods.

The second approach consists in the so called *decomposition* [11] of a time series where each element x_t of the time series is decomposed into the following components

$$x_t = Tr_t + S_t + C_t + E_t \tag{9}$$

where Tr_t, S_t, C_t, E_t are the t-th *trend, seasonal, cyclic* and *error* components, respectively.

Trend and seasonal components may be analyzed. The cyclic component is a bit problematic. The name "cyclic" comes from the economical cycles which are not regular and they are dependent on many outer factors to be analyzed from the past. The error component is a random noise which basically cannot be forecasted and therefore is omitted from our further considerations.

After these omissions, for further investigation, usually the following simplified decomposed model is considered

$$x_t = Tr_t + S_t. \tag{10}$$

Remark 2.6 Formulas (9) and (10) describe so called additive decomposition. Using multiplications instead of summations, we would get the multiplicative decomposition. Since the treatment is analogous, for the sake of simplicity we restrict our focus on the additive ones but the further methods and results are similar.

3 Proposed approach

Let us introduce our suggested approach directly employing the techniques briefly recalled in Section 2.

3.1 Trend and Seasonal Analysis

Traditional approach to the trend analysis assumes the trend to be an a priori given function, e.g., linear, polynomial, exponential or a kind of a saturation function such as sigmoidal function, for instance. This approach simplifies the analysis, which consists in a regressive determination of parameters of the predetermined function, and even the prediction, which consists in a simple prolongation – i.e. in an evaluation of the determined trend function at nodes $T+1, \ldots, T+k$.

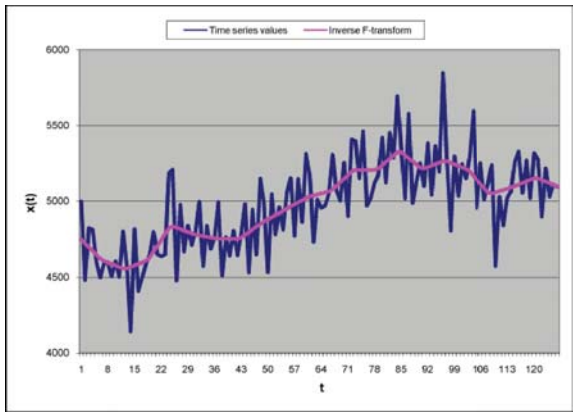


Figure 3: Time series with its inverse F-transform. Standard trend analysis would be inappropriate due to irregular cyclic changes.

But such an approach is not always the most proper one, see Fig. 3. The problem is that such an approach where we fix a trend function for the whole domain is too restrictive. The trend may change during the measurements of a time series especially in case of a long time series. For such cases, complicated adaptive trend changing models are constructed. As a typical example, we may especially in these days of the financial crisis recall the equity indexes where we may hardly simply prolong the trends while very often after robust growth radical and dramatic falls come followed by stagnation and then again by growths, due to the cyclic influences. Here, the prolongation might be very often the worst thing we may apply in predictions.

We propose the F-transform method for the trend description. The F-transform does not fix any shape of the curve, but it possess powerful approximation, noise reduction and computational properties [12].

The time series x_t may be viewed as a function x on the interval $[0, T]$ which we are not given analytically, but we are provided by measurements $x(t) = x_t$ at nodes $t = 1, \ldots, T$. Let us build a uniform fuzzy partition according to Definition 2.1 such that each basic function $A_2, \ldots, A_{n-1}$ "cover" the number of nodes equal to the number of nodes belonging to a season. For example, in case of a time series on the monthly basis, each basic function covers 12 point except basic functions A_1, A_n which cover the first and the last 6 nodes, respectively. So, the set of nodes is sufficiently dense with respect to the fuzzy partition. From now on, we will explicitly refer to such a time series on the monthly basis since everything may be easily generalized for a different seasonality.

Remark 3.1 Usually, we are provided by some additional information about a time series, for example that it is a GDP growth per month. In cases, when such information is not at disposal or does not lead to a sufficient knowledge to determine the seasonality (diesel engine pressure etc. [13]), the perception-based logical deduction with appropriate rules may be successfully used. For details, we refer the reader to [14].

Let $F_n[x] = [X_1, \ldots, X_n]$ be the F-transform of the function x w.r.t. the given fuzzy partition and let $x_{F,n}$ be its inverse F-transform. The inverse F-transform will serve us as a model

ISBN: 978-989-95079-6-8

of the trend component (including cyclic changes). It does not fix a priori some shape of the function. Therefore, we may easily from (9) with omitted error component determine the seasonal components S_t

$$S_t = x_t - x_{F,n}(t) \tag{11}$$

where $x_{F,n}(t) = Tr_t + C_t$.

3.2 Trend Forecast

The suggested approach to the trend analysis implicitly treats also possible cyclic component influences without complicated adaptive trend changing mechanisms. This is a significant difference to the classical approach, where we first model the trend only and then determine the seasonal components which are influenced by the cyclic irregular changes. Our approach treats the problem the other way around and the trend model $x_{F,n}$ primarily serves us to get pure seasonal components without the cyclic influences.

On the other hand, we cannot easily forecast such a trend model by the prolongation, i.e. by the evaluation of the predetermined fixed trend function at nodes $t = T+1, \ldots, T+k$. Due to the drawback of such a traditional approach, it is not disadvantage but vice-versa, as explained below.

We follow the idea of [15] and for the trend forecast, we employ the perception-based logical deduction. As antecedent variables, we consider the F-transform components of the given time series X_i, $i = 1, \ldots, n$ as well as their differences of the first and the second order

$$\begin{array}{rcll} \Delta X_i & = & X_i - X_{i-1}, & i = 2, \ldots, n \\ \Delta^2 X_i & = & \Delta X_i - \Delta X_{i-1}, & i = 3, \ldots, n \end{array} \tag{12}$$

respectively.

So, we will deal, e.g., with fuzzy rules such as the following one

$$\textbf{IF } \Delta X_{i-1} \text{ is } \mathcal{A}_{\Delta i-1} \textbf{ AND } X_i \text{ is } \mathcal{A}_i \textbf{THEN } \Delta X_{i+1} \text{ is } \mathcal{B}. \tag{13}$$

The differences of the F-transform components expressing the time series trend (and the cycle) of distinct orders are capable to describe the dynamics of the time series better than the F-transform components itself. Fuzzy rules such as (13) may describe logical dependencies of the trend changes (hidden cycle influences) which is highly desirable and suggested in comparison with the standard prolongation of the trend observed in the past. The advantage of the transparently interpretable form of the rules using fragments of natural language is doubtless. This might be helpful in better understanding of functionalities and motive factors determining the changes in a process yielding the analyzed times series.

Let us mention that fuzzy rules such as (13) are automatically generated by the *linguistic learning* algorithm [16] implemented in the software package LFLC 2000 [17] from the F-transform components of the time series and their differences.

Remark 3.2 Although the fuzzy rules are also generated from the past, the suggested approach may be able to learn, describe and successfully predict the future of equity indexes mentioned as a motivation example at the beginning of Subsection 3.1. Of course, if a similar progress has been observed and measured in the past. Prolongation of a trend function is not capable of this task.

3.3 Independent Models

Fuzzy rules and the perception-based logical deduction may help us to forecast next F-transform components $X_{n+1}, \ldots, X_{n+\ell}$ of the time series from which we easily determine trend components of the time series as the values of the inverse F-transform $x_{F,n+\ell}(T+1), \ldots, x_{F,n+\ell}(T+k)$, where $\ell < k$.

The problem may appear when the forecast is supposed to be for a too long term. In case if we want to predict the time series x_t on a monthly basis for the next 18 months, we first have to forecast X_{n+1}. For this, we use components $X_1, \ldots, X_n$ and differences $\Delta X_2, \ldots, \Delta X_n$, $\Delta^2 X_3, \ldots, \Delta^2 X_n$. Then, using the same fuzzy rules and antecedent variables $X_1, \ldots, X_{n+1}$, $\Delta X_2, \ldots, \Delta X_{n+1}$ and $\Delta^2 X_3, \ldots, \Delta^2 X_{n+1}$ we may forecast X_{n+2}. And finally, we are able to forecast X_{n+3} based on $X_1, \ldots, X_{n+2}$, $\Delta X_2, \ldots, \Delta X_{n+2}$ and $\Delta^2 X_3, \ldots, \Delta^2 X_{n+2}$.

Clearly, there is a danger of the forecast error propagation since we forecast from forecasted values. The longer-term prediction, the higher is the damage.

To overcome this problem, we suggest to build a finite number of independent trend forecasting models using the technique described in Subsection 3.2. For the sake of simplicity, let us consider ΔX_{i-1} and X_i as the antecedent variables. Then set of rules (13) will serve us as a model for the prediction the next F-transform component X_{n+1}.

Similarly, a set of rules

$$\textbf{IF } \Delta X_{i-1} \text{ is } \mathcal{A}_{\Delta i-1} \textbf{ AND } X_i \text{ is } \mathcal{A}_i \textbf{ THEN } \Delta X_{i+2} \text{ is } \mathcal{B} \tag{14}$$

generated again by the linguistic learning algorithm from the differences of the F-transform components, will serve us as a model for the prediction the next but one F-transform component X_{n+2}. And analogously, we may build the third model

$$\textbf{IF } \Delta X_{i-1} \text{ is } \mathcal{A}_{\Delta i-1} \textbf{ AND } X_i \text{ is } \mathcal{A}_i \textbf{ THEN } \Delta X_{i+3} \text{ is } \mathcal{B} \tag{15}$$

and further models, if necessary because of a longer-term forecast than just to next 18 values $x_{T+1}, \ldots, x_{T+18}$.

The suggested approach fully avoids the problem of the error accumulation due to the "forecasting from forecasted".

3.4 Seasonal Component Forecast

Forecasted F-transform components $X_{n+1}, \ldots, X_{n+\ell}$ determine the trend and cyclic components of the time series predicted for the future which are given as the values of the inverse F-transform $x_{F,n+\ell}(T+1), \ldots, x_{F,n+\ell}(T+k)$. To forecast the seasonal components and to compose them with the latter values is the last step in the time series prediction procedure.

The seasonal components are forecasted as follows. Let

$$\mathbf{S}_\xi = [S_{p\cdot(\xi-1)}, S_{p\cdot(\xi-1)+1}, S_{p\cdot(\xi-1)+2}, \ldots, S_{p\cdot(\xi-1)+p-1}] \tag{16}$$

be the ξ-th vector of the seasonal components (11) where p denotes the period of the seasonality, i.e., in our case of a time series on a monthly basis $p = 12$ and the $\mathbf{S}_\xi$ is the vector of the January, February etc. to December measurements of the ξ-th year.

Stationarity assumption on the seasonal component of the time series is considered, i.e., we assume that $\mathbf{S}_\xi$ is a linear

ISBN: 978-989-95079-6-8

combination of previous θ vectors $\mathbf{S}_{\xi-\theta}, \ldots, \mathbf{S}_{\xi-1}$. It means that we generate the following system of equations

$$\mathbf{S}_\xi = \sum_{j=1}^{\theta} d_j \cdot \mathbf{S}_{\xi-j}, \quad \xi > \theta \tag{17}$$

to which we search for the optimal solution with respect to coefficients $d_1, \ldots, d_\theta$. The computed coefficients are then used to determine the $\mathbf{S}_\xi$

Let us mention that the stationarity is a standard assumption which might be easily checked, see [6, 18].

Composition of both forecasts, of the forecasted F-transform components and the seasonal components is the last step to get the overall time series forecast. It is done inversely to the original decomposition which was either additive or multiplicative.

3.5 Optimization

There are some unknown parameters in the whole procedure which are to be determined individually for every single time series. Basically, it is the list of antecedent variables for the prediction of the F-transform components, the number of the antecedent variables and the parameter θ.

The time series is divided into the *learning set* and the *validation set* in such a way that the latter one is given by the last values of the times series of the length equal to the expected forecast. It means, the learning set is given by $\{x_1, \ldots, x_{T-k-1}\}$ and we cut $\{x_{T-k}, \ldots, x_T\}$ off the time series to determine the validation.

There is a software [19], developed at the *Institute for Research and Applications of Fuzzy Modeling*, where the parameters θ and the maximal number of antecedent variables are set up by a user. Then the software computes all possible combinations of antecedent variables up to the maximal number and combines it with seasonal components determined based on the optimal solution of (17) for all possible θ, again up to the pre-specified maximal one. For this computations, only the learning set is used.

All computed models are used to forecast $\{x'_{T-k}, \ldots, x'_T\}$ and these forecasts are compared with the validation set $\{x_{T-k}, \ldots, x_T\}$. As a suggestion to a user, all models are ordered according to a pre-specified error criterion which may be generally arbitrarily. A user may then employ any of the optimized and tuned models for the forecast of $x_{T+1}, \ldots, x_{T+k}$.

4 Demonstration

We applied our approach to analyze and forecast 111 time series on a monthly basis from the *Artificial Neural Network & Computational Intelligence Forecasting Competition NN3*. All the time series were "real-world" and chosen from different fields such as economy, meteorology, industry etc. while no such an information about the meaning or an origin of any time series was provided to participants.

There were 580 participants who registered to the competition and downloaded the time series but only 25 finally afforded to submit their results which illustrates the complicacy of the task to deal with such a huge and heterogenous set of time series. This fact is even supported by the average 15.44% error result (measured by SMAPE - Symmetric Mean Absolute Percent Error) obtained by commercial statistical software *ForecastPro* for the whole set of time series. Let us note that the *ForecastPro* software was used only as a benchmark together with five other benchmarks. Let us also remark that it employs all the existing standard techniques from the Box-Jenkins methodology while participants were limited by a single method of a non-statistical nature.

Our approach (without the construction of independent models at that times) took the 12th position with the overall average error 18.81%. In an unofficially evaluated subset of time series with a seasonal character, our approach took the 3rd rank among all participants and even beaten four of the six statistical benchmarks. The average error in this category measured by SMAPE was 12.7%.

The construction of independent models may even improve the results. For an illustration, let us turn the attention to the time series NN3_106 from the competition, see Fig. 4.

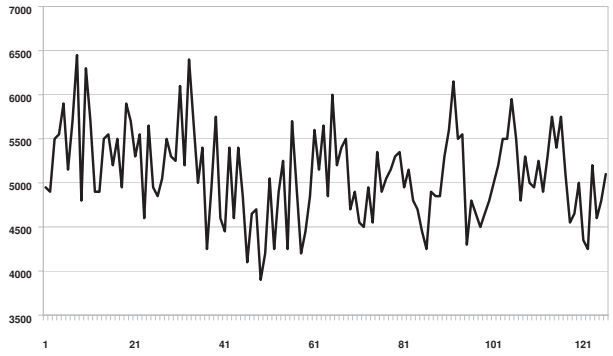


Figure 4: Time series NN3_106 from the NN3 competition.

It is a 126 months long time series with a hardly determinable trend - a typical representative of an appropriate time series for our approach. We cut off last 18 months to create a validation set and we use the first 114 months as a learning set to determine our model. The model may be used to forecast the last 18 months and then compared with the original time series values forming the validation set. There is the black line displaying the validation set of the original time series in Fig. 5.

The dotted grey line in the same figure denotes the forecast made by our methodology with a single trend prediction model optimized as described in Subsection 3.5. The generated model consisted of 15 fuzzy rules with the following 3 input variables $X_i, \Delta^2 X_i, \Delta^2 X_{i-1}$ and the output variable ΔX_{i+1}. The SMAPE error was 5.70%.

If we generate 3 independent models – all of them with the following 3 input variables $X_i, \Delta^2 X_i, \Delta^2 X_{i-1}$, every one of them predicting a different F-transform component – $\Delta X_{i+1}, \Delta X_{i+2}, \Delta X_{i+3}$, respectively, we improve the results. Due to the F-transform, the above mentioned 3 models predicting independently 3 first order differences of the F-transform components are fully sufficient to forecast values

ISBN: 978-989-95079-6-8

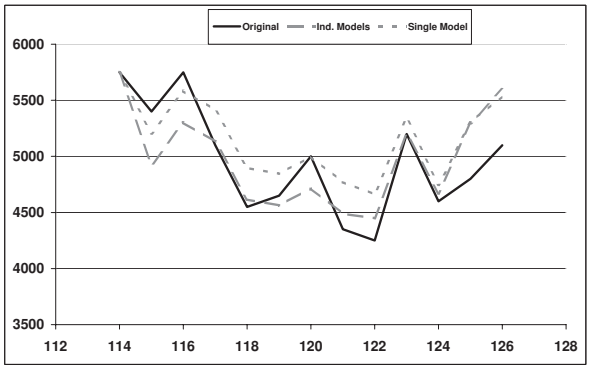


Figure 5: Time series NN3_106 (black line) and its two forecasts, one based on our methodology with a single model (dotted grey line), one forecast employs our methodology with 3 independent models (dashed grey line).

of 18 months ahead. The improve results are depicted by the grey dashed line on Fig. 5 and yielded error 4.77%.

5 Conclusions

We have proposed a novel fuzzy approach to analysis and prediction of time series. It directly uses two techniques, the F-transform and the perception-based logical deduction. Their combination helps to avoid the classical problem with prolongation of a pre-specified curve. Moreover, due to the interpretability of fuzzy rules used by the perception-based logical deduction, a user may get into a deeper understanding not only of the automatically generated model but also to the whole process whose measured samples yielded the analyzed time series.

Furthermore, due to the F-transform and the possibility to construct several independent trend forecasting models using fuzzy rules, we may easily produce long-time forecasts without the influence of the propagated and accumulated error due to the forecasting from forecasted values.

Acknowledgment

This work has been supported by project DAR 1M0572 of the MŠMT ČR.

References

[1] I. Perfilieva. Fuzzy transforms: theory and applications. *Fuzzy Sets and Systems*, 157:993–1023, 2006.

[2] V. Novák and I. Perfilieva. On the semantics of perception-based fuzzy logic deduction. *International Journal of Intelligent Systems*, 19:1007–1031, 2004.

[3] V. Novák. Perception-based logical deduction. In B. Reusch, editor, *Computational Intelligence, Theory and Applications (Advances in Soft Computing)*, pages 237–250. Springer, Berlin, 2005.

[4] V. Novák. A comprehensive theory of trichotomous evaluative linguistic expressions. *Fuzzy Sets and Systems*, 159(22):2939–2969, 2008.

[5] V. Novák and S. Lehmke. Logical structure of fuzzy IF-THEN rules. *Fuzzy Sets and Systems*, 157:2003–2029, 2006.

[6] G. Box and G. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.

[7] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15:116–132, 1985.

[8] J.L.M. Aznarte, J.M. Bentez, and J.L. Castro. Fuzzy identification of systems and its applications to modeling and control. *Fuzzy Sets and Systems*, 158:2734–2745, 2007.

[9] Hai-Jun Rong, N. Sundararajan, Guang-Bin Huang, and P. Saratchandran. Sequential adaptive fuzzy inference system (safis) for nonlinear system identification and prediction. *Fuzzy Sets and Systems*, 157:1260–127, 2006.

[10] G. Leng, T.M. McGinnity, and G. Prasad. An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network. *Fuzzy Sets and Systems*, 150:211–243, 2005.

[11] A. Bovas and J. Ledolter. *Statistical Methods for Forecasting*. Wiley, New York, 2003.

[12] I. Perfilieva and R. Valášek. Fuzzy transforms in removing noise. In B. Reusch, editor, *Computational Intelligence, Theory and Applications (Advances in Soft Computing)*, pages 221–230. Springer, Berlin, 2005.

[13] M. Štěpnička, E. Lughofer, and V. Pavliska. Comparison of data-driven fuzzy modelling methods tested on nox data. In *Abstracts of the 2nd FLLL/SCCH Master and PhD Seminar*, pages 33–39, 2006.

[14] V. Novák, M. Štěpnička, I. Perfilieva, and V. Pavliska. Analysis of periodical time series using soft computing methods. In Da Ruan, J. Montero, J. Lu, L. Martí nez, P. D'hondt, and E.E. Kerre, editors, *Computational Intelligence in Decision and Control*, pages 55–60. World Scientific, New Jersey, 2008.

[15] I. Perfilieva, V. Novák, V. Pavliska, A. Dvořák, and M. Štěpnička. Analysis and prediction of time series using fuzzy transform. In *Proc. IEEE World Congress on Computational Intelligence*, pages 3875–3879, 2008.

[16] R. Bělohlávek and V. Novák. Learning rule base of the linguistic expert systems. *Soft Computing*, 7:79–88, 2002.

[17] A. Dvořák, H. Habiballa, V. Novák, and V. Pavliska. The software package LFLC 2000 - its specificity, recent and perspective applications. *Computers in Industry*, 51:269–280, 2003.

[18] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods (Springer Series in Statistics), 2nd ed.* Springer-Verlag, Heidelberg, 1998.

[19] H. Habiballa, V. Pavliska, and A. Dvořák. Software system for time series prediction based on F-transform and linguistic rules. *APLIMAT - Journal of Applied Mathematics*, accepted, 2009.

ISBN: 978-989-95079-6-8

On the use of aggregation operators for location privacy

Aida Valls[1], Jordi Nin[2] and Vicenç Torra[2]

1. ITAKA Research Group - Intelligent Tech. for Advanced Knowledge Acquisition
Department of Computer Science and Mathematics
Universitat Rovira i Virgili
43007 Tarragona, Catalonia, Spain

2. IIIA, Artificial Intelligence Research Institute
CSIC, Spanish National Research Council
Campus UAB s/n
08193 Bellaterra, Catalonia, Spain

Email: aida.valls@urv.cat, {jnin,vtorra}@iiia.csic.es

***Abstract*—**
Nowadays, the management of sequential and temporal data is an increasing need in many data mining processes. Therefore, the development of new privacy preserving data mining techniques for sequential data is a crucial need to ensure that sequence data analysis is performed without disclosure sensitive information. Although data analysis and protection are very different processes, they share a few common components such as similarity measurement.

In this paper we propose a new similarity function for categorical sequences of events based on OWA operators and fuzzy quantifiers. The main advantage of this new similarity function is the possibility of incorporating the user preferences in the similarity computation. We describe the implications of the application of different user preference policies in the similarity measurement when microaggregation, a well-known data protection method, is applied to sequential data.

***Keywords*—** Microaggregation, Privacy, Sequence aggregation, OWA operators.

1 Introduction

The development of new methods for managing sequential and temporal relationships is becoming an strategic area. Nowadays, thanks to the new technologies, categorical sequences of data are easily available in many scenarios. Initially, sequential data was originated by the study of gene information (DNA), but novel classes of applications based on geo-spatial information are currently appearing [1]. For instance, many applications use personal or vehicular tracking data, this information allows us to find interesting patterns to be used in many different applications, such as traffic control, sustainable mobility management, accessibility of services or even tracking patients with Alzheimer [2]. This type of data is known as Event Sequences [3].

Classical data mining algorithms and decision making techniques were developed for static data (*i.e.* data whose feature values do not depend on time). Consequently, the sequential nature of temporal data makes difficult to apply classical data mining and decision making methods. Some areas of Artificial Intelligence have started to develop new methods to deal with event sequences, such as temporal data mining algorithms [4, 5] or sequence data analysis, both interested in studying how feature values change with regards to time, in order to identify interesting temporal patterns.

As in classical data mining, in order to ensure that sequence data analysis is performed without disclosure sensitive information of data owners, statistical disclosure control (SDC) [6] and privacy preserving data mining (PPDM) [7] techniques should be applied. In this way, the privacy of the individuals can be guaranteed. A very common way to achieve a certain level of privacy on a database is to ensure k-anonymity [8, 9]. Microaggregation [10, 11] is the standard SDC technique to achieve the k-anonymity in a database. Essentially, microaggregation builds clusters of at least k records and replaces the original records with the centroid of the cluster that the record belongs to. In this way, k-anonymity is ensured because in the protected database there are at least k identical records.

In this paper we address the problem of measuring the similarity between sequences of events. Similarity measurement is a key point in data analysis but also in microaggregation. This problem was already studied in [12], where a new approach to calculate the similarity for event sequences was presented: the *Ordered-based Sequence Similarity* (OSS). This similarity is greately improved in this paper. The new version presented in this paper, applies the OWA operator to let the user to define different preference policies for the similarity comparison of two event sequences, thus, it is called OSS_{OWA}. Preference policies can be expressed using different fuzzy measures. We will present the implications that some policies (fuzzy measure) have on the results of a microaggregation procedure with an illustrative example.

This paper is organized as follows. Firstly, in Section 2 we give some preliminaries about aggregation functions and microaggregation methods for SDC. Then, in Section 3 we present a detailed description of our new similarity function for categorical sequences of events. Section 4 provides some experiments combining our similarity function with microaggregation. Finally, Section 5 draws some conclusions and describes some lines for future work.

ISBN: 978-989-95079-6-8

2 Preliminaries

In this section we give a short overview of aggregation operators, focused on the OWA operator, and microaggregation (a data protection method).

2.1 Aggregation functions

Aggregation functions [13] are functions used for information fusion. They typically combine N data values supplied by N data sources into a single datum.

In this paper, we use them to define a new similarity measure among sequences of categorical events. In particular, we consider the Ordered Weighted Aggregation (OWA) operator.

Two different definitions for this operator can be found in the literature. One applicable when the number of data sources is known in advance, and another that does not require to know how many data sources are combined. We use this latter definition, which is based on fuzzy quantifiers.

Definition 1 *A function $Q : [0, 1] \to [0, 1]$ is a* regular monotonically non-decreasing fuzzy quantifier *(non-decreasing fuzzy quantifiers for short) if it satisfies: (i) $Q(0) = 0$; (ii) $Q(1) = 1$; (iii) $x > y$ implies $Q(x) \geq Q(y)$.*

Two examples of families of fuzzy quantifiers Q_1 and Q_2 are given below. Q_1 corresponds to Yager α-quantifiers, for $\alpha > 0$.

$$Q_1^\alpha(x) = x^\alpha \tag{1}$$

$$Q_2^\alpha(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1/(1 + e^{(\alpha - x)*10}) \text{ for } \alpha > 0 & \text{if } 0 < x < 1 \\ 1 & \text{if } x = 1 \end{cases} \tag{2}$$

A graphical representation of these two fuzzy quantifiers is given in Figure 1 for some particular α. $\alpha = \{0.2, 0.4, \ldots, 1.8, 2.0\}$ are used for $Q_1^\alpha(x)$, and $\alpha = \{0, 0.1, \ldots 0.9\}$ are used for $Q_2^\alpha(x)$. We can observe that for small α values, the function increases quickly near $x = 0$, whereas the increase is smoothly for larger values of α.

Using fuzzy quantifiers, the OWA operator [14, 15] is defined as follows.

Definition 2 *Let Q be a non-decreasing fuzzy quantifier, then $OWA_Q : \mathbb{R}^N \to \mathbb{R}$ is an* Ordered Weighted Averaging (OWA) operator *if*

$$OWA_Q(a_1, ..., a_N) = \sum_{i=1}^{N} (Q(i/N) - Q((i-1)/N)) a_{\sigma(i)}$$

where σ is a permutation such that $a_{\sigma(i)} \geq a_{\sigma(i+1)}$.

The interest of the OWA operators is that they permit the user to aggregate the values giving importance to large (or small) values. In the case of the quantifiers given above, the smaller the α is, the larger the importance for the largest values being aggregated. In contrast, the higher the α is, the lower the importance of the largest values (and the higher the importance given to low values). This different behaviour can be seen in Figure 1.

2.2 Microaggregation

Microaggregation [10, 11] is one of the most commonly employed data protection methods [16]. From the operational point of view, given a database with A attributes, microaggregation proceeds by building clusters of at least k records and, then, replacing each record by the centroid of the cluster to which the record belongs to. A certain level of privacy is ensured because for each attribute, there are always k records with an identical value. This assures the k-anonymity property at the attribute level [8, 9]. The goal of a microaggregation is to find clusters of at least k records while minimizing the total Sum of the Square Error:

$$SSE = \sum_{i=1}^{c} \sum_{x_{ij} \in C_i} (x_{ij} - \bar{x}_i)^T (x_{ij} - \bar{x}_i), \tag{3}$$

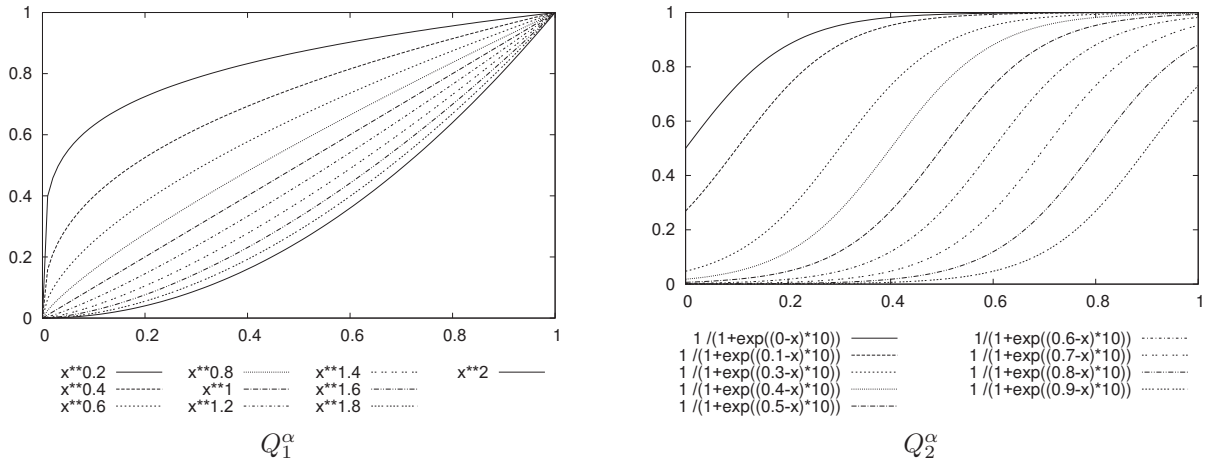
where c is the total number of clusters, C_i is the i-th cluster and $\bar{x}_i$ is the centroid of C_i. The restriction is $|C_i| \geq k$, for all $i = 1, \ldots, c$.

In the simplest case (when $A = 1$), there are polynomial time algorithms to obtain the optimal microaggregation [17]. However, for the general case ($A > 1$), the problem of finding the optimal microaggregation is NP-hard [18]. For this reason, microaggregation methods are heuristic.

It is worth to mention that most of the classical microaggregation techniques are based on the assumption that an average record for all the elements in the database can be computed, like MDAV algorithm [19]. For the case of dealing with sequences of events, this is not a feasible assumption, since sequences can be very diverse and the definition of a single average sequence is not meaningful. Other methods rely on an Euclidean space, which is also not the case of the event sequences space.

In [3] the use of hierarchical clustering algorithms for sequences of events was presented. Those methods are sequential algorithms based on a similarity measure between pairs of elements [20]. The result is a hierarchical tree of non-overlapping clusters that can be cut at any level to obtain a partition. For the experiments of this paper, the k-Sized Hierarchical Clustering method (KSHC) has been applied [21]. This method is a modification of the classical hierarchical sequential agglomerative non-overlapping clustering methods in data mining [20]. The clustering process is structured in two main steps: (1) the construction of a dissimilarity matrix for all pairs of individuals, using an appropriate similarity measure according to the type of data and problem characteristics and (2) the exploitation of the similarity matrix to build a tree of non-overlapping clusters. In this second stage, an iterative process is executed, which consists of finding the pair of elements (i.e. individuals or clusters) that are at minimum dissimilarity (i.e. the most similar) and build a new cluster with these elements. Then, those elements are removed from the dissimilarity matrix, and the dissimilarity of the new cluster with respect to the rest of elements is computed. Different methods for calculating this dissimilarity have been defined: Single Linkage (if the minimum distance with respect to the elements in the cluster is taken), Complete Linkage (if the maximum distance is taken), Average Linkage, Median Linkage, etc. The resulting tree can be cut at any level to produce a partition of the individuals. The KSHC approach keeps the

ISBN: 978-989-95079-6-8



Q_1^α Q_2^α

Figure 1: Graphical representation of Q_1^α and Q_2^α.

tradictional clustering process, but modifies it properly in order to produce a partition that respects the k-anonymity property.

2.3 The Cluster Prototype

Methods for defining cluster prototypes (or centroids) are needed in microaggregation as well as in some unsupervised machine learning techniques (*e.g.* in clustering). In general, a prototype is defined as a typical example, basis, or standard for other elements of the same cluster. Prototypes combine the most representative values for the attributes of the elements that belong to the cluster. Consequently, prototypes are typical instances that represent the content of the cluster. They are calculated using an averaging function (in most of the cases, a weighted mean is used). However, when dealing with categorical sequences of events, those averaging functions are not applicable.

In [22] a new method for generating prototypes of event sequences was presented. This method is based on the calculation of an *Element Scoring Table (EST)*, where the elements can be either individual events or patterns of consecutive events considered as an indivisible unit. So, the method is called *Ordered Element Scoring Prototyping (OESP)*. Different scores are assigned to each event depending on its position in the sequences of the cluster. The ranking given by those scores is used to iteratively append an event to the prototype (initially the prototype is an empty sequence) until the prototype has a length similar to the average lengths of the sequences in the cluster.

This method will be used to calculate the centroids of the clusters obtained by the new similarity measure, which is needed to calculate the Sum of Square Error (3).

3 OSS_{OWA}: A new dissimilarity function for categorical sequences of events

From a formal point of view, a dissimilarity function f over two sequences of events s_1 and s_2 has to fulfill the following conditions:

1. Symmetry: $f(s_1, s_2) = f(s_2, s_1)$
2. Positivity: $f(s_1, s_2) \geq 0$ for any s_1 and s_2
3. Reflexivity: $f(s_1, s_1) = 0$

Some dissimilarity measures for sequences of symbols have been defined. For instance, Hamming distance [23] or Edit (Levenstein) distance [24] are usually considered [3]. However, such distances are quite simple when it is necessary to deal with the relative ordering positions of the symbols in the sequence, as it is needed in tracking studies such as the ones presented in the introduction. Another drawback is the difficulty of incorporating user (expert) preference policies in the computation of these measures.

As we are interested in comparing temporal event sequences, the results of those traditional categorical distances are not good since they disregard a lot of information that should be taken into account to obtain a more appropriate similarity value [12]. Two good examples of temporal event sequences are web logs and the itineraries of the tourists. In these scenarios, there are two key issues to be considered:

- The number of common elements inside the sequences.
- The order (position) among the common elements inside the sequences.

On one hand, the first consideration allows us to evaluate the common part of two different sequences, if this common part is large, then the two sequences have to be similar, because such sequences show that their owners have done the same things. On the other hand, the second consideration takes into account whether the events have been done in the same order or not. This second information is very useful when applying data mining processes. For example, when detecting common itineraries for planning new bus lines. Note that in this latter case, if two sequences have the same events placed in the same order, they have to be more similar than two sequences in which the events are placed in different order. Based on those two assumptions, in [12], the Ordered-based Sequence Similarity was defined (OSS).

However, in the context of event sequence comparison, apart from these two issues, it is also interesting to incorporate any user (expert) preferences on the similarity function. In some scenarios, it is more important to group sequences sharing a large number of events even though the events are not in the same order. On the contrary, in other scenarios, the

ISBN: 978-989-95079-6-8

user is interested in tracking the sub-sequences of common events, considering that two sequences with a few events in the same order are more similar than two sequences with a larger common set of unsorted events. Let us take as an example, an analysis of the tourist behaviour in a theme park. In the first case, the manager wants to find groups of visitors that have been at the same attractions (same locations) in spite of the order of visiting them. In the second case, the manager is interested in studying the flow of visitors in the park, and to know the common paths between attractions for planning or user-recommendation purposes.

The OWA operator with a fuzzy quantifier can be used to introduce the user aggregation policy into the computation of the similarity between pairs of temporal event sequences. This permits us to tackle the issues explained before. Based on this approach, a new similarity function can be defined, $d_{OSS-OWA}$. It is calculated in two steps.

Firstly, partial distances between the elements of the two sequences are calculated as follows.

Definition 3 *Let s and r be two event sequences. Let s_i be the i-th event of the sequence s, then the event distance of s_i with respect to r is defined as*

$$d(s_i, r) = \begin{cases} \frac{min_{s_i=r_j}|i-j|}{|s|+|r|} & if\ s_i \in r \\ 1 & otherwise \end{cases} \quad (4)$$

where $|s|$ and $|r|$ stand for the sequence lengths of sequences s and r respectively.

If s_i appears n times in s and m times in r, the occurrences in s and r are matched forming pairs that minimize $|i-j|$, if some occurrences do not have a couple ($n \neq m$), they are treated as if they do not appear in r, that is, they have a distance to r of 1.

Then, those partial distances are aggregated using the OWA operator, which permits to apply different aggregation policies[14].

Formally, given sequences s and r, we compute the event distances $d(s_i, r)$ for all $i = 1, \ldots, |s|$ and $d(r_i, s)$ for all $i = 1, \ldots, |r|$. Then, the OSS_{OWA} measure is defined as follows.

Definition 4 *Let s and r be two temporal event sequences with length $|s|$ and $|r|$, respectively, then the $d_{OSS-OWA}$ dissimilarity function of s and r is defined as*

$$d_{OSS-OWA(Q)}(s, r) =$$

$$OWA_Q(d(s_1, r), \ldots, d(s_{|s|}, r), d(r_1, s), \ldots, d(r_{|r|}, s))$$

where Q is a non-decreasing fuzzy quantifier, and $d(s_i, r)$ stands for the event distance of s_i with regards to r.

Following Definition 3, all event distances equal to 1 represents the case of two sequences with non-common events. On the contrary, distances smaller than 1 represent the shift between the common events in both sequences. In this case, the larger the event distance, the larger the shift. OWA operators permit the user to aggregate the values giving more importance to large or small values when appropriate quantifiers are selected. This stands for considering more important non-common events (large distance values) or common events (small distance values). In this way we are modeling the two scenarios described above.

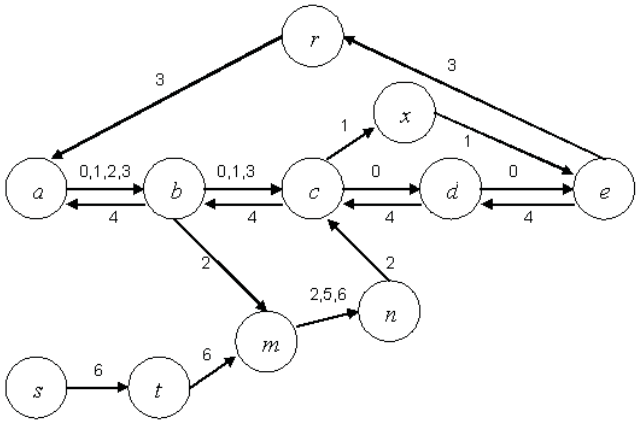


Figure 2: Example with 6 sequences.

Table 1: Original sequences.

id	Sequence
0	abcde
1	abcxe
2	abmnc
3	erabc
4	edcba
5	mn
6	stmn

4 Experiments

In this section, we present the results obtained with a synthetic sequences database (see Table 1). Figure 2 shows a graph representation of those sequences as paths between a set of 11 locations. These could be places inside a theme attractions park. The arrows indicate a transition from one attraction to the next one. The labels refer to the tourists that have followed each path step.

In order to test the $d_{OSS-OWA}$ function, the k-sized hierarchical clustering (KSHC) microaggreagtion algorithm has been used on this database. It has been applied to a k-anonymity value of $k = 2$, which means that cluster of 2 or 3 elements can be generated (recall that the number of elements is constrained by k and $2k-1$). Then, the method explained in section 2.3 for prototype generation has been performed to obtain the centroids of the clusters and calculate the information loss, that is, the SSE (3) .

Six different fuzzy quantifiers have been tested: $Q_1^{0.4}$, Q_1^1, Q_1^2, $Q_2^{0.1}$, $Q_2^{0.5}$ and $Q_2^{0.8}$. They define different aggregation policies in order to show how to introduce expert knowledge in the sequence similarity computation. Quantifiers with small α parameter consider more similar those sequences with a large set of common events. On the contrary, quantifiers with large α parameter consider more similar those sequences with few common events but placed in the same order. Note that, quantifiers with medium α parameter consider common events as important as events placed in the same order.

Table 2 shows the results obtained for the dataset given in Table 1. Each row corresponds to a different parametrization of the fuzzy measure used to compute the OSS_{OWA} dissimilarity function between pairs of sequences. The first three

ISBN: 978-989-95079-6-8

Table 2: Microaggregated sequences with $k = 2$.

Q	SSE	Sequences
$Q_1^{0.4}$	0.515	abcxe, erabc
	0.767	abcde, edcba, abmnc
	0.907	mn, stmn
Q_1^{2}	0.032	erabc, edcba
	0.063	abmnc, abcde, abcxe
	0.081	mn, stmn
Q_1^{4}	0.002	erabc, edcba
	0.0	abcde, abcxe
	0.157	mn, abmnc, stmn
$Q_2^{0.1}$	0.934	abcxe, erabc
	1.137	abcde, edcba, abmnc
	1.614	mn, stmn
$Q_2^{0.5}$	0.087	erabc, edcba
	0.162	abmnc, abcde, abcxe
	0.159	mn, stmn
$Q_2^{0.8}$	0.002	erabc, edcba
	0.0	abcde, abcxe
	0.146	mn, abmnc, stmn

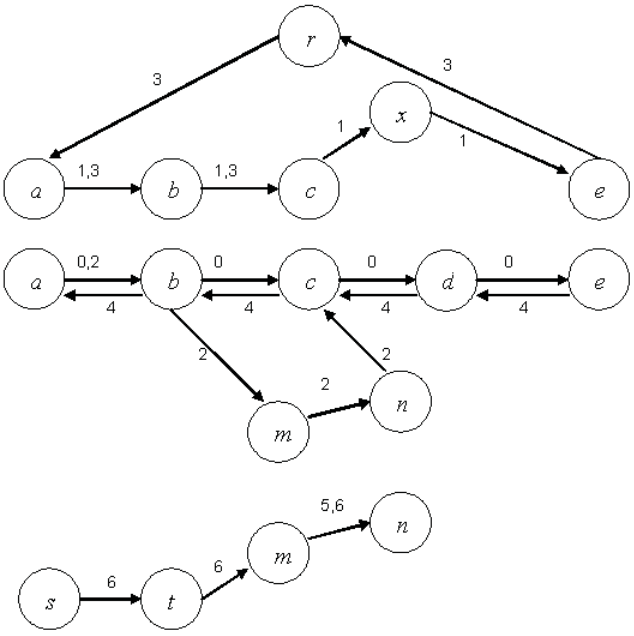


Figure 3: Clusters obtained with small α.

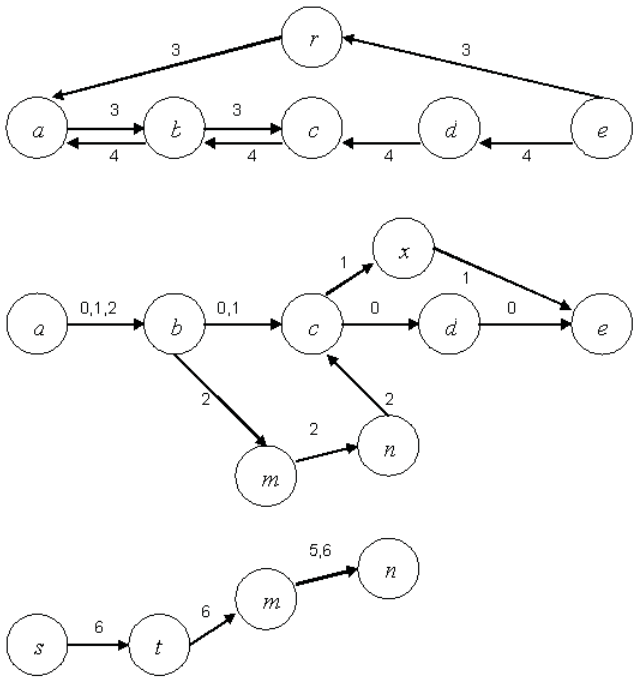


Figure 4: Clusters obtained with a medium α.

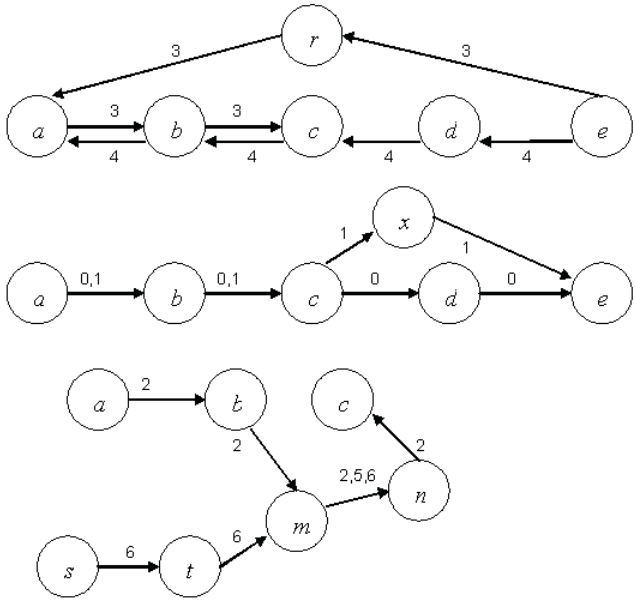


Figure 5: Clusters obtained with large α.

rows correspond to the fuzzy measure Q_1 (1) and the three last rows correspond to the fuzzy measure Q_2 (2).

The first thing to note is that the results for Q_1 and Q_2 are the same for low, medium and high α values. Using low values in the parameterization of the fuzzy measures (*e.g.* α equal to 0.4 for Q_1 and 0.1 for Q_2) generates a similarity function that stress the importance of visiting the same places, disregarding the event order. For this reason, the sequence $abcde$ is similar to the sequence $edcba$. In the same way, the sequence $abcxe$ is more similar to $erabc$ (with 4 equal elements) than to $abmnc$ (with only 3 common elements).

Oppositely, when large α values are selected (*e.g.* α equal to 4 for Q_1 and 0.8 for Q_2), the order among the common events plays a more important role in the similarity measurement. Consequently, the sequence $edcba$ is more similar to the sequence $erabc$ than to the sequence $abcde$, because the path between elements a , b and c is exactly the same. Note also that the sequences $abcxe$ and $abmnc$ also change their cluster with regards to the case of low α values.

To balance these two polices, a medium α value can be selected (*e.g.* α equal to 2 for Q_1 and 0.5 for Q_2). In this case, the groups obtained take into account both issues (common events and event order) in a similar way. Notice that in this case, the final groups are different from the ones in the other two cases.

ISBN: 978-989-95079-6-8

The clusters obtained for the different α values are graphically displayed in Figures 3, 4 and 5, for small, medium and large α, respectively. From them, it can be more easily seen that the itineraries are grouped in a different manner depending on the aggregation policy.

To sum up, it can be said that the $d_{OSS-OWA}$ function permits to model different similarity measurement polices. This fact is crucial for data mining analysis and for privacy preserving data mining. If a microaggregation method does not consider the different semantics that the similarity function can have, the partition obtained may put together individuals that should be distinguished. Moreover, if the information of those individuals is then replaced by their prototype values, the posterior analysis of these masked data will not be able to retrieve some relevant knowledge for the user.

5 Conclusions and future work

In this paper, we have introduced a new similarity function for categorical event sequences using OWA operators. We have illustrated how it is possible to introduce expert knowledge inside the similarity computation using well-known fuzzy quantifiers. We have also presented some results about the application of this similarity function to a very recent microaggregation algorithm for categorical event sequences.

As future work we include the analysis of the microaggregation with respect to information loss and disclosure risk measures when it is applied to event sequences. These measures are required to properly evaluate the performance of the microaggregation when it is applied to event sequences and compare it with other approaches. Moreover, the effect of different fuzzy measures must be studied.

Acknowledgment

Partial support by the Spanish MEC (projects ARES – CONSOLIDER INGENIO 2010 CSD2007-00004 – and eAEGIS – TSI2007-65406-C03-02), the Generalitat de Catalunya (grant 2005-SGR-00093) and Universitat Rovira i Virgili (project 2008TURISME-02). Jordi Nin wants to thank the Spanish Council for Scientific Research (CSIC) for his I3P grant.

References

[1] O. Abul, M. Atzori, F. Bonchi, and F. Giannotti. Hiding sequences. In *ICDE Workshops*, pages 147–156. IEEE Computer Society, 2007.

[2] N. Shoval, G. K. Auslander, T. Freytag, R. Landau, F. Oswald, U. Seidl, H.-W. Wahl, S. Werner, and J. Heinik. The use of advanced tracking technologies for the analysis of mobility in alzheimer's disease and related cognitive diseases. *BMC Geriatrics*, 8(7), 2008.

[3] G. Dong and J. Pei. *Sequence Data Mining*. Springer, 2007.

[4] T. G. Dietterich. Machine learning for sequential data: A review. In *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, pages 15–30, London, UK, 2002. Springer-Verlag.

[5] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers, 2nd edition, 2006.

[6] L. Willenborg and T. de Waal. *Elements of Statistical Diclosure Control*, volume 155 of *Lecture Notes in Statistics*. Springer, 2001.

[7] R. Agrawal and R. Srikant. Privacy-preserving data mining. In *SIGMOD '00: Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 439–450, 2000.

[8] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *Int. J. of Unc., Fuzz. and Knowledge Based Systems*, 10(5):571–588, 2002.

[9] L. Sweeney. k-anonymity: a model for protecting privacy. *Int. J. of Unc., Fuzz. and Knowledge Based Systems*, 10(5):557–570, 2002.

[10] D. Defays and P. Nanopoulos. Panels of enterprises and confidentiality: The small aggregates method. In *Proceedings of 92th Symposium on Design and Analysis of Longitudinal Surveys*, pages 195–204. Statistics Canada, Ottawa, 1993.

[11] J. Domingo-Ferrer and V. Torra. A quantitative comparison of disclosure control methods for microdata. In *Confidentiality, disclosure, and data access: theory and practical applications for statistical agencies*, pages 111–133. Elsevier, 2001.

[12] C. Gómez-Alonso and A. Valls. *Lecture Notes on Artificial Intelligence: Modeling Decisions for Artificial Intelligence (MDAI 2008)*, chapter A similarity measure for sequences of categorical data based on the ordering of common elements, pages 134–145. Springer, 2008.

[13] V. Torra and Y. Narukawa. *Modeling decisions: information fusion and aggregation operators*. Springer, 2007.

[14] R. R. Yager. On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Trans. on Systems, Man and Cybernetics*, 18:183–190, 1988.

[15] R. R. Yager. Families of owa operators. *Fuzzy Sets and Systems*, 59:125–148, 1993.

[16] F. Felsö, J. Theeuwes, and G. Wagner. Disclosure limitation in use: Results of a survey. In *Confidentiality, disclosure, and data access: theory and practical applications for statistical agencies*, pages 17–42, 2001.

[17] S. Hansen and S. Mukherjee. A polynomial algorithm for optimal univariate microaggregation. *IEEE Trans. on Kwnoledge and Data Engineering*, 15(4):1043–1044, 2003.

[18] A. Oganian and J. Domingo-Ferrer. On the complexity of optimal microaggregation for statistical disclosure control. *Statistical Journal United Nations Economic Commission for Europe*, 18(4):345–354, 2000.

[19] J. Domingo-Ferrer and J. M. Mateo-Sanz. Practical data-oriented microaggregation for statistical disclosure control. *IEEE Trans. on Kwnoledge and Data Engineering*, 14(1):189–201, 2002.

[20] A. K. Jain, M. N. Murty, and P.J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[21] A. Valls and C. Gómez-Alonso. The k-sized hierarchical clustering method for microaggregation. Technical Report DEIM-RR-09-01, Universitat Rovira i Virgili, 2009.

[22] A. Valls, C. Gómez-Alonso, and V. Torra. Generation of prototypes for masking sequences of events. In *3rd Int. Workshop on Advances in Information Security, 4th Int. Conf. on Availability, Reliability and Security (ARES)*, pages 947–952. IEEE Computer Society, 2009.

[23] R. W. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal*, 26(2):147–160, 1950.

[24] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966.

ISBN: 978-989-95079-6-8

Application of Fuzzy Vectors of Normalized Weights in Decision Making Models

Ondřej Pavlačka, Jana Talašová

Department of Mathematical Analysis and Applications of Mathematics,
Faculty of Science, Palacký University Olomouc
tř. Svobody 26, 771 46 Olomouc, Czech Republic
Email: {pavlacko,talasova}@inf.upol.cz

Abstract— *In this paper, the notion of a fuzzy vector of normalized weights is introduced, and its application in decision making models based on the weighted average operation is studied. The fuzzy vector of normalized weights represents a generalization of normalized fuzzy weights that authors formerly applied for modeling uncertain weights of criteria and/or uncertain probabilities of states of the world. Illustrative examples show that fuzzy vectors of normalized weights extend the possibilities of utilizing the vague expert information concerning weights of criteria. For three particular forms of the fuzzy vectors of normalized weights, fuzzy weighted averages of given values are calculated and the obtained results are compared. Finally, the fuzzy models of multiple criteria decision making and decision making under risk based on the fuzzy weighted average operation are described.*

Keywords— Decision making under risk; Fuzzy vectors; Fuzzy weighted average; Multiple criteria decision making; Normalized fuzzy weights.

1 Introduction

In decision making models, the weighted average is the most often used aggregation operator. In multiple criteria decision making, overall evaluations of alternatives are mostly calculated as weighted averages of their evaluations with respect to particular criteria. The weighted average operator is also used in discrete stochastic models of decision making under risk where the alternatives are usually ranked according to their expected evaluations. The expected evaluations are in fact weighted averages of evaluations of alternatives under possible states of the world; probabilities of the states of the world play the role of normalized weights.

Decision making models based on the weighted average operation can be fuzzified in two aspects: First, uncertain weighted values, i.e. uncertain evaluations with respect to particular criteria in the models of multiple criteria decision making and/or uncertain evaluations under possible states of the world in decision making under risk, can be considered. Such fuzzification is useful because it enables to process correctly e.g. the vague expert evaluations of alternatives with respect to qualitative criteria. This approach is described in detail e.g. in [1]. Second, normalized weights, i.e. weights of criteria and/or probabilities of states of the worlds, can be fuzzified. It makes the mathematical models substantially more complicated. But as the values of normalized weights are commonly set expertly, such fuzzification is very eligible.

The fuzzy extension of the weighted average operation for the case when both the weighted values and the weights are fuzzy has been studied since the second half of 70's. In [2, 3, 4], the fuzzy weights are supposed to be non-negative fuzzy numbers that express the importance of particular criteria; they are normalized only in the process of the fuzzy weighted average calculation. Later, it was shown in [5, 6] that uncertain weights expressing shares of the partial evaluation in the overall one or uncertain probabilities of states of the world have to be modeled by means of a special structure of fuzzy numbers which is called a tuple of normalized fuzzy weights. The tuple of normalized fuzzy weights is formally identical with a so called feasible tuple of fuzzy probabilities mentioned in [7]. The fuzzy weights normalization, i.e. the transformation of fuzzy weights into the normalized fuzzy weights, was studied in [8, 9].

This paper describes a new approach to modeling the uncertain normalized weights of criteria and/or the uncertain probabilities of states of the world - a notion of a fuzzy vector of normalized weights is introduced here. It will be shown that, in comparison with normalized fuzzy weights, fuzzy vectors of normalized weights extend the possibilities of utilizing the vague expert information concerning the weights.

2 Preliminaries

A fuzzy set A on a non-empty set X is characterized by its membership function $A : X \to [0,1]$. By $Ker\,A$ and $Supp\,A$, we denote a kernel of A, $Ker\,A = \{x \in X \mid A(x) = 1\}$, and a support of A, $Supp\,A = \{x \in X \mid A(x) > 0\}$, respectively. For any $\alpha \in [0,1]$, A_α means an α-cut of A, $A_\alpha = \{x \in X \mid A(x) \geq \alpha\}$. The family of all fuzzy sets on X will be denoted by $\mathcal{F}(X)$.

The intersection of two fuzzy sets $A, B \in \mathcal{F}(X)$ is defined as a fuzzy set $A \cap B \in \mathcal{F}(X)$ whose membership function is for all $x \in X$ given by $(A \cap B)(x) = \min\{A(x), B(x)\}$. Since the minimum operation is used, $(A \cap B)_\alpha = A_\alpha \cap B_\alpha$ holds for all $\alpha \in [0,1]$.

Let $X_1, X_2, \ldots, X_m$ be non-empty sets. An m-ary fuzzy relation on $X_1 \times X_2 \times \ldots \times X_m$ is defined as a fuzzy set $R \in \mathcal{F}(X_1 \times X_2 \times \ldots \times X_m)$. A particular case of an m-ary fuzzy relation is the Cartesian product of fuzzy sets, $R^C = R_1 \times R_2 \times \ldots \times R_m$, where $R_i \in \mathcal{F}(X_i)$, $i = 1, 2, \ldots, m$, and $R^C(x_1, x_2, \ldots, x_m) = \min\{R_1(x_1), R_2(x_2), \ldots, R_m(x_m)\}$ for all $(x_1, x_2, \ldots, x_m) \in X_1 \times X_2 \times \ldots \times X_m$. Since the minimum operation is used, $R^C_\alpha = R_{1\alpha} \times R_{2\alpha} \times \ldots \times R_{m\alpha}$ holds for any $\alpha \in [0,1]$.

Let us denote the index set $\{1, 2, \ldots, m\}$ by N_m throughout the paper. For any $i \in N_m$, the i-th projection of an m-ary

ISBN: 978-989-95079-6-8

fuzzy relation $R \in \mathcal{F}(X_1 \times X_2 \times \ldots \times X_m)$ is the fuzzy set $[R]_i \in \mathcal{F}(X_i)$ whose membership function is for all $x_i \in X_i$ given by

$$[R]_i(x_i) = \sup_{\substack{y_j \in X_j,\ j \in N_m \\ y_i = x_i}} R(y_1, y_2, \ldots, y_m). \qquad (1)$$

A fuzzy number is a fuzzy set C on the set of all real numbers $\mathbb{R}$ that fulfils the following conditions: a) $Ker\ C \neq \emptyset$, b) for all $\alpha \in (0,1]$, C_α are closed intervals, c) $Supp\ C$ is bounded. If $Supp\ C \subseteq [a,b]$, then C is referred to as a fuzzy number on $[a,b]$. The family of all fuzzy numbers will be denoted by $\mathcal{F}_N(\mathbb{R})$, and the family of all fuzzy numbers on $[a,b]$ will be denoted by $\mathcal{F}_N([a,b])$.

In [9, 10], it was shown that any fuzzy number C can be also determined by a couple of functions $\underline{c} : [0,1] \to \mathbb{R}$ and $\overline{c} : [0,1] \to \mathbb{R}$ that describe the minimal and maximal values of the α-cuts of C and of the closure of the support of C. The functions $\underline{c}$ and $\overline{c}$ are left-continuous on $(0,1]$, right-continuous at 0 and satisfy $\underline{c}(\alpha) \leq \underline{c}(\beta) \leq \overline{c}(\beta) \leq \overline{c}(\alpha)$ for all $0 \leq \alpha < \beta \leq 1$. In the sequel, the notation $C = \{[\underline{c}(\alpha), \overline{c}(\alpha)]\}_{\alpha \in [0,1]}$ will be used for the fuzzy number C such that $C_\alpha = [\underline{c}(\alpha), \overline{c}(\alpha)]$ for all $\alpha \in (0,1]$, and $Cl(SuppC) = [\underline{c}(0), \overline{c}(0)]$ where $Cl(SuppC)$ means the closure of the support of C. For the membership function of the fuzzy number $C = \{[\underline{c}(\alpha), \overline{c}(\alpha)]\}_{\alpha \in [0,1]}$, the following relation holds

$$C(x) = \begin{cases} \max\limits_{\alpha \in [0,1]:\ x \in [\underline{c}(\alpha), \overline{c}(\alpha)]} \alpha, & \text{for } x \in [\underline{c}(0), \overline{c}(0)], \\ 0, & \text{elsewhere.} \end{cases} \qquad (2)$$

Let us note that if $\underline{c}(\alpha) = \overline{c}(\alpha) = c$ for all $\alpha \in [0,1]$, then C represents a real number c. Furthermore, if $\underline{c}(\alpha) = a$ and $\overline{c}(\alpha) = b$ for all $\alpha \in [0,1]$, where $a < b$, then C represents a closed interval $[a,b]$.

In this paper, a fuzzy number $C = \{[\underline{c}(\alpha), \overline{c}(\alpha)]\}_{\alpha \in [0,1]}$ is called linear if both the functions $\underline{c}$ and $\overline{c}$ are linear. Any linear fuzzy number C can be fully characterized by a quadruple of real numbers $c^1 \leq c^2 \leq c^3 \leq c^4$, where $Cl(Supp\ C) = [c^1, c^4]$ and $Ker\, C = [c^2, c^3]$. In this case, the functions $\underline{c}$ and $\overline{c}$ are given for all $\alpha \in [0,1]$ as follows

$$\underline{c}(\alpha) = c^1 + \alpha(c^2 - c^1) \quad \text{and} \quad \overline{c}(\alpha) = c^4 - \alpha(c^4 - c^3). \qquad (3)$$

The notation $C = \langle c^1, c^2, c^3, c^4 \rangle$ will be used for such a linear fuzzy number C. For $c^2 \neq c^3$, the linear fuzzy number C is usually referred to as a trapezoidal fuzzy number; for $c^2 = c^3$, as a triangular fuzzy number.

The extension from functions having crisp arguments to functions with fuzzy set arguments is done according to the extension principle. Let X and Y be non-empty sets and let $f : X \to Y$ be a mapping. Then a fuzzy extension of f is the mapping $f_F : \mathcal{F}(X) \to \mathcal{F}(Y)$ such that for any $A \in \mathcal{F}(X)$, the membership function of $f_F(A)$ is defined for all $y \in Y$ as follows

$$f_F(A)(y) = \begin{cases} \sup\limits_{x \in X:\ f(x) = y} A(x), & \text{if } f^{-1}(y) \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases} \qquad (4)$$

3 Fuzzy Vectors

Since an uncertain value of one continuous variable is expressed by a fuzzy number, then, by means of analogy, a fuzzy vector is used for expressing uncertain values of m continuous variables. An m-dimensional fuzzy vector $\mathbf{V}$ is a fuzzy set on $\mathbb{R}^m$ that fulfils the following conditions: a) $Ker\ \mathbf{V} \neq \emptyset$, b) for all $\alpha \in (0,1]$, $\mathbf{V}_\alpha$ are closed and convex subsets of $\mathbb{R}^m$, c) $Supp\ \mathbf{V}$ is bounded. If $Cl(Supp\ \mathbf{V}) \subseteq Q \subset \mathbb{R}^m$, then $\mathbf{V}$ is called a fuzzy vector on Q. Let us remark that the α-cuts $\mathbf{V}_\alpha$ are compact subsets of $\mathbb{R}^m$ for all $\alpha \in (0,1]$. The family of all m-dimensional fuzzy vectors will be denoted by $\mathcal{F}_V(\mathbb{R}^m)$, and the family of all m-dimensional fuzzy vectors on $Q \subset \mathbb{R}^m$ will be denoted by $\mathcal{F}_V(Q)$. Obviously, since a closed and convex subset of $\mathbb{R}$ is a closed interval, for $m = 1$ a fuzzy number is obtained, i.e. $\mathcal{F}_V(\mathbb{R}) = \mathcal{F}_N(\mathbb{R})$.

In literature, the requirement of convexity of $\mathbf{V}_\alpha$ is often weakened, e.g. in [11], the α-cuts of fuzzy vectors are supposed to be closed and simply connected subsets of $\mathbb{R}^m$. As convexity is the characteristic property of expertly given uncertain values, the convexity of α-cuts of fuzzy vectors will be considered in the paper.

For any $i \in N_m$, the i-th projection of the m-dimensional fuzzy vector $\mathbf{V}$ will be denoted by $[\mathbf{V}]_i$. It was proved in [9, 10] that $[\mathbf{V}]_i$ is a fuzzy number and its membership function is defined for all $y \in \mathbb{R}$ by the following formula

$$[\mathbf{V}]_i(y) = \max_{\substack{x_j \in \mathbb{R},\ j \in N_m \\ x_i = y}} \mathbf{V}(x_1, x_2, \ldots, x_m). \qquad (5)$$

In the fuzzy models of decision making discussed in this paper, the expert setting of uncertain input data by fuzzy vectors is based on the following results that were proved in [9, 10].

If $X_1, X_2, \ldots, X_m$ are fuzzy numbers, then the Cartesian product $\mathbf{X} = X_1 \times X_2 \times \ldots \times X_m$ is an m-dimensional fuzzy vector, and $[\mathbf{X}]_i = X_i$ for all $i \in N_m$. Similarly, if $\mathbf{X}_i \in \mathcal{F}_V(\mathbb{R}^{n_i})$ for all $i \in N_m$, then the Cartesian product $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_m$ is an n-dimensional fuzzy vector where $n = \sum_{i=1}^m n_i$.

If the uncertain values of m variables are expressed by fuzzy numbers $X_1, X_2, \ldots, X_m$, and the set of all admissible combinations of their values is given by a crisp nonseparable m-ary relation $Q \subset \mathbb{R}^m$, then the maximal fuzzy relation (in the sense of ordering given by inclusion) expressing available information concerning the values of the m-tuple of variables is given by $\mathbf{X}_Q = (X_1 \times X_2 \times \ldots \times X_m) \cap Q$. If Q is closed and convex, then the fuzzy set $\mathbf{X}_Q$ is an m-dimensional fuzzy vector on Q if and only if

$$(Ker\ X_1 \times Ker\ X_2 \times \ldots \times Ker\ X_m) \cap Q \neq \emptyset. \qquad (6)$$

Moreover, $[\mathbf{X}_Q]_i \subseteq X_i$ for all $i \in N_m$. The equality $[\mathbf{X}_Q]_i = X_i$ holds for all $i \in N_m$ if and only if for all $i \in N_m$ and for all $\alpha \in (0,1]$ the following holds:

$$\text{For any } x_i \in X_{i\alpha},\ \text{there exist } x_j \in X_{j\alpha},\ j \in N_m,\ j \neq i, \\ \text{such that } (x_1, \ldots, x_i, \ldots, x_m) \in Q. \qquad (7)$$

Calculations with fuzzy vectors are based on the following well-known result, the proof of which can be found e.g. in [11]. Let $D \subseteq \mathbb{R}^m$, $m \geq 1$, let $f : D \to \mathbb{R}$ be a continuous function, and let $\mathbf{X} \in \mathcal{F}_V(D)$. Then $f_F(\mathbf{X})$ is a fuzzy number whose membership function is given by

$$f_F(\mathbf{X})(y) = \begin{cases} \max\limits_{\mathbf{x} \in D:\ f(\mathbf{x}) = y} \mathbf{X}(\mathbf{x}), & \text{if } f^{-1}(y) \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases} \qquad (8)$$

ISBN: 978-989-95079-6-8

and for all $\alpha \in (0,1]$ the following holds

$$f_F(\mathbf{X})_\alpha = f(\mathbf{X}_\alpha). \tag{9}$$

Employing the notation $f_F(\mathbf{X}) = \{[\underline{y}(\alpha), \overline{y}(\alpha)]\}_{\alpha\in[0,1]}$, then it follows from (9) that for any $\alpha \in (0,1]$, the values $\underline{y}(\alpha)$ and $\overline{y}(\alpha)$ can be obtained by solving the following problems of mathematical programming

$$\underline{y}(\alpha) = \min_{\mathbf{x}\in\mathbf{X}_\alpha} f(\mathbf{x}), \tag{10}$$

$$\overline{y}(\alpha) = \max_{\mathbf{x}\in\mathbf{X}_\alpha} f(\mathbf{x}). \tag{11}$$

Furthermore, since the functions $\underline{y}(\alpha)$ and $\overline{y}(\alpha)$ are right-continuous at 0, the values $\underline{y}(0)$ and $\overline{y}(0)$ are given as follows

$$\underline{y}(0) = \lim_{\alpha\to 0+} \underline{y}(\alpha) \quad \text{and} \quad \overline{y}(0) = \lim_{\alpha\to 0+} \overline{y}(\alpha). \tag{12}$$

If $\mathbf{X} = X_1 \times X_2 \times \ldots \times X_m$ where $X_i \in \mathcal{F}_N(\mathbb{R})$ for any $i \in N_m$, $X_i = \{[\underline{x}_i(\alpha), \overline{x}_i(\alpha)]\}_{\alpha\in[0,1]}$, then for all $\alpha \in [0,1]$ the following holds

$$\underline{y}(\alpha) = \min_{x_i\in[\underline{x}_i(\alpha),\overline{x}_i(\alpha)],\ i\in N_m} f(x_1, x_2, \ldots, x_m), \tag{13}$$

$$\overline{y}(\alpha) = \max_{x_i\in[\underline{x}_i(\alpha),\overline{x}_i(\alpha)],\ i\in N_m} f(x_1, x_2, \ldots, x_m). \tag{14}$$

If $\mathbf{X}_Q = (X_1 \times X_2 \times \ldots \times X_m) \cap Q$, where the crisp relation $Q \subset \mathbb{R}^m$ and fuzzy numbers $X_i = \{[\underline{x}_i(\alpha), \overline{x}_i(\alpha)]\}_{\alpha\in[0,1]}$, $i \in N_m$, fulfil the condition (6), then the values $\underline{y}(\alpha)$ and $\overline{y}(\alpha)$ can be obtained for all $\alpha \in [0,1]$ as follows

$$\underline{y}(\alpha) = \min_{\substack{x_i\in[\underline{x}_i(\alpha),\overline{x}_i(\alpha)],\ i\in N_m \\ (x_1,x_2,\ldots,x_m)\in Q}} f(x_1, x_2, \ldots, x_m), \tag{15}$$

$$\overline{y}(\alpha) = \max_{\substack{x_i\in[\underline{x}_i(\alpha),\overline{x}_i(\alpha)],\ i\in N_m \\ (x_1,x_2,\ldots,x_m)\in Q}} f(x_1, x_2, \ldots, x_m). \tag{16}$$

The formulas (15) and (16) correspond to the concept of constrained fuzzy arithmetics that was studied e.g. in [12].

4 Fuzzy Vectors of Normalized Weights

An m-dimensional fuzzy vector of normalized weights is an arbitrary m-dimensional fuzzy vector $\mathbf{W}$, whose support $Supp\ \mathbf{W}$ is a subset of the m-dimensional probability simplex $\mathcal{S}_m = \big\{(w_1, w_2, \ldots, w_m) \in \mathbb{R}^m \mid w_i \geq 0 \text{ for all } i \in N_m,\ \sum_{i=1}^m w_i = 1\big\}$. The family of all m-dimensional fuzzy vectors of normalized weights will be denoted by $\mathcal{F}_V(\mathcal{S}_m)$. Figure 1 illustrates an example of a 3-dimensional fuzzy vector of normalized weights.

From the general point of view, an m-dimensional fuzzy vector of normalized weights expresses an uncertain division of a unit into m fraction. In this paper, we will apply the fuzzy vectors of normalized weights to model the uncertain shares of partial objectives of evaluation in the overall one in fuzzy models of multiple criteria decision making, and to model the uncertain probabilities of states of the world in fuzzy-stochastic models of decision making under risk.

In the decision making models, direct expert setting of a fuzzy vector of normalized weights could be quite complicated. Therefore, three procedures of setting of fuzzy vectors of normalized weights will be proposed now.

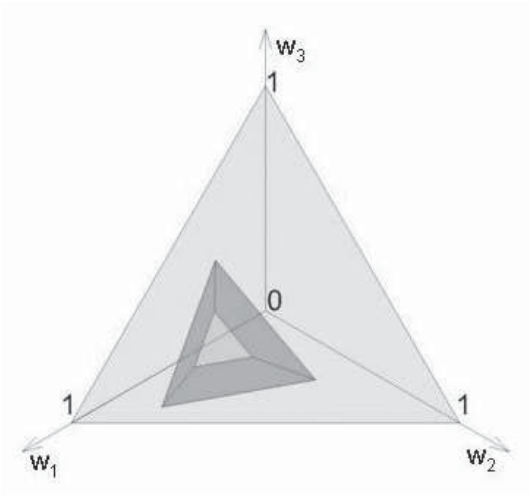


Figure 1: A 3-dimensional fuzzy vector of normalized weights.

Firstly, an expert can describe the uncertain values of normalized weights by a special structure of fuzzy numbers called normalized fuzzy weights. The normalized fuzzy weights are defined as fuzzy numbers $W_1, W_2, \ldots, W_m \in \mathcal{F}_N([0,1])$ that satisfy for all $\alpha \in (0,1]$ and for all $i \in N_m$ the following condition:

$$\text{For any } w_i \in W_{i\alpha},\ \text{there exist } w_j \in W_{j\alpha},\ j \in N_m,\ j \neq i,\ \text{such that } w_i + \sum_{j=1,\ j\neq i}^{m} w_j = 1. \tag{17}$$

The condition (17) represents the special case of (7) for $Q = \mathcal{S}_m$. Since the probability simplex $\mathcal{S}_m$ is closed and convex, the fuzzy set $\mathbf{W} = (W_1 \times W_2 \times \ldots \times W_m) \cap \mathcal{S}_m$ is a fuzzy vector of normalized weights, and $[\mathbf{W}]_i = W_i$ for all $i \in N_m$. The procedures of practical setting of normalized fuzzy weights were described in [6, 9].

Example 1 Let $m = 3$. An expert describes the uncertain values of normalized weights by the triple of linear normalized fuzzy weights $W_1 = \langle 0.4, 0.5, 0.65, 0.8\rangle$, $W_2 = \langle 0.15, 0.2, 0.35, 0.55\rangle$ and $W_3 = \langle 0.05, 0.15, 0.3, 0.45\rangle$. The corresponding fuzzy vector of normalized weights is the fuzzy set $\mathbf{W} = (W_1 \times W_2 \times W_3) \cap \mathcal{S}_3$. For all $\alpha \in (0,1]$, the α-cuts $\mathbf{W}_\alpha$ are given as follows

$$\begin{aligned} \mathbf{W}_\alpha = \big\{&(w_1, w_2, w_3) \in \mathcal{S}_3 \mid \\ &w_1 \in [0.4 + 0.1\alpha, 0.8 - 0.15\alpha], \\ &w_2 \in [0.15 + 0.05\alpha, 0.55 - 0.2\alpha], \\ &w_3 \in [0.05 + 0.1\alpha, 0.45 - 0.15\alpha]\big\}. \end{aligned} \tag{18}$$

The normalized fuzzy weights W_1, W_2, W_3 and the fuzzy vector of normalized weights $\mathbf{W} = (W_1 \times W_2 \times W_3) \cap \mathcal{S}_3$ are depicted in Figure 2. □

Secondly, an expert can give the information about normalized weights by a crisp relation $Q \subseteq \mathcal{S}_m$. If Q is non-empty, closed and convex, then it represents a special type of the fuzzy vector of normalized weights. For any $i \in N_m$, the

ISBN: 978-989-95079-6-8

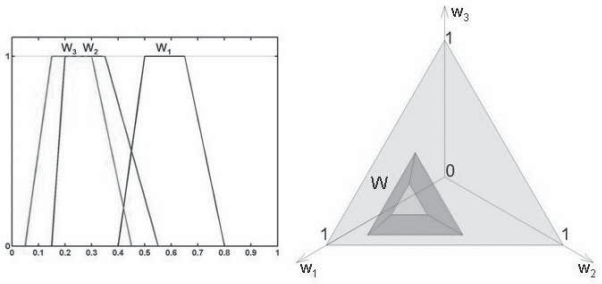


Figure 2: Normalized fuzzy weights W_1, W_2 and W_3 and the fuzzy vector of normalized weights $\mathbf{W} = (W_1 \times W_2 \times W_3) \cap \mathcal{S}_3$.

uncertain value of the i-th weight is expressed by

$$[Q]_i = \big\{w_i \in [0,1] \mid w_i \text{ is the } i\text{-th component of at least one vector } \mathbf{w} \in Q\big\}. \tag{19}$$

Example 2 Let $m = 3$. An expert gives only the ordinal information about the normalized weights by the crisp relation $T = \{(w_1, w_2, w_3) \in \mathcal{S}_3 \mid w_1 \geq w_2 \geq w_3\}$. Since T is closed and convex, T forms a 3-dimensional fuzzy vector of normalized weights $\mathbf{T}$ (see Fig. 3). Obviously, it follows from (19) that $[\mathbf{T}]_1 = [\frac{1}{3}, 1]$, $[\mathbf{T}]_2 = [0, \frac{1}{2}]$ and $[\mathbf{T}]_3 = [0, \frac{1}{3}]$. □

Thirdly, an expert can give both information about the uncertain values of weights, i.e. normalized fuzzy weights $W_1, W_2, \ldots, W_m$, and information about the crisp relations among normalized weights, i.e. the crisp relation $Q \subset \mathcal{S}_m$. If Q is non-empty, closed and convex and if $(Ker\ W_1 \times Ker\ W_2 \times \ldots \times Ker\ W_m) \cap Q \neq \emptyset$, then $\mathbf{W}_Q = (W_1 \times W_2 \times \ldots \times W_m) \cap Q$ is the corresponding fuzzy vector of normalized weights. For any $i \in N_m$, the true uncertain value of the i-th weight is expressed by $[\mathbf{W}_Q]_i \subseteq W_i$; the α-cuts of $[\mathbf{W}_Q]_i$ can be obtained for all $\alpha \in (0,1]$ in the following way

$$[\mathbf{W}_Q]_{i\alpha} = \big\{w_i \in W_{i\alpha} \mid (W_{1\alpha} \times \ldots \times W_{i-1\alpha} \times \{w_i\} \times W_{i+1\alpha} \times \ldots \times W_{m\alpha}) \cap Q \neq \emptyset\big\}. \tag{20}$$

Example 3 Let $m = 3$. First, an expert gives the ordinal information about the normalized weights by a crisp relation $T = \{(w_1, w_2, w_3) \in \mathcal{S}_3 \mid w_1 \geq w_2 \geq w_3\}$. Later on, he/she adds an uncertain cardinal information by a triple of linear normalized fuzzy weights $W_1 = \langle 0.4, 0.5, 0.65, 0.8\rangle$, $W_2 = \langle 0.15, 0.2, 0.35, 0.55\rangle$ and $W_3 = \langle 0.05, 0.15, 0.3, 0.45\rangle$. As $W_1 \geq W_2 \geq W_3$, i.e. $W_{1\alpha} \geq W_{2\alpha} \geq W_{3\alpha}$ for all $\alpha \in (0,1]$, it holds that $(KerW_1 \times KerW_2 \times KerW_3) \cap T \neq \emptyset$. The fuzzy set $\mathbf{W}_T = (W_1 \times W_2 \times W_3) \cap T$ represents the corresponding fuzzy vector of normalized weights (see Fig. 3). For all $\alpha \in (0,1]$, the following holds

$$\begin{aligned}\mathbf{W}_{T\alpha} = \big\{&(w_1, w_2, w_3) \in \mathcal{S}_3 \mid w_1 \geq w_2 \geq w_3,\\ &w_1 \in [0.4 + 0.1\alpha, 0.8 - 0.15\alpha],\\ &w_2 \in [0.15 + 0.05\alpha, 0.55 - 0.2\alpha],\\ &w_3 \in [0.05 + 0.1\alpha, 0.45 - 0.15\alpha]\big\}.\end{aligned} \tag{21}$$

According to (20), the projections of $\mathbf{W}_T$ are given as follows: $[\mathbf{W}_T]_1 = W_1$, $[\mathbf{W}_T]_2 = \{[\underline{w}_{T2}(\alpha), \overline{w}_{T2}(\alpha)]\}_{\alpha\in[0,1]}$, where $\underline{w}_{T2}(\alpha) = 0.15 + 0.05\alpha$, $\overline{w}_{T2}(\alpha) = 0.475 - 0.05\alpha$ for $0 \leq \alpha \leq 0.5$ and $\overline{w}_{T2}(\alpha) = 0.55 - 0.2\alpha$ for $0.5 < \alpha \leq 1$, and $[\mathbf{W}_T]_3 = \langle 0.05, 0.15, 0.25, 0.3\rangle$.

Let us note that although the fuzzy numbers W_1, W_2 and W_3 satisfy $W_1 \geq W_2 \geq W_3$, the crisp relation T represents an additional information for the model. Since the normalized fuzzy weights W_1, W_2 and W_3 satisfy with respect to T only the condition (6), but not (7), the projections $[\mathbf{W}_T]_1, [\mathbf{W}_T]_2$ and $[\mathbf{W}_T]_3$ are less uncertain then the normalized fuzzy weights W_1, W_2 and W_3. □

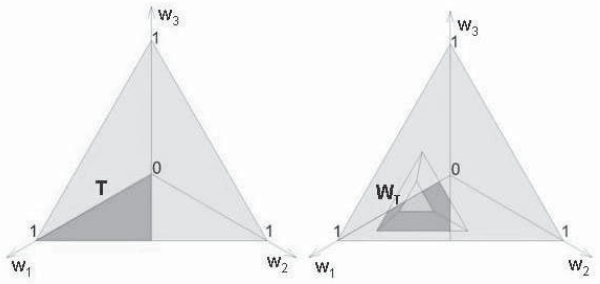


Figure 3: Fuzzy vectors of normalized weights $\mathbf{T} = \{(w_1, w_2, w_3) \in \mathcal{S}_3 \mid w_1 \geq w_2 \geq w_3\}$ and $\mathbf{W}_T = (W_1 \times W_2 \times W_3) \cap T$.

5 Fuzzy Weighted Average of Fuzzy Numbers with a Fuzzy Vector of Normalized Weights

A weighted average of real numbers $u_1, u_2, \ldots, u_m$ with a vector of normalized weights $\mathbf{w} = (w_1, w_2, \ldots, w_m)$ is defined by the following formula

$$a(\mathbf{w}, u_1, u_2, \ldots, u_m) = \sum_{i=1}^{m} w_i u_i. \tag{22}$$

Since the weighted average operation is a continuous real function defined on $\mathcal{S}_m \times \mathbb{R}^m$, its fuzzy extension a_F has the property that $a_F(\mathbf{X})$ is a fuzzy number for all $\mathbf{X} \in \mathcal{F}_V(\mathcal{S}_m \times \mathbb{R}^m)$. In decision making models, the vector of normalized weights is usually supposed to be independent of the weighted values, and the weighted values to be mutually independent as well. Therefore, only the fuzzy vectors $\mathbf{X} = \mathbf{W} \times U_1 \times U_2 \times \ldots \times U_m$, where $\mathbf{W} \in \mathcal{F}_V(\mathcal{S}_m)$ and $U_i \in \mathcal{F}_N(\mathbb{R})$ for all $i \in N_m$, will be considered in this paper.

According to (8), the fuzzy weighted average of fuzzy numbers $U_1, U_2, \ldots, U_m$ with a fuzzy vector of normalized weights $\mathbf{W}$ is a fuzzy number $U = a_F(\mathbf{W}, U_1, U_2, \ldots, U_m)$ whose membership function is for all $u \in \mathbb{R}$ given by the formula

$$U(u) = \max_{\substack{\mathbf{w}\in\mathcal{S}_m \\ u_i\in\mathbb{R},\, i\in N_m \\ \sum_{i=1}^m w_i\cdot u_i = u}} \min\big\{\mathbf{W}(\mathbf{w}), U_1(u_1), \ldots, U_m(u_m)\big\}. \tag{23}$$

Since the weighted average operation is an increasing function in variables $u_1, u_2, \ldots, u_m$, it was proved in [9] that the general mathematical programming problems (10) and (11) can be simplified. Let us denote $U_i = \{[\underline{u}_i(\alpha), \overline{u}_i(\alpha)]\}_{\alpha\in[0,1]}$ for any $i \in N_m$. For the fuzzy weighted average U, $U =$

ISBN: 978-989-95079-6-8

$\{[\underline{u}(\alpha), \overline{u}(\alpha)]\}_{\alpha\in[0,1]}$, the values of the functions $\underline{u}$ and $\overline{u}$ can be calculated for all $\alpha \in (0, 1]$ as follows

$$\underline{u}(\alpha) = \min_{(w_1,w_2,\ldots,w_m)\in \mathbf{W}_\alpha} \sum_{i=1}^{m} w_i \cdot \underline{u}_i(\alpha), \qquad (24)$$

$$\overline{u}(\alpha) = \max_{(w_1,w_2,\ldots,w_m)\in \mathbf{W}_\alpha} \sum_{i=1}^{m} w_i \cdot \overline{u}_i(\alpha). \qquad (25)$$

Hence, complexity of the calculation of U depends only on the form of the α-cuts $\mathbf{W}_\alpha$.

It is worth noting that in the case of fuzzy weighted average of fuzzy numbers with normalized fuzzy weights $W_1, W_2, \ldots, W_m$, i.e. when $\mathbf{W} = (W_1 \times W_2 \times \ldots \times W_m) \cap \mathcal{S}_m$, it is not necessary to solve the mathematical programming problems (24) and (25). For such a case, the following effective algorithm for computing the values $\underline{u}(\alpha)$ and $\overline{u}(\alpha)$, $\alpha \in [0, 1]$, was described in [6]. Let us denote $W_i = \{[\underline{w}_i(\alpha), \overline{w}_i(\alpha)]\}_{\alpha\in[0,1]}$ for any $i \in N_m$. For each $\alpha \in [0, 1]$, let $\{i_k\}_{k=1}^m$ be such a permutation on an index set N_m that $\underline{u}_{i_1}(\alpha) \le \underline{u}_{i_2}(\alpha) \le \ldots \le \underline{u}_{i_m}(\alpha)$. For $k \in N_m$, let us denote

$$w_{i_k}(\alpha) = 1 - \sum_{j=1}^{k-1} \overline{w}_{i_j}(\alpha) - \sum_{j=k+1}^{m} \underline{w}_{i_j}(\alpha). \qquad (26)$$

Let $k^* \in N_m$ be such an index that $\underline{w}_{i_{k^*}}(\alpha) \le w_{i_{k^*}}(\alpha) \le \overline{w}_{i_{k^*}}(\alpha)$. Then

$$\underline{u}^N(\alpha) = \sum_{j=1}^{k^*-1} \overline{w}_{i_j}(\alpha) \cdot \underline{u}_{i_j}(\alpha) + w_{i_{k^*}}(\alpha) \cdot \underline{u}_{i_{k^*}}(\alpha) + \sum_{j=k^*+1}^{m} \underline{w}_{i_j}(\alpha) \cdot \underline{u}_{i_j}(\alpha). \qquad (27)$$

Let $\{i_h\}_{h=1}^m$ be such a permutation on an index set N_m that $\overline{u}_{i_1}(\alpha) \ge \overline{u}_{i_2}(\alpha) \ge \ldots \ge \overline{u}_{i_m}(\alpha)$. For $h \in N_m$, let us denote

$$w_{i_h}(\alpha) = 1 - \sum_{j=1}^{h-1} \overline{w}_{i_j}(\alpha) - \sum_{j=h+1}^{m} \underline{w}_{i_j}(\alpha). \qquad (28)$$

Let $h^* \in N_m$ be such an index that $\underline{w}_{i_{h^*}}(\alpha) \le w_{i_{h^*}}(\alpha) \le \overline{w}_{i_{h^*}}(\alpha)$. Then

$$\overline{u}^N(\alpha) = \sum_{j=1}^{h^*-1} \overline{w}_{i_j}(\alpha) \cdot \overline{u}_{i_j}(\alpha) + w_{i_{h^*}}(\alpha) \cdot \overline{u}_{i_{h^*}}(\alpha) + \sum_{j=h^*+1}^{m} \underline{w}_{i_j}(\alpha) \cdot \overline{u}_{i_j}(\alpha). \qquad (29)$$

Example 4 Let us consider the fuzzy vectors of normalized weights $\mathbf{W}$ from Example 1, $\mathbf{T}$ from Example 2 and $\mathbf{W}_T$ from Example 3. Let the weighted values be, for simplicity, given by real numbers $u_1 = 0.3$, $u_2 = 0.1$ and $u_3 = 0.9$. Let us denote $U_W = a_F(\mathbf{W}, u_1, u_2, u_3)$, $U_T = a_F(\mathbf{T}, u_1, u_2, u_3)$ and $U_{W_T} = a_F(\mathbf{W}_T, u_1, u_2, u_3)$.

Since $\mathbf{W} = (W_1 \times W_2 \times W_3) \cap \mathcal{S}_3$, where W_1, W_2, W_3 are normalized fuzzy weights, by the special algorithm mentioned above we obtain that $U_W = \langle 0.22, 0.32, 0.44, 0.54\rangle$.

As $\mathbf{T} = \{(w_1, w_2, w_3) \in \mathcal{S}_3 \mid w_1 \ge w_2 \ge w_3\}$ is a crisp relation on $\mathcal{S}_3$ and weighted values are real numbers, the mathematical programming problems (24) and (25) does not depend on α. Therefore, the fuzzy weighted average U_T is equal to a closed interval, $U_T = \{[\underline{u}_T, \overline{u}_T]\}_{\alpha\in[0,1]}$. The values $\underline{u}_T = 0.2$ and $\overline{u}_T = 0.433$ were obtained by solving the following linear programming problems

$$\underline{u}_T = \min_{\substack{\sum_{i=1}^3 w_i=1 \\ w_1-w_2\ge 0 \\ w_2-w_3\ge 0 \\ w_3\ge 0}} 0.3w_1 + 0.1w_2 + 0.9w_3, \qquad (30)$$

$$\overline{u}_T = \max_{\substack{\sum_{i=1}^3 w_i=1 \\ w_1-w_2\ge 0 \\ w_2-w_3\ge 0 \\ w_3\ge 0}} 0.3w_1 + 0.1w_2 + 0.9w_3. \qquad (31)$$

The fuzzy weighted average of u_1, u_2 and u_3 with the fuzzy vector of normalized weights $\mathbf{W}_T$, $\mathbf{W}_T = (W_1 \times W_2 \times W_3) \cap T$, is the fuzzy number $U_{W_T} = \{[\underline{u}_{W_T}(\alpha), \overline{u}_{W_T}(\alpha)]\}_{\alpha\in[0,1]}$ where

$$\underline{u}_{W_T} = \min_{\substack{w_i\in W_{i\alpha},\ i\in N_3 \\ \sum_{i=1}^3 w_i=1 \\ w_1-w_2\ge 0 \\ w_2-w_3\ge 0 \\ w_3\ge 0}} 0.3w_1 + 0.1w_2 + 0.9w_3, \qquad (32)$$

$$\overline{u}_{W_T} = \max_{\substack{w_i\in W_{i\alpha},\ i\in N_3 \\ \sum_{i=1}^3 w_i=1 \\ w_1-w_2\ge 0 \\ w_2-w_3\ge 0 \\ w_3\ge 0}} 0.3w_1 + 0.1w_2 + 0.9w_3. \qquad (33)$$

From (32) and (33), we obtain that $\underline{u}_{W_T}(\alpha) = 0.235 + 0.07\alpha$ for $0 \le \alpha < 0.5$, $\underline{u}_{W_T}(\alpha) = 0.22 + 0.1\alpha$ for $0.5 \le \alpha \le 1$, and $\overline{u}_{W_T}(\alpha) = 0.42 - 0.02\alpha$.

Fig. 4 shows the fuzzy weighted averages U_W, U_T and U_{W_T}. □

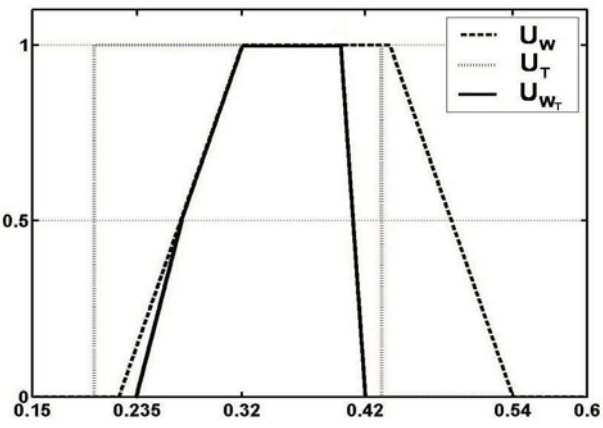


Figure 4: Fuzzy weighted averages U_W, U_T and U_{W_T}.

6 Fuzzy Models of Decision Making Based on the Fuzzy Weighted Average

In this section, a multiple criteria decision making model and a model of decision making under risk, where the fuzzy weighted average of fuzzy numbers with fuzzy vector of normalized weights is applied, will be described.

ISBN: 978-989-95079-6-8

First, let us consider a problem of multiple criteria decision making, where the best of alternatives $x_1, x_2, \ldots, x_n$ is to be chosen. The alternatives are being evaluated with respect to a given overall objective that is partitioned into m partial objectives associated with criteria $C_1, C_2, \ldots, C_m$. Let the uncertain information about the shares of the partial objectives in the overall one be given by an m-dimensional fuzzy vector of normalized weights $\mathbf{W}$. Let uncertain partial fuzzy evaluations of alternatives x_i, $i \in N_n$, with respect to the criteria C_j, $j \in N_m$, be expressed by fuzzy numbers $U_{i,j}$ defined on $[0,1]$ that represent the fuzzy degrees of satisfaction of the corresponding partial objectives of evaluation.

Then the overall fuzzy evaluation U_i of the alternative x_i, $i \in N_n$, can be expressed as follows

$$U_i = a_F(\mathbf{W}, U_{i,1}, U_{i,2}, \ldots, U_{i,m}). \tag{34}$$

The overall fuzzy evaluations $U_1, U_2, \ldots, U_n$ express the fuzzy degrees of satisfaction of the overall objective of evaluation. The best alternative is the first alternative in an ordering of the fuzzy numbers $U_1, U_2, \ldots, U_n$ or the closest to the ideal alternative whose evaluation is equal to 1. For more details on metrics and ordering of fuzzy numbers see e.g. [13]. The overall fuzzy evaluations of alternatives can also be approximated linguistically by linearly ordered elements of a proper linguistic evaluation scale defined on $[0,1]$ (see [1]).

Second, let us consider a problem of decision making under risk that is described by the following fuzzy decision matrix (see Table 1), where $x_1, x_2, \ldots, x_n$ are alternatives, $S_1, S_2, \ldots, S_r$ states of the world, $\mathbf{P}$ is an r-dimensional fuzzy vector of normalized weights that expresses the uncertain probabilities of the states of the world, and for any $i \in N_n$ and $k \in N_r$, $U_{i,k} \in \mathcal{F}_N([0,1])$ denotes the fuzzy degree in which the alternative x_i satisfies a given decision objective if the state S_k occurs.

Table 1: Fuzzy decision matrix.

	S_1	S_2	...	S_r	$\mathbf{P}$
x_1	$U_{1,1}$	$U_{1,2}$	...	$U_{1,r}$	FEU_1
x_2	$U_{2,1}$	$U_{2,2}$	...	$U_{2,r}$	FEU_2
...	...	...	...	...	...
x_n	$U_{n,1}$	$U_{n,2}$	...	$U_{n,r}$	FEU_n

It was shown in [5] that the fuzzy expected value of evaluations of alternatives $x_1, x_2, \ldots, x_n$ are for all $i \in N_n$ expressed by

$$FEU_i = a_F(\mathbf{P}, U_{i,1}, U_{i,2}, \ldots, U_{i,r}). \tag{35}$$

The best alternative is usually determined by the rule of the maximum fuzzy expected value of evaluation. The maximum fuzzy expected value of evaluation is selected from $FEU_1, FEU_2, \ldots, FEU_n$ analogously as it was shown in the case of multiple criteria decision making.

A similar approach can be applied also to multiple criteria decision making under risk (see [14]).

7 Conclusion

In this paper, a new fuzzification of the weighted average operation where uncertain normalized weights are modeled by a general fuzzy vector defined on the probability simplex was presented. Several ways of expert setting the fuzzy vector of normalized weights in fuzzy models were proposed. It was shown that, in comparison with a tuple of normalized fuzzy weights, the fuzzy vector of normalized weights extends the possibilities of modeling the expert's knowledge concerning the weights. Besides the formal definition of the fuzzy weighted average of fuzzy numbers with a fuzzy vector of normalized weights, also a general algorithm for computing the fuzzy weighted average was presented; its three particular forms were illustrated by examples. Finally, applications of the fuzzy weighted average operation in models of multiple criteria decision making and decision making under risk were described.

References

[1] J. Talašová. NEFRIT - multicriteria decision making based on fuzzy approach. *Central European Journal of Operations Research*, 8(4):297–319, 2000.

[2] S. Baas and H. Kwakernaak. Rating and ranking of multiple-aspect alternatives using fuzzy sets. *Automatica*, 13:47–58, 1977.

[3] W. M. Dong and F. S. Wong. Fuzzy weighted averages and implementation of the extension principle. *Fuzzy Sets and Systems*, 21:183–199, 1987.

[4] Y.-Y. Guh, Ch.-Ch. Hon, and E. S. Lee. Fuzzy weighted average: The linear programming approach via charness and cooper's rule. *Fuzzy Sets and Systems*, 117:157–160, 2001.

[5] J. Talašová and O. Pavlačka. Fuzzy probability spaces and their applications in decision making. *Austrian Journal of Statistics*, 35(2&3):347–356, 2006.

[6] O. Pavlačka and J. Talašová. Application of the fuzzy weighted average of fuzzy numbers in decision making models. *New Dimensions in Fuzzy Logic and Related Technologies (Eds.: Štěpnička, M., Novák, V., Bodenhofer, U.). Proceedings of the 5th EUSFLAT Conference, Ostrava, Czech Republic, September 11-14, 2007*, *II*:455–462, 2007.

[7] Y. Pan and B. Yuan. Bayesian inference of fuzzy probabilities. *Int. J. General Systems*, 26(1–2):73–90, 1997.

[8] Y.-M. Wang and T. M. S. Elhag. On the normalization of interval and fuzzy weights. *Fuzzy Sets and Systems*, 157:2456–2471, 2006.

[9] O. Pavlačka. *Fuzzy methods of decision making (in Czech)*. Dissertation thesis, Palacký University, Olomouc, 2007.

[10] O. Pavlačka and J. Talašová. Fuzzy vectors as a tool for modeling uncertain multidimensional quantities. *Submited to Fuzzy Sets and Systems in December 2008*.

[11] R. Viertl. *Statistical Methods for Non-Precise Data*. CRC Press, Boca Raton, Florida, 1996.

[12] G. J. Klir and Y. Pan. Constrained fuzzy arithmetic: basic questions and some answers. *Soft Computing*, 2:100–108, 1998.

[13] D. Dubois and H. Prade. *Fundamentals of fuzzy sets*. Kluwer Academic Publishers, Boston-London-Dordrecht, 2000.

[14] J. Talašová. Fuzzy sets in decision-making under risk (in Czech). *Proceedings of the 4th Conference Aplimat (Eds. M. Kováčová), STU, Bratislava*, pages 532–544, 2005.

ISBN: 978-989-95079-6-8

Integrating Complex User Preferences into a Route Planner: A Fuzzy-Set-Based Approach

Amine Mokhtari Olivier Pivert Allel Hadjali

IRISA/ENSSAT, Université de Rennes 1
6, Rue de Kerampont, BP 80518, 22305 Lannion Cedex, France
Email: {mokhtari, pivert, hadjali}@enssat.fr

Abstract— Route planners are systems which help users select a route between two locations. In such a context, personalization mechanisms notably aim at taking into account user preferences so as to identify the best route(s) among a set of possible answers. In this paper, we present a framework based on fuzzy set theory for modeling complex user preferences. We provide a typology of preferences which make sense in the domain considered and outline a query language integrating such flexible features. The paper also includes a discussion about query evaluation.

Keywords— Route planning, preferences, fuzzy sets.

1 Introduction

During the last century, and particularly the last fifty years, the impact of transportation means on our daily lives has constantly increased. However, it is only recently, with the tremendous development of the Internet, that an alternative solution to the use of paper maps has appeared to help users choose a route. Route planners are systems which aim at computing the "best" route between a location A and a location B, using different types of information about a given road network (structural information, of course, but also, in some cases, real-time data about traffic jams, road conditions, etc).

Even though there now exists a relatively large set of commercially available route planners, these systems have very limited capabilities — if any — when it comes to taking into account sophisticated user preferences.

The need for personalization has been felt for several years in domains such as databases or information retrieval, but not much research effort has been devoted to customizing route planners so far [20]. Currently, in the most "intelligent" commercial systems and research prototypes, only a small set of predefined preferences (e.g., avoid turnpikes, prefer freeways) are made available to the user who can sometimes attach weights to them.

In this paper, we describe a fuzzy-set-based approach to the modeling of sophisticated user preferences concerning a route planning task. Fuzzy set theory has already been used to deal with some transportations problems; in particular route choice[1], where several models based on fuzzy sets have been proposed in order to take into account uncertainty and the subjectivity of the user [17, 12].

[1]For a given origin-destination pair and a given transport mode, the route choice problem (also called traffic assignment) deals with identifying which route a given traveler *would take*. It represents the fourth step in the conventional transportation forecasting model, that aims at estimating the number of vehicles or travelers that will use a specific transportation facility in the future.

The choice of this theoretical framework is motivated by several reasons, in particular:

- the fact that it is very well suited to the interpretation of linguistics terms, which constitute a convenient way for a user to express his/her preferences,
- the fact that fuzzy set theory relies on a commensurability assumption which makes it possible to aggregate several preferences regarding different attributes, thus leading to a complete pre-order.

The remainder of the paper is organized as follows. After a survey of related work in Section 2, we deal with data representation issues in Section 3. In Section 4, we present a typology of user preferences in the context of unimodal point-to-point route planning. The main features of a query language involving such user preferences are presented in Section 5. Query evaluation is discussed in Section 6. The conclusion recalls the main contributions and outlines some perspectives for future work.

2 Related work

Building a personalized route planner is not an easy task, for several reasons. The large size and complexity of modern road networks, and the number of possible preferences present many a difficulty to overcome.

In the last decade, a few propositions have been made to personalize route planners. Some of them exploit a driver's route history to learn individual preferences that can then be used in future planning tasks. Most of the approaches assume that preferences are not given explicitly due to the difficulty of their modeling and elicitation [13].

Liu [14] proposes a route planning system which combines knowledge about the road network with case-based reasoning and brute-force search. He describes how geographical knowledge isolates the search for useful route segments to a local map region. The approach makes the strong assumption that users prefer routes that follow major roads, and the planning algorithm explicitly seeks out major roads to form the plan of a target route.

Rogers and Langley [18] propose a route planning system which can learn preferences from user feedback. During each interactive session, the user is asked to express his/her preferences among recommended routes. The feedback resulting from this interaction is used as the training data for a perceptron-style training algorithm. The authors assume a fixed user preference model which only concern route length,

ISBN: 978-989-95079-6-8

driving time and turn angles. A numeric weight can be associated to the corresponding preferences.

Let us also mention the work by McGinty and Smyth [15] about a case-based route planning approach. The system they describe generates routes which reflect implicit preferences of individual users. The main aspect that distinguishes this system from that proposed by Rogers and Langley [18] is the fact that it does not assume a fixed preference model. Every user preferences are represented as a collection of previous route cases that the user considered satisfactory. Thus, new routes are generated by reusing and combining relevant sections of multiple cases.

Letchner et al. [13] propose a route planner, named TRIP, which produces route plans that more closely match the routes chosen by people who have extensive experience travelling within a region. To do so, it incorporates time-variant road speeds learned from large amounts of driver-collected GPS data. It also exploits a driver's past GPS logs when responding to future route queries in order to provide routes which are more suited to the driver's individual driving preferences.

The work most similar to our proposal is that by Balke et al. [4, 2, 3]. These authors propose a route planning system integrating user preferences over four characteristics of a route: length, traffic jams, road works and weather conditions. To aggregate the scores related to these preferences into an overall degree, a weighted function F is used. The weights are user-defined and express the importance of each of the route characteristics. The user may use a five-level linguistic scale for asserting the importance, and the levels are automatically mapped onto numerical weights w_i. In response to a route planning query, the top k results are delivered to the user.

The main aspect which distinguishes our approach from the works [14, 15, 18] is that we choose — like Balke et al. [4, 2, 3] — to *explicitly* model user preferences. On the other hand, unlike Balke et al., we aim at taking into account a wide range of atomic user preferences, and fuzzy-set theory provides us with a very rich set of connectors for combining such preferences, leading to a highly expressive query language.

3 Representation issues

Modeling user preferences requires a clear representation of concepts, relations and geographic entities which can be involved in a route planning query. Different spatial data models have been proposed in the literature, and road networks can be represented in several ways. This section provides some elements on these topics.

3.1 Road network modeling

A road network is generally modeled as a (directed) graph. Several granularity levels can be considered, according to what the components of the graph (vertices and edges) represent. In the case of an interurban road network, a vertex represents a city, and an edge describes an interurban link (freeway, national roads, etc). However, in a urban road network, a vertex represents an intersection between two or more routes, and an edge corresponds to a street.

In our work, we consider the finest granularity: a vertex represents any intersection/bifurcation between routes or dead ends, and an edge corresponds to a road of any kind.

3.2 Geographic Data Files (GDF)

We assume a *a priori* that any spatial feature may be concerned by a user preference. Thus, interpreting and processing queries involving such a wide range of potential preferences imply to have available a model suited to the representation of these features. Geographic Data Files (GDF) [10] is international standard that specifies the conceptual and logical data model for geographic databases for Intelligent Transportation Systems (ITS) applications. It includes a specification of potential contents of such databases (Features, Attributes and Relationships), a specification of how these contents shall be represented, and of how relevant information about the database itself can be specified (meta data).

3.3 Time and its representation

We consider the following model of time: time is linear (precedence) and dense (for any two time elements there is always a third element between them). We assume time representation based on time points. A time point is represented by a real number.

4 Typology of user preferences

In the context of unimodal, point-to-point, route planning query, we distinguish three families of user preferences:

spatial preferences p_s they express preferences about roads, places or parts of the road network;

spatio-temporal preferences p_{st} they are spatial preferences involving a time component which expresses the moment or period when the spatial preference is relevant.

intrinsic preferences p_{int} they concern some global properties of a route, such as comfort, length, duration or safety.

In the next subsection, we present each family of user preferences in more detail.

4.1 Spatial preferences

Spatial preferences represent the largest class of preference in terms of diversity. Every entity of the road network may be concerned by such a preference.

Definition (Spatial preferences): a preference p is said to be spatial if it concerns a geographic entity, i.e., an entity which has spatial coordinates.

Two categories of spatial preferences may be identified: *i)* preferences on one (or several) specific element(s) of a road network, *ii)* preferences on an induced part of a road network (induced subgraph). The first category corresponds to preferences on one or several vertices/edges of a road network, i.e., on intersection(s) (resp. section(s)) of a road network. This category is used by a user to favour (respectively, disfavour) some specific elements of the network. An example is "avoid turnpikes". The second category corrresponds to preferences which concern an induced part of a road network, that we call a zone in the following. A zone is a geographic area, for instance an administrative area. Preferences on a zone are about the environment that surrounds a route. An example is "prefer a route which passes across Brittany". Two aspects of spatial entities are worthy of discussion. The first one concerns the boundary of a zone. While some zones have clear-cut (crisp)

ISBN: 978-989-95079-6-8

boundaries, as administrative areas, states or regions, others have fuzzy boundaries, such as "a polluted zone". The second aspect concerns the durability of some spatial entities. Depending on what it represents, the definition of a zone can evolve with time. It is the case, for instance, of zones defined in terms of weather conditions such as "a foggy area".

4.1.1 Spatial relationships

Two approaches exist for designating road network elements concerned by spatial user preferences. The explicit approach consists in designating the different entities concerned by the preference by means of their inherent, individual, characteristics. On the other hand, the implicit approach uses references to other spatial entities, by means of spatial relations. An example is "prefer routes which have gas stations along them".

Definition (Spatial relationship): A spatial relationship is relationship between two or more spatial objects.

Several primitive relationships exist. They can be classified into the following non-exhaustive subcategories [16]:

Topological relationships such as *adjacent*, *inside*, *disjoint*, they are invariant under topological transformations like translation, scaling or rotation.

Spatial (strict) order relationships they are based on the definition of an order, either large or strict: *left to/right to*, *above/below*, *behind/in front of*, *near/far*, *inside/outside* or *surround* [11].

Metric relationships they are based on the concept of distance in the considered space.

Directional relationships these relationships are based on an (absolute/relative) frame of reference. Cardinal direction relationships that describe orientation in space (e.g. *north*, *northeast*) are the most commonly used such relationships.

Several mathematical frameworks have been proposed for each type of spatial relations, independently or in association with one another, notably based on fuzzy set theory [5].

In a route planning query with preferences, different spatial relationships may be used. Many of them may be interpreted in a fuzzy manner while some are intrinsically crisp [21].

4.2 Spatio-temporal preferences

Time has an important place in route planning since the duration of a trip is in general a major factor as to the satisfaction of the user. Thus, it is essential to enable the user to express his/her preferences regarding this aspect.

The time dimension may appear with many semantics according to the problem to be represented. As mentioned before, in a spatio-temporal preference, time is used to express the validity period of a spatial preference. An example is "avoid the city center around noon". In other words, a spatio-temporal preference involves both spatial and temporal entities. Thus, a spatio-temporal preference p_{st} can be seen as a complex preference, formed of a spatial preference p_s and a temporal preference p_t. In the following subsections we concentrate on the temporal part of spatio-temporal preferences. We discuss the different types of temporal entities and relations that they may involve.

4.2.1 Temporal entities and representation

Three types of temporal statements may be involved in a route planning query:

- statements referring to a specific point in time, called instant, denoted by i.
- statements which make reference to a continuous time period with duration, called interval, denoted by I.
- statements which make reference to a duration (i.e., the absolute distance between two instants), denoted by d.

A classical numeric representation of time is the set of real numbers $\mathbb{R}$ [1]. Thus, an instant i is considered to be a singleton representing a date/hour (for instance: 8pm). A (time/duration) interval is described by a lower and an upper bound (beginning instant, ending instant) $[i^-, i^+]$ or a beginning instant and a duration) $[i^-, i^- + d]$.

In many cases, users tend to employ linguistic terms such as *around noon*, *early morning*, to express their temporal preferences. Fuzzy set theory provides a suitable symbolic/numeric interface to the representation of such preferences.

4.2.2 Qualitative/Quantitative temporal preferences

A route planning query may contain two types of temporal preferences: quantitative and qualitative ones.

Quantitative preferences express absolute bounds or restrict the temporal distance between two instants [6]. In other terms, they express preferences on the duration of events or their timing. Two types of quantitative preferences can be distinguished:

- unary preferences, which express a constraint on an instant i by means of a set of intervals, and are expressed as: $(i \in I_1) \vee (i \in I_2) \vee \ldots \vee (i \in I_n)$;
- binary preferences, which concern two instants i_1 and i_2 and constrain the distance $i_2 - i_1$: $(i_2 - i_1) \in I_1 \vee (i_2 - i_1) \in I_2 \vee \ldots \vee (i_2 - i_1) \in I_n$.

Example 1. Find a route from Paris to Rennes such that the duration of the trip is less than three hours.□

Qualitative preferences provide a means to specify the relative position of a pair of temporal entities (instants or intervals) t_1 and t_2 [1]. A qualitative temporal preference is expressed as $t_1 \, R \, t_2$, such that: $R \in \Re$ where $\Re$ is a finite set of basic temporal relations. The set $\Re$ contains the following types of relations:

- point-to-point relations: $<$, $\leq$, $=$, $>$, $\geq$.
- interval-to-interval relations between $I = [i^-, i^+]$ and $J = [j^-, j^+]$: *after/before* $(i^+ < j^-)$, *overlaps* $(i^- < j^- < i^+ < j^+)$, *equals* $(i^- = j^- \wedge i^+ = j^+)$.
- point-to-interval and interval-to-point relations between a point i and an interval $I = [i^-, i^+]$: *after/before* $(i < i^-)$, *during/contains* $(i^- < i < i^+)$.

ISBN: 978-989-95079-6-8

Example 2. Find a route from Paris to Berlin, with the preference "to arrive before night". Processing such a query requires to compare the arrival time with the linguistic term *night*.□

Notice that in the framework considered, temporal criteria intervene in spatio-temporal preferences, i.e., are related to a given spatial entity. In the example above, the criterion about the arrival time is related to the spatial entity "Berlin".

Let us also mention the existence of some theoretical work about the fuzzification of temporal relations [7].

4.3 Intrinsic preferences

A route can be seen as a particular geographic entity within the framework we propose. Contrary to spatial and spatio-temporal preferences which can be called *local* inasmuch as they concern a part of a route, intrinsic preferences characterize the route *as a whole*. They involve qualitative criteria such as expensive, fast, safe, and so on. Among the most intuitive properties of a route, let us mention: *rapidity*, *length*, *safety*, *cost* and *comfort*. Other properties may be of interest in some specific contexts, for example: robustness (for military applications or rescue services) or pollution produced.

5 Outline of a query language

In the previous section, we have presented the different kinds of preferences, elementary entities and possible relations involved in a route planning query with preferences. These elements constitute the basic components of a query language dedicated to route planning. In the sequel, we outline the syntax of this language as well as the different semantics that a route query can convey.

5.1 Query formalization

Let $Q^{\delta_d}_{\delta_a}(P)$ be a unimodal point-to-point route planning query with user preferences where: $\delta_d = (s_d,\ t_d)$ represents departure parameters (departure place s_d and departure time t_d), $\delta_a = (s_a,\ t_a)$ represents arrival parameters, and $P = p_1 \otimes p_2 \otimes \ldots \otimes p_n$ is a compound preference where $\otimes$ stands for a fuzzy connector (conjunction, mean, etc) and p_i is an atomic user preference. Two types of atomic preferences may be distinguished: those which concern the route as a whole, and those which concern the segment of a route. The first type takes the forms:

- $attribute\ \theta\ constant$, where θ is a crisp or fuzzy comparator, e.g., duration $\approx$ 5 hours.
- $attribute\ is\ fuzzy\_term$, e.g., cost is *high*.

The second type may take the following forms:

- $[\exists \mid \forall]$ RoadElement $R\ Entity\ \{temporal\ constraint\}$ where R is a spatial relation. An example is: prefer a route passing next to at least one gas station: $\exists$ RoadElement near gas_station. Another example is: avoid the city center between 11am and 2pm: $\forall$ RoadElement outside city_center when $t \in$ [11am, 2pm].
- $[\exists \mid \forall]$ RoadElement.attribute $\theta\ value$ where θ is a comparator and $value$ is a (linguistic or numeric) constant. For instance, avoid highways: $\forall$ RoadElement.type $\neq$ 'highway'. Another example is: prefer a route with no portion with a speed limit under 60mph: $\forall$ RoadElement.speedlimit ≥ 60.

A first query representation that could be envisaged is:

GO	**FROM**	δ_d
	TO	δ_a
	PREFERENCES	P

Example 3. Find a fast, comfortable, and inexpensive route from Rennes to Paris, with a departure at 16:00, and with the constraint that the cost of the trip must be low. This query may be formulated as follows:

GO	**FROM**	Rennes, 16:00
	TO	Paris
	PREFERENCES	duration is short **and** comfort is high **and** cost is low□

However, this representation does not distinguish between constraints and wishes (in the spirit of the concept of bipolarity [8]). Consequently, in the example above, a route can be selected even though its cost is almost not low at all, as soon as it has a very short duration and a very high level of comfort. To avoid this, user preferences may be partitioned into two sets representing two distinct components of user requirements. The first one, introduced by the keyword *preferring*, gathers the preferences, denoted by P^+, which describe the wishes of the user. The second component, introduced by the keyword *avoiding*, denoted by P^-, specifies the constraints, i.e., the characteristics of routes considered as unacceptable or undesirable by the user. The following formulation makes explicit the bipolar nature of route planning queries:

GO	**FROM**	δ_d
	TO	δ_a
	PREFERRING	P^+
	AVOIDING	P^-

Thus, the previous route query can be expressed as:

GO	**FROM**	Rennes, 16:00
	TO	Paris
	PREFERRING	duration is short **and** comfort is high
	AVOIDING	cost is high.

5.2 Route query semantics

Let $Q^{\delta_d}_{\delta_a}(P)$ be a route query. The parameter t_d (resp. t_a) may be an instant or an interval. Depending on the value of t_d (resp. t_a), such a query may take two different interpretations: static or dynamic.

Static queries (SQ): consist in computing the relevant routes between s_d and s_a for a set of preferences P, at a given instant t_d (resp. t_a). Such a query involves either t_d or t_a or none of them, but not both. If the query includes t_a, an answer will provide t_d, and reciprocally.

Dynamic queries (DQ): the parameter t_d (resp. t_a) is an interval. Such a query specifies either t_a or t_d. The objective is not only to return the relevant routes, but also to determine the best moment for the trip. An answer is thus a triple (departure time, arrival time, route). Here, the satisfaction degree associated with an answer depends on the state of the network at the moment considered (traffic jams, construction works, etc) thus takes into account dynamic data.

From a processing point of view, a dynamic query $Q^{\delta_d}_{\delta_a}(P)$ boils down to a set of static ones. Assuming a discrete time

ISBN: 978-989-95079-6-8

representation, an interval t_d (resp. t_a) corresponds to a finite set of instants $\{i_1, \ldots, i_n\}$ (resp. $\{i'_1, \ldots, i'_m\}$). Evaluating Q with $\delta_d = (s_d,\ t_d)$ (resp. $\delta_a = (s_a,\ t_a)$) comes down to evaluate $Q^{\delta'_d}_{\delta'_a}(P)$ with $\delta'_d = (s_d, t_i)$ for $i = 1..n$, (resp. with $\delta'_a = (s_a, t_{i'})$ for $i' = 1..m$), and to keep the best answer(s).

6 Route query processing

In this section, we describe the evaluation procedure of a route query with preferences. Two aspects are tackled: *i)* how to define the semantics (in terms of trapezoidal membership functions, t.m.f.[2]) of the fuzzy sets which model preferences; *ii)* how to proceed for finding the best k answers. For the former, two cases are considered:

Absolute predicates The semantics of such predicates is explicitly given by the user. It is assumed that a route planning system is endowed with an interface that allows the user to define the t.m.f. of such fuzzy terms in a convenient way. Examples of conditions involving absolute predicates are: departure time is $around\_2pm$, RoadElement is $close\_to$ seashore.

Relative predicates Examples of such predicates are: *as fast as possible*, *as short as possible*. The definition of relative predicates (i.e., their t.m.f.) is automatically computed by the system on the basis of a set of values (that we call the *context*) returned by another query (for example, the context of the relative predicates present in a user query $Q = (\delta_d,\ \delta_a,\ P)$ may be defined as the result of the query which returns the top-k' shortest routes between δ_d and δ_a without taking into account the preferences P). The idea is to define the t.m.f. of a relative predicate using the minimum, average and maximum values of the context, as illustrated in the example given further.

As to query evaluation, finding a best path in a graph w.r.t. a set of preferences is known as a multi-objective shortest path problem, where preferences represent the set of objectives to optimize. Depending on the preference model used, the goal can then be either to compute the set of non dominated routes (preference model based on Pareto order where the preferences are not commensurable) or the k best answers (as in the fuzzy-set-based preference model used here). In any case, computing the answer $\mathcal{R}$ to such a query is an NP-hard problem [19]. In the following, we propose a method which leads to an approximation of $\mathcal{R}$ in an acceptable time.

Let us consider a route planning query $Q^{\delta_d}_{\delta_a}(P)$ and assume that the user is interested in the k best route plans. The evaluation procedure proposed involves four steps:

Step 1 compute the k' shortest paths ($k' > k$) from departure place δ_d to arrival place δ_a. This can be done using one of the classical algorithms from the literature [9]. Let $\mathcal{R}_{k'}$ be the set of resulting routes;

Step 2 build the t.m.f. of the relative predicates involved in $Q^{\delta_d}_{\delta_a}(P)$ using $\mathcal{R}_{k'}$ as their context; it is assumed that "length is short" is an implicit constraint present in every query. So, the list of relative predicates always includes $short_d$ (related to the length of the route);

[2] A t.m.f is expressed by a quadruplet (A, B, a, b), where $[A, B]$ defines the core of the fuzzy set and $[A - a, B + b]$ its support.

Step 3 evaluate each route $r \in S_{k'}$ w.r.t. each atomic user preference present in $Q^{\delta_d}_{\delta_a}(P)$;

Step 4 aggregate the preference degrees into a single score (or two: one for the constraints, one for the wishes, depending on how bipolarity is handled), sort $\mathcal{R}_{k'}$ according to this score (or these two scores) and return the k best answers.

Example 4. Let us consider a person who wants to go from city A to city B, with a departure at 16:00 and an arrival around 21:00 (wish), preferably by a route which is fast (wish) and not expensive (constraint). The route query Q is:

GO	**FROM**	A, 16:00
	TO	B, around_21h
	PREFERRING	duration is short
	AVOIDING	cost is high

The fuzzy predicates involved in this query are:

- $around\_21h$: It is a user-defined (absolute) predicate. Its semantics may be given by the following trapezoidal membership function (t.m.f.): (20:30, 21:30, 1:00, 1:00).
- $short_t$: It is a relative predicate. Its t.m.f. is obtained the following way: *i)* one computes the duration of each route plan in $\mathcal{R}_{k'}$; *ii)* denoting by d_{min} and d_{avg} the minimum and the average duration respectively, then the t.m.f. associated with the predicate $short_t$ is $(d_{min}, d_{min}, 0, d_{avg} - d_{min})$.
- $high$: It is also a relative predicate. Its t.m.f. is built as follows: *i)* for each route $r \in \mathcal{R}_{k'}$, one computes the cost of the trip; *ii)* denoting by c_{avg} and c_{max} the average and the maximum of the computed costs, then the t.m.f. associated to the predicate "*high*" is $(c_{max}, \infty, c_{max} - c_{avg}, 0)$.
- implicit relative predicate $short_d$ (about length) is defined the same way as $short_t$ using the length instead of the duration.□

So as to obtain a unique scalar index for rank-ordering the routes from $\mathcal{R}_{k'}$ according to their overall satisfaction degrees w.r.t. Q, one may merge the satisfaction degree μ_{Pos} coming from the positive preferences (i.e., the wishes) with the degree μ_{Neg} related to the negative preferences (i.e., the constraints) the following way [8]:

$$\mu_Q(r) = min(\mu_{Neg}(r),\ \lambda.\mu_{Neg}(r) + (1 - \lambda).\mu_{Pos}(r))$$

where $\lambda \in]0, 1]$ is a parameter expressing some trade-off between μ_{Pos} and μ_{Neg}. In the case of our example, we have:

$$\mu_{Neg}(r) = min(1 - \mu_{high}(r.cost),\ \mu_{short_d}(r.length)),$$
$$\mu_{Pos}(r) = max(\mu_{short_t}(r.duration),\ \mu_{around\_21h}(r.t_a)).$$

In case two degrees are kept, one may rank the answers using the lexicographic order (with a priority given to the constraints, which means that the wishes are only used to break ties).

One may think that in general the value k specified by the user will be ≤ 5. A choice that we think *a priori* reasonable for k' is $4 \times k$, but experimentations will have to be performed

ISBN: 978-989-95079-6-8

so as to determine which is the "best" choice for k'. The value of k' should be big enough to include the most interesting routes, and small enough to have a reasonable computational cost.

7 Conclusion

In this paper, we have outlined a fuzzy-set-based approach to the modeling and handling of complex user preferences in the context of route planning. A typology of preferences covering a variety of intuitive preferences has been proposed and discussed. We have also provided the basis of a query language dedicated to route planning on the one hand, and described the main steps of the evaluation of a route query, on the other hand.

This is still a preliminary work, and many perspectives exist for future research. First, we intend to provide a more detailed specification of the formal query language which was just outlined here. An aspect strongly connected with this aspect, and which is worthy of investigation, is that of a user-friendly interface, which is all the more crucial as fuzzy predicates are to be handled. Also, a user study is necessary to determine the preferences which really make sense in practice (i.e the most relevant ones). As to query evaluation, it would be of interest to devise optimization mechanisms aimed at reducing the search space further. In particular, it is worth investigating whether some pruning criteria could be infered from non-intrinsic preferences, which would enable computing $\mathcal{R}_{k'}$ more efficiently. Our final objective, of course, is to implement a prototype based on the principles described here and to assess its performances by means of experimentations.

References

[1] J.F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, 1983.

[2] W.-T. Balke, W. Kießling, and C. Unbehend. Performance and quality evaluation of a personalized route planning system. In *Proc. of the Brazilian Symposium on Databases (SBBD)*, pages 328–340, 2003.

[3] W.-T. Balke, W. Kießling, and C. Unbehend. Personalized services for mobile route planning: a demonstration. In *Proc. of the 19th International Conference on Data Engineering*, pages 771–773, 2003.

[4] W.-T. Balke, W. Kießling, and C. Unbehend. A situation-aware mobile traffic information system. In *Proc. of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03)*, Washington, DC, USA, 2003.

[5] I. Bloch. Fuzzy spatial relationships for image processing and interpretation: a review. *Image and Vision Computing*, 23(2):89–110, 2005.

[6] R. Dechter, I. Meiri, and J. Pearl. Temporal constraint networks. *Artif. Intell.*, 49(1-3):61–95, 1991.

[7] D. Dubois, A. Hadjali, and H. Prade. A possibility theory-based approach to the handling of uncertain relations between temporal points. *Int. J. Intell. Syst.*, 22(2):157–179, 2007.

[8] D. Dubois and H. Prade. *Bipolarity in Flexible Querying*, volume 2522/2002 of *Lecture Notes in Computer Science*, pages 174–182. Springer Berlin / Heidelberg, 2002.

[9] D. Eppstein. Finding the k Shortest Paths. In *Proc. of the Annual Symposium on Foundations of Computer Science*, volume 35, pages 154–154, 1994.

[10] International Organization for Standardization. Intelligent transport systems – geographic data files (gdf), 2007.

[11] J. Freeman. The modelling of spatial relations. *Computer Graphics and Image Processing*, 4(2):156–171, 1975.

[12] E. Frejinger. *Route choice analysis: data, models, algorithms and applications*. PhD thesis, Lausanne, 2008.

[13] J. Letchner, J. Krumm, and E. Horvitz. Trip router with individualized preferences (trip): Incorporating personalization into route planning. In *Proc. of the 18th Innovative Applications of Artificial Intelligence Conference (IAAI 2006)*, 2006.

[14] B. Liu. Intelligent route finding: Combining knowledge, cases and an efficient search algorithm. In *Proc. of the 12th European Conference on Artificial Intelligence*, pages 380–384, 1996.

[15] L. McGinty and B. Smyth. Personalized route planning: A case-based approach. In *Proc. of the 5th European Workshop on Advances in Case-Based Reasoning*, pages 431–442. Springer-Verlag, 2000.

[16] D. Pullar and M.J. Egenhofer. Toward formal definitions of topological relations among spatial objects. In *Proc. of the 3rd International Symposium on Spatial Data Handling*, pages 225–241, 1988.

[17] M. Ridwan. Fuzzy preference based traffic assignment problem. *Transportation Research Part C: Emerging Technologies*, 12(3-4):209 – 233, 2004.

[18] S. Rogers and P. Langley. Personalized driving route recommendations. In *Proc. of the American Association of Artificial Intelligence Workshop on Recommender Systems*, pages 96–100, 1998.

[19] P. Serafini. Some considerations about computational complexity for multiobjective combinatorial problems. In J. Jahn and W. Krabs, editors, *Recent advances and historical development of vector optimization*, volume 294 of *Lecture Notes in Economics and Mathematical Systems*, Berlin, 1986. Springer-Verlag.

[20] T.C. Son and E. Pontelli. Planning with preferences using logic programming. *Theory and Practice of Logic Programming*, 6(5):559–607, 2006.

[21] H. Tveite. *Data Modelling and Database Requirements for Geographical Data*. Phd thesis, Norwegian Institute of Technology, 1997.

ISBN: 978-989-95079-6-8

Basic Properties of L-fuzzy Quantifiers of the Type $\langle 1 \rangle$ Determined by Fuzzy Measures

Antonín Dvořák, Michal Holčapek

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech Republic
Email: {antonin.dvorak,michal.holcapek}@osu.cz

Abstract— The aim of this paper is to study monadic **L**-fuzzy quantifiers of the type $\langle 1 \rangle$ determined by fuzzy measures. These fuzzy quantifiers are defined using a novel notion of $\otimes$-fuzzy integral. Several semantic properties of these **L**-fuzzy quantifiers are studied.

Keywords— fuzzy integral, fuzzy logic, fuzzy measure, fuzzy quantifier, generalized quantifier

1 Introduction

This paper studies semantic properties of one class of monadic **L**-fuzzy quantifiers [1, 2] by studying one specific but important class of them, namely fuzzy quantifiers of the type $\langle 1 \rangle$ determined by fuzzy measures.

Quantifiers of the type $\langle 1 \rangle$ are denotations of important noun phrases of natural language, e.g. "something" in "Something is broken.", "everyone" in "Everyone likes Bob.", "nobody" in "Nobody knows everything.", etc. Moreover, classical logical quantifiers "for all" and "there exists" also belong to this type. It is claimed (e.g. in [3]) that from the point of view of natural language semantics, quantifiers of the type $\langle 1, 1 \rangle$ (e.g. "every" in "Every book has leaves.", "most" in "Most birds fly.") are more basic and more important. However, it is usual and advantageous to start with the type $\langle 1 \rangle$ quantifiers, because they are simpler and there are important relationships between them and quantifiers of the type $\langle 1, 1 \rangle$.

Generalized quantifiers evolved, from pioneering works of Mostowski [4], Lindström [5], Barwise and Cooper [6], into quite large research field with deep results. For overview as well as new results see a recent monograph [3]. A quantifier of the type $\langle 1 \rangle$ is usually modeled, given a universe M, as a mapping $Q_M : \mathcal{P}_M \longrightarrow \{true, false\}$ (or, equivalently, as subsets of the power set $\mathcal{P}_M$). It is possible to introduce many properties of (models of) quantifiers, characterizing their behavior from various points of view, for example permutation invariance (PI), isomorphism invariance (ISOM), extension (EXT), and others, see Section 4.

When we think about a definition and properties of generalized quantifiers (like e.g. *many*, *a few* and others), we feel that their truth values should not change abruptly if we gradually change cardinalities of corresponding sets of objects. Consider for example a sentence "Many people read books." If the number of people who read books increase by 1, it would be very strange if the truth value of this sentence changes from false to true. Therefore, it was inevitable that researchers started to consider more than two truth values in this context, and so-called *fuzzy quantifiers* emerged, starting from a generalization of definition from the previous paragraph, where instead of $\{true, false\}$ we consider some other structure of truth values, notably the interval $[0, 1]$.

Research in the field of fuzzy quantifiers started with works of Zadeh [7], Thiele [8], Ralescu [9] and others, see also [10, 11, 12]. An important contribution was made by Hájek in [13]. A comprehensive study of fuzzy quantifiers was undertaken by Glöckner [1] (see also [14]). In the recent paper [15], Novák studies so-called intermediate quantifiers, mainly from the syntactic point of view in the frame of fuzzy type theory [16]. An attempt to model linguistic quantifiers by fuzzy (Sugeno) integral was presented by Ying in [17].

The semantic interpretation of many generalized quantifiers is connected to measurement of "size" of sets in concern. Consider e.g. quantifier "many". The truth value of a proposition "many books have red cover" clearly depends on the "size" of the set of red books. Therefore, it is natural to consider measures (and integrals) of (fuzzy) sets as natural tools for the modeling of important classes of monotonically non-decreasing and monotonically non-increasing generalized quantifiers.

Fuzzy measures and integrals ([18], see also [19, 20]) are important tools allowing us to compare sets with respect to their size. Standardly, fuzzy measures are set functions defined on some algebra of sets which are monotone with respect to inclusion and they assign zero to the empty set. In our approach, fuzzy measures are defined on algebras of fuzzy sets (fuzzy measure spaces) and, generally, they attain values from a complete residuated lattice **L**. Details can be found in [21] in these proceedings, here we present only basic ideas.

Two types of fuzzy integral, namely the $\otimes$-fuzzy integral and the $\rightarrow$-fuzzy integral, are defined on an arbitrary fuzzy measure space. Integrals of $\otimes$ type will be used as models of quantifiers like *all* and *some*, while integrals of $\rightarrow$ type as models of *no* and *not all*, etc. However, in this contribution we will concentrate only on $\otimes$-fuzzy integrals and quantifiers defined by means of them. Nevertheless, if the structure of truth values **L** is a complete MV-algebra, then it is possible to define the $\rightarrow$-fuzzy integral from the $\otimes$-fuzzy integral [21].

2 Preliminaries

For details we refer to our contribution [21] in this volume, here we review only a few necessary notions.

2.1 Structures of truth values

In this paper, we suppose that the structure of truth values is a *complete residuated lattice* (see e.g. [22]), i.e., an algebra $\mathbf{L} = \langle L, \wedge, \vee, \rightarrow, \otimes, \bot, \top \rangle$ with four binary operations and

ISBN: 978-989-95079-6-8

two constants such that $\langle L, \wedge, \vee, \bot, \top\rangle$ is a complete lattice, where $\bot$ is the least element and $\top$ is the greatest element of L, respectively, $\langle L, \otimes, \top\rangle$ is a commutative monoid (i.e., $\otimes$ is associative, commutative and the identity $a \otimes \top = a$ holds for any $a \in L$) and the adjointness property is satisfied, i.e.,

$$a \leq b \to c \quad \text{iff} \quad a \otimes b \leq c \tag{1}$$

holds for each $a, b, c \in L$, where $\leq$ denotes the corresponding lattice ordering. The operations $\otimes$ and $\to$ are usually called multiplication and residuum, respectively.

2.2 L-fuzzy sets

Let $\mathbf{L} = \langle L, \wedge, \vee, \to, \otimes, \bot, \top\rangle$ be a complete residuated lattice and M be a universe of discourse (possibly empty). A mapping $A : M \to L$ is called an **L**-*fuzzy set on* M. A value $A(m)$ is called a *membership degree of* m *in the* **L**-*fuzzy set* A. The set of all **L**-fuzzy sets on M is denoted by $\mathcal{F}_{\mathbf{L}}(M)$. Obviously, if $M = \emptyset$, then the empty mapping $\emptyset$ is the unique **L**-fuzzy set on $\emptyset$ and thus $\mathcal{F}(\emptyset) = \{\emptyset\}$. An **L**-fuzzy set A on M is called *crisp*, if there is a subset X of M such that $A = 1_X$, where 1_X denotes the characteristic function of X. Particularly, $1_\emptyset$ denotes the empty **L**-fuzzy set on M, i.e., $1_\emptyset(m) = \bot$ for any $m \in M$. This convention will be also kept for $M = \emptyset$. The set of all crisp **L**-fuzzy sets on M is denoted by $\mathcal{P}_{\mathbf{L}}(M)$. An **L**-fuzzy set A is *constant*, if there is $c \in L$ such that $A(m) = c$ for any $m \in M$. For simplicity, a constant **L**-fuzzy set is denoted by the corresponding element of L, e.g., a, b, c.[1]

Let A be an **L**-fuzzy set on M. The *complement* of A is an **L**-fuzzy set $\overline{A}$ on M defined by $\overline{A}(m) = \neg A(m)$ for any $m \in M$. Finally, an extension of the operations $\otimes$ and $\to$ on L to the operations on $\mathcal{F}_{\mathbf{L}}(M)$ is given by

$$(A \otimes B)(m) = A(m) \otimes B(m) \tag{2}$$

$$(A \to B)(m) = A(m) \to B(m) \tag{3}$$

for any $A, B \in \mathcal{F}_{\mathbf{L}}(M)$ and $m \in M$, respectively. A mapping $f^{\to} : \mathcal{F}_{\mathbf{L}}(M) \to \mathcal{F}_{\mathbf{L}}(M')$ defined by $f^{\to}(A)(m) = \bigvee_{m' \in f^{-1}(m)} A(m')$ is called a *fuzzy extension* of the mapping f. Obviously, if f is a bijective mapping, then $f^{\to}(A)(f(m)) = A(m)$ for any $m \in M$.

3 Fuzzy measures and integrals

In this section, we will review a notion of a fuzzy measure of **L**-fuzzy sets and of a $\otimes$-fuzzy integral that will be used to define **L**-fuzzy quantifiers. For details and examples see [21], where also definitions of a complementary fuzzy measure and a $\to$-fuzzy integral can be found. For more information about fuzzy integrals, we refer to [18, 19].

3.1 Fuzzy measures of L-fuzzy sets

For our purposes we will consider algebras of **L**-fuzzy sets as a base for defining fuzzy measures of **L**-fuzzy sets.

Definition 3.1 ([18])**.** Let M be a non-empty universe of discourse. A subset $\mathcal{M}$ of $\mathcal{F}_{\mathbf{L}}(M)$ is an *algebra of* **L**-*fuzzy sets on* M, if the following conditions are satisfied

(i) $1_\emptyset, 1_M \in \mathcal{M}$,

(ii) if $A \in \mathcal{M}$, then $\overline{A} \in \mathcal{M}$,

(iii) if $A, B \in \mathcal{M}$, then $A \cup B \in \mathcal{M}$.

A couple $(M, \mathcal{M})$ is called a *fuzzy measurable space*, if $\mathcal{M}$ is an algebra of **L**-fuzzy sets on M.

Let us introduce the concept of fuzzy measure as follows. The first definition is a modification of the definition of a normed measure with respect to truth values (see e.g. [19, 20]).

Definition 3.2. Let $(M, \mathcal{M})$ be a fuzzy measurable space. A mapping $\mu : \mathcal{M} \to L$ is called a *fuzzy measure* on $(M, \mathcal{M})$, if

(i) $\mu(1_\emptyset) = \bot$ and $\mu(1_M) = \top$,

(ii) if $A, B \in \mathcal{M}$ such that $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

A triplet $(M, \mathcal{M}, \mu)$ is called the *fuzzy measure space*, if $(M, \mathcal{M})$ is a fuzzy measurable space and μ is a fuzzy measure on $(M, \mathcal{M})$.

Example 3.1. Let **L** be a complete residuated lattice with the support $[0, 1]$ and $\mathbb{N}$ be the set of natural numbers with 0. For any non-empty countable (finite or denumerable) universe M, injective mapping $f : M \to \mathbb{N}$, $n \in \mathbb{N}$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$, denote

$$A_{f,n}(m) = \begin{cases} A(m), & \text{if } f(m) \leq n; \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Further, for any injective mapping $f : M \to \mathbb{N}$ and $n \in \mathbb{N}$, define $\mu_{f,n} : \mathcal{F}_{\mathbf{L}}(M) \to [0, 1]$ as follows

$$\mu_{f,n}(A) = \frac{\sum_{m \in \mathrm{Supp}(A_{f,n})} A_{f,n}(m)}{|\mathrm{Supp}(1_{M_{f,n}})|} \tag{5}$$

and, finally, define $\underline{\mu}_f, \overline{\mu}_f : \mathcal{F}_{\mathbf{L}}(M) \to [0, 1]$ as follows

$$\underline{\mu}_f = \liminf_{n \to \infty} \mu_{f,n}(A), \tag{6}$$

$$\overline{\mu}_f = \limsup_{n \to \infty} \mu_{f,n}(A). \tag{7}$$

It is easy to see that $\mu_{f,n}$, $\underline{\mu}_f$ and $\overline{\mu}_f$ are fuzzy measures on $(M, \mathcal{F}_{\mathbf{L}}(M))$ determined by an injective mapping f. If, for example, $M = \mathbb{N}$ and $f = \mathrm{id}$, then $\underline{\mu}_f(A) = \overline{\mu}_f(A) = \bot$ for any **L**-fuzzy set on a finite universe. For the set of all even or odd numbers, both fuzzy measures give $\frac{1}{2}$ and, for the set of all prime numbers, we obtain 0.

If M is finite, then $\underline{\mu}_f = \underline{\mu}_g = \overline{\mu}_f = \overline{\mu}_g$ for any injective mappings $f, g : M \to \mathbb{N}$ and

$$\underline{\mu}_f(A) = \overline{\mu}_f(A) = \frac{\sum_{m \in M} A(m)}{|M|}. \tag{8}$$

Let $(M, \mathcal{M})$ be a fuzzy measurable space and $X \in \mathcal{F}_{\mathbf{L}}(M)$. Denote $\mathcal{M}_X$ the set of all $\mathcal{M}$-measurable sets which are contained in X, i.e.,

$$\mathcal{M}_X = \{A \mid A \in \mathcal{M} \text{ and } A \subseteq X\}. \tag{9}$$

Note that $1_\emptyset \in \mathcal{M}_X$ for each $X \in \mathcal{F}_{\mathbf{L}}(M)$ and if X is $\mathcal{M}$-measurable set, then also $X \in \mathcal{M}_X$. If $X = M$, then we will write only $\mathcal{M}$ instead of $\mathcal{M}_M$.

In the following part we will define an isomorphism between fuzzy measure spaces.

[1] We suppose that the meaning of this symbol will be unmistakable from the context, that is, it should be clear when an element of L is considered and when a constant **L**-fuzzy set is assumed.

ISBN: 978-989-95079-6-8

Definition 3.3. Let $(M, \mathcal{M})$ and $(M', \mathcal{M}')$ be fuzzy measurable spaces. We say that a mapping $g : \mathcal{M} \to \mathcal{M}'$ is an *isomorphism between* $(M, \mathcal{M})$ *and* $(M', \mathcal{M}')$, if

(i) g is a bijective mapping with $g(1_\emptyset) = 1_\emptyset$,

(ii) $g(A \cup B) = g(A) \cup g(B)$ and $g(\overline{A}) = \overline{g(A)}$ hold for any $A, B \in \mathcal{M}$,

(iii) there exists a bijective mapping $f : M \to M'$ with $A(m) = g(A)(f(m))$ for any $A \in \mathcal{M}$ and $m \in M$.

Definition 3.4. Let $(M, \mathcal{M})$ and $(M', \mathcal{M}')$ be fuzzy measurable spaces. We say that a mapping $g : \mathcal{M} \to \mathcal{M}'$ is an *isomorphism between* $(M, \mathcal{M}, \mu)$ *and* $(M', \mathcal{M}', \mu')$, if

(i) g is an isomorphism between $(M, \mathcal{M})$ and $(M', \mathcal{M}')$,

(ii) $\mu(A) = \mu'(g(A))$.

If g is an isomorphism between fuzzy measure spaces $(M, \mathcal{M}, \mu)$ and $(M', \mathcal{M}', \mu')$, then we write $g(M, \mathcal{M}, \mu) = (M', \mathcal{M}', \mu')$.

Let $[(M, \mathcal{M}, \mu)]$ denote the class of all fuzzy measure spaces defined on M that are isomorphic with $(M, \mathcal{M}, \mu)$. Obviously, we can write

$$[(M, \mathcal{M}, \mu)] = \{(M, f^{\rightarrow}(\mathcal{M}), \mu_{f^{\rightarrow}}) \mid \\ \mid f : M \to M \text{ is a bijective mapping}\}.$$

3.2 $\otimes$-fuzzy integral

Definition 3.5. Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $X \in \mathcal{M}$. The *$\otimes$-fuzzy integral of A on X* is given by

$$\int_X^{\otimes} A \, d\mu = \bigvee_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \bigwedge_{m \in \mathrm{Supp}(Y)} (A(m) \otimes \mu(Y)). \quad (10)$$

If $X = 1_M$, then we write $\int^{\otimes} A \, d\mu$.

Theorem 3.1. *Let $(M, \mathcal{M}, \mu)$ be a fuzzy measure space. If $X \in \mathcal{M}$ is such that $1_{\mathrm{Supp}(Y)} \in \mathcal{M}_X$ for any $Y \in \mathcal{M}_X$, then for any $A \in \mathcal{F}_{\mathbf{L}}(M)$*

$$\int_X^{\otimes} A \, d\mu = \bigvee_{1_Y \in \mathcal{P}_X \setminus \{1_\emptyset\}} \bigwedge_{m \in Y} (A(m) \otimes \mu(1_Y)), \quad (11)$$

where $\mathcal{P}_X = \{1_{\mathrm{Supp}(Z)} \mid Z \in \mathcal{M}_X\}$.

Theorem 3.2. *Let $\mathbf{L}$ be a complete MV-algebra, $(M, \mathcal{M}, \mu)$ be a fuzzy measure space, $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $X \in \mathcal{M}$. Then*

$$\int_X^{\otimes} A \, d\mu = \bigvee_{Y \in \mathcal{M}_X \setminus \{1_\emptyset\}} \big(\mu(Y) \otimes \bigwedge_{m \in \mathrm{Supp}(Y)} A(m)\big). \quad (12)$$

Moreover,

$$\int_X^{\otimes} (c \otimes A) \, d\mu = c \otimes \int_X^{\otimes} A \, d\mu \quad (13)$$

for any $c \in L$.

Theorem 3.3. *Let g be an isomorphism between fuzzy measure spaces $(M, \mathcal{M}, \mu)$ and $(M', \mathcal{M}', \mu')$ and $X \in \mathcal{M}$. Then*

$$\int_X^{\otimes} A \, d\mu = \int_{g(X)}^{\otimes} g(A) \, d\mu' \quad (14)$$

for any $A \in \mathcal{F}_{\mathbf{L}}(M)$.

In the end of this part we will show that the Sugeno integral is a special case of our proposed integral. Let $\mathbf{L}$ be a complete residuated lattice and $(M, \mathcal{M})$ be a fuzzy measurable space such that $A \cap B \in \mathcal{M}$ for any $A, B \in \mathcal{M}$. Denote $A_a = \{m \mid m \in M \ \& \ A(m) \geq a\}$. We say that an $\mathbf{L}$-fuzzy set A is *$\mathcal{M}$-Sugeno measurable*, if $1_{A_a} \in \mathcal{M}$ for any $a \in L$. The *Sugeno integral* is given, for any fuzzy measure space $(M, \mathcal{M}, \mu)$ with $A \cap B \in \mathcal{M}$ for any $A, B \in \mathcal{M}$, for any $\mathcal{M}$-Sugeno measurable $\mathbf{L}$-fuzzy set A and for any $X \in \mathcal{M}$, by

$$\int_X A \, d\mu = \bigvee_{a \in L} (a \wedge \mu(1_{A_a} \cap X)). \quad (15)$$

Theorem 3.4. *Let $\mathbf{L}$ be a complete Heyting algebra, $(M, \mathcal{M}, \mu)$ be a fuzzy measure space with $A \cap B \in \mathcal{M}$ for any $A, B \in \mathcal{M}$, A be an $\mathcal{M}$-Sugeno measurable $\mathbf{L}$-fuzzy set and $X \in \mathcal{M}$. Then $\int_X A \, d\mu = \int_X^{\otimes} A \, d\mu$.*

For a definition and properties of $\rightarrow$-fuzzy integral we refer to [21].

4 L-fuzzy quantifiers of the type $\langle 1 \rangle$

In [2], we defined the monadic $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1^n, 1 \rangle$. Here, we restrict ourselves to their special subclass, namely, to the monadic $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1 \rangle$ that can be defined as follows.

Definition 4.1. Let $\mathbf{L}$ be a complete residuated lattice, M be a universe (possibly empty[2]). A mapping $Q_M : \mathcal{F}_{\mathbf{L}}(M) \to L$ is called a *monadic $\mathbf{L}$-fuzzy quantifier of the type $\langle 1 \rangle$ limited to M*.

Definition 4.2. An *unlimited (finite, countable) monadic $\mathbf{L}$-fuzzy quantifier of the type $\langle 1 \rangle$* is a functional Q assigning to each (finite, countable) universe M a monadic $\mathbf{L}$-fuzzy quantifier Q_M of the type $\langle 1 \rangle$ limited to M.

In the following text, we will usually omit the terms "unlimited", "monadic" and "of the type $\langle 1 \rangle$" and we will say only "$\mathbf{L}$-fuzzy quantifier". Let us demonstrate several examples of unlimited $\mathbf{L}$-fuzzy quantifiers that are interpretations of well-known quantifiers in natural language (see [2]). We will use expressions all, some, not all and no as generic expressions which stand for natural language quantifiers of the type $\langle 1 \rangle$, e.g. "everything", "someone", "not everyone" and "nothing", respectively.

[2] To define the behavior of generalized quantifiers for the empty universe is important in some situations. It happens, for example, when we study type $\langle 1, 1 \rangle$ quantifiers which are obtained from type $\langle 1 \rangle$ quantifiers by means of *relativization*. Then it is vital to have values of e.g. $\mathsf{some}_\emptyset(1_\emptyset)$ defined, see discussion in [3], p. 137.

ISBN: 978-989-95079-6-8

Example 4.1. Let $\mathbf{L}$ be a complete residuated lattice. Then

$$(\mathsf{all})_M(A) = \bigwedge_{m\in M} A(m),$$
$$(\mathsf{some})_M(A) = \bigvee_{m\in M} A(m),$$
$$(\mathsf{not\ all})_M(A) = \bigvee_{m\in M} \neg A(m),$$
$$(\mathsf{no})_M(A) = \bigwedge_{m\in M} \neg A(m),$$

where M is an arbitrary universe and $A \in \mathcal{F}_{\mathbf{L}}(M)$, define unlimited $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1\rangle$. Obviously, the definitions of all and some (interpretations of quantifiers *all* and *some*) are the same as the interpretations of $\forall$ and $\exists$ in fuzzy logic, respectively. The others are negations of the previous ones. Notice that $(\mathsf{all})_\emptyset(\emptyset) = (\mathsf{no})_\emptyset(\emptyset) = \top$ and $(\mathsf{some})_\emptyset(\emptyset) = (\mathsf{not\ all})_\emptyset(\emptyset) = \bot$.

Now, let us recall some well-known semantics properties that are usually investigated in the case of the $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1\rangle$. For more information as well as examples we refer to [1, 2].

Definition 4.3. Let Q, P be $\mathbf{L}$-fuzzy quantifiers. Then we say that Q *is less than or equal to* P and denote it by $Q \le P$, if, for any non-empty universe M and $A \in \mathcal{F}_{\mathbf{L}}(M)$,

$$Q_M(A) \le P_M(A). \tag{16}$$

Further, we say that Q *is equal to* P and denote it by $Q = P$, if $Q \le P$ and $P \le Q$.

Definition 4.4. Let Q, P be $\mathbf{L}$-fuzzy quantifiers. We say that Q *is identical to* P and denote it by $Q \equiv P$, if for any (possibly empty) universe M and $A \in \mathcal{F}_{\mathbf{L}}(M)$,

$$Q_M(A) = P_M(A). \tag{17}$$

Remark 4.2. Note that the behavior of $\mathbf{L}$-fuzzy quantifiers for the empty universe is often unpredictable (e.g., $\mathsf{all}_M(A) \le \mathsf{some}_M(A)$ for all $M \neq \emptyset$, but $\mathsf{some}_\emptyset(1_\emptyset) \le \mathsf{all}_\emptyset(1_\emptyset)$), therefore, we require only non-empty universes for their comparison in the first definition. Moreover, this restriction seems to be insignificant from the practical point of view. The second definition of identity of $\mathbf{L}$-fuzzy quantifiers gives useful denotation.

Definition 4.5. An $\mathbf{L}$-fuzzy quantifier Q is *permutation-invariant*, if for arbitrary universe M, bijective mapping $f : M \to M$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$

$$Q_M(A) = Q_M(f^{\rightarrow}(A)). \tag{18}$$

The set of all permutation-invariant $\mathbf{L}$-fuzzy quantifiers is denoted by PI.

Definition 4.6. An $\mathbf{L}$-fuzzy quantifier Q is *isomorphism-invariant*, if for arbitrary universe M, bijective mapping $f : M \to M'$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$,

$$Q_M(A) = Q_{M'}(f^{\rightarrow}(A)). \tag{19}$$

The set of all isomorphism-invariant $\mathbf{L}$-fuzzy quantifiers is denoted by ISOM.

Definition 4.7. An $\mathbf{L}$-fuzzy quantifier Q satisfies *extension*, if for arbitrary universes M, M' with $M \subseteq M'$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$,

$$Q_M(A) = Q_{M'}(A). \tag{20}$$

The set of all $\mathbf{L}$-fuzzy quantifiers satisfying extension is denoted by EXT.

Definition 4.8. Let Q be an $\mathbf{L}$-fuzzy quantifier. We say that Q is *monotonically non-decreasing*, if for arbitrary universe M and $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $A' \in \mathcal{F}_{\mathbf{L}}(M)$ with $A \subseteq A'$,

$$Q_M(A) \le Q_M(A') \tag{21}$$

and Q is *monotonically non-increasing*, if for arbitrary universe M and $A \in \mathcal{F}_{\mathbf{L}}(M)$ and $A' \in \mathcal{F}_{\mathbf{L}}(M)$ with $A' \subseteq A$,

$$Q_M(A) \le Q_M(A'). \tag{22}$$

For our purpose we will consider the following stronger definition of $\mathbf{L}$-similarity of fuzzy sets. Recall that a mapping $R : \mathcal{F}_{\mathbf{L}}(M) \times \mathcal{F}_{\mathbf{L}}(M) \to L$ is called an $\mathbf{L}$-*fuzzy relation on* $\mathcal{F}_{\mathbf{L}}(M)$. Let $[A\ R\ B]$ denote the degree in which $\mathbf{L}$-fuzzy sets A and B belongs to $\mathbf{L}$-fuzzy relation, i.e., $[A\ R\ B] = R(A, B)$. Let us define an $\mathbf{L}$-fuzzy relation $\equiv_M$: $\mathcal{F}_{\mathbf{L}}(M) \times \mathcal{F}_{\mathbf{L}}(M) \to L$ by $[A \equiv_M B] = \top$, if there is a bijective mapping f of M onto M such that $f^{\rightarrow}(A) = B$, and $[A \equiv_M B] = \bot$, otherwise. The following definition generalizes $\equiv_M$.

Definition 4.9. An $\mathbf{L}$-fuzzy relation $\approx_M$: $\mathcal{F}_{\mathbf{L}}(M) \times \mathcal{F}_{\mathbf{L}}(M) \to L$ is called an $\mathbf{L}$-*permutation equivalence* on $\mathcal{F}_{\mathbf{L}}(M)$, if

$$[A \approx_M B] \ge [A \equiv_M B] \tag{23}$$
$$[A \approx_M B] = [B \approx_M A] \tag{24}$$
$$[A \approx_M B] \le [\overline{A} \approx_M \overline{B}] \tag{25}$$
$$[A \approx_M B] \otimes [B \approx_M C] \le [A \approx_M C] \tag{26}$$

hold for arbitrary $A, B, C \in \mathcal{F}_{\mathbf{L}}(M)$.

Obviously, (24) and (26) are the common axioms of symmetry and transitivity, respectively. Let $A, B \subseteq M$ are $\mathbf{L}$-fuzzy sets which are similar. Then one could wish that the complements of A and B are also similar (at least in the degree in which A and B are $\mathbf{L}$-equivalent). This idea is expressed in (25).

Example 4.3 (see [2])**.** Let $\mathbf{L}$ be a complete residuated lattice, M be any universe and $\mathrm{Perm}(M)$ denote the set of all bijective mappings of M onto M. Then

$$[A \approx^{\wedge}_M B] = \bigvee_{f\in \mathrm{Perm}(M)} \bigwedge_{m\in M} (A(m) \leftrightarrow B(m)) \tag{27}$$

defines the $\mathbf{L}$-permutation equivalence $\approx^{\wedge}_M$ on $\mathcal{F}_{\mathbf{L}}(M)$.

Definition 4.10. Let $\approx$ be a class of $\mathbf{L}$-permutation equivalences such that for each (finite, countable) universe M there is a unique $\approx_M$ from $\approx$ defined on $\mathcal{F}_{\mathbf{L}}(M)$. A (finite, countable) $\mathbf{L}$-fuzzy quantifier Q of the type $\langle 1\rangle$ is *extensional with respect to* $\approx$, if

$$[A \approx_M A'] \le Q_M(A) \leftrightarrow Q_M(A') \tag{28}$$

for each (finite, countable) universe M and $A, A' \in \mathcal{F}_{\mathbf{L}}(M)$. The set of all extensional $\mathbf{L}$-fuzzy quantifiers with respect to $\approx$ is denoted by $\mathrm{EXTENS}(\approx)$.

ISBN: 978-989-95079-6-8

5 L-fuzzy quantifiers of the type $\langle 1\rangle$ determined by fuzzy measures

Let $\mathcal{S}(M)$ denote a set of fuzzy measure spaces defined on M. For better readability, we will denote by

$$\int_{(M,\mathcal{M})}^{\otimes} A\, d\mu \tag{29}$$

the $\otimes$-fuzzy integral $\int^{\otimes} A\, d\mu$ defined over a fuzzy measure space $(M,\mathcal{M},\mu)$. Now we can define $\mathbf{L}$-fuzzy quantifiers limited to M using fuzzy measure spaces from a set $\mathcal{S}(M)$ as follows.

Definition 5.1. Let $\mathcal{S}(M)$ be a (possibly empty) set of fuzzy measure spaces defined on a non-empty universe M. An $\mathbf{L}$-*fuzzy quantifier of the type* $\langle 1\rangle$ *limited to* M *determined by the fuzzy measure spaces from* $\mathcal{S}(M)$ is a mapping $Q_{\mathcal{S}(M)} : \mathcal{F}_{\mathbf{L}}(M) \to L$ defined by

$$Q_{\mathcal{S}(M)}(A) = \bigvee_{(M,\mathcal{M},\mu)\in\mathcal{S}(M)} \int_{(M,\mathcal{M})}^{\otimes} A\, d\mu. \tag{30}$$

Remark 5.1. It is easy to see that if $\mathcal{S}(M) = \emptyset$ for some (possibly non-empty) set M, then $Q_{\mathcal{S}(M)}(A) = \bot$ for any $A \in \mathcal{F}_{\mathbf{L}}(M)$.

It is easy to see that $\mathcal{S}(\emptyset) = \emptyset$ (there is no fuzzy measure space with $M = \emptyset$). Hence, each unlimited fuzzy quantifier Q based only on the formula (30) has $Q_\emptyset(1_\emptyset) = Q_{\mathcal{S}(\emptyset)}(1_\emptyset) = \bot$. However, for example, it holds that $(\mathsf{all})_\emptyset(1_\emptyset) = \top$. This motivates us to exclude the determination of $Q_\emptyset(1_\emptyset)$ by (30) in the following definition of unlimited $\mathbf{L}$-fuzzy quantifier.

Definition 5.2. Let $\mathcal{S}$ be a functional assigning to each universe M a set $\mathcal{S}(M)$ of fuzzy measure spaces defined on M. An *unlimited* $\mathbf{L}$-*fuzzy quantifier of the type* $\langle 1\rangle$ *determined by fuzzy measures over* $\mathcal{S}$ is an unlimited $\mathbf{L}$-fuzzy quantifier of the type $\langle 1\rangle$ assigning an $\mathbf{L}$-fuzzy quantifier $Q_{\mathcal{S}(M)}$ determined by the fuzzy measure spaces from $\mathcal{S}(M)$ to each *non-empty* universe M.

Example 5.2. Let M be a non-empty universe and $\mathcal{S}(M)_i = \{(M,\mathcal{F}_{\mathbf{L}}(M),\mu_i)\}$, where, for $i = 1, 2$,

$$\mu_1(A) = \begin{cases} \bot, & \text{if } A = 1_\emptyset, \\ \top, & \text{otherwise} \end{cases} \tag{31}$$

and

$$\mu_2(A) = \begin{cases} \top, & \text{if } A = 1_M, \\ \bot, & \text{otherwise.} \end{cases} \tag{32}$$

If Q is determined by fuzzy measures from $Q_{\mathcal{S}(M)_1}$ for all $M \neq \emptyset$, then $Q = \mathsf{some}$. In fact, if $M \neq \emptyset$ and $A \in \mathcal{F}_{\mathbf{L}}(M)$, then

$$Q_M(A) = Q_{\mathcal{S}(M)_1}(A) = \int_{(M,\mathcal{F}_{\mathbf{L}}(M))}^{\otimes} A\, d\mu_1 =$$
$$= \bigvee_{m\in M} \mu_1(\{m\}) \otimes A(m) =$$
$$= \bigvee_{m\in M} \top \otimes A(m) = \bigvee_{m\in M} A(m) = (\mathsf{some})_M(A).$$

According to Definition 4.3, $Q = \mathsf{some}$. One checks easily that Q determined by $Q_{\mathcal{S}(M)_2}$ for all $M \neq \emptyset$ is equal to all.

Example 5.3. Let $\mathbf{L}$ be the standard Gödel algebra and

$$\mathcal{S}(M) = \{(M,\mathcal{F}_{\mathbf{L}}(M),\mu)\}.$$

One checks easily, using Theorem 3.1, that

$$Q_{\mathcal{S}(M)}(A) = \bigvee_{1_Y\in\mathcal{P}_{\mathbf{L}}(M)\setminus\{1_\emptyset\}} \Big(\mu(1_Y) \wedge \bigwedge_{m\in Y} A(m)\Big),$$

because $\otimes$ coincides with $\wedge$ in Gödel algebras. Notice that because $\mathbf{L}$ is the Gödel algebra, it is also a Heyting algebra and, according to Theorem 3.4, the $\otimes$-fuzzy integral coincides with the Sugeno integral in this case.

Let M be a non-empty countable universe and μ_f denote one of the fuzzy measures on $(M,\mathcal{F}_{\mathbf{L}}(M))$ defined by (6) and (7) in Example 3.1. Putting

$$Q_{\mathcal{S}(M)}(A) = \bigvee_{1_Y\in\mathcal{P}_{\mathbf{L}}(M)\setminus\{1_\emptyset\}} \Big(\mu_f(1_Y) \wedge \bigwedge_{m\in Y} A(m)\Big)$$

for any non-empty countable universe M and $Q_\emptyset(1_\emptyset) = \top$, we obtain a countable $\mathbf{L}$-fuzzy quantifier which is an interpretation of the quantifier *many things*. Define

$$\mu_f^{1/2}(A) = \begin{cases} \top, & \text{if } \mu_f(A) \geq \frac{1}{2}, \\ \bot, & \text{otherwise,} \end{cases}$$

for any $A \in \mathcal{F}_{\mathbf{L}}(M)$. Then putting

$$Q_{\mathcal{S}(M)}(A) = \bigvee_{1_Y\in\mathcal{P}_{\mathbf{L}}(M)\setminus\{1_\emptyset\}} \Big(\mu_f^{1/2}(1_Y) \wedge \bigwedge_{m\in Y} A(m)\Big)$$

for any non-empty countable universe M and $Q_\emptyset(1_\emptyset) = \top$, we obtain a countable $\mathbf{L}$-fuzzy quantifier which is an interpretation of the quantifier *at least half things*. If we restrict ourselves to the class of all finite $\mathbf{L}$-fuzzy quantifiers, then one checks easily (using the equality $\mu_f(A) = \mu_f(h^{\rightarrow}(A))$ from Example 3.1) that both defined quantifiers are PI and ISOM. Moreover, they are $\mathrm{EXTENS}(\approx^{\wedge})$ (see Theorem 5.6).

Intuitively, it is obvious that all is the smallest and some the greatest $\mathbf{L}$-fuzzy quantifier Q determined by fuzzy measures with respect to the ordering from Definition 4.3.

Theorem 5.1. *For each* $\mathbf{L}$*-fuzzy quantifier* Q *determined by fuzzy measures over* $\mathcal{S}$*, it holds that*

$$\mathsf{all} \leq Q \leq \mathsf{some}. \tag{33}$$

In the following part, we will present some results on the semantic properties of $\mathbf{L}$-fuzzy quantifiers determined by fuzzy measures. The following theorem states a sufficient condition for $\mathbf{L}$-fuzzy quantifiers to be permutation invariant.

Theorem 5.2. *Let* Q *be an unlimited* $\mathbf{L}$*-fuzzy quantifier of the type* $\langle 1\rangle$ *determined by fuzzy measures over* $\mathcal{S}$ *such that for each non-empty universe* M *it holds that* $\mathcal{S}(M) = [(M,\mathcal{M},\mu)]$*. Then* $Q \in \mathrm{PI}$.

Note that the specification of a necessary condition for $\mathbf{L}$-fuzzy quantifiers being permutation invariant seems to be immensely complicated and it is still an open problem. In the following theorem, let us denote fuzzy measure spaces $(M,\mathcal{M},\mu)$ and $(M',\mathcal{M}',\mu')$ by $\mathbf{M}$ and $\mathbf{M}'$, respectively.

ISBN: 978-989-95079-6-8

Theorem 5.3. *Let Q be an unlimited* $\mathbf{L}$*-fuzzy quantifier of the type $\langle 1\rangle$ determined by fuzzy measures over $\mathcal{S}$ such that, for any universes M, M' with the same cardinality,*

(i) if $\mathbf{M} \in \mathcal{S}(M)$ and $f : M \to M'$ is a bijection, then $f^{\to}(\mathbf{M}) \in \mathcal{S}(M')$,

(ii) if $\mathbf{M} \in \mathcal{S}(M)$ and $\mathbf{M}' \in \mathcal{S}(M')$, then $\mathbf{M}$ and $\mathbf{M}'$ are isomorphic.

Then $Q \in$ ISOM.

From the definition of $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1\rangle$ determined by fuzzy measures it should be clear that they possess non-decreasing behavior, which is expressed by the following theorem. Naturally, if we base $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1\rangle$ on complementary fuzzy measures and $\to$-fuzzy integrals [21], they will be non-increasing.

Theorem 5.4. *Let Q be an unlimited* $\mathbf{L}$*-fuzzy quantifier of the type $\langle 1\rangle$ determined by fuzzy measures over $\mathcal{S}$. Then Q is a non-decreasing* $\mathbf{L}$*-fuzzy quantifier.*

The following theorem shows that some is the only type $\langle 1\rangle$ quantifier with the extension (EXT) property which can be successfully modeled by our $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1\rangle$ determined by fuzzy measures. However, the quantifier all and quantifiers which are interesting from the point of view of fuzzy logic, for example "at least half things", "many things", "most things" do not possess the extension property, and we can model them, see Examples 5.2 and 5.3. Nevertheless, quantifiers which refer to absolute cardinalities, e.g. "at least three things", possess the extension property, therefore they cannot be successfully modeled by our quantifiers.

Theorem 5.5. *Let Q be an unlimited* $\mathbf{L}$*-fuzzy quantifier of the type $\langle 1\rangle$ determined by fuzzy measures over $\mathcal{S}$. Then $Q \in$* EXT *if and only if $Q \equiv$* some.

Finally, we show that fuzzy quantifiers of type $\langle 1\rangle$ determined by fuzzy measures are behaving well with respect to extensionality (see Definition 4.10).

Theorem 5.6. *Let Q be an unlimited* $\mathbf{L}$*-fuzzy quantifier of the type $\langle 1\rangle$ determined by fuzzy measures over $\mathcal{S}$ such that $\mathcal{S}(M) = [(M, \mathcal{M}, \mu)]$ for each non-empty universe M. Then $Q \in$* EXTENS($\approx^{\wedge}$).

6 Conclusion

In the future we will concentrate on studying of $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1, 1\rangle$ (and possibly also of the type $\langle 1^n, 1\rangle$) generated by fuzzy measures. Quantifiers of the type $\langle 1, 1\rangle$ serve as models of very important class of natural language determiners (cf. e.g. [3]), for example "a few X are Y", "almost all X are Y", etc. Our definitions of $\otimes$-fuzzy integral and $\to$-fuzzy integral allow us to define these quantifiers, and we believe that they provide an important class of models with interesting properties.

Acknowledgment

This paper has been supported by the Grant IAA108270901 of the GA AV ČR.

References

[1] I. Glöckner. *Fuzzy Quantifiers in Natural Language: Semantics and Computational Models*. Der Andere Verlag, Osnabrück, Germany, 2004.

[2] M. Holčapek. Monadic $\mathbf{L}$-fuzzy quantifiers of the type $\langle 1^n, 1\rangle$. *Fuzzy Sets and Systems*, 159:1811–1835, 2008.

[3] S. Peters and D. Westerståhl. *Quantifiers in Language and Logic*. Oxford University Press, New York, 2006.

[4] A. Mostowski. On a generalization of quantifiers. *Fund. Math.*, 44:12–36, 1957.

[5] P. Lindström. First ordered predicate logic with generalized quantifiers. *Theoria*, 32:186–195, 1966.

[6] J. Barwise and R. Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219, 1981.

[7] L.A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Comp. Math. with Applications*, 9:149–184, 1983.

[8] H. Thiele. On fuzzy quantifiers. In Z. Bien and K.C. Min, editors, *Fuzzy Logic and its Applications, Information Sciences, and Intelligent Systems*, pages 343–352. Kluwer Academic Publisher, Dordrecht, 1996.

[9] D. Ralescu. Cardinality, quantifiers and the aggregation of fuzzy criteria. *Fuzzy Sets and Systems*, 69:355–365, 1995.

[10] Y. Liu and E.E. Kerre. An overview of fuzzy quantfiers. part I: Interpretation. *Fuzzy Sets and Systems*, 95(1):1–21, 1998.

[11] Y. Liu and E.E. Kerre. An overview of fuzzy quantfiers. part II: Reasoning and application. *Fuzzy Sets and Systems*, 95(2):135–146, 1998.

[12] V. Novák. Antonyms and linguistic quantifiers in fuzzy logic. *Fuzzy Sets and Systems*, 124(3):335–351, 2001.

[13] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, Dordrecht, 1998.

[14] I. Glöckner. *Fuzzy Quantifiers: A Computational Theory*. Springer-Verlag, Berlin, 2006.

[15] V. Novák. A formal theory of intermediate quantifiers. *Fuzzy Sets and Systems*, 159(10):1229 1246, 2008.

[16] V. Novák. On fuzzy type theory. *Fuzzy Sets and Systems*, 149:235–273, 2005.

[17] M. Ying. Linguistic quantifiers modeled by Sugeno integrals. *Artificial Intelligence*, 170:581–606, 2006.

[18] Z. Wang and G.J. Klir. *Fuzzy measure theory*. Plenum Press, New York, 1992.

[19] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy Measures and Integrals. Theory and Applications*. Studies in Fuzziness and Soft Computing. Physica Verlag, Heidelberg, 2000.

[20] E.P. Klement, R. Mesiar, and E. Pap. A universal integral based on measures of level sets. *IEEE Transactions on Fuzzy Systems*, to appear.

[21] A. Dvořák and M Holčapek. Fuzzy integral over complete residuated lattices. In *IFSA/EUSFLAT 2009*, Lisboa, 2009.

[22] R. Bělohlávek. *Fuzzy Relational Systems: Foundations and Principles*. Kluwer Academic Publisher, New York, 2002.

ISBN: 978-989-95079-6-8

On the Definition of Extended Norms and Co-norms to Aggregate Fuzzy Bipolar Conditions

Ludovic Liétard[1] Daniel Rocacher[2]

1 IRISA/IUT, rue Edouard Branly, BP 30219,
22302 Lannion Cedex, France
2 IRISA/ENSSAT, rue de Kérampont, BP 447,
22305 Lannion Cedex, France
Email: ludovic.lietard@univ-rennes1.fr, rocacher@enssat.fr

Abstract— *Fuzzy queries addressed to a relational database management system are based on the expression of user preferences which leads to obtain more appropriate answers. Fuzzy set theory provides powerful tools to define such conditions (fuzzy conditions). In this context, it is possible to consider fuzzy bipolar conditions which are made of two fuzzy sets to describe user's preferences: i) a fuzzy constraint to represent a mandatory requirement, ii) a fuzzy wish to represent an optional requirement. A Fuzzy constraint is mandatory in the sense that it* ***must*** *be fulfilled (and it can be used to discard elements). A fuzzy wish is optional since it cannot be used to discard elements (but its satisfaction brings a bonification). This paper is devoted to the aggregation of fuzzy bipolar conditions and to the definition of extended norms and co-norms.*

Keywords— *flexible querying, fuzzy bipolar condition, relational database.*

1 Introduction

When querying a relational database management system (DBMS), the expression of user preferences is a manner to integrate the personalized needs of the user. As a consequence, it leads to obtain more appropriate answers, since they can be ranked from the more relevant (the more preferred) to the less relevant (a set of discriminated answers being obtained). The problem of expressing and managing user preferences in the relational model of data has received more and more attention in the last few years [1, 2, 3, 4, 5, 6, 7] and we consider the study of bipolar conditions [8, 9, 10, 11, 12]. Such conditions are made of two components: a mandatory condition (also called a constraint) and an optional condition (also called a wish). More precisely, we consider fuzzy bipolar conditions where the constraint and the wish express user preferences and are defined by fuzzy sets.

The main interrogation systems proposed to integrate user preferences do not take into account such bipolar conditions. The SQLf language considers preferences defined by fuzzy sets and proposes a commensurable approach without bipolarity. It does not distinguish between constraints and wishes, each atomic fuzzy condition being a commensurable constraint. On the contrary, PreferenceSQL deals with a particular case of bipolarity where constraints are Boolean while wishes are not commensurable preferences. The preferences being not commensurable in PreferenceSQL, this proposition leads to a partial order of answers while SQLf language delivers a total order (the preferences being commensurable). One of the main differences [10] between these two approaches is that SQLf is set in the algebraic framework while it is not the case for PreferenceSQL. As a consequence, this article is a contribution to the extension of the relational algebra (and SQL) to fuzzy bipolar conditions. Our final aim is to propose an SQL-like query language where fuzzy bipolar conditions can be defined as an extension of fuzzy conditions. As a consequence, a total order of answers is expected.

Fuzzy bipolar conditions are introduced in section 2 and it is shown that a fuzzy condition defined by a fuzzy set is a particular case of a fuzzy bipolar condition. Section 3 is interested in extending logical operators to define more complex fuzzy bipolar conditions. These operators are soundness from an user's point of view but they cannot be considered as extended norms and co-norms. Section 4 proposes definition for an extended norm and an extended co-norm to aggregate fuzzy bipolar conditions.

2 Fuzzy bipolar conditions

Bipolarity can be defined as the human capability to evaluate situations in terms of pros and cons or positive and negative aspects. Depending of the relationship between these two aspects, Dubois and Prade [8, 9] distinguish between three categories of bipolarity, called types I, II and III.

The type I is called *symmetric univariate bipolarity*. It corresponds to the case where the scale to evaluate situations is bipolar which means defined by an ordered set (total order) with a neutral elements to separate positive and negative information (*symmetric bipolarity*). As a consequence a single value (*univariate*) from this scale expresses, at the same time the positive and negative aspects since the opposite evaluation is obtained by symmetry. An example of type I bipolarity is provided by probability measures, the neutral element being 0.5. The type II is called *symmetric bivariate unipolarity*. In this case, the scale to evaluate situations is unipolar which means that the neutral element is at one end of the scale and two dual values (symmetric bivariate) is necessary to judge a situation. Unlike the symmetric univariate bipolarity, it is possible for a situation to be evaluated positive and negative at the same time. An example of type II bipolarity is provided by possibility theory where a couple of dual unipolar indexes (possibility and possibility of opposite event) is used to evaluate a situation. The type III is called *asymmetric bipolarity*. In this case two bipolar scales are used to evaluate a situation, one expressing the positive, the other one the negative aspects. As a consequence, positive and negative evaluations are of different nature and can be provided by different kind of sources.

ISBN: 978-989-95079-6-8

A bipolar condition is made of two components: a *constraint* (a condition denoted C) whose negation defines not acceptable values (negative evaluation) and a *wish* (a condition denoted W) to define desired values (positive evaluation). The satisfaction to the constraint and the wish are based on two different bipolar scales, the two extremum being the total satisfaction 1 and the total non satisfaction 0. In addition, constraints and wishes are neither dual nor symmetric. As a consequence, a bipolar condition expresses a type III bipolarity.

This paper considers that a bipolar condition denoted (C, W) means "to satisfy C and, if possible, to satisfy W" and this meaning has two important consequences. Firstly, it implies that the satisfaction with respect to the constraint is mandatory while the satisfaction with respect to the wish is optional. Secondly, the set of wished values must be a subset of that of mandatory values (W $\subseteq$ C). This property states that it is incoherent to wish non-acceptable values.

In the context of database querying with such conditions, tuples which do not satisfy the constraint are discarded (since this condition is mandatory), while tuples which do not satisfy the wish can be proposed to the user (the wish being optional). When C and W are Boolean conditions, the satisfaction with respect to the bipolar condition (C,W) is a couple (c, w) where c and w are values in {0, 1}. Tuples satisfying the constraint and the wish are returned in priority to the user. If such answers do not exist, the tuples satisfying only the constraint are delivered. When the two components express preferences and are defined by fuzzy sets, their satisfactions are sets in [0,1]. As a consequence, the satisfaction of a given element x with respect to a fuzzy bipolar condition is a couple of degrees ($\mu_C(x)$, $\mu_W(x)$), and their processing is more tricky. These two degrees are not commensurable (due to the chosen meaning for the fuzzy bipolar conditions) and, consequently, they cannot be aggregated to compute an overall degree of satisfaction. In addition, the constraint being mandatory, its satisfaction is firstly used to discriminate among answers. The satisfactions with respect to the wish being not mandatory, they can only be used to discriminate among answers having same evaluation with respect to the constraint. Thus, the wishes allow to differentiate between tuples which are equal with respect to their constraints and a *total order* is obtained on C and W (with (1, 1) as greatest element and (0, 0) as least element). In other words, it is possible to rank the tuples by using a *lexicographical* order on the constraints and the wishes:

a tuple t_i is preferred to a tuple t_j
if ($\mu_C(t_i) > \mu_C(t_j)$)
or (($\mu_C(t_i) = \mu_C(t_j)$) and ($\mu_W(t_i) > \mu_W(t_j)$)),

Obviously, a tuple t_i is similar to a tuple t_j (they have same level of preference) if and only if $\mu_C(t_i) = \mu_C(t_j)$ and $\mu_W(t_i) = \mu_W(t_j)$.

For the sake of simplicity, "t_i is preferred to t_j" is denoted $t_i > t_j$ while "t_i is similar to t_j" is denoted $t_i = t_j$.

Example 1. We consider relation Sales from Table 1 where attributes #Sale, Date, Town, Benefit are respectively the identification of a sale, the date and place where it has been concluded and the percentage of benefit for the salesman.

A salesman wants to know the town where a sale has occurred with a *high* benefit and, if possible, a benefit *around 100%*. A *high* benefit is a mandatory condition while condition *around 100%* is optional. These two fuzzy conditions are given by Figure 1 and define the fuzzy bipolar condition (C, W) where C is "*high* benefit" and W is "benefit *around 100%*".

Table 1 : Relation Sales.

#Sale	Date	Town	Benefit
#1	01-05-07	Paris	82%
#2	04-06-07	Paris	85%
#3	08-09-07	Londres	59%
#4	15-06-07	Nice	94%
#5	17-07-07	Londres	97%
#6	08-09-07	Madrid	56%

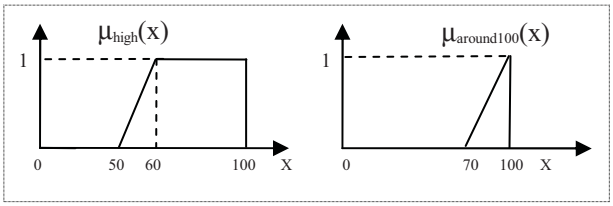


Figure 1 : *high* benefit and benefit *around 100%*.

The satisfactions with respect to the constraint C and the wish W are given by Table 2.

Table 2 : Satisfactions with respect to the fuzzy bipolar condition : "*high* benefit" and if possible "*around 100%*".

#Sale	*(high* benefit, benefit *around 100%*) (μ_C(#Sale), μ_W(#Sale)),
#1	(1, 0.4)
#2	(1, 0.5)
#3	(0.9, 0)
#4	(1, 0.8)
#5	(1, 0.9)
#6	(0.6, 0)

The lexicographical order gives :

#5 > #4 > #2 > #1 > #3 > #6.

Sale #5 is the best sale with respect to the bipolar condition (C, W) because, among the best sales according to the constraint (at degree 1), it is the best one according to the wish. ◆

The property of constraints and wishes expressed on a same universe is that the set of wished values must be a subset of that of mandatory values (W $\subseteq$ C). In case of fuzzy sets, this property of inclusion is rewritten :

for each value x in X, $\mu_W(x) \leq \mu_C(x)$,

ISBN: 978-989-95079-6-8

such that X is the domain set where the fuzzy bipolar condition applies. Such a property can be checked on the fuzzy bipolar condition provided by example 1. This property is natural and is always implicit in the formulation of fuzzy bipolar conditions. As a consequence, a condition formulated as "a *cheap* and, if possible, a *low* consuming car" should be rewritten : "a *cheap* car and, if possible, a *cheap* and *low* consuming car".
An important particular case arises when a fuzzy bipolar condition is a fuzzy predicate (a pure constraint). Since a fuzzy predicate C can be rewritten "C and, if possible, C", a pure constraint is defined by a fuzzy bipolar condition where the constraint equals the wish ((C, C)). As a consequence, a fuzzy bipolar condition is a generalization of a fuzzy condition.

3 Logical operators and fuzzy bipolar conditions

In this section, we consider several fuzzy bipolar conditions and their aggregation using conjunction and disjunction operators.
A first idea to define the conjunction of two fuzzy bipolar conditions (C_1, W_1) and (C_2, W_2) is to compute the minimum of both the satisfactions to the constraints and to the wishes. In this case "(C_1, W_1) and (C_2, W_2)" is the fuzzy bipolar condition "to satisfy (C_1 and C_2) and if possible to satisfy (W_1 and W_2)". Since the minimum is a triangular norm, this operation can be generalized to define parameterized operators $O_{op1,op2}$ using various norms and conorms :

$$(C_1,W_1)\ O_{op1,\,op2}\ (C_2,W_2) \equiv (op_1(C_1, C_2), op_2(W_1,W_2)), \quad (1)$$

where op_1 and op_2 are either a norm or a co-norm. Depending on the choice for op_1 and op_2 , different meanings can be obtained for the operator (as shown by Table 3).

Table 3 : Different meanings for $O_{op1,op2}$.

	op_1 is a norm	op_1 is a co-norm
op_2 is a norm	$(C_1,W_1)\ O_{op1,op2}\ (C_2,W_2)$ ≡ (C_1 and C_2) and if possible (W_1 and W_2)	$(C_1,W_1)\ O_{op1,\,op2}\ (C_2,W_2)$ ≡ (C_1 or C_2) and if possible (W_1 and W_2)
op_2 is a co-norm	$(C_1,W_1)\ O_{op1,op2}\ (C_2, W_2)$ ≡ (C_1 and C_2) and if possible (W_1 or W_2)	$(C_1,W_1)\ O_{op1,op2}\ (C_2,W_2)$ ≡ (C_1 or C_2) and if possible (W_1 or W_2)

Example 2. A client looks for a trip in a travel agency. His request is expressed as: "a *cheap* journey, if possible with a *small* flight time, and an hotel in a *small* town, if possible located *in the countryside*". This query is a selection R:Θ where relation R describes trips proposed by the agency and condition Θ is the conjunction between two fuzzy bipolar conditions "(C_1,W_1) and (C_2,W_2)". Constraint C_1 is the fuzzy condition "a *cheap* journey", W_1 is the fuzzy wish "a *cheap* journey with a *small* flight time" and C_2 is the fuzzy constraint "hotel in a *small* town" while W_2 is the fuzzy wish "hotel in a *small* town and *in the countryside*". Each one of these two atomic bipolar fuzzy conditions is defined on several attributes (on the Cartesian product of their respective domains). The different satisfactions with respect to the constraints and wishes are given by Table 4. Obviously, the result depends on the chosen operator to represent the conjunction. We propose the two different conjunction operators $O_{min,\,min}$ and $O_{min,\,max}$.

Table 4 : Satisfactions with respect to (C_1,W_1) and (C_2,W_2)

#Trip	C_1 "a *cheap* journey"	W_1 "a *cheap* journey and a *small* flight time"	C_2 "*small* town"	W_2 "a *small* town and *in the countryside*"
#1	0.9	0.7	1	0.1
#2	1	0.5	0.9	0.5
#3	0.9	0.3	1	1
#4	0.7	0.6	0.6	0.5
#5	0.8	0.2	0.6	0.5

When the operator $O_{min,\,min}$ is chosen, the meaning of the query becomes:

"a *cheap* journey with an hotel in a *small* town, if possible with a *small* flight time ***and*** an hotel located *in the countryside*".

The results are given by Table 5.

Table 5 : Results provided by $O_{min,\,min}$

#Trip	μ_C(#Trip)	μ_W(#Trip)
#2	0.9	0.5
#3	0.9	0.3
#1	0.9	0.1
#4	0.6	0.5
#5	0.6	0.2

We obtain the order : #2 > #3 > #1 > #4 > #5.

When the operator $O_{min,\,max}$ is chosen, the meaning of the query becomes:

"a *cheap* journey with an hotel in a *small* town, if possible with a *small* flight time ***or*** an hotel located *in the countryside*".

The results are given by Table 6:

Table 6 : Results provided by $O_{min,\,max}$

#Trip	μ_C(#Trip)	μ_W(#Trip)
#3	0.9	1
#1	0.9	0.7
#2	0.9	0.5
#4	0.6	0.6
#5	0.6	0.5

ISBN: 978-989-95079-6-8

We obtain the order : #3 > #1 > #2 > #4 > #5.

The obtained order is different from the ones provided by $O_{min, min}$ (the meanings of the operator being different).
This example shows that the two different operators considered here lead to two different results since they have two different meanings ($O_{min, min}$ means "*all* the constraints and *all* the wishes", while $O_{min, max}$ means "*all* the constraints and *at least one* wish"). ♦

Last example has shown that, when $O_{min, min}$ and $O_{min, max}$ are used, the fundamental property of constraints and wishes ($W \subseteq C$) is lost (since the result are based on ad-hoc operators). However, we think that this loss is not an issue since it reflects the user's attitude with respect to the aggregation of wishes (either all, either at least one). However, it stresses the need for other operators (as extended norms and co-norms).

4 Extended norms and co-norms

This section is interested in defining operators acting as extended norms and co-norms for fuzzy bipolar conditions. Subsection 4.1 shows that it is not possible to define an extended norm (and an extended co-norm) using any operator $O_{op1, op2}$ where op_1 and op_2 are either a norm or a co-norm (cf. section 3). Section 4.2 defines a couple of extended norm and co-norm (and their associated negation).

4.1 Extended norms, co-norms and $O_{op1, op2}$

The different properties bring by norms and co-norms are necessary conditions to keep the algebraic framework of computations (or at least to achieve computations in an human-consistent framework). The operator $O_{min,min}$ is a rather rational candidate for an extended norm but next example shows that it does not respect the monotonicity property of norms and it concludes that this loss is very damageable.

Example 3. The property of monotonicity of a given norm $\wedge$ is defined by:

$$a \geq b \text{ and } c \geq d \Rightarrow (a \wedge c) \geq (b \wedge d)$$

The following counter example shows it is not respected by $O_{min,min}$:

Let x and y be two elements and two fuzzy bipolar conditions FBC1 and FBC2. The respective satisfaction of x and y to these two fuzzy bipolar conditions are shown by Figure 2.

	FBC1	FBC2
x	(0.7, 0.7)	(0.9, 0.4)
y	(0.7, 0.6)	(0.8, 0.5)

Figure 2 : Satisfactions of x and y to FBC1, FBC2.

The property of monotonicity does not hold in this case since:

$$(0.7, 0.7) \geq (0.7, 0.6) \text{ and } (0.9 , 0.4) \geq (0.8, 0.5)$$

while :

$$(0.7, 0.7)\ O_{min,min}\ (0.9, 0.4)$$

$$< (0.7, 0.6)\ O_{min,min}\ (0.8, 0.5).$$

This behaviour is not acceptable for a conjunction since x is preferred to y when considering criteria FBC1 and when considering criteria FBC2, but x is not preferred to y when considering the conjunction "FBC1 and FBC2". ♦

More generally, an extended norm (or co-norm) cannot be defined by any operator $O_{op1,op2}$ as demonstrated hereafter (the demonstration is only provided for the norm and it can be easily adapted for a co-norm).

Demonstration. We assume an extended norm defined as $O_{op1,op2}$:

$$(C_1,W_1)\ O_{op1, op2}\ (C_2,W_2) \equiv (op_1(C_1, C_2), op_2(W_1,W_2)),$$

where op_1 and op_2 are either a norm or a co-norm. The demonstration is based on two steps. At first, two properties which are necessarily satisfied by such an operator are pointed out. Secondly, it is shown that these properties are not compatible with an extended norm since the property of monotonicity does not hold.

First step: Since a fuzzy condition (a fuzzy set) is a particular case of a fuzzy bipolar condition, an extended norm for fuzzy bipolar conditions should revert to a norm in case of a fuzzy set. A fuzzy set being defined as a fuzzy bipolar condition where the constraint equals the wish, the application of an extended norm $O_{op1,op2}$ can be rewritten (in case of fuzzy sets):

$$(C_1,C_1)\ O_{op1, op2}\ (C_2,C_2) \equiv (op_1(C_1, C_2), op_2(C_1,C_2)),$$

In addition $(op_1(C_1, C_2), op_2(C_1,C_2))$ is nothing but the expression of a norm $\wedge$ between C_1 and C_2 . Since $\wedge (C_1, C_2)$ can be represented by $(\wedge (C_1, C_2), \wedge (C_1,C_2))$ we immediately get two properties:

P1) op1 equals op2,
P2) op1 is a norm (denoted $\wedge$).

Second step: Due to the previous result, any $O_{op1,op2}$ operator defining an extended norm is rewritten $O_{\wedge,\wedge}$ where $\wedge$ is a norm. As a consequence, the property of monotonicity does not hold since:

$$(0, 1) \geq (0, 1)$$

and $(1, 0) \geq (0, 1)$

while:

$$(0, 1)\ O_{\wedge,\wedge}\ (1, 0) < (0, 1)\ O_{\wedge,\wedge}\ (0, 1)$$

(since $(0, 1)\ O_{\wedge,\wedge}\ (1, 0) = (0 \wedge 1, 1 \wedge 0) = (0, 0)$
and $(0, 1)\ O_{\wedge,\wedge}\ (0, 1) = (0 \wedge 0, 1 \wedge 1) = (0, 1)$). ♦

ISBN: 978-989-95079-6-8

This subsection has shown that it is necessary to define operators satisfying the properties of norms and co-norms (we recall that the total order considered here is the lexicographical order with (1, 1) as greatest element and (0, 0) as least element).

4.2 Definition for a couple of extended norm and co-norm

An extended norm $\wedge$ and an extended co-norm $\vee$ can be defined by the minimum and maximum with respect to the lexicographical order (which provides a total order):

$$(x, y) \wedge (x', y') = \min((x, y), (x', y')), \quad (2)$$

$$(x, y) \vee (x', y') = \max((x, y), (x', y')). \quad (3)$$

Due to the total order, it is easy to prove that $\wedge$ and $\vee$ are respectively an extended norm and an extended co-norm.

Similarly, a negation operator $\neg$ can be defined by:

$$\neg(x, y) = (1-x, 1-y). \quad (4)$$

Next proof shows that the De Morgan laws are preserved:

$$\neg(x, y) \vee \neg(x', y') = \neg[(x, y) \wedge (x', y')]$$

$$\neg(x, y) \wedge \neg(x', y') = \neg[(x, y) \vee (x', y')].$$

Proof. We consider $(x, y) \wedge (x', y') = \min((x, y), (x', y')) = (x, y)$. The properties to demonstrate obviously hold when ($x = x'$) and ($y = y'$). When ($x \neq x'$) or ($y \neq y'$), we get ($x < x'$ or (($x = x'$) and ($y < y'$)). The complementation to one gives : ($1-x > 1-x'$) or ((1- x = 1-x') and ($1-y > 1-y'$)). As a consequence $(1 - x, 1 -y) \vee (1-x', 1-y') = (1-x, 1-y)$ which can be rewritten $\neg(x, y) \vee \neg(x', y') = \neg(x, y)$. Then we get $\neg(x, y) \vee \neg(x', y') = \neg(x, y) = \neg[(x, y) \wedge (x', y')]$. The first law is then demonstrated and the second one is valid .♦

Example 4. We consider the framework and the query of example 2 ("a *cheap* journey, if possible with a *small* flight time, and an hotel in a *small* town, if possible located *in the countryside*). Constraint C_1 is the fuzzy condition "a *cheap* journey", W_1 is the fuzzy wish "a *cheap* journey with a *small* flight time" and C_2 is the fuzzy constraint "hotel in a *small* town" while W_2 is the fuzzy wish "hotel in a *small* town and *in the countryside*".
When the norm $\wedge$ is chosen, the conjunction returns the worst case between (C_1,W_1) and (C_2,W_2) and the meaning of the query is a pure conjunction (the worst evaluation being retained). We get the results provided by Table 7.

Table 7 : Results provided by norm $\wedge$

#Trip	μ_C(#Trip)	μ_W(#Trip)
#1	0.9	0.7
#2	0.9	0.5
#3	0.9	0.3
#4	0.6	0.5
#5	0.6	0.5

We obtain the order : #1 > #2 > #3 > #4 = #5.

Here again, the obtained order is different from the ones provided by $O_{min,\ max}$ and $O_{min,\ min}$ (the meanings of the operator being different, cf example 3). In addition, one may remark that the fundamental property of fuzzy wish and constraint ($W \subseteq C$) holds in this case (since $\mu_W(\#Trip) \leq \mu_C(\#Trip)$ for any trip). ♦

5 Particular case of fuzzy conditions

This subsection shows the behaviours of the operators defined in the previous sections in the particular case where the fuzzy bipolar conditions are pure constraints ($W = C$).
When dealing with pure constraints, the extended norm $\wedge$ and co-norm $\vee$ are nothing but the norm min and the co-norm max of fuzzy sets.

Proof. We consider $(x, x) \wedge (y, y) = (x, x)$. It means (x, x) is the minimum when the lexicographical order is used which can be rewritten ($x < y$) or ($x = y$) and then:

$$(x, x) \wedge (y, y) = (x, x) = (\min(x, y), \min(x, y)).$$

The expression $(\min(x, y), \min(x, y))$ is the representation (in terms of constraints and wishes) of the minimum of the two degrees x and y which demonstrate that $(x, x) \wedge (y, y)$ represents the norm min. A similar proof holds for $\vee$ and the co-norm max.♦

The extended norm $\wedge$ and co-norm $\vee$ can also be defined by the parameterized operator $O_{op1,op2}$ where op_1 equals op_2 and is either the minimum or the maximum (it is not the case in general since $\wedge$ and $\vee$ cannot be expressed by $O_{op1,op2}$, as example $(0.6, 0.8) \wedge (0.5, 1) = (0.5, 1)$ while $(0.6, 0.8)\ O_{min,\ min}\ (0.5, 1) = (0.5, 0.8)$).

Proof. We consider $(x, x) \wedge (y, y) = (x, x)$ and $(x, x) \vee (y, y) = (y, y)$. It means ($x < y$) or ($x = y$) and then :

$$(x, x) \wedge (y, y) = (x, x) = (\min(x, y), \min(x, y))$$
$$= (x, x)\ O_{min,\ min}\ (y, y),$$
$$(x, x) \vee (y, y) = (y, y) = (\max(x, y), \max(x, y))$$
$$= (x, x)\ O_{max,\ max}\ (y, y).\ ♦$$

More generally, when op_1 equals op_2 the parameterized operator $O_{op1,op2}$ is either a norm or a co-norm between two fuzzy sets (the one chosen to implement op_1 and op_2).
The main question is about the meaning of $O_{op1,op2}$ when op_1 and op_2 differs. First of all, for a sake of simplicity, we assume that op_1 is a norm. Since the lexicographical order is used, it is possible to claim that $O_{op1,op2}$ allows for a refinement of norm op_1 since results are firstly ranked on the degree provided by op1, op_2 being only used to discriminate among elements having same degrees for the norm op_1. Similarly, when op1 is a co-norm, the lexicographical order provides a refinement of this co-norm.

ISBN: 978-989-95079-6-8

Example 5. We consider four items x,y,z and t which are valuated using a conjunction between the satisfactions of two fuzzy bipolar conditions FBC1 and FBC2 (cf. Table 8).

Table 8 : Satisfactions with respect to FBC1 and FBC2

item	FBC1	FBC2
x	(0.7, 0.7)	(0.8,0.8)
y	(0.5, 0.5)	(1, 1)
z	(0.5, 0.5)	(0.8, 0.8)
t	(0.3, 0.3)	(1, 1)

The satisfactions with respect to the conjunction are given by Table 9.

Table 9 : Satisfactions with respect to the conjunction

item	FBC1 $O_{min, min}$ FBC2	FBC1 $O_{min, max}$ FBC2
x	(0.7, 0.7)	(0.7,0.8)
y	(0.5, 0.5)	(0.5, 1)
z	(0.5, 0.5)	(0.5, 0.8)
t	(0.3, 0.3)	(0.3, 1)

When the minimum ($O_{min, min}$) is considered we get $x > y = z > t$, while $O_{min, max}$ gives $x > y > z > t$. The operator $O_{min, max}$ allows to discriminate elements y and z which are undistinguishable when $O_{min, min}$ is considered. ♦

6 Conclusion

This paper has dealt with the evaluation of fuzzy bipolar conditions into fuzzy queries addressed to a regular relational database (where the stored data are precisely known). Such conditions are made of two components: a mandatory condition (also called a constraint) and an optional condition (also called a wish). More precisely, we have considered fuzzy bipolar conditions where the constraint and the wish express user preferences and are defined by fuzzy sets. In particular, we have shown that a fuzzy bipolar condition is a generalization of a fuzzy condition (defined by a fuzzy set) and several operators to aggregate fuzzy bipolar conditions have been defined.

This paper has pointed out the importance to define an extended norm and an extended co-norm for fuzzy bipolar conditions. It proposes a couple of extended norm and co-norm which is a generalization of the couple (min, max) of fuzzy sets. In the particular case where fuzzy bipolar conditions are fuzzy conditions, some proposed operators leads to a refinement of norm and conorm as it is the case for the leximin and discrimin orders [13].

References

[1] Agrawal A. and Wimmers E.L. "A Framework for Expressing and Combining Preferences". *In Proc. of the 2000 ACM SIGMOD International Conference on Management of Data,* Dallas, USA, pp. 297–306, 2000.

[2] Bordogna G., Pasi G. "Linguistic aggregation operators of selection criteria in fuzzy information retrieval" In *Int. Jour. of Intelligent Systems*, vol. 10(2), pp.233–248, 1995.

[3] Borzsnyi, S. Kossmann D. and Stoker K. "The Skyline Operator". In Proc. of the 17th International Conference on Data Engineering, (ICDE01), Heidelberg, Germany, pp. 421–430, 2001.

[4] Bosc P., Pivert O. "SQLf: a relational database language for fuzzy querying". In *IEEE Transactions on Fuzzy Systems*, vol(3) pp.1–17, 1995.

[5] Chomicki J. "Preference Formulas in Relational Queries". In ACM Transactions on Database Systems, (TODS'03), 28(4):1–39, 2003.

[6] Kießling W. "Foundations of Preferences in Database Systems". In Proc. of the 28th International Conference on Very Large Data bases, (VLDB), Hong Kong, China, pp. 331–322, 2002.

[7] Kießling W. and Kostler G. "Preference SQL - Design, Implementation, Experiences". In Proc. of the 28th International Conference on Very Large Databases, (VLDB), Hong Kong, China, pp. 990–1001, 2002.

[8] Dubois D. and Prade H. "Handling Bipolar Queries in Fuzzy Information Processing". In Handbook of Research on Fuzzy Information Processing in Databases, J. Galindo (Ed.), 97–114, 2008.

[9] Dubois D. and Prade H. "An Introduction to Bipolar Representations of Information and Preference". In International Journal of Intelligent Systems, vol. 23, 866-877, 2008.

[10] Liétard L., Rocacher D. and Tbahriti S.-E., Preferences and Bipolarity in Query Language, International Conference of the North American Fuzzy Information Processing Society (NAFIPS 2008), New-York, USA.

[11] Zadrozny S. "Bipolar Queries Revisited". In Modeling Decisions for Artificial Intelligence, V. Torra et al. (Eds.), 387-398, 2005.

[12] Zadrozny S. and Kacprzyk J. "Bipolar Queries and Queries with Preferences". In Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA'06), 415-419, 2006.

[13] Dubois D., Fargier H. and Prade H. "Beyond min aggregation in multi-criteria decision: (ordered) weighted Min, Discrimin, Leximin". In The Ordered Weighted Averaging Operators - Theory and Applications. R.R. Yager, J. Kacprzyk (Eds.), Kluwer Academic Publ., Boston, 181–192, 1997.

ISBN: 978-989-95079-6-8

An abstract approach to fuzzy logics: implicational semilinear logics

Petr Cintula[1] Carles Noguera[2]

1. Institute of Computer Science, Czech Academy of Sciences
Prague, Czech Republic
2. Department of Mathematics and Computer Science, University of Siena
Siena, Italy
Email: cintula@cs.cas.cz, cnoguera@iiia.csic.es

Abstract— This paper presents a new abstract framework to deal in a uniform way with the increasing variety of fuzzy logics studied in the literature. By means of notions and techniques from Abstract Algebraic Logic, we perform a study of non-classical logics based on the kind of generalized implication connectives they possess. It yields the new *hierarchy of implicational logics*. In this framework the notion of *implicational semilinear logic* can be naturally introduced as a property of the implication, namely a logic L is an implicational semilinear logic iff it has an implication such that L is complete w.r.t. the matrices where the implication induces a linear order, a property which is typically satisfied by well-known systems of fuzzy logic. The hierarchy of implicational logics is then restricted to the semilinear case obtaining a classification of implicational semilinear logics that encompasses almost all the known examples of fuzzy logics and suggests new directions for research in the field.

Keywords— Abstract Algebraic Logic, Implicative logics, Leibniz Hierarchy, Mathematical Fuzzy Logic, Semilinear logics.

1 Introduction

Mathematical Fuzzy Logic is the subdiscipline of Mathematical Logic which studies the logical systems that, since the inception of the theory of fuzzy sets, have been proposed to deal with the reasoning with predicates that can be modelled by fuzzy sets. The first ones, coming from the many-valued logic tradition, were Łukasiewicz and Gödel-Dummett logics, both complete w.r.t. the semantics given by a continuous t-norm. Later, a third system with this feature was introduced: product logic. Starting from these three main examples, the area has followed a long process of increasing generalization that has led to wider and wider classes of fuzzy logics. The first step was taken by Hájek [14] when he proposed the Basic fuzzy Logic BL, which turn out to be complete w.r.t. the semantics of all continuous t-norms. Later on, to put it in Hájek's words, scholars kept *removing legs from the flea* by considering weaker notions of fuzzy logic: divisibility was removed in the logic MTL [8] which is complete w.r.t. the semantics of all left-continuous t-norms, negation was removed when considering fuzzy logics based on hoops [9], commutativity of t-norms was disregarded in [13], and t-norms were replaced by *uninorms* in [16]. On the other hand, logics with a higher expressive power were introduced by considering expanded real-valued algebras (with projection Δ, involution $\sim$, truth-constants, etc.), and in recent works fuzzy logics have started emancipating from the real-valued algebras as the only intended semantics by considering systems complete w.r.t. rational, finite or hyperreal linearly ordered algebras [5].

When dealing with this huge variety of fuzzy logics one may want to have some tools to prove general results that apply not only to a particular logic, but to a class of logics. To some extent this has been achieved by means of the notions of core and Δ-core fuzzy logics [15] and results for these classes can be already found in a number of papers. However, those classes contain roughly just expansions of MTL and $\mathrm{MTL}_\triangle$ logics, so they do not cover weaker systems such as those from [16]. This shows that general notions of fuzzy logics are very useful, but we need to look for a more abstract framework to cope with all known examples and with other new logics that may arise in the near future.

In doing so, one certainly needs some intuition about the class of objects he would like to mathematically determine, namely some intuition of what are the minimal properties that should be required for a logic to be fuzzy. The evolution outlined above shows that almost no property of these systems was essential as they were step-by-step disregarded. Nevertheless, there is one that has remained untouched so far: *completeness w.r.t. a semantics based on linearly ordered algebras*. It actually corresponds to the main thesis of [1] that defends that *fuzzy logics are the logics of chains*. Such a claim must be read as a methodological statement, pointing at a roughly defined class of logics, rather than a precise mathematical description of what fuzzy logics are, since there could be many different ways in which a logic might enjoy a complete semantics based on chains. The aim of the present paper is to use some notions and techniques from Abstract Algebraic Logic (AAL) to provide a new framework where we can develop in a natural way a particular technical notion corresponding to the intuition of fuzzy logics as the logics of chains. Namely, we will present the *hierarchy of implicational logics* as a new classification of non-classical logics extending the well-known *Leibniz hierarchy* and encompassing other important classes such as implicative logics [18] and weakly implicative logics [4]. Inside this new hierarchy we will build a very general class of fuzzy logics that we will call *implicational semilinear logics*.

The technical aspects of this paper are based on the submitted work [6] however it concentrates on presenting the justification of our new framework from the point of view of Mathematical Fuzzy Logic. Section 2 describes our general setting and Section 3 informally summarizes our arguments. The remaining sections present samples of technical arguments supporting our thesis. Finally, Appendix A recalls crucial preliminary notions from the theory of logical calculi.

ISBN: 978-989-95079-6-8

2 The hierarchy of implicational logics

Although implication in the vast majority of existing fuzzy logics is given by a single (primitive or derived) connective, we follow a long-established tradition of Abstract Algebraic Logic and consider that implication could be given by a (possibly parameterized) set of formulae. However the following convention will allow us to hide this feature of our approach, providing a high level of abstraction without any apparent increase of complexity. Let $\Rightarrow(p, q, \overrightarrow{r})$ be a set of $\mathcal{L}$-formulae in two variables and, possibly, with a sequence of parameters $\overrightarrow{r}$. Given formulae φ, ψ and a sequence of formulae $\overrightarrow{\alpha}$, $\Rightarrow(\varphi, \psi, \overrightarrow{\alpha})$ denotes the set obtained by substituting the variables in $\Rightarrow(p, q, \overrightarrow{r})$ by the corresponding formulae, and $\varphi \Rightarrow \psi$ denotes the set $\bigcup\{\Rightarrow(\varphi, \psi, \overrightarrow{\alpha}) \mid \overrightarrow{\alpha} \in \mathrm{Fm}_{\mathcal{L}}^{<\omega}\}$.

We generalize the following properties, typically satisfied by an implication, to sets of (parameterized) formulae. However a reader can always understand these conditions as if $\Rightarrow$ would be just a single binary connective.

Definition 1. *Let* L *be a logic and* $\Rightarrow(p, q, \overrightarrow{r}) \subseteq \mathrm{Fm}_{\mathcal{L}}$ *be a parameterized set of formulae. We say that* $\Rightarrow$ *is a* weak p-implication *in* L *if:*

$$
\begin{array}{ll}
\text{(R)} & \vdash_{\mathrm{L}} \varphi \Rightarrow \varphi \\
\text{(MP)} & \varphi, \varphi \Rightarrow \psi \vdash_{\mathrm{L}} \psi \\
\text{(T)} & \varphi \Rightarrow \psi, \psi \Rightarrow \chi \vdash_{\mathrm{L}} \varphi \Rightarrow \chi \\
\text{(sCng)} & \varphi \Rightarrow \psi, \psi \Rightarrow \varphi \vdash_{\mathrm{L}} c(\chi_1, \ldots, \chi_i, \varphi, \ldots, \chi_n) \\
 & \Rightarrow c(\chi_1, \ldots, \chi_i, \psi, \ldots, \chi_n) \\
 & \text{for each } \langle c, n\rangle \in \mathcal{L} \text{ and each } i < n
\end{array}
$$

We change the prefix 'weak' to 'algebraic' if there is a set $\mathcal{E}(p)$ *of equations in one variable such that*

$$\text{(Alg)} \quad p \dashv\vdash_{\mathrm{L}} E[\mathcal{E}(p)],$$

where $E(p, q, \overrightarrow{r}) = \Rightarrow(p, q, \overrightarrow{r}) \cup \Rightarrow(q, p, \overrightarrow{r})$
We change the prefix 'weak' to 'regular' if:

$$\text{(Reg)} \quad \varphi, \psi \vdash_{\mathrm{L}} \psi \Rightarrow \varphi$$

We change the prefix 'weak' to 'Rasiowa' if:

$$\text{(W)} \quad \varphi \vdash_{\mathrm{L}} \psi \Rightarrow \varphi$$

Finally, if $\Rightarrow$ *is parameter-free we drop the prefix 'p-'.*

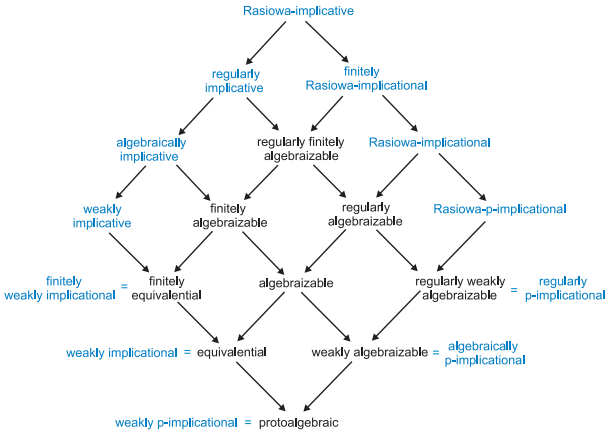
We can easily show the relative strength of defined notions: each Rasiowa p-implication is a regular p-implication and each regular p-implication is an algebraic p-implication.

Definition 2. *We say that a logic* L *is a* weakly/algebraically/regularly/Rasiowa- (p-)implicational logic *if there is a (parameterized) set of formulae* $\Rightarrow$ *which is a weak/algebraic/regular/Rasiowa (p-)implication in* L. *We add the prefix* 'finitely' *if* $\Rightarrow$ *is finite and we use the adjective* 'implicative' *instead of* 'implicational' *if* $\Rightarrow$ *is a* parameter-free singleton.

Each class of the well-known Leibniz hierarchy [7] coincides with some of our newly defined classes. In fact, our new taxonomy extends it, incorporates other already existing classes of logics,[1] and offers a more systematic way of classification: in one axis we go from p-implicational, implicational, finitely implicational to implicative (depending on the form of the implication set); in the second one we use prefixes 'weakly', 'algebraically', 'regularly', or 'Rasiowa-' (depending on extra properties fulfilled by that set). The translation table is:

Classes of Leibniz hierachy	*Our systematic names*
protoalgebraic	weakly p-implicational
(finitely) equivalential	(finitely) weakly impl.
weakly algebraizable	algebraically p-impl.
regularly weakly algebraizable	regularly p-implicational
(finitely) algebraizable	(finitely) algebraically impl.
(finitely) regularly algebraizable	(finitely) regularly impl.

Our new classification of logics, *the hierarchy of implicational logics* is depicted below (the arrows correspond to the class subsumption relation). We can show that almost all classes of logics in this hierarchy are mutually different; only the difference between Rasiowa-implicational and Rasiowa-p-implicational logics remains to be shown.



The syntactical notion of weak p-implication that we have introduced has a natural semantical counterpart: a preorder in the models that becomes an order in reduced models.

Definition 3. *Let* $\Rightarrow$ *be a parameterized set of formulae and* $\mathbf{A} = \langle \mathcal{A}, F\rangle$ *a matrix. We define a binary relation* $\leq_{\mathbf{A}}^{\Rightarrow}$ *on* A*:*

$$a \leq_{\mathbf{A}}^{\Rightarrow} b \qquad \textit{iff} \qquad a \Rightarrow^{\mathcal{A}} b \subseteq F.$$

Proposition 1. *Let* L *be a logic and* $\mathbf{A} \in \mathbf{MOD}(\mathrm{L})$*. Then a parameterized set* $\Rightarrow$ *is a weak p-implication in* L *iff* $\leq_{\mathbf{A}}^{\Rightarrow}$ *is a preorder and its symmetrization of* $\leq_{\mathbf{A}}^{\Rightarrow}$ *is the Leibniz congruence of* $\mathbf{A}$.

Clearly $\leq_{\mathbf{A}}^{\Rightarrow}$ is an order iff $\mathbf{A}$ is reduced. Thus (by virtue of Theorem 10) we can say that a L is complete w.r.t. the class of ordered matrices. Our main interest are the logics complete w.r.t. linearly ordered matrices in the following sense:

Definition 4. *Let* L *be a logic and* $\mathbf{A} = \mathbf{MOD}(\mathrm{L})$*. We say that* $\mathbf{A}$ *is a* linear model *w.r.t.* $\Rightarrow$ *if* $\leq_{\mathbf{A}}^{\Rightarrow}$ *is a linear order. The class of linear models of* L *is denoted by* $\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$.

Observe that the class of linear models is not intrinsically defined for a given logic as it depends on the chosen implication. However, we will see later that in a reasonably wide class of logics all *semilinear implications* define the same linear models. But even in a general case we can prove:

Theorem 1. *Let* L *be a protoalgebraic logic. Then, for any weak p-implication* $\Rightarrow$*,* $\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L}) \subseteq \mathbf{MOD}^{*}(\mathrm{L})_{\mathrm{RFSI}}$.

[1] Rasiowa-implicative logics were already defined in 1974 by Rasiowa [18] and weakly implicative logics in 2006 by Cintula [4].

ISBN: 978-989-95079-6-8

3 Semilinear implications and logics

Given a logic L and a weak p-implication $\Rightarrow$, we say that $\Rightarrow$ is a weak *semilinear* p-implication if the logic is complete w.r.t. the class of its corresponding linear models. Formally:

Definition 5. *Let* L *be a logic and* $\Rightarrow$ *a weak p-implication. We say that* $\Rightarrow$ *is a weak* semilinear *p-implication if*

$$\vdash_{\mathrm{L}} \;=\; \models_{\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})}.$$

Later we define *implicational semilinear logics* as those possessing some weak semilinear p-implication. Obviously, they will be fuzzy logics in the sense of [1]. However, we choose the term 'semilinear' instead of 'fuzzy' in spite of the fact that a first step towards the general definition we are offering here had been done by the first author in [4], when he defined the class of *weakly implicative fuzzy logics* (in our new framework: logics with a weak semilinear implication given a single binary connective). We have realized that the attempt of [4] of using the term 'fuzzy' to formally define a class of logics was rather futile as such word is heavily charged with many conflicting potential meanings.

Therefore, we have opted now for the new neutral name 'semilinear'. The term was first used by Olson and Raftery in [17] in the context of residuated lattices; it refers to the Universal Algebra tradition of calling a class of algebras 'semi$X$' whenever its subdirectly irreducible members have the property X (how this is related to our case will be apparent after Theorem 3). Despite using a new neutral name our intention remains the same: to formally delimit the class of fuzzy logics inside some existing abstract class of formal non-classical logics (originally, among weakly implicative ones now among the protoalgebraic ones). Of course it should include almost all the prominent examples of fuzzy logics known so far and exclude non-classical logics which are usually not recognized as fuzzy logics in the Logic community.

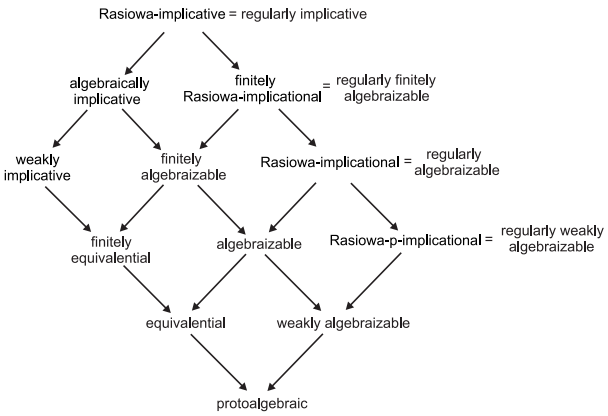
However let us stress that we *do not* expect to capture in a mathematical definition *the whole* intuitive notion of *arbitrary* fuzzy logic. Even if we would agree that linearity of semantics is crucial for a formal logic to be *fuzzy* there could be several other ways in which a logic might have a complete semantics somehow based on chains (see e.g. [2] or some recent work on modal fuzzy logics).

We formally define classes of implicational semilinear logics based on the form of semilinear implication they possess.

Definition 6. *We say that* L *is a* weakly/algebraically/ Rasiowa- (p-)implicational semilinear logic *if there is a (parameterized) set of formulae* $\Rightarrow$ *such that it is a weak/algebraic/Rasiowa semilinear (p-)implication in* L. *We add the* 'finitely' *if the set* $\Rightarrow$ *is finite and we use* 'implicative' *instead of* 'implicational' *if* $\Rightarrow$ *is a parameter-free singleton.*

We have not defined the class of *regularly* (p-)implicational (implicative) semilinear logics, because (as we will see in Corollary 2) we would obtain that each regularly p-implicational semilinear logic is a Rasiowa-p-implicational semilinear logic (and analogously for the other three Rasiowa-classes in the hierarchy of implicational logics). See all the classes and their inclusions in the next figure.

We can prove the mutual difference of many classes, but three differences remain to be seen: Rasiowa-implicational



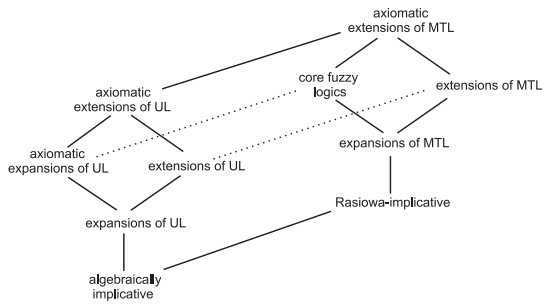
semilinear logics $\neq$ Rasiowa-p-implicational semilinear logics, algebraizable semilinear logics $\neq$ weakly algebraizable semilinear logics, and equivalential semilinear logics $\neq$ protoalgebraic semilinear logics.

Proposition 2. *Let* $\mathcal{X}$ *be any class in the hierarchy of implicational logics. Then, there is an* $\mathcal{X}$ *logic which is not an* $\mathcal{X}$ *semilinear logic.*

The three main logics based on continuous t-norms (Łukasiewicz, Gödel-Dummett, and Product logics) as well as the logic of all continuous t-norms BL are clearly Rasiowa-implicative semilinear logics. The same can be said in general as regards to left-continuous t-norm-based logics such as MTL and its t-norm based axiomatic extensions, and even for all *axiomatic extensions of MTL* (even those which are not complete w.r.t. a semantics of t-norms) as all of them are complete w.r.t. a subvariety of MTL-algebras generated by its linearly ordered members. Two incomparable superclasses of this one have been considered in the literature. On one hand, we have the so-called *core fuzzy logics* introduced in [15] as finitary logics expanding MTL or $\mathrm{MTL}_{\triangle}$, satisfying (sCng) for $\rightarrow$, and one of the following forms of Deduction Theorem: (i) $T, \varphi \vdash_{\mathrm{MTL}_{\triangle}} \psi$ iff $T \vdash_{\mathrm{MTL}_{\triangle}} \Delta\varphi \rightarrow \psi$, for expansions of $\mathrm{MTL}_{\triangle}$, or (ii) $T, \varphi \vdash_{\mathrm{MTL}} \psi$ iff there is $n \in \mathbb{N}$ such that $T \vdash_{\mathrm{MTL}} \varphi^n \rightarrow \psi$, for expansions of MTL. On the other hand, we can consider the class of all *semilinear finitary extensions of MTL*. Their equivalent quasivariety semantics are the subquasivarieties of MTL-algebras generated by chains. Since there are such quasivarieties that are not varieties, we have that this class is strictly bigger than that of axiomatic extensions of MTL. Both incomparable classes are included in the class of *semilinear expansions of MTL*, and finally this one is included in the Rasiowa-implicative semilinear logics. In the recent paper [16] the fuzzy logic UL based on uninorms instead of t-norms has been studied. It is an algebraizable logic without weakening, so it belongs to the class of algebraically implicative semilinear logics. We can consider the same structure of classes as above without weakening by replacing MTL for UL. See the resulting hierarchy of classes of semilinear logics in the next figure. We realize that all of them lie on the top of our classification, above Rasiowa-implicative or algebraically implicative semilinear logics. But if, by means of our definition of semilinear implication presented in this paper, we have succeeded in capturing an interesting way by which a logic can be fuzzy this means that fuzzy logics are a much wider class than those studied so far. Thus,

ISBN: 978-989-95079-6-8

future research in the field will probably bring new significant examples of fuzzy logics throughout the whole hierarchy of implicational semilinear logics.



4 Characterizations of semilinearity

This section provides some useful mathematical characterization of semilinear implications. First we define:

Definition 7. *Let* $\mathbf{A} = \langle \mathcal{A}, F \rangle \in \mathbf{MOD}(\mathrm{L})$. *The filter* F *is called* $\Rightarrow$-linear *if* $\leq^{\Rightarrow}_{\mathbf{A}}$ *is a total preorder.*

We say that L *has the* Linear Extension Property (LEP) *w.r.t.* $\Rightarrow$ *if for every theory* $T \in Th(\mathrm{L})$ *and every formula* $\varphi \in \mathrm{Fm}_{\mathcal{L}} \setminus T$, *there is a* $\Rightarrow$*-linear theory* $T' \supseteq T$ *s.t.* $\varphi \notin T'$.

Notice that a matrix $\langle \mathcal{A}, F \rangle \in \mathbf{MOD}(\mathrm{L})$ is in $\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$ iff it is reduced and F is $\Rightarrow$-linear. Clearly the (LEP) says that the $\Rightarrow$-linear theories form a basis of the closure system $Th(\mathrm{L})$. Next theorem shows that an analogous statement holds for other than Lindenbaum matrices, as one of the so-called 'transfer principles' from AAL.

Theorem 2. *Let* L *be a finitary logic with* (LEP) *w.r.t.* $\Rightarrow$ *and* $\mathcal{A} \in \mathbf{ALG}^{*}(\mathrm{L})$. *Then* $\Rightarrow$*-linear filters form a basis of* $\mathcal{F}i_{\mathrm{L}}(\mathcal{A})$.

Next we generalize the 'Prelinearity property' from [4]. However, here we prefer the new name 'Semilinearity Property' following our new terminology.

Definition 8. *We say that* L *has the* Semilinearity Property (SLP) *w.r.t.* $\Rightarrow$ *if the following meta rule is valid:*

$$\frac{\Gamma, \varphi \Rightarrow \psi \vdash_{\mathrm{L}} \chi \qquad \Gamma, \psi \Rightarrow \varphi \vdash_{\mathrm{L}} \chi}{\Gamma \vdash_{\mathrm{L}} \chi}$$

Theorem 3 (Characterization of semilinear implications). *The following are equivalent:*

1. $\Rightarrow$ *is semilinear in* L,
2. L *has the* (LEP) *w.r.t.* $\Rightarrow$.

Furthermore, if L *is finitary we can add:*

3. L *has the* (SLP) *w.r.t.* $\Rightarrow$,
4. $\mathbf{MOD}^{*}(\mathrm{L})_{\mathrm{RSI}} \subseteq \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$.

Moreover, if $\Rightarrow$ *is finite we can add:*

5. $\mathbf{MOD}^{*}(\mathrm{L})_{\mathrm{RFSI}} \subseteq \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$.

The previous theorem has several important corollaries. Using Theorem 1 we obtain that, at least in a reasonably wide class of logics, being the class of linear models w.r.t. any finite semilinear implication is an intrinsic property of a logic.

Corollary 1. *Let* L *be a finitary protoalgebraic logic and* $\Rightarrow$ *a finite weak semilinear p-implication. Then* $\mathbf{MOD}^{*}(\mathrm{L})_{\mathrm{RFSI}} = \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$.

Corollary 2. *Each regular semilinear p-implication is a Rasiowa p-implication.*

Corollary 3. *Let* $\Rightarrow$ *a weak semilinear p-implication in* L. *Then,* $\Rightarrow$ *is semilinear in all axiomatic extensions of* L.

This last corollary will be particulary useful for showing that some logic has *no* semilinear implication. It is quite easy to show that an implication in some logic is not semilinear, consider e.g. the normal implication of the intuitionistic logic; the well-know fact that the linear Heyting algebras do not generate the variety of Heyting algebras does the job. However using part 5. of the characterization theorem we can show much more: there is *no weak semilinear p-implication definable* in the intuitionistic logic, i.e. not only the standard nice Rasiowa implication given by a single formula is not semilinear but even using an infinite set with parameters we could never obtain an implication whose linearly ordered Heyting algebras would generate the variety of Heyting algebras.

Proposition 3. *Let* L *be the logic of a quasivariety of pointed residuated lattices containing the variety of Heyting algebras. Then,* L *is not weakly semilinear p-implicational logic.*

Many well-known logics fall under the scope of the previous proposition: Full Lambek logic (possibly extended with structural rules), multiplicative-additive fragment of (Affine) Intuitionistic Linear logic, Relevance logic R, etc.

Corollary 4. *Let* L *be a logic and* $\Rightarrow$ *a weak p-implication. Then, there is the weakest logic extending* L *where* $\Rightarrow$ *is semilinear. Let us denote this logic as* $\mathrm{L}^{\ell}_{\Rightarrow}$.

In Section 6 we will show how to axiomatize $\mathrm{L}^{\ell}_{\Rightarrow}$. However, to determine a complete semantics is simple:

Proposition 4. *Let* $\Rightarrow$ *be a weak p-implication in* L. *Then,* $\mathrm{L}^{\ell}_{\Rightarrow} = \models_{\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})}$ *and* $\mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L}^{\ell}_{\Rightarrow}) = \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$.

5 Disjunctions

In order to provide additional characterizations of semilinearity (and to fill the gap between our abstract setting and real-life logics) we need to study a generalized notion of disjunction. As in the case of implication, given a parameterized set of formulae $\nabla(p, q, \overrightarrow{r})$ and formulae φ and ψ we define $\varphi \nabla \psi$.

Definition 9. *Given a logic* L *and a parameterized set of formulae* ∇, *we define the following properties:*

(PD)	$\varphi \vdash_{\mathrm{L}} \varphi \nabla \psi \quad$ *and* $\quad \psi \vdash_{\mathrm{L}} \varphi \nabla \psi$
(C)	$\varphi \nabla \psi \vdash_{\mathrm{L}} \psi \nabla \varphi$
(I)	$\varphi \nabla \varphi \vdash_{\mathrm{L}} \varphi$
(A)	$\varphi \nabla (\psi \nabla \chi) \dashv\vdash_{\mathrm{L}} (\varphi \nabla \psi) \nabla \chi$
(PCP)	*If* $\Gamma, \varphi \vdash_{\mathrm{L}} \chi$ *and* $\Gamma, \psi \vdash_{\mathrm{L}} \chi$, *then* $\Gamma, \varphi \nabla \psi \vdash_{\mathrm{L}} \chi$.

They correspond to well known usual properties of disjunction connectives. (C), (I) and (A) are respectively commutativity, idempotency and associativity, which are typically also satisfied by conjunction connectives. In contrast, the (PCP) is typically satisfied only by disjunction connectives. In [6] we also study a weaker variant of (PCP) for $\Gamma = \emptyset$.

ISBN: 978-989-95079-6-8

Definition 10. *Given a logic* L *and a parameterized set of formulae* $\nabla(p, q, \overrightarrow{r})$ *and a set of properties* $\sigma \subseteq \{(\mathrm{C}), (\mathrm{I}), (\mathrm{A})\}$*, we say that* ∇ *is a* σ-p-protodisjunction *in* L *if* (PD) *and the properties of* σ *are satisfied. Furthermore a p-protodisjunction* ∇ *is a* p-disjunction *if it satisfies* (PCP)*. Finally, if* ∇ *has no parameters we drop the prefix 'p-'.*

All these defined notions are mutually distinct and any p-disjunction is in fact a $\{(\mathrm{C}), (\mathrm{I}), (\mathrm{A})\}$-p-protodisjunction. Moreover, the notion of p-disjunction is intrinsic as any two p-disjunctions are mutually interderivable.

Definition 11. *We call a logic* (p-)disjunctional *if it has a (p-)disjunction. Furthermore, we call a logic* disjunctive *if it has a disjunction given by a single parameter-free formula.*

The classes of disjunctive and disjunctional logics are mutually different. E.g. the implication fragment of Gödel logic is not disjunctive but the set $\{(p \to q) \to q, (q \to p) \to p\}$ is its disjunction.

Definition 12. *Let* L *be a logic,* ∇ *a parameterized set of formulae,* $\mathcal{A} \in \mathbf{ALG}^*(\mathrm{L})$*, and* $F \in \mathcal{F}i_{\mathrm{L}}(\mathcal{A})$*. F is called* ∇-prime *if for every* $a, b \in A$*,* $a\nabla^{\mathcal{A}}b \subseteq F$ *iff* $a \in F$ *or* $b \in F$*.*

The *prime extension property* (PEP) is defined as the (LEP) by substituting the notion of $\Rightarrow$-linear filter for that of ∇-prime filter. Then we can prove:

Theorem 4. *Let* L *be a finitary logic and* ∇ *a p-protodisjunction. Then* L *has the* (PEP) *iff* ∇ *is p-disjunction.*

Theorem 5. *Let* L *be a finitary p-disjunctional logic. Then* ∇*-prime filters form a basis of* $\mathcal{F}i_{\mathrm{L}}(\mathcal{A})$ *for any* $\mathcal{A} \in \mathbf{ALG}^*(\mathrm{L})$*.*

Now we provide a syntactical characterization of (PCP). Let us by R^{∇} (for an $\mathcal{L}$-consecution $R = \Gamma \rhd \varphi$) denote the set $\{\Gamma\nabla\chi \rhd \delta \mid \chi \in \mathrm{Fm}_{\mathcal{L}}$ and $\delta \in \varphi\nabla\chi\}$.

Theorem 6. *Let* L *be a finitary logic with a presentation* $\mathcal{AS}$ *and* ∇ *a* $\{(\mathrm{C}), (\mathrm{I})\}$*-p-protodisjunction. Then, the following are equivalent:*

1. ∇ *is a p-disjunction,*
2. $R^{\nabla} \subseteq \mathrm{L}$ *for each (finitary)* $R \in \mathrm{L}$*,*
3. $R^{\nabla} \subseteq \mathrm{L}$ *for each* $R \in \mathcal{AS}$*.*

Corollary 5. *Let* ∇ *be a p-disjunction in a finitary logic* L_1 *and* L_2 *an expansion of* L_1 *by a set of consecutions* $\mathcal{C}$*. Then:*

- ∇ *is a p-disjunction in* L_2 *if* $R^{\nabla} \subseteq \mathrm{L}_2$ *for each* $R \in \mathcal{C}$*.*
- *If all the consecutions from* $\mathcal{C}$ *are finitary, then* $R^{\nabla} \subseteq \mathrm{L}_2$ *for each* $R \in \mathcal{C}$ *iff* ∇ *is a p-disjunction in* L_2*.*
- *If all the consecutions from* $\mathcal{C}$ *are axioms, then* ∇ *is a p-disjunction.*

Definition 13. *Let* L *be a logic and* ∇ *a parameterized set of formulae. We denote by* L^{∇} *the least logic extending* L *where* ∇ *is a p-disjunction.*

Theorem 7. *Let* L *be a finitary logic with a finitary presentation* $\mathcal{AS}$ *and* ∇ *a* $\{(\mathrm{C}), (\mathrm{I}), (\mathrm{A})\}$*-p-protodisjunction. Then,* L^{∇} *is axiomatized by* $\mathcal{AS} \cup \bigcup\{R^{\nabla} \mid R \in \mathcal{AS}\}$*.*

6 Disjunctions and semilinearity

In this section we consider the interesting relationships between the several kinds of disjunctions and implications we have defined and their corresponding properties. First, we introduce two natural syntactical conditions: a version of Modus Ponens with disjunction (DMP): $\varphi \Rightarrow \psi, \varphi\nabla\psi \vdash_{\mathrm{L}} \psi$ and $\varphi \Rightarrow \psi, \psi\nabla\varphi \vdash_{\mathrm{L}} \psi$, and a generalization of the prelinearity axiom used in fuzzy logics (P): $\vdash_{\mathrm{L}} (\varphi \Rightarrow \psi)\nabla(\psi \Rightarrow \varphi)$.

Theorem 8. *Let* L *be a logic,* ∇ *a p-protodisjunction, and* $\Rightarrow$ *a weak p-implication.*

- *If* L *fulfills* (DMP)*, we have:*
  1. *each* $\Rightarrow$*-linear theory is* ∇*-prime,*
  2. *if* $\Rightarrow$ *has the* (LEP)*, then* ∇ *has the* (PEP)*,*
  3. *if* $\Rightarrow$ *has the* (SLP)*, then* ∇ *has the* (PCP)*.*
- *If* L *fulfills* (P)*, we have:*
  4. *each* ∇*-prime theory is* $\Rightarrow$*-linear,*
  5. *if* ∇ *has the* (PEP)*, then* $\Rightarrow$ *has the* (LEP)*.*
- *If* L *fulfills* (P) *and either it is finitary or* $\Rightarrow$ *is finite and parameter-free, we have:*
  6. *if* ∇ *has the* (PCP)*, then* $\Rightarrow$ *has the* (SLP)*.*

This theorem together with known relations of the properties (SLP), (PCP), (PEP), (LEP) and semilinearity (Theorems 3 and 4) allows us to formulate numerous corollaries about their mutual relationships.

Corollary 6. *If* L *is finitary,* ∇ *is a p-protodisjunction, and* $\Rightarrow$ *a weak p-implication, the following are equivalent:*

1. L *satisfies* (DMP) *and* $\Rightarrow$ *is semilinear.*
2. L *satisfies* (DMP) *and* $\Rightarrow$ *has the* (SLP)*.*
3. L *satisfies* (DMP) *and* $\Rightarrow$ *has the* (LEP)*.*
4. L *satisfies* (P) *and* ∇ *has the* (PEP)*.*
5. L *satisfies* (P) *and* ∇ *has the* (PCP)*.*

Thus e.g. a weak p-implication in a finitary p-disjunctional logic L is semilinear iff L satisfies (P). In p-disjunctional logic we can strengthen two important results from the previous section. First, we can remove the precondition of finiteness of implication in Part 5. of Theorem 3.

Corollary 7. *Let* L *be a finitary p-disjunctional logic and* $\Rightarrow$ *a weak p-implication. Then the following are equivalent:*

1. $\Rightarrow$ *is semilinear in* L*,*
2. $\mathbf{MOD}^*(\mathrm{L})_{\mathrm{RFSI}} \subseteq \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$*.*

Furthermore, in any finitary p-disjunctional protoalgebraic logic it holds that: $\mathbf{MOD}^*(\mathrm{L})_{\mathrm{RFSI}} = \mathbf{MOD}^{\ell}_{\Rightarrow}(\mathrm{L})$ *for any semilinear p-implication* $\Rightarrow$*.*

Theorem 9. *If* L *is finitary,* ∇ *is a* $\{(\mathrm{C}),(\mathrm{I}),(\mathrm{A})\}$*-p-protodisjunction,* $\Rightarrow$ *is a weak p-implication and* L *satisfies* (DMP)*, then* $\mathrm{L}^{\ell}_{\Rightarrow}$ *is the extension of* L^{∇} *by* (P)*.*

Corollary 8. *Let* L *be a finitary p-disjunctional logic and* $\Rightarrow$ *a weak p-implication. Then,* $\mathrm{L}^{\ell}_{\Rightarrow}$ *is extension of* L *by* (P)*.*

ISBN: 978-989-95079-6-8

A Bits of the theory of logical calculi

We recall some basic definitions and results of Abstract Algebraic Logic.[2] The notion of *propositional language* $\mathcal{L}$ is defined in the usual way (a set of connectives with finite arity). By $\mathbf{Fm}_{\mathcal{L}}$ we denote the free term algebra over a denumerable set of variables in the language $\mathcal{L}$, by $\mathrm{Fm}_{\mathcal{L}}$ we denote its universe and we call its elements $\mathcal{L}$*-formulae*.

A $\mathcal{L}$*-consecution* is a pair $\Gamma \rhd \varphi$, where $\Gamma \subseteq \mathrm{Fm}_{\mathcal{L}}$ and $\varphi \in \mathrm{Fm}_{\mathcal{L}}$. A consecution $\Gamma \rhd \varphi$ is *finitary* if Γ is finite. For a set of consecutions L we write $\Gamma \vdash_{\mathrm{L}} \varphi$ rather than $\Gamma \rhd \varphi \in \mathrm{L}$. A *propositional logic* is a pair $\mathrm{L} = \langle \mathcal{L}, \vdash_{\mathrm{L}} \rangle$ where $\vdash_{\mathrm{L}}$ is a structural consequence relation.

A logic L is *finitary* if for every $\Gamma \cup \{\varphi\} \subseteq \mathrm{Fm}_{\mathcal{L}}$ such that $\Gamma \vdash_{\mathrm{L}} \varphi$ there is a finite $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \vdash_{\mathrm{L}} \varphi$. We write $\Gamma \vdash_{\mathrm{L}} \Delta$ when $\Gamma \vdash_{\mathrm{L}} \varphi$ for every $\varphi \in \Delta$. A *theory* of a logic L is a set of formulae T such that if $T \vdash_{\mathrm{L}} \varphi$ then $\varphi \in T$. By $Th(\mathrm{L})$ we denote the set of all theories of L.

Given a *finitary* logic $\mathrm{L} = \langle \mathcal{L}, \vdash_{\mathrm{L}} \rangle$, we say that a set $\mathcal{AS}$ of $\mathcal{L}$-consecutions whose left member is finite is a *presentation* of L if the relation $\vdash_{\mathrm{L}}$ coincides with the provability relation given by $\mathcal{AS}$ as a Hilbert-style axiomatic system.

Given a language $\mathcal{L}$, an $\mathcal{L}$*-matrix* is a pair $\mathbf{A} = \langle \mathcal{A}, F \rangle$ where $\mathcal{A}$ is an $\mathcal{L}$-algebra and F is a subset of A called the *filter* of $\mathbf{A}$. A homomorphism from $\mathbf{Fm}_{\mathcal{L}}$ to $\mathcal{A}$ is called an $\mathbf{A}$*-evaluation*. The semantical consequence w.r.t. a class of matrices $\mathbb{K}$ is defined as: $\Gamma \models_{\mathbb{K}} \varphi$ iff for each $\mathbf{A} \in \mathbb{K}$ and each $\mathbf{A}$-evaluation e we obtain $e(\varphi) \in F$ whenever $e[\Gamma] \subseteq F$. Clearly, $\langle \mathcal{L}, \models_{\mathbb{K}} \rangle$ is a logic. We say that a matrix $\mathbf{A}$ is a *model* of L if $\vdash_{\mathrm{L}} \subseteq \models_{\mathbf{A}}$ and write $\mathbf{A} \in \mathbf{MOD}(\mathrm{L})$.

Given an $\mathcal{L}$-algebra $\mathcal{A}$, a subset $F \subseteq A$ is an L*-filter* if $\langle \mathcal{A}, F \rangle \in \mathbf{MOD}(\mathrm{L})$. Let $\mathcal{F}i_{\mathrm{L}}(\mathcal{A})$ be the set of all L-filters over $\mathcal{A}$. Observe that for every set of formulae T, we have $T \in Th(\mathrm{L})$ iff $\langle \mathbf{Fm}_{\mathcal{L}}, T \rangle \in \mathbf{MOD}(\mathrm{L})$; these models are called the *Lindenbaum matrices* for L.

It is straightforward to check that $\mathcal{F}i_{\mathrm{L}}(\mathcal{A})$ is closed under arbitrary intersections and hence it is a closure system. Recall that a family $\mathcal{B} \subseteq \mathcal{C}$ is a *basis* of a closure system $\mathcal{C}$ if for every $X \in \mathcal{C}$ there is a $\mathcal{D} \subseteq \mathcal{C}$ such that $X = \bigcap \mathcal{D}$ (which can be equivalent formulated as: for every $Y \in \mathcal{C}$ and every $a \in A \setminus Y$ there is $Z \in \mathcal{B}$ such that $Y \subseteq Z$ and $a \notin Z$).

Given a matrix $\mathbf{A} = \langle \mathcal{A}, F \rangle$, a binary relation $\Omega_{\mathcal{A}}(F)$ is defined as $\langle a, b \rangle \in \Omega_{\mathcal{A}}(F)$ if, and only if, for every sequence of parameters $\overrightarrow{z}$, $\mathcal{L}$-formula $\varphi(x, \overrightarrow{z})$, and $\overrightarrow{c} \in A^{<\omega}$ we have $\varphi^{\mathcal{A}}(a, \overrightarrow{c}) \in F$ iff $\varphi^{\mathcal{A}}(b, \overrightarrow{c}) \in F$. Inspired by the famous Leibniz's identity of indiscernibles principle, $\Omega_{\mathcal{A}}(F)$ is called the *Leibniz congruence* of $\langle \mathcal{A}, F \rangle$.

A matrix is said to be *reduced* if its Leibniz congruence is the identity relation. Given a logic L, the class of its reduced models is denoted by $\mathbf{MOD}^*(\mathrm{L})$, and the class of algebraic reducts of $\mathbf{MOD}^*(\mathrm{L})$ is denoted by $\mathbf{ALG}^*(\mathrm{L})$. They are enough to provide a complete semantics for the logic:

Theorem 10. *Let* L *be a logic. Then* $\Gamma \vdash_{\mathrm{L}} \varphi$ *if, and only if,* $\Gamma \models_{\mathbf{MOD}^*(\mathrm{L})} \varphi$*, for every set of formulae* $\Gamma \cup \{\varphi\}$.

A matrix $\mathbf{A} \in \mathbf{MOD}^*(\mathrm{L})$ is called *(finitely) subdirectly irreducible* if it is not a non-trivial (finite) subdirect product of reduced matrices. The corresponding classes of matrices are denoted as $\mathbf{MOD}^*(\mathrm{L})_{\mathrm{RSI}}$ and $\mathbf{MOD}^*(\mathrm{L})_{\mathrm{RFSI}}$ respectively.

[2] For a comprehensive survey see [7, 10, 11]. Any necessary background on Universal Algebra can be found e.g. in [3].

Acknowledgments

Petr Cintula was partly supported by ESF EUROCORES project ICC/08/E018 of the Grant Agency of the Czech Republic and partly by Institutional Research Plan AVOZ10300504. Carles Noguera acknowledges partial support from the grant 2006-BP-A-10043 of the Departament d'Educació i Universitats of the Generalitat de Catalunya and the Spanish project MULOG2 (TIN2007-68005-C04).

References

[1] Libor Běhounek and Petr Cintula. Fuzzy logics as the logics of chains. *Fuzzy Sets and Systems*, 157(5):604–610, 2006.

[2] Félix Bou, Francesc Esteva, Josep Maria Font, Àngel Gil, Lluís Godo, Antoni Torrens, and Ventura Verdú. Logics preserving degrees of truth from varieties of residuated lattices. To appear in *Journal of Logic and Computation*, 2009.

[3] Stanley Burris and H.P. Sankappanavar. *A Course in Universal Algebra*, volume 78 of *Graduate Texts in Mathematics*. Springer-Verlag, 1981.

[4] Petr Cintula. Weakly implicative (fuzzy) logics I: Basic properties. *Archive for Mathematical Logic*, 45(6):673–704, 2006.

[5] Petr Cintula, Francesc Esteva, Joan Gispert, Lluís Godo, Franco Montagna, and Carles Noguera. Distinguished algebraic semantics for t-norm based fuzzy logics: Methods and algebraic equivalencies. *Annals of Pure and Applied Logic*, 160(1):53–81, 2009

[6] Petr Cintula and Carles Noguera. A hierarchy of implicational (semilinear) logics: the propositional case. Submitted, 2009. (Preprint available at http://www.cs.cas.cz/mathfuzzlog)

[7] Janusz Czelakowski. *Protoalgebraic Logics*, volume 10 of *Trends in Logic*. Kluwer, Dordercht, 2001.

[8] Francesc Esteva and Lluís Godo. Monoidal t-norm based logic: Towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124(3):271–288, 2001.

[9] Francesc Esteva, Lluís Godo, Petr Hájek, and Franco Montagna. Hoops and Fuzzy Logic, *Journal of Logic and Computation*, 13(4):531–555, 2003.

[10] Josep Maria Font and Ramon Jansana. *A General Algebraic Semantics for Sentential Logics*. Springer-Verlag, 1996.

[11] Josep Maria Font, Ramon Jansana, and Don Pigozzi. A survey of Abstract Algebraic Logic. *Studia Logica*, 74(1–2, Special Issue on Abstract Algebraic Logic II):13–97, 2003.

[12] Nikolaos Galatos, Peter Jipsen, Tomasz Kowalski, and Hiroakira Ono. *Residuated Lattices: An Algebraic Glimpse at Substructural Logics*, volume 151 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, Amsterdam, 2007.

[13] Petr Hájek. Fuzzy Logics with Noncommutative Conjunctions. *Journal of Logic and Computation* 13(4):469–479, 2003.

[14] Petr Hájek. *Metamathematics of Fuzzy Logic*, volume 4 of *Trends in Logic*. Kluwer, Dordercht, 1998.

[15] Petr Hájek and Petr Cintula. On theories and models in fuzzy predicate logics. *Journal of Symbolic Logic*, 71(3):863–880, 2006.

[16] George Metcalfe and Franco Montagna. Substructural fuzzy logics. *Journal of Symbolic Logic*, 72(3):834–864, 2007.

[17] Jeffrey S. Olson and James G. Raftery. Positive Sugihara monoids. *Algebra Universalis*, 57:75–99, 2007.

[18] Helena Rasiowa. *An Algebraic Approach to Non-Classical Logics*. North-Holland, Amsterdam, 1974.

ISBN: 978-989-95079-6-8

Polynomial functions on bounded chains

Miguel Couceiro Jean-Luc Marichal

Mathematics Research Unit, University of Luxembourg
162A, avenue de la Faïencerie, L-1511 Luxembourg, Luxembourg
Email: {miguel.couceiro, jean-luc.marichal}@uni.lu

Abstract— *We are interested in representations and characterizations of lattice polynomial functions $f : L^n \to L$, where L is a given bounded distributive lattice. In an earlier paper [4, 5], we investigated certain representations and provided various characterizations of these functions both as solutions of certain functional equations and in terms of necessary and sufficient conditions. In the present paper, we investigate these representations and characterizations in the special case when L is a chain, i.e., a totally ordered lattice. More precisely, we discuss representations of lattice polynomial functions given in terms of standard simplices and we present new axiomatizations of these functions by relaxing some of the conditions given in [4, 5] and by considering further conditions, namely comonotonic minitivity and maxitivity.*

Keywords— Lattice polynomial function, discrete Sugeno integral, normal form, homogeneity, strong idempotency, median decomposability, comonotonicity.

1 Introduction

In [5], the class of (*lattice*) *polynomial functions*, i.e., functions representable by combinations of variables and constants using the lattice operations $\wedge$ and $\vee$, was considered and characterized both as solutions of certain functional equations and in terms of necessary and sufficient conditions rooted in aggregation theory.

Formally, let L be a bounded distributive lattice with operations $\wedge$ and $\vee$, and with least and greatest elements 0 and 1, respectively. An *n-ary polynomial function* on L is any function $f : L^n \to L$ which can be obtained by finitely many applications of the following rules:

(i) For each $i \in [n] = \{1, \ldots, n\}$ and each $c \in L$, the projection $\mathbf{x} \mapsto x_i$ and the constant function $\mathbf{x} \mapsto c$ are polynomial functions from L^n to L.

(ii) If f and g are polynomial functions from L^n to L, then $f \vee g$ and $f \wedge g$ are polynomial functions from L^n to L.

Polynomial functions are also called lattice functions (Goodstein [9]), algebraic functions (Burris and Sankappanavar [3]) or weighted lattice polynomial functions (Marichal [15]). Polynomial functions obtained from projections by finitely many applications of (ii) are referred to as *lattice term functions*.

The recent interest by aggregation theorists in this class of polynomial functions is partially motivated by its connection to noteworthy aggregation functions such as the (*discrete*) *Sugeno integral*, which was introduced by Sugeno [18, 19] and widely investigated in aggregation theory, due to the many applications in fuzzy set theory, data fusion, decision making, image analysis, etc. As shown in [15], the discrete Sugeno integrals are nothing other than those polynomial functions $f : L^n \to L$ which are idempotent, that is, satisfying $f(x, \ldots, x) = x$. For general background on aggregation theory, see [1, 11] and for a recent reference, see [10].

In this paper, we refine our previous results in the particular case when L is a chain, by relaxing our conditions and proposing weak analogues of those properties used in [4, 5], and then providing characterizations of polynomial functions, accordingly. Moreover, and motivated by the axiomatizations of the discrete Sugeno integrals established by de Campos and Bolaños [6] (in the case when $L = [0, 1]$ is the unit real interval), we present further and alternative characterizations of polynomial functions given in terms of comonotonic minitivity and maxitivity. As particular cases, we consider the subclass of discrete Sugeno integrals.

The current paper is organized as follows. We start in §2 by introducing the basic notions needed in this paper and presenting the characterizations of lattice polynomial functions on arbitrary (possibly infinite) bounded distributive lattices, established in [4, 5]. Those characterizations are reassembled in Theorem 1. We discuss representations of polynomial functions in normal form (such as the classical disjunctive and conjunctive normal forms) and introduce variant representations in the case of chains and given in terms of standard simplices in §3. In §4, we present characterizations of polynomial functions on chains given in terms of weak analogues of the properties used in Theorem 1 as well as in terms of comonotonic minitivity and maxitivity. Axiomatizations for the subclass of discrete Sugeno integrals are then presented in §5.

2 Basic notions and terminology

Throughout this paper, let L be a bounded chain with operations $\wedge$ and $\vee$, and with least and greatest elements 0 and 1, respectively. A subset S of a chain L is said to be *convex* if for every $a, b \in S$ and every $c \in L$ such that $a \leqslant c \leqslant b$, we have $c \in S$. For any subset $S \subseteq L$, we denote by $\overline{S}$ the convex hull of S, that is, the smallest convex subset of L containing S. For every $a, b \in S$ such that $a \leqslant b$, the *interval* $[a, b]$ is the set $[a, b] = \{c \in L : a \leqslant c \leqslant b\}$. For any integer $n \geqslant 1$, let $[n] = \{1, \ldots, n\}$.

For any bounded chain L, we regard the Cartesian product L^n, $n \geqslant 1$, as a distributive lattice endowed with the operations $\wedge$ and $\vee$ given by

$$\begin{aligned}(a_1, \ldots, a_n) \wedge (b_1, \ldots, b_n) &= (a_1 \wedge b_1, \ldots, a_n \wedge b_n),\\ (a_1, \ldots, a_n) \vee (b_1, \ldots, b_n) &= (a_1 \vee b_1, \ldots, a_n \vee b_n).\end{aligned}$$

The elements of L are denoted by lower case letters $a, b, c, \ldots$, and the elements of L^n, $n > 1$, by bold face letters $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$.

ISBN: 978-989-95079-6-8

We also use $\mathbf{0}$ and $\mathbf{1}$ to denote the least element and greatest element, respectively, of L^n. For $c \in L$ and $\mathbf{x} = (x_1, \ldots, x_n) \in L^n$, set

$$\mathbf{x} \wedge c = (x_1 \wedge c, \ldots, x_n \wedge c) \quad \text{and} \quad \mathbf{x} \vee c = (x_1 \vee c, \ldots, x_n \vee c).$$

The *range* of a function $f\colon L^n \to L$ is defined by $\mathcal{R}_f = \{f(\mathbf{x}) : \mathbf{x} \in L^n\}$. A function $f\colon L^n \to L$ is said to be *nondecreasing* (*in each variable*) if, for every $\mathbf{a}, \mathbf{b} \in L^n$ such that $\mathbf{a} \leqslant \mathbf{b}$, we have $f(\mathbf{a}) \leqslant f(\mathbf{b})$. Note that if f is nondecreasing, then $\overline{\mathcal{R}}_f = [f(\mathbf{0}), f(\mathbf{1})]$.

Let S be a nonempty subset of L. A function $f\colon L^n \to L$ is said to be

- *S-idempotent* if for every $c \in S$, we have $f(c, ..., c) = c$.
- *S-min homogeneous* if for every $\mathbf{x} \in L^n$ and $c \in S$,
$$f(\mathbf{x} \wedge c) = f(\mathbf{x}) \wedge c. \tag{1}$$
- *S-max homogeneous* if for every $\mathbf{x} \in L^n$ and $c \in S$,
$$f(\mathbf{x} \vee c) = f(\mathbf{x}) \vee c. \tag{2}$$
- *horizontally S-minitive* if for every $\mathbf{x} \in L^n$ and $c \in S$,
$$f(\mathbf{x}) = f(\mathbf{x} \vee c) \wedge f([\mathbf{x}]^c), \tag{3}$$
where $[\mathbf{x}]^c$ is the n-tuple whose ith component is 1, if $x_i \geqslant c$, and x_i, otherwise.
- *horizontally S-maxitive* if for every $\mathbf{x} \in L^n$ and $c \in S$,
$$f(\mathbf{x}) = f(\mathbf{x} \wedge c) \vee f([\mathbf{x}]_c), \tag{4}$$
where $[\mathbf{x}]_c$ is the n-tuple whose ith component is 0, if $x_i \leqslant c$, and x_i, otherwise.
- *median decomposable* if, for every $\mathbf{x} \in L^n$ and $k \in [n]$,
$$f(\mathbf{x}) = \operatorname{median}\big(f(\mathbf{x}_k^0), x_k, f(\mathbf{x}_k^1)\big), \tag{5}$$
where $\operatorname{median}(x, y, z) = (x \wedge y) \vee (y \wedge z) \vee (z \wedge x)$ and $\mathbf{x}_k^c = (x_1, \ldots, x_{k-1}, c, x_{k+1}, \ldots, x_n)$ for $c \in L$.
- *strongly idempotent* if, for every $\mathbf{x} \in L^n$ and $k \in [n]$,
$$f(x_1, \ldots, x_{k-1}, f(\mathbf{x}), x_{k+1}, \ldots, x_n) = f(\mathbf{x}).$$

Remark 1. In the case when $S = L$ is the real interval $[0, 1]$, the concepts of S-min and S-max homogeneity were introduced by Fodor and Roubens [8] to specify certain classes of aggregation functions, and the concept of horizontal S-maxitivity was introduced by Benvenuti et al. [2] as a general property of the Sugeno integral. The concept of median decomposability was introduced in [15] and that of strong idempotency in [4, 5] as properties of polynomial functions.

We say that a function $f\colon L^n \to L$ has a *componentwise convex range* if either $n = 1$ and f has a convex range, or $n > 1$ and for every $\mathbf{a} = (a_1, \ldots, a_n) \in L^n$ and every $k \in [n]$, the unary function $f_{\mathbf{a}}^k\colon L \to L$, given by $f_{\mathbf{a}}^k(x) = f(\mathbf{a}_k^x)$ has a convex range.

The following theorem reassembles the various characterizations of polynomial functions, established in [4], in the particular case when L is a bounded chain.

Theorem 1. *Let $f\colon L^n \to L$ be a function. The following conditions are equivalent:*

(i) *f is a polynomial function.*

(ii) *f is median decomposable.*

(iii) *f is nondecreasing, strongly idempotent, has a convex range and a componentwise convex range.*

(iv) *f is nondecreasing, $\overline{\mathcal{R}}_f$-min homogeneous, and $\overline{\mathcal{R}}_f$-max homogeneous.*

(v) *f is nondecreasing, $\overline{\mathcal{R}}_f$-min homogeneous, and horizontally $\overline{\mathcal{R}}_f$-maxitive.*

(vi) *f is nondecreasing, horizontally $\overline{\mathcal{R}}_f$-minitive, and $\overline{\mathcal{R}}_f$-max homogeneous.*

(vii) *f is nondecreasing, $\overline{\mathcal{R}}_f$-idempotent, horizontally $\overline{\mathcal{R}}_f$-minitive, and horizontally $\overline{\mathcal{R}}_f$-maxitive.*

3 Representations of polynomial functions

Polynomial functions are known to be exactly those functions which can be represented by formulas in disjunctive and conjunctive normal forms. This fact was first observed by Goodstein [9] who, in fact, showed that each polynomial function $f\colon L^n \to L$ is uniquely determined by its restriction to $\{0, 1\}^n$. For a recent reference, see Rudeanu [17].

In this section we recall and refine some known results concerning normal forms of polynomial functions and, in the special case when L is a chain, we present variant representations given in terms of standard simplices of L^n.

The following three results are due to Goodstein [9].

Corollary 2. *Every polynomial function is completely determined by its restriction to $\{0, 1\}^n$.*

Corollary 3. *A function $g\colon \{0, 1\}^n \to L$ can be extended to a polynomial function $f\colon L^n \to L$ if and only if it is nondecreasing. In this case, the extension is unique.*

Proposition 4. *Let $f\colon L^n \to L$ be a function. The following conditions are equivalent:*

(i) *f is a polynomial function.*

(ii) *There exists $\alpha\colon 2^{[n]} \to L$ such that*
$$f(\mathbf{x}) = \bigvee_{I \subseteq [n]} \big(\alpha(I) \wedge \bigwedge_{i \in I} x_i\big).$$

(iii) *There exists $\beta\colon 2^{[n]} \to L$ such that*
$$f(\mathbf{x}) = \bigwedge_{I \subseteq [n]} \big(\beta(I) \vee \bigvee_{i \in I} x_i\big).$$

We shall refer to the expressions given in (ii) and (iii) of Proposition 4 as the *disjunctive normal form* (DNF) representation and the *conjunctive normal form* (CNF) representation, respectively, of the polynomial function f.

Remark 2. By requiring α and β to be nonconstant functions from $2^{[n]}$ to $\{0, 1\}$ and satisfying $\alpha(\varnothing) = 0$ and $\beta(\varnothing) = 1$, respectively, we obtain the analogue of Proposition 4 for term functions.

ISBN: 978-989-95079-6-8

For each polynomial function $f\colon L^n \to L$, set

$$\mathrm{DNF}(f) = \Big\{\alpha \in L^{2^{[n]}} : f(\mathbf{x}) = \bigvee_{I\subseteq[n]} \big(\alpha(I) \wedge \bigwedge_{i\in I} x_i\big)\Big\},$$

$$\mathrm{CNF}(f) = \Big\{\beta \in L^{2^{[n]}} : f(\mathbf{x}) = \bigwedge_{I\subseteq[n]} \big(\beta(I) \vee \bigvee_{i\in I} x_i\big)\Big\}.$$

A complete description of the sets $\mathrm{DNF}(f)$ and $\mathrm{CNF}(f)$ can be found in [5, §3.1]. As we are concerned by the case when L is a chain, we recall the description only in this special case; see [15, §3].

For each $I \subseteq [n]$, let $\mathbf{e}_I$ be the element of L^n whose ith component is 1, if $i \in I$, and 0, otherwise. Let $\alpha_f\colon 2^{[n]} \to L$ be the function given by $\alpha_f(I) = f(\mathbf{e}_I)$ and consider the function $\alpha_f^*\colon 2^{[n]} \to L$ defined by

$$\alpha_f^*(I) = \begin{cases} \alpha_f(I), & \text{if } \bigvee_{J\subsetneq I} \alpha_f(J) < \alpha_f(I),\\ 0, & \text{otherwise.}\end{cases}$$

Dually, Let $\beta_f\colon 2^{[n]} \to L$ be the function given by $\beta_f(I) = f(\mathbf{e}_{[n]\setminus I})$ and consider the function $\beta_f^*\colon 2^{[n]} \to L$ defined by

$$\beta_f^*(I) = \begin{cases} \beta_f(I), & \text{if } \bigwedge_{J\subsetneq I} \beta_f(J) > \beta_f(I),\\ 1, & \text{otherwise.}\end{cases}$$

Proposition 5. *Let $f\colon L^n \to L$ be a polynomial function. Then*

(i) $\mathrm{DNF}(f) = [\alpha_f^*, \alpha_f]$ and f has a unique DNF representation if and only if $\bigvee_{J\subsetneq I} \alpha_f(J) < \alpha_f(I)$ for every $I \subseteq [n]$,

(ii) $\mathrm{CNF}(f) = [\beta_f, \beta_f^*]$ and f has a unique CNF representation if and only if $\bigwedge_{J\subsetneq I} \beta_f(J) > \beta_f(I)$ for every $I \subseteq [n]$.

In particular, α_f and β_f are the unique isotone and antitone, respectively, maps in $\mathrm{DNF}(f)$ and $\mathrm{CNF}(f)$, respectively.

In the case of chains, the DNF and CNF representations of polynomial functions $f\colon L^n \to L$ can be refined and given in terms of standard simplices of L^n (see Proposition 6 below). Recall that

$$\mathrm{median}(x_1, \ldots, x_{2n+1}) = \bigvee_{\substack{I\subseteq[2n+1]\\ |I|=n+1}} \bigwedge_{i\in I} x_i.$$

Let σ be a permutation on $[n]$. The *standard simplex* of L^n associated with σ is the subset $L_\sigma^n \subset L^n$ defined by

$$L_\sigma^n = \{(x_1, \ldots, x_n) \in L^n : x_{\sigma(1)} \leqslant \cdots \leqslant x_{\sigma(n)}\}.$$

For each $i \in [n]$, define $S_\sigma^{\uparrow}(i) = \{\sigma(i), \ldots, \sigma(n)\}$ and $S_\sigma^{\downarrow}(i) = \{\sigma(1), \ldots, \sigma(i)\}$. As a matter of convenience, set $S_\sigma^{\uparrow}(n+1) = S_\sigma^{\downarrow}(0) = \varnothing$.

Proposition 6. *Let $f\colon L^n \to L$ be a function. The following conditions are equivalent:*

(i) f is a polynomial function.

(ii) For any permutation σ on $[n]$ and every $\mathbf{x} \in L_\sigma^n$, we have

$$\begin{aligned} f(\mathbf{x}) &= \bigvee_{i\in[n+1]} \big(\alpha_f(S_\sigma^{\uparrow}(i)) \wedge x_{\sigma(i)}\big)\\ &= \bigwedge_{i\in[n+1]} \big(\alpha_f(S_\sigma^{\uparrow}(i)) \vee x_{\sigma(i-1)}\big)\\ &= \mathrm{median}\big(x_1, ..., x_n, \alpha_f(S_\sigma^{\uparrow}(1)), ..., \alpha_f(S_\sigma^{\uparrow}(n+1))\big),\end{aligned}$$

where $x_{\sigma(0)} = 0$ and $x_{\sigma(n+1)} = 1$.

(iii) For any permutation σ on $[n]$ and every $\mathbf{x} \in L_\sigma^n$, we have

$$\begin{aligned} f(\mathbf{x}) &= \bigvee_{i\in[n+1]} \big(\beta_f(S_\sigma^{\downarrow}(i-1)) \wedge x_{\sigma(i)}\big)\\ &= \bigwedge_{i\in[n+1]} \big(\beta_f(S_\sigma^{\downarrow}(i-1)) \wedge x_{\sigma(i-1)}\big)\\ &= \mathrm{median}\big(x_1, \ldots, x_n, \beta_f(S_\sigma^{\downarrow}(0)), \ldots, \beta_f(S_\sigma^{\downarrow}(n))\big),\end{aligned}$$

where $x_{\sigma(0)} = 0$ and $x_{\sigma(n+1)} = 1$.

Remark 3. The equivalence between (i) and (ii) of Proposition 6 was already observed in [15, §5]. Prior to this, Propositions 5 and 6 were already established in [14] for idempotent polynomial functions (discrete Sugeno integrals) in the case when L is the unit real interval $[0, 1]$; see also [13, §4.3].

4 Characterizations of polynomial functions

In this section, we propose weak analogues of the properties used in Theorem 1 and provide characterizations of polynomial functions on chains, accordingly. Moreover, we introduce further properties, namely, comonotonic minitivity and maxitivity, which we then use to provide further characterizations of polynomial functions.

For integers $0 \leqslant p \leqslant q \leqslant n$, define $L_n^{(p,q)} = \{\mathbf{x} \in L^n : |\{x_1, \ldots, x_n\} \cap \{0, 1\}| \geqslant p \text{ and } |\{x_1, \ldots, x_n\}| \leqslant q\}$. For instance, $L_n^{(0,2)}$ is the set of Boolean vectors of L^n that are two-sided trimmed by constant vectors, that is

$$L_n^{(0,2)} = \bigcup_{\substack{\mathbf{e}\in\{0,1\}^n\\ c,d\in L}} \{\mathrm{median}(c, \mathbf{e}, d)\},$$

where the median is taken componentwise.

4.1 Weak homogeneity

Let S be a nonempty subset of L. We say that a function $f\colon L^n \to L$ is *weakly S-min homogeneous* (resp. *weakly S-max homogeneous*) if (1) (resp. (2)) holds for every $\mathbf{x} \in L_n^{(0,2)}$ and every $c \in S$.

For every integer $m \geqslant 1$, every $\mathbf{x} \in L^m$, and every $f\colon L^n \to L$, we define $\langle \mathbf{x} \rangle_f \in L^m$ as the m-tuple

$$\langle \mathbf{x} \rangle_f = \mathrm{median}(f(\mathbf{0}), \mathbf{x}, f(\mathbf{1})),$$

where the right-hand side median is taken componentwise. As observed in [5], for every nonempty subset $S \subseteq L$, we have that f is S-min homogeneous and S-max homogeneous if and only if it satisfies

$$f(\mathrm{median}(r, \mathbf{x}, s)) = \mathrm{median}(r, f(\mathbf{x}), s)$$

ISBN: 978-989-95079-6-8

for every $\mathbf{x} \in L^n$ and every $r, s \in S$. In particular, if $f(\mathbf{0}), f(\mathbf{1}) \in S$ then, for any $\mathbf{x} \in L^n$ such that $f(\mathbf{0}) \leqslant f(\mathbf{x}) \leqslant f(\mathbf{1})$, we have $f(\mathbf{x}) = f(\langle \mathbf{x} \rangle_f)$.

It was also shown in [5] that, for every nonempty subset $S \subseteq L$, if f is S-min homogeneous and S-max homogeneous, then it is S-idempotent. The following lemma shows that the weak analogue also holds.

Lemma 7. *Let S be a nonempty subset of L. If $f\colon L^n \to L$ is weakly S-min homogeneous and weakly S-max homogeneous, then it is S-idempotent. Moreover, if $f(\mathbf{0}), f(\mathbf{1}) \in S$ then, for any $\mathbf{x} \in L_n^{(0,2)}$ such that $f(\mathbf{0}) \leqslant f(\mathbf{x}) \leqslant f(\mathbf{1})$, we have $f(\mathbf{x}) = f(\langle \mathbf{x} \rangle_f)$.*

As we are going to see, of particular interest is when $S = \overline{\mathcal{R}}_f$. The following result characterizing the class of polynomial functions shows that, in the case of chains, the conditions in (iv) of Theorem 1 can be replaced with their weak analogues.

Theorem 8. *A function $f\colon L^n \to L$ is a polynomial function if and only if it is nondecreasing, weakly $\overline{\mathcal{R}}_f$-min homogeneous, and weakly $\overline{\mathcal{R}}_f$-max homogeneous.*

Remark 4. Note that Theorem 8 does not generally hold in the case of bounded distributive lattices. To see this, let $L = \{0, a, b, 1\}$ where $a \wedge b = 0$ and $a \vee b = 1$, and consider the binary function $f\colon L^2 \to L$ defined by

$$f(x_1, x_2) = \begin{cases} 1, & \text{if } x_1 = 1 \text{ or } x_2 = 1, \\ 1, & \text{if } x_1 = x_2 = b, \\ a, & \text{if } (x_1 = a \text{ and } x_2 \neq 1) \\ & \text{or } (x_2 = a \text{ and } x_1 \neq 1), \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to verify that f is nondecreasing and both weakly $\overline{\mathcal{R}}_f$-min homogeneous and weakly $\overline{\mathcal{R}}_f$-max homogeneous, but it is not $\overline{\mathcal{R}}_f$-max homogeneous. For instance, $f(a \vee a, b \vee a) = f(a, 1) = 1 \neq a = f(a, b) \vee a$. Thus, f is not a polynomial function by Theorem 1.

4.2 Weak horizontal minitivity and maxitivity

Let S be a nonempty subset of L. We say that a function $f\colon L^n \to L$ is *weakly horizontally S-minitive* (resp. *weakly horizontally S-maxitive*) if (3) (resp. (4)) holds for every $\mathbf{x} \in L_n^{(0,2)}$ and every $c \in S$.

The following result provides characterizations of the n-ary polynomial functions on a chain L, given in terms of weak homogeneity and weak horizontal minitivity and maxitivity.

Theorem 9. *Let $f\colon L^n \to L$ be a function. The following conditions are equivalent:*

(i) *f is a polynomial function.*

(ii) *f is nondecreasing, weakly $\overline{\mathcal{R}}_f$-min homogeneous, and weakly $\overline{\mathcal{R}}_f$-max homogeneous.*

(iii) *f is nondecreasing, weakly $\overline{\mathcal{R}}_f$-min homogeneous, and weakly horizontally $\overline{\mathcal{R}}_f$-maxitive.*

(iv) *f is nondecreasing, weakly horizontally $\overline{\mathcal{R}}_f$-minitive, and weakly $\overline{\mathcal{R}}_f$-max homogeneous.*

(v) *f is nondecreasing, $\overline{\mathcal{R}}_f$-idempotent, weakly horizontally $\overline{\mathcal{R}}_f$-minitive, and weakly horizontally $\overline{\mathcal{R}}_f$-maxitive.*

4.3 Weak median decomposability

As we saw in Theorem 1, in the case of distributive lattices L, the n-ary polynomial functions on L are exactly those which satisfy the median decomposition formula (5). As we are going to see, in the case of chains, this condition can be relaxed by restricting the satisfaction of (5) by a function $f : L^n \to L$ to the vectors of $L_n^{(0,2)} \cup L_n^{(1,3)}$. In the latter case, we say that $f : L^n \to L$ is *weakly median decomposable*.

Lemma 10. *Let $f\colon L^n \to L$ be a nondecreasing function. If f is weakly median decomposable, then it is $\overline{\mathcal{R}}_f$-idempotent.*

Proposition 11. *Let $f\colon L^n \to L$ be a nondecreasing function. The following conditions are equivalent:*

(i) *f is weakly median decomposable.*

(ii) *f weakly $\overline{\mathcal{R}}_f$-min homogeneous and weakly $\overline{\mathcal{R}}_f$-max homogeneous.*

Remark 5. Using the binary function f given in Remark 4, we can see that Proposition 11 does not hold in the general case of bounded distributive lattices. Indeed, as observed, f is nondecreasing and both weakly $\overline{\mathcal{R}}_f$-min homogeneous and weakly $\overline{\mathcal{R}}_f$-max homogeneous, but $f(b, b) = 1 \neq b = \operatorname{median}(f(0, b), b, f(1, b))$ which shows that f is not weakly median decomposable.

From Proposition 11 and Theorem 8, we obtain the following description of polynomial functions given in terms of weak median decomposability.

Theorem 12. *A nondecreasing function $f\colon L^n \to L$ is a polynomial function if and only if it is weakly median decomposable.*

Remark 6. Note that Theorem 12 does not hold if weak median decomposability would have been defined in terms of vectors in $L_n^{(0,2)}$ only. To see this, let $L = \{0, c, 1\}$ and consider the following nondecreasing function $f : L^3 \to L$, defined by

$$f(x_1, x_2, x_3) = \begin{cases} 1, & \text{if } \operatorname{median}(x_1, x_2, x_3) = 1, \\ c, & \text{if } \operatorname{median}(x_1, x_2, x_3) \\ & = x_1 \wedge x_2 \wedge x_3 = c, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to this that f is median decomposable for vectors in $L_n^{(0,2)}$, but it is not a polynomial function, e.g., we have $f(0, c, c) = 0$ but $f(0, 1, 1) \wedge c = c$.

4.4 Strong idempotency and componentwise range convexity

By Theorem 1, a nondecreasing function $f\colon L^n \to L$ is a polynomial function if and only if it is strongly idempotent, has a convex range, and a componentwise convex range. In the case of chains, the condition requiring a convex range becomes redundant, since it becomes a consequence of componentwise range convexity. Thus, we obtain the following characterization, which weakens condition (iii) of Theorem 1 when L is a chain.

ISBN: 978-989-95079-6-8

Theorem 13. *A function $f\colon L^n \to L$ is a polynomial function if and only if it is nondecreasing, strongly idempotent, and has a componentwise convex range.*

Remark 7. None of the conditions provided in Theorem 13 can be dropped off. For instance, let L be the real interval $[0,1]$. Clearly, the unary function $f(x) = x^2$ is nondecreasing and has a componentwise convex range, but it is not strongly idempotent. On the other hand, the function $f\colon L^2 \to L$ defined by

$$f(x_1, x_2) = \begin{cases} 1, & \text{if } x_1 = x_2 = 1, \\ 0, & \text{otherwise,} \end{cases}$$

is nondecreasing and strongly idempotent but it does not have a componentwise convex range, e.g., both $f^1_{\mathbf{1}}$ and $f^2_{\mathbf{1}}$ do not have convex ranges.

In the special case of real interval lattices, i.e., where $L = [a,b]$ for reals $a \leqslant b$, the property of having a convex range, as well as the property of having a componentwise convex range, are consequences of continuity. More precisely, for nondecreasing functions $f : [a,b]^n \to \mathbb{R}$, being continuous reduces to being continuous in each variable, and this latter property is equivalent to having a componentwise convex range. In fact, since polynomial functions are continuous, the condition of having a componentwise convex range can be replaced in Theorem 13 by continuity in each variable. Also, we can add continuity and replace $\overline{\mathcal{R}}_f$ by $\mathcal{R}_f$ in Theorems 8 and 9.

Corollary 14. *Assume that L is a bounded real interval $[a,b]$. A function $f\colon L^n \to L$ is a polynomial function if and only if it is nondecreasing, strongly idempotent, and continuous (in each variable).*

4.5 Comonotonic maxitivity and minitivity

Two vectors $\mathbf{x}$ and $\mathbf{x}'$ in L^n are said to be *comonotonic* if $\mathbf{x}, \mathbf{x}' \in L^n_\sigma$ for some permutation σ on $[n]$. A function $f\colon L^n \to L$ is said to be

- *comonotonic minitive* if, for any two comonotonic vectors $\mathbf{x}, \mathbf{x}' \in L^n$, we have
$$f(\mathbf{x} \wedge \mathbf{x}') = f(\mathbf{x}) \wedge f(\mathbf{x}').$$
- *comonotonic maxitive* if, for any two comonotonic vectors $\mathbf{x}, \mathbf{x}' \in L^n$, we have
$$f(\mathbf{x} \vee \mathbf{x}') = f(\mathbf{x}) \vee f(\mathbf{x}').$$

Note that for any $\mathbf{x} \in L^n$ and any $c \in L$, we have that $\mathbf{x}$ and $(c, \ldots, c)$ are comonotonic and that $\mathbf{x} \vee c$ and $[\mathbf{x}]^c$ are comonotonic. These facts lead to the following result.

Lemma 15. *Let S be a nonempty subset of L. If a function $f\colon L^n \to L$ is comonotonic minitive (resp. comonotonic maxitive), then it is horizontally S-minitive (resp. horizontally S-maxitive). Moreover, if f is S-idempotent, then it is S-min homogeneous (resp. S-max homogeneous).*

Let σ be a permutation on $[n]$. Clearly, every comonotonic minitive (or comonotonic maxitive) function $f\colon L^n \to L$ is nondecreasing on the standard simplex L^n_σ. The following lemma shows that this fact can be extended to the whole domain L^n.

Lemma 16. *If $f\colon L^n \to L$ is comonotonic minitive or comonotonic maxitive, then it is nondecreasing. Furthermore, every nondecreasing unary function is comonotonic minitive and comonotonic maxitive.*

We now have the following characterization of polynomial functions.

Theorem 17. *Let $f\colon L^n \to L$ be a function. The following conditions are equivalent:*

(i) *f is a polynomial function.*

(ii) *f is weakly $\overline{\mathcal{R}}_f$-min homogeneous and comonotonic maxitive.*

(iii) *f is comonotonic minitive and weakly $\overline{\mathcal{R}}_f$-max homogeneous.*

(iv) *f is $\overline{\mathcal{R}}_f$-idempotent, weakly horizontally $\overline{\mathcal{R}}_f$-minitive, and comonotonic maxitive.*

(v) *f is $\overline{\mathcal{R}}_f$-idempotent, comonotonic minitive, and weakly horizontally $\overline{\mathcal{R}}_f$-maxitive.*

(vi) *f is $\overline{\mathcal{R}}_f$-idempotent, comonotonic minitive, and comonotonic maxitive.*

Remark 8. (i) As already observed in the remark following Theorem 9, the weak horizontal $\overline{\mathcal{R}}_f$-minitivity (resp. weak horizontal $\overline{\mathcal{R}}_f$-maxitivity) can be replaced with weak horizontal L-minitivity (resp. weak horizontal L-maxitivity) in the assertions (iv)–(v) of Theorem 17.

(ii) The condition requiring $\overline{\mathcal{R}}_f$-idempotency is necessary in conditions (iv)–(vi) of Theorem 17. For instance, let L be the unit interval $[0,1]$. Clearly, the unary function $f(x) = x^2$ is nondecreasing and thus comonotonic minitive and comonotonic maxitive. By Lemma 15, it is also horizontally $\overline{\mathcal{R}}_f$-minitive and horizontally $\overline{\mathcal{R}}_f$-maxitive. However, it is not a polynomial function.

(iii) The concept of comonotonic vectors appeared as early as 1952 in Hardy et al. [12]. Comonotonic minitivity and maxitivity were introduced in the context of Sugeno integrals in de Campos et al. [7]. An interpretation of these properties was given by Ralescu and Ralescu [16] in the framework of aggregation of fuzzy subsets.

5 Discrete Sugeno integrals

In this final section, we consider a noteworthy subclass of polynomial functions, namely, that of discrete Sugeno integrals, and provide its characterizations.

A function $f\colon L^n \to L$ is said to be *idempotent* if it is L-idempotent.

Fact 18. *A polynomial function is $\{0,1\}$-idempotent if and only if it is idempotent.*

In [15, §4], $\{0,1\}$-idempotent polynomial functions are referred to as *discrete Sugeno integrals*. They coincide exactly with those functions $\mathcal{S}_\mu\colon L^n \to L$ for which there is a fuzzy measure μ such that

$$\mathcal{S}_\mu(\mathbf{x}) = \bigvee_{I \subseteq [n]} \big(\mu(I) \wedge \bigwedge_{i \in I} x_i\big).$$

ISBN: 978-989-95079-6-8

(A *fuzzy measure* μ is simply a set function $\mu\colon 2^{[n]} \to L$ satisfying $\mu(I) \leqslant \mu(I')$ whenever $I \subseteq I'$, and $\mu(\varnothing) = 0$ and $\mu([n]) = 1$.)

The following proposition shows how polynomial functions relate to Sugeno integrals; see [15, Proposition 12].

Proposition 19. *For any polynomial function $f\colon L^n \to L$ there is a fuzzy measure $\mu\colon 2^{[n]} \to L$ such that $f(\mathbf{x}) = \langle \mathcal{S}_\mu(\mathbf{x}) \rangle_f$.*

We say that a function $f\colon L^n \to L$ is *Boolean min homogeneous* (resp. *Boolean max homogeneous*) if (1) (resp. (2)) holds for every $\mathbf{x} \in \{0,1\}^n$ and every $c \in L$. Note that every weakly L-min homogeneous (resp. weakly L-max homogeneous) function is Boolean min homogeneous (resp. Boolean max homogeneous). The following result provides a variant of Theorem 8.

Theorem 20. *A function $f\colon L^n \to L$ is a discrete Sugeno integral if and only if it is nondecreasing, Boolean min homogeneous, and Boolean max homogeneous.*

Remark 9. (i) Theorem 20 as well as the characterization of the discrete Sugeno integrals obtained by combining $\{0,1\}$-idempotency with (vi) in Theorem 17 were presented in the case of real variables by [13, §4.3]; see also [14].

(ii) Even though Theorem 20 can be derived from condition (ii) of Theorem 9 by simply modifying the two homogeneity properties, to proceed similarly with conditions (iii) and (iv), it is necessary to add the conditions of $\{1\}$-idempotency and $\{0\}$-idempotency, respectively (and similarly with conditions (ii) and (iii) of Theorem 17). To see this, let L be a bounded chain with at least three elements and consider the unary functions $f(x) = x \wedge d$ and $g(x) = x \vee d$, where $d \in L \setminus \{0,1\}$. Clearly, f is L-min homogeneous and horizontally L-maxitive and g is L-max homogeneous and horizontally L-minitive. However, neither f nor g is a discrete Sugeno integral. To see that these additions are sufficient, just note that L-min homogeneity (resp. L-max homogeneity) implies $\{0\}$-idempotency (resp. $\{1\}$-idempotency).

(iii) Marichal [13, §2.2.3] showed that, when L is a chain, a nondecreasing and idempotent function $f\colon L^n \to L$ is Boolean min homogeneous (resp. Boolean max homogeneous) if and only if we have $f(\mathbf{e} \wedge c) \in \{f(\mathbf{e}), c\}$ (resp. $f(\mathbf{e} \vee c) \in \{f(\mathbf{e}), c\}$) for every $\mathbf{e} \in \{0,1\}^n$ and every $c \in L$.

References

[1] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners*. Studies in Fuziness and Soft Computing. Springer, Berlin, 2007.

[2] P. Benvenuti, R. Mesiar, and D. Vivona. Monotone set functions-based integrals. In *Handbook of measure theory, Vol. I, II*, pages 1329–1379. North-Holland, Amsterdam, 2002.

[3] S. Burris and H. P. Sankappanavar. *A course in universal algebra*, volume 78 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1981.

[4] M. Couceiro and J.-L. Marichal. Polynomial functions over bounded distributive lattices. `http://arxiv.org/abs/0901.4888`

[5] M. Couceiro and J.-L. Marichal. Characterizations of discrete Sugeno integrals as polynomial functions over distributive lattices. `http://arxiv.org/abs/0808.2619`

[6] L. M. de Campos and M. J. Bolaños. Characterization and comparison of Sugeno and Choquet integrals. *Fuzzy Sets and Systems*, 52(1):61–67, 1992.

[7] L. M. de Campos, M. T. Lamata, and S. Moral. A unified approach to define fuzzy integrals. *Fuzzy Sets and Systems*, 39(1):75–90, 1991.

[8] J. C. Fodor and M. Roubens. Characterization of weighted maximum and some related operations. *Inform. Sci.*, 84(3-4):173–180, 1995.

[9] R. L. Goodstein. The solution of equations in a lattice. *Proc. Roy. Soc. Edinburgh Sect. A*, 67:231–242, 1965/1967.

[10] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation functions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, UK, 2009.

[11] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy measures and integrals*, volume 40 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2000.

[12] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge, at the University Press, 1952. 2d ed.

[13] J.-L. Marichal. *Aggregation operators for multicriteria decision aid*. PhD thesis, Institute of Mathematics, University of Liège, Liège, Belgium, December 1998.

[14] J.-L. Marichal. On Sugeno integral as an aggregation function. *Fuzzy Sets and Systems*, 114(3):347–365, 2000.

[15] J.-L. Marichal. Weighted lattice polynomials. *Discrete Mathematics*, 309(4):814–820, 2009.

[16] A. L. Ralescu and D. A. Ralescu. Extensions of fuzzy aggregation. *Fuzzy Sets and Systems*, 86(3):321–330, 1997.

[17] S. Rudeanu. *Lattice functions and equations*. Springer Series in Discrete Mathematics and Theoretical Computer Science. Springer-Verlag London Ltd., London, 2001.

[18] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, Tokyo, 1974.

[19] M. Sugeno. Fuzzy measures and fuzzy integrals—a survey. In M. M. Gupta, G. N. Saridis, and B. R. Gaines, editors, *Fuzzy automata and decision processes*, pages 89–102. North-Holland, New York, 1977.

ISBN: 978-989-95079-6-8

Quasi-polynomial functions on bounded chains

Miguel Couceiro Jean-Luc Marichal

Mathematics Research Unit, University of Luxembourg
162A, avenue de la Faïencerie, L-1511 Luxembourg, Luxembourg
Email: {miguel.couceiro, jean-luc.marichal}@uni.lu

Abstract— *Two emergent properties in aggregation theory are investigated, namely horizontal maxitivity and comonotonic maxitivity (as well as their dual counterparts) which are commonly defined by means of certain functional equations. We present complete descriptions of the function classes axiomatized by each of these properties, up to weak versions of monotonicity, in the cases of horizontal maxitivity and minitivity. While studying the classes axiomatized by combinations of these properties, we introduce the concept of quasi-polynomial function which appears as a natural extension of the well-established notion of polynomial function. We present further axiomatizations for this class both in terms of functional equations and natural relaxations of homogeneity and median decomposability. As noteworthy particular cases, we investigate those subclasses of quasi-term functions and quasi-weighted maximum and minimum functions, and present characterizations accordingly.*

Keywords— Discrete Sugeno integral, quasi-polynomial function, horizontal maxitivity and minitivity, comonotonic maxitivity and minitivity, functional equation.

1 Introduction

Aggregation functions arise wherever aggregating information is important: applied and pure mathematics (probability, statistics, decision theory, functional equations), operations research, computer science, and many applied fields (economics and finance, pattern recognition and image processing, data fusion, etc.). For recent references, see Beliakov et al. [1] and Grabisch et al. [10].

A noteworthy aggregation function is the so-called discrete Sugeno integral, which was introduced by Sugeno [15, 16] and which has been widely investigated in aggregation theory, due to its many applications for instance in fuzzy set theory, decision making, and image analysis. For general background, see also the edited book [11].

A convenient way to introduce the discrete Sugeno integral is via the concept of (lattice) polynomial functions, i.e., functions which can be expressed as combinations of variables and constants using the lattice operations $\wedge$ and $\vee$. As shown by Marichal [13], the discrete Sugeno integrals are exactly those polynomial functions $f : L^n \to L$ which are idempotent, that is, satisfying $f(x, \ldots, x) = x$. Several axiomatizations of the class of discrete Sugeno integrals (as idempotent polynomial functions) have been recently given; see [4].

Of particular interest in aggregation theory, are the so-called horizontal maxitivity and comonotonic maxitivity (as well as their dual counterparts), usually expressed in terms of certain functional equations, and which we now informally describe.

Let L be a bounded chain. For every $\mathbf{x} \in L^n$ and every $c \in L$, consider the horizontal maxitive decomposition of $\mathbf{x}$ obtained by "cutting" it with c, namely $\mathbf{x} = (\mathbf{x} \wedge c) \vee [\mathbf{x}]_c$, where $[\mathbf{x}]_c$ is the n-tuple whose ith component is 0, if $x_i \leqslant c$, and x_i, otherwise. A function $f\colon L^n \to L$ is said to be *horizontally maxitive* if

$$f(\mathbf{x}) = f(\mathbf{x} \wedge c) \vee f([\mathbf{x}]_c)$$

for every $\mathbf{x} \in L^n$ and every $c \in L$.

A function $f\colon L^n \to L$ is said to be *comonotonic maxitive* if, for any two vectors $\mathbf{x}$ and $\mathbf{x}'$ in the same standard simplex of L^n, we have

$$f(\mathbf{x} \vee \mathbf{x}') = f(\mathbf{x}) \vee f(\mathbf{x}').$$

As we are going to see (Lemma 6 below), these (as well as their duals) are closely related and constitute properties shared by discrete Sugeno integrals. Still, and as it will become evident, no combination of these with their dual forms suffices to fully describe the class of Sugeno integrals. Thus, and given their emergence in aggregation theory, it is natural to ask which classes of functions are axiomatized by combinations of these properties or, in fact, by each of these properties.

In this paper, we answer this question for both the maxitive and minitive comonotonic properties, and for horizontal maxitivity and minitivity properties, up to certain weak variants of monotonicity. While looking at combinations of the latter properties, we reach a natural generalization of polynomial functions, which we call *quasi-polynomial functions* and which are best described by the following equation

$$f(x_1, \ldots, x_n) = p(\varphi(x_1), \ldots, \varphi(x_n)),$$

where p is a polynomial function and φ a nondecreasing function (see Theorem 10 below). Searching for alternative descriptions, we introduce weaker versions of well-established properties, such as homogeneity and median decomposability, to provide further axiomatizations of the class of quasi-polynomial functions, accordingly.

This paper is organized as follows. We start by recalling basic notions and terminology in lattice function theory, as well as present some known results, needed throughout this paper (Section 2). In Section 3, we study the properties of horizontal maxitivity and comonotonic maxitivity, as well as their dual forms, and determine those function classes axiomatized by each of this properties. Combinations of the latter are then considered in Section 4.1, where the notion of quasi-polynomial function is introduced. In Section 4.2, we propose weaker versions of homogeneity and median decomposability, and provide further characterizations of quasi-polynomial functions, accordingly. In Section 5, we introduce and axiomatize few noteworthy subclasses of quasi-polynomial functions, namely, those of quasi-term functions and those quasi-weighted maximum and minimum functions.

ISBN: 978-989-95079-6-8

2 Basic notions and preliminary results

In this section we recall basic terminology as well as some results needed in the current paper. For general background we refer the reader to, e.g., Burris and Sankappanavar [3] and Rudeanu [14].

2.1 General background

Throughout this paper, let L be a bounded chain with operations $\wedge$ and $\vee$, and with least and greatest elements 0 and 1, respectively. A subset S of a chain L is said to be *convex* if for every $a, b \in S$ and every $c \in L$ such that $a \leqslant c \leqslant b$, we have $c \in S$. For any subset $S \subseteq L$, we denote by $\overline{S}$ the convex hull of S, that is, the smallest convex subset of L containing S. For every $a, b \in S$ such that $a \leqslant b$, the *interval* $[a, b]$ is the set $[a, b] = \{c \in L : a \leqslant c \leqslant b\}$. For any integer $n \geqslant 1$, let $[n] = \{1, \ldots, n\}$.

For any bounded chain L, we regard the Cartesian product L^n, $n \geqslant 1$, as a distributive lattice endowed with the operations $\wedge$ and $\vee$ given by

$$\begin{aligned}(a_1, \ldots, a_n) \wedge (b_1, \ldots, b_n) &= (a_1 \wedge b_1, \ldots, a_n \wedge b_n),\\ (a_1, \ldots, a_n) \vee (b_1, \ldots, b_n) &= (a_1 \vee b_1, \ldots, a_n \vee b_n).\end{aligned}$$

The elements of L are denoted by lower case letters $a, b, c, \ldots$, and the elements of L^n, $n > 1$, by bold face letters $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$. We also use $\mathbf{0}$ and $\mathbf{1}$ to denote the least element and greatest element, respectively, of L^n. For $c \in L$ and $\mathbf{x} = (x_1, \ldots, x_n) \in L^n$, set

$$\mathbf{x} \wedge c = (x_1 \wedge c, \ldots, x_n \wedge c) \quad \text{and} \quad \mathbf{x} \vee c = (x_1 \vee c, \ldots, x_n \vee c).$$

The *range* of a function $f\colon L^n \to L$ is defined by $\mathcal{R}_f = \{f(\mathbf{x}) : \mathbf{x} \in L^n\}$. A function $f\colon L^n \to L$ is said to be *nondecreasing* (*in each variable*) if, for every $\mathbf{a}, \mathbf{b} \in L^n$ such that $\mathbf{a} \leqslant \mathbf{b}$, we have $f(\mathbf{a}) \leqslant f(\mathbf{b})$. The *diagonal section* of f, denoted δ_f, is defined as the unary function given by $\delta_f(x) = f(x, \ldots, x)$. Note that if f is nondecreasing, then δ_f is nondecreasing and $\overline{\mathcal{R}}_{\delta_f} = \overline{\mathcal{R}}_f = [f(\mathbf{0}), f(\mathbf{1})]$.

2.2 Polynomial functions and their representations

In this paper the so-called polynomial functions will play a fundamental role. Formally, an *n-ary polynomial function* on L is any function $f : L^n \to L$ which can be obtained by finitely many applications of the following rules:

(i) For each $i \in [n]$ and each $c \in L$, the projection $\mathbf{x} \mapsto x_i$ and the constant function $\mathbf{x} \mapsto c$ are polynomial functions from L^n to L.

(ii) If f and g are polynomial functions from L^n to L, then $f \vee g$ and $f \wedge g$ are polynomial functions from L^n to L.

Polynomial functions are also called lattice functions (Goodstein [9]), algebraic functions (Burris and Sankappanavar [3]) or weighted lattice polynomial functions (Marichal [13]). Idempotent polynomial functions (i.e., satisfying $f(c, \ldots, c) = c$ for every $c \in L$) are referred to by aggregation theorists as (*discrete*) *Sugeno integrals*, and those obtained from projections by finitely many applications of (ii) are usually referred to as (*lattice*) *term functions*. A noteworthy term function that we shall make use in the sequel is the median function, defined by

$$\operatorname{median}(x, y, z) = (x \wedge y) \vee (y \wedge z) \vee (z \wedge x).$$

As observed by Goodstein [9] (see also Rudeanu [14]), polynomial functions are exactly those functions which can be represented by formulas in disjunctive and conjunctive normal forms. In fact, each polynomial function $f\colon L^n \to L$ is uniquely determined by its restriction to $\{0, 1\}^n$. Due to their relevance in the sequel, we recall some known results concerning normal form representations of polynomial functions in the special case where L is a chain. The following result is due to Goodstein [9].

Proposition 1. *(a) Every polynomial function is completely determined by its restriction to $\{0, 1\}^n$.*

(b) A function $g\colon \{0, 1\}^n \to L$ can be extended to a polynomial function $f\colon L^n \to L$ if and only if it is nondecreasing. In this case, the extension is unique.

(c) For any $f\colon L^n \to L$, the following are equivalent:

(i) f is a polynomial function.

(ii) There exists $\alpha\colon 2^{[n]} \to L$ such that

$$f(\mathbf{x}) = \bigvee_{I \subseteq [n]} \big(\alpha(I) \wedge \bigwedge_{i \in I} x_i\big). \tag{1}$$

(iii) There exists $\beta\colon 2^{[n]} \to L$ such that

$$f(\mathbf{x}) = \bigwedge_{I \subseteq [n]} \big(\beta(I) \vee \bigvee_{i \in I} x_i\big). \tag{2}$$

The expressions given in (1) and (2) are usually referred to as the *disjunctive normal form* (DNF) representation and the *conjunctive normal form* (CNF) representation, respectively, of the polynomial function f.

Remark 1. By requiring α and β to be nonconstant functions from $2^{[n]}$ to $\{0, 1\}$ and satisfying $\alpha(\varnothing) = 0$ and $\beta(\varnothing) = 1$, respectively, we obtain the analogue of (c) of Proposition 1 for term functions.

As observed by Marichal [13], the DNF and CNF representations of polynomial functions $f\colon L^n \to L$ are not necessarily unique. For instance, we have

$$x_1 \vee (x_1 \wedge x_2) = x_1 = x_1 \wedge (x_1 \vee x_2).$$

However, from among all the possible set functions α (resp. β) defining the DNF (resp. CNF) representation of f, only one is isotone (resp. antitone), namely the function $\alpha_f\colon 2^{[n]} \to L$ (resp. $\beta_f\colon 2^{[n]} \to L$) defined by

$$\alpha_f(I) = f(\mathbf{e}_I) \qquad (\text{resp. } \beta_f(I) = f(\mathbf{e}_{[n] \setminus I})), \tag{3}$$

where $\mathbf{e}_I$ denotes the element of $\{0, 1\}^n$ whose ith component is 1 if and only if $i \in I$.

In the case when L is a chain, it was shown in [4] that the DNF and CNF representations of polynomial functions $f\colon L^n \to L$ can be refined and given in terms of standard simplices of L^n. Let σ be a permutation on $[n]$. The *standard simplex* of L^n associated with σ is the subset $L^n_\sigma \subset L^n$ defined by

$$L^n_\sigma = \{(x_1, \ldots, x_n) \in L^n : x_{\sigma(1)} \leqslant x_{\sigma(2)} \leqslant \cdots \leqslant x_{\sigma(n)}\}.$$

For each $i \in [n]$, define $S^{\uparrow}_\sigma(i) = \{\sigma(i), \ldots, \sigma(n)\}$ and $S^{\downarrow}_\sigma(i) = \{\sigma(1), \ldots, \sigma(i)\}$. As a matter of convenience, set $S^{\uparrow}_\sigma(n+1) = S^{\downarrow}_\sigma(0) = \varnothing$.

ISBN: 978-989-95079-6-8

Proposition 2. *For any function $f\colon L^n \to L$, the following conditions are equivalent:*

(i) *f is a polynomial function.*

(ii) *For any permutation σ on $[n]$ and every $\mathbf{x} \in L^n_\sigma$, we have*

$$f(\mathbf{x}) = \bigvee_{i=1}^{n+1} \big(\alpha_f(S^{\uparrow}_{\sigma}(i)) \wedge x_{\sigma(i)}\big),$$

where $x_{\sigma(n+1)} = 1$.

(iii) *For any permutation σ on $[n]$ and every $\mathbf{x} \in L^n_\sigma$, we have*

$$f(\mathbf{x}) = \bigwedge_{i=0}^{n} \big(\beta_f(S^{\downarrow}_{\sigma}(i)) \vee x_{\sigma(i)}\big),$$

where $x_{\sigma(0)} = 0$.

3 Motivating characterizations

Even though horizontal maxitivity and comonotonic maxitivity, as well as their dual counterparts, play an important role in aggregation theory (as properties shared by noteworthy classes of aggregation functions), they have not yet been described independently. In this section we investigate each of these properties and determine their corresponding function classes (up to weak versions of monotonicity, in the cases of horizontal maxitivity and minitivity).

3.1 Horizontal maxitivity and minitivity

Recall that a function $f\colon L^n \to L$ is said to be

- *horizontally maxitive* if, for every $\mathbf{x} \in L^n$ and every $c \in L$, we have

$$f(\mathbf{x}) = f(\mathbf{x} \wedge c) \vee f([\mathbf{x}]_c),$$

where $[\mathbf{x}]_c$ is the n-tuple whose ith component is 0, if $x_i \leqslant c$, and x_i, otherwise.

- *horizontally minitive* if, for every $\mathbf{x} \in L^n$ and every $c \in L$, we have

$$f(\mathbf{x}) = f(\mathbf{x} \vee c) \wedge f([\mathbf{x}]^c),$$

where $[\mathbf{x}]^c$ is the n-tuple whose ith component is 1, if $x_i \geqslant c$, and x_i, otherwise.

Let us consider the following weak forms of nondecreasing monotonicity:

$(\mathbf{P_1})$ $f(\mathbf{e} \wedge c) \leqslant f(\mathbf{e}' \wedge c)$ for every $\mathbf{e}, \mathbf{e}' \in \{0,1\}^n$ such that $\mathbf{e} \leqslant \mathbf{e}'$ and every $c \in L$.

$(\mathbf{D_1})$ $f(\mathbf{e} \vee c) \leqslant f(\mathbf{e}' \vee c)$ for every $\mathbf{e}, \mathbf{e}' \in \{0,1\}^n$ such that $\mathbf{e} \leqslant \mathbf{e}'$ and every $c \in L$.

$(\mathbf{P_2})$ $f(\mathbf{e} \wedge c) \leqslant f(\mathbf{e} \wedge c')$ for every $\mathbf{e} \in \{0,1\}^n$ and every $c, c' \in L$ such that $c \leqslant c'$.

$(\mathbf{D_2})$ $f(\mathbf{e} \vee c) \leqslant f(\mathbf{e} \vee c')$ for every $\mathbf{e} \in \{0,1\}^n$ and every $c, c' \in L$ such that $c \leqslant c'$.

Theorem 3. *A function $f\colon L^n \to L$ is horizontally maxitive and satisfies $\mathbf{P_1}$ if and only if there exists $g\colon L^n \to L$ satisfying $\mathbf{P_2}$ such that*

$$f(\mathbf{x}) = \bigvee_{I \subseteq [n]} g\Big(\mathbf{e}_I \wedge \bigwedge_{i \in I} x_i\Big).$$

In this case, we can choose $g = f$.

Similarly, we obtain the following dual characterization:

Theorem 4. *A function $f\colon L^n \to L$ is horizontally minitive and satisfies $\mathbf{D_1}$ if and only if there exists $g\colon L^n \to L$ satisfying $\mathbf{D_2}$ such that*

$$f(\mathbf{x}) = \bigwedge_{I \subseteq [n]} g\Big(\mathbf{e}_{[n]\setminus I} \vee \bigvee_{i \in I} x_i\Big).$$

In this case, we can choose $g = f$.

From Theorems 3 and 4 we have the following corollary.

Corollary 5. *A function $f\colon L^n \to L$ is horizontally maxitive (resp. horizontally minitive) and satisfies $\mathbf{P_1}$ (resp. $\mathbf{D_1}$) if and only if there are unary nondecreasing functions $\varphi_I\colon L \to L$, for $I \subseteq [n]$, such that*

$$\begin{aligned} f(\mathbf{x}) &= \bigvee_{I \subseteq [n]} \big(\alpha_f(I) \wedge \bigwedge_{i \in I} \varphi_I(x_i)\big) \\ \text{(resp. } f(\mathbf{x}) &= \bigwedge_{I \subseteq [n]} \big(\beta_f(I) \vee \bigvee_{i \in I} \varphi_I(x_i)\big)\text{)}, \end{aligned}$$

where the set function α_f (resp. β_f) is defined in (3). In this case, we can choose $\varphi_I(x) = f(\mathbf{e}_I \wedge x)$ (resp. $\varphi_I(x) = f(\mathbf{e}_{[n]\setminus I} \vee x)$) for every $I \subseteq [n]$.

Remark 2. (i) Theorem 3 (resp. Theorem 4) provides the description of those horizontally maxitive (resp. horizontally minitive) functions which are nondecreasing.

(ii) Every Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ satisfying $f(\mathbf{0}) \leqslant f(\mathbf{x})$ (resp. $f(\mathbf{x}) \leqslant f(\mathbf{1})$) is horizontally maxitive (resp. horizontally minitive). Moreover, not all such functions are nondecreasing, thus showing that condition $\mathbf{P_1}$ (resp. $\mathbf{D_1}$) is necessary in Theorem 3 (resp. Theorem 4).

(iii) As shown in [4], polynomial functions $f\colon L^n \to L$ are exactly those $\overline{\mathcal{R}}_f$-idempotent (i.e., satisfying $f(c, \ldots, c) = c$ for every $c \in \overline{\mathcal{R}}_f$) which are nondecreasing, horizontally maxitive, and horizontally minitive.

(iv) The concept of horizontal maxitivity was introduced, in the case when L is the real interval $[0,1]$, by Benvenuti et al. [2] as a general property of the Sugeno integral.

3.2 Comonotonic maxitivity and minitivity

Two vectors $\mathbf{x}, \mathbf{x}' \in L^n$ are said to be *comonotonic* if there exists a permutation σ on $[n]$ such that $\mathbf{x}, \mathbf{x}' \in L^n_\sigma$. A function $f\colon L^n \to L$ is said to be

- *comonotonic maxitive* if, for any two comonotonic vectors $\mathbf{x}, \mathbf{x}' \in L^n$, we have

$$f(\mathbf{x} \vee \mathbf{x}') = f(\mathbf{x}) \vee f(\mathbf{x}').$$

ISBN: 978-989-95079-6-8

- *comonotonic minitive* if, for any two comonotonic vectors $\mathbf{x}, \mathbf{x}' \in L^n$, we have

$$f(\mathbf{x} \wedge \mathbf{x}') = f(\mathbf{x}) \wedge f(\mathbf{x}').$$

Note that for any $\mathbf{x} \in L^n$ and any $c \in L$, the vectors $\mathbf{x} \vee c$ and $[\mathbf{x}]^c$ are comonotonic. As a consequence, if a function $f\colon L^n \to L$ is comonotonic maxitive (resp. comonotonic minitive), then it is horizontally maxitive (resp. horizontally minitive). It was also observed in [4] that if f is comonotonic maxitive or comonotonic minitive, then it is nondecreasing. Moreover, we obtain the following result.

Lemma 6. *A function $f\colon L^n \to L$ is comonotonic maxitive (resp. comonotonic minitive) if and only if it is horizontally maxitive (resp. horizontally minitive) and satisfies $\mathbf{P_1}$ (resp. $\mathbf{D_1}$).*

Combining Theorems 3 and 4 with Lemma 6, we immediately obtain the descriptions of the classes of comonotonic maxitive and comonotonic minitive functions.

Theorem 7. *A function $f\colon L^n \to L$ is comonotonic maxitive if and only if there exists $g\colon L^n \to L$ satisfying $\mathbf{P_2}$ such that*

$$f(\mathbf{x}) = \bigvee_{I \subseteq [n]} g\Big(\mathbf{e}_I \wedge \bigwedge_{i \in I} x_i\Big).$$

In this case, we can choose $g = f$.

Theorem 8. *A function $f\colon L^n \to L$ is comonotonic minitive if and only if there exists $g\colon L^n \to L$ satisfying $\mathbf{D_2}$ such that*

$$f(\mathbf{x}) = \bigwedge_{I \subseteq [n]} g\Big(\mathbf{e}_{[n]\setminus I} \vee \bigvee_{i \in I} x_i\Big).$$

In this case, we can choose $g = f$.

As before, we have the following corollary.

Corollary 9. *A function $f\colon L^n \to L$ is comonotonic maxitive (resp. comonotonic minitive) if and only if there are unary nondecreasing functions $\varphi_I\colon L \to L$, for $I \subseteq [n]$, such that*

$$\begin{aligned} f(\mathbf{x}) &= \bigvee_{I \subseteq [n]} \big(\alpha_f(I) \wedge \bigwedge_{i \in I} \varphi_I(x_i)\big) \\ \text{(resp. } f(\mathbf{x}) &= \bigwedge_{I \subseteq [n]} \big(\beta_f(I) \vee \bigvee_{i \in I} \varphi_I(x_i)\big)\text{)}, \end{aligned}$$

where the set function α_f (resp. β_f) is defined in (3). In this case, we can choose $\varphi_I(x) = f(\mathbf{e}_I \wedge x)$ (resp. $\varphi_I(x) = f(\mathbf{e}_{[n]\setminus I} \vee x)$) for every $I \subseteq [n]$.

Remark 3. (i) An alternative description of comonotonic maxitive (resp. comonotonic minitive) functions was obtained in Grabisch et al. [10, §2.5] in the case when L is a real interval.

(ii) It was shown in [4] that polynomial functions $f\colon L^n \to L$ are exactly those $\overline{\mathcal{R}}_f$-idempotent functions which are comonotonic maxitive and comonotonic minitive.

(ii) Comonotonic minitivity and maxitivity were introduced in the context of Sugeno integrals in de Campos et al. [5].

4 Quasi-polynomial functions

Motivated by the results of Section 3 concerning horizontal maxitivity and comonotonic maxitivity, as well as their dual counterparts, we now study combinations of these properties. This will lead to a relaxation of the notion of polynomial function, which we will refer to as *quasi-polynomial function*. Accordingly, we introduce weaker variants of well-established properties, such as homogeneity and median decomposability, which are then used to provide further axiomatizations of the class of quasi-polynomial functions.

4.1 Motivation and definition

We start by looking at combinations of those properties studied in Section 3. These are considered in the following result.

Theorem 10. *Let $f\colon L^n \to L$ be a function. The following assertions are equivalent:*

(i) f is horizontally maxitive, horizontally minitive, and satisfies $\mathbf{P_1}$ or $\mathbf{D_1}$.

(ii) f is comonotonic maxitive and comonotonic minitive.

(iii) f is horizontally maxitive and comonotonic minitive.

(iv) f is comonotonic maxitive and horizontally minitive.

(v) There exist a polynomial function $p\colon L^n \to L$ and a nondecreasing function $\varphi\colon L \to L$ such that

$$f(x_1, \ldots, x_n) = p(\varphi(x_1), \ldots, \varphi(x_n)).$$

If these conditions hold then we can choose for p the unique polynomial function p_f extending $f|_{\{0,1\}^n}$ and for φ the diagonal section δ_f of f.

Theorem 10 motivates the following definition.

Definition 11. We say that a function $f\colon L^n \to L$ is a *quasi-polynomial function* (resp. a *discrete quasi-Sugeno integral*, a *quasi-term function*) if there exist a polynomial function (resp. a discrete Sugeno integral, a term function) $p\colon L^n \to L$ and a nondecreasing function $\varphi\colon L \to L$ such that $f = p \circ \varphi$, that is,

$$f(x_1, \ldots, x_n) = p(\varphi(x_1), \ldots, \varphi(x_n)). \tag{4}$$

Remark 4. (i) Note that each quasi-polynomial function $f\colon L^n \to L$ can be represented as a combination of constants and a nondecreasing unary function φ (applied to the projections $\mathbf{x} \mapsto x_i$) using the lattice operations $\vee$ and $\wedge$.

(ii) In the setting of decision-making under uncertainty, the nondecreasing function φ in (4) can be thought of as a *utility function* and the corresponding quasi-polynomial function as a (qualitative) *global preference functional*; see for instance Dubois et al. [6].

Note that the functions p and φ in (4) are not necessarily unique. For instance, if f is a constant $c \in L$, then we could choose $p \equiv c$ and φ arbitrarily, or p idempotent and $\varphi \equiv c$. We now describe all possible choices for p and φ. For any integers $m, n \geqslant 1$, any vector $\mathbf{x} \in L^m$, and any function $f\colon L^n \to L$, we define $\langle \mathbf{x} \rangle_f \in L^m$ as the m-tuple

$$\langle \mathbf{x} \rangle_f = \operatorname{median}(f(\mathbf{0}), \mathbf{x}, f(\mathbf{1})),$$

where the right-hand side median is taken componentwise.

ISBN: 978-989-95079-6-8

Proposition 12. *Let $f\colon L^n \to L$ be a quasi-polynomial function and let $p_f : L^n \to L$ be the unique polynomial function extending $f|_{\{0,1\}^n}$. We have*

$$\{(p,\varphi)\colon f = p \circ \varphi\} = \{(p,\varphi)\colon p_f = \langle p\rangle_f \text{ and } \delta_f = \langle\varphi\rangle_p\},$$

where p and φ stand for polynomial and unary nondecreasing functions, respectively. In particular, we have $f = p_f \circ \delta_f$.

It was shown in Marichal [13] that every polynomial function $p\colon L^n \to L$ can be represented as $\langle q\rangle_p$ for some discrete Sugeno integral $q\colon L^n \to L$. Combining this with Proposition 12, we obtain the next result.

Corollary 13. *The class of quasi-polynomial functions is exactly the class of discrete quasi-Sugeno integrals.*

4.2 Further axiomatizations

We now recall some properties of polynomial functions, namely homogeneity and median decomposability, and we propose weaker variants of these to provide alternative axiomatizations of the class of quasi-polynomial functions.

4.2.1 Quasi-homogeneity

Let S be a subset of L. A function $f\colon L^n \to L$ is said to be *S-max homogeneous* (resp. *S-min homogeneous*) if for every $\mathbf{x} \in L^n$ and every $c \in S$, we have

$$f(\mathbf{x} \vee c) = f(\mathbf{x}) \vee c \qquad (\text{resp. } f(\mathbf{x} \wedge c) = f(\mathbf{x}) \wedge c).$$

Although polynomial functions $p\colon L^n \to L$ share both of these properties for any $S \subseteq \overline{\mathcal{R}}_p$, this is not the case for quasi-polynomial functions. For instance, let $f_1, f_2\colon [0,1] \to [0,1]$ be respectively given by $f_1(x) = x^2$ and $f_2(x) = \sqrt{x}$. Clearly, f_1 and f_2 are quasi-polynomial functions but, e.g., for $x = c$, we have

$$f_1(x \vee c) < f_1(x) \vee c \quad \text{and} \quad f_2(x \wedge c) > f_2(x) \wedge c.$$

This example motivates the following relaxations. We say that a function $f\colon L^n \to L$ is *quasi-max homogeneous* (resp. *quasi-min homogeneous*) if for every $\mathbf{x} \in L^n$ and $c \in L$, we have

$$f(\mathbf{x} \vee c) = f(\mathbf{x}) \vee \delta_f(c) \qquad (\text{resp. } f(\mathbf{x} \wedge c) = f(\mathbf{x}) \wedge \delta_f(c)).$$

Observe that if f is $\overline{\mathcal{R}}_f$-idempotent (i.e., satisfying $f(c,\ldots,c) = c$ for every $c \in \overline{\mathcal{R}}_f$), then $\overline{\mathcal{R}}_f$-min homogeneity (resp. $\overline{\mathcal{R}}_f$-max homogeneity) is equivalent to quasi-min homogeneity (resp. quasi-max homogeneity).

Lemma 14. *Let $f\colon L^n \to L$ be nondecreasing and quasi-min homogeneous (resp. quasi-max homogeneous). Then f is quasi-max homogeneous (resp. quasi-min homogeneous) if and only if it is horizontally maxitive (resp. horizontally minitive).*

Combining Theorem 10 and Lemma 14, we obtain a characterization of quasi-polynomial functions in terms of quasi-min homogeneity and quasi-max homogeneity.

Theorem 15. *A function $f\colon L^n \to L$ is a quasi-polynomial function if and only if it is nondecreasing, quasi-max homogeneous, and quasi-min homogeneous.*

4.2.2 Quasi-median decomposability

A function $f\colon L^n \to L$ is said to be *median decomposable* [13] if, for every $\mathbf{x} \in L^n$ and every $k \in [n]$, we have

$$f(\mathbf{x}) = \operatorname{median}\big(f(\mathbf{x}_k^0), x_k, f(\mathbf{x}_k^1)\big),$$

where $\mathbf{x}_k^c = (x_1,\ldots,x_{k-1},c,x_{k+1},\ldots,x_n)$ for any $c \in L$. As Marichal [13] showed, the class of polynomial functions are exactly those functions which are median decomposable.

In complete analogy with the previous subsection we propose the following weaker variant of median decomposability. We say that a function $f\colon L^n \to L$ is *quasi-median decomposable* if, for every $\mathbf{x} \in L^n$ and every $k \in [n]$, we have

$$f(\mathbf{x}) = \operatorname{median}\big(f(\mathbf{x}_k^0), \delta_f(x_k), f(\mathbf{x}_k^1)\big).$$

Note that every nondecreasing unary function is quasi-median decomposable.

Observe that $\vee$ and $\wedge$, as well as any nondecreasing function $\varphi\colon L \to L$, are quasi-median decomposable. Also, it is easy to see that any combination of constants and a nondecreasing unary function φ using $\vee$ and $\wedge$ is quasi-median decomposable and hence, by Remark 4 (i), every quasi-polynomial function is quasi-median decomposable. Our following result asserts that quasi-median decomposable functions $f\colon L^n \to L$ with a nondecreasing diagonal section δ_f are exactly the quasi-polynomial functions.

Theorem 16. *A function $f\colon L^n \to L$ is a quasi-polynomial function if and only if δ_f is nondecreasing and f is quasi-median decomposable.*

5 Some special classes of quasi-polynomial functions

In this final section we consider few noteworthy subclasses of quasi-polynomial functions, namely those of quasi-term functions and quasi-weighted maximum and minimum functions, and provide characterizations accordingly.

5.1 Quasi-term functions

We say that a function $f\colon L^n \to L$ is

- *conservative* if, for every $\mathbf{x} \in L^n$, we have $f(\mathbf{x}) \in \{x_1,\ldots,x_n\}$.
- *quasi-conservative* if, for every $\mathbf{x} \in L^n$, we have $f(\mathbf{x}) \in \{\delta_f(x_1),\ldots,\delta_f(x_n)\}$.

Note that, if f is idempotent, then it is quasi-conservative if and only if it is conservative.

Theorem 17. *A quasi-polynomial function $f\colon L^n \to L$ is a quasi-term function if and only if it is quasi-conservative.*

5.2 Quasi-weighted maximum and minimum functions

A function $f\colon L^n \to L$ is said to be a *weighted maximum function* if there are $v_0, v_1, \ldots, v_n \in L$ such that

$$f(\mathbf{x}) = v_0 \vee \bigvee_{i\in[n]} (v_i \wedge x_i). \tag{5}$$

Similarly, $f\colon L^n \to L$ is said to be a *weighted minimum function* if there are $w_0, w_1, \ldots, w_n \in L$ such that

$$f(\mathbf{x}) = w_0 \wedge \bigwedge_{i\in[n]} (w_i \vee x_i). \tag{6}$$

ISBN: 978-989-95079-6-8

We say that a function $f\colon L^n \to L$ is a *quasi-weighted maximum function* (resp. a *quasi-weighted minimum function*) if there exist a weighted maximum function (resp. a weighted minimum function) $p\colon L^n \to L$ and a nondecreasing function $\varphi\colon L \to L$ such that $f = p \circ \varphi$.

To present a axiomatization of each of these classes, we need to recall some terminology. We say that a function $f\colon L^n \to L$ is

- *maxitive* if, for every $\mathbf{x}, \mathbf{x}' \in L^n$, we have $f(\mathbf{x} \vee \mathbf{x}') = f(\mathbf{x}) \vee f(\mathbf{x}')$.
- *minitive* if, for every $\mathbf{x}, \mathbf{x}' \in L^n$, we have $f(\mathbf{x} \wedge \mathbf{x}') = f(\mathbf{x}) \wedge f(\mathbf{x}')$.

We first recall the descriptions of maxitive and minitive functions; see Dubois and Prade [8] and Marichal [12].

Proposition 18. *A function $f\colon L^n \to L$ is maxitive (resp. minitive) if and only if there are nondecreasing unary functions $f_i\colon L \to L$ ($i \in [n]$) such that, for every $\mathbf{x} \in L^n$,*

$$f(\mathbf{x}) = \bigvee_{i\in[n]} f_i(x_i) \qquad (\textit{resp. } f(\mathbf{x}) = \bigwedge_{i\in[n]} f_i(x_i)).$$

Theorem 19. *Let $f\colon L^n \to L$ be a quasi-polynomial function. Then f is a quasi-weighted maximum function (resp. quasi-weighted minimum function) if and only if it is maxitive (resp. minitive).*

Remark 5. (i) Idempotent weighted maximum functions $f\colon L^n \to L$ are those functions (5) for which $v_0 = 0$ and $\vee_{i\in[n]} v_i = 1$. Dually, idempotent weighted minimum functions $f\colon L^n \to L$ are those functions (6) for which $w_0 = 1$ and $\wedge_{i\in[n]} w_i = 0$. These functions were introduced on real intervals by Dubois and Prade [7] in fuzzy set theory.

(ii) As observed in Proposition 12, the underlying weighted maximum function (resp. weighted minimum function) defining a given quasi-weighted maximum function (resp. quasi-weighted minimum function) can be chosen to be idempotent.

References

[1] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners*. Studies in Fuziness and Soft Computing. Springer, Berlin, 2007.

[2] P. Benvenuti, R. Mesiar, and D. Vivona. Monotone set functions-based integrals. In *Handbook of measure theory, Vol. I, II*, pages 1329–1379. North-Holland, Amsterdam, 2002.

[3] S. Burris and H. P. Sankappanavar. *A course in universal algebra*, volume 78 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1981.

[4] M. Couceiro and J.-L. Marichal. Representations and characterizations of polynomial functions on chains. `http://arxiv.org/abs/0811.0309`

[5] L. M. de Campos, M. T. Lamata, and S. Moral. A unified approach to define fuzzy integrals. *Fuzzy Sets and Systems*, 39(1):75–90, 1991.

[6] D. Dubois, J.-L. Marichal, H. Prade, M. Roubens, and R. Sabbadin. The use of the discrete Sugeno integral in decision-making: a survey. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, 9(5):539–561, 2001.

[7] D. Dubois and H. Prade. Weighted minimum and maximum operations in fuzzy set theory. *Inform. Sci.*, 39(2):205–210, 1986.

[8] D. Dubois and H. Prade. Aggregation of possibility measures. In *Multiperson decision making models using fuzzy sets and possibility theory*, volume 18 of *Theory Decis. Lib. Ser. B Math. Statist. Methods*, pages 55–63. Kluwer Acad. Publ., Dordrecht, 1990.

[9] R. L. Goodstein. The solution of equations in a lattice. *Proc. Roy. Soc. Edinburgh Sect. A*, 67:231–242, 1965/1967.

[10] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation functions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, UK, to appear in 2009.

[11] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy measures and integrals*, volume 40 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2000. Theory and applications.

[12] J.-L. Marichal. On Sugeno integral as an aggregation function. *Fuzzy Sets and Systems*, 114(3):347–365, 2000.

[13] J.-L. Marichal. Weighted lattice polynomials. *Discrete Mathematics*, 309(4):814–820, 2009.

[14] S. Rudeanu. *Lattice functions and equations*. Springer Series in Discrete Mathematics and Theoretical Computer Science. Springer-Verlag London Ltd., London, 2001.

[15] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, Tokyo, 1974.

[16] M. Sugeno. Fuzzy measures and fuzzy integrals—a survey. In M. M. Gupta, G. N. Saridis, and B. R. Gaines, editors, *Fuzzy automata and decision processes*, pages 89–102. North-Holland, New York, 1977.

ISBN: 978-989-95079-6-8

Novel Methods for the Design of General Type-2 fuzzy Sets based on Device Characteristics and Linguistic Labels Surveys

Christian Wagner Hani Hagras

The Computational Intelligence Centre, School of Computer Science and Electronic Engineering, University of Essex
Colchester, United Kingdom
Email: chwagn@essex.ac.uk, hani@essex.ac.uk

Abstract—*Fuzzy Logic Systems are widely recognized to be successful at modelling uncertainty in a large variety of applications. While recently interval type-2 fuzzy logic has been credited for the ability to better deal with large amounts of uncertainty, general type-2 fuzzy logic has been a steadily growing research area. All fuzzy logic systems require the accurate specification of the membership functions' (MFs) parameters. While some work for automatic or manual design of these parameters has been proposed for type-1 and interval type-2 fuzzy logic, the problem has not yet been widely addressed for general type-2 fuzzy logic. In this paper we propose two methods which allow the automatic design of general type-2 MFs using either data gathered through a survey on the linguistic variables required or, in the case of physical devices (e.g. sensors, actuators), using data directly gathered from the specific devices. As such, the proposed methods allow for the creation of general type-2 MFs which directly model the uncertainty incorporated in the respective applications. Additionally, we demonstrate how interval type-2, type-1 and the recently introduced zSlices based general type-2 MFs can be extracted from the automatically designed general type-2 MFs. We also present a recursive algorithm that computes the convex approximation of generated fuzzy sets.*

Keywords— Type-2 fuzzy systems, general type-2 fuzzy sets, specification of fuzzy sets.

1 Introduction

Fuzzy Logic is widely applied because of its ability to model the uncertainties which are inherent to most real world applications and environments. The high levels of uncertainty encountered by such fuzzy logic systems can generally be attributed to one or more of the categories below:

- Uncertainties in inputs to FLCs which translate to uncertainties in the antecedent Membership Functions (MFs) as for example sensor measurements are typically noisy and affected by the conditions of observation (i.e. their characteristics are changed by the environmental conditions such as wind, sunshine, humidity, rain, etc.).
- Uncertainties in control outputs which translate to uncertainties in the consequent MFs of FLCs. Such uncertainties can result from the change of the actuators characteristics which can be due to wear, tear, environmental changes, etc.
- Linguistic uncertainties associated with the meaning of words (that are used as antecedents or consequents in FLCs or used as part of Computing With Words (CWW) applications) which can be uncertain as "words mean different things to different people" [1]. For example the perception of the linguistic label "warm" varies for different people while also being heavily subject to context, e.g. "warm" in the Caribbean will usually be associated with a different temperature than "warm" at the North Pole etc.

While the capability of modelling uncertainty provided by type-1 fuzzy sets has been intensively researched and shown to provide good results, in recent years interval type-2 fuzzy sets have been found to provide an even higher potential at modelling large amounts of uncertainty [1],[2],[3],[4]. Additionally, a variety of significant advances has been made in the area of general type-2 fuzzy logic which is hoped to provide the framework for an accurate model of uncertainty.

One of the main challenges during the design of any fuzzy based system (general type-2, interval type-2 or type-1) is the specification of the fuzzy Membership Functions (MFs). The choice of type of membership function (such as gaussian, triangular, etc.) as well as the choice of their specific parameters strongly affects the performance of the fuzzy system. A variety of methods to alleviate this problem have been researched for mainly type-1 and interval type-2 FLCs. Such methods are generally based on the use of expert knowledge, evolutionary techniques (such as genetic algorithms) and neural networks [5], [6]. However, there is still much work needed to standardize and simplify the selection of appropriate MFs.

Recent developments in general type-2 fuzzy theory [7], [8], [9], [10], [11] have made general type-2 applications a realistic option for the future. Hence, we have developed novel techniques to determine the appropriate parameters for general type-2 MFs which due to their complex nature are highly difficult to specify. The proposed techniques are based on directly using existing knowledge about the uncertainty contained within the variables represented by the MFs. Additionally, the proposed techniques have the advantage that they can be used for the generation of all currently common types of fuzzy sets: type-1, interval type-2 and general type-2.

Section 2 describes the design of fuzzy MFs modelling linguistic concepts using surveys while Section 3 details the design of fuzzy MFs based on physical device or variable characteristics. Both sections present real world examples to demonstrate the techniques. Section 4 describes how to extract zSlice based general type-2, interval type-2 and type-1 MFs from the general type-2 MFs generated in Section 2 and 3. In Section 5 we present the recursive convexity algorithm employed throughout the paper and give a detailed example of its application. Finally, in Section 6 we present

ISBN: 978-989-95079-6-8

the conclusions and address the work planned for the near future.

2 Design of general type-2 fuzzy sets based on the perception of linguistic labels determined through surveys.

2.1 The Human Perception of Linguistic Labels.

Fuzzy sets are frequently used to specify linguistic variables such as *temperature* or *height* used by people on an everyday basis. As such, the variable *temperature* can for example be modelled using three linguistic labels, *cold*, *moderate* and *warm* which in turn are represented using fuzzy MFs. While type-1 sets can be used to model such labels, it is clear that as type-1 MFs associate each input with a crisp membership value, hence a given type-1 MF will need to be constantly re-tuned for different perceptions of the same linguistic label. For example, the perception of the linguistic label *warm* can easily be seen to depend on factors such as the country of origin of a person (someone from Alaska will consider warm a different temperature than someone from the Caribbean), the age, sex, current season, etc.

Interval type-2 fuzzy sets can alleviate the above need for the continuous re-tuning of individual MFs as the incorporated Footprint of Uncertainty (FOU) allows for the model to include the full range of values associated with a specific linguistic label by different people in different contexts. Interval type-2 fuzzy sets nevertheless have the significant shortcoming that the uncertainty is spread evenly across the FOU which does not allow for variations of the uncertainty within this FOU to be modelled. For example, considering the example of *warm* temperature, if two people associate warm with 25°C and one person associates it with 28°C, a FOU incorporating 25°C and 28°C will not reflect that the majority of the people (here two) associate 25°C with the linguistic label *warm* - we have effectively lost this information.

This additional information can be modelled using general type-2 fuzzy sets by employing the third dimension associated with the FOU to encode the amount of uncertainty associated with each point within the FOU. In terms of our example, the FOU would be associated in its third dimension with more uncertainty towards 28°C and less uncertainty towards 25°C.

In the next few subsections we will describe how a survey on linguistic labels can be used to gather information on a specific linguistic variable using the example of *Indoor Temperature*. While this example is chosen to illustrate and simplify the process, the mechanism can be applied for any linguistic variable which is modelled by a series of consecutive linguistic labels. We will show how the information can be used to generate fuzzy MFs which model the linguistic labels and their associated uncertainty accurately using general type-2 fuzzy sets. In addition, we will present the mechanisms used to extract appropriate interval type-2 and type-1 MFs as well as zSlices based general type-2 MFs which were introduced in [11].

2.2 The survey on the linguistic variable "Indoor Temperature".

As part of our work on the generation of MFs from surveys we have conducted a survey investigating the perception of the linguistic variable *Indoor Temperature*. We have presented the participants with a questionnaire asking them to identify the position of the three linguistic labels *Cold*, *Moderate* and *Warm* in the context of an indoor environment on a temperature scale ranging from 0°C to 40°C. Further details on the survey and its results are given in Table 1. It should be noted that participants could supply additional information during the survey, specifically their age, sex and country of origin. While maintaining anonymity, we consider this information highly useful as it can be expected that factors such as the country of origin significantly affect people's perception for example of temperature.

Table 1: Details on the linguistic survey on the perception of *Indoor Temperature*. (* based on information supplied)

Number of participants:	62
Age range of participants:	18-58 years*
Mean age of participants:	30.145 years*
Number of different countries of origin:	22 countries*
Location of survey:	Colchester, UK
Date of survey:	January 2009
Minimum values with *cold, moderate, warm*:	0°C, 10°C, 16°C
Maximum values with *cold, moderate, warm*:	24°C, 27°C, 35°C
Mean values for *cold, medium, warm*:	12.062°C, 19.664°C, 25.664°C
Standard deviations for *cold, medium, warm*:	4.637, 3.552, 3.926

2.3 From a Survey to a General Type-2 Fuzzy Set model of the Linguistic Variable

In order to utilise the data provided through the survey to generate a general type-2 fuzzy set model of the linguistic variable which includes all the uncertainty information gathered in the data, we follow the following steps:

(1) Map all the values associated with a specific linguistic label to the frequency with which the values occur in the data. This step can be likened to creating a continuous histogram for the data where the xAxis represents the values (in our case, the temperature) and the yAxis represents the number of occurrences. Fig. 1 shows this mapping for our data using interpolation between the data points.

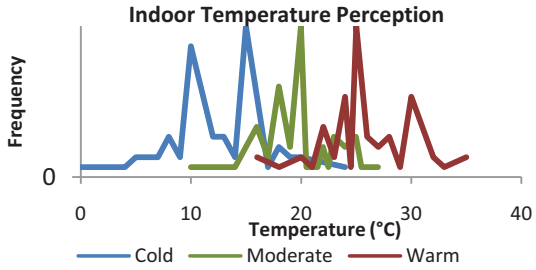


Figure 1: Mapping of survey data for the three linguistic labels *cold, moderate and warm*, showing the large amounts of uncertainty contained within the perceptions of the linguistic labels (e.g. large overlap).

(2) The data mapped in Step (1) and illustrated in Fig. 2 illustrates the significant variation in the perception of the individual linguistic labels as well as their strongly differing interrelation for different people. In order to

ISBN: 978-989-95079-6-8

utilize the information to define the third dimension of general type-2 fuzzy sets we normalise the frequency so it correlates with the uncertainty modelled in the third dimension of a fuzzy set, ie. the more frequent a specific sample was identified in the survey, the less uncertainty is associated with it.

(3) As convexity is a general requirement in several general type-2 fuzzy logic operations such as the join and meet computed on the membership grade presented in [12], we convert the normalised plots created in Step 2 to their convex approximation using a recursive algorithm which we describe in Section 5 in more detail. The original, normalised data as well as its convex approximations using our recursive algorithm and basic triangular approximation are shown for all three linguistic labels in Figs. 2a, 2b and 2c.

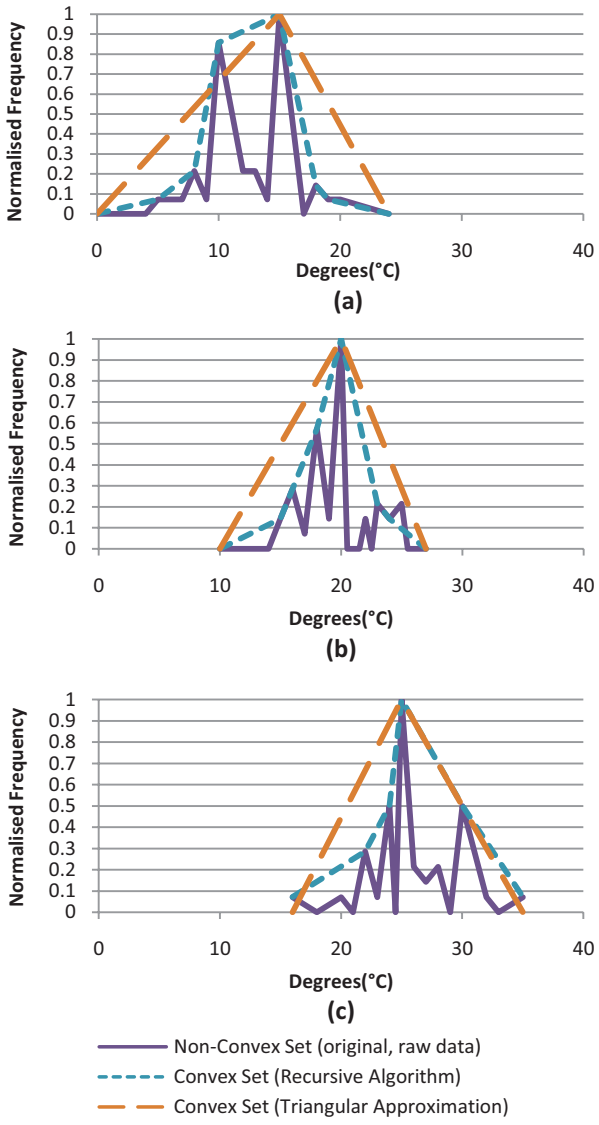


Figure 2: Representations of the linguistic labels, (a) H(cold), (b) H(moderate), (c) H(warm).

Fig. 2a, Fig. 2b and Fig. 2c clearly show that a more accurate convex approximation of the original sets is achieved when employing the proposed recursive algorithm (explained in Section 5) when compared to a simple triangular approximation. We refer to the "histogram based" representations of the linguistic labels cold, moderate and warm as H(cold), H(moderate) and H(warm).

(4) In order to create a fuzzy model of the linguistic variable *Indoor Temperature* we proceed one linguistic label at a time:

a. For *cold*, we associate H(cold) with the level 1 on the yAxis and H(moderate) with level 0 on the yAxis. We e*mploy* interpolation to recreate the general type-2 set for *cold* as shown in Fig. 3a, Fig. 3b and Fig 3c. It s*houl*d be noted that the positioning of H(cold) and H(modera*te)* at *thi*s step is not arbitrary but follows logically from the linguistic labels, i.e. H(cold) represe*nts* the temperature values associated with *cold*, hence they should be related to a yLevel of 1. H(moderate) on the other hand represents the *temp*erature values which are associated with *moderate* and as such not anymore related with *cold*, thus, in terms of the fuzzy set **cold**, it should be ass*ociated* with yLevel 0 at H(moderate).

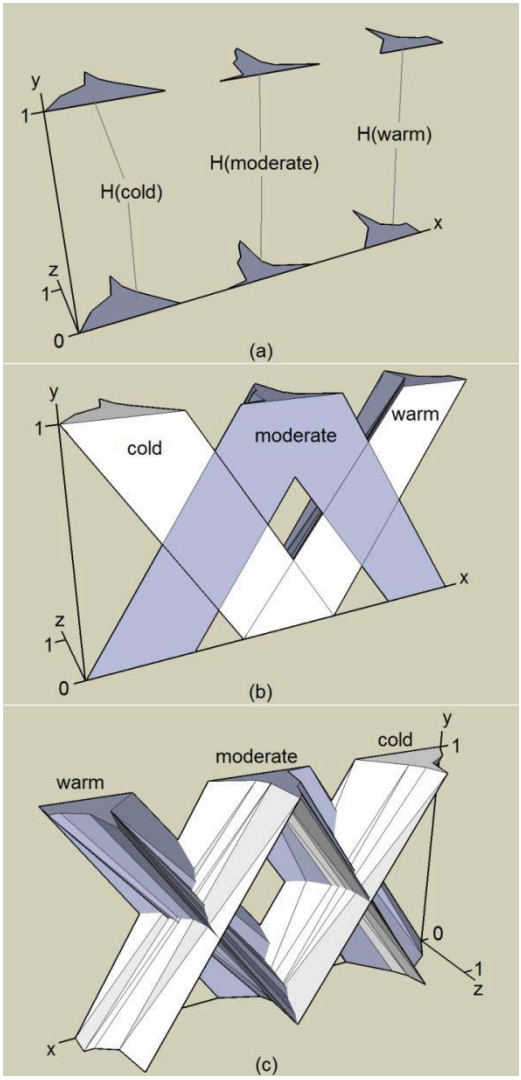


Figure 3: Schematic representation of the steps involved in creating the fuzzy set representation of linguistic labels. (a) "Histogram representations" of the linguistic labels associated with the respective levels on the yAxis before interpolation. (b) Fuzzy sets after interpolation from the front. (c) Fuzzy sets after interpolation from the rear.

ISBN: 978-989-95079-6-8

b. For *moderate*, we associate H(cold) with the level 0 on the yAxis, H(moderate) with level 1 on the yAxis and H(warm) with yLevel 0 on the yAxis. We employ interpolation to recreate the general type-2 set for *moderate* as shown in Fig. 3a, Fig. 3b and Fig. 3c.The reasoning of the positioning on the yAxis of H(cold), H(mod*erate) a*nd H(warm) is analogous to Step a.

c. Fo*r warm,* we associate H(moderate) with level 0 on the yAxis and H(warm) with yLevel 1 on the yAxis. We employ interpolation to recreate the general type-2 set for *warm* as shown in Fig 3(a)-(c).
The reasoning of the positioning on the yAxis of H(moderate) and H(war*m) i*s analogous to Step a.

We have shown in a series of steps how the data on a linguistic variable such as temperature can be used to generate a general type-2 representation of the given variable. It is straightforward to extract zSlices based general type-2 fuzzy sets as introduced in [11], interval type-2 and type-1 sets. The individual steps are shown for the general case in Section 4.

3 Design of fuzzy sets based on device or variable characteristics

In the context of Fuzzy Logic Controllers (FLCs), fuzzy logic sets are generally employed to model the uncertainty associated with the inputs and outputs of the FLC. In real world applications these inputs and outputs are commonly physical devices such as sensors and actuators which are susceptible to a wide range of sources for uncertainty, from wear and tear, to environmental conditions, signal noise etc. General type-2 fuzzy sets are well suited to model this uncertainty and in this section we are describing how information about the device or variable to be modelled can be directly incorporated into the fuzzy sets modelling. We will be using the illustrative example of generating the MFs for a sonar sensor from information which is directly retrieved from the device. While the example is referring to a sonar sensor, the overall method is applicable to every device or artefact that is to be modelled.

Generation of two MFs *Near* and *Far* which model the distance value measured using a Sonar sensor.

1. As a first step, we discretize the maximum sensor range into X steps. A finer level of discretization will result in a more accurate model of the uncertainty contained within the sensor outputs. The process of discretizing the sensor range of a Pioneer robot side sonar is shown in Fig. 4.
2. At every step, the device, in our case the sonar sensor is sampled (for example by presenting the sonar with an obstacle at the specified distance in our case) Y times. A larger number Y of samples will result in a more accurate model of the uncertainty contained within the sensor outputs. The series of Y readings are analysed to create a mapping detailing the normalized frequency of each specific sample. As such, in our case, the reading (or close group of readings) which was measured the largest number of times out of Y samples is mapped to 1, readings that have never been measured are mapped to 0 etc. This approach can be likened to a continuous (high discretization level) histogram. An example of this can be seen in Fig. 4. At each sampling step x, Y samples are taken which are translated to the respective histograms H(x).

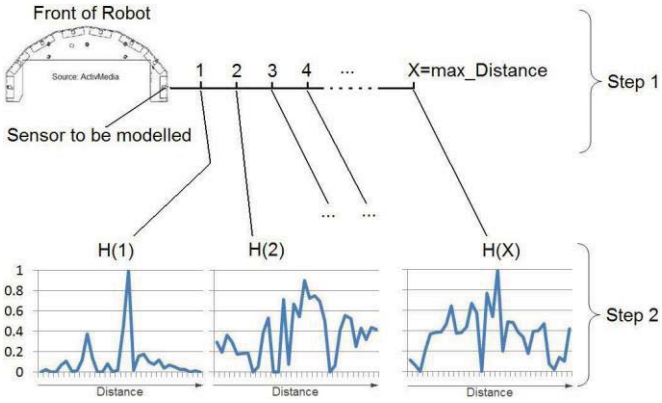


Fig. 4 Step 1 and 2 for the automatic generation of MFs for the sonar sensor mounted on a Pioneer robot.

3. The histograms generated in Step 2 are representative of the uncertainty encountered while sampling the specific distance x, i.e. while usually the majority of samples return the accurate reading, some will return a different reading because of sensor noise, environmental impact, etc. The information contained in the histograms can thus be employed to define the third dimension of general type-2 fuzzy sets which models the uncertainty within the sensor. In order to use the histograms to define the third dimension of general type-2 fuzzy sets, the histograms need to be strictly convex. The convexity of the third dimension of fuzzy sets is a general requirement in several general type-2 fuzzy logic operations such as the join and meet computed on the membership grade presented in [12]. While we are planning to address the direct handling of non-convex sets in the future, at this stage we are converting the existing histogram into a convex function. While this can be done through simple triangular approximation, this results in a very poor representation of the information contained in the initial histogram. We have devised a recursive algorithm which converts the histograms H(x) to their convex counterparts C(x) while preserving as much information as possible. An example comparing an initial histogram to its convex approximations using the triangular approach and our algorithm can be seen in Fig. 5 and an in depth

ISBN: 978-989-95079-6-8

description of the algorithm and further comparisons can be found in Section 5.

It should be noted that the algorithm can be applied to any type-1 fuzzy set and as such to the histograms as presented here as well as any vertical slices of the general type-2 fuzzy set if required.

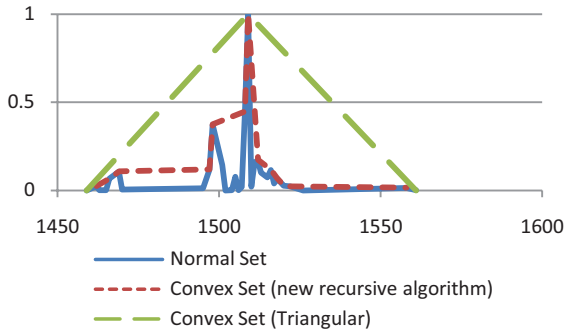


Figure 5: Comparison between an initial non-convex histogram (set) and its convex approximation using the triangular and our recursive algorithm approach.

4. For each linguistic label, we will need to determine the position of the linguistic labels within the device input/output range. In our example we are using triangular MFs and as such would define three points for every linguistic variable: its start, its maximum and its end. Fig.6 shows the Near MF for which in our example we defined the start as Ns, the maximum as Nm and the end as Ne. All previously generated convex H(x) where $x \in [start, end]$ are mapped to the triangle formed by {start, maximum, end}, "raising" them to the respective level on the yAxis. This process is illustrated in Fig. 6 for 5 histograms. In case there is uncertainty (for example due to linguistic uncertainties) about the positions of points Ns, Nm and Ne, we will have various triangles, each one is associated with the relevant histograms as shown in Fig.6. These triangles are then interpolated (as explained in the following step) in the x-y and the y-z domain to generate the FOU and third dimension of the general type-2 fuzzy set.

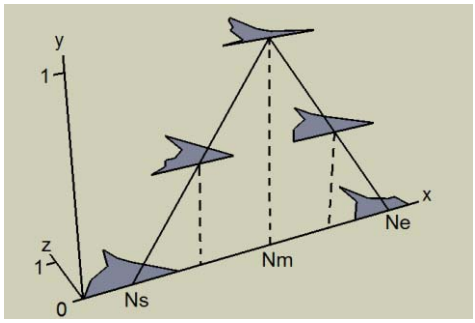


Fig. 6 The fuzzy MF *Near* is composed of the endpoints defined by the linguistic variable combined with the histograms of any sampling po*ints* which lie within the range of the linguistic variable. Here, five histograms are shown which have been "raised" to their respective level on the yAxis which is defined by the intersection of their sampling point x with the triangular {Ns,Nm,Ne}.

5. In the final step towards the generation of a general type-2 MF, the different histograms are linked through interpolation. It should be noted that the choice of interpolation method and the number of histograms directly affect the quality of the uncertainty model of the MF. Fig. 7 illustrates the front and side view of our exemplary MF *Near*. If it is requirement for the vertical slices of the resulting set to be strictly convex, the same convexity algorithm as described above can be applied to the vertical slice.

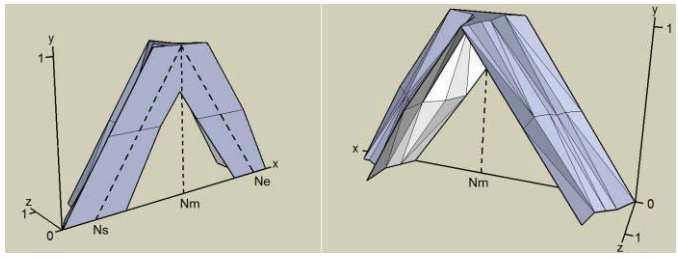


Figure 7: (a) Front view of the generated membership function for *Near*; (b) rear view of the same set.

The procedure *for* the automatic generation of general type-2 MFs presented in this section allows the generation of MFs which reflect and model the uncertainty encountered by or inherent to a specific device. The generated general type-2 MFs can easily be used to extract type-1, interval type-2 and zSlices based general type-2 fuzzy sets as we will be shown in Section 4.

4 Extraction of zSlice based general, interval type-2 and type-1 fuzzy sets.

It is straightforward to extract type-1 MFs and interval type-2 MFs from the general type-2 MFs generated in Sections 2 and 3 by restricting the zLevel (i.e. the value on the zAxis) to 1 or 0 respectively (in the case of interval type-2 MFs, they are subsequently "raised" to a zLevel of 1). This allows the process to generate type-1 and interval type-2 MFs from data without altering the procedure, which should facilitate the performance comparison between different levels of complexity in FLCs. An example of the MFs extracted from the general type-2 fuzzy sets shown in Fig. 7 is shown in Fig. 8.

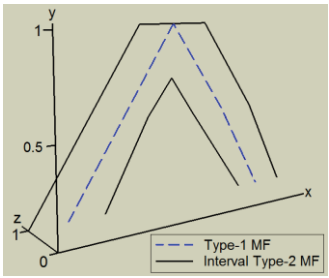


Figure 8: The type-1 and interval type-2 MFs extracted from the general type-2 MF shown in Fig. 7

ISBN: 978-989-95079-6-8

Finally, it is worth noting that the fuzzy sets generated in Sections 2 and 3 can easily be transformed to zSlices based MFs which were introduced in [11]. Fig. 9 shows an example of the generated zSlice based MF from the general type-2 MF shown in Fig. 7.

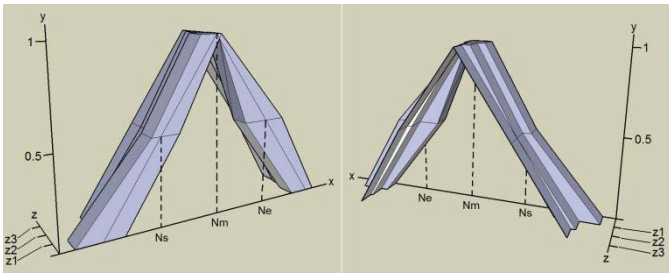


Figure 9: zSlice based general type-2 version (3 zSlices) of the general type-2 set shown in Fig. 15. It should be noted that only the zSlices 1-3 are shown, zSlice 0 has been omitted to improve visibility.

5 Approximation of a convex set from a non-convex set.

5.1 Description of the algorithm.

In order to generate a convex function from a non-convex function (more specifically, the discretized points from this function) while maintaining a high level of similarity to the original function, we have devised a recursive algorithm which is described in detail in this section.

The algorithm takes a discretized non-convex function (such as a type-1 fuzzy set) F as input. The discretization level along the xAxis is referred to as D. The function F matches each point on the xAxis x_d to a point y_d on the yAxis (where $d \in [1, D]$). As such, the function is represented as a series of points P_D in two-dimensional space where the x and y coordinates of a point P_d are referred to as P_{d_x} and P_{d_y}.

In our application of the convex approximation of "continuous histograms" (as discussed in Sections 2 & 3), the x-Axis represents the input axis, for example the distance, if we are considering a distance sensor, while the yAxis represents the number of samples. We denote the convex output set as O.

The recursive algorithm performs the following steps:

0.
   a. Traverse set from left to right (arbitrary- or from right to left) and find the coordinates of the discretized point P_{Max} (P_{Max} is defined in two-dimensional space by its coordinates P_{Max_x} and P_{Max_y}), where
      $P_{Max} \in P_D$ and
      $P_{Max_y} = \max\left(P_{d_y}\right), d \in [1, D]$.
      If more than one point with the same maximum is found, the middle point (according to its x coordinates P_{d_x}) is chosen. If the number of points with the same maximum is even, one of the middle points is chosen arbitrarily.
   b. Add point P_{Max} to the output set O.
   c. Split the series of points into two series: $\{P_1, \dots, P_{Max}\}$ and $\{P_{Max}, \dots, P_D\}$.
   d. Pass the series $\{P_1, \dots, P_{Max}\}$ to Step 1 and the series $P_{Max}, \dots, P_D\}$ to Step 2.

1.
   a. Refer to the given series of points as $[P_{start}, P_{end}]$.
   b. $IF\ (P_{start} == P_{end})$ THEN STOP.
   c. Find the new maximum point $P_{Max'}$ of the given series such that:
      $P_{Max'_y} > P_{start_y}\ AND\ P_{Max'_y} < P_{end_y}$
      $P_{Max'_y} = \max\left(P_{d_y}\right), d \in]start, end[$.
   d. Add point $P_{Max'}$ to the output set O.
   e. Split the given series into two series: $\{P_{start}, \dots, P_{Max'}\}$ and $\{P_{Max'}, \dots, P_{end}\}$ and pass both series separately to Step 1 (recursion).

2.
   a. Refer to the given series of points as $[P_{start}, P_{end}]$.
   b. $IF\ (P_{start} == P_{end})$ THEN STOP.
   c. Find the maximum point $P_{Max'}$ of the given series with endpoints such that
      $P_{Max'_y} < P_{start_y}\ AND\ P_{Max'_y} > P_{end_y}$
      $P_{Max'_y} = \max\left(P_{d_y}\right), d \in]start, end[$.
   d. Add point $P_{Max'}$ to the output set O.
   e. Split the given series into two series: $\{P_{start}, \dots, P_{Max'}\}$ and $\{P_{Max'}, \dots, P_{end}\}$ and pass both series separately to Step 2 (recursion).

Below we give an example of the application of the recursive algorithm.

5.2 Example application of the recursive algorithm.

In Table 2 we indicate the coordinates of a non convex set shown in Fig.10 taken from real world experiments and compare it to the coordinates computed by our algorithm and the triangular approximation.

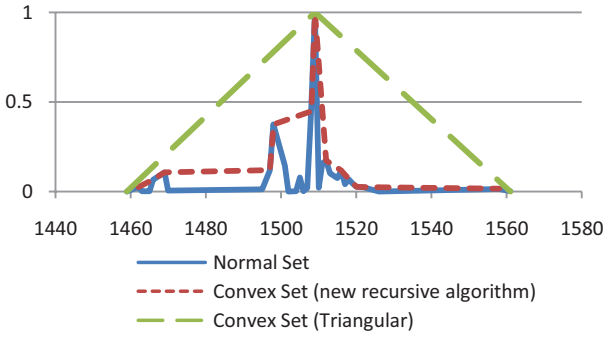


Fig. 10 A plot of the non convex input function as well as both the recursive and triangular approximated

From Table 2 it can be seen that both the recursive and the triangular algorithms omit a significant amount of detail in order to generate a convex approximation of the input set. Nevertheless the number of preserved samples and thus detail and information is much larger in the recursive approximation. This is also visible in Fig. 10 which shows that the recursive algorithm provides a vastly superior fit to

ISBN: 978-989-95079-6-8

the original function in comparison to the triangular approximation.

Table 2 Comparison of z and y tuples preserved by our recursive algorithm and a triangular algorithm while converting the non-convex set shown in Fig. 10.

Non-convex Set				Convex Set (Recursive Algorithm)		Convex Set (Triangular Algorithm)	
x	y	x	y	x	y	x	y
0.00	1459	0.44	1508	0.02	1462	0	1459
0.02	1462	1.00	1509	0.06	1466	1	1509
0.00	1463	0.02	1510	0.10	1469	0	1561
0.00	1465	0.15	1511	0.11	1497		
0.06	1466	0.17	1512	0.37	1498		
0.10	1469	0.10	1513	0.44	1508		
0.00	1470	0.07	1515	1.00	1509		
0.01	1495	0.11	1516	0.17	1512		
0.11	1497	0.04	1517	0.11	1516		
0.37	1498	0.06	1518	0.06	1518		
0.14	1501	0.04	1519	0.04	1519		
0.00	1502	0.02	1520	0.02	1520		
0.00	1504	0.02	1522	0.02	1522		
0.07	1505	0.00	1526	0.01	1558		
0.00	1506	0.01	1558				
0.02	1507	0.00	1561				

6 Conclusions

In this paper we have addressed the reoccurring problem in fuzzy systems of specifying the parameters for fuzzy membership functions (MFs). The correct selection of parameter of fuzzy MFs is crucial in order for the fuzzy systems to model the uncertainty contained within the system correctly and thus to allow the generated fuzzy system to provide good performance. While the problem has been previously addressed to some extent for interval type-2 and type-1 fuzzy systems, the use of general type-2 fuzzy sets requires the specification of an even larger number of parameters for each MF and as such is set to exasperate the problem of choosing those parameters.

We have presented two novel methods which allow the design of general type-2 fuzzy MFs directly from data and as such eliminate the need for a manual design of the MFs or the use of artificial intelligence techniques in order to tune the MFs.

The first method allows the design of general type-2 fuzzy MFs using data on linguistic labels which is collected through a survey. The technique allows for an accurate representation of the perception of linguistic variables gathered through the survey without any loss of information.

The second method specifies the general type-2 MFs based on data which is directly sampled from the device (such as a sensor) which is to be modelled and as such creates an accurate representation of the uncertainty associated with the specific device.

Both methods further allow the extraction of type-1, interval type-2 and the recently introduced zSlices based general type-2 fuzzy sets.

Additionally, we have presented a recursive algorithm that computes a convex approximation of fuzzy sets while preserving significantly more detail than simple triangular approximation which in turn enables a more accurate modelling of the uncertainty within the fuzzy MFs.

In the future, we are planning to further refine the algorithms for the automatic definition of fuzzy sets while investigating their performance in real world applications. Particularly, we are looking to extend the theoretical frameworks to non-convex fuzzy sets and to investigate non-singleton fuzzification in conjunction with the automatically created sets.

Acknowledgment

This work was supported in part by the European Commission under Grant FP7-ICT-1-8.2/216837 entitled “Adaptive and Trusted Ambient Ecologies” (ATRACO).

References

[1] Mendel and R. John; “Type-2 Fuzzy Sets Made Simple” ; IEEE Transactions on Fuzzy Systems, Vol. 10, pp.117-127, April 2002.

[2] H. Hagras, "A Hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots,” *IEEE Transactions On Fuzzy Systems*, vol. 12 No. 4, pp. 524-539, August 2004.

[3] J. Figueroa, J. Posada, J. Soriano, S. Rojas and M. Melgarejo, “A Type-2 Fuzzy Controller for Tracking Mobile Objects in the Context of Robotic Soccer Games”, Proceedings of the International Conference of Fuzzy System, May 2005, pp.359 – 364.

[4] J. Mendel, Uncertain Rule-Based Fuzzy Logic Systems, Prentice Hall 2001.

[5] A. Homaifar and E. McCormick, “Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms,” IEEE Transactions Fuzzy Systems, vol. 3, No. 2, pp. 129-139 May 1995.

[6] C. Lynch. H. Hagras and V. Callaghan, “Embedded Interval type-2 Neuro-Fuzzy Speed Controller for Marine Diesel Engines,” Proceedings of the 2006 Information Processing and Management of Uncertainty in Knowledge-based Systems conference, Paris, France, July 2006, pp. 1340-1347.

[7] S. Coupland and R. John, “Geometric type-1 and type-2 fuzzy logic systems,” IEEE Transactions on Fuzzy Systems, vol. 15, pp. 3-15, Feb. 2007.

[8] S. Greenfield and R. John, “Optimised Generalised Type-2 Join and Meet Operations,” Proceedings of the IEEE International Conference of Fuzzy Systems, London, UK, July 2007, pp. 141-146.

[9] L. Lucas, T. Centeno and M. Delgado, “General type-2 fuzzy inference systems: analysis, design and computational aspects,” Proceedings of the IEEE International Conference of Fuzzy Systems, London, UK, July 2007, pp. 1107-1112.

[10] J. Mendel, Feilong Liu, “On new quasi-type-2 fuzzy logic systems,” Proceedings of the IEEE International Conference of Fuzzy Systems, Hong Kong, China, June 2008, pp. 354-360.

[11] C. Wagner and H. Hagras, “zSlices – Towards bridging the gap between Interval and General Type-2 Fuzzy Logic,“ Proceedings of the IEEE International Conference of Fuzzy Systems, Hong Kong, China, June 2008, pp. 489-497.

[12] N. Karnik and J. Mendel, "Operations on type-2 fuzzy sets," Fuzzy Sets and Systems, vol. 122, pp. 327-348, 2001.

ISBN: 978-989-95079-6-8

Flow Line Systems with Possibilistic Data: a System with Waiting Time in Line Uncertain

David de la Fuente [1] María José Pardo [2]

1 Polytechnic School of Engineering-Gijon, University of Oviedo
Gijón, Spain
2 Faculty of Economics, University of Basque Country
Bilbao, Spain
Email: david@uniovi.es, mjose.pardo@ehu.es

***Abstract*—** *This paper proposes to analyze two flow line systems in which we include possibilistic data -the priority-discipline is possibilistic instead of probabilistic- and measure the performances of the systems with the effectiveness measure "waiting time in queue". In a previous work we have analyzed and developed a queuing model with uncertain priority-discipline, using Zadeh's extension principle. Because of it in this paper the analysis of the model takes the work realized by Prade as a starting point to incorporate an uncertain priority-discipline in a flow line system by means of a possibilistic distribution. To verify the validity of the proposed method, we calculate the possibility distribution of the priority-discipline and the possibility distribution of the performance measure "waiting time in queue", in two flow line systems with different waiting lines: the determinist flow line system $D/D/1$ and the classic flow line system $M/M/1/N$ with finite capacity. Since the performance measure is expressed by membership function rather than by crisp value, the fuzziness of input information is conserved completely, and more information is provided for capacity planning in flow line systems.*

Keywords— Flow Line Systems, Fuzzy Numbers, Possibility Theory, Priority-Discipline

1 Introduction

In this work we present two flow line systems to which we add a possibilistic priority-discipline instead of a probabilistic priority-discipline. For the development of the models we will follow the methodology exposed by *Prade* [1] about possibilistic models of queuing systems. In the literature regarding the queuing models with uncertain data we have not found any article that uses and develops his method in order to incorporate uncertainty in the queuing models with priority-discipline (except for *Prade*'s work). For example, the works of Li and Lee [2], Negi and Lee [3], or more recently Chen [4] or Pardo and de la Fuente [5,6] among others, study the flow line systems with uncertain data according to the Zadeh's extension principle [7], and in the work of Pardo and de la Fuente [8] the optimization of the fuzzy queuing model with priority-discipline is also studied according to the Zadeh's extension principle.

2 Calculation of the membership function of possibilistic priority-discipline

The possibilistic priority-discipline indicates the possibility that has the unit that arrives to the system in the position *n* to be overtaken, displaced or served in a different position that *n* in the waiting line. To develop the proposed models in this work, we need to define a possibility distribution of the priority-discipline but the possibility distribution can be calculated from the different existing methods to construct possibility distributions from real data or from the opinion of the experts (for example in Dubois et al. [9]).

In this way, we define the following possibility distribution (we denote by μ_i^n) and it provides us with the possibility that has the unit that has arrived in the position *n* to be served in the position $n+i$. Logically, this is valid when there is a queue length of one or more units at the arrival of one unit to the system, depending on the case. Besides and because of simplicity, the possibility distribution of the priority-discipline will be considered independent of the unit *n*. For this reason, the possibility of the unit *n* to advance one or more positions in the queue is:

1. The possibility of being served in the position *n* is: $\mu_0^n = 1$.
2. The possibility of being served in the position $n-1$ is: $\mu_{-1}^n = \sigma_1 \left(\sigma_1 \le 1\right)$.
3. The possibility of being served in the position $n-2$ is: $\mu_{-2}^n = \sigma_2 \left(\sigma_2 \le 1\right)$. In this case we consider that the unit $n-1$ and the unit $n-2$ move back one position in the waiting line.
4. The possibility of being served in the position $n+i$ with $i = -3,-4,...$ is: $\mu_i^n = 0$. In this way one unit can not advance more than two positions in the waiting line.

From the following hypotheses, we calculate the possibility that has the waiting unit to be moved back one position in the waiting line because of the arrival of one unit with higher priority. Then:

1. There are two reasons for the unit *n* to move back one position in the waiting line:
   i) Because the unit $n+1$ overtakes the unit *n*. That happens because it has priority to advance one position in the queue or because it has priority to advance two positions in the queue. The

ISBN: 978-989-95079-6-8

possibility of the event "the unit $n+1$ overtakes the unit n" is the association of the two previous events and it is know that it has a certain rate according to the fuzzy logic (Zimmermann [10]):

$$\max\{\sigma_1, \sigma_2\} \tag{1}$$

ii) If the unit $n+1$ does not have higher priority to the unit n, but the unit $n+2$ has higher priority to the units n and $n+1$ and advances two positions, the units n and $n+1$ move back one position and that happens with a possibility grade of σ_2.

Then, the possibility grade of the unit n to move one position back once it has arrived to the system is the association of the previous events, and according to the fuzzy logic (Zimmermann [10]):

$$\mu_1^n = \max\{\max\{\sigma_1, \sigma_2\}, \sigma_2\} = \max\{\sigma_1, \sigma_2\} \tag{2}$$

2. The unit n can move back two positions in the queue if it moves back twice consecutively, that means if the event "the unit n moves back one position" happens twice, what in fuzzy logic has a certain rate (Zimmermann [10]):

$$\mu_2^n = \min\{\mu_1^n, \mu_1^n\} = \mu_1^n = \max\{\sigma_1, \sigma_2\} \tag{3}$$

3. Likewise, the possibility grade of the unit to be moved back three or more positions in the waiting line can be calculated and in all of them we have:

$$\mu_i^n = \max\{\sigma_1, \sigma_2\} \quad \text{with} \quad i = 3, 4, \ldots \tag{4}$$

Then, the possibility grade of the unit n to move back in the waiting line, regardless of the number of positions is:

$$\mu_i^n = \max\{\sigma_1, \sigma_2\} \quad \text{with} \quad i = 1, 2, \ldots \tag{5}$$

The possibility distribution of the priority-discipline in the waiting line is:

$$\mu_i^n = \begin{cases} 0 & \text{if} \quad i = -3, -4, \ldots \\ \sigma_2 & \text{if} \quad i = -2 \\ \sigma_1 & \text{if} \quad i = -1 \\ 1 & \text{if} \quad i = 0 \\ \max\{\sigma_1, \sigma_2\} & \text{if} \quad i = 1, 2, \ldots \end{cases} \tag{6}$$

Once we know μ_i^n we obtain the "possibility distribution of the permanence time in line of the unit n" (denoted $\tilde{W}_q^n$), taking into account that if the unit n advances two positions, it does not have to wait in the waiting line to the service of the units $n-2$ and $n-1$. I it advances one position, we have to take away the service of the unit $n-1$. If it maintains its arrival position, its permanence time in queue does not change. And if it moves back in the queue it has to wait its permanence time and also so many service times as units that had overtaken it. Using Zadeh's notation to denote a fuzzy number (Zadeh [7]) (the possibility grade is previous to the symbol "/", the possible value is posterior, and the symbol "+" represents the association), if it is denoted by b the service time of one unit, the possibility distribution of the permanence time in queue of the unit n is:

$$\begin{aligned} \tilde{W}_q^n = {} & \mu_{-2}^n / \left(W_q^n - 2b\right) + \mu_{-1}^n / \left(W_q^n - b\right) + \\ & + \mu_0^n / W_q^n + \mu_1^n / \left(W_q^n + b\right) + \mu_2^n / \left(W_q^n + 2b\right) + \ldots \end{aligned} \tag{7}$$

Depending on the particular features of the flow line system to which is added this possibilistic priority-discipline, we obtain some results that differ among the different models. In the following sections we show the obtained theoretical results with two flow line systems with possibilistic priority-discipline.

3 Determinist flow line system $D/D/1$ with times among arrivals longer than the service time and possibilistic priority-discipline

In the determinist flow line system, the units arrive to the system with time among arrivals constant and are also served in constant times. It is supposed that the system is organised in a way that the arrivals happen in regular intervals and they are also regularly served in an only channel. The arrival source is considered unlimited and there is no limitation in the capacity of the system.

Saaty [11] analyses the determinist queuing system D/D/1 in transitory state. That means that the initial conditions (the number of units that there are in the system at the initial moment) affect the obtained results in the model. In addition to this, Saaty studies the system with times among arrivals longer than the service times. This condition supposes that the units take in arriving to the system more time that the system needs to serve them, that is why there will never be a waiting line when starting the system there are no units in it (initial condition $I = 0$ when $t = 0$, with I the number of initial units in the system). Each unit that arrives to the system is served before the arrival of the next unit. Due to this, Saaty studies the determinist model supposing that it is able to empty (the units in the system are not accumulated) under the initial existence conditions of I units in the system at the initial moment.

To know the evolution of this system and to calculate its performance measure, we have to analyse the behaviour of this flow line. The I units are found in the facilities when the system start to work and since the system is able to absorb the next arrivals (they are served very fast) then there will be an moment (denoted T) in which there will be no units in the system and each one arriving to the system will be served before the arrival of the next one. In T, the system reaches the steady state: the server stays idle until a new arrival, there are no units in the waiting line and the permanence time in queue is zero. In order to define the system completely, we need to calculate the number of units that arrive to the system before T (denoted A). They are the units that have to stay in the waiting line together with the initials.

ISBN: 978-989-95079-6-8

Regarding the performance measure "waiting time in queue of the unit *n*", W_q^n Saaty does not calculate it for the different values of *n*. That is why before incorporating a possibilistic priority-discipline, we will first calculate the variable *A* and the performance measure W_q^n. It will be so denoted:

a: Time among consecutive arrivals.

b: Service time.

I: Number of units in the system at the initial moment (including the unit that arrives to the system at the initial moment and the units that were already in it). The case $I=1$ means that at the arrival of the first unit, the system was empty.

The system reaches the steady state in *T* once the server has attended the initial units *I* and the new arrivals *A*. The arrival $A+1$ is the first one that finds the server free and goes directly to be attended. This arrival happens in $(A+1)a$. From here, the moment in that the first unit $A+1$ arrives to the system has to be longer than the time of service of the $I+A$ units, otherwise it is no possible that such arrival finds the system free. Then:

$$(A+1)a > (I+A)b \Leftrightarrow A > \frac{Ib-a}{a-b} \tag{8}$$

The variable *A* has to be the first entire number that fulfils the previous expression, then:

$$A = \left[\frac{Ib-a}{a-b}\right] + 1 = \left[\frac{b(I-1)}{a-b}\right] \tag{9}$$

where $[x]$ represents the whole part of the number *x*, that is to say, the largest integer less than or equal to a number *x* (floor function).

The performance measure of the permanence time in queue of the unit *n*, W_q^n is:

$$W_q^n = \begin{cases} (n-1)b & \text{if} \quad n \le I \\ (n-1)b-(n-I)a & \text{if} \quad I < n \le I+A \\ 0 & \text{if} \quad n > I+A \end{cases} \tag{10}$$

We obtain the result (10) from the following reasoning: If the arrival time to the system of the unit *n* is:

$$T_{ll}^n = \begin{cases} 0 & \text{if} \quad n \le I \\ (n-I)a & \text{if} \quad n > I \end{cases} \tag{11}$$

and the time when the unit *n* is served is:

$$T_s^n = \begin{cases} (n-1)b & \text{if} \quad n \le I+A \\ (n-I)a & \text{if} \quad n > I+A \end{cases} \tag{12}$$

then $W_q^n = T_s^n - T_{ll}^n$.

3.1 Example

Once we know the behaviour of the determinist flow line systems, following we show the incorporation of a possibilistic priority-discipline in the system $D/D/1$ with practice application, calculating the possibility distribution of the permanence time in queue for each unit that when arrive to the system have to stay in the waiting line for a flow line system with time among arrivals of 30 min., service time 20 min., 8 initial units and with the possibility distribution of the priority-discipline: $\mu_{-2}^n = 0.2$; $\mu_{-1}^n = 0.4$ and $\mu_0^n = 1$, from where we have that:

$$\mu_i^n = \max\{\sigma_1, \sigma_2\} = \max\{0.4, 0.2\} = 0.4\,,\ i = 1,2,...$$

First of all with the data of the example it will be $I=8$, $A=14$, and:

$$T_{ll}^n = \begin{cases} 0 & \text{if} \quad n \le 8 \\ (n-8)30 & \text{if} \quad n > 8 \end{cases} \qquad T_s^n = \begin{cases} (n-1)20 & \text{if} \quad n \le 22 \\ (n-8)30 & \text{if} \quad n > 22 \end{cases}$$

$$W_q^n = \begin{cases} (n-1)20 & \text{if} \quad n \le 8 \\ (n-1)20-(n-8)30 & \text{if} \quad 8 < n \le 22 \\ 0 & \text{if} \quad n > 22 \end{cases}$$

Before calculating the possibility distribution of the permanence time in queue of the client *n* and given the features of the system, we have to make the following remarks:

- Because of the evolution of this system there is a unit *N*, such that $I < N \le I+A$ $(8 < N \le 22)$, which fulfils that every unit $n > N$ is served before the arrival of the next unit and it is not possible for the unit $n+1$ to overtake the unit *n*. That means, if it is denoted by T_s^N at the same moment that the unit *N* is served and by T_{ll}^{N+1} the arrival time of the unit $N+1$, then *N* is the first unit that fulfils: $T_s^N < T_{ll}^{N+1}$, so there is no possibility for the position *n* to be moved back in the waiting line, then $\mu_j^n = 0$, $j = 1,2,...$ and $n > N$, with:

  $$T_s^N = (N-1)20 < (N+1-8)30 = T_{ll}^{N+1}$$

  From here, we obtain $N > 19$, then in the example it is $N = 20$. In this way we obtain:

  1. If $n < 20$ is $\mu_j^n = 0.4$ with $j = 1,2,...,20-n$ and $\mu_j^n = 0$ with $j = 20-n+1,....$
  2. If $n \ge 20$ is $\mu_j^n = 0$ with $j = 1,2,...$
  3. And if from the unit 20 it is not possible to move back in the waiting line, then it is not possible to overtake in the line from the unit 21 and the following ones. Then, if $n \ge 21$ is $\mu_{-2}^n = \mu_{-1}^n = 0$.

- The possibilities μ_{-2}^n and μ_{-1}^n when $n \le 20$ depend on the unit *n*:

ISBN: 978-989-95079-6-8

1. If $n=1$, that is, the first unit in the system, then: $\mu_{-2}^{1}=\mu_{-1}^{1}=0$.
2. If $n=2$, then: $\mu_{-2}^{2}=0$ and $\mu_{-1}^{2}=0.4$, given that the second unit arrives at the same time that the first one and there is a possibility that it advances in the waiting line.
3. If $3\le n\le I=8$, they are units that arrive to the system in the moment zero and there is a possibility that all of them change their position advancing one or two positions in the waiting line, then: $\mu_{-2}^{n}=0.2$ and $\mu_{-1}^{n}=0.4$
4. If $9\le n\le 20$, because of the features of that system we know that the waiting line diminishes for any unit that enters, that means that it is possible for the last arrivals that the level of the queue is lower than 2 or 1 unit and they can not advance one or two positions respectively. To determine what units can make its priority effective, we follow the next reasoning: We denote N_1 the unit with priority to advance two positions in the waiting line, in order to make this priority effective, the arrival time of the unit N_1 must be lower or the same that the service time of the unit N_1-2 (the unit previous to N_1 in two positions) so it can occupy its position, then $T_{ll}^{N_1}\le T_s^{N_1-2}$, and from here we get $N_1\le 18$. And we denote N_2 the unit with priority to advance one position in the waiting line, this priority can be likewise effective if the arrival time of the unit N_2 is lower or the same that the service time of the unit N_2-1 (the unit previous to N_2 one position), then $T_{ll}^{N_2}\le T_s^{N_2-1}$, from where it is $N_2\le 20$. Then:

   i) If $9\le n\le 18$, then $\mu_{-2}^{n}=0.2$ and $\mu_{-1}^{n}=0.4$

   ii) If $n=19$ or $n=20$, then $\mu_{-2}^{n}=0$ and $\mu_{-1}^{n}=0.4$

With all this we calculate the possibility distribution of the permanence time in queue of the unit *n*:

$$\begin{aligned}\tilde{W}_q^n=\ &\mu_{-2}^{n}/\left(W_q^n-40\right)+\mu_{-1}^{n}/\left(W_q^n-20\right)+\\ &+1/W_q^n+\mu_1^n/\left(W_q^n+20\right)+\mu_2^n/\left(W_q^n+40\right)+\\ &+\ldots+\mu_{20-n}^{n}/\left(W_q^n+(20-n)20\right)\end{aligned}$$

Following is detailed this possibility distribution for some values of *n*:

$$\begin{aligned}\tilde{W}_q^1&=1/0+0.4/20+\ldots+0.4/380\\ \tilde{W}_q^2&=0.4/0+1/20+0.4/40+\ldots+0.4/380\\ \tilde{W}_q^8&=0.2/100+0.4/120+1/140+0.4/160+\ldots+0.4/380\\ \tilde{W}_q^{15}&=0.2/30+0.4/50+1/70+0.4/90+\ldots+0.4/170\\ \tilde{W}_q^{20}&=0.4/0+1/20\\ \tilde{W}_q^{21}&=1/10\\ \tilde{W}_q^n&=1/0\quad\text{if}\quad n\ge 22\end{aligned}$$

Just as the possibilistic priority-discipline has been adapted to the determinist flow line system $D/D/1$, with time among arrivals longer than the service time, it can be incorporated to other determinist queuing systems only taking into account the features of the system and the evolution of the waiting line.

4 Flow line systems $M/M/1/N$ with possibilistic priority-discipline

The classic flow line system $M/M/1/N$ with finite capacity is considered, with times among arrivals and service time distributed according to an exponential of parameters λ and μ respectively, equals to all the units and regardless the kind of priority. The system has an unlimited enter source and the capacity is *N* units in the system.

The system is analysed in steady state so we know the probabilities of the system to be empty [12]:

$$P_0=\left(\sum_{n=0}^{N}\left(\frac{\lambda}{\mu}\right)^n\right)^{-1} \tag{13}$$

and that the units *n* are found in the system:

$$P_n=P_0\left(\frac{\lambda}{\mu}\right)^n \text{ with } n=1,2,\ldots,N \tag{14}$$

and we also know that the average permanence time in queue of each unit is:

$$W_q=\frac{L_q}{\lambda\left(1-P_N\right)} \tag{15}$$

where:

$$L_q=P_0\frac{\lambda}{\mu}\left(\sum_{n=2}^{N}(n-1)\left(\frac{\lambda}{\mu}\right)^{n-1}\right) \tag{16}$$

is average length of the units in queue.

To incorporate the possibilistic priority-discipline before indicated we have to take into account that in the stochastic flow line systems we do not know for sure the moment when the unit arrives to the system, which the length of the queue is when the unit arrives and if, after that, one ore more units can arrive and make that the previous unit move back in the queue from its arrival position. Then, we should calculate the average possibility to advance one or two

ISBN: 978-989-95079-6-8

positions in line, the average possibility to remain in the arrival position or the average possibility of moving back in the queue:

1. When the unit *n* arrives to the system, if it has priority to overtake 2 units and besides there is a waiting line with 2 or more units in the queue (3 units or more in the system), the priority-discipline will be made effective. The possibility of overtaken two units is μ_{-2}^{n} and the probability that 3 or more units are found in the system is $P_{n\geq 3} = 1 - P_0 - P_1 - P_2$. For this reason the average possibility of one unit advancing two positions in the queue (denoted p_{-2}) happens when both happen at the same time, and using the definition of Zadeh [13] of the probability of a fuzzy event is:

$$p_{-2} = \mu_{-2}^{n} \cdot P_{n\geq 3} \tag{17}$$

2. With a similar reasoning to the previous one, we conclude that the average possibility of one unit advancing one position is:

$$p_{-1} = \mu_{-1}^{n} \cdot P_{n\geq 2} \tag{18}$$

with $P_{n\geq 2} = 1 - P_0 - P_1$

3. The possibility of remaining in its position in the waiting line always happens, regardless the number of units that there are in the system at the moment that the unit *n* arrives, then:

$$p_0 = \mu_0^{n} = 1 \tag{19}$$

4. The average possibility of moving back in the queue, p_i with $i = 1, 2, ..., N-1$ happens when the unit has to wait to be attended since another unit can arrive meanwhile with higher priority. One unit has to wait if the system is busy when it arrives to it and this happens with probability $P_{n>0} = 1 - P_0$, then:

$$p_i = \mu_i^{n} \cdot P_{n>0} \tag{20}$$

Once we know the average possibility of advancing or moving back in the waiting line, the possibility distribution of the average permanence time in queue is:

$$\begin{aligned} \tilde{W}_q = {} & p_{-2} / \left(W_q - 2/\mu\right) + p_{-1} / \left(W_q - 1/\mu\right) + \\ & + p_0 / W_q + p_1 / \left(W_q + 1/\mu\right) + p_2 / \left(W_q + 2/\mu\right) + ... \end{aligned} \tag{21}$$

4.1 Example

To conclude, we illustrate with an example the incorporation of a possibilistic priority-discipline calculating the possibility distribution of the average permanence time in queue for a flow line system $M/M/1/10$ of parameters $\lambda = 6$ and $\mu = 8$ units per hour, and $N = 10$ units. The possibility distribution of the waiting discipline is: $\mu_{-2}^{n} = 0.4$, $\mu_{-1}^{n} = 0.6$ and $\mu_0^{n} = 1$. From here: $\mu_i^{n} = \max\{0.4, 0.6\} = 0.6$, with $i = 1, 2, ..., 9$.

With this data we have: $P_0 = 0.261$, $P_1 = 0.1958$ and $P_2 = 0.1468$, then $P_{n\geq 3} = 0.3964$, $P_{n\geq 2} = 0.5432$ and $P_{n>0} = 0.739$. The average possibility of advancing or moving back in the waiting line is: $p_{-2} = 0.1586$, $p_{-1} = 0.3259$, $p_0 = 1$ and $p_i = 0.4434$ with $i = 1, 2, ..., 9$ The average permanence time in the waiting line of one unit is: $W_q = 0.3004$. And the possibility distribution of the average permanence time in the queue is:

$$\begin{aligned} \tilde{W}_q = {} & 0.1586/0.0504 + 0.3259/0.1754 + \\ & + 1/0.3004 + 0.4434/0.4254 + \\ & + 0.4434/0.5504 + ... \end{aligned}$$

The possibility distribution $\tilde{W}_q$ is represented in the Fig. 1.

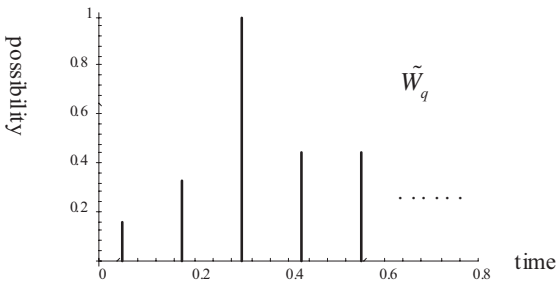


Figure 1: Possibility distribution $\tilde{W}_q$

The possibilistic priority-discipline can be adapted to other stochastic flow line system only by following a similar procedure to the one carried out in the last example.

5 Conclusions

In this work, taking the work started by *Prade* [1] as a starting point and incorporating it a possibilistic priority-discipline, we have given a more realistic representation of the flow line systems. For this we have defined a possibility distribution for the waiting-discipline but it can be calculated from real data or from the opinion of the experts. To illustrate the proposed method and the obtained theoretical results we have carried out two applications: one, to incorporate the possibilistic priority-discipline in a pure determinist flow line system and the other one to incorporate the uncertain priority-discipline in a classic flow line model. In both applications we have developed the technique to incorporate the priority-discipline given the particular features of the system and on the other hand, we have exposed an example that explains the proposed methodology.

Even so, the flow line systems with uncertain waiting line presented and developed in this work provide the decision maker with more complete and informative results and also with a better knowledge about the behaviour of the system due to the results that are possibility distributions that

ISBN: 978-989-95079-6-8

include all the uncertainty contained in the system. It is also a simple methodology that can be adapted to the different flow line systems with uncertain priority-discipline. For this reason we think that the proposed flow line system with uncertain data can have more applications than the classic one.

References

[1] H.M. Prade, *An outline of fuzzy or possibilistic models for queuing systems*, in: Fuzzy Sets: Theory and Applications to Policy Analysis and Information Systems. Wang, P.P. and Chang, S.K. (eds.) New York: Plenum Press, 1980.

[2] R.J. Li and E.S. Lee, Analysis of fuzzy queues. *Computers and Mathematics with Applications*, 17 (7):1143-1147, 1989.

[3] D.S. Negi and E.S. Lee, Analysis and simulation of fuzzy queues. *Fuzzy Sets and Systems*, 46:321-330, 1992.

[4] S.P. Chen, Solving fuzzy queuing decision problems via a parametric mixed integer nonlinear programming method. *European Journal of Operational Research*, 177 (1):445-457, 2007.

[5] M.J. Pardo and D. de la Fuente, Optimizing a priority-discipline queuing model using fuzzy set theory. *Computers and Mathematics with Applications*, 54 (2):267-281, 2007.

[6] M.J. Pardo and D. de la Fuente, Optimal selection of the service rate for a finite input source fuzzy queuing system. *Fuzzy Sets and Systems*,. 159 (3):325-342, 2008.

[7] L.A. Zadeh, Fuzzy sets. *Information and control*, 8:338-353, 1965.

[8] M.J. Pardo and D. de la Fuente, Design of a fuzzy finite capacity queuing model based on the degree of customer satisfaction: analysis and fuzzy optimization. *Fuzzy Sets and Systems*, 159 (24):3313-3332, 2008.

[9] D. Dubois, H.M. Prade and R.R. Yager, *Merging fuzzy information*, in: Fuzzy Sets in Approximate Reasoning and Information Systems. Boston: Kluwer, 1999.

[10] H.J. Zimmermann, *Fuzzy Set Theory and its applications*, 4th ed. Boston: Kluwer, 2001.

[11] T.L. Saaty, *Elements of queuing theory with applications*. New York: Dover, 1983.

[12] W.L. Winston, *Operations Research: Applications and Algorithms*, 4th ed. Boston: Duxbury Press, 2003.

[13] L.A. Zadeh, Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23:421-427, 1968.

ISBN: 978-989-95079-6-8

Valuations on the Algebra of Intervals

Mirko Navara

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University in Prague
Technická 2, CZ-166 27, Praha, Czech Republic
Email: navara@cmp.felk.cvut.cz

Abstract— *Interval-valued fuzzy sets are based on the algebra of subintervals of the unit interval* $[0,1]$. *We study valuations as a special type of measures on this algebra. We present a description of all valuations which preserve the standard fuzzy negation and extend the identity on the elements of the form* (x,x). *Consequences for sublattices are formulated.*

Keywords— Girard algebra, interval-valued fuzzy set, intuitionistic fuzzy set, measure, valuation

1 Introduction

In order to work with vague quantities, fuzzy sets were suggested [13]; their membership degrees are in the unit interval $I = [0,1]$. Sometimes this approach appeared too restrictive; instead of one value, the membership degree of a point was restricted to an interval in I. This led to the notion of *interval-valued fuzzy sets*. The original term was *intuitionistic fuzzy sets* [1, 2], but it causes misunderstandings because there is no relevance to the intuitionistic logic. Other synonyms found in the literature are *IFS events* [8], etc. The idea of using more general membership degrees is rather old, cf. [4].

Another motivation of interval-valued fuzzy sets is the following: If we determine the membership degree of a point p by an interval $(x,y) \in I^2$, $x \le y$, we say that p *belongs* to the fuzzy set with the degree at least x and *does not belong* to it with the degree at least $z = 1-y$; the remainder $1-(x+z) = y-x \ge 0$ represents missing knowledge. This approach works with pairs $(x,z) \in I^2$, $x+z \le 1$, representing *bipolar* information. It is distinguished for philosophical, psychological, and methodological reasons; from the mathematical point of view, the descriptions by the pair (x,y) or (x,z) lead to isomorphic models and results can be easily translated from one context to the other.

We refer to [3, 5, 6, 7, 8] for recent investigation of properties of interval-valued fuzzy sets, operations and measures on them.

2 Standard interval algebra

Here we define the standard algebra of all intervals in I. In the sequel, we shall also consider its subalgebras.

We shall work with the set

$$I_2 = \left\{(x,y) \in I^2 \mid x \le y\right\} . \tag{1}$$

Equipped with the usual pointwise ordering,

$$(x_1,y_1) \le (x_2,y_2) \iff (x_1 \le x_2 \text{ and } y_1 \le y_2) , \tag{2}$$

I_2 becomes a lattice with the least element $\mathbf{0} = (0,0)$ and the greatest element $\mathbf{1} = (1,1)$ and the lattice operations

$$(x_1,y_1) \wedge (x_2,y_2) = (\min(x_1,y_1), \min(x_2,y_2)) , \tag{3}$$
$$(x_1,y_1) \vee (x_2,y_2) = (\max(x_1,y_1), \max(x_2,y_2)) . \tag{4}$$

The sublattice

$$D = \{(x,x) \mid x \in I\} \subset I_2 \tag{5}$$

consists of the elements $(x,x) \in I_2$ which represent real membership degrees corresponding to $x \in I$ in standard fuzzy sets. Thus I_2 represents an extension of $D \cong I$ by elements $(x,y) \in I^2$, $x < y$.

The standard fuzzy negation $'\colon I \to I$, $x' = 1-x$, has a natural extension to the negation $'\colon I_2 \to I_2$,

$$(x,y)' = (y',x') . \tag{6}$$

Thus we have a bounded lattice with an antitone involutive operation $'$. We call this structure the *standard interval algebra* and—with some abuse of notation—we denote it again by I_2.

The standard interval algebra is often considered with additional operations. These are often derived from the *standard MV-algebra* $([0,1],0,\oplus,')$, where $'$ is the standard fuzzy negation and $\oplus$ is the Łukasiewicz t-conorm $a \oplus b = \min(1, a+b)$. Its dual operation is the Łukasiewicz t-norm $a \odot b = (a' \oplus b')' = \max(0, a+b-1)$. These operations are applied to elements of I_2 coordinatewise,

$$(x_1,y_1) \odot (x_2,y_2) = (x_1 \odot x_2, y_1 \odot y_2) , \tag{7}$$
$$(x_1,y_1) \oplus (x_2,y_2) = (x_1 \oplus x_2, y_1 \oplus y_2) . \tag{8}$$

Then $D \subset I_2$ becomes isomorphic to the standard MV-algebra.

Remark 1. *Other t-norms and t-conorms on* I_2, *not originated this way, are studied, e.g., in [3]. We do not deal with them here.*

Another line of research is due to S. Weber [10, 11]. He extends the MV-algebra D to a Girard algebra with operations

$$(x_1,y_1) \sqcap (x_2,y_2) = (x_1 \odot x_2, \max((x_1 \odot y_2), (y_1 \odot x_2))) , \tag{9}$$
$$(x_1,y_1) \sqcup (x_2,y_2) = (\min((x_1 \oplus y_2), (y_1 \oplus x_2)), y_1 \oplus y_2) . \tag{10}$$

Although rather strange, these operations arise as a unique solution of well-motivated algebraic conditions [11]. On D, they

ISBN: 978-989-95079-6-8

coincide with the Łukasiewicz operations and with (7), (8). The lattice reduct $(I_2, \vee, \wedge)$ of the Girard algebra is the same as in the former approach.

As we shall deal only with the lattice operations, the (standard) negation, and with the Łukasiewicz operations on D, we need not deal with the differences between the above two approaches; we study features which are common to both of them.

3 Measures on the interval algebra

In order to study probability of fuzzy events, we need to introduce (probability) measures on fuzzy sets. First of all, we need measures on the algebras of membership degrees; in our case, on the standard interval algebra I_2. Different conditions are imposed in order to extend the notion of probability from classical sets to fuzzy sets and interval-valued fuzzy sets. Here we follow one of the approaches suggested by S. Weber in [11], but analogous conditions are studied in many other papers. We refer to [12] for a comparison of many approaches to measures on collections of fuzzy sets.

Without referring to a specific definition, let us accept that the standard MV-algebra admits one very natural candidate for a measure—the identity. It commutes with the standard negation and satisfies all forms of additivity conditions (see [9] for details). Thus here we assume that the only measure on the standard MV-algebra is the identity. Also S. Weber in [11] studies measures m which are extensions of this measure on D, i.e.,

$$\forall x \in I : m(x,x) = x\,. \tag{M1+}$$

(For simplicity, we write $m(x,x)$ instead of $m((x,x))$.) For all elements $(x,y) \in I_2$, he imposes the following conditions as minimal requirements on so-called *uncertainty measures*:

$$m(\mathbf{0}) = 0\,, \qquad m(\mathbf{1}) = 1\,, \tag{M1}$$

$$A \le B \implies m(A) \le m(B)\,, \tag{M2}$$

$$m\,(A') = 1 - m(A)\,. \tag{M3}$$

(Condition (M1) follows from (M1+).) One of the additional conditions studied in [11] is the *valuation property*:

$$m(A \wedge B) + m(A \vee B) = m(A) + m(B)\,. \tag{V}$$

In this paper, we study mappings $m\colon I_2 \to [0,1]$ satisfying (M1+), (M2), (M3), and (V); we call them *valuation measures*.

3.1 Characterization of valuation measures on the standard interval algebra

Our principal contribution is the following (μ is a $\frac{1}{2}$-*Lipschitz* function if $|\mu(u) - \mu(v)| \le \frac{1}{2}\ |u - v|$ for all u, v):

Theorem 2. *A mapping $m\colon I_2 \to [0,1]$ is a valuation measure if and only if there is a $\frac{1}{2}$-Lipschitz function μ on $\left[0, \frac{1}{2}\right]$ such that m is of the form*

$$m(x,y) = \frac{x+y}{2} + \mu\left(\left|x - \tfrac{1}{2}\right|\right) - \mu\left(\left|y - \tfrac{1}{2}\right|\right)\,. \tag{12}$$

Proof: First, let us suppose that μ is a $\frac{1}{2}$-Lipschitz function and m is given by (12). If $x = y$, we obtain (M1+). Condition (M3) is verified as follows:

$$\begin{aligned} &m(1-y, 1-x)\\ &= \frac{1-y+1-x}{2} + \mu\left(\left|(1-y) - \tfrac{1}{2}\right|\right) - \mu\left(\left|(1-x) - \tfrac{1}{2}\right|\right)\\ &= 1 - \frac{x+y}{2} - \mu\left(\left|x - \tfrac{1}{2}\right|\right) + \mu\left(\left|y - \tfrac{1}{2}\right|\right)\\ &= 1 - m(x,y)\,. \end{aligned} \tag{13}$$

Monotonicity (M2) can be verified in each coordinate separately and follows from the $\frac{1}{2}$-Lipschitz condition. For $x_1 < x_2$, we obtain

$$\begin{aligned} &m(x_2,y) - m(x_1,y)\\ &= \frac{x_2 - x_1}{2} + \mu\left(\left|x_2 - \tfrac{1}{2}\right|\right) - \mu\left(\left|x_1 - \tfrac{1}{2}\right|\right)\\ &\ge \frac{x_2 - x_1}{2} - \frac{1}{2}\left|\left|x_2 - \tfrac{1}{2}\right| - \left|x_1 - \tfrac{1}{2}\right|\right|\\ &\ge \frac{x_2 - x_1}{2} - \frac{|x_2 - x_1|}{2} = 0\,. \end{aligned} \tag{14}$$

For $y_1 < y_2$, we obtain

$$\begin{aligned} &m(x,y_2) - m(x,y_1)\\ &= \frac{y_2 - y_1}{2} + \mu\left(\left|y_2 - \tfrac{1}{2}\right|\right) - \mu\left(\left|y_1 - \tfrac{1}{2}\right|\right)\\ &\ge \frac{y_2 - y_1}{2} - \frac{1}{2}\left|\left|y_2 - \tfrac{1}{2}\right| - \left|y_1 - \tfrac{1}{2}\right|\right|\\ &\ge \frac{y_2 - y_1}{2} - \frac{|y_2 - y_1|}{2} = 0\,. \end{aligned} \tag{15}$$

To check the valuation property (V), take $A = (x_1, y_1)$, $B = (x_2, y_2)$. Notice that, independently of all possible orderings of x_1, x_2 (resp. y_1, y_2), $m(A \wedge B) + m(A \vee B)$ expressed by (12) contains the same summands as $m(A) + m(B)$:

$$\begin{aligned} &m(A \wedge B) + m(A \vee B)\\ &= \frac{x_1 + x_2 + y_1 + y_2}{2}\\ &+ \mu\left(\left|x_1 - \tfrac{1}{2}\right|\right) + \mu\left(\left|x_2 - \tfrac{1}{2}\right|\right)\\ &- \mu\left(\left|y_1 - \tfrac{1}{2}\right|\right) - \mu\left(\left|y_2 - \tfrac{1}{2}\right|\right)\\ &= m(A) + m(B)\,. \end{aligned} \tag{16}$$

Thus m is a valuation measure.

To prove the reverse implication, let us assume that m is a valuation measure. Whenever $x + y = 1$, the element (x,y) is invariant to the negation and, due to (M3), $m(x,y) = \frac{1}{2}$.

We may define

$$\mu(z) = \tfrac{z}{2} + m\left(\tfrac{1}{2} - z, \tfrac{1}{2}\right)\,, \tag{17}$$

in particular,

$$\mu(0) = m\left(\tfrac{1}{2}, \tfrac{1}{2}\right) = \tfrac{1}{2}\,. \tag{18}$$

For $x \le y = \frac{1}{2}$, the right-hand side of (12) is

$$\begin{aligned} \tfrac{x}{2} + \tfrac{1}{4} + \mu\left(\tfrac{1}{2} - x\right) - \mu\left(0\right) &= \tfrac{x}{2} - \tfrac{1}{4} + \mu\left(\tfrac{1}{2} - x\right)\\ &= m\left(x, \tfrac{1}{2}\right) \end{aligned} \tag{19}$$

as desired.

ISBN: 978-989-95079-6-8

For $x \le y < \frac{1}{2}$, we take $A = \left(x, \frac{1}{2}\right)$, $B = (y, y)$. Then $A \wedge B = (x, y)$, $A \vee B = \left(y, \frac{1}{2}\right)$. Using the valuation property, (19), and (M1+), we obtain

$$\begin{aligned} m(x,y) &= m(A \wedge B) = m(A) + m(B) - m(A \vee B) \\ &= m\left(x, \tfrac{1}{2}\right) + m(y,y) - m\left(y, \tfrac{1}{2}\right) \\ &= \tfrac{x}{2} + \mu\left(\tfrac{1}{2} - x\right) + y - \tfrac{y}{2} - \mu\left(\tfrac{1}{2} - y\right) \\ &= \tfrac{x}{2} + \mu\left(\tfrac{1}{2} - x\right) + \tfrac{y}{2} - \mu\left(\tfrac{1}{2} - y\right) \end{aligned} \tag{20}$$

in accordance with (12).

The case $\frac{1}{2} < x \le y$ is obtained by duality.

In the remaining case $x \le \frac{1}{2} < y$, without loss of generality we may suppose that $x + y \le 1$ (the other case is dual). We take $A = (x, y)$, $B = \left(1 - y, \frac{1}{2}\right)$. Then $A \wedge B = \left(x, \frac{1}{2}\right)$, $A \vee B = (1 - y, y)$. Using the valuation property and (19), we obtain

$$\begin{aligned} m(x,y) = m(A) &= m(A \wedge B) + m(A \vee B) - m(B) \\ &= m\left(x, \tfrac{1}{2}\right) + m(1-y, y) - m\left(1-y, \tfrac{1}{2}\right) \\ &= \tfrac{x}{2} + \mu\left(\tfrac{1}{2} - x\right) + \tfrac{1}{2} - \tfrac{1-y}{2} + \mu\left(\tfrac{1}{2} - (1-y)\right) \\ &= \tfrac{x}{2} + \mu\left(\tfrac{1}{2} - x\right) + \tfrac{y}{2} - \mu\left(y - \tfrac{1}{2}\right) . \end{aligned} \tag{21}$$

It remains to prove that μ is $\frac{1}{2}$-Lipschitz. Suppose that $0 \le x_1 \le x_2 \le \frac{1}{2}$. One inequality follows from the monotonicity of m:

$$\begin{aligned} \mu(x_2) - \mu(x_1) &= \tfrac{x_2}{2} + m\left(\tfrac{1}{2} - x_2, \tfrac{1}{2}\right) - \tfrac{x_1}{2} - m\left(\tfrac{1}{2} - x_1, \tfrac{1}{2}\right) \\ &\le \tfrac{x_2 - x_1}{2} . \end{aligned} \tag{22}$$

To prove the second inequality, we take $A = \left(\frac{1}{2} - x_2, \frac{1}{2}\right)$, $B = \left(\frac{1}{2} - x_1, \frac{1}{2} - x_1\right)$. Then $A \wedge B = \left(\frac{1}{2} - x_2, \frac{1}{2} - x_1\right)$, $A \vee B = \left(\frac{1}{2} - x_1, \frac{1}{2}\right)$. We obtain

$$m(A \vee B) - m(A) = m(B) - m(A \wedge B) , \tag{23}$$

$$\begin{aligned} & m\left(\tfrac{1}{2} - x_1, \tfrac{1}{2}\right) - m\left(\tfrac{1}{2} - x_2, \tfrac{1}{2}\right) \\ &= m\left(\tfrac{1}{2} - x_1, \tfrac{1}{2} - x_1\right) - m\left(\tfrac{1}{2} - x_2, \tfrac{1}{2} - x_1\right) \\ &\le m\left(\tfrac{1}{2} - x_1, \tfrac{1}{2} - x_1\right) - m\left(\tfrac{1}{2} - x_2, \tfrac{1}{2} - x_2\right) = x_2 - x_1 , \end{aligned} \tag{24}$$

$$\begin{aligned} & \mu(x_1) - \mu(x_2) \\ &= \tfrac{x_1}{2} + m\left(\tfrac{1}{2} - x_1, \tfrac{1}{2}\right) - \tfrac{x_2}{2} - m\left(\tfrac{1}{2} - x_2, \tfrac{1}{2}\right) \\ &\le \tfrac{x_1 - x_2}{2} + x_2 - x_1 = \tfrac{x_2 - x_1}{2} . \end{aligned} \tag{25}$$

Corollary 3. *Every valuation measure on I_2 is continuous, even 1-Lipschitz.*

Function μ in (12) is determined by m uniquely up to an additive constant. Thus we may choose, e.g., $\mu(0) = 0$, then μ attains values from $\left[-\frac{1}{4}, \frac{1}{4}\right]$.

Example 4. *For $\mu_0 = 0$, we obtain the valuation measure*

$$m_0(x,y) = \frac{x+y}{2} \tag{26}$$

which is additive with respect to $\oplus, \odot$.

The extreme cases are $\mu_1(z) = \frac{z}{2}$, $\mu_{-1}(z) = -\frac{z}{2}$, which lead to

$$m_1(x,y) = \begin{cases} y & \text{if } x \le y \le \frac{1}{2} , \\ \frac{1}{2} & \text{if } x \le \frac{1}{2} < y , \\ x & \text{if } \frac{1}{2} < x \le y , \end{cases} \tag{27}$$

$$m_{-1}(x,y) = \begin{cases} x & \text{if } x \le y \le \frac{1}{2} , \\ x + y - \frac{1}{2} & \text{if } x \le \frac{1}{2} < y , \\ y & \text{if } \frac{1}{2} < x \le y . \end{cases} \tag{28}$$

The choice of function μ is not limited to linear functions in (12).

3.2 Convex structure of the space of valuation measures

In this section, we consider the set $\mathcal{M}$ of all valuation measures on I_2. We consider $\mathcal{M}$ equipped with the product topology inherited from $[0,1]^{I_2} \supseteq \mathcal{M}$. As the defining conditions of valuation measures can be written in the form of equations using only continuous operations, the set $\mathcal{M}$ is closed. It can be easily checked that a convex combination of valuation measures is a valuation measure; thus $\mathcal{M}$ is also convex. Valuation measures μ_1, μ_{-1} of Example 4 are (some of) its extreme points.

Moreover, convex combinations are preserved by (12) in the following sense:

Proposition 5. *Let m_a, resp. m_b, be a valuation measure on I_2 and μ_a, resp. μ_b, the corresponding $\frac{1}{2}$-Lipschitz function satisfying (12). Let $\lambda \in [0,1]$. Then the convex combination $\lambda\, m_a + (1-\lambda)\, m_b$ is a valuation measure which satisfies (12) for $\mu = \lambda\, \mu_a + (1-\lambda)\, \mu_b$.*

4 Generalizations

Theorem 2 described valuation measures on the standard interval algebra. This case is important, but very special. It describes all interval-valued fuzzy subsets of a singleton. Here we extend Theorem 2 to more general cases.

4.1 Valuation measures on the set of all interval-valued fuzzy sets

Now we shall extend Theorem 2 to finite products of algebras isomorphic to the standard interval algebra. These products are isomorphic to the collections of all interval-valued fuzzy subset of a finite set.

Let $n \in \mathbb{N}$. We denote by I_2^n the product of n standard interval algebras, i.e., the cartesian product $\prod_{i=1}^n I_2$ with the lattice operations and negation applied to each coordinate separately.

Theorem 6. *Let $n \in \mathbb{N}$. A mapping $m\colon I_2^n \to [0,1]$ is a valuation measure if and only if there are valuation measures $m_1, \ldots, m_n$ on I_2 and real coefficients $c_1, \ldots, c_n \in [0,1]$ such that $\sum_{i=1}^n c_i = 1$ and m is of the form*

$$m(A_1, \ldots, A_n) = \sum_{i=1}^n c_i\, m_i(A_i) . \tag{29}$$

Proof: Suppose first that m is of the form (29), i.e., a convex combination of valuation measures $m_1, \ldots, m_n$ on I_2. For each $i = 1, \ldots, n$, we denote by π_i the projection on the ith coordinate,

$$\pi_i(A_1, \ldots, A_n) = A_i . \tag{30}$$

ISBN: 978-989-95079-6-8

It is a lattice homomorphism of I_2^n onto I_2. Thus the composition $m_i \circ \pi_i \colon I_2^n \to [0,1]$,

$$(m_i \circ \pi_i)(A_1, \ldots, A_n) = m_i(\pi_i(A_1, \ldots, A_n)) = m_i(A_i)\,, \tag{31}$$

is a valuation measure on I_2^n. As a convex combination of valuation measures, m is also a valuation measure.

Second, suppose that m is a valuation measure on I_2^n. Let $\mathcal{B}$ be the collection of all sharp (=crisp) elements of I_2^n, i.e., all n-tuplets $(A_1, \ldots, A_n)$ such that $A_i \in \{\mathbf{0}, \mathbf{1}\}$ for all $i = 1, \ldots, n$. Then $\mathcal{B}$ is a finite Boolean algebra with n atoms $(\delta_{i1}, \ldots, \delta_{in})$, $i = 1, \ldots, n$, where

$$\delta_{ik} = \begin{cases} \mathbf{1} & \text{if } i = k\,, \\ \mathbf{0} & \text{if } i \neq k\,. \end{cases} \tag{32}$$

This means that $(\delta_{i1}, \ldots, \delta_{in}) = (\mathbf{0}, \ldots, \mathbf{0}, \mathbf{1}, \mathbf{0}, \ldots, \mathbf{0}) \in I_2^n$, where the unit is in the ith coordinate. A valuation measure on a Boolean algebra is a classical probability measure. Thus m it is of the form

$$m(A_1, \ldots, A_n) = \sum_{i=1}^{n} c_i\, s_i(A_1, \ldots, A_n)\,, \tag{33}$$

where $c_i = m(\delta_{i1}, \ldots, \delta_{in})$, $i = 1, \ldots, n$, and $s_1, \ldots, s_n$ are valuation measures on I_2^n such that

$$s_i(\delta_{i1}, \ldots, \delta_{in}) = 1\,. \tag{34}$$

As

$$s_i((\delta_{i1}, \ldots, \delta_{in})') = s_i(\delta'_{i1}, \ldots, \delta'_{in}) = 0\,, \tag{35}$$

we obtain

$$s_i(B_1, \ldots, B_n) = 0 \tag{36}$$

whenever

$$(B_1, \ldots, B_n) \leq (\delta_{i1}, \ldots, \delta_{in})'\,. \tag{37}$$

Thus $s_i(A_1, \ldots, A_n)$ depends only on the ith coordinate, $A_i = \pi_i(A_1, \ldots, A_n)$. We may express s_i in the form $s_i = m_i \circ \pi_i$ for some mapping $m_i \colon I_2 \to [0,1]$. The image of I_2^n under π_i is isomorphic to I_2. As π_i is a homomorphism, m_i must be a valuation measure on I_2. The proof is finished.

The characterization can be extended also to infinite products of the standard interval algebra, thus to collections of interval-valued fuzzy subsets of an infinite universe. This requires valuation measures concentrated not only in coordinates, but also in the points of the underlying Stone space of the Boolean algebra in question. This way, we may also admit subalgebras of products, so that the underlying Boolean algebra is not the whole power set. We leave these extensions for future research.

Theorems 2 and 6 can be combined. We obtain a characterization of valuation measures on the product I_2^n using $\frac{1}{2}$-Lipschitz functions μ_i, $i = 1, \ldots, n$. These functions can be different in different coordinates.

4.2 Extension to subalgebras

S. Weber in [11] investigates valuation measures on finite Girard algebras, mainly on finite subalgebras of I_2 of the form

$$\tilde{\mathbb{C}}_n = \left\{ \left(\tfrac{j}{n}, \tfrac{k}{n}\right) \middle| j, k \in \mathbb{Z},\ 0 \leq j \leq k \leq n \right\}\,, \tag{38}$$

where $n \geq 1$ is a fixed integer. Valuation measures on I_2, when restricted to $\tilde{\mathbb{C}}_n$, become valuation measures on $\tilde{\mathbb{C}}_n$. Conversely, each valuation measure on $\tilde{\mathbb{C}}_n$ admits an extension to a valuation measure on I_2. This holds also for other sublattices of I_2:

Theorem 7. *Let L be a sublattice of I_2. Suppose that $(0,0), (1,1) \in L$ and $(x,y) \in L$ implies (x,x), (y,y), $(y',x') \in L$. A mapping $m \colon L \to [0,1]$ is a valuation measure if and only if it satisfies the condition of Theorem 2.*

The proof follows the pattern of Theorem 2. Analogously, we can also generalize Theorem 6 and obtain a characterization of valuation measures on subalgebras of I_2^n. These subalgebras are isomorphic to general collections of (not all possible) interval-valued fuzzy sets. For this, it is necessary to add assumptions on closedness of these collections under the lattice operations. We leave further details to future research.

5 Conclusions

We studied measures on finite products of the algebra of subintervals of $[0,1]$. We gave a characterization of measures which (besides the obvious properties of monotonicity and boundary conditions)

1. are compatible with the negation,
2. extend the identity on elements of the form (x,x), $x \in [0,1]$,
3. are valuations.

This is aimed as the first step towards the characterization of such measures on collections of interval-valued fuzzy sets.

Our conditions do not represent the only choice, rather a necessary minimum. Other conditions could be added, depending on the application domain. Our general characterization forms a basis for description of measures restricted by additional conditions, too.

Acknowledgment

The author acknowledges the support by the Czech Ministry of Education under project MSM 6840770038.

References

[1] K. T. Atanassov: Intuitionistic fuzzy sets. *Fuzzy Sets Syst.* 20(1) (1986) 87–96.

[2] K. T. Atanassov: *Intuitionistic Fuzzy Sets*. Physica-Verlag, Heidelberg/New York, 1999.

[3] G. Deschrijver, C. Cornelis, and E. E. Kerre: On the representation of intuitionistic fuzzy t-norms and t-conorms. *IEEE Trans Fuzzy Syst.* 12(1) (2004) 45–61.

[4] J. Goguen: L-fuzzy sets. *J. Math. Anal. Appl.* 18(1) (1967) 145–174.

[5] P. Grzegorzewski and E. Mrówka: Probability of intuitionistic fuzzy events. *Soft Methods in Probability, Statistics and Data Analysis*, Physica Verlag, New York, 2002, 105–115.

ISBN: 978-989-95079-6-8

[6] M.G. Karunambigai, P. Rangasamy, K. Atanassov, N. Palaniappan: An intuitionistic fuzzy graph method for finding the shortest paths in networks. In: O. Castillo, P. Melin, O.M. Ross, R.S. Cruz, W. Pedrycz, J. Kacprzyk: *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing*, Proc. IFSA Congress 2007, Springer, Berlin/Heidelberg, 2007, 3–10.

[7] B. Riečan B.: A descriptive definition of the probability on intuitionistic fuzzy sets. *Proc. EUSFLAT'2003*, Zittau, Goerlitz Univ. Appl. Sci., 2003, 263–266.

[8] B. Riečan B.: Representation of probabilities on IFS events. *Advances in Soft Computing, Soft Methodology and Random Information Systems*, Springer, Berlin, 2004, 243–246.

[9] B. Riečan and D. Mundici: Probability on MV-algebras. In: E. Pap (ed.), *Handbook of Measure Theory*, Elsevier Science, Amsterdam, 2002, Chapter 21, 869–910.

[10] S. Weber: Uncertainty measures, decomposability and admissibility. *Fuzzy Sets Syst.* 40 (1991) 395–405.

[11] S. Weber: Uncertainty measures—Problems concerning additivity. *Fuzzy Sets Syst.* 160(3) (2009) 371–383.

[12] S. Weber and E. P. Klement: Fundamentals of a generalized measure theory. In: U. Höhle et al. (eds.), *Mathematics of Fuzzy Sets. Logic, Topology, and Measure Theory*, Kluwer Academic Publishers, Dordrecht, 1999, 633–651.

[13] L. A. Zadeh: Fuzzy Sets: *Inform. Control* 8 (1965) 338–353.

ISBN: 978-989-95079-6-8

Impute Missing Assessments by Opinion Clustering in Multi-Criteria Group Decision Making Problems

Jun Ma[1], Guangquan Zhang[1], Jie Lu[1] Da Ruan[2,3]

1. School of Software Engineering, Faculty of Engineering and Information Technology
University of Technology, Sydney (UTS), PO Box 123, Broadway, NSW 2007, Australia
2. Beglian Nuclear Research Centre (SCK•CEN), 2400, Boeretong, Mol, Belgium
3. Department of Applied Math. & Computer Science, Ghent University, 9000 Gent, Belgium
Email: {junm, zhangg, jielu}@it.uts.edu.au, druan@sckcen.be, da.ruan@ugent.be

Abstract— Multi-criteria group decision-making and evaluation (MCGDME) method typically aggregates information in evaluation tables. For various reasons, evaluation tables (decision matrix) often include missing data that highly affect correct decision-making and evaluation. Most existing imputation methods of missing data are based on statistical features which do not exist in an MCGDME setting. This paper proposes an imputation method of missing data (IMD) in evaluation tables. The IMD method measures the similarity betweent two evaluators' mental models. Evaluators are then classed into several groups based on their similarities by using fuzzy clustering methods. Finally, missing data are imputated under the assumption that the imputated value of missing data does not change the previous clustering results. The proposed IMD method is implemented and tested in two numerical experiments.

Keywords— decision making, missing data, multi-criteria evaluation, opinion clustering, aggregation

1 Introduction

Multi-criteria group decision-making (MCGDM) methods have been widely used in various fields such as new product development in industries [9], clinical diagnose, service strategies selection in business, as well as public economic policy evaluation [11]. In a typical MCGDM process, evaluating alternatives and aggregating assessments based on decision matrices are two crucial processes. Evaluators (or experts) assess alternatives in terms of a set of identified criteria. The assessments are recorded into several decision matrices, also known as decision tables, which are further analyzed and integrated. The aggregation process implements a specific aggregation model to achieve an overall assessment for each alternative and based on which the best alternative(s) is/are selected. Majority of existing research focuses particularly on the problem of how to efficiently integrate heterogeneous quantitative or qualitative information from multiple information sources. Most processes are conducted mainly on the assumption that those obtained decision matrices are intact, i.e., those tables do not include missing values. However, this assumption may not be completely satisfied in real situations.

Various reasons result in missing values in decision matrices. Values may be missed in the course of data collection by device faults or human errors. Evaluators who are lack of relevant background for some criteria or are not willing to provide opinions on some criteria may leave blanks in their decision tables. Security or privacy policies may also prevent evaluators from providing answers to some criteria. An evidently result of missing values reduce the reliability of decision making or evaluation and to produce extraordinary bias towards appropriate decisions. Moreover, missing values may result in invalid evaluation and decision making. Hence, necessary and efficient missing data imputation is required in real applications such as in nuclear material safeguards information management.

Literature indicates missing data processing has a long research history [4–6, 8, 12]. However, most research is mainly conducted on the theory and application of statistics sciences [5]. Existing techniques, such as the LOCF (last observation carried forward), best or worst case imputation, mean imputation, and case deletion [1], generally require the collected data has specific and identifiable features. Such features include, for examples, the data is time-related or follows a particular statistics distribution, as well as has sufficient volume. These features do not hold in a typical MCGDM settings in general. Assessments in a typical MCGDM problem have uniqueness in the sense that each assessment is uniquely linked to a unique alternative, a unique evaluator, and a unique criterion. Moreover, because evaluators are generally not asked or are not willing to repeat same evaluation procedures, the obtained assessment values are always onetime data which is irrelevant to observation time. Therefore, not enough volume of data exists with the same features of a missing value. Besides the uniqueness and small volume of assessments, assessment values are generated through information processing in evaluators' mental model of the given decision problem which cannot be strictly expressed through a formalized and quantified model. Moreover, the links among the selected criteria are generally weak and unclear, because evaluation criteria are always deliberately selected to represent different considerations for alternatives. These features make it very hard to use statistics-based missing data imputation methods to MCGDM problems.

This paper presents an imputation method for missing assessments (IMD) in decision tables for MCGDM. The presented method mainly includes two steps. A fuzzy clustering method is first used to classify evaluators into several groups based on their opinions reflected in their assessments. Then missing assessments are imputed according to those assessments of evaluators belonging to the same group. The rest of the paper is organized as follows. Section 2 reviews the main steps of the presented IMD method. Details of each step is introduced in Section 3. To further illustrate the IMD method,

ISBN: 978-989-95079-6-8

two case studies are conducted. Finally, main results and future works are discussed in Section 5.

2 Overview

2.1 Formalization of typical MCGDM problems

A typical MCGDM problem is composed of a set of alternatives (e.g., action choices, policies), a hierarchy of attributes (i.e., decision criteria), weights associated with criteria, a set of decision matrices (i.e., decision tables), as well as a set of evaluators who present those decision matrices. Three main steps are involved in selections of alternatives [13]: 1) determine the relevant criteria and alternatives; 2) evaluate the relative impacts of alternatives on those criteria; and 3) determine a ranking of each alternatives. Roughly, an MCGDM process can be expressed by a 4-tuple model:

$$\mathcal{M} = (\mathcal{C}, \mathcal{E}, \mathcal{A}, \mathcal{T}), \tag{1}$$

where $\mathcal{C} = \{(c_j, wc_j)|j = 1, 2, \ldots, n\}$ is a set of criteria and their corresponding weights (importance); $\mathcal{E} = \{(e_k, we_k)|k = 1, 2, \ldots, m\}$ is a set of evaluators (experts) and their corresponding weights (reliabilities); $\mathcal{A} = \{a_i|i = 1, 2, \ldots, p\}$ is a set of alternatives (action choices); and $\mathcal{T} = \{T_i = (v^i_{jk})_{n\times m}|i = 1, 2, \ldots, p\}$ is a set of decision matrices. For each a_i, the overall assessment y_i based on the model $\mathcal{M}$ is explicitly expressed as

$$\begin{aligned} y_i &= W_C \circ T_i \diamond W_E^{\mathrm{T}} \\ &= (wc_1, \cdots, wc_n) \circ \begin{pmatrix} v^i_{11} & v^i_{12} & \cdots & v^i_{1m} \\ v^i_{21} & v^i_{22} & \cdots & v^i_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v^i_{n1} & v^i_{n2} & \cdots & v^i_{nm} \end{pmatrix} \diamond \begin{pmatrix} we_1 \\ we_2 \\ \vdots \\ we_m \end{pmatrix} \end{aligned}$$

where $\circ$ and $\diamond$ are operations in $\mathcal{M}$, W_C is the weight vector for criteria, and W_E the weight vector for evaluators. Once $\circ$ and $\diamond$ are given in explicit forms, the model $\mathcal{M}$ is then determined. For example, let $\circ$ and $\diamond$ be a weighted sum. For any vectors $\boldsymbol{w}$ and $\boldsymbol{v}$

$$\boldsymbol{w} \circ \boldsymbol{v}^{\mathrm{T}} = (w_1, \cdots, w_n) \circ (v_1, \cdots, v_n)^{\mathrm{T}} = \sum_{i=1}^{n} w_i v_i, \tag{2}$$

then y_i can be written as

$$y_i = \sum_{j=1}^{n} \left(wc_i \cdot \sum_{k=1}^{m} v_{jk} we_k \right) = \sum_{k=1}^{m} \left(\sum_{j=1}^{n} wc_j v_{jk} \right) \cdot we_k. \tag{3}$$

Note that most research works on MCGDM focus on how to define these two operations which are mostly studied in the framework of aggregation operators.

In the following, let $T_i = (\boldsymbol{v}^i_1, \boldsymbol{v}^i_2, \cdots, \boldsymbol{v}^i_k)$, where $\boldsymbol{v}^i_k = (v^i_{1k}, v^i_{2k}, \cdots, v^i_{nk})^{\mathrm{T}}$ is called the evaluation vector of evaluator e_k on alternative a_i. T_i is called complete if no missing values in T_i. T_i is called incomplete if at least one evaluation vector includes missing values. For convenience, the following analysis will omit the alternative index i.

2.2 Distance and similarity of evaluator opinions

For each $\boldsymbol{v}_k$ in T, $\boldsymbol{v}_k$ can be seen as an output of evaluator e_k's mental model with the physical nature of alternative as input and denoted by x, i.e., this can be formally described as

$$\boldsymbol{v}_k = f_k(x), \quad k = 1, 2, \ldots, m. \tag{4}$$

Figure 1 illustrates a generalized procedure of obtaining an evaluation table. Notice that the input of each evaluator's mental model is the same, the difference among the obtained evaluation vectors can only be introduced by the difference among evaluators' mental models. Assume that the relationship between evaluators' mental models exists and can be found through their outputs, i.e., the evaluation vectors, then missing values can be approximately imputed with the aid of those relationships. In the light of this idea, the following strategy is adopted to impute missing values. First, a distance measure d between two evaluation vectors is defined, which is then taken as the distance between two evaluators' opinions. Next, the similarity of two evaluators' mental models is induced from the distance. Based on the similarities between each pair of evaluators, a fuzzy clustering method is used to segment evaluators to several clusters. For each of missing values, a predicted value is obtained according to the evaluation vectors provided by evaluators belonging to the same group. This value is the imputed value.

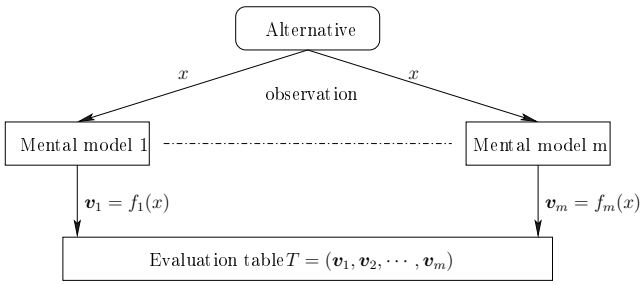


Figure 1: Producing of an evaluation table.

To implement the above-mentioned strategy, we consider some relevant issues. First, the difference of criteria weights is considered when defining a distance measurement. The weight of a criterion indicates the importance degree or relevance degree of that criterion to the decision goal. Some criteria are more important than others. Hence, criteria with bigger weights should contribute more distance to the overall distance between two evaluation vectors. For this reason, we use an aggregation operator to define the overall distance between two evaluation vectors. Another issue is concerned about criteria to that not all missing values must be imputed. For criteria with very small weights, the imputation of missing data is not essential because the bias from appropriate decision induced by discrepancy in those criteria is insignificant. Thus, a parameter α is defined to filter criteria with small weights. Missing data with respect to those filtered criteria are not imputed. Moreover, a second parameter β is used in the segmentation of evaluators, which indicates the at least similarity required when two evaluators are treated with the similar opinion. It can be proved that the bigger the β is, the more obtained clusters and the reduced number of evaluators belong to a certain cluster.

Because the presented strategy adopts two extra parameters α and β, the presented method is thus called an α-β imputation. Next section will give details of the α-β imputation method.

ISBN: 978-989-95079-6-8

3 An α-β imputation method

This section illustrates the details of the α-β imputation method for missing data (assessments, evaluations). For convenience, let all weights for criteria and evaluators be real numbers in the real unit interval [0, 1] and all assessment values be real numbers in [0, 1], too. Without loss of generality, the following discussion will consider the decision matrix for only one alternative.

3.1 Distance between evaluation vectors

An evaluation vector refers to a column in a decision matrix T^i which records the assessments presented by a certain evaluator. As each evaluation vector reflects the opinion of an evaluator, distance of evaluation vectors could be used to reflect the difference between two evaluators' opinions. Based on this idea, we first measure the distance between evaluation vectors. Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are two evaluation vectors.

For each criterion c_j ($j = 1, 2, \ldots, n$), let u and v be two assessments from different evaluators about c_j. Then the distance between u and v can be calculated. Suppose the distance is calculated through a distance function d_j which is with respect to c_j and $d_j(u, v)$ is a real number in $[0, 1]$, then d_j can be treated as a function from $[0, 1] \times [0, 1]$ to $[0, 1]$. In the following, let $d_1, d_2, \ldots, d_n$ be the selected distance functions with respect to criteria $c_1, c_2, \ldots, c_n$, respectively without other specifications. Here, we do not require that all d_j be of the same form.

Based on the selected d_j, $j = 1, 2, \ldots, n$, the overall distance between two evaluation vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ can be calculated as follows. Let Agg be an aggregation operator [10] defined from $[0, 1]^n$ to $[0, 1]$ and $\beta_1, \beta_2, \ldots, \beta_n$ be the normalized weights of those criteria, i.e.,

$$\beta_j = \frac{wc_j}{\sum_{j=1}^{n} wc_j}, \quad j = 1, 2, \ldots, n. \tag{5}$$

The distance function $\hat{d}$ between the two evaluation vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is a function from $[0, 1]^n$ to $[0, 1]$ such that

$$\hat{d}(\boldsymbol{u}, \boldsymbol{v}) = \mathrm{Agg}\left\{(\beta_j, d_j(u_j, v_j))|j = 1, 2, \ldots, n\right\}. \tag{6}$$

Example 3.1 *Suppose Table 1 gives an example setting. Let* Agg *be the weighted-mean. Then the distance between evaluation vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *is:*

$$\frac{0.12}{1.52} \times 0.75 + \frac{0.47}{1.52} \times 0.11 + \frac{0.93}{1.52} \times 0.57 = 0.44.$$

If Agg *is the OWA operator, then the distance becomes*

$$\frac{0.12}{1.52} \times 0.75 + \frac{0.47}{1.52} \times 0.57 + \frac{0.93}{1.52} \times 0.11 = 0.30.$$

Table 1: An example settings

	d_j	wc_j	$\boldsymbol{u}$	$\boldsymbol{v}$
c_1	$\lvert u - v\rvert$	0.12	0.07	0.82
c_2	$\lvert u - v\rvert^2$	0.47	0.64	0.97
c_3	$\lvert u - v\rvert^3$	0.93	0.93	0.1

Equation (6) indicates the distance between two evaluation vectors that are determined by the selected aggregation operator and the distance functions for each of criteria. It is possible to choose different operators and measures according to the real problem.

3.2 Similarity between two evaluation vectors

Similar to the definition of distance, similarity for individual criterion is first given and then an overall similarity is calculated by a selected aggregation operator.

Let g_j ($j = 1, 2, \ldots, n$) be the selected similarity function with respect to criterion c_j. g_j should satisfy the following two conditions: for any u and v,

$$g_{j_1}(u, v) \geqslant g_{j_2}(u, v) \text{ if } wc_{j_1} \leqslant wc_{j_2}, \quad j_1, j_2 \in \{1, \cdots, n\}; \tag{7}$$

and g_j is a non-increasing function with respect to d_j. Then function g_j is also called a similarity measure with respect to criterion c_j, $j = 1, 2, \ldots, n$. (7) indicates that a more important criterion has less similarity with respect to the same distance between two assessments.

For example, the following functions can be treated as a set of similarity measures.

$$g_j(u, v) = 1 - d_j(u, v)^{\sigma/wc_j}, \quad j = 1, 2, \ldots, n, \tag{8}$$

where $\sigma \in \{wc_j | j = 1, 2, \ldots, n\}$.

In the light of individual similarity measures $\{g_j\}$, an overall similarity $\hat{g}$ between two evaluation vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined, which is a non-increasing function from $[0, 1]^n$ to $[0, 1]$ such that

$$\hat{g}(\boldsymbol{u}, \boldsymbol{v}) = \mathrm{Agg}\{(\beta_j, g_j(u_j, v_j)), j = 1, 2, \ldots, n\}, \tag{9}$$

where Agg is a selected aggregation operators, β_j is the normalized weights of those criteria, $j = 1, 2, \ldots, n$.

In Example 3.1, suppose $g_j(u, v) = 1 - d_j(u, v)$, then the similarities for the weighted-sum aggregation and OWA aggregation are 0.56 and 0.70, respectively. Obviously $\hat{g}(\boldsymbol{u}, \boldsymbol{v}) = 1 - \hat{d}(\boldsymbol{u}, \boldsymbol{v})$. However, this does not hold if we change the relationship between g_j and d_j for any j. This indicates that in real problems we can select the appropriate operators to define the similarity. Moreover, we should notice that $\hat{g}(\boldsymbol{u}, \boldsymbol{v})$ is given through individual distance d_j and individual similarity g_j, $j = 1, 2, \ldots, n$, rather than the overall distance $\hat{d}(\boldsymbol{u}, \boldsymbol{v})$. The reason for separately defining them is to provide more flexibilities to consider the individual influence of criteria.

3.3 Evaluator segmentation

The similarity between two evaluation vectors reflects to what extend two evaluators' opinions are similar, which is used to define the similarity between two evaluators' mental models. Evaluators, thus, can be clustered into several groups based on the similarity of their mental models.

Suppose the similarities between any two pair of evaluators are calculated and recorded in a similarity matrix $M = (s_{jk})_{m \times m}$ where s_{jk} is the similarity between evaluators e_j and e_k. To segment evaluators based on M, we use the fuzzy clustering technique here. Let $\bar{M}$ be the transitive closure of M obtained by

$$\bar{M} = \bigcup_{i=1} M^i, \tag{10}$$

where M^i is the i-th power of M by the max-min composition [11]. Therefore, a clustering result can be obtained from the

ISBN: 978-989-95079-6-8

matrix $\bar{M}$. For example, suppose the matrix $\bar{M}$ is

$$\bar{M} = \begin{pmatrix} 1.0 & 0.6 & 0.8 & 0.6 \\ 0.6 & 1.0 & 0.6 & 0.7 \\ 0.8 & 0.6 & 1.0 & 0.6 \\ 0.6 & 0.7 & 0.6 & 1.0 \end{pmatrix}. \tag{11}$$

Then the clustering result is shown in the following figure.

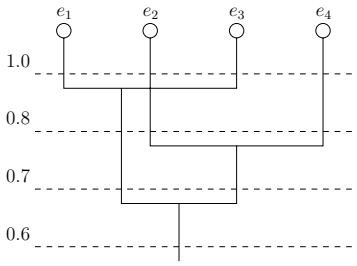


Figure 2: An evaluator clustering results.

From Figure 2, the segmentation results may change with different levels of similarity. For instance, the segmentation result at $\hat{g} = 0.8$ is $\{e_1, e_3\}$, $\{e_2\}$, and $\{e_4\}$; while all evaluators belong to the unique group at $\hat{g} = 0.6$. Hence, a parameter β is used here to indicate at which level the segmentation result is concerned about. In the following let the segmentation result with respect to the given β be $\Gamma(\beta)$.

3.4 Imputation of missing data

To impute missing data, we consider relevant issues and illustrate corresponding solutions here.

First, some missing data need not to be imputed when the associated criteria have infinitesimal effects to the decision G. Therefore, a parameter α is introduced to filter criteria with smaller weights than α. For those criteria, missing data are not imputed.

Next, the evaluator segmentation in Section 3.3 is conducted under the assumption that the decision matrix is a complete one. When the decision matrix is incomplete, we make the following adjustment. Without loss of generality, set

$$T = \begin{pmatrix} A \\ B \end{pmatrix}, \tag{12}$$

where A is a sub-matrix of T which is complete and B is the remainder of T in which each row contains at least one missing data. The evaluator segmentation is obtained based on A, i.e., taking A as the basis to calculate distance and similarity between evaluation vectors. Values in B are used to generate imputations of missing values. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two evaluation vectors and divide them into two parts according to A, B,

$$\boldsymbol{u} = \begin{pmatrix} u_A \\ u_B \end{pmatrix}, \qquad \boldsymbol{v} = \begin{pmatrix} v_A \\ v_B \end{pmatrix}.$$

Suppose $u_j \in u_B$ is a missing value and $v_j \in v_B$ exists. To estimate u_j from v_j, we follow the principle that adding missing value does not change the established similarity between the two evaluation vectors. That means if the established similarity between u_A and v_A is g, then the similarity between $u_{A\cup\{u_j\}}$ and $v_{A\cup\{v_j\}}$ is still g after imputation the missing value u_j. Formally, this can be expressed as the following equation

$$\begin{aligned} &\mathrm{Agg}\Big\{(\beta'_l, g_l(u_l, v_l)), l \in \delta(A)\cup\{j\}\Big\} \\ &\quad = \mathrm{Agg}\Big\{(\beta_l, g_l(u_l, v_l)), l \in \delta(A)\Big\}, \end{aligned} \tag{13}$$

where $\delta(A) = \{l | u_l \in A\}$, β' is the normalized weight after adding the criterion c_j and is calculated by

$$\beta'_l = \frac{wc_l}{\sum_{r\in\delta(A)\cup\{j\}} wc_r} = \frac{wc_l}{wc_j + \sum_{r\in\delta(A)} wc_r}. \tag{14}$$

Notice in (13) only u_j is unknown, therefore, the solution(s) of the equation is/are possible estimation(s) of the missing value u_j in $\boldsymbol{u}$ in terms of $\boldsymbol{v}$.

For example, suppose the Agg is the weighted mean and $g_j(\boldsymbol{u}, \boldsymbol{v}) = 1 - d_j(\boldsymbol{u}, \boldsymbol{v})$, $j = 1, 2, \ldots, n$. Then by (13), we have

$$\begin{aligned} &\mathrm{Agg}\Big\{(\beta'_l, g_l(u_l, v_l)), l \in \delta(A)\cup\{j\}\Big\} \\ &\quad = \mathrm{Agg}\Big\{(\beta_l, g_l(u_l, v_l)), l \in \delta(A)\Big\}, \end{aligned}$$

Hence,

$$\sum_{l\in\delta A\cup\{j\}} \beta'_l g_l(u_l, v_l) = \sum_{l\in\delta(A)} \beta_l g_l(u_l, v_l),$$

$$\sum_{l\in\delta(A)\cup\{j\}} \frac{wc_l \cdot g_l(u_l, v_l)}{wc_j + \sum_{r\in\delta(A)} wc_r} = \sum_{l\in\delta(A)} \frac{wc_l \cdot g_l(u_l, v_l)}{\sum_{r\in\delta(A)} wc_r},$$

$$1 - \sum_{l\in\delta(A)\cup\{j\}} \frac{wc_l \cdot d_l(u_l, v_l)}{wc_j + \sum_{r\in\delta(A)} wc_r} = 1 - \sum_{l\in\delta(A)} \frac{wc_l \cdot d_l(u_l, v_l)}{\sum_{r\in\delta(A)} wc_r}$$

So,

$$d_j(u_j, v_j) = \frac{\sum_{l\in\delta(A)} wc_l \cdot d_l(u_l, v_l)}{\sum_{r\in\delta(A)} wc_r} = \hat{d}(u_A, v_A). \tag{15}$$

Equation (15) indicates that u_j can be obtained simply by the d_j and $\hat{d}$ under the above assumptions. Notice that (15) is obtained from g_j and $\hat{g}$, this equation also indicates that the estimation of u_j is determined by the selection of g_j and $\hat{g}$. Hence given different definitions of them, the relationship in (15) may vary.

3.5 Discussions

To find solution of (15), we notice that the equation may produce more than one estimation of the missing value. The question, then, raises naturally that how to determine a unique one. This paper partly resolves this problem through calculating the change $\tilde{d}$ of overall distance induced by an obtained estimation. The estimation that produces the least $\tilde{d}$ will be taken as the imputation of missing data. The overall distance change $\tilde{d}$ is defined as

$$\tilde{d}(u_j) = \sum_{v\in\Gamma(u)\setminus\{u\}} d_j(u_j, v_j), \tag{16}$$

where $\Gamma(u)$ is the cluster of evaluation vectors in which u belongs to. By (16), the best estimation is the one which produces the least distance sum between itself to other evaluation

ISBN: 978-989-95079-6-8

vectors belonging to the same cluster. Examples in the following section have shown that this method can reduce some estimations although they cannot grantee the uniqueness of estimation. Hence, more work is being done to provide other strategy to resolve this problem.

Another question raising from the above imputation procedure is the obtained evaluator segmentation may not be correct because the segmentation is conducted based on A only. Accounting for this problem, we extend the above procedure as follows. For any $\boldsymbol{u}$ and $\boldsymbol{v}$, the distance $\hat{d}(\boldsymbol{u}, \boldsymbol{v})$ and the similarity $\hat{g}(\boldsymbol{u}, \boldsymbol{v})$ are calculated in terms of criteria on which both $\boldsymbol{u}$ and $\boldsymbol{v}$ are not missed rather than limited to criteria corresponding to A. That means the obtained distance and similarity between $\boldsymbol{u}$ and $\boldsymbol{v}$ are the most possible values from the current incomplete decision table. An experiment of the extended process is given in the next section.

4 Experiments and observations

The presented α-β imputation method of missing data is programmed in the Java^{TM} programming language running on a Linux system.

4.1 Experiment 1

Experiment 1 is conducted on a complete decision table with 16 criteria and 5 evaluators. Missing data are randomly selected from the decision table.

To impute the missing data, we set the distance measure for each criteria as $d_j(u_j, v_j) = |u_j - v_j|$ $(j = 1, 2, \ldots, 16)$ and the similarity measure for each criterion as $g_j(u_j, v_j) = 1 - d_j(u_j, v_j)$, the aggregation operators used in calculating overall distance and similarity are the same, say the weighted mean; and set the two parameters $\alpha = 0.2$ and $\beta = 0.6$.

Table 2: The original evaluation table in which missing data are marked by bold font.

wc_j	e_1	e_2	e_3	e_4	e_5
0.262	0.877	0.035	0.88	0.772	0.941
0.334	0.977	0.458	0.903	0.898	0.227
0.823	0.845	0.065	0.894	0.122	0.555
0.216	0.532	0.630	**0.676**	0.353	0.858
0.803	0.862	0.073	0.183	0.618	0.784
0.227	0.797	0.046	**0.721**	0.551	0.950
0.794	0.159	0.995	0.621	0.808	0.307
0.811	0.227	0.936	0.994	0.973	0.134
0.945	0.304	0.901	0.636	0.225	**0.927**
0.892	0.967	0.717	0.851	0.835	**0.035**
0.344	0.641	**0.336**	0.666	0.281	0.598
0.016	**0.475**	0.915	0.743	0.876	0.651
0.235	0.026	0.999	0.501	0.431	0.586
0.905	0.307	0.371	0.757	0.832	0.66
0.316	0.541	0.003	0.438	0.113	0.275
0.068	0.042	0.27	0.929	0.675	0.206

By the given distance and similarity measure, we get the distance matrix $\mathcal{D}$ and the similarity matrix $\mathcal{M}$ as follows:

$$\mathcal{D} = \begin{pmatrix} 0 & 0.571 & 0.35 & 0.392 & 0.233 \\ 0.571 & 0 & 0.348 & 0.332 & 0.55 \\ 0.35 & 0.348 & 0 & 0.255 & 0.382 \\ 0.392 & 0.332 & 0.255 & 0 & 0.394 \\ 0.233 & 0.55 & 0.382 & 0.394 & 0 \end{pmatrix} \tag{17}$$

and

$$\mathcal{M} = \begin{pmatrix} 1 & 0.429 & 0.65 & 0.608 & 0.767 \\ 0.429 & 1 & 0.652 & 0.668 & 0.45 \\ 0.65 & 0.652 & 1 & 0.745 & 0.618 \\ 0.608 & 0.668 & 0.745 & 1 & 0.606 \\ 0.767 & 0.45 & 0.618 & 0.606 & 1 \end{pmatrix}. \tag{18}$$

By the fuzzy clustering method, we get the following clustering result.

$$\bar{\mathcal{M}} = \begin{pmatrix} 1 & 0.65 & 0.65 & 0.65 & 0.767 \\ 0.65 & 1 & 0.668 & 0.668 & 0.65 \\ 0.65 & 0.668 & 1 & 0.745 & 0.65 \\ 0.65 & 0.668 & 0.745 & 1 & 0.65 \\ 0.767 & 0.65 & 0.65 & 0.65 & 1 \end{pmatrix} \tag{19}$$

i.e., we have only one class $\{e_1, e_2, e_3, e_4, e_5\}$. By the imputation method in Section 3.4, the following result is obtained (shown in Table 3).

Table 3: Decision table after imputation of missing data (bold number for estimations and # for still missing data).

wc_j	e_1	e_2	e_3	e_4	e_5
0.262	0.877	0.035	0.88	0.772	0.941
0.334	0.977	0.458	0.903	0.898	0.227
0.823	0.845	0.065	0.894	0.122	0.555
0.216	0.532	0.63	**0.608**	0.353	0.858
0.803	0.862	0.073	0.183	0.618	0.784
0.227	0.797	0.046	**0.296**	0.551	0.95
0.794	0.159	0.995	0.621	0.808	0.307
0.811	0.227	0.936	0.994	0.973	0.134
0.945	0.304	0.901	0.636	0.225	**0.071**
0.892	0.967	0.717	0.851	0.835	**0.734**
0.344	0.641	#	0.666	0.281	0.598
0.016	**0.418**	0.915	0.743	0.876	0.651
0.235	0.026	0.999	0.501	0.431	0.586
0.905	0.307	0.371	0.757	0.832	0.66
0.316	0.541	0.003	0.438	0.113	0.275
0.068	0.042	0.27	0.929	0.675	0.206

4.2 Experiment 2

In the presented imputation method of missing data, aggregation operators are used for calculation of overall distance and overall similarity of two evaluation vectors. Different aggregation operators may lead to different clustering results and, in turn, may result in different imputations. Experiment 2 compares three widely-used aggregation operators, i.e., the weighted mean (Agg_w); the arithmetic mean (Agg_a)

$$\mathrm{Agg}_a(x_1, x_2, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i; \tag{20}$$

and the OWA operator (Agg_o)

$$\mathrm{Agg}_o(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} w_i x_{\sigma(j)}, \tag{21}$$

where σ is a permutation such that $x_{\sigma(1)} \geqslant x_{\sigma(2)} \geqslant \cdots \geqslant x_{\sigma n}$. To conduct this experiment, we change the distance measure and the similarity measure for each individual criterion,

ISBN: 978-989-95079-6-8

i.e.,

$$d_j(u_j, v_j) = |u_j - v_j|, \quad g_j(u_j, v_j) = 1 - d_j(u_j, v_j).$$

The original evaluation table is still the one shown in Table 2. The comparison results are shown in Table 4.

Table 4: Comparison results.

missing data		original value	imputations		
row	col		Agg_w	Agg_o	Agg_a
3	2	0.676	0.608	0.527	0.581
5	2	0.721	0.296	0.756	0.323
8	4	0.927	0.071	/	0.557
9	4	0.035	0.734	/	0.714
10	1	0.336	/	/	/
11	0	0.475	0.418	/	0.398

From the comparison, we can observe that different aggregation operators may produce different imputations. Even more, some aggregation operators may not produce estimation of missing data. Moreover, the bias between an imputation and an original value is bigger for some criteria. The independence nature of the data and the unsupervised nature of the presented method are two possible reasons for these results.

5 Conclusions and Future Works

Imputation of missing data is a vital step in decision making and evaluation problems. Due to various reasons, decision tables are often including missing data. Therefore, how to efficient impute the missed data will affect correct decision making and evaluation. However, few work is reported on imputation of missing data for decision tables. In this paper, an α-β imputation method of missing data (IMD) in a decision table is proposed. The IMD method estimates missing data based on the similarity between evaluators' mental models. First, the IMD method uses an overall distance to measure the similarity of two evaluators mental models through their evaluation vectors. Then a fuzzy clustering method is used to segment evaluators based on the similarity of their mental models. Finally, missing data are imputed by the clustering results under the assumption that an imputed data should not change the clustering results. Two numerical experiments are conducted to test the proposed IMD method. Experiments indicate the IMD method can efficiently estimate some missing data. Moreover, those experiments also show that selecting different aggregation operators, distance measure, and similarity measure may change the imputed values. However, from these experiments, we notice that some estimations of missing data have bigger bias from the originals due to the independence nature of the data and lack of backgrounds as well as other possible reasons. Hence, more work still need to be done such as how to improve the accuracy of an estimation of missing data and how to combine relevant background to the imputation. Now, some related work is undertaking.

Currently, we notice that another missing value processing problem in fuzzy preference relations for group decision-making problem has been discussed in [2, 3]. Method based on additive consistency has been discussed [7]. Although the missing value in that method is different from the one discussed in the presented work, ideas in those work can be introduced into the presented work to improve the efficiency and accuracy. Some work is conducting.

Acknowledgment

The work presented in this paper was supported by Australian Research Council (ARC) under Discovery Projects DP0559213 and DP0880739.

The authors sincerely appreciate the reviewers comments and suggestions to improve the presented work.

References

[1] Points to consider on missing data, 2001. Committee for Proprietary Medicinal Products (CPMP),The European Agency for the Evaluation of Medicinal Products.

[2] S. Alonso, F. J. Cabrerizo, F. Chiclana, F. Herrera, and E. Herrera-Viedma. Group decision making with incomplete fuzzy linguistic preference relations. *International Journal of Intelligent Systems*, 24(2):201–222, 2009.

[3] S. Alonso, F. Herrera-Viedma, F. Chiclana, and F. Herrera. Individual and social strategies to deal with ignorance situations in multi-person decision making. *International Journal of Information Technology and Decision Making*, 2009. in press.

[4] Q. Chen, J. G. Ibrahim, M.-H. Chen, and P. Senchaudhuri. Theory and inference for regression models with missing responses and covariates. *Journal of Multivariate Analysis*, 99:1302–1331, 2008.

[5] A. Farhangfar, L. Kurgan, and J. Dy. Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition*, 41:3692–3705, 2008.

[6] O. Harel. Inferences on missing information under multiple imputation and two-stage multiple imputation. *Statistical Methodology*, 4(1):75–89, January 2007.

[7] E. Herrera-Viedma, F. Chiclana, F. Herrera, and S. Alonso. Group decision-making model with incomplete fuzzy preference relations based on additive consistency. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 37(1):176–189, 2007.

[8] J. Liesiö, P. Mild, and A. Salo. Robust portfolio modeling with incomplete cost information and project interdependencies. *European Journal of Operational Research*, 190:679–695, 2008.

[9] J. Lu, G. Zhang, D. Ruan, and F. Wu. *Multi-Objective Group Decision Making – Methods, Software and Applications with Fuzzy Set Technology*. Imperial College Press, London, 2007.

[10] R. Mesiar, A. Kolesárová, T. Calvo, and M. Komorníková. A review of aggregation functions. In H. Bustince, F. Herrera, and J. Montero, editors, *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, volume 220 of *Studies in Fuzziness and Soft Computing*, pages 121–144. Springer Berlin/Heidelberg, 2008.

[11] G. Munda. A conflict analysis approach for illuminating distributional issues in sustainability policy. *European Journal of Operational Research*, 194(1):307–322, 2009.

[12] J. Siddique and T. R. Belin. Using an approximate bayesian bootstrap to multiply impute nonignorable missing data. *Computational Statistics & Data Analysis*, 53:405–415, 2008.

[13] E. Triantaphyllou. *Multi-Criteria Decision Making Methods: A Comparative Study*. Kluwer Academic Publishers, Dordrecht/Boston/London, 2000.

ISBN: 978-989-95079-6-8

Application of Zadeh's Impossibility Principle to Approximate Explanation

Vesa A. Niskanen

Dept. of Economics & Management, University of Helsinki, PO Box 28, 00014 Helsinki, Finland
Email : vesa.a.niskanen@helsinki.fi

Abstract: *We consider application of Zadeh's impossibility principle and extended logic FLe to approximate scientific explanation from the standpoint of the philosophy of science. First, traditional explanation models are sketched. Second, Zadeh's principle is considered in the broader context which includes reasoning and theory formation. Third, application ideas for using the impossibility principle to approximate explanation are provided. We focus on the role of approximate reasoning, approximate hypothesis assessment and approximate probability in explanation. It seems that approximate explanation can be usable when imprecise entities are involved in the conduct of inquiry.*

Keywords: *approximate explanation, approximate reasoning, impossibility principle.*

1 Introduction

Scientific explanations make our objects of research intelligible for us [28]. In practice this means that we aim to find the causes, functions or purposes of the phenomena under study. An explanation constitutes of two parts, the phenomenon or problem to be explained (explanandum) and our explanation for it (explanans). Explanations play an important role in theory formation and hypothesis assessment [12,17,18,28,29].

The study on the causes of the phenomena is a central issue in explanation. If we study the causes, Aristotle already suggested four types of causes, viz. material, formal, effective and final causes. Today, only two of these are usually applied in the conduct of inquiry. The natural sciences focus on the effective causes, whereas in the human sciences final causes are also considered [18].

Today, the mainstream seems to aim at precise explanation models, but thanks for the fuzzy systems and Zadeh's impossibility principle, it seems that we can also construct corresponding approximate models in the future.

Below we consider the prospects for applying the impossibility principle and other Zadeh's recent ideas to scientific explanation from the standpoint of the philosophy of science. Due to our philosophical approach, we mainly provide some general ideas for applying these methods. Lack of space puts also restrictions on our examination. Section 2 briefly presents the prevailing explanation models. In section 3 we integrate the impossibility principle with reasoning, theory formation, hypothesis testing and explanation. Section 4 considers aspects of approximate explanation, and section 5 concludes our examination.

2 Traditional explanation models

The pragmatic approach to explanation, which seems widely-used today, considers such phenomena as facts, regularities and occurrences. Our task is to provide the cause, origin, history, function or goal of a given phenomenon. In practice we provide such questions as "why?", "how?", "what is the purpose for?", "what is the reason for?" and "what is the function of?" [18,27,29].

The positivistic tradition, in turn, has emphasized syntactical and semantical aspects of explanation in which case concept analysis and logical structures are essential [18,27].

Another categorization of explanation stems from the causes. If we attempt to find effective causes of a phenomenon, we consider which agents, factors or events have initiated the phenomena under study. The well-known example of this idea is the subsumption model in the natural sciences [12,13,17].

The final causes, in turn, are those goals, purposes, intentions or motives of agents, in particular those of the human beings, which guide their behavior. Aristotle's and von Wright's practical syllogism models [29] apply this idea. The subsumption approach mainly provides why?- and how?-questions, and the latter tradition usually asks what is the purpose for or what is the reason for something.

We usually search for effective causes in the inanimated world, but in the animated world the problem arises whether we should also search for final causes. From the philosophical standpoint, the former approach mainly stems from the positivistic tradition and Marxism and the latter, which studies both of these causes, is known as the Geisteswissenschaft approach (hermeneutics, phenomenology etc.). In [29] these competitive approaches are referred to as the Galilean and Aristotelian traditions, respectively, and we use these terms below.

The explanation of the behavior of the human being has been a long-lasting controversy between these two traditions. The Galilean tradition assumes that our behavior is based on our stimulus-response functions and such feedback systems which are analogous to homeostats. In this framework we can even deal with *goal-directed* behavior of beings.

The Aristotelian tradition presupposes that the human beings are *goal-oriented* agents and thus we must also study their aims, intentions, motives and other underlying causes of behavior. Our goals can be both conscious and subconscious, and intentionality is usually related to the former category. A classical example within this tradition is that we should explain (inanimate) nature but understand history, i.e., historical research also takes into account the intentional factors of beings [2,7,15,26].

If we consider the nature of causes on the time axis, the distinction between the Galilean and Aristotelian explanations essentially means that in the former case the cause precedes and in the latter case it follows the effect. Another way to state this is that we use causal and teleological explanations, respectively.

ISBN: 978-989-95079-6-8

For example, if we state in the first day of the IFSA '09 Congress that Lotfi Zadeh is present because he was requested to deliver the keynote speech, we use the Galilean explanation. If, in turn, we state that Lotfi Zadeh is present in order to deliver the keynote speech, the Aristotelian approach is adopted. Naturally, we can often use these explanations simultaneously.

Hence, in this sense we can establish that

1. Requested for speech (causal explanans) → **Zadeh is present in the IFSA '09 (explanandum).**
2. **Zadeh is present in the IFSA '09 (explanandum)** ← In order to deliver speech (teleological explanans):

Other models for explanation are also available. For example, the genetic explanations aim to discover causes by examining the history or chronological order of events of a phenomenon. In the historical research and forensic sciences, inter alia, abduction is usually applied in this context.

The statistical and probabilistic explanations, in turn, are based on statistical results and probabilistic models, respectively. An example of the former type would be the reasoning that most of the participants in the IFSA '07 Congress were males because about 90 % of the researchers in the fuzzy community are males. In the latter case we could provide an explanation that if we are tossing a coin, about 50 % of the outcomes are heads because the frequency probability for this event is .5. In particular in the human sciences these explanation models are relevant because the human behavior is usually indeterministic by nature.

The traditional explanation models aim to operate with precise entities and traditional deduction, and thus certain problems can arise. First, they usually apply bivalent logic and this approach can yield overly rough models. Second, if mathematical expressions are used, they can be oversimplified or too idealized for modeling the corresponding real-world phenomena. Finally, this approach lacks an appropriate uniform language which can be used in the context of imprecise entities. Fuzzy systems, in turn, can provide some resolutions to these problems.

3 Impossibility principle and FLe

Recently, Zadeh has established the principles of the extended fuzzy logic, *FLe* (a.k.a. FL+), which is a combination of "traditional" provable and precisiated fuzzy logic, *FLp*, as well as a novel meta-level unprecisiated fuzzy logic, *Flu* [31,38]. He maintains that in the *FLp* the objects of discourse and analysis can be imprecise, uncertain, unreliable, incomplete or partially true, whereas the results of reasoning, deduction and computation are expected to be provably valid. In the *Flu*, in turn, membership functions and generalized constraints are not specified, and they are a matter of perception rather than measurement.

In our framework below this means that we can apply both traditional bivalent-based and novel approximate validity, definitions, axioms, theories and explanations, inter alia. In the former case we thus operate with precise theorems, classical deducibility, syllogisms and formal logic, whereas the latter is related to "f-entities" within informal and approximate reasoning and incomplete information (Fig. 1).

For example [31, p. 2],

"A simple example of a f-theorem in f-geometry is: f-medians of f-triangle are f-concurrent. This f-theorem can be f-proved by fuzzification of the familiar proof of the crisp version of the theorem."

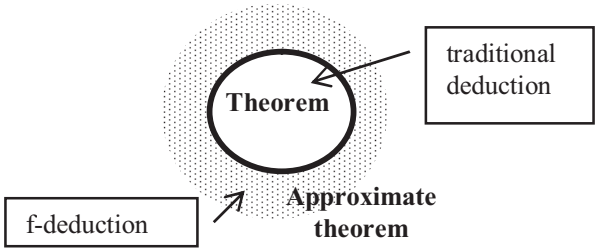


Figure 1. Precise and approximate theorem by applying the *FLe*.

F-validity and f-theoremhood in the *Flu* are examples of Zadeh's Impossibility Principle [31]. This principle informally states that in an environment of imprecision, uncertainty, incompleteness of information, conflicting goals and partiality of truth, p-validity is not, in general, an achievable objective.

The fuzzy modus ponens is the well-known application of this principle in semantic validity, i.e., we can replace the traditional syllogism [8,11,14,19]

A
If A, then B
Thus, B

with its approximate counterpart

A'
If A, then B
Thus B'

in which A≈A' and B≈B'. In the latter case we can thus draw approximate conclusions, i.e., if A' resembles the antecedent A, the conclusion B' resembles the consequent B. In other words, the approximate conclusion is in the proximity to its true counterpart. F-validity is the corner stone in the *FLe* and it is also essential in approximate theory formation and explanation.

In [21,22,23], Zadeh's *FLe* is considered from the standpoint of reasoning and theory formation and it was assumed that the approximation stems essentially from the former. We assume below that if our statements are true, it means that they are identical with their true counterparts. The non-true statements, in turn, are more or less distinct from their true counterparts, or in other words, their distances between their true counterparts are non-zero.

For example, given that John is young (true counterpart), we can reason that the statement

"John is young"

is true. On the other hand, "John is fairly young" is non-true because the predicates "is young" and "is fairly young" are dissimilar.

In fact, we can thus operate with a f-similarity relation of the *Flu,* and then the truth values are specified according

ISBN: 978-989-95079-6-8

to this operation. The higher this similarity, the more true the statement in consideration [23].

Then it seems plausible that, for example, the truth value of the statement

"John is fairly young", provided that he is young

is fairly true [23].

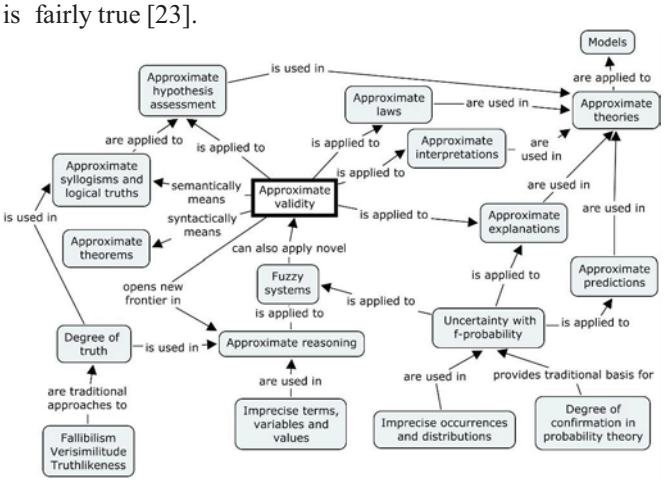


Figure 2. The role of Zadeh's *FLe* in the conduct of inquiry.

We thus assume that approximate theorems are in the neighborhood of their true counterparts. We can also assume that in theory formation approximate laws and theories are in the proximity to their true counterparts and approximate explanations are close to the corresponding true explanations. Figure 2 depicts these ideas at a general level when we establish that our examination on the *FLe* is mainly based approximate validity.

In the case of theory formation, for example, the constituent or "law" of one physical theory states that the boiling point of water is 100 ºC at sea level, but we can also apply it as an approximate law in the low altitudes in general, i.e., we can state that the boiling point of water is *approximately 100 ºC close to* sea level [19,22].

In addition, within the traditional approach to the theory formation and model construction we only use their approximations and idealizations on some occasions. In the former case the degree of truth of a theory or model often increases when we replace the approximate theory, law or model with its less approximate counterpart [18,19]. Hence, in this respect the law

the boiling point of water is 100 ºC at sea level

is a more true than

the boiling point of water is 100 ºC.

Idealizations, in turn, are typical when mathematical or statistical models are used. For example, in statistics such ideal curves as the Gaussian curve are used when the empirical data sets, the distributions of which are only more or less similar to their mathematical counterparts, are examined. Thus, we actually draw statistical conclusions on approximate grounds (c.f. also section 4).

Popper has already considered the foregoing aspects in his theory of verisimilitude from the bivalent standpoint [24,25]. He presupposes that we aim to formulate true theories (cognitivism) but in practice our theories can be non-true. If we perform successful research, our theories will approach their true counterparts and thus they are always corrigible by nature. The similar idea is applied in Peirce's fallibilism [17,18], and in general, in scientific realism (Peirce, Lenin, Popper, Hempel). Niiniluoto [20] has considered approximate theories and explanations within his theory of truthlikeness, but he has also applied bivalent logic in this context. In fact, if we replace bivalent logic with *FLe*, our approach is fairly close to his ideas [23].

4 Approximate explanation

As was stated above, our examination is based on the idea of approximate reasoning in the *FLe*. Thus, in an approximate explanation model the explanans is in the neighborhood of its true counterpart. In practice we now assess the degree of similarity between our explanans and its true counterpart. The higher the degree of similarity, the higher the degree of truth for our explanans (Fig. 3).

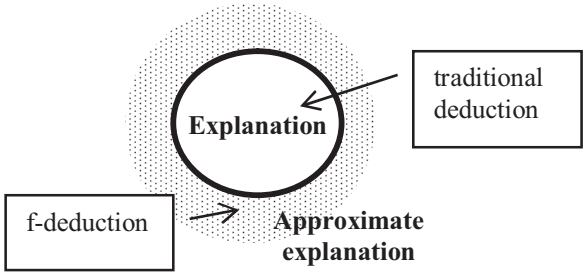


Figure 3. Precise and approximate explanation by applying the *FLe*.

Hence, we have to consider again which logical tools are appropriate to us. As above, within simple statements we assess the similarity between the predicates, i.e.,

truth(John is fairly young), provided that he is young,

is based on the similarity between "young" and "fairly young".

The more complex statements can also be examined within the *FLe* because we can apply the fuzzy similarity measure to compound and quantified statements [31-38]. Then, for example, we should be able to assess such truth values as

truth(John is fairly young and very decent),

provided that he is young and decent.

Another example is the assessment of such quantified expressions as

truth(all Swedes are fairly tall),

provided that most Swedes are tall.

In the former case some concrete resolutions are already available in the *FLe*, whereas fuzzy quantification still awaits usable truth evaluation results. Approximate quantification is problematic because then we operate at two levels, with approximate quantifiers and (compound) statements. Application of the extension principle and Zadeh's idea on

ISBN: 978-989-95079-6-8

f-transformation [38] can be a possible resolution for these, and one method is suggested in [23]. Naturally, these approaches are context-dependent by nature. Hence, more concrete results are expected for assessing the truth values and validity within the *FLe*.

In the probability and statistical explanation models mentioned above we can apply corresponding approximate theories. For example, if in the former case we are tossing the coin in the real world, only approximately 50 % will be heads. Hence, if we ask why about 50 % of these outcomes are heads, we can provide an approximate explanation that it is due to this approximate frequency probability. In other words, our approximate probability statements are close to their true counterparts.

The same goes with the statistical explanations. If we explain why most of the attendants in the IFSA '07 Congress were males, we rely on the approximate statistical result that about 90 % of persons in the fuzzy community are males.

We have to bear in mind that both probabilistic and statistical explanation models above are distinct from the forecasting models. In the former case we know the conclusion and we derive it from the relevant premises, whereas in forecasting the conclusion is unknown for us. Naturally, the systematic power of a theory presupposes both of these features.

In the foregoing models fuzzy probability approach usually means approximate probability variables and values of these variables, i.e., statements of the type

the probability that John is young is very high,

and this area has already been studied quite much in the literature [4,5,6,9,31,37]. In Zadeh's *FLe* we can also apply such approximate probability distributions as approximate normal distributions by using the f-transformation. In this case even such precise variable values as

probability(John's age is 20 years)

yield approximate probabilities. With imprecise values (e.g., John is fairly old) we obtain even more imprecise outputs (Fig. 4).

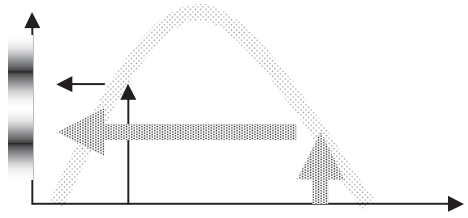


Figure 4. Approximate probability distributions yield approximate outputs even from precise inputs.

The foregoing examples of probability are mainly based on the physicalistic cases, i.e., on the facts and data of the real world. However, Zadeh's theories seem also applicable to epistemic probability in which case these outcomes base on our inference and beliefs [1,30]. In practice we consider the relationship between the hypotheses and their evidence, and we thus assess the degrees of truth, confirmation or belief of hypotheses according to the available relevant evidence.

The gradation of truth was already studied above. As regards the hypothesis testing in general, due to the lack of space we only focus on the degree of confirmation in the context of the explanatory hypotheses and the hypothetico-deductive method [12,13,17,18].

In this hypothesis assessment we usually apply three methods. First, Mill's method of difference [17,18], i.e.,

> *"If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon."*

Second, the disjunctive syllogism

A or B
not A
thus, B.

Third, the modus tollens syllogism

if A, then B
not B
thus, not A.

Mill's principle and the disjunctive syllogism provide us with the general roadmap for excluding the false hypotheses and within the hypothetico-deductive method this task is carried out in practice by applying the modus tollens syllogism.

Hence, given the hypothesis A and its observable or testable logical consequence B, the justifiability of B is assessed according to our knowledge, experiments and observations. If these facts are inconsistent with B, the modus tollens leads us to the conclusion that our hypothesis is false. According to the disjunctive syllogism, in turn, this hypothesis can be rejected or excluded.

If, on the other hand, our experiments and observations correspond with B, the traditional modus tollens will not provide us with any resolution. Hence, in the latter case we have to replace deduction with induction, and we can use such expressions as "our hypothesis is confirmed" in this context [17,18,24,25]. Sufficient degree of confirmation, in turn, will lead to the truth or acceptance of the hypothesis (c.f. below).

In the traditional conduct of inquiry this method thus uses the hypotheses, which stem from the researcher's context of discovery and inventions, deduces tests and experiments from these hypotheses and finally either rejects or confirms the hypotheses according to the known facts and empirical evidence. However, this approach does not provide any resolution to the problem how we can invent or discover novel hypotheses.

Statistical tests also apply the disjunctive method and physicalistic probability when we consider the acceptance of the null and alternative hypotheses according to the tests of significance. We accept the null hypothesis if the value of our test variable does not deviate too much from the "usual" case. Hence, in fact we consider the so-called type 1 error, i.e., we attempt to find justified reasons to reject the null hypotheses. In practice this error is estimated with the p-

ISBN: 978-989-95079-6-8

value in which case we consider the rejection of the null hypotheses if the p-value is sufficiently small (e.g., $p < .05$), i.e., the p-value is our risk to make an erroneous decision if we reject the null hypothesis.

In practice, we have three alternatives in this case [10,16], viz.

1. the null hypothesis is true but our data set is exceptional
2. the null hypothesis is false and our data set is typical
3. the null hypothesis is false and our data set is exceptional

In the conventional case the statistical hypotheses are mutually exclusive, and thus the rejection of one means the acceptance of the other. Hence, according to the disjunctive method, our reasoning boils down to the metarules

1. if the p-value is not small, accept the null hypothesis
2. if the p-value is small, reject the null hypothesis (and accept the alternative hypothesis)

The statistical decision making is thus in this respect based on approximate reasoning and probability. Zadeh's *FLe* can make this reasoning more formal and informative if we operated with the degrees of acceptance and rejection in this context. Then, we could apply such metarule as

> the smaller the p-value, the lower the risk to reject the null hypothesis (the higher the degree of rejection for the null hypothesis)

With the *FLe* we could also take into account better the borderline of these two, viz. the area in which we hesitate over these alternatives. This would mean the fuzzy rules which we have already stated above, i.e.,

1. If the p-value is not small, then accept the null hypothesis.
2. If the p-value is small, then reject the null hypothesis.

With these rules the corresponding fuzzy system could also operate fluently with the borderline cases of p. However, more concrete tools for considering this problem are still expected within the *FLe*.

As regards the modus tollens syllogism in hypothesis assessment, we already considered its bivalent version above. If we apply its approximate version, we can also use linguistic and approximate constituents. Thus, given that $A \approx A'$ and $B \approx B'$, these syllogisms are such as [23]

> if A, then B
> B'
> thus, A'

In this case the truth values of the premises may also be between true and false. This syllogism is analogous to the bivalent case when A' and B' are the antonyms of A and B, respectively, because given that the implication and B' are true, B is false, and thus A must be false (and A' is thus true).

Within the *FLe* the other extreme could be that B' is false in which case B is true and A is thus anything from false to true (and A' is also anything from false to true). In practice this could mean that our hypothesis, A, may be true at least to some extent.

For example,

> If John is young, then he is still a schoolboy.
> John is retired.
> Thus, John is old and reject the hypothesis "John is young".

> If John is young, then he is still a schoolboy.
> John is still a schoolboy.
> Thus, the degree of truth for John being young can be non-zero.

It seems that the approximate version of the modus tollens is more versatile because it is also usable when the second premise is non-false. Then we can at least approximately apply the metarule that the more convincing the evidence for the hypothesis, the higher the degree of truth of our hypothesis. In addition, the more various experiments and observations support our hypotheses, the more true hypothesis is obtained. We can also use the concepts "degree of confirmation" or "degree of acceptance / rejection" in this context if necessary.

Within the *FLe* we can also assume that the implication in the approximate modus tollens syllogism is non-true. For example, if this implication is only fairly true, we can establish that even the false B does not necessarily lead to mere false hypotheses. Equally the truth value of B close to true may already lead to conclusion that the truth value of our hypothesis is anything between false and true. In general, we may thus assume that the non-true implications cause more "dispersion", granulation or imprecision to our conclusions, and loose reasoning links of this type are typical in the human sciences in which we usually operate with noisy data and the complex interrelationships between the variables.

We may also apply the foregoing hypothesis testing to interpretation, which is a widely-used technique in the qualitative research. In this context we attempt to provide a "true" interpretation according to our preliminary interpretation hypothesis (*foreknowledge* in hermeneutics). This procedure of assessing hypotheses is often more subjective by nature than in the quantitative research. However, the *FLe* also seems appropriate to qualitative hypotheses assessment because this inquiry is usually linguistic and approximate by nature [2,3,23].

Zadeh has also considered the idea of the second-order probability under the *FLe* [31,37]. In this context we examine such statements as

> The probability that "the probability of John being very young is fairly high" is high.

This subject matter would also extend a new frontier within both the fuzzy systems and the probability theory, but it also awaits much further studies.

Approximate scientific explanation is related to reasoning, theory formation and hypothesis testing. These, in turn, base much on approximate and probabilistic reasoning. Hence, this object of research is very wide. Zadeh's *FLe* seems to provide promising tools for resolving many of the foregoing problems, but much further studies are still expected in this area.

ISBN: 978-989-95079-6-8

5 Conclusions

We have provided some guidelines how Zadeh's impossibility principle and the extended logic *FLe* could be applied to scientific explanation. We have also examined reasoning, hypothesis assessment and theory formation because these are closely related to explanation.

First traditional reasoning, theory formation and explanation, which use precise entities and bivalent logic in particular in the quantitative research, were considered.

Second, we dealt with Zadeh's *FLe* and his other recent theories, and we examined how we could design usable approximate theories and explanations which essentially base on approximate reasoning and probability.

In particular, we provided examples on our approximation approach to scientific laws, theory formation, statistical and probabilistic reasoning, hypothesis assessment and two well-known syllogisms. These ideas, in turn, were relevant when approximate explanation was considered.

It seems that the foregoing methods are usable in approximate explanation and in this context we could apply the general principle that approximate explanations are in the neighborhood of their true counterparts. This aim is essentially achieved by using approximate reasoning.

Zadeh's novel ideas seem promising in particular in the qualitative research because then we operate with imperfect information, and in this context we still lack a uniform and usable logico-linguistic system. It also seems possible to apply his ideas to the quantitative research, in particular if we attempt to mimic human reasoning in a computer environment.

We must bear in mind that several of the foregoing issues are still controversial even in the philosophy of science, and the same problem will likely concern Zadeh's and Author's ideas above.

However, due to the lack of space, this paper only provides at metalevel some general guidelines for approximate scientific explanation and related topics from the standpoint of philosophy of science, and hence this relevant subject matter still awaits much further studies in the future.

Acknowledgements

The author expresses his deep gratitude to Prof. Lotfi Zadeh at the UC Berkeley for the original ideas, articles, several discussions and encouragement which made it possible to write this article. I also express my thanks to the anonymous referees for their valuable comments.

References

[1] Carnap, R. (1962): The Logical Foundations of Probability. The University of Chicago Press, Chicago.

[2] Denzin, N. & Lincoln, Y. (1994): Handbook of Qualitative Research. Sage, Thousand Oaks.

[3] Dimitrov, V. & Hodge, B. (2002): Social Fuzziology – Study of Fuzziness of Social Complexity. Physica Verlag, Heidelberg.

[4] Dubois, D. & al. (1991): Fuzzy Sets in Approximate Reasoning, Part 2: Logical Approaches. Fuzzy Sets and Systems 40, 203 - 244.

[5] Dubois, D. & Prade, H. (1980): Fuzzy Sets and Systems, Theory and Applications. Academic Press, New York.

[6] Dubois, D. & Prade, H. (1991): Fuzzy Sets in Approximate Reasoning, Part 1: Inference with Possibility Distributions. Fuzzy Sets and Systems 40, 143-202.

[7] Fetzer, J. (1993): Philosophy of Science. Paragon House, New York, N. Y.

[8] Gadamer, H.-G. (1975): Truth and Method. Sheed & Ward, London.

[9] Grzegorzewski, P. & al. (Eds.) (2002): Soft Methods in Probability, Statistics and Data Analysis. Physica Verlag, Heidelberg.

[10] Guilford, J. & Fruchter, B. (1978): Fundamental Statistics in Psychology and Education. McGraw-Hill, London.

[11] Haack, S. (1978): Philosophy of Logics. Cambridge University Press, Cambridge.

[12] Hempel, C. (1965): Aspects of Scientific Explanation and Other Essays in the Philosophy of Science. The Free Press, New York.

[13] Hempel, C. (1966): Philosophy of Natural Science. Prentice Hall, Englewood Cliffs.

[14] Hintikka, J. & Suppes, P. (Eds.) (1970): Information and Inference. Reidel, Dordrecht.

[15] Kaplan, A. (1973): The Conduct of Inquiry: Methodology for Behavioral Sciences. Chandler, New York, N. Y.

[16] Kakkuri-Knuuttila, M.-L. (1998): Argumentti ja kritiikki (Argument and Criticism). Tammer-Paino, Tampere.

[17] Nagel, E. (1961): The Structure of Science. Routledge & Kegan Paul, London.

[18] Niiniluoto, I. (1983): Tieteellinen päättely ja selittäminen (Scientific Reasoning and Explanation). Otava, Keuruu.

[19] Niiniluoto, I. (1984): Johdatus tieteenfilosofiaan (Introduction to the Philosophy of Science). Otava, Keuruu.

[20] Niiniluoto, I. (1987): Truthlikeness. Reidel, Dordrecht.

[21] Niskanen, V. A. (2009): Fuzzy Systems and Scientific Method – Meta-Level Reflections and Prospects. In Seising, R. (Ed.): Fuzzy Set Theory - Philosophy, Logics, and Criticism. Springer.

[22] Niskanen, V. A. Meta-Level Prospects for Scientific Theory Formation by Adopting the Information Granulation Approach. Proceedings of the IMS '08 Conference, Sakarya University, pp. 33-41.

[23] Niskanen, V. A. (2004): Soft Computing Methods in Human Sciences. Studies in Fuzziness and Soft Computing, Vol. 134. Springer Verlag, Berlin.

[24] Popper, K. (1963): Conjectures and Refutations. Routledge & Kegan Paul, London.

[25] Popper, K. (1959): The Logic of Scientific Discovery. Hutchinson, London.

[26] Rescher, N. (1980): Empirical Inquiry. The Athlone Press, London.

[27] Rescher, N. (1970): Scientific Explanation. The Free Press, New York.

[28] Runes, D. (Ed.) (1976): Dictionary of Philosophy. Littlefield, Adams & Co., Totowa, New Jersey.

[29] Von Wright, G. H. (1971): Explanation and Understanding. Cornell University Press, Cornell.

[30] Von Wright, G. H. (1957): The Logical Foundations of Probability. Blackwell, Oxford.

[31] Zadeh, L. Computation with Imprecise Probabilities. Plenary lecture, International Conference on Information Processing and Management of Uncertainty for Knowledge-Based Systems (IPMU), Malaga, Spain, June 22-27, 2008.

[32] Zadeh, L. (2006): From Search Engines to Question Answering Systems? The Problems of World Knowledge, Relevance, Deduction and Precisiation. In E. Sanchez, (Ed.), Fuzzy Logic and the Semantic Web, Elsevier, Amsterdam.

[33] Zadeh, L. (1975): Fuzzy Logic and Approximate Reasoning. Synthese 30, 407-428.

[34] Zadeh, L. (1996): Fuzzy Logic = Computing with Words. IEEE Transactions on Fuzzy Systems, 2, 103-111.

[35] Zadeh, L. (1997): Toward a Theory of Fuzzy Information Granulation and Its Centrality in Human Reasoning and Fuzzy Logic. Fuzzy Sets and Systems 90/2, 111-127.

[36] Zadeh, L. (1999): From Computing with Numbers to Computing with Words - From Manipulation of Measurements to Manipulation of Perceptions. IEEE Transactions on Circuits and Systems 45, 105-119.

[37] Zadeh, L. (2002): Toward a Perception-Based Theory of Probabilistic Reasoning with Imprecise Probabilities. Journal of Statistical Planning and Inference 105/2, 233-264.

[38] Zadeh, L., (2009): Toward Extended Fuzzy Logic – A First Step. Fuzzy Sets and Systems (to appear).

ISBN: 978-989-95079-6-8

On the completion mechanism produced by the Choquet integral on some decision strategies

Christophe Labreuche[1]

1. Thales Research & Technology
Campus de Polytechnique, 1 avenue Augustin Fresnel
91767 Palaiseau Cedex- France
christophe.labreuche@thalesgroup.com

Abstract— Preference modeling consists in constructing a preference relation from initial preferences given by a decision maker. We are interested in the preference relation obtained from the use of the Choquet integral. The necessity preference is constructed as the intersection of all preference relations corresponding to a Choquet integral which are compatible with the initial preferences of the decision maker. We study some properties of this necessity relation on some classes of initial preferences. This gives some properties on the completion or extension ability of the Choquet integral.

Keywords— Choquet integral, capacity, initial preferences.

1 Introduction

Multi-Criteria Decision Aid aims at representing the preferences of a Decision Maker (DM) over some options described by several points of view or attributes. We are interested in this paper in the case where the preference relation can be described by an overall utility in which the aggregation of the attributes is obtained with the Choquet integral. The preference model is then thoroughly constructed once the parameters of the Choquet integral - namely the fuzzy measure - are determined.

To this end, the facilitator asks the DM to provide some preferential information. The most broadly used type of preferential information is when the DM says that an option is preferred to another one. The major difficulty the facilitator faces is that there usually does not exist one single fuzzy measure that fulfils this preferential information. Most of the elicitation methods based on the Choquet integral consist in selecting a fuzzy measure (fulfilling the preferential information) that maximizes some functional, which may be an entropy for instance [1, 2, 3]. This is not quite satisfactory for the DM since he does not usually understand what maximizing the functional really means. The use of a maximization problem introduces some additional information that does not come from the DM.

The facilitator shall rather stick strictly to what the DM says and add no further information. One then looks for a robust way to recommend some comparisons among the options from the preferential information [4]. The concept of *necessity preference relation* has been recently introduced for robust quantitative multi-criteria decision models [5, 6]. It has been applied to the Choquet integral in [7]. An option is *necessarily* preferred to another option according to the necessity preference relation, if the first option is preferred to the second one according to all models that fulfill the preferential information provided by the DM. This necessity preference relation is usually incomplete, unless the model is completely specified from the preferential information of the DM. The necessity preference relation is cautious in the sense that its outcomes cannot be contradicted.

The goal of this paper is to study the necessity preference relation for several examples of typical preferential information, in the case where the underlying model is the Choquet integral. In particular, we want to analyze the strength of the completion mechanism obtained from the Choquet integral. This will give complementary insights (compare to the axiomatic characterizations) on the ability of the Choquet integral to represent DM preferences. For each preferential information, the layout is the following:

- Statement of the preferential information and precise description in terms of binary relations.
- Equivalence conditions on the model. We determine the necessary and sufficient conditions on the parameters of the model for this latter to satisfy the preferential information.
- Description of the necessity preference relation derived from the preferential information.

Section 2 presents a more formal description of the problem. The Choquet integral is defined in Section 3. Section 4 describes the process elicitation of the preferential information. Three decision strategies are then studied: the case of empty preference information (see Section 5), the relative importance (see Section 6) and the veto (see Section 7).

2 Description of the problem

Multi-Criteria Decision Aid aims at modeling the preferences of a Decision Maker (DM) over alternatives described by several points of view $N = \{1, \ldots, n\}$. The points of view to be taken into account in the decision making process are denoted by $X_1, \ldots, X_n$. An alternative is characterized by a value w.r.t. each point of view

ISBN: 978-989-95079-6-8

and is thus identified to a point in the Cartesian product X of the points of view: $X = X_1 \times \cdots \times X_n$.

Preference modeling aims at helping a DM to compare the options of a set $\Xi \subseteq X$ containing the alternatives of interest for him. We denote by $\succsim$ the preference relation over X that we wish to define. We are interested in the case where $\succsim$ results from an algebraic utility model. According to this model, the preference relation $\succsim$ can be represented by an overall utility $U : X \to \mathbb{R}$. Hence for all $x, y \in X$, we have

$$x \succsim y \iff U(x) \geq U(y).$$

A standard representation of U is the so-called transitive decomposable model $U(x) = F(u(x))$, where $u_i : X_i \to \mathbb{R}$ is the utility function on attribute $i \in N$, $u(x) = (u_1(x_1), \ldots, u_n(x_n))$, and $F : \mathbb{R}^N \to \mathbb{R}$ is the aggregation function. We assume here that u is known so that the options can be described by a vector in $\mathbb{R}^n$. Hence $X = \mathbb{R}^n$ and the utility functions are the identity function. The aggregation function is characterized by some parameter vector w that need to be determined. Let $\mathcal{W}$ the set of values of the parameters. The preference relation obtained from the parameter vector w is denoted by $\succsim_w$. We denote by $\succ_w$ and $\sim_w$ the asymmetric and symmetric parts of $\succsim_w$.

Any binary relation on $\mathbb{R}^n$ can be identified to a subset of $(\mathbb{R}^n)^2$ composed of the pairs (a, b) such that a is in relation with b.

The fact that the DM needs some help to compare the options in Ξ, means that the comparison of at least some of the options in Ξ is complex. Some preferential information, containing the preferences of the DM over several options in $X \backslash \Xi$, is asked to the DM. This preferential information is encoded by three partial orders $\trianglerighteq, \triangleright, \equiv$ over X. For $x, y \in X$, relation $x \trianglerighteq y$ means that the DM finds x at least as good as y, $x \triangleright y$ means that the DM finds x strictly better than y, and $x \equiv y$ means that the DM finds x similar to y. We set $\square = (\trianglerighteq, \triangleright, \equiv)$. In order to deduce a comparison of the elements of Ξ from $\square$, one necessarily needs some completion mechanism. We want to construct three binary relations $\succsim$, $\succ$ and $\sim$ on X that extend $\square$. The strength of the completion is measured through the differences $\succsim \backslash \trianglerighteq$, $\succ \backslash \triangleright$ and $\sim \backslash \equiv$ compare to $\trianglerighteq, \triangleright, \equiv$ respectively. For a very weak completion, we have $\succsim \backslash \trianglerighteq = \emptyset$, $\succ \backslash \triangleright = \emptyset$ and $\sim \backslash \equiv = \emptyset$.

Let $\mathcal{W}(\square)$ be the set of parameter vectors w for which $x \succsim_w y$ whenever $x \trianglerighteq y$, $x \succ_w y$ whenever $x \triangleright y$, and $x \sim_w y$ whenever $x \equiv y$. The completion is necessarily based on the family $\{\succsim_w\}_{w \in \mathcal{W}(\square)}$. A classical completion is the preference relation $\succsim_w$ corresponding to a particular $w \in \mathcal{W}(\square)$ solution to an optimization problem [1, 2, 3]

$$w = \text{argmax}_{w \in \mathcal{W}(\square)} G_\square(w) \tag{1}$$

where $G_\square$ is the function depending on $\square$ to be maximized. As said in the introduction, we are interested in the *necessity* preference relation $\succsim^{\mathcal{W}}_{\text{N},\square}$ that is constructed from $\square$ as follows [7, 5, 6]:

$$x \succsim^{\mathcal{W}}_{\text{N},\square} y \iff [\forall w \in \mathcal{W}(\square) \;\; x \succsim_w y]\,. \tag{2}$$

One may also define $\succ^{\mathcal{W}}_{\text{N},\square}$ and $\sim^{\mathcal{W}}_{\text{N},\square}$ by

$$x \succ^{\mathcal{W}}_{\text{N},\square} y \iff [\forall w \in \mathcal{W}(\square) \;\; x \succ_w y]$$
$$x \sim^{\mathcal{W}}_{\text{N},\square} y \iff [\forall w \in \mathcal{W}(\square) \;\; x \sim_w y]\,.$$

Note that one may have $\succsim^{\mathcal{W}}_{\text{N},\square} \neq \succ^{\mathcal{W}}_{\text{N},\trianglerighteq} \cup \sim^{\mathcal{W}}_{\text{N},\square}$.

Another interesting order relation derived from $\square$ is the so-called *possibility preference relation* defined as follows [7, 5, 6]:

$$x \succsim^{\mathcal{W}}_{\Pi,\square} y \iff [\exists w \in \mathcal{W}(\square) \;\; x \succsim_w y]\,.$$

Let us mention that these concepts of necessary and possible preference representation are originated from earlier works on preference modelling. In the context of decision rules, the necessary and possibility preference relations are the lower and upper approximations of a rough set [8]. Moreover, in qualitative decision models, these concepts correspond to the dominance and ordering queries for the Ceterus Paribus nets [9].

The choice of the representation of $\succsim_w$ (and thus of the family of aggregation functions F) is very important since it characterizes directly the completion mechanism. The choice of the family F results from a compromise: if the completion is too strong, we deduce much more than what the DM expresses and the DM may not agree with the outcome. If the completion is too weak, we basically do not deduce really more than what the DM says. Moreover, the stronger the completion mechanism, the less information $\square$ needs to be given.

As an example, the completion resulting from a simple model such a weighted sum is stronger than that obtained from a more general model such as the Choquet integral. We will illustrate this in the main part of the paper. Consider the case where F is the usual weighted sum: $F(x) = \sum_{i \in N} w_i\, x_i$ where w_i is the weight of criterion i. Thanks to the additivity properties of this model, one can deduce that $(1, \frac{1}{2}, 0) \succsim^{\mathcal{W}}_{\text{N},\square} (0, 1, \frac{1}{2})$, from the simplier comparisons $(\frac{1}{2}, \frac{1}{2}, 0) \trianglerighteq (0, \frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 0, 0) \trianglerighteq (0, \frac{1}{2}, 0)$ provided by the DM. The previous deduction is wrong for the Choquet integral. The weighted sum, which is a very simple aggregation model, implies a very strong completion mechanism.

The necessity preference relation $\succsim^{\mathcal{W}}_{\text{N},\square}$ corresponds to the comparisons among the options in X that can be made for sure. If the initial preferences $\square$ of the DM are fixed, the recommendation given by the facilitator shall be based only on $\succsim^{\mathcal{W}}_{\text{N},\square}$.

Now, if $\succsim^{\mathcal{W}}_{\text{N},\square}$ is not enough complete, the facilitator will probably try to enrich $\square$ with new comparisons provided by the DM. The candidate comparisons must be picked up from the possibility preference relation. Assume indeed that $\square$ is enriched with a new information "x is at least as good as y" such that $x \not\succsim^{\mathcal{W}}_{\Pi,\square} y$. The enriched preferential information is $\square' = (\trianglerighteq', \triangleright', \equiv')$ with $\trianglerighteq' = \trianglerighteq \cup \{(x, y)\}$, $\triangleright' = \triangleright$ and $\equiv' = \equiv$. Since $x \not\succsim^{\mathcal{W}}_{\Pi,\square} y$, we obtain $\mathcal{W}(\square') = \emptyset$. Hence the new preferential information to add to $\square$ must belong to $\succsim^{\mathcal{W}}_{\Pi,\square}$, $\succ^{\mathcal{W}}_{\Pi,\square}$ or $\sim^{\mathcal{W}}_{\Pi,\square}$. One clearly has [5, Proposition 4.1]

$$\succsim^{\mathcal{W}}_{\text{N},\square} \subseteq \succsim^{\mathcal{W}}_{\Pi,\square}\,.$$

ISBN: 978-989-95079-6-8

The comparisons contained in the necessity preference relation are automatically deduced from $\Box$, and do not therefore enrich $\Box$. In other words, if $(x,y) \in \succsim^{\mathcal{W}}_{\mathrm{N},\Box}$, then $\mathcal{W}(\Box') = \mathcal{W}(\Box)$, where $\Box' = (\trianglerighteq', \rhd', \equiv')$ with $\trianglerighteq' = \trianglerighteq \cup \{(x,y)\}$, $\rhd' = \rhd$ and $\equiv' = \equiv$. This proves that the new relevant preferential information to add to $\Box$ actually belongs to

$$\succsim^{\mathcal{W}}_{\Pi,\Box} \setminus \succsim^{\mathcal{W}}_{\mathrm{N},\Box}\ ,\ \succ^{\mathcal{W}}_{\Pi,\Box} \setminus \succ^{\mathcal{W}}_{\mathrm{N},\Box} \quad \text{or} \quad \sim^{\mathcal{W}}_{\Pi,\Box} \setminus \sim^{\mathcal{W}}_{\mathrm{N},\Box}\ .$$

In this paper, we will not study specifically the possibility preference relation. We will focus only on the necessity preference relation. The reason is that $\succsim^{\mathcal{W}}_{\mathrm{N},\Box}$ and $\succsim^{\mathcal{W}}_{\Pi,\Box}$ are strongly linked together. More precisely, since $\succsim_w$ is a complete preorder, we have

$$x \succsim^{\mathcal{W}}_{\Pi,\Box} y \iff y \not\succ^{\mathcal{W}}_{\mathrm{N},\Box} x.$$

In other words, $\succsim^{\mathcal{W}}_{\Pi,\Box} = X^2 \setminus (\succ^{\mathcal{W}}_{\mathrm{N},\Box})^{-1}$.

3 Notation and the Choquet integral

We hereafter restrict to the case where F is a Choquet integral.

3.1 Choquet integral

A fuzzy measure (also called *capacity*) on a set N of criteria is a set function $\mu : 2^N \to [0,1]$ such that [10]

- $\mu(\emptyset) = 0$, $\mu(N) = 1$,
- $\forall A \subseteq B \subseteq N\ ,\ \mu(A) \leq \mu(B)$

Let $\mathcal{M}$ be the set of all fuzzy measures. A fuzzy measure is said to be additive if $\mu(A \cap B) = \mu(A) + \mu(B)$ for every pair (A,B) of disjoint coalitions.

The Choquet integral of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ defined w.r.t. a capacity μ has the following expression [11] :

$$C_\mu(x_1, \ldots, x_n) = \sum_{i=1}^{n} \left(x_{\tau(i)} - x_{\tau(i-1)}\right) \times \mu\left(\{\tau(i), \cdots, \tau(n)\}\right), \qquad (3)$$

where τ is a permutation on N such that $x_{\tau(1)} \leq x_{\tau(2)} \leq \cdots \leq x_{\tau(n)}$, and $x_{\tau(0)} := 0$. The Choquet integral has been proved to be able to model both the importance of criteria and the interaction between criteria. One fundamental property of the Choquet integral is that

$$C_\mu(1_A, 0_{N \setminus A}) = \mu(A). \qquad (4)$$

3.2 k-additive capacities

We introduce a useful linear transformation of fuzzy measures. The *Möbius transform* m of a fuzzy measure μ is the unique solution of the equation

$$\forall A \subseteq N \quad \mu(A) = \sum_{B \subseteq A} m(B), \qquad (5)$$

and is given by:

$$m^\mu(A) := \sum_{B \subseteq A} (-1)^{|A|-|B|} \mu(B). \qquad (6)$$

A fuzzy measure is defined by 2^n coefficients, which is much more than a weighted sum. The concept of k-additive fuzzy measure has a complexity in-between a fuzzy measure and a weighted sum. More precisely, a fuzzy measure μ is said to be k-additive [12] if $m^\mu(A) = 0$ whenever $|A| > k$ and there exists A with $|A| = k$ such that $m^\mu(A) \neq 0$. We denote by $\mathcal{M}^k$ the set of fuzzy measures that are at most k-additive (i.e. 1, or 2, or ..., or k additive). Note that $\mathcal{M} = \mathcal{M}^n$ and that $\mathcal{M}^1$ are the additive fuzzy measures and corresponds to the usual weighted sum. An interesting particular case is when $k = 2$.

3.3 Interpretation

A capacity contains 2^n terms, which makes its interpretation complex for a DM. In order to ease its interpretation, several global indices have been defined. The first index called *importance index* aims at measuring the degree to which a criterion is important for the decision. The importance index of criterion $i \in N$ is defined by [13]:

$$v_i := \sum_{A \subset N \setminus \{i\}} \frac{(n - |A| - 1)!|A|!}{n!} [\mu(A \cup \{i\}) - \mu(A)].$$

Note that the definition is originated from Cooperative Game Theory in which it corresponds to a fair share of a common wealth among several players.

The interaction index I_{ij} of the couple of criteria $\{i,j\}$ is defined by [14]

$$I_{ij} := \sum_{A \subset N \setminus \{i,j\}} \frac{(n - |A| - 2)!|A|!}{(n-1)!} [\mu(A \cup \{i,j\}) - \mu(A \cup \{i\}) - \mu(A \cup \{j\}) + \mu(A)].$$

The Shapley index v_i can be interpreted as a kind of average value of the contribution of criterion i alone in all coalitions. The interaction index I_{ij} can be interpreted as a kind of average value of the *added value* obtained by putting criterion i and j together. When I_{ij} is positive (resp. negative), the interaction is said to be positive (resp. negative). These two concepts correspond to a mean behavior and are anyhow not so easy to understand for a DM.

3.4 Initial preferential information

We denote by $\mathcal{M}^k(\Box)$ the set of capacities in $\mathcal{M}^k$ that satisfy the preferential information $\Box = (\trianglerighteq, \rhd, \equiv)$:

$$\mathcal{M}^k(\Box) = \Big\{ \mu \in \mathcal{M}^k\ , \forall a, b \in \mathbb{R}^n \quad a \trianglerighteq b \ \Rightarrow\ C_\mu(a) \geq C_\mu(b) \quad a \rhd b \ \Rightarrow\ C_\mu(a) > C_\mu(b) \quad a \equiv b \ \Rightarrow\ C_\mu(a) = C_\mu(b) \Big\}$$

The necessity preference relation on $\mathbb{R}^n$ is a cautious way to take into account the fact that the aggregation model is not uniquely determined from a knowledge of $\Box$, and

ISBN: 978-989-95079-6-8

is defined by:

$$\forall a, b \in \mathbb{R}^n \quad , \quad a \succsim_{\mathrm{N},\Box}^{\mathcal{M}^k} b \iff [\forall \mu \in \mathcal{M}^k(\Box) \quad C_\mu(a) \geq C_\mu(a)] \, .$$

It is easy to see that $\succsim_{\mathrm{N},\Box}^{\mathcal{M}^k}$ is usually incomplete, but is transitive and reflexive. One can define $\succ_{\mathrm{N},\Box}^{\mathcal{M}^k}$ and $\sim_{\mathrm{N},\Box}^{\mathcal{M}^k}$ as in Section 2. One has

$$\sim_{\mathrm{N},\Box}^{\mathcal{M}^k} = \succsim_{\mathrm{N},\Box}^{\mathcal{M}^k} \cap (\succsim_{\mathrm{N},\Box}^{\mathcal{M}^k})^{-1} \quad \text{and} \quad \succ_{\mathrm{N},\Box}^{\mathcal{M}^k} \subseteq \succsim_{\mathrm{N},\Box}^{\mathcal{M}^k} \setminus \sim_{\mathrm{N},\Box}^{\mathcal{M}^k} .$$

Let us define finally the incomparability as $I_{\mathrm{N},\Box}^{\mathcal{M}^k} = (\mathbb{R}^n)^2 \setminus (\succsim_{\mathrm{N},\Box}^{\mathcal{M}^k} \cup (\succsim_{\mathrm{N},\Box}^{\mathcal{M}^k})^{-1})$.

4 Elicitation process

The preference model is usually constructed iteratively through a decision aiding process [15]. This process is centered around the interaction between the DM and the decision aiding system. The information that the DM expresses regarding his preferences can be of two types:

- The first type of information can be easily produced and interpreted by the DM. It is usually admitted that the DM can naturally compare some options. In this paper, the only information of this type corresponds to the three binary relations $\Box = (\trianglerighteq, \rhd, \equiv)$. It is called the *Preferential information*.

- The second type of information (called *Decision strategies*) corresponds to abstract information concerning the preference model. It can typically concern concepts such as the mean importance, veto, relative importance,... The DM can understand them but only to some extend, but it is unexpected that he will naturally provide information of this type. If the decision analyst guesses that the DM fulfills to a decision strategy, he needs to convert it to preferential information in order to check the validity of his guess.

At the end, the only learning data that is considered from the DM corresponds to some preferential information $\Box$. It is transformed in the *internal representation* of the model. This is the set of parameter vectors fulfilling the preferential information, that is $\mathcal{M}^k(\Box)$. It is not understood by the DM. Consistency and completeness is checked in this representation. The necessity preference relation can then be constructed. This relation can be presented to the DM in different ways. It can be directly applied to the set Ξ of options of interest for the DM. It is also possible to interpret the necessity preference relation in terms of decision strategies. The interpretation may concern some indices regarding the mean importance, veto, relative importance, if all the admissible capacities $\mathcal{M}^k(\Box)$ share some common properties on one of these concepts.

The elicitation process can thus be seen as a flow diagram between the three concepts and is organized on the steps *a* to *f* in Figure 1:

- *a*: The DM can express some abstract information regarding his decision strategies. He may express for instance that a criterion is more important than another one.

- *b*: The DM can express information in the format of the order relations $\Box$. He may say for instance that an option is preferred to another one. If the DM has already given some preferential information (denoted by $\Box'$) in a previous step of the elicitation process and he wishes to enrich $\Box'$, the new preferential information shall lie in the sets $\succsim_{\Pi,\Box'}^{\mathcal{M}^k} \setminus \succsim_{\mathrm{N},\Box'}^{\mathcal{M}^k}$, $\succ_{\Pi,\Box'}^{\mathcal{M}^k} \setminus \succ_{\mathrm{N},\Box'}^{\mathcal{M}^k}$ or $\sim_{\Pi,\Box'}^{\mathcal{M}^k} \setminus \sim_{\mathrm{N},\Box'}^{\mathcal{M}^k}$.

- *c*: When the DM expresses abstract information (decision strategies), they shall be converted into preferential information through $\Box$ in order to be precisely set and taken into account by the elicitation method.

- *d*: The preferential information is transformed into constraints in the parameter set. This corresponds to the set $\mathcal{M}^k(\Box)$. This step is sometimes called disaggregation step.

- *e*: An interpretation of the set $\mathcal{M}^k(\Box)$ is presented in terms of, for instance, veto, importance, interaction indices. It can also provide the necessity and possibility preference relations. This step is sometimes called aggregation step.

- *f*: An interpretation of the results of the elicitation is displayed to the DM in some understandable way.

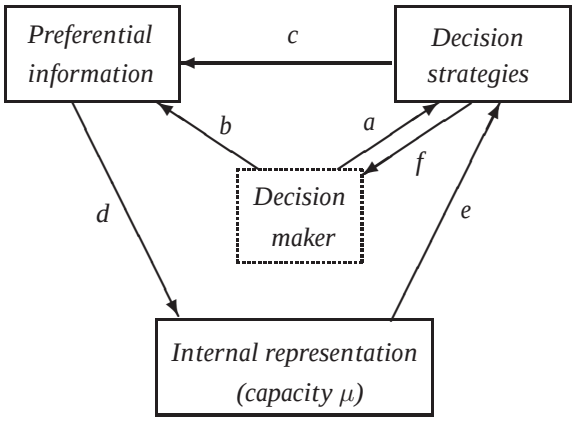


Figure 1: Eliciation process.

In the remaining of this paper, we are interested in some abstract decision strategies that the DM often wishes to express in practice. For each type of preferential information, we determine how it is converted into order relations $\Box$ (step *c*). Then, we construct the necessity preference relation corresponding to this preferential information (steps *d* and *e*).

5 Case of empty preference information

We denote by $\Box_\emptyset$ the empty preference information. When the DM says nothing, there is of course no new

ISBN: 978-989-95079-6-8

constraint on the fuzzy measure:

$$\mathcal{M}^k(\Box_\emptyset) = \mathcal{M}^k.$$

Moreover, it is easy to check that for all $k \in \{1, \ldots, n\}$

$$\succsim^{\mathcal{M}^k}_{\mathrm{N},\Box_\emptyset} = \mathcal{P}_{\mathrm{a}} := \{(a,b) \in \mathbb{R}^n \ , \ \forall i \in N \ \ a_i \geq b_i\} \ .$$

This relation extends [5, Remark 4.1] to the case of the Choquet integral. This is the standard Pareto dominance order. When no preferential information is given, the necessary preferential information is similar to the Pareto dominance order, which is what one expects.

6 Relative importance between criteria

6.1 Statement of the preferential information

We want to express a preferential information regarding the relative importance between two criteria – namely that criterion p is more important than criterion q. The first idea is to interpret this as the relation $v_p > v_q$ based on the Shapley value. This relation is considered in [1, 7]. However, as explained in Section 3.3, the Shapley value is not a simple concept for a DM. It is not clear at all whether the statement *"criterion p is more important than criterion q"* expressed by a DM can be represented by the relation $v_p > v_q$. We wish only to express the previous statement only as binary relations denoted by $\Box^{p,q}_{\mathrm{RelImp}} = (\trianglerighteq^{p,q}_{\mathrm{RelImp}}, \triangleright^{p,q}_{\mathrm{RelImp}}, \equiv^{p,q}_{\mathrm{RelImp}})$. The following definition seems to express quite accurately the previous statement.

Definition 1 ([16]) *A criterion $p \in N$ is said to be* more important *than criterion $q \in N \setminus \{p\}$ iff*

$$\left(\gamma_p, \beta_q, a_{N\setminus\{p,q\}}\right) \triangleright^{p,q}_{\mathrm{RelImp}} \left(\beta_p, \gamma_q, a_{N\setminus\{p,q\}}\right) \qquad (7)$$

for all $\gamma, \beta \in \mathbb{R}$ with $\gamma > \beta$, and all $a_{N\setminus\{p,q\}} \in \mathbb{R}^{n-2}$, where $\left(\alpha_p, \beta_q, a_{N\setminus\{p,q\}}\right)$ denotes the alternative having value α on criterion p, value β on criterion q, and value $a_{N\setminus\{p,q\}}$ on the remaining criteria.

Note that (7) makes sense only if the scales of the criteria are commensurate.

6.2 Conditions on the fuzzy measure

Let $\delta_{p,q}\mu(A) := \mu(A \cup \{p\}) - \mu(A \cup \{q\})$.

Lemma 1 *For all $k \in \{2, \ldots, n\}$, criterion p is more important than criterion q with the Choquet integral with a capacity in $\mathcal{M}^k$ iff*

$$\forall A \subseteq N \setminus \{p,q\} \quad \delta_{p,q}\mu(A) > 0. \qquad (8)$$

6.3 Completion

We have the following result.

Proposition 1 *Let $p, q \in N$ with $p \neq q$. Then for all $k \in \{2, \ldots, n\}$*

$$\begin{aligned}
&\succsim^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
&= \Big\{ \Big((\gamma_p, \beta_q, a_{N\setminus\{p,q\}}), (\beta'_p, \gamma'_q, a'_{N\setminus\{p,q\}}) \Big) \in (\mathbb{R}^n)^2 \ , \\
&\quad \gamma > \beta \ , \ \gamma' > \beta' \ , \gamma' \leq \gamma \ , \ \beta' \leq \beta \text{ and} \\
&\quad \text{for all } i \in N \setminus \{p,q\} \ \ a'_i \leq a_i \Big\} \bigcup \mathcal{P}_{\mathrm{a}} \\
&\succ^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
&= \Big\{ \Big((\gamma_p, \beta_q, a_{N\setminus\{p,q\}}), (\beta'_p, \gamma'_q, a'_{N\setminus\{p,q\}}) \Big) \in (\mathbb{R}^n)^2 \ , \\
&\quad \gamma > \beta \ , \ \gamma' > \beta' \ , \gamma' \leq \gamma \ , \ \beta' \leq \beta \ \text{and} \\
&\quad \text{for all } i \in N \setminus \{p,q\} \ \ a'_i \leq a_i \Big\} \bigcup \mathcal{P}^p_{\mathrm{a}>} \\
&\sim^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} = \Big\{ (a,a) \ , \ \ a \in \mathbb{R}^n \Big\}
\end{aligned}$$

where $\mathcal{P}^p_{\mathrm{a}>} = \{(a,b) \in \mathcal{P}_{\mathrm{a}} \ , \ a_p > b_p\}$.

When $\gamma > \beta$, $(\gamma_p, \beta_q, a_{N\setminus\{p,q\}})$ is at least as good as $(\beta_p, \gamma_q, a_{N\setminus\{p,q\}})$ from $\Box^{p,q}_{\mathrm{RelImp}}$. From the Pareto relation $\mathcal{P}_{\mathrm{a}}$, $(\beta_p, \gamma_q, a_{N\setminus\{p,q\}})$ is at least as good as $(\beta'_p, \gamma'_q, a'_{N\setminus\{p,q\}})$ whenever $\gamma' \leq \gamma$, $\beta' \leq \beta$ and $a'_k \leq a_k$ for all $k \in N \setminus \{p,q\}$. The completion of $\Box^{p,q}_{\mathrm{RelImp}}$ by a k-additive Choquet integral (with $k \geq 2$) is just the combination of the definition of $\Box^{p,q}_{\mathrm{RelImp}}$ and the Pareto dominance order. Therefore, there is no side effect coming from the Choquet model.

Lemma 2 *Let $p, q \in N$ with $p \neq q$. Then*

$$\begin{aligned}
&\succsim^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
&= \Big\{ \Big((\gamma_p, \beta_q, a_{N\setminus\{p,q\}}), (\beta'_p, \gamma'_q, a'_{N\setminus\{p,q\}}) \Big) \in (\mathbb{R}^n)^2 \\
&\quad \gamma \geq \beta' \ , \ \gamma - \beta + \beta' - \gamma' \geq 0 \\
&\quad \text{and for all } i \in N \setminus \{p,q\} \ \ a'_i \leq a_i \Big\} \bigcup \mathcal{P}_{\mathrm{a}} \\
&\succ^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
&= \Big\{ \Big((\gamma_p, \beta_q, a_{N\setminus\{p,q\}}), (\beta'_p, \gamma'_q, a'_{N\setminus\{p,q\}}) \Big) \in (\mathbb{R}^n)^2 \\
&\quad \gamma > \beta' \ , \ \gamma - \beta + \beta' - \gamma' \geq 0 \ , \ \gamma' \leq \gamma \ , \ \beta' \leq \beta \\
&\quad \text{and for all } i \in N \setminus \{p,q\} \ \ a'_i \leq a_i \Big\} \cup \mathcal{P}^p_{\mathrm{a}>} \\
&\sim^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} = \Big\{ (a,a) \ , \ \ a \in \mathbb{R}^n \Big\}
\end{aligned}$$

There is a strange condition $\gamma - \beta + \beta' - \gamma' \geq 0$ in the necessity preference relation of the model $\mathcal{M}^1$.

One has for all $k \in \{2, \ldots, n\}$

$$\begin{aligned}
\succsim^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} &\subset \succsim^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
\succ^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} &\subset \succ^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \\
\sim^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} &\subset \sim^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}}
\end{aligned}$$

The information $\succsim^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \setminus \succsim^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}}$ and $\succ^{\mathcal{M}^1}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}} \setminus \succ^{\mathcal{M}^k}_{\mathrm{N},\Box^{p,q}_{\mathrm{RelImp}}}$ is contained neither in Definition

ISBN: 978-989-95079-6-8

1 nor in the Pareto set $\mathcal{P}_{\rm a}$. This proves that the completion resulting from the Choquet integral fits exactly with the preferential information provided by the DM, whereas the completion resulting from the weighted sum is strong. The Choquet is thus able to represent the relative importance without adding new information not provided by the DM.

7 Veto criteria

7.1 Statement of the preferential information

As in group decision, one can define a veto criterion. The preferential information corresponding to a veto criterion p is denoted by $\Box^p_{\rm Veto} = (\trianglerighteq^p_{\rm Veto}, \triangleright^p_{\rm Veto}, \equiv^p_{\rm Veto})$.

Definition 2 *Criterion p is a* veto *criterion if for all $x, y \in \mathbb{R}^n$*

$$(x_{N\setminus\{p\}}, 0_p) \equiv^p_{\rm Veto} (y_{N\setminus\{p\}}, 0_p).$$

It is straightforward to see that the evaluation of the option $(x_{N\setminus\{p\}}, 0_p)$ is zero. This means that a bad score on a veto criterion cannot be saved by good scores on the remaining criteria.

7.2 Conditions on the fuzzy measure

Lemma 3 *Criterion $p \in N$ is a veto iff*

$$\forall S \subset N \setminus \{p\}\ , \quad \mu(S) = 0\ . \tag{9}$$

7.3 Completion

There is a unique additive fuzzy measure for which p is a veto. It is defined by $\mu(A) = 1$ if $p \in A$ and $= 0$ otherwise. Criterion p is then interpreted as a dictator. All the other criteria are merely discarded. Hence $\succsim^{\mathcal{M}^1}_{{\rm N},\Box^p_{\rm Veto}}$ is a complete order.

The following result shows that the necessity preference relation does not depend on k for $k \geq 2$.

Proposition 2 *Let P be the preferential information stating that criterion p is a veto. Then for all $k \in \{2, \ldots, n\}$*

$$\begin{aligned}
\succsim^{\mathcal{M}^k}_{{\rm N},\Box^p_{\rm Veto}} &= \Big\{(a,b) \in (\mathbb{R}^n)^2\ ,\ a_p \geq b_p \textit{ and} \\
&\quad \forall i \in N \setminus \{p\} \textit{ either } a_i \geq b_i \textit{ or } [a_i < b_i \textit{ and } a_i \geq b_p]\Big\} \\
\succ^{\mathcal{M}^k}_{{\rm N},\Box^p_{\rm Veto}} &= \Big\{(a,b) \in (\mathbb{R}^n)^2\ ,\ a_p > b_p \textit{ and} \\
&\quad \forall i \in N \setminus \{p\} \textit{ either } a_i > b_i \textit{ or } [a_i \leq b_i \textit{ and } a_i > b_p]\Big\} \\
\sim^{\mathcal{M}^k}_{{\rm N},\Box^p_{\rm Veto}} &= \Big\{(a,b) \in (\mathbb{R}^n)^2\ ,\ a_p = b_p \textit{ and} \\
&\quad \forall i \in N \setminus \{p\} \textit{ either } a_i = b_i \textit{ or } [a_i \geq a_p \textit{ and } b_i \geq a_p]\Big\}
\end{aligned}$$

The condition $a_i < b_i$ and $a_i \geq b_p$ means that the relatively bad score of a on criterion i is hidden by the score of b on the veto criterion p. Hence the necessity preference relation is completely natural. The completion mechanism of the Choquet integral is once more easily interpretable.

8 Conclusion

We have studied the completion mechanism produced by the Choquet integral on three types of preferential information: empty prior information, relative importance and veto. The result is trivial on the empty prior information since one recovers the Pareto ordering. Concerning the two other types of preferential information, we have shown that the Choquet integral does not introduce weird phenomena in the robust completion corresponding to the necessity preference relation. Moreover, the completion does not depend on order k for k-additive fuzzy measures, provided that $k \geq 2$. On the contrary, the weighted sum produced a strong completion.

References

[1] M. Grabisch, I. Kojadinovic, and P. Meyer. A review of capacity identification methods for Choquet integral based multi-attribute utility theory — applications of the Kappalab R package. *Eur. J. of Operational Research*, 186:766–785, 2008.

[2] I. Kojadinovic. Minimum variance capacity identification. *Quaterly J. of Operations Research (4OR)*, 12:23–36, 2006.

[3] Ch. Labreuche. Identification of a fuzzy measure with an l^1 entropy. In *Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, Malaga, Spain, June 22-27 2008.

[4] B. Roy. *Multicriteria methodology for decision aiding*. Kluwer Academic Publishers, Dordrecht, 1996.

[5] S. Greco, B. Matarazzo, and R. Słowinski. Ordinal regression revisited: Multiple criteria ranking with a set of additive value functions. *European Journal of Operational Research*, 191:416–436, 2008.

[6] S. Greco, R. Słowinski, J. Figueira, and V. Mousseau. Robust ordinal regression. In M. Ehrgott J. Figueira, S. Greco, editor, *New Trends in Multiple Criteria Decision Analysis*. Springer Verlag, to appear.

[7] S. Angilella, S. Greco, and B. Matarazzo. Non-additive robust ordinal regression: a multiple criteria decision model based on the choquet integral. *European Journal of Operational Research*, to appear.

[8] S. Greco, B. Matarazzo, and R. Słowinski. Decision rule approach. In M. Ehrgott J. Figueira, S. Greco, editor, *Multiple Criteria Decision Analysis - State of the Art Surveys*, pages 507–561. Springer International Series, 2005.

[9] C. Boutilier, R. Brafman, C. Domshlak, H. Hoos, and D. Poole. CP-nets: a tool for representing and reasoning with conditional Ceteris Paribus preference statements. *J. of Artificial Intelligence Research*, 21:135–191, 2004.

[10] M. Grabisch, T. Murofushi, and M. Sugeno. *Fuzzy Measures and Integrals. Theory and Applications (edited volume)*. Studies in Fuzziness. Physica Verlag, 2000.

[11] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[12] M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *Int. J. of Uncertainty, Fuzziness, and Knowledge Based Systems*, 5:587–607, 1997.

[13] L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.

[14] T. Murofushi and S. Soneda. Techniques for reading fuzzy measures (III): interaction index. In *9th Fuzzy System Symposium*, pages 693–696, Sapporo, Japan, May 1993. In Japanese.

[15] A. Tsoukiàs. From decision theory to decision aiding methodology. *European Journal of Operational Research*, 187:138–161, 2008.

[16] S. Kaci and H. Prade. Constraints associated with choquet integrals and other aggregation-free ranking devices. In *Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 1344–1351, Malaga, Spain, June 22-27 2008.

ISBN: 978-989-95079-6-8

Fuzzy Implications and the Weak Law of Importation

Sebastià Massanet Joan Torrens

Departament de Ciències Matemàtiques i Informàtica
Universitat de les Illes Balears
07122 Palma de Mallorca, Spain
Email: sebasmassanet@gmail.com, dmijts0@uib.es

Abstract— *Some open problems on fuzzy implications dealing with the so-called importation law are studied and totally or partially solved in this work. In particular, it is proved that such property (in fact a weaker version than the law of importation) is stronger than the exchange principle. Along this study, new characterizations of (S, N)-implications and R-implications involving the law of importation are showed.*

Keywords— Implication function, exchange principle, law of importation, (S, N)-implication, R-implication.

1 Introduction

Fuzzy implications are an essential tool in fuzzy control and approximate reasoning, as well as in many fields where these theories apply. This is because they are used not only to model fuzzy conditionals, but also to make inferences in any fuzzy rule based system (see for instance [1] or [2]) through the modus ponens and modus tollens. Moreover, fuzzy implications are also useful in many other fields like fuzzy relational equations and fuzzy mathematical morphology ([3]), fuzzy DI-subsethood measures and image processing ([4] and [5]), and data mining ([6]). Due to this great quantity of applications many authors have focused their interest in the theoretical study of fuzzy implications. See for instance the recent book [7], exclusively devoted to fuzzy implications, and the references therein.

From this theoretical study some open problems on fuzzy implications have been recently posed in some works about this topic. In this paper we want to deal with some of them involving the so-called law of importation, that is derived from the tautology in classical logic

$$(p \wedge q) \to r \equiv p \to (q \to r) \tag{1}$$

which, translated to fuzzy logic, becomes the functional equation:

$$I(T(x, y), z) = I(x, I(y, z)) \quad \text{for all } x, y, z \in [0, 1] \tag{2}$$

where T is a t-norm and I is a fuzzy implication. This property has been extensively studied in [8] and [9] for many kinds of implications derived from t-norms and t-conorms. Moreover, an extension of (2) involving uninorms instead of t-norms has been also studied in [10]. From this study some open problems for fuzzy implications involving the law of importation have been posed (see [11]) that we will try to solve in next sections. During the process, new characterizations of (S, N)-implications and R-implications involving the law of importation are showed.

2 Preliminaries

We will suppose the reader to be familiar with the theory of t-norms, t-conorms and fuzzy negations (all necessary results and notations can be found in [12]). We recall here only some facts on implications.

Definition 1 *A binary operator $I : [0, 1]^2 \to [0, 1]$ is said to be an* implication function*, or an* implication*, if it satisfies:*

(I1) $I(x, z) \geq I(y, z)$ *when* $x \leq y$*, for all* $z \in [0, 1]$.

(I2) $I(x, y) \leq I(x, z)$ *when* $y \leq z$*, for all* $x \in [0, 1]$.

(I3) $I(0, 0) = I(1, 1) = 1$ *and* $I(1, 0) = 0$.

Note that, from the definition, it follows that $I(0, x) = 1$ and $I(x, 1) = 1$ for all $x \in [0, 1]$ whereas the symmetrical values $I(x, 0)$ and $I(1, x)$ are not derived from the definition.

Special interesting properties for implication functions are:

- The *law of importation* with a t-norm T,

$$I(T(x, y), z) = I(x, I(y, z)), \quad \text{for all} \quad x, y, z \in [0, 1]. \tag{LI}$$

- The *exchange principle*,

$$I(x, I(y, z)) = I(y, I(x, z)), \quad \text{for all} \quad x, y, z \in [0, 1]. \tag{EP}$$

- The *ordering property*,

$$x \leq y \iff I(x, y) = 1, \quad \text{for all} \quad x, y \in [0, 1]. \tag{OP}$$

Among other models (see [3] and [7]), the most used fuzzy implications are derived from t-norms and t-conorms and they are:

- R-implications derived from a left-continuous t-norm T,

$$I_T(x, y) = \sup\{z \in [0, 1] \mid T(x, z) \leq y\} \tag{3}$$

  for all $x, y \in [0, 1]$.

- (S, N)-implications derived from a t-conorm S and a strong negation[1] N,

$$I_{S,N}(x, y) = S(N(x), y) \qquad \text{for all} \quad x, y \in [0, 1]. \tag{4}$$

[1] This kind of implications has been characterized also when N is strict or continuous, see [13]. However, (LI) has been studied only in the cases when N is strict or strong, but not when N is only continuous.

ISBN: 978-989-95079-6-8

But also QL-implications and D-implications respectively given by

$$I_{QL}(x,y) = S(N(x), T(x,y)) \tag{5}$$

for all $x, y \in [0,1]$, and

$$I_D(x,y) = S(T(N(x), N(y)), y) \tag{6}$$

for all $x, y \in [0,1]$, where T is a t-norm, S a t-conorm and N a strong negation. For all these kinds of implications the (LI) is equivalent to the (EP) and moreover, there is one and only one t-norm for which the (LI) holds (see [8] and [9]).

Finally let us recall the characterizations of (S,N) and R-implications. Several of these characterizations can be found in the literature:

Theorem 1 *([13]) For a function $I : [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i). I is an (S,N)-implication generated from some t-conorm S and some continuous (strict, strong) fuzzy negation N

(ii). I satisfies (I1), (EP) and $I(x,0) = N(x)$ is a continuous (strict, strong) fuzzy negation.

Theorem 2 *([14], [15] or [7]) Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is an R-implication derived from a left-continuous t-norm, if and only if, I satisfies (I2), (OP), (EP) and I is right-continuous with respect to the second variable.*

With the assumption of continuity we have a characterization of the following subclass of R-implications, that are also (S,N)-implications, known as the Smets-Magrez Theorem, see [16] and see also [17] for the current version.

Theorem 3 *Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is a continuous function satisfying (OP), (EP), if and only if, I is conjugate with the Łukasiewicz implication, that is, there exists a unique increasing bijection $\varphi : [0,1] \to [0,1]$ such that*

$$I(x,y) = \varphi^{-1}(\min\{1, 1 - \varphi(x) + \varphi(y)\}) \tag{7}$$

for all $x, y \in [0,1]$.

3 Main results

Due to the commutativity of the t-norm T (or the uninorm), it is clear that the law of importation implies the exchange principle, but in all known cases the converse also holds and the t-norm for which (LI) holds is uniquely determined. The first two open problems (posed in [11], Problem 8.1) that we want to study are just the following:

1. **(OP1)** Does the exchange principle imply that there exists a t-norm such that the law of importation holds? If not, give an example and characterize all fuzzy implications for which the answer is positive.

2. **(OP2)** Is the t-norm in the law of importation uniquely determined?

Note that for R, (S,N), QL and D-implications derived from t-norms and t-conorms the answer of both problems is positive (see [8] and [9]). With respect to (OP1) a negative answer has been done in [10] (see also [7], Section 7.3) by using implications derived from uninorms. The examples given in these works are implications satisfying (EP) such that they do not satisfy (LI) for any t-norm. However, in all these cases there exists a conjunctive uninorm for which the (LI) holds and thus, the problems above are just translated to the following more general versions:

1. **(OP1')** Does the exchange principle imply that there exists a uninorm (or a more general binary operation) such that the law of importation holds? If not, give an example and characterize all fuzzy implications for which the answer is positive.

2. **(OP2')** Is the uninorm in the law of importation uniquely determined?

Moreover, the answer of both problems is again positive for R, S, QL and D-implications derived from uninorms (see [10]).

Along the paper, we will prove that (OP1') has a negative answer in general and we will characterize some particular cases for which the answer is positive. On the other hand, with respect to (OP2) some counterexamples have been proved in [7] for implications I for which $I(x,0)$ equals the greatest or the least fuzzy negation (the same examples work for (OP2')). We will discuss however the case when $I(x,0)$ is at least a continuous fuzzy negation.

3.1 The Weak Law of Importation (WLI)

To deal with these problems let us introduce a weaker law of importation than (LI), by reducing the requirements on the function T. To maintain its relation with the exchange principle we need to take T commutative, and due to the monotonicity of implication functions it seems also adequate to take T nondecreasing, but no more conditions are necessary. Thus we introduce the following

Definition 2 *Let $I : [0,1]^2 \to [0,1]$ be a function. We say that I satisfies the* weak law of importation *if there exists a commutative and nondecreasing function $F : [0,1]^2 \to [0,1]$ such that*

$$I(F(x,y),z) = I(x, I(y,z)) \quad \textit{for all } x,y,z \in [0,1] \tag{WLI}$$

and then we say that I satisfies (WLI) with the function F.

First of all, note that (WLI) clearly implies (EP).

On the other hand, when I is an (S,N), a QL or a D-implication it satisfies the following boundary property

$$I(x,0) = N(x), \quad \text{for all} \quad x \in [0,1] \tag{BP}$$

where N is a (continuous, strict, strong) fuzzy negation.

We want to study the properties of those functions $I : [0,1]^2 \to [0,1]$ that satisfy (WLI) and (BP). These properties will be useful in order to find a counterexample to the equivalence of (EP) and (LI).

Proposition 1 *Let $I : [0,1]^2 \to [0,1]$ be a function which satisfies (WLI) with a commutative and nondecreasing function F and (BP) with a fuzzy negation N. Then*

ISBN: 978-989-95079-6-8

- *I satisfies (I3).*
- *I satisfies right contrapositive symmetry, that is,*

$$I(x, N(y)) = I(y, N(x)) \text{ for all } x, y \in [0,1]. \quad (8)$$

Moreover, if N is continuous, then

- *I satisfies (I1) and (I2), and consequently I is a fuzzy implication.*
- *I satisfies left neutrality principle, that is,*

$$I(1, y) = y \text{ for all } y \in [0,1]. \quad (9)$$

Finally, if N is strong then

- *I satisfies contrapositive symmetry, that is,*

$$I(N(y), N(x)) = I(x, y) \text{ for all } x, y \in [0,1]. \quad (10)$$

Proof: I satisfies (EP) due to the commutativity of F and (WLI). Furthermore, I satisfies also (I3) because

$$\begin{array}{ll} I(0,0) & = N(0) = 1 \\ I(1,0) & = N(1) = 0 \\ I(1,1) & = I(1, I(0,0)) = I(0, I(1,0)) = I(0,0) = 1. \end{array}$$

In addition, I satisfies right contrapositive symmetry because

$$I(x, N(y)) = I(x, I(y,0)) = I(y, I(x,0)) = I(y, N(x)).$$

From now on, N will be a continuous negation. Let us prove (I1). If $x \leq y$, using the continuity of N, we only need to show that $I(x, N(z)) \geq I(y, N(z))$ for all $z \in [0,1]$.

$$\begin{array}{ll} I(x, N(z)) & = I(x, I(z,0)) = I(F(x,z), 0) = N(F(x,z)) \\ & \geq N(F(y,z)) = I(F(y,z), 0) = I(y, I(z,0)) \\ & = I(y, N(z)) \end{array}$$

A similar argument shows (I2). If $y \leq z$, using the continuity of N, we will show that $I(x, N(y)) \geq I(x, N(z))$ for all $x, y, z \in [0,1]$.

$$\begin{array}{ll} I(x, N(y)) & = I(x, I(y,0)) = I(F(x,y), 0) = N(F(x,y)) \\ & \geq N(F(x,z)) = I(F(x,z), 0) = I(x, I(z,0)) \\ & = I(x, N(z)) \end{array}$$

Furthermore, I satisfies the left neutrality principle

$$\begin{array}{ll} I(1, N(y)) & = I(1, I(y,0)) = I(y, I(1,0)) \\ & = I(y,0) = N(y) \end{array}$$

Finally, if N is strong, we have

$$\begin{array}{ll} I(N(y), N(x)) & = I(N(y), I(x,0)) = I(x, I(N(y),0)) \\ & = I(x, N(N(y))) = I(x,y). \end{array}$$

■

It is known that in general, (EP) and (BP) with a continuous (even strong) fuzzy negation N do not imply (I1). For example, the function

$$I(x,y) = \begin{cases} 1-x & \text{if } y = 0 \\ y & \text{if } x = 1 \\ 0.5 & \text{otherwise} \end{cases} \quad (11)$$

satisfies (EP) and (BP) with $N(x) = 1 - x$ the classical strong negation, but it does not satisfy the nonincreasingness in the first variable, (I1) (see [13]). Consequently, I can not satisfy (WLI), by the previous proposition. So, we have just found a counterexample proving that, for functions $I : [0,1]^2 \to [0,1]$ in general, (EP) does not imply (WLI). In section 3.3 we will prove that a counterexample for fuzzy implications is also available.

3.2 (WLI) and (S,N)-implications

The main target of this section is the study of (WLI) on (S,N)-implications. We will show a new characterization of (S,N)-implications based on (WLI) and we will partially solve (OP2).

Proposition 2 *Let $I : [0,1]^2 \to [0,1]$ be a binary operator satisfying (WLI) with a commutative and nondecreasing function F and (BP) with a continuous (strict, strong) fuzzy negation N. Then I is an (S,N)-implication generated from a t-conorm S and the negation N.*

Proof: Due to proposition 1, it follows that I satisfies (I1) and (EP). So, by using theorem 1, I is an (S,N)-implication generated from some t-conorm S and some continuous (strict, strong) fuzzy negation N. ■

After that let us take a look to the verification of (WLI) on (S,N)-implications.

Proposition 3 *An (S,N)-implication I generated from a t-conorm S and a strict (strong) negation N satisfies (WLI) with a function $F : [0,1]^2 \to [0,1]$ if and only if F is the N-dual t-norm of S, that is*

$$F(x,y) = N^{-1}(S(N(x), N(y))) \text{ for all } x, y \in [0,1]. \quad (12)$$

Proof: Let I be an (S,N)-implication generated from a t-conorm S and a strict negation N. If I satisfies (WLI) with a function F, then for all $x, y, z \in [0,1]$ we have

$$\begin{array}{ll} I(F(x,y), z) & = S(N(F(x,y)), z) \\ I(x, I(y,z)) & = S(N(x), I(y,z)) = S(N(x), S(N(y), z)) \end{array}$$

Taking $z = 0$ in both expressions and using (WLI), we obtain $N(F(x,y)) = S(N(x), N(y))$, and consequently, $F(x,y) = N^{-1}(S(N(x), N(y)))$ for all $x, y \in [0,1]$. Conversely, if F is the N-dual t-norm of S, a straightforward computation proves that I satisfies (WLI) with F. ■

Remark 1 *Proposition 3 is the counterpart for (WLI) of an analogous result for (LI) in [8] and [13]. Note that none property of F is used in the proof. Thus, for (S,N)-implications with N strict or strong (LI) and (WLI) are equivalent and in fact, the N-dual t-norm of S is the only function $F : [0,1]^2 \to [0,1]$ (without any further assumption) for which (WLI) is satisfied.*

ISBN: 978-989-95079-6-8

We will see that the equivalence in the previous remark also holds for (S,N)-implications with N continuous (not necessarily strict), but something different occurs in this case. We will see that they also satisfy (WLI), but the function F for which they satisfy (WLI) needs not to be unique. However, among the functions F for which the (S,N)-implication satisfies (WLI), there is at least one of them that is a t-norm.

To do this, let us consider the following definition.

Definition 3 *(see [13]) Given any continuous fuzzy negation N, we consider the function $\mathfrak{R}_N : [0,1] \to [0,1]$ defined by*

$$\mathfrak{R}_N(x) = \begin{cases} N^{(-1)}(x) & \text{if } x \in (0,1] \\ 1 & \text{if } x = 0 \end{cases} \tag{13}$$

where $N^{(-1)}$ stands for the pseudo-inverse of N given by

$$N^{(-1)}(x) = \sup\{z \in [0,1] \mid N(z) > x\} \quad \text{for all } x \in [0,1]. \tag{14}$$

In [13], it is proved that $\mathfrak{R}_N$ is a strictly decreasing negation, $\mathfrak{R}_N^{(-1)} = N$, $N \circ \mathfrak{R}_N = id_{[0,1]}$ and

$$\mathfrak{R}_N \circ N|_{\text{Ran}(\mathfrak{R}_N)} = id|_{\text{Ran}(\mathfrak{R}_N)},$$

where $\text{Ran}(\mathfrak{R}_N)$ stands for the range of function $\mathfrak{R}_N$. This function $\mathfrak{R}_N$ plays the role of the "inverse" function of N and therefore, we have the following result:

Proposition 4 *An (S,N)-implication I generated from a t-conorm S and a continuous fuzzy negation N satisfies (WLI) with the function $F(x,y) = \mathfrak{R}_N(S(N(x),N(y)))$ for all $x,y \in [0,1]$.*

Proof: It is clear from its definition that F is commutative and nondecreasing. Now,

$$\begin{array}{ll} I(T(x,y),z) & = S(N(T(x,y)),z) \\ & = S(N(\mathfrak{R}_N(S(N(x),N(y)))),z) \\ (N \circ \mathfrak{R} = id_{[0,1]}) & = S(S(N(x),N(y)),z) \\ & = S(N(x),S(N(y),z)) \\ & = S(N(x),I(y,z)) = I(x,I(y,z)) \end{array}$$

■

Contrary to what happens in the case when N is strict (or strong), the function F in the proposition above is not the only one for which the (S,N)-implication satisfies (WLI). Note that the key fact in the proof of the previous proposition is that $N \circ \mathfrak{R}_N = id_{[0,1]}$ (in fact it is only necessary that $N(F(x,y)) = S(N(x),N(y))$ for all $x,y \in [0,1]$). When N is continuous but non-strict there are other functions than $\mathfrak{R}_N$ satisfying this condition, and consequently different functions F for which I satisfies the (WLI). See for instance the following example.

Example 1 *(see [13]) Consider the continuous fuzzy negation N given by*

$$N(x) = \begin{cases} -2x+1 & \text{if } x \in [0,0.25] \\ 0.5 & \text{if } x \in (0.25,0.75) \\ -2x+2 & \text{if } x \in [0.75,1]. \end{cases} \tag{15}$$

An easy calculation shows that

$$\mathfrak{R}_N(x) = \begin{cases} -0.5x+1 & \text{if } x \in [0,0.5) \\ -0.5x+0.5 & \text{if } x \in [0.5,1], \end{cases} \tag{16}$$

but taking N_1 given by

$$N_1(x) = \begin{cases} -0.5x+1 & \text{if } x \in [0,0.5] \\ -0.5x+0.5 & \text{if } x \in (0.5,1], \end{cases} \tag{17}$$

we also obtain $N \circ N_1 = id_{[0,1]}$. Consequently, given I any (S,N)-implication, derived from a t-conorm S and the negation N, defining $F_1(x,y) = N_1(S(N(x),N(y)))$ we have that I satisfies (WLI) with both functions, F (obtained from Proposition 4) and F_1.

Proposition 5 *Let S be a t-conorm and N a continuous fuzzy negation. Then the function $F(x,y) = \mathfrak{R}_N(S(N(x),N(y)))$ for all $x,y \in [0,1]$ is a t-subnorm, i.e., it is nondecreasing, commutative, associative and such that $F(x,y) \leq \min(x,y)$ for all $x,y \in [0,1]$.*

Proof: F is clearly nondecreasing and commutative. To see associativity, note that from the associativity of S we easily obtain

$$\begin{array}{ll} F(F(x,y),z) & = \mathfrak{R}_N(S(S(N(x),N(y)),N(z)) \\ & = F(x,F(y,z)). \end{array}$$

Finally, using the definition of $\mathfrak{R}_N$ we easily obtain

$$\mathfrak{R}_N(S(N(x),N(y))) \leq x \text{ for all } x \in [0,1]$$

and so $F(x,y) \leq \min(x,y)$ for all $x,y \in [0,1]$. ■

However, they are not t-norms in general because we have

$$F(1,y) = \mathfrak{R}_N(S(0,N(y))) = \mathfrak{R}_N(N(y))$$

and $\mathfrak{R}_N \circ N$ is the identity function only for values in $\text{Ran}(\mathfrak{R}_N)$. Nevertheless, one can always find a t-norm for which the (S,N)-implication satisfies (WLI) as follows.

Proposition 6 *Let I be a (S,N)-implication generated from a t-conorm S and a continuous fuzzy negation N. Then*

$$T(x,y) = \begin{cases} \mathfrak{R}_N(S(N(x),N(y))) & \text{if } \max\{x,y\} < 1 \\ \min\{x,y\} & \text{if } \max\{x,y\} = 1 \end{cases}$$

is a t-norm such that I satisfies (LI) with T.

Proof: T is clearly a t-norm since it is constructed in the usual way to obtain a t-norm from a given t-subnorm (see for instance, [12], Proposition 1.6). Thus, we only need to prove that I satisfies (LI) with T. From Proposition 4 it is sufficient to verify (LI) when $x = 1$ or $y = 1$. But, for $x = 1$ we have

$$I(T(1,y),z) = I(y,z) = I(1,I(y,z))$$

and similarly for $y = 1$. ■

Corollary 1 *Let I be an implication satisfying (BP) with a continuous fuzzy negation N. Then*

$$I \text{ satisfies (LI)} \quad \Leftrightarrow \quad I \text{ satisfies (WLI)}$$

Remark 2 *Note that the construction given in the Propositions 5 and 6 works only for the function $\mathfrak{R}_N$. For the other functions N_1 satisfying $N \circ N_1 = id_{[0,1]}$ the monotonicity of T is not guaranteed (it is enough to take $S = \max$, N and N_1 as in Example 1 and $x = 0.5 > y$, then $N_1(\max(N(0.5),N(y))) = 0.75 > 0.5$). Thus, for (S,N)-implications with N continuous we have found many functions F for which (WLI) holds, but only one t-norm for which (LI) holds. However, we are convinced that there can be found t-conorms S such that even the t-norm is not unique.*

ISBN: 978-989-95079-6-8

At this point, we propose a new characterization of (S,N)-implications based on (WLI).

Theorem 4 *For a function $I:[0,1]^2\to[0,1]$ the following statements are equivalent:*

(i). I is an (S,N)-implication generated from a t-conorm S and a continuous (strict, strong) fuzzy negation N

(ii). I satisfies (WLI) with a commutative and nondecreasing function F and (BP) with a continuous (strict, strong) fuzzy negation N.

In this case possible functions F are given by $F(x,y)=N_1(I(x,N(y)))$ for all $x,y\in[0,1]$, where N_1 is a fuzzy negation with $N\circ N_1=id_{[0,1]}$. Moreover, when N is strict (strong) then such F is unique with $N_1=N^{-1}$.

Proof: If I satisfies $i)$, then $I(x,0)=S(N(x),0)=N(x)$ and I satisfies (BP) with N. Moreover, it also satisfies (WLI) from Proposition 4. The converse is guaranteed by Proposition 2. ∎

Finally and returning to the question about the equivalence of (EP) and (LI), we have the following result.

Proposition 7 *Let I be an implication satisfying (BP) with a continuous fuzzy negation N. Then*

$$I \text{ satisfies (EP)} \Leftrightarrow I \text{ satisfies (LI)} \Leftrightarrow I \text{ satisfies (WLI)}$$

Proof: We already know that (LI) implies (EP). To see the converse, note that if I is an implication it satisfies in particular (I1) and then Theorem 1 implies that I is an (S,N)-implication generated from a t-conorm S and a continuous fuzzy negation N. Finally, Proposition 4 and Corollary 1 ensures that I satisfies (LI). The other equivalence has been already proved in Corollary 1. ∎

3.3 Counterexample for implications

To sum up, we have proved that (EP) and (LI) are equivalent for implications satisfying (BP) with a continuous fuzzy negation N. So, to find a counterexample we need to search among those implications I such that $I(x,0)=N(x)$ is non-continuous (it is always a fuzzy negation). In next proposition we give such a counterexample which proves not only that (EP) does not imply (LI), but also that it does not imply (WLI).

Proposition 8 *Let S be a nilpotent t-conorm and N a strict negation. Let $I:[0,1]^2\to[0,1]$ be the function given by*

$$I(x,y)=\begin{cases}0 & \text{if } y=0 \text{ and } x\neq 0\\ S(N(x),y) & \text{otherwise.}\end{cases} \tag{18}$$

Then I is a fuzzy implication that satisfies (EP), but there is no function F for which I satisfies (WLI).

Proof: We divide the proof in two steps.

- I satisfies (EP). When $z\neq 0$ we have $I(x,z)\neq 0$ for all $x\in[0,1]$ and then I satisfies (EP) because it is given as an (S,N)-implication.

  When $z=0$ and $x,y\neq 0$ then

  $$I(x,I(y,0))=I(y,I(x,0))=0.$$

  Finally, when $z=0$ and $\min\{x,y\}=0$ then

  $$I(x,I(y,0))=I(y,I(x,0))=1.$$

- I does not satisfy (WLI) with any F. Suppose on the contrary that I satisfies (WLI) with a function F. Then for all $x,y\in[0,1]$ and $z\neq 0$ (WLI) derives into

  $$S(N(F(x,y)),z)=S(S(N(x),N(y)),z). \tag{19}$$

  If we denote by N_S the strong negation associated to S (i.e., $N_S(x)=\varphi^{-1}(1-\varphi(x))$ where φ is the normalized additive generator of S), the left hand side of the previous equation equal 1 if and only if $N(F(x,y))\geq N_S(z)$, whereas the right hand side equals 1 if and only if $S(N(x),N(y))\geq N_S(z)$. Consequently for all $x,y\in[0,1]$ and $z\neq 0$

  $$N(F(x,y))\geq N_S(z)\Longleftrightarrow S(N(x),N(y))\geq N_S(z).$$

  Let us now prove from this equivalence that

  $$F(x,y)=N^{-1}(S(N(x),N(y)) \quad \text{for all } x,y\in[0,1]. \tag{20}$$

  Note that when $F(x,y)=0$, $N(F(x,y))\geq N_S(z)$ for all $z\in[0,1]$ and, from the equivalence before we obtain $S(N(x),N(y))=1$ and equation (20) holds. On the other hand, when $F(x,y)>0$ there exists $z\in[0,1]$ such that $N_S(z)>N(F(x,y))$ and then equation (19) can be written as

  $$\begin{aligned}&\varphi^{-1}(\varphi(N(F(x,y)))+\varphi(z))\\ &=\varphi^{-1}(\varphi(S(N(x),N(y)))+\varphi(z))\end{aligned}$$

  and consequently equation (20) also holds.

  Thus, F should be given by equation (20). However, I does not satisfy (WLI) with this F when $z=0$. To see it just take $x,y\neq 0$ such that $S(N(x),N(y))=1$ since then $F(x,y)=0$ and we obtain

  $$\begin{aligned}I(x,I(y,0)) &= I(x,0)=0\\ I(F(x,y),0) &= I(0,0)=1\end{aligned}$$

  ∎

3.4 R-implications derived from left-continuous t-norms

A characterization of those functions $I:[0,1]^2\to[0,1]$ that are R-implications derived from a left-continuous t-norm is given in Theorem 2, and the case of continuous R-implications is characterized in Theorem 3. Note that in this last case condition (I2) is not necessary to derive the conclusion. It was claimed in [17] that the same happens in the general case. Specifically it was claimed that (EP) and (OP) imply (I2). This is not true and a counterexample was recently proved in [18].

In this section, we will prove that the situation is completely different if we change (EP) by (WLI). We begin by proving that (I2) can be derived from (WLI) and (OP).

Proposition 9 *Let $I:[0,1]^2\to[0,1]$ be a function satisfying (OP) and (WLI) with a commutative and nondecreasing function F. Then I satisfies (I2), i.e, it is nondecreasing in the second variable.*

ISBN: 978-989-95079-6-8

Proof: Consider $x, y, z \in [0,1]$ with $y \leq z$ and let us divide our argument in two cases.

- If $x \leq z$, then $I(x,y) \leq I(x,z) = 1$.
- If $y \leq z < x$, then by (OP) and (WLI),

  $$1 = I(I(x,y), I(x,y)) = I(F(I(x,y),x), y)$$

  and consequently $F(I(x,y),x) \leq y \leq z$. Thus

  $$1 = I(F(I(x,y),x), z) = I(I(x,y), I(x,z))$$

  and therefore, $I(x,y) \leq I(x,z)$. ∎

Moreover, we have also the following properties.

Proposition 10 *Let $I : [0,1]^2 \to [0,1]$ be a function satisfying (OP) and (WLI) with a commutative and nondecreasing function F. Then*

i) I satisfies (I1), i.e, it is nonincreasing in the first variable.

ii) $I(x,-)$ is right-continuous for all $x \in [0,1]$.

Proof: To prove i) consider $x, y, z \in [0,1]$ with $x \leq y$. We divide our argument in two cases.

- If $x \leq z$, then $I(y,z) \leq I(x,z) = 1$.
- If $x > z$, then by (OP) and (WLI),

  $$1 = I(I(y,z), I(y,z)) = I(F(I(y,z),y), z)$$

  and consequently $F(I(y,z),y) \leq z$. Since F is nondecreasing we also have $F(I(y,z),x) \leq z$. Thus

  $$1 = I(F(I(y,z),x), z) = I(I(y,z), I(x,z))$$

  and therefore, $I(y,z) \leq I(x,z)$.

To prove ii), since I is nondecreasing in the second variable, it is enough to show that $\inf\{I(x,y_n)\} = I(x, \inf\{y_n\})$ for all $x \in [0,1]$. Say $y = \inf\{y_n\}$, it is clear that $I(x,y) \leq I(x,y_n)$ for all n and consequently $I(x,y) \leq \inf\{I(x,y_n)\}$. Moreover, for all n we have

$$1 = I(I(x,y_n), I(x,y_n)) = I(F(I(x,y_n),x), y_n)$$

which implies $F(I(x,y_n),x) \leq y_n$ for all n and thus $F(I(x,y_n),x) \leq y$. Finally, we have

$$1 = I(F(I(x,y_n),x), y)) = I(I(x,y_n), I(x,y)),$$

that is, $I(x,y_n) \leq I(x,y)$ and thus $\inf\{I(x,y_n)\} \leq I(x,y)$.∎

Due to the previous results we can reformulate the characterization of R-implications given by theorem 2.

Theorem 5 *Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is an R-implication derived from a left-continuous t-norm, if and only if, I satisfies (OP) and (WLI) with a commutative and nondecreasing function F.*

Moreover, in this case F must be the t-norm from which I is obtained by residuation.

Finally, it is well know that if a function $I : [0,1]^2 \to [0,1]$ satisfies (OP), (EP) and (BP) with a continuous fuzzy negation N, then N must be strong. However, note that from the previous results more properties are derived.

Corollary 2 *Let $I : [0,1]^2 \to [0,1]$ satisfy (OP), (EP) and (BP) with a continuous fuzzy negation N. Then I satisfies (I2), I is right-continuous in the second variable and so, I is an R-implication derived from a left-continuous t-norm.*

Acknowledgment

The authors want to thank the referees for their valuable comments. This paper has been partially supported by the Spanish grant MTM2006-05540 (with FEDER support), and the Government of the Balearic Islands grant PCTIB-2005GC1-07.

References

[1] E.E. Kerre, C. Huang, and D. Ruan. *Fuzzy Set Theory and Approximate Reasoning*. Wu Han University Press, Wu Chang, 2004.

[2] S. Gottwald. *A Treatise on Many-Valued Logic*. Research Studies Press, Baldock, 2001.

[3] M. Mas, M. Monserrat, J. Torrens, and E. Trillas. A survey on fuzzy implication functions. *IEEE Transactions on Fuzzy Systems*, 15(6):1107–1121, 2007.

[4] H. Bustince, V. Mohedano, E. Barrenechea, and M. Pagola. Definition and construction of fuzzy DI-subsethood measures. *Information Sciences*, 176:3190–3231, 2006.

[5] H. Bustince, M. Pagola, and E. Barrenechea. Construction of fuzzy indices from fuzzy DI-subsethood measures: application to the global comparison of images. *Information Sciences*, 177:906–929, 2007.

[6] P. Yan and G. Chen. Discovering a cover set of ARsi with hierarchy from quantitative databases. *Information Sciences*, 173:319–336, 2005.

[7] M. Baczyński and B. Jayaram. *Fuzzy Implications*. Springer, Berlin Heidelberg, 2008.

[8] B. Jayaram. On the law of importation $(x \wedge y) \to z \equiv (x \to (y \to z))$ in fuzzy logic. *IEEE Transactions on Fuzzy Systems*, 16:130–144, 2008.

[9] M. Mas, M. Monserrat, and J. Torrens. The law of importation for QL and D-implications. *Submitted*, 2008.

[10] M. Mas, M. Monserrat, and J. Torrens. On the law of importation for some kinds of fuzzy implications derived from uninorms. *In Proceedings of IPMU'08*, Torremolinos (Málaga):1303–1310, 2008.

[11] E.P. Klement and R. Mesiar. Open problems posed at the eighth international conference in fuzzy set theory and applications (FSTA 2006, Liptovský Ján, Slovakia). *Kybernetika*, 42:225–235, 2006.

[12] E.P. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Publishers, Dordrecht, 2000.

[13] M. Baczyński and B. Jayaram. On the characterization of (S,N)-implications. *Fuzzy Sets and Systems*, 158:1713–1727, 2007.

[14] J.C. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers, Dordrecht, 1994.

[15] M. Baczyński and B. Jayaram. (S,N)- and R-implications: A state-of-the-art survey. *Fuzzy Sets and Systems*, 159:1836–1859, 2008.

[16] P. Smets and P. Magrez. Implications in fuzzy logic. *International Journal of Approximate Reasoning*, 1:327–347, 1987.

[17] M. Baczyński. Residual implications revisited. notes on the smets-magrez theorem. *Fuzzy Sets and Systems*, 145:267–277, 2004.

[18] R. Smogura. A note on the article: Residual implications revisited. Notes on the Smets-Magrez theorem. *Fuzzy Sets and Systems*, 157:466–469, 2006.

ISBN: 978-989-95079-6-8

Fuzzy Transform and Smooth Functions

Luciano Stefanini
DEMQ, Department of Economics and Quantitative Methods
University of Urbino "Carlo Bo", ITALY
Email: lucste@uniurb.it

Abstract— *We describe parametric fuzzy partitions based on parametric shape functions and we show that the Perfilieva's Fuzzy Transform defined on parametric fuzzy partitions has the properties of a smoother. We also illustrate some examples.*

Keywords— Parametric Fuzzy Partition, Fuzzy Transform, Smoothing Functions, Fuzzy Numbers

1 Introduction

The fuzzy transform has been recently introduced by I. Perfilieva and the theoretical investigations and several connections between theory and applications are covered e.g. in [3], [4], [5] and [9].

In a recent paper (see [6]), we suggested the use of a parametrized class of fuzzy numbers (see [7], [8]) to define general parametric partitions for the fuzzy transform. The proposed flexible family of fuzzy partitions allows the possibility of estimating the shapes of the fuzzy partition to obtain better approximation properties.

In this paper we suggest an extension of the fuzzy partitions (section 3) and we describe the corresponding direct and inverse fuzzy transforms (section 4). Then, in section 5, we show that the inverse fuzzy transform based on extended partitions has an important smoothing property and we illustrate it by numerical examples.

2 Fuzzy Transform and Parametric Fuzzy Partitions

We briefly recall the basic definitions and properties

Definition: (Perfilieva[3]) A *fuzzy partition* of a given real compact interval $[a,b]$ is constructed by a decomposition $\mathbb{P} = \{a = x_1 < x_2 < ... < x_n = b\}$ of $[a,b]$ into $n-1$ subintervals $[x_{k-1}, x_k]$, $k = 2, ..., n$ and by a family $\mathbb{A} = \{A_1, A_2, ..., A_n\}$ of n fuzzy numbers (the *basic functions*), identified by their membership functions $A_1(x), A_2(x), ..., A_n(x)$ for $x \in [a,b]$ and with the properties (to complete this notation, we set $x_0 = a$ and $x_{n+1} = b$):

1. $A_k : [a,b] \longrightarrow [0,1]$ is continuous with $A_k(x_k) = 1$, $A_k(x) = 0$ for $x \notin [x_{k-1}, x_{k+1}]$;
2. for $k = 2, 3, ..., n$, A_k is increasing on $[x_{k-1}, x_k]$ and decreasing on $[x_k, x_{k+1}]$; A_1 is decreasing on $[a, x_2]$; A_n is increasing on $[x_{n-1}, b]$;
3. for all $x \in [a,b]$ the following partition of unity condition holds

$$\sum_{k=1}^{n} A_k(x) = 1.$$

We denote a fuzzy partition by the pair $(\mathbb{P}, \mathbb{A})$. The fuzzy numbers $A_1, A_2, ..., A_n$ giving the family of basic functions can be defined by $n-1$ increasing functions $L_2(x), ..., L_n(x)$ as follows

$$A_1(x) = 1 - L_2(x) \text{ if } x \in [a, x_2],$$

$$A_k(x) = \begin{cases} L_k(x) & \text{if} \quad x \in [x_{k-1}, x_k] \\ 1 - L_{k+1}(x) & \text{if} \quad x \in [x_k, x_{k+1}] \end{cases} \quad \text{for } k = 2, ..., n-1 \tag{1}$$

$$A_n(x) = L_n(x) \text{ if } x \in [x_{n-1}, b]$$

where $L_k(x_{k-1}) = 0$ and $L_k(x_k) = 1$.

In the standard case, the support of each basic function $A_k(x)$ is the compact interval $[x_{k-1}, x_{k+1}]$ so that, on each subinterval $[x_{k-1}, x_k]$ of the decomposition $\mathbb{P}$ only two basic functions $A_{k-1}(x)$ and $A_k(x)$ are non zero for $k = 2, ..., n$.

Definition 1: (Perfilieva[3]) Given a continuous function $f : [a,b] \longrightarrow R$ and a fuzzy partition (P, A), the direct fuzzy transform (F-transform) of f with respect to (P, A) is the n-tuple of real numbers $F = (F_1, F_2, ..., F_n)^T$ given by (notation $(\cdot)^T$ means transposition)

$$\text{For } k = 1, 2, ..., n \tag{2}$$

$$F_k = \frac{\int_a^b f(x)A_k(x)dx}{\int_a^b A_k(x)dx} = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x)A_k(x)dx}{\int_{x_{k-1}}^{x_{k+1}} A_k(x)dx}$$

Definition 2: (Perfilieva[3]) Given a continuous function $f : [a,b] \longrightarrow R$, a fuzzy partition (P, A) and the direct fuzzy transform $(F_1, F_2, ..., F_n)^T$ of f with respect to (P, A), the inverse F-transform is the continuous function $\widehat{f}_F : [a,b] \longrightarrow R$ given by

$$\widehat{f}_F(x) = \sum_{k=1}^{n} F_k A_k(x) \text{ for } x \in [a,b]. \tag{3}$$

The following property (Th. 2 in Perfilieva[3]) is one of the fundamentals of the F-transform setting.

Property 1: If $f : [a,b] \longrightarrow R$ is a continuous function then, for any positive real ε, there exists a fuzzy partition $(P_\varepsilon, A_\varepsilon)$ such that the corresponding F-transform $F_\varepsilon = (F_{1,\varepsilon}, F_{2,\varepsilon}, ..., F_{n_\varepsilon,\varepsilon})^T$ and inverse fuzzy transform $\widehat{f}_{F_\varepsilon} : [a,b] \longrightarrow R$ satisfy

$$\left| f(x) - \widehat{f}_{F_\varepsilon}(x) \right| < \varepsilon \text{ for all } x \in [a,b].$$

To define a fuzzy partition $(\mathbb{P}, \mathbb{A})$ we can use a general decomposition $\mathbb{P} = \{a = x_1 < x_2 < ... < x_n = b\}$ of $[a,b]$ into $n-1$ subintervals $[x_{k-1}, x_k]$, and a general set of basic functions $\mathbb{A} = \{A_1, A_2, ..., A_n\}$ of n fuzzy numbers

ISBN: 978-989-95079-6-8

$A_1, A_2, ..., A_n$ on $[a, b]$; denote, in general, $h_k = x_k - x_{k-1}$, $k = 2, ..., n$.

If the decomposition $\mathbb{P}$ is uniform, we have $x_k = a + (k-1)h$ where $h = \frac{b-a}{n-1}$.

We can have uniform $\mathbb{P}$ and general $\mathbb{A}$.

To obtain families of basic functions $A_1, A_2, ..., A_n$ in (1), defined by $n-1$ increasing functions $L_2, ..., L_n$, we suggested to use parametric monotonic functions $p(t; \beta_0, \beta_1)$, $t \in [0, 1]$ such that, for any parameters $\beta_0, \beta_1 \geq 0$, the following Hermite-type interpolation conditions are satisfied:

$$\begin{aligned} p(0; \beta_0, \beta_1) &= 0,\ p(1; \beta_0, \beta_1) = 1, \\ p'(0; \beta_0, \beta_1) &= \beta_0,\ p'(1; \beta_0, \beta_1) = \beta_1 \end{aligned}$$

and $p'(t; \beta_0, \beta_1) \geq 0\ \forall t \in [0, 1]$ (derivatives at $t = 0$ and $t = 1$ are, respectively, right and left derivatives).

Valid examples of function $p(t; \beta_0, \beta_1)$ are reported in see [7]; to exemplify, we will use here the (2,2)-rational spline:

$$p_{rat22}(t; \beta_0, \beta_1) = \frac{t^2 + \beta_0 t(1-t)}{1 + (\beta_0 + \beta_1 - 2)t(1-t)}. \quad (4)$$

By using parametric monotonic functions $p(t; \beta_0, \beta_1)$, $t \in [0, 1]$ we obtain increasing functions $L_2, ..., L_n$ that define the basic membership functions of the fuzzy partition $A_1, A_2, ..., A_n$:

$$\begin{aligned} L_k(x) &= p(\frac{x - x_{k-1}}{x_k - x_{k-1}}; \beta_{k,0}, \beta_{k,1}) \quad (5) \\ \text{for } x &\in [x_{k-1}, x_k] \text{ and } k = 2, ..., n \end{aligned}$$

and each fuzzy partition $(\mathbb{P}, \mathbb{A})$ contains $2(n-1)$ parameters $\beta_{k,0}, \beta_{k,1}$, $k = 2, ..., n$ giving the first derivatives $L'_k(x_{k-1}) = \beta_{k,0}/(x_k - x_{k-1})$ and $L'_k(x_k) = \beta_{k,1}/(x_k - x_{k-1})$ (to be intended as right and left derivatives, respectively).

We define a *uniform fuzzy partition* $(\mathbb{P}, \mathbb{A})$ if $\mathbb{P}$ is uniform and $\beta_{k,0} = \beta_{j,0}, \beta_{k,1} = \beta_{j,1}$ for all j, k.

We define a *balanced fuzzy partition* $(\mathbb{P}, \mathbb{A})$ if $\mathbb{P}$ is uniform and $\beta_{k,0} = \beta_{k+r,0}, \beta_{k,1} = \beta_{k+r,1}$ for all k..

We obtain a uniform and balanced partition $(\mathbb{P}, \mathbb{A})$ by requiring that $\beta_{k,0} = \beta_{k,1} = \beta_{j,0} = \beta_{j,1}$ for all j, k.

3 Extended Fuzzy Partitions

On the interval $[a, b]$ with a decomposition $\mathbb{P} = \{a = x_1 < x_2 < ... < x_n = b\}$ into $n-1$ subintervals $[x_{k-1}, x_k]$, $k = 2, ..., n$ the basic functions of the fuzzy partition can be defined such that the support of each A_k is more extended than two intervals, identified by three points x_{k-1}, x_k, x_{k+1}. Consider an integer $r \geq 1$ and $2r+1$ consecutive points of $\mathbb{P}$, $x_{k-r}, ..., x_k, ..., x_{k+r}$ for all $k = 1, 2, ..., n$ (to complete the notation, we extend the points to $x_{1-r} < ... < x_0 < a$ and $b < x_{n+1} < ... < x_{n+r}$).

Definition 3: we define a *fuzzy r-partition* of $[a, b]$ to be $(\mathbb{P}, \mathbb{A}^{(r)})$ with the family of $n + 2r - 2$ continuous, normal, convex fuzzy numbers

$$\begin{aligned} \mathbb{A}^{(r)} &= \{A_k^{(r)} : [a, b] \longrightarrow [0, 1] | \\ k &= -r+2, ..., 1, 2, ..., n, ..., n+r-1\} \end{aligned}$$

where $A_k^{(r)}$ has core $\{x_k\}$ and support $[x_{k-r}, x_{k+r}]$ i.e.

a. for $k = 1, 2, ..., n$, $A_k^{(r)}$ is a continuous fuzzy number with $A_k^{(r)}(x_k) = 1$ (i.e. the core is $\{x_k\}$) and $A_k^{(r)}(x) = 0$ for $x \notin [x_{k-r}, x_{k+r}]$ (i.e. the support is $[x_{k-r}, x_{k+r}]$);

b. for $k = 1, 2, ..., n$, $A_k^{(r)}$ is increasing on $[x_{k-r}, x_k]$ and decreasing on $[x_k, x_{k+r}]$;

c. for $k = -r+2, ..., 0$, $A_k^{(r)}$ is decreasing on $[x_k, x_{k+r}]$;

d. for $k = n+1, ...n+r-1$, $A_k^{(r)}$ is increasing on $[x_{k-r}, x_k]$;

e. for all $x \in [a, b]$ the following *$r-$partition of unity* condition holds

$$\frac{1}{r} \sum_{k=-r+2}^{n+r-1} A_k^{(r)}(x) = 1.$$

If $r = 1$ we have the standard partition of unity.

We can construct parametric *fuzzy r-partition* of $[a, b]$ by considering $n + r - 2$ shape functions $p(t; \beta_0, \beta_1)$ of types (4); then by defining

$$\begin{aligned} \text{for } k &= 2, ..., n+r-1 \\ L_k^{(r)}(x) &= p(\frac{x - x_{k-r}}{x_k - x_{k-r}}; \beta_{k,0}, \beta_{k,1}) \end{aligned}$$

the basic functions are

$$\begin{aligned} \text{for } k &= 2, ..., n-1 \\ A_k^{(r)}(x) &= \begin{cases} L_k^{(r)}(x) & \text{if } x \in [x_{k-r}, x_k] \\ 1 - L_{k+r}^{(r)}(x) & \text{if } x \in [x_k, x_{k+r}] \\ 0 & \text{otherwise} \end{cases} \quad (6) \end{aligned}$$

$$\begin{aligned} \text{for } k &= -r+2, ..., 1 \\ A_k^{(r)}(x) &= \begin{cases} 1 - L_{k+r}^{(r)}(x) & \text{if } x \in [x_k, x_{k+r}] \\ 0 & \text{otherwise} \end{cases} \quad (7) \end{aligned}$$

$$\begin{aligned} \text{for } k &= n, ..., n+r-1 \\ A_k^{(r)}(x) &= \begin{cases} L_{k-r}^{(r)}(x) & \text{if } x \in [x_{k-r}, x_k] \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

It is a matter of simple algebraic manipulations to show that for all $x \in [a, b]$ the *$r-$partition of unity* condition holds; in fact for $k = 1, 2, ..., n$ and $x \in [x_{k-1}, x_k]$, only $A_{k-r}^{(r)}, ..., A_k^{(r)}, ..., A_{k+r-1}^{(r)}$ are different from zero and

$$\begin{aligned} \sum_{j=k-r}^{k+r-1} A_j^{(r)}(x) &= \sum_{j=k-r}^{k-1} (1 - L_{j+r}^{(r)}(x)) + \sum_{j=k}^{k+r-1} L_j^{(r)}(x) \\ &= \sum_{j=k}^{k+r-1} (1 - L_j^{(r)}(x)) + \sum_{j=k}^{k+r-1} L_j^{(r)}(x) \\ &= r. \end{aligned}$$

Consider for example $r = 2$; we need $n + r - 2 = n$ shape functions and (we remove the superscript $^{(r)}$ for simplicity) define $L_k(x) = p(\frac{x - x_{k-2}}{x_k - x_{k-2}}; \beta_{k,0}, \beta_{k,1})$ for $k = 2, ..., n+1$. So (in the relevant subintervals, otherwise the functions are null)

$$\begin{aligned} \text{for } k &= 2, ..., n-1 \\ A_k(x) &= \begin{cases} L_k(x) & \text{if } x \in [x_{k-2}, x_k] \\ 1 - L_{k+2}(x) & \text{if } x \in [x_k, x_{k+2}] \end{cases} \end{aligned}$$

ISBN: 978-989-95079-6-8

$$\begin{aligned} \text{for } k &= 0, 1, n, n+1 \\ A_0(x) &= 1 - L_2(x) \text{ if } x \in [x_0, x_2] \\ A_1(x) &= 1 - L_3(x) \text{ if } x \in [x_1, x_3] \\ A_n(x) &= L_n(x) \text{ if } x \in [x_{n-2}, x_n] \\ A_{n+1}(x) &= L_{n+1}(x) \text{ if } x \in [x_{n-1}, x_{n+1}]. \end{aligned}$$

The figures below illustrate three *fuzzy r-partition* $(\mathbb{P}, \mathbb{A}^{(r)})$ for $r = 1$ (standard case), $r = 2$ (each $A_k(x)$ covers four intervals) and $r = 3$ (each $A_k(x)$ covers six intervals).

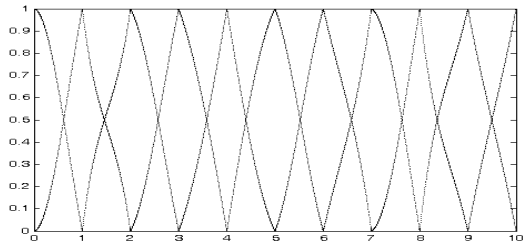


Figure 1a) *fuzzy r-partition* $(\mathbb{P},\mathbb{A}^{(r)})$ with $r = 1$

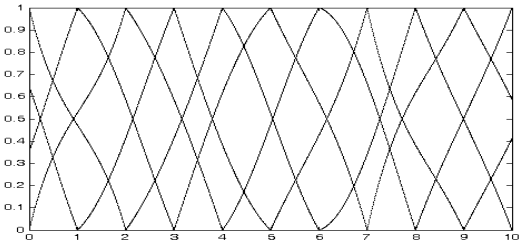


Figure 1b) *fuzzy r-partition* $(\mathbb{P},\mathbb{A}^{(r)})$ with $r = 2$

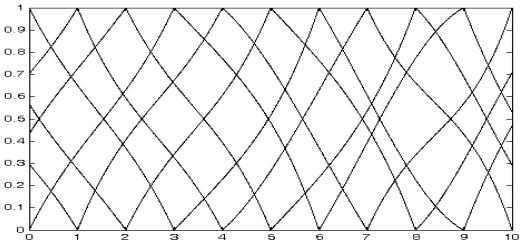


Figure 1c) *fuzzy r-partition* $(\mathbb{P},\mathbb{A}^{(r)})$ with $r = 3$

4 Fuzzy Transform with Extended Partitions

The direct F$^{(r)}$-transform (of degree r) based on the given *fuzzy r-partition* $(\mathbb{P}, \mathbb{A}^{(r)})$ can be defined by the vector $F^{(r)} = (F_1^{(r)}, F_2^{(r)}, ..., F_n^{(r)})^T$

$$F_k^{(r)} = \frac{\int_a^b f(x)A_k^{(r)}(x)dx}{\int_a^b A_k^{(r)}(x)dx}, k = 1, 2, ..., n \tag{8}$$

and the inverse F$^{(r)}$-transform by

$$\widehat{f}^{(r)}(x) = \frac{1}{r}\sum_{k=1}^{n} F_k^{(r)} A_k^{(r)}(x). \tag{9}$$

We see that each $\widehat{f}^{(r)}(x_k)$ has the structure of a moving average of the values $\{F_j^{(r)}, j = 1, ..., n\}$; in fact, at the points $x = x_k$ we have (assume $F_k^{(r)} = 0$ if $k < 1$ or $k > n$)

$$\widehat{f}^{(r)}(x_k) = \frac{1}{r}\sum_{j=k-r}^{k+r} F_j^{(r)} A_j^{(r)}(x_k) \tag{10}$$

i.e. a weighted average of $F_{k-r}^{(r)}, ..., F_k^{(r)}, ..., F_{k+r}^{(r)}$ with weights $\frac{A_{k-r}^{(r)}(x_k)}{r}, ..., \frac{1}{r}, ..., \frac{A_{k+r}^{(r)}(x_k)}{r}$, respectively.

It is possible to see that the F$^{(r)}$-transform has analogues properties to theorem 1 and theorem 2 of Perfilieva ([3]) i.e.

1. the local (for any $k = 1, 2, ..., n$) "error" function $\frac{1}{r}\int_a^b [f(x) - y]^2 A_k^{(r)}(x)dx$ is minimized by $y = F_k^{(r)}$;

2. the minimizer of $\frac{1}{r}\sum_{k=1}^{n}\int_a^b [f(x) - y_k]^2 A_k^{(r)}(x)dx$ is $y = F^{(r)} = (F_1^{(r)}, F_2^{(r)}, ..., F_n^{(r)})^T$ (notation $(\cdot)^T$ means transposition);

3. for fixed $r \geq 1$ and for any $\varepsilon > 0$ there exists a *fuzzy r-partition* $(\mathbb{P}_\varepsilon, \mathbb{A}_\varepsilon^{(r)})$ with the corresponding inverse F$^{(r)}$-transform $\widehat{f}_\varepsilon^{(r)}$ such that $\sup_x \left| f(x) - \widehat{f}_\varepsilon^{(r)}(x) \right| < \varepsilon$.

For a given family of parametric basic functions $A_k^{(r)}$ consider the linear combination (with $\theta = (\theta_1, \theta_2, ..., \theta_n)^T$)

$$C_{\mathbb{A}^{(r)}}(x; \theta, \beta) = \frac{1}{r}\sum_{j=1}^{n} \theta_j A_j^{(r)}(x; \beta) \text{ for } x \in [a, b] \tag{11}$$

where β is the vector of the non negative slope parameters $\beta_{j,0}$, $\beta_{j,1}$ giving $L_j(x; \beta_{j,0}, \beta_{j,1})$.

The following minimization defines $C_{\mathbb{A}^{(r)}}(x; \widehat{\theta}, \widehat{\beta})$ as the *least squares approximation* of $f(x)$ over interval $[a, b]$

$$\underset{\theta,\beta}{Minimize} \quad \int_a^b [f(x) - C_{\mathbb{A}^{(r)}}(x; \theta, \beta)]^2 dx \tag{12}$$

where $(\widehat{\theta}, \widehat{\beta})$ is the minimizing point.

For fixed β, the linear parameters $\widehat{\theta}$ are obtained by putting to zero the partial derivatives (for $i = 1, 2, ..., n$) and we obtain the equations

$$\int_a^b A_i^{(r)}(x; \beta) f(x)dx = \frac{1}{r}\sum_{j=1}^{n} \widehat{\theta}_j \int_a^b A_i^{(r)}(x; \beta) A_j^{(r)}(x; \beta)dx \tag{13}$$

From definition of the F$^{(r)}$-transform $F^{(r)}(\beta)$ as a function of β, we have

$$\int_a^b A_i^{(r)}(x; \beta) f(x)dx = F_i^{(r)}(\beta) \int_a^b A_i^{(r)}(x; \beta)dx \tag{14}$$

and we finally obtain the linear system of n equations

$$\begin{aligned} C^{(r)}(\beta)\widehat{\theta} &= rF^{(r)}(\beta) \quad i.e. \\ \sum_{j=1}^{n} c_{i,j}^{(r)}(\beta)\widehat{\theta}_j &= rF_i^{(r)}(\beta), \quad i = 1, 2, ..., n \end{aligned} \tag{15}$$

where the matrix $C^{(r)}(\beta)$ has elements

$$c_{i,j}^{(r)}(\beta) = \frac{1}{r}\int_a^b A_i^{(r)}(x; \beta) A_j^{(r)}(x; \beta)dx. \tag{16}$$

For fixed parameters β, we can obtain a matrix form for θ in the definition of function (11).

Given $f : [a, b] \longrightarrow \mathbb{R}$ and a fuzzy r-partition $(\mathbb{P}, \mathbb{A}^{(r)}(\beta))$ with the associated direct F$^{(r)}$-transform $F^{(r)}(\beta) = (F_1^{(r)}, F_2^{(r)}, ..., F_n^{(r)})^T$ and inverse F$^{(r)}$-transform

$$\widehat{f}_F^{(r)}(x; \beta) = \frac{1}{r}\sum_{j=1}^{n} F_j^{(r)}(\beta) A_j^{(r)}(x; \beta). \tag{17}$$

ISBN: 978-989-95079-6-8

Define the following vectors (we omit reference to β and r for simplicity of notation), obtained by evaluating $f(x)$ and $\widehat{f}_F(x)$ at points $x_1, x_2, ..., x_n$

$$f = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \\ f(x_n) \end{bmatrix}, \quad \widehat{f} = \begin{bmatrix} \widehat{f}_F(x_1) \\ \widehat{f}_F(x_2) \\ \\ \widehat{f}_F(x_n) \end{bmatrix}, \quad F = \begin{bmatrix} F_1^{(r)} \\ F_2^{(r)} \\ \\ F_n^{(r)} \end{bmatrix} \tag{18}$$

and the following $n \times n$ matrix

$$\begin{aligned} A^{(r)} &= [a_{i,j}] \text{ with} \\ a_{i,j} &= A_j^{(r)}(x_i). \end{aligned} \tag{19}$$

Matrix $A^{(r)}$ is a $(2r-1)$-band matrix and its $i-th$ row has the form ($i = 1, 2, ..., n$)

$$a_{i,j} = \begin{cases} 0 & j = 1, ..., i-r \\ A_j^{(r)}(x_i) & j = i-r+1, ..., i-1 \\ 1 & j = i \\ A_j^{(r)}(x_i) & j = i+1, ..., i+r-1 \\ 0 & j = i+r, ..., n \end{cases} \tag{20}$$

It is immediate that

$$\widehat{f} = \frac{1}{r}(A^{(r)})^T F \tag{21}$$

and $\frac{1}{r}(A^{(r)})^T$ represents a *moving average operator* acting on F to produce $\widehat{f}$.

4.1 Discrete $F^{(r)}$-Transform

Matrix A in (19) can be defined also for the discrete direct and inverse F$^{(r)}$-transforms. For m points $(t_i, y_i), i = 1, 2, ..., m$, where $y_i = f(t_i)$ and defining $A^{(r)}(\beta) = [a_{i,j}(\beta) = A_j^{(r)}(t_i, \beta)]$ for the family $\{A_j^{(r)}(t; \beta) | j = 1, 2, ..., n\}$ of basic functions, the discrete F$^{(r)}$-transform is

$$\begin{aligned} \text{For } j &= 1, 2, ..., n \\ F_j^{(r)}(\beta) &= \frac{g_j(\beta)}{s_j(\beta)} \text{ where} \\ g_j(\beta) &= \sum_{i=1}^{m} y_i A_j^{(r)}(t_i; \beta) \text{ and } s_j = \sum_{i=1}^{m} A_j^{(r)}(t_i; \beta) > 0 \end{aligned}$$

If the elements s_j are organized into a diagonal $n-$matrix

$$S = diag(s_1, s_2, ..., s_n)$$

so that

$$S^{-1} = diag(\frac{1}{s_1}, \frac{1}{s_2}, ..., \frac{1}{s_n})$$

and denote

$$\begin{aligned} y &= (y_1, ..., y_m)^T, \\ F^{(r)}(\beta) &= (F_1^{(r)}(\beta), ..., F_n^{(r)}(\beta))^T, \\ g(\beta) &= (g_1(\beta), ..., g_n(\beta))^T, \end{aligned}$$

then we have $(A^{(r)}(\beta))^T y = g(\beta)$ so that

$$F^{(r)}(\beta) = S(\beta)^{-1}(A^{(r)}(\beta))^T y.$$

Denote also, for $i = 1, 2, ..., m$ (note that for $i = r, ..., n-r+1$ we have $d_i = r$),

$$\begin{aligned} d_i(\beta) &= \sum_{j=1}^{n} A_j(t_i; \beta) > 0 \\ D &= diag(d_1, d_2, ..., d_m). \end{aligned}$$

For the discrete inverse F$^{(r)}$-transform

$$f_{F,\beta}^{(r)}(x) = \frac{\sum_{j=1}^{n} F_j^{(r)} A_j^{(r)}(x; \beta)}{\sum_{j=1}^{n} A_j^{(r)}(x; \beta)}$$

and denoting $f_{F,\beta}^{(r)} = (f_{F,\beta}^{(r)}(t_1), f_{F,\beta}^{(r)}(t_2), ..., f_{F,\beta}^{(r)}(t_m))^T$, we get

$$\begin{aligned} f_{F,\beta}^{(r)} &= D(\beta)^{-1} A^{(r)}(\beta) F^{(r)} \\ &= D(\beta)^{-1} A^{(r)}(\beta) S(\beta)^{-1} (A^{(r)}(\beta))^T y. \end{aligned} \tag{22}$$

5 Smoothing functions from $F^{(r)}$-Transform

In this section we investigate the smoothing effect of the inverse $F^{(r)}$-Transform $f^{(r)}$ on a finite set of data points (t_i, y_i), $i = 1, 2, ..., m$; the discrete direct and inverse F$^{(r)}$-transform are obtained with a partition $(\mathbb{P}, \mathbb{A}^{(r)}(\beta))$, where

$$\begin{aligned} \mathbb{P} &= \{a = x_1 < x_2 < ... < x_n = b\} \text{ and} \\ a &= \min\{t_i\}, \ b = \max\{t_i\} \\ x_i &= a + (i-1)/(n-1), i = 1, 2, ..., n. \end{aligned}$$

We first consider the parameters β as fixed for the family of basic functions $\mathbb{A}^{(r)}(\beta)$.

For a given function $f : [a, b] \longrightarrow \mathbb{R}$ a measure of smoothness can be given by its total variation

$$V(f) = \sup\{\sum_{j=1}^{k-1} |f(\alpha_{j+1}) - f(\alpha_j)|; \forall k \geq 2, \forall \alpha_j \in [a, b]\}$$

or by the following *total* variation (see [1], [2] and the references therein)

$$\widetilde{V}(f) = \sum_{j=1}^{m-1} |f(t_{j+1}) - f(t_j)|. \tag{23}$$

By increasing the integer value of r, the total variation $\widetilde{V}(f^{(r)})$ of the inverse transform functions $f^{(r)}$ tend to decrease to zero, as illustrated by the graphical representation in Fig. 2. we can obtain inverse fuzzy transform functions $f^{(r)}$ with arbitrarily small $\widetilde{V}(f^{(r)})$. A graphical representation is in Fig. 2.

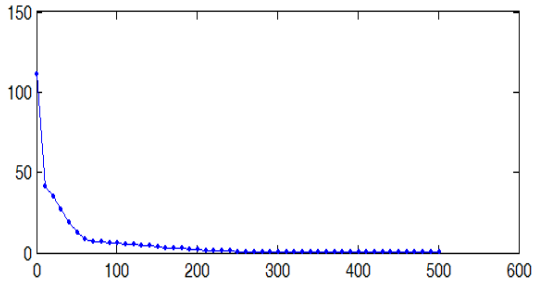


Figure 2. *Total variation $\widetilde{V}(f^{(r)})$ of inverse fuzzy transform function for increasing r.*

ISBN: 978-989-95079-6-8

It is interesting to see the combined effect of changing n and r; in figures 3a)-3c), $m = 101$ data points are generated from function $f(t) = 1+0.2t+sin(t).cos(t)$ by adding a Gaussian random noise with zero mean and standard deviation $\sigma = 2$, i.e. $y_i = f(t_i) + z_i$ and $z_i \in N(0, \sigma^2)$ (the t_i are uniform on $[-2\pi, 2\pi]$). In all the figures of this section, the continuos line is the original data, the dotted line is the inverse fuzzy transform $f^{(r)}$ and the circles are the direct $F^{(r)}$-transform.

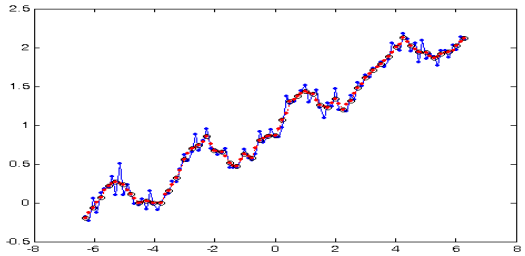


Figure 3a) Smoothing of m=101 data points, n=51, r=1.

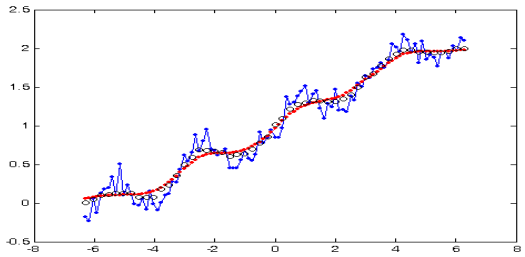


Figure 3b) Smoothing of m=101 data points, n=51, r=5.

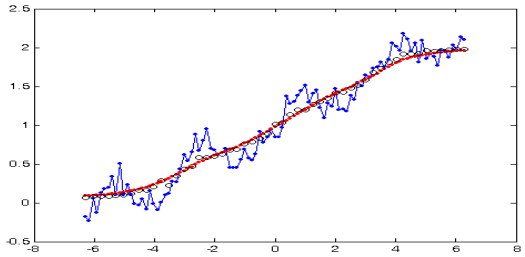


Figure 3c) Smoothing of m=101 data points, n=51, r=9.

In figures 4a)-4c) the data are generated from the same model but $n = 21$ is smaller then in previous case. When n is reduced, the smoothing effect is increased but it is more visible at a global scale: compare figures 3a) with 4a), 3b) with 4b) and 3c) with 4c).

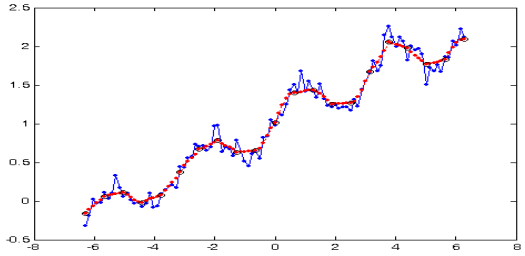


Figure 4a) Smoothing of m=101 data points, n=21, r=1.

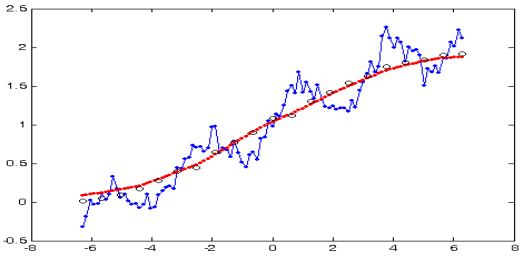


Figure 4b) Smoothing of m=101 data points, n=21, r=5.

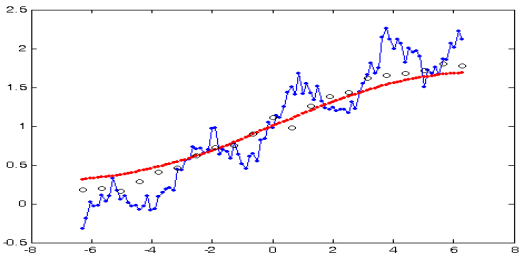


Figure 4c) Smoothing of m=101 data points, n=21, r=9.

If n is small with respect to m as in figures 5a) and 5b), the global smoothing is increased further.

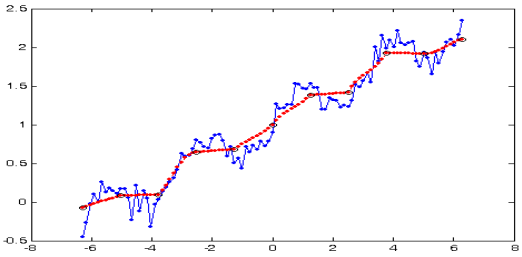


Figure 5a) Smoothing of m=101 data points, n=11, r=1.

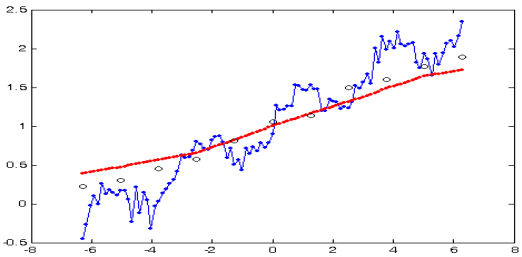


Figure 5b) Smoothing of m=101 data points, n=11, r=5.

So, according to the general properties of the $F^{(r)}$-transform, the discussion above illustrates that, in general, there are two possible sources for smoothness: for fixed m and r, reducing n will increase the smoothing effect; for fixed m and n, increasing r will increase the smoothness.

5.1 A Smoothing Procedure

Let (t_i, y_i), $i = 1, 2, ..., m$ be given observations and define $T_1 = \{i | x_1 \leq t_i \leq x_2\}$, $T_k = \{i | x_{k-1} < t_i \leq x_k\}$ for $k = 2, ..., n-1$.

The parameters β (if not fixed) can be estimated from the data by solving the following criterion, a penalized nonlinear least squares problem where the objective function is penalized by the total variation and a positive tuning parameter $\lambda > 0$ is introduced

$$\underset{\beta}{\text{Minimize}} \quad \left\| f^{(r)}_{F,\beta} - y \right\|^2 + \lambda \widetilde{V}(f^{(r)}_{F,\beta}). \qquad (24)$$

ISBN: 978-989-95079-6-8

Suppose that each T_k has a sufficient number of elements to allow the following approximation problem be well defined: determine $\beta = (\beta_{2,0}, ..., \beta_{n+r-1,0}\beta_{2,1}, ..., \beta_{n+r-1,1})$ to minimize the functional

$$\Phi^{(r)}(\beta) = \left\| L^{(r)}(\beta)F^{(r)}(\beta) - y \right\|^2 + \lambda \widetilde{V}(f^{(r)}_{F,\beta}) \quad (25)$$

where matrix $L^{(r)}(\beta)$ is

$$L^{(r)}(\beta) = D^{(r)}(\beta)^{-1}A^{(r)}(\beta) \text{ and}$$
$$F^{(r)}(\beta) = S^{(r)}(\beta)^{-1}(A^{(r)}(\beta))^T y$$

Alternatively, we can tune the total variation $\widetilde{V}(f^{(r)}_{F,\beta})$ by fixing a variation level $\overline{V} > 0$ and solving the following constrained problem

$$\Phi^{(r)}(\widehat{\beta}) = Minimize \left\| L^{(r)}(\beta)F^{(r)}(\beta) - y \right\|^2 \quad (26)$$
$$s.t. \quad \widetilde{V}(f^{(r)}_{F,\beta}) \leq \overline{V}$$

We show the procedure with the following example and with the value of $\overline{V} = \gamma \widetilde{V}(y)$ (a fraction γ of the total variation of the data); the generated data are $y_i = f(t_i) + 2z_i$ and $z_i \in N(0,1)$ (the $m = 101$ points t_i are uniform on $[0,2]$) with the function $f(t) = 5e^{-0.5t}sin^2(\pi t)$. For the generated data, $\widetilde{V}(y) = 210.25$. Figs 6a)-6c) illustrate the best smoothing obtained by solving (26) for different values of $\gamma \in \{0.5, 0.2, 0.1\}$ and $n = 21$; Figs 7a)-7b) show the best smoothing for $\gamma \in \{0.2, 0.1\}$ and $n = 51$.

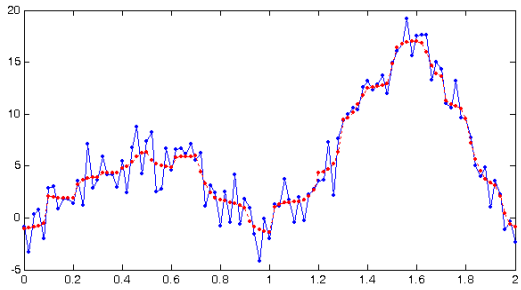


Figure 6a) Best smoothing fit with $n = 21, \gamma = 0.5$ ($r = 1$)

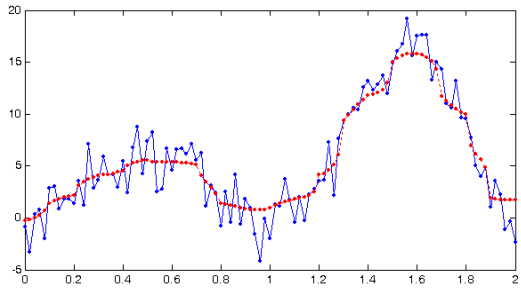


Figure 6b) Best smoothing fit with $n = 21, \gamma = 0.2$ ($r = 2$)

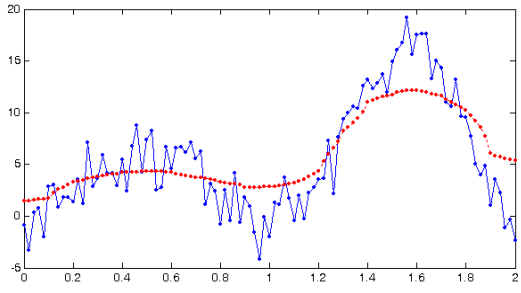


Figure 6c) Best smoothing fit with $n = 21, \gamma = 0.1$ ($r = 6$)

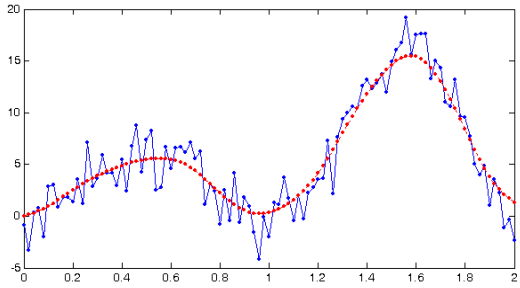


Figure 7a) Best smoothing fit with $n = 51, \gamma = 0.2$ ($r = 4$)

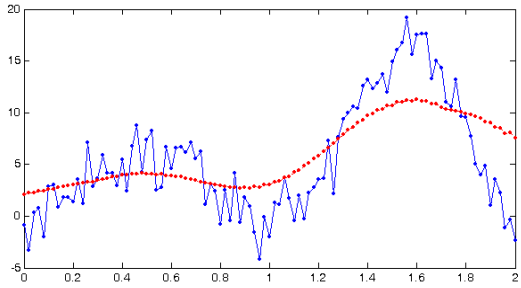


Figure 7b) Best smoothing fit with $n = 51, \gamma = 0.1$ ($r = 8$)

References

[1] P.L. Davies, A. Kovac, Local Extremes, runs, strings and multiresolution, *The Annals of Statistics*, 29 (2001) 1-65.

[2] A. Kovac, Smooth functions and local extreme values, *Computational Statistics & Data Analysis*, 51 (2007) 5155-5171.

[3] I. Perfilieva, Fuzzy Transforms: Theory and Applications, *Fuzzy Sets and Systems*, 157 (2006) 993-1023.

[4] I. Perfilieva, Fuzzy Transforms: A Challenge to conventional transforms. In P.W. Hawkes (Ed.), *Advances in Images and Electron Physics*, 147, Elsevier Academic Press, 2007, 137-196.

[5] I. Perfilieva, Fuzzy Transforms and their applications to data compression. In I. Bloch et al. (Eds), *Fuzzy Logic and Applications, LNAI 3849, Springer,* 2006, 19-31.

[6] L. Stefanini, Fuzzy Transform with Parametric LU-Fuzzy Partitions, Proceeding of the 8-th International FLINS Conference, (Madrid, Sept. 2008), in D. Ruan et Al. (Eds) "Computational Intelligence in Decision and Control", World Scientific, 2008, 399-404.

[7] L. Stefanini, L. Sorini, M.L. Guerra, Parametric representations of fuzzy numbers and application to fuzzy calculus, *Fuzzy Sets and Systems*, 157 (2006) 2423–2455.

[8] L. Stefanini, L. Sorini, M.L. Guerra, Fuzzy Numbers and Fuzzy Arithmetic. In W. Pedrycz, A. Skowron, V. Kreynovich (Eds), *Handbook of Granular Computing*, Chapter 12, J. Wiley & Sons, 2009.

[9] M. Stepnicka, R. Valasek, Fuzzy Transforms for functions of two variables. In J. Ramik, V. Novak (Eds), *Methods for decision support in environment with uncertainty*, University of Ostrava, IRAFM, 2003, 96-102.

ISBN: 978-989-95079-6-8

Towards F-transform of a Higher Degree

Irina Perfilieva, Martina Daňková

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech Republic
Email: Irina.Perfilieva, Martina.Dankova@osu.cz

Abstract— The aim of this contribution is to show how the F-transform technique can be generalized from the case of constant components to the case of polynomial components. For this purpose, we choose complete functional spaces with inner products. After a general presentation of F^m-transform where $m \geq 0$, a detailed characterization of the F^1-transform is given. We applied a technique of numeric integration in order to simplify computation of F^1-transform components. The inverse F^m-transform, $m \geq 0$, is defined in the same way as the ordinary F-transform.

Keywords— F-transform, F^m-transform, fuzzy partition, orthogonal basis, Gaussian quadrature

1 Introduction

The goal of this paper is to provide a deeper analysis of fuzzy modeling and its contribution to general mathematics. In [1], we introduced the notion of a fuzzy transform (F-transform, for short) which explains modeling with fuzzy IF-THEN rules as a specific transformation. This enabled us to compare the success of fuzzy modeling with the success of classical transforms (Fourier, Laplace, integral, wavelet etc.). From this point of view, fuzzy transforms contribute to the theory of linear and, in particular, integral transforms. Moreover, they initiated a theory of semi-linear spaces (see [2]).

In [1], the approximation property of fuzzy transform has been described and then, in [3], it has been shown how shapes of basic functions influence the approximation quality. The F-transform has many other useful properties and great potential for various applications, such as special numerical methods, solution of ordinary and partial differential equations with fuzzy initial condition [4, 5], mining dependencies from numerical data [6], applications to signal processing, compression and decompression of images [7, 8], and fusion of images [9].

We aim at showing how the F-transform technique can be generalized in the sense that constant components considered as polynomials of 0 degree are replaced by polynomials of degree $m \geq 1$. For this purpose, we choose complete functional spaces with inner products. After presentation of the generalized approach, a detailed characterization of the F^1-transform is given. We applied a technique of numeric integration in order to simplify computation of F^1-transform components. The inverse F^m-transform, $m \geq 1$, is defined in the same way as the inverse F-transform.

The paper is organized as follows: in Section 2, we introduce the concept of F^m-transform of finite degree $m \geq 0$, and show some of its properties and its relation to the ordinary F-transform. In Section 3, F^1-transform is introduced in details. Moreover, a simplification of F^1-transform components computation is discussed. The inverse F^m-transform is discussed in Section 4.

2 F-transform of an Arbitrary Finite Degree

Let us recall that the (direct) F-transform of an integrable function is a certain vector with real components. In [1], we proposed various formulas which represent components of the F-transform and showed a relationship between a given function and its F-transform. In this contribution, we propose to generalize our view on the F-transform and consider it as a vector of components that are orthogonal projections of an original function onto a linear subspace of certain functions that have polynomial representation. In the particular, if the degree of polynomials is zero, we obtain the originally proposed F-transform which will be referred to as F^0-transform in the sequel.

2.1 Fuzzy partition

Let $[a, b]$ be an interval on the real line $\mathbb{R}$. Fuzzy sets on $[a, b]$ will be identified with their membership functions, i.e. mappings from $[a, b]$ into $[0, 1]$. The notion of fuzzy partition is a principle one for our construction so that we will repeat it from [1] and slightly change below.

Definition 1
Let $[a, b]$ be an interval on $\mathbb{R}$, $n \geq 2$ and $x_1, \ldots, x_n$ nodes such that $a = x_1 < \ldots < x_n = b$. We say that fuzzy sets $A_1, \ldots, A_n$, identified with their membership functions, constitute a fuzzy partition *of $[a, b]$ if for $k = 1, \ldots, n$, they fulfill the following conditions :*

1. *$A_k : [a, b] \longrightarrow [0, 1]$, $A_k(x_k) = 1$;*
2. *for $k = 2, \ldots, n$, $A_k(x) = 0$ if $x \in [a, x_{k-1}]$ and for $k = 1, \ldots, n-1$, $A_k(x) = 0$ if $x \in [x_{k+1}, b]$,*
3. *$A_k(x)$ is continuous;*
4. *for $k = 2, \ldots, n$, $A_k(x)$ strictly increases on $[x_{k-1}, x_k]$ and for $k = 1, \ldots, n-1$, $A_k(x)$ strictly decreases on $[x_k, x_{k+1}]$;*
5. *for all $x \in [a, b]$*

$$\sum_{k=1}^{n} A_k(x) = 1. \tag{1}$$

The membership functions $A_1, \ldots, A_n$ are called basic functions.

ISBN: 978-989-95079-6-8

Let us extend $[a,b]$ by two extra nodes $x_0 = 2a - x_2$ and $x_{n+1} = 2b - x_{n-1}$ and for $x \in [x_0, a)$ define $A_1(x) = A_1(2a - x)$, and for $(b, x_{n+1}]$ define $A_n(x) = A_n(2b - x)$. Further on we will always assume that $[a,b]$ and membership functions A_1, A_n are extended.

We say that the fuzzy partition $A_1, \ldots, A_n$, $n \geq 3$, is *h-uniform* if nodes $x_0, \ldots, x_{n+1}$ are h-equidistant, i.e. for all $k = 1, \ldots, n$, $x_k = a + h(k-1)$ where $h = (b-a)/(n-1)$, and two additional properties are met:

6. $A_k(x_k - x) = A_k(x_k + x)$, for all $x \in [0, h]$, $k = 1, \ldots, n$,
7. $A_k(x) = A_{k-1}(x - h)$, for all $k = 2, \ldots, n$ and $x \in [x_{k-1}, x_{k+1}]$.

2.2 F-transform in a space of functions with scalar product

Let us fix $[a,b]$, $n \geq 3$, nodes $x_0 < \ldots < x_{n+1}$ and fuzzy partition $A_1, \ldots, A_n$ of $[a,b]$. For every $k = 1, \ldots, n$, let us denote $L_2(A_k)$ a set of functions $f : [x_{k-1}, x_{k+1}] \longrightarrow \mathbb{R}$ for which the following integral

$$\int_{x_{k-1}}^{x_{k+1}} f(x)^2 A_k(x) dx$$

exists. Let

$$(f,g)_k = \int_{x_{k-1}}^{x_{k+1}} f(x) g(x) A_k(x) dx, \tag{2}$$

be a weighted scalar product of functions f and g. Then $L_2(A_k)$ is a *linear space of functions with scalar product*.

For every integer $m \geq 0$, let $\varphi_0^k, \varphi_1^k, \ldots, \varphi_m^k \in L_2(A_k)$, be an orthogonal system of polynomials where $\varphi_0^k = 1$ and orthogonality is considered with respect to (2). Denote $L_2^m(A_k)$ a linear subspace of $L_2(A_k)$ with the basis $\varphi_0^k, \varphi_1^k, \ldots, \varphi_m^k$.

The following trick allows us to extend arbitrary function $f : [a,b] \longrightarrow \mathbb{R}$ to function $f^{ex} : [x_0, x_{n+1}] \longrightarrow \mathbb{R}$:

$$f^{ex}(x) = \begin{cases} f(a-x) = 2f(a) - f(a+x), \text{ if } x \in [0, x_2 - a], \\ f(x), \text{ if } x \in [a,b], \\ f(b+x) = 2f(b) - f(b-x), \text{ if } x \in [0, b - x_{n-1}] \end{cases}$$

Definition 2
Let $f : [a,b] \longrightarrow \mathbb{R}$ be a given function and f^{ex} its extension to $[x_0, x_{n+1}]$. Assume that for all $k = 1, \ldots, n$, $f^{ex}|_{[x_{k-1}, x_{k+1}]} \in L_2(A_k)$. Let F_k^m be the k-th orthogonal projection of f^{ex} on $L_2^m(A_k)$, $k = 1, \ldots, n$. Then the n-tuple $[F_1^m, \ldots, F_n^m]$ of functions is the F^m-transform of f with respect to $A_1, \ldots, A_n$. Every function F_k^m, $k = 1, \ldots, n$, is called the F^m-transform component.

Remark 1
Definition 2 does not provide us with a formal representation of components $F_1^m, \ldots, F_n^m$. Let us show how they can be obtained. According to the definition above, every component F_k^m minimizes the scalar product $(f^{ex} - g, f^{ex} - g)_k$ where g is an arbitrary function from $L_2^m(A_k)$. Taking into account that $\varphi_0^k, \varphi_1^k, \ldots, \varphi_m^k$ is the basis of $L_2^m(A_k)$, we can represent g by a linear combination $x_0\varphi_k^0 + x_1\varphi_k^1 + \ldots + x_m\varphi_k^m$ of basis functions. Assume that $c_0, c_1 \ldots, c_m$ are coefficients that minimize the following integral

$$\int_{x_{k-1}}^{x_{k+1}} (f^{ex}(x) - (x_0\varphi_k^0(x) + x_1\varphi_k^1(x) + \ldots + x_m\varphi_k^m(x)))^2 A_k(x) dx. \tag{3}$$

Hence, $F_k^m = c_0\varphi_k^0 + c_1\varphi_k^1 + \ldots + c_m\varphi_k^m$.

Below, we will prove the following fact: the original F-transform (see, e.g. [1]) is actually (up to the first and the last components) the F^0-transform. This requires to show that every F-transform component F_k, $k = 2, \ldots, n-1$, that has been originally introduced by

$$F_k = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x) A_k(x) dx}{\int_{x_{k-1}}^{x_{k+1}} A_k(x) dx}, \tag{4}$$

is the k-th orthogonal projection of f on $L_2^0(A_k)$. In order to prove this fact, we recall that the basis of $L_2^0(A_k)$ consists of the constant function φ_k^0 that has the value 1. Then the assertion above immediately follows from the proposition given below which has been proved in [1]:

Theorem 1
Let f be a continuous function on $[a,b]$ and $A_1, \ldots, A_n$ be basic functions which constitute a fuzzy partition of $[a,b]$. Then the k-th component F_k (4) of the F-transform gives minimum to the function

$$\Phi(y) = \int_a^b (f(x) - y)^2 A_k(x) dx$$

defined on $[f(a), f(b)]$.

3 F^1-transform

On the example of F^1-transform, we will show how an arbitrary component of F^m, $m \geq 1$, can be computed. Actually, one possibility of computation directly follows from minimization of integral (3) (see Remark 1). We will use another approach which corresponds to the definition of F^1-transform.

Throughout this section, we fix an h-uniform partition $A_1, \ldots, A_n$ of $[a,b]$ where $n \geq 3$, and assume that $f : [a,b] \longrightarrow \mathbb{R}$ is a function such that for all $k = 1, \ldots, n$, $f^{ex}|_{[x_{k-1}, x_{k+1}]} \in L_2(A_k)$. The explanation below will be given for an arbitrary k, $k = 1, \ldots, n$.

3.1 F^1-transform components

Since component F_k^1 of the F^1-transform is the k-th orthogonal projection of f^{ex} on $L_2^m(A_k)$, and $L_2^m(A_k)$ is a linear subspace of $L_2(A_k)$ with orthogonal basis φ_k^0, φ_k^1, we will begin this subsection with a characterization of orthogonal polynomials φ_k^0, φ_k^1 of degrees 0 and 1.

Lemma 1
Polynomials $\varphi_k^0, \varphi_k^1 : [x_{k-1}, x_{k+1}] \longrightarrow \mathbb{R}$ such that for all $x \in [x_{k-1}, x_{k+1}]$, $\varphi_k^0(x) = 1$, $\varphi_k^1(x) = x - x_k$ are orthogonal with weight A_k, $k = 1, \ldots, n$.

PROOF: The proof is technical and follows from the following assertions:

(i) $\int_{x_{k-1}}^{x_{k+1}} A_k(x) dx = h$,

ISBN: 978-989-95079-6-8

(ii) $\int_{x_{k-1}}^{x_{k+1}} xA_k(x)dx = hx_k$,

and properties of the uniform partition. □

Remark 2
Orthogonal polynomials of degrees 0 *and* 1, *described in Lemma 1, are invariant to shapes of* A_k.

Theorem 2
Under the assumptions above, the F^1*-transform of* f *with respect to* $A_1, \ldots, A_n$ *is the vector* $[F_1^1, \ldots, F_n^1]$ *of linear functions such that an arbitrary component* F_k^1, $k = 1, \ldots, n$, *is represented as follows:*

$$F_k^1(x) = c_k^0 + c_k^1(x - x_k), \quad x \in [x_{k-1}, x_{k+1}]$$

where

$$c_k^0 = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x)A_k(x)dx}{h}, \tag{5}$$

$$c_k^1 = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x)(x - x_k)A_k(x)dx}{\int_{x_{k-1}}^{x_{k+1}} (x - x_k)^2 A_k(x)dx}. \tag{6}$$

PROOF: The proof will be given for one fixed component F_k^1 where $k = 1, \ldots, n$. By Definition 2, F_k^1 is the k-th orthogonal projection of f^{ex} on $L_2^1(A_k)$ where the orthogonality is determined by (2). Therefore,

$$f^{ex}|_{[x_{k-1}, x_{k+1}]} = c_k^0 + c_k^1(x - x_k) + R_k(x)$$

where R_k is orthogonal to each basis functions φ_k^0 and φ_k^1. Therefore,

$$c_k^0 = \frac{(f^{ex}, \varphi_k^0)_k}{(\varphi_k^0, \varphi_k^0)_k},$$

$$c_k^1 = \frac{(f^{ex}, \varphi_k^1)_k}{(\varphi_k^1, \varphi_k^1)_k}$$

which after substitution gives the required expressions for c_k^0 and c_k^1. □

Corollary 1
Under the assumptions above, $c_k^0 = F_k$ *where* F_k *is the* k*-th component (4) of the ordinary* F*-transform.*

PROOF: The proof follows from expressions (4), (6) and equality $\int_{x_{k-1}}^{x_{k+1}} A_k(x)dx = h$. □

3.2 Simplification of F^1*-transform components computation*

In this subsection, we will show how Gaussian quadratures and the properties of orthogonal polynomials can be used for replacing integral $\int_{x_{k-1}}^{x_{k+1}} (x - x_k)^2 A_k(x)dx$ in the denominator of c_k^1 by its precise value.

At first, let us recall the notion of Gaussian quadratures (see e.g.[10]). The approximate equality

$$\int_{x_{k-1}}^{x_{k+1}} f(x)A_k(x)dx \approx \frac{h}{2} \sum_{i=1}^{N} d_i f(t_i) \tag{7}$$

which is precise for all polynomials of the highest possible degree is a Gaussian quadrature. We will put $N = 2$ and characterize arguments t_1, t_2 and the highest possible degree of polynomials which turn (7) into a precise equality. The following statement (see e.g.[10]) gives the required characterization: if

(i) t_1, t_2 are roots of polynomial φ_k^2 which is orthogonal to φ_k^0 and φ_k^1, and

(ii) equality

$$\int_{x_{k-1}}^{x_{k+1}} P_l(x)A_k(x)dx = h(d_1 P_l(t_1) + d_2 P_l(t_2)), \tag{8}$$

holds true for some coefficients d_1, d_2 and all polynomials P_l of degrees $l = 0, 1$,

then (8) holds true for all polynomials P_l of degrees $0 \leq l \leq 3$.

Thus, our next purpose is to find a polynomial of the degree 2, orthogonal to φ_k^0 and φ_k^1, as well as to find its roots. Due to positivity and symmetry of A_k, two lemmas below hold true.

Lemma 2
If t_1, t_2 *are symmetrical with respect to* x_k *then equalities*

$$\int_{x_{k-1}}^{x_{k+1}} A_k(x)dx = h(\frac{1}{2} + \frac{1}{2}) = h,$$

$$\int_{x_{k-1}}^{x_{k+1}} (x - x_k)A_k(x)dx = \frac{h}{2}((t_1 - x_k) + (t_2 - x_k)) = 0,$$

hold true for $l = 0, 1$.

Lemma 3
If we denote

$$I_2 = \int_{x_{k-1}}^{x_{k+1}} (x - x_k)^2 A_k(x)dx$$

then polynomial φ_k^2, *represented by*

$$\varphi_k^2(x) = (x - x_k)^2 - \frac{I_2}{h} \tag{9}$$

is orthogonal to φ_k^0 *and* φ_k^1 *on* $[x_{k-1}, x_{k+1}]$.

The roots t_1, t_2 *of* φ_k^2 *belong to* (x_{k-1}, x_{k+1}) *and are symmetrical with respect to* x_k, *i.e. for some* δ, $t_1 = x_k - \delta$ *and* $t_2 = x_k + \delta$.

The statement above together with Lemmas 2, 3 leads to the equality

$$\int_{x_{k-1}}^{x_{k+1}} (x - x_k)^2 A_k(x)dx = h\delta^2.$$

Thus, the desired simplified representation of c_k^1 is as follows:

$$c_k^1 = \frac{\int_{x_{k-1}}^{x_{k+1}} f(x)(x - x_k)A_k(x)dx}{h\delta^2}.$$

ISBN: 978-989-95079-6-8

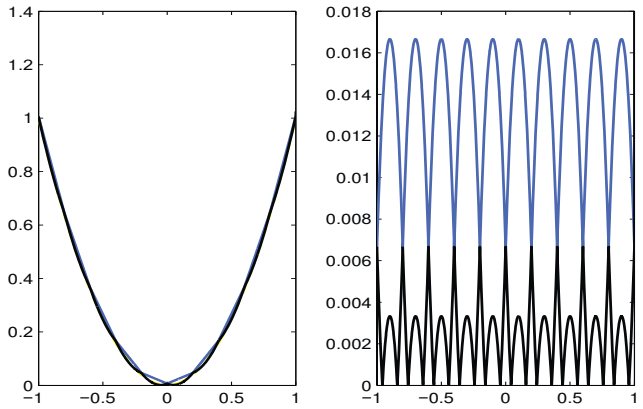


Figure 1: *Left.* The function x^2 and its inverse F^0 (gray line) and F^1 (black line) transforms. *Right.* Graphs of the error functions. Maximal errors of approximation are: 0.017 (the inverse F^0-transform) and 0.062 (the inverse F^1-transform).

4 Inverse F^m-transform

Similarly to the ordinary F-transform, the inverse F^m-transform is defined as a linear combination of the basic functions with "coefficients" given by the F^m-transform components.

Definition 3
Let $f : [a, b] \longrightarrow \mathbb{R}$ be a given function and f^{ex} its extension to $[x_0, x_{n+1}]$ such that for all $k = 1, \ldots, n$, $f^{ex}|_{[x_{k-1}, x_{k+1}]} \in L_2(A_k)$. For a certain $m \geq 0$, let $[F_1^m, \ldots, F_n^m]$ be the F^m-transform of f with respect to $A_1, \ldots, A_n$. Then the function

$$f_{F,m,n}(x) = \sum_{k=1}^{n} F_k^m A_k(x) \tag{10}$$

is called the inverse F^m-transform.

The following recurrent formula easily follows from Definition 3 and the whole structure of the F^m-transform, $m \geq 1$:

$$f_{F,m,n}(x) = f_{F,m-1,n}(x) + \sum_{k=1}^{n} c_m \varphi_k^m(x) A_k(x). \tag{11}$$

By Remark 1, the components F_k^m, $k = 1, \ldots, n$, $m \geq 0$, are the best approximation of f among all elements represented by linear combinations $a_0\varphi_k^0 + a_1\varphi_k^1 + \ldots + a_m\varphi_k^m$. Therefore, for each $k = 1, \ldots, n$, F_k^m is a better approximation of f than F_k^{m-1}. This observation together with (11) implicitly demonstrates that the quality of approximation of f by $f_{F,m,n}$ is better than that of approximation by $f_{F,m-1,n}$. The same conclusion follows also from our experiments (see Figure 1). The proof of this assertion is a matter of future investigation.

5 Conclusion

We have generalized the F-transform technique to the case where its components are polynomials. A detailed characterization of the F^1-transform with linear components was given. We have shown how a computation of F^1-transform components can be simplified if the technique of Gaussian quadratures is used. The inverse F^m-transform, $m \geq 1$, is defined in the same way as the inverse F-transform.

Acknowledgement

The paper has been supported partially by the grant IAA108270902 of GA AV and partially by the project MSM 6198898701 of the MŠMT ČR.

References

[1] I. Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy Sets and Systems*, 157:993–1023, 2006.

[2] A. Di Nola, A. Lettieri, I. Perfilieva, and V. Novák. Algebraic analysis of fuzzy systems. *Fuzzy Sets and Systems*, 158:1–22, 2007.

[3] L. Stefanini. Fuzzy transform with parametric lu-fuzzy partitions. In D. Ruan and et al., editors, *Computational Intelligence in Decision and Control*, pages 399–404. World Scientific, Paris, France, 2008.

[4] M. Štěpnička and R. Valášek. Fuzzy transforms for functions with two variables. In J. Ramík and V. Novák, editors, *Methods for Decision Support in Environment with Uncertainty*, pages 96–102. IRAFM, University of Ostrava, 2003.

[5] I. Perfilieva, H. De Meyer, B. De Baets, and D. Plskova. Cauchy problem with fuzzy initial condition and its approximate solution with the help of fuzzy transform. In *Proc. of WCCI 2008, IEEE Int. Conf. on Fuzzy Systems*, pages 2285–2290, Hong Kong, 2008.

[6] V. Novák, I. Perfilieva, A. Dvořák, G. Chen, Q. Wei, and P. Yan. Fuzzy transform in the analysis of data. *Int. Journ. of Appr. Reasoning*, 48:4–22, 2008.

[7] S. Sessa, F. Di Martino, V. Loia, and I. Perfilieva. An image coding/decoding method based on direct and inverse fuzzy transforms. *Int. Journ. of Appr. Reasoning*, 48:110—-131, 2008.

[8] I. Perfilieva. Fuzzy transforms: A challenge to conventional transforms. In P. W. Hawkes, editor, *Advances in Images and Electron Physics, 147*, pages 137–196. Elsevier Academic Press, San Diego, 2007.

[9] I. Perfilieva and M. Daňková. Image fusion on the basis of fuzzy transforms. In D. Ruan and et al., editors, *Computational Intelligence in Decision and Control*, pages 471–476. World Scientific, Paris, France, 2008.

[10] N. S. Bachvalov, N. P. Zhidkov, and G. M. Kobelkov. *Numerical Methods*. Nauka, Moscow, 1987.

ISBN: 978-989-95079-6-8

Some chaotic and mixing properties of Zadeh's extension

Jiří Kupka

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
30. dubna 22, 701 33 Ostrava, Czech Republic
Email: Jiri.Kupka@osu.cz

Abstract— *Let X be a compact metric space, let φ be a continuous self-map on X, and let $\mathbb{F}(X)$ denote the space of fuzzy sets on X equipped with the levelwise topology. In this paper we study relations between various dynamical properties of a given (crisp) dynamical system (X, φ) and its Zadeh's extension Φ on $\mathbb{F}(X)$. Among other things we study various (weak, strong, mild etc.) mixing properties and also several kinds of chaotic behaviors (Li-Yorke chaos, ω-chaos, distributional chaos, topological chaos etc.).*

Keywords— Zadeh's extension, fuzzification, chaos, mixing, transitivity, topological entropy.

1 Introduction

Throughout this paper, let (X, d_X) be a compact metric space and let $C(X)$ denote the space of continuous maps $\varphi : X \to X$. A discrete dynamical system is a pair (X, φ). For other notions and notations mentioned in this section, we refer to Section **??**. It is well known ([**?**]) that the discrete dynamical system (X, φ) naturally induces a dynamical system $(\mathbb{F}(X), \Phi)$ on the space $\mathbb{F}(X)$ of all fuzzy compact subsets of X. The map Φ is called the *fuzzification* (or *Zadeh's extension*) (see (**??**)).

It is natural to ask the following **question:** *how is the dynamical complexity of the fuzzified (resp. crisp) dynamical system related to the dynamical properties of the original (resp. fuzzy) one*. There are only a few papers devoted to this question so far – for example, [**?**], [**?**] and [**?**], where different chaotic properties of fuzzy discrete dynamical systems were considered.

In this paper, we consider the space $\mathbb{F}(X)$ of upper semi-continuous fuzzy sets with compact supports. This space is equipped with the topology induced by the levelwise metric d_∞ (see (**??**)), since this topology is stronger than the other topologies commonly used in fuzzy topological dynamics (e.g. see [**?**]). We especially deal with the subspace $\mathbb{F}^1(X) \subseteq \mathbb{F}(X)$ of all normal fuzzy sets on X (see (**??**)). The reason for this is the following: no fuzzification $\Phi : \mathbb{F}(X) \to \mathbb{F}(X)$ admits one of the simplest chaotic behaviors (namely the transitivity, see Proposition **??**) and, consequently, it does not admit more complex behavior.

This paper is a partial answer to the question mentioned above. Our results concerning the most commonly used chaotic and mixing properties can be summarized as follows:

• if P denotes either the distributional or Li-Yorke or topological or ω-chaos then

$$(\varphi \text{ has } P \Rightarrow \Phi \text{ has } P), \text{ but } (\Phi \text{ has } P \not\Rightarrow \varphi \text{ has } P), \quad (1)$$

• if P denotes either transitivity or total transitivity, then

$$(\Phi \text{ has } P \Rightarrow \varphi \text{ has } P), \text{ but } (\varphi \text{ has } P \not\Rightarrow \Phi \text{ has } P), \quad (2)$$

• if P denotes one of the following properties: exactness, sensitive dependence, weak mixing, mild mixing, or strong mixing, then

$$(\Phi \text{ has } P \Rightarrow \varphi \text{ has } P), \quad (3)$$

but the validity of the converse implication is unknown.

This paper is organized as follows: in Section **??**, we introduce notation and definitions used in this paper. Then, in Section **??**, some preliminary results are proven, showing also connections between the set-valued and fuzzified system induced by the same original system. Finally, the chaotic and mixing properties are studied in Section **??**.

2 Definitions and notation

Further we denote by $\mathbb{N}$ and $\mathbb{R}$ the set of integers and real numbers, respectively. Now we define some classic notions from topological dynamics. For a given dynamical system (X, φ) and a given point $x \in X$, we define the n-th *iteration* of the point x inductively by $\varphi^0(x) = x$, $\varphi^{n+1}(x) = \varphi(\varphi^n(x))$ for any $n \in \mathbb{N}$. Then, the sequence $\{\varphi^n(x)\}_{n\in\mathbb{N}}$ of all iterations of x is called the *trajectory* of the point x. Any limit point of the trajectory of the point x is called an *ω-limit point* of the point x, and the union $\omega_\varphi(x)$ of all ω-limit points of the point x is the *ω-limit set* of the point x. The iterations of a given set $A \subseteq X$ are defined analogously. The point $x \in X$ is called *fixed* if $\varphi(x) = x$ or *periodic* if $\varphi^k(x) = x$ for some $k \in \mathbb{N}$. We denote by $\omega(\varphi)$, $P(\varphi)$ and $Fix(\varphi)$ the set of ω-limit, periodic and fixed points, respectively.

A map $\varphi \in C(X)$ is called *transitive* if for any non-empty open subsets $U, V \subseteq X$, there exists some $k \in \mathbb{N}$ such that $\varphi^k(U) \cap V \neq \emptyset$. The map φ is *totally transitive* if the n-th iteration of φ is transitive for any $n \in \mathbb{N}$. The map φ is *weakly mixing* if the product map $\varphi \times \varphi$ is transitive. The map φ is *strongly mixing* if for any non-empty open subsets $U, V \subseteq X$ there exists some $m \in \mathbb{N}$ such that $\varphi^k(U) \cap V \neq \emptyset$ for any $k \geq m$. The map φ is *topologically exact* (or simply *exact*) if for any non-empty subset $U \subseteq X$, there exists some $k \in \mathbb{N}$, such that $\varphi^k(U) = X$.

2.1 Chaotic properties

When defining chaotic properties we follow the notation introduced in [**?**]. The notion of distributional chaos was introduced in [**?**]. For any $x, y \in X$, $t \in \mathbb{R}$ and $n \in \mathbb{N}$, set

$$\xi(x, y, t, n) = \#\{i,\ 0 \leq i < n \wedge d(\varphi^i(x), \varphi^i(y)) < t\}. \quad (4)$$

Set

$$F^\star_{xy}(t) = \limsup_{n\to\infty} \frac{1}{n}\xi(x, y, t, n) \quad (5)$$

ISBN: 978-989-95079-6-8

and

$$F_{xy}(t) = \liminf_{n\to\infty} \frac{1}{n}\xi(x,y,t,n). \quad (6)$$

Obviously, both maps $F^{\star}_{xy}$ and F_{xy} are nondecreasing, $0 \leq F_{xy}(t) \leq F^{\star}_{xy}(t) \leq 1$ for all $t \in \mathbb{R}$, $F^{\star}_{xy}(t) = 0$ if $t \leq 0$ and $F^{\star}_{xy}(t) = 1$ if $t \geq diam(X)$. The map $F^{\star}_{xy}$ ($F_{xy}(t)$) is an upper (a lower) distribution function for $x, y \in X$.

The map φ is *distributionally chaotic* of type 1 (d_1C) if $F^{\star}_{xy} \equiv 1$ and $F_{xy}(t) = 0$ for some $t > 0$. The map φ is *distributionally chaotic* of type 2 (d_2C) if $F^{\star}_{xy} \equiv 1$ and $F^{\star}_{xy}(t) > F_{xy}(t)$ for some $t > 0$. Finally, the map φ is *distributionally chaotic* of type 3 (d_3C) if $F^{\star}_{xy}(t) > F_{xy}(t)$ for all $t \in J$, where J is a nondegenerate interval.

It follows from the definition that

$$d_1C \Rightarrow d_2C \Rightarrow d_3C. \quad (7)$$

However, the converse implications are not valid (see, for instance, [**?**] and [**?**]).

Two points $x, y \in X$ form a *Li-Yorke pair* if

$$\limsup_{n\to\infty} d_X(\varphi^n(x), \varphi^n(y)) > 0 \quad (8)$$

and

$$\liminf_{n\to\infty} d_X(\varphi^n(x), \varphi^n(y)) = 0. \quad (9)$$

A set $S \subseteq X$ is a *LY-scrambled set* for the map φ if $\#S \geq 2$ and every pair from S is Li-Yorke. The map φ is *Li-Yorke chaotic* (shortly *LYC*) if there exists an uncountable LY-scrambled set.

A map $\varphi \in C(X)$ is *ω-chaotic* ([**?**]) (shortly ωC) if there exists an uncountable *ω-scrambled* set $S \subseteq X$, i.e. for any two points $x, y \in S$, the following conditions are satisfied: (i) $\omega_\varphi(x) \setminus \omega_\varphi(y)$ is uncountable, (ii) $\omega_\varphi(x) \setminus \omega_\varphi(y) \neq \emptyset$ and (iii) $\omega_\varphi(x) \cap P(\varphi) \neq \emptyset$.

If a map $\varphi : X \to X$ is transitive and $P(\varphi)$ is dense in X then φ is called *Devaney chaotic*. It should be mentioned that in the original definition of Devaney, φ *depends sensitively on initial conditions*,i.e. there exists $\delta > 0$ such that for any $x \in X$ and any open neighborhood U of x there is $y \in U$ satisfying $d_X(\varphi^k(x), \varphi^k(y)) > \delta$ for some $k \in \mathbb{N}$. But it was proved that this condition is implied by the transitivity and density of periodic points (see [**?**] and [**?**]).

The notion of positive topological entropy was firstly defined by Bowen ([**?**]). The topological entropy of a map φ is a number $h(\varphi) \in [0, \infty]$, defined by

$$h(\varphi) = \lim_{\varepsilon>0} \limsup_{n\to\infty} \#E(n, \varphi, \varepsilon), \quad (10)$$

where $E(n, \varphi, \varepsilon)$ is a $(n, \varphi, \varepsilon)$-*span* with a minimal possible number of points, i.e. a set such that for any $x \in X$ there exists a $y \in E(n, \varphi, \varepsilon)$ satisfying $d(\varphi^k(x), \varphi^k(y)) < \varepsilon$ for any j, $1 \leq j \leq n$. A map φ is *topologically chaotic* (shortly PTE) if $h(\varphi) > 0$. It is well-known that the topological entropy is monotone in the following way: for any $A, B \subseteq X$,

$$A \subseteq B \quad \Rightarrow \quad h(\varphi|_A) \leq h(\varphi|_B). \quad (11)$$

A map $\varphi \in C(X)$ has the *specification property* if for any $\varepsilon > 0$ there is a positive $M \in \mathbb{N}$ such that for any integer $k \geq 2$ and any k points $x_i \in X$, $i = 1, 2, \ldots, k$ and any $2k$ integers $a_1 \leq b_1 < a_2 \leq b_2 < \ldots < a_k \leq b_k$ with $a_i - b_{i-1} \geq M$, there exists $z \in X$ for which

$$d(\varphi^n(z), \varphi^n(x_i)) < \varepsilon \quad (12)$$

for any $n = a_i, \ldots, b_i$ and any $i = 1, 2, \ldots, k$.

The following implications are currently known among the chaotic and mixing properties mentioned above:

$$\begin{gathered}\text{specification property} \Rightarrow \text{strong mixing} \Rightarrow \\ \text{mild mixing} \Rightarrow \text{weak mixing} \Rightarrow \\ \text{total transitivity} \Rightarrow \text{transitivity}.\end{gathered} \quad (13)$$

For further details and relations among the chaotic properties, we refer to [**?**] and to the references therein.

2.2 Metric spaces of fuzzy sets

Let (X, d) denote a compact metric space, and let A, B be non-empty closed subsets of X. The *Hausdorff metric* D_X between A and B is defined, as usual, by

$$D_X(A, B) = \inf\{\varepsilon > 0 \,|\, A \subseteq U_\varepsilon(B) \wedge B \subseteq U_\varepsilon(A)\}, \quad (14)$$

where

$$U_\varepsilon(A) = \{x \in X \,|\, D(x, A) < \varepsilon\}, \quad (15)$$

and

$$D(x, A) = \inf_{a\in A} d(x, a). \quad (16)$$

By $\mathbb{K}(X)$ we denote the space of all nonempty compact subsets of X, equipped with the Hausdorff metric D_X. It is well known (c.f. [**?**]) that $(\mathbb{K}(X), D_X)$ is compact, complete and separable whenever X is compact, complete and separable.

A fuzzy set A on the space X is a function $A : X \to I$ where I denotes the closed unit interval $[0, 1]$. The *α-cuts* (or the *α-level sets*) $[A]_\alpha$ and the *support* $supp(A)$ of a given fuzzy set A are defined as usual by -

$$[A]_\alpha = \{x \in X \,|\, A(x) \geq \alpha\}, \ \alpha \in [0, 1], \quad (17)$$

and

$$supp(A) = \overline{\{x \in X \,|\, A(x) > 0\}}. \quad (18)$$

Further, we define $\mathbb{F}(X)$ as the system of all upper semi-continuous fuzzy sets $A : X \to I$ having compact supports. Moreover, let

$$\mathbb{F}^1(X) = \{A \in \mathbb{F}(X) \,|\, A(x) = 1 \text{ for some } x \in X\} \quad (19)$$

denote the system of all *normal* fuzzy sets on X. Finally, we define $\emptyset_X$ as the *empty fuzzy set* ($\emptyset_X(x) = 0$ for each $x \in X$) on the space X, and $\mathbb{F}_0(X)$ as the system of all nonempty fuzzy sets.

Let us define a *levelwise* metric d_∞ on $\mathbb{F}_0(X)$ by

$$d_\infty(A, B) = \sup_{\alpha\in(0,1]} D_X([A]_\alpha, [B]_\alpha). \quad (20)$$

This equality defines the levelwise metric correctly only for non-empty fuzzy sets $A, B \in \mathbb{F}_0(X)$ whose maximal values are identical, since the Hausdorff distance D_X is only measured between two non-empty closed subsets of the space X.

ISBN: 978-989-95079-6-8

Thus, we consider the following extension of the Hausdorff metric D_X:

$$D_X(\emptyset,\emptyset) = 0 \text{ and } D_X(\emptyset, A) = diam\,(X) \qquad (21)$$

for any $A \in \mathbb{K}(X)$. With this extension, **(??)** correctly defines the levelwise metric on $\mathbb{F}(X)$. It is obvious that

$$d_\infty(\emptyset_X,\emptyset_X) = 0 \text{ and } d_\infty(\emptyset_X, A) = diam\,(X) \qquad (22)$$

for any $A \in \mathbb{F}_0(X)$.

It should be noted that the metric d_∞ is one of the three most commonly used metrics in fuzzy topological dynamics. We also recall that the metric space $(\mathbb{F}(X), d_\infty)$ is complete but is not separable and not compact and that the levelwise topology induced by d_∞ is stronger than the remaining (*sendograph* and *endograph*) ones. For more details we refer to [**?**] and to the references therein.

2.3 Zadeh's extension

Let X be a compact metric space and $\varphi \in C(X)$. Then a *fuzzification* (or *Zadeh's extension*) of the (crisp) dynamical system (X, φ) is a map $\Phi : \mathbb{F}(X) \to \mathbb{F}(X)$ defined by

$$(\Phi(A))(x) = \sup_{y\in\varphi^{-1}(x)} \{A(y)\} \qquad (23)$$

for any $A \in \mathbb{F}(X)$ and $x \in X$.

It is shown recently by [**?**] that, if X is a compact metric space, then the fuzzification $\Phi : \mathbb{F}(X) \to \mathbb{F}(X)$ is continuous if and only if $\varphi : X \to X$ is continuous. The last statement was generalized about the case of locally compact metric spaces in [**?**] recently.

It is known that, for any $\alpha \in (0, 1]$ and any $A \in \mathbb{F}(X)$,

$$\varphi([A]_\alpha) = [\Phi(A)]_\alpha. \qquad (24)$$

Similarly, $\varphi(supp(A)) = supp(\Phi(A))$ holds.

3 Preliminary results

Inspired by the results mentioned, for instance, in [**?**], we define some basic properties of generalized extensions. For any $U \subseteq X$, we define

$$e(U) = \{B \in \mathbb{F}(X) \,|\, supp(B) \subseteq U\} \qquad (25)$$

It is obvious that $e(U) \neq \emptyset$ if and only if $U \neq \emptyset$. Moreover, we have the following assertion (Lemma **??**) that was partially proved in [**?**]:.

Lemma 1 *A subset U is a non-empty open subset of X if and only if $e(U)$ is a non-empty open subset of $\mathbb{F}(X)$.*

Proof. Since the implication "⇒" has been proven in [**?**], the converse remains to be proven. So let $e(U)$ be a non-empty open subset of $(\mathbb{F}(X), d_\infty)$. Assume by contradiction that U is not open. Take any $x \in U \setminus int(U)$ and consider a fuzzy set $\chi_{\{x\}}$. Then, for any $\varepsilon > 0$, an open ε-neighborhood $V \subseteq X$ of x intersects the exterior of U. Consequently, χ_V is ε-close to $\chi_{\{x\}}$ (by the definition of d_∞), but $\chi_V \notin e(U)$. Thus no ε-neighborhood V of $\chi_{\{x\}}$ is a subset of $e(U)$. This contradicts the fact that $e(U)$ is open in $(\mathbb{F}(X), d_\infty)$. □

Lemma 2 (Representation theorem of Negoita-Ralescu, e.g. [**?**]) *Consider a family $\{B_\alpha \,|\, \alpha \in [0, 1]\}$ of closed subsets of X satisfying the following two conditions:*
(a) $B_\beta \subseteq B_\alpha \subseteq B_0$ if $0 \leq \alpha \leq \beta$,
(b) if $\{\alpha_n\}$ is an increasing sequence in I converging to α_0 then $B_{\alpha_0} = \bigcap_{n\in\mathbb{N}} B_{\alpha_n}$.
Then there exists $B \in \mathbb{F}(X)$ such that $[B]_\alpha = B_\alpha$.

Conversely, if B is a fuzzy set on X then the system $\{B_\beta\}_{\beta\in I}$ defined by $B_\beta = [B]_\beta$ for any $\beta \in (0, 1]$ and $B_0 = supp(B)$ satisfies conditions (a) and (b).

Lemma 3 *Let U, V be two subsets of X and $\varphi \in C(X)$. Then*
(i) $e(U \cap V) = e(U) \cap e(V)$,
(ii) $\Phi(e(U)) \subseteq e(\varphi(U))$,
(iii) $\Phi(e(U)) = e(\varphi(U))$ whenever U is closed.

Proof. The statements (i) and (ii) were already proved in [**?**].

The statement (iii) still remains to be proven. The inclusion $\Phi(e(U)) \subseteq e(\varphi(U))$ in (iii) follows from (ii). Let us prove $e(\varphi(U)) \subseteq \Phi(e(U))$ if U is closed. Take any $A \in e(\varphi(U))$. We want to show that there exists $B \in e(U)$ such that $\Phi(B) = A$. Since $A \neq \emptyset_X$ and upper semi-continuous, there exists $\alpha_0 = \max_{x\in X}\{A(x) \,|\, A(x) > 0\}$. Moreover, for any $\alpha \in (0, \alpha_0]$, $[A]_\alpha$ is nonempty, closed and, consequently by the continuity of φ, $\varphi^{-1}([A]_\alpha) \cap U$ is also nonempty and closed.

By the definition of the fuzzification Φ, $\max(A) = \max(B)$ whenever $B \in \mathbb{F}(X)$ is any preimage of A. So a fuzzy set $B \in e(U)$ can be defined as follows. For any $\beta \in (0, \alpha_0]$ we put

$$[B]_\beta = \varphi^{-1}([A]_\beta) \cap U. \qquad (26)$$

We also put $[B]_0 = U$ and $[B]_\beta = \emptyset$ for any $\beta \in (\alpha_0, 1]$. Obviously the system $\{[B]_\beta\}_{\beta\in I}$ satisfies the condition (a) of Lemma **??**. We shall now show that the condition (b) of Lemma **??** is satisfied, i.e. $B \in \mathbb{F}(X)$. Then, by the definition of B, we obtain that $B \in e(U)$.

Assume that $\{\beta_n\} \subseteq I$ is an increasing sequence that converges to $\beta_0 \leq \alpha_0$. Suppose by contradiction that $[B]_{\beta_0} \neq \bigcap_{n\in\mathbb{N}}[B]_{\beta_n}$. By the monotonicity of $\{\beta_n\}$ and Lemma **??** (a), the only possibility is that $[B]_{\beta_0}$ is a proper subset of $\bigcap_{n\in\mathbb{N}}[B]_{\beta_n}$, i.e.

$$\Big(\bigcap_{n\in\mathbb{N}} [B]_{\beta_n}\Big) \setminus [B]_{\beta_0} \neq \emptyset. \qquad (27)$$

Take any $x_0 \in \bigcap[B]_{\beta_n} \setminus [B]_{\beta_0}$ and then take $x_n \in [B]_{\beta_n} \setminus [B]_{\alpha_0}$ such that $\{x_n\}$ converges to x_0. Obviously, since $x_0 \notin [B]_{\beta_0}$ we obtain from **(??)** that

$$\varphi(x_0) \notin [A]_{\beta_0}. \qquad (28)$$

On the other hand, $x_n \in [B]_{\beta_n}$ for any $n \in \mathbb{N}$, i.e.,

$$\varphi(x_n) \in [A]_{\beta_n}, \text{ for any } n \in \mathbb{N}. \qquad (29)$$

Now the continuity of φ implies that $\{\varphi(x_n)\}$ converges to $\varphi(x_0)$. Hence **(??)** and **(??)** imply that $[A]_{\beta_0}$ is a proper subset of $\bigcap_{n\in\mathbb{N}}[A]_{\beta_n}$, i.e. $A \notin \mathbb{F}(X)$ by Lemma **??** – a contradiction. Thus, we have shown that $B \in \mathbb{F}(X)$.

ISBN: 978-989-95079-6-8

Finally, by the construction of B, we have $\Phi(B) = A$. □

We will need some further notation. For any $\alpha \in (0,1]$ and $U \subseteq X$, set

$$e_\alpha(U) = \{A \in \mathbb{F}(X) \mid [A]_\alpha \neq \emptyset \text{ and } [A]_\alpha \subseteq U\} \tag{30}$$

and

$$\vartheta(U) = e_1(U) \cap e(U). \tag{31}$$

Lemma 4 *For any $\alpha \in (0,1]$, $e_\alpha(U)$ is open in $(\mathbb{F}(X), \tau_\infty)$ if and only if $U \subseteq X$ is open in X.*

Proof. Let $\alpha \in (0,1]$ be fixed. We shall show that $e_\alpha(U)$ is open in $(\mathbb{F}(X), \tau_\infty)$ if U is open. Take any $A \in e_\alpha(U)$. Since $[A]_\alpha \subseteq U$ is closed and U is open, there exists an open ε-neighborhood V of $[A]_\alpha$ lying in U for any $\varepsilon > 0$. Consequently, by the definition of d_∞, if we consider an open ε-neighborhood $V' \subseteq (\mathbb{F}(X), d_\infty)$ of A with $\varepsilon < diam(X)$, then for any $B \in V'$,

$$[B]_\alpha \subseteq V \subseteq U. \tag{32}$$

Thus, there exists an open neighborhood V' of A in $e_\alpha(U)$, i.e., $e_\alpha(U)$ is open in $(\mathbb{F}(X), \tau_\infty)$.

Let us prove the converse implication. Assume by contradiction that U is not open and take any $x \in U \setminus int(U)$. Then, by the definition of $e_\alpha(U)$, $\chi_{\{x\}}$ belongs to $e_\alpha(U)$ but no ε-neighborhood of $\chi_{\{x\}}$ is a subset of $e_\alpha(U)$, i.e. $e_\alpha(U)$ is not open - a contradiction. □

Corollary 1 *A subset $U \subseteq X$ is open in X if and only if $\vartheta(U)$ is open in $(\mathbb{F}(X), \tau_\infty)$ (and therefore also in $(\mathbb{F}^1(X), d_\infty)$).*

Lemma 5 *Let X be a compact metric space and $\varphi \in C(X)$ Then, for any $\alpha \in (0,1]$ and $U, V \subseteq X$,*
(i) $e_\alpha(U \cap V) = e_\alpha(U) \cap e_\alpha(V)$,
(ii) $\Phi(e_\alpha(U)) \subseteq e_\alpha(\varphi(U))$.

Proof. Clearly, for any $\alpha \in (0,1]$, $A \in e_\alpha(U \cap V)$ if and only if $[A]_\alpha \subseteq U \cap V$, if and only if $[A]_\alpha \subseteq U$ and $[A]_\alpha \subseteq V$, if and only if $A \in e_\alpha(U)$ and $A \in e_\alpha(V)$, if and only if $A \in e_\alpha(U) \cap e_\alpha(V)$. Thus, (i) holds.

Let us prove (ii). Consider any $A \in \Phi(e_\alpha(U))$. Then there exists $B \in e_\alpha(U)$ for which $\Phi(B) = A$. Since $[B]_\alpha \subseteq U$ it follows from **(??)** and from the continuity of φ that

$$\varphi([B]_\alpha) = [\Phi(B)]_\alpha = [A]_\alpha \subseteq \varphi(U), \tag{33}$$

i.e. $A \in e_\alpha(\varphi(U))$, and the inclusion is valid for any $\alpha \in (0,1]$. □

We are ready to modify Lemma **??** and to prove the next lemma, which is used for the further study of dynamics in the space of normal fuzzy sets on X. For completeness, we note the obvious fact that $\vartheta(U) \neq \emptyset$ if and only if $U \neq \emptyset$.

Lemma 6 *Let U, V be two subsets of X and $\varphi \in C(X)$. Then*
(i) $\vartheta(U \cap V) = \vartheta(U) \cap \vartheta(V)$,
(ii) $\Phi(\vartheta(U)) \subseteq \vartheta(\varphi(U))$,
(iii) $\Phi(\vartheta(U)) = \vartheta(\varphi(U))$ whenever U is closed.

Proof. The statements (i) and (ii) are easy consequences of Lemmas **??** and **??**. Moreover, the proof of the statement (iii) is a slight modification of the proof of Lemma **??** ($\max_{x \in X} A(x) = \alpha_0 = 1$). □

At the end of this section we mention two simple properties of the usual fuzzification. By χ_A we denote the characteristic function of a given set $A \subseteq X$.

Lemma 7 *Let X be a compact metric space, $\varphi \in C(X)$ let Φ be a fuzzification of φ. Then*

$$\Phi(\alpha\chi_A) = \alpha\chi_{\varphi(A)} \tag{34}$$

for any $\alpha \in (0,1]$.

Proof. Obvious. □

Lemma 8 *Let X be a compact metric space. Then the map $i : (\mathbb{K}(X), D_X) \to (\mathbb{F}(X), d_\infty)$ defined by $i(A) = \chi_A$ for any $A \in \mathbb{K}(X)$ is an isometrical embedding.*

Proof. It is obvious that the map i is correctly defined. It follows directly from the definition of the metric d_∞ on the space of fuzzy sets that

$$d_X(x,y) = d_\infty(\chi_{\{x\}}, \chi_{\{y\}}) \tag{35}$$

for any $x, y \in X$, and also

$$D_X(U,V) = d_\infty(\chi_U, \chi_V) \tag{36}$$

for any $U, V \in \mathbb{K}(X)$. Thus, by Lemma **??**, i is injective, continuous and isometric. □

Remark 1 *The same assertion is true when we consider either the sendograph or endograph topology instead of the levelwise one, since the equalities used in the proof are valid also in those topologies. Consequently, some results (Propositions **??** and **??**) presented in the next section are valid also in all mentioned topologies. The property mentioned in Lemma **??** was firstly mentioned in Kloeden's pioneering paper [**?**] in a little bit different setting.*

4 Chaotic and mixing properties.

Firstly we study some chaotic properties. It follows from Lemma **??** that any fuzzified system is an extension of the set-valued dynamical system $(\mathbb{K}(X), \tilde{\varphi})$ induced by the original dynamical system (X, φ). Consequently, we can use results published in [**?**] to easily provide some answers (Propositions **??** and **??**) to the question mentioned at the beginning of this paper.

Lemma 9 [**?**] *Let X be a compact metric space and $\varphi \in C(X)$.*

(i) If there is a set $S \subseteq X$ that is d_jC-scrambled ($j = 1,2,3$) (ω-scrambled, LY-scrambled, resp.) for the map φ then there exists d_jC-scrambled ($j = 1,2,3$) (ω-scrambled, LY-scrambled, resp.) set for $\tilde{\varphi}$ with the same cardinality as S.

(ii) If φ is d_jC (LYC, ωC, resp.) then the same holds for $\tilde{\varphi}$.

ISBN: 978-989-95079-6-8

Lemma 10 [**?**] *There exists a compact metric space X with a zero topological entropy map φ for which there exist no LY pairs, d_3C-scrambled sets, or ω-scrambled sets, such that $\tilde{\varphi}$ is PTE, d_1C, ωC and LYC.*

Proposition 1 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be a fuzzification of φ. Then*

(i) If there is a set $S \subseteq X$ that is d_jC-scrambled ($j = 1, 2, 3$) (ω-scrambled, LY-scrambled, resp.) for the map φ, then there exists a d_jC-scrambled ($j = 1, 2, 3$) (resp. ω-scrambled, LY-scrambled) set S' for Φ, with the same cardinality as S.

(ii) If φ is d_jC (resp. LYC, ωC), then the same holds for Φ.

Proof. Let $S' = \{A \in \mathbb{F}(X) \mid A = \chi_{\{x\}} \text{ and } x \in S\}$. Then this proposition is a corollary of Lemmas **??**, **??** and **??**. □

Proposition 2 *Let Φ denote a fuzzification of φ. There exists a compact metric space X with a zero topological entropy map φ for which there exists no LY pairs, neither d_3C-scrambled set, nor ω-scrambled set such that Φ is PTE, d_1C, ωC and LYC.*

Proof. This proposition is an easy consequence of Lemmas **??**, **??** and **??**, and of the fact that the three versions of distributional chaos are comparable (**??**). □

The following proposition justifies why we study dynamical properties of fuzzified dynamical systems on the space of normal fuzzy sets in the rest of this paper.

Proposition 3 *Let X be a compact metric space and $\varphi \in C(X)$. Then no fuzzification $\Phi : (\mathbb{F}(X), d_\infty) \to (\mathbb{F}(X), d_\infty)$ of φ is transitive.*

Proof. Take any $A, B \in \mathbb{F}(X)$ such that $\max(A) \neq \max(B)$. By the definition of d_∞, there is an open ε-neighborhood U_ε of A (resp., an open ϑ-neighborhood V_ϑ of B) for some $\varepsilon, \vartheta > 0$ such that $\max(A') = \max(A)$ for any $A' \in U_\varepsilon$ (resp., $\max(B') = \max(B)$ for any $B' \in V_\vartheta$). Moreover, Φ preserves the maxima of fuzzy sets from U_ε and V_ϑ. Thus, by the choice of A and B, $\Phi^n(U_\varepsilon) \cap V_\vartheta = \emptyset$ (resp., $U_\varepsilon \cap \Phi^n(V_\vartheta) = \emptyset$) for any $n \in \mathbb{N}$, i.e., Φ is not transitive. □

Corollary 2 *Let X be a compact metric space and $\varphi \in C(X)$. Then a fuzzification $\Phi : (\mathbb{F}(X), d_\infty) \to (\mathbb{F}(X), d_\infty)$ of φ has none of the properties listed in (**??**) (specification property, strong mixing, mild mixing, weak mixing and total transitivity).*

Proof. The proof is an obvious consequence of Proposition **??** and (**??**). □

For completeness, it follows directly from Lemmas **??** and (**??**) that if the original dynamical system (X, φ) is topologically chaotic then also the fuzzified system $(\mathbb{F}^1(X), \Phi)$ (and hence $(\mathbb{F}(X), \Phi)$) is topologically chaotic. Now we would like to mention a recent paper [**?**] where the conditions defining Devaney chaos ([**?**]) were studied on the space of all fuzzy sets on X. We recall that we extended their results since no such fuzzification can be Devaney chaotic by Lemma **??**.

Moreover, dynamical properties of fuzzifications on the space of normal fuzzy sets on X have been never studied before. So we study them in the rest of this section.

Proposition 4 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is transitive then φ is transitive, but the converse implication does not hold.*

Proof. Let the assumptions be fulfilled. Consider any open subsets $U, V \subseteq X$. We want to show that φ is transitive, i.e. there exists $n \in \mathbb{N}$ such that

$$\varphi^n(U) \cap V \neq \emptyset. \tag{37}$$

First we take open subsets $U', V' \subseteq X$ such that $U' \subseteq \overline{U'} \subseteq U$ and $V' \subseteq \overline{V'} \subseteq V$. Then, by Corollary **??**, $\vartheta(U')$ and $\vartheta(V')$ are open subsets of $(\mathbb{F}^1(X), \tau_\infty)$. Since Φ is transitive, there exists $n \in \mathbb{N}$ for which $\Phi^n(\vartheta(U')) \cap \vartheta(V') \neq \emptyset$. Then $\Phi^n(\vartheta(\overline{U'})) \cap \vartheta(\overline{V'}) \neq \emptyset$, and by using results (i) and (iii) of Lemma **??**, we obtain

$$\vartheta(\varphi^n(\overline{U'})) \cap \vartheta(\overline{V'}) \neq \emptyset \quad \Rightarrow \quad \vartheta(\varphi^n(\overline{U'}) \cap \overline{V'}) \neq \emptyset. \tag{38}$$

Therefore $\varphi^n(\overline{U'}) \cap \overline{V'} \neq \emptyset$, and consequently, (**??**) is proved.

A counterexample (an irrational rotation on the unit circle) showing that the converse implication does not hold was presented in [**?**] for the fuzzification Φ defined on $\mathbb{F}(X)$, but it can be applied also to $\Phi|_{\mathbb{F}^1(X)}$. □

Proposition 5 *Let X be a compact metric space, $\varphi \in C(X)$, and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is totally transitive, then φ is totally transitive, but the converse implication does not hold.*

Proof. This proposition is an easy corollary of Proposition **??**, since the transitivity of any iteration Φ^n of Φ on $\mathbb{F}^1(X)$ implies the transitivity of φ^n on X for any $n \in \mathbb{N}$.

As a counterexample to the converse, we again can use the example mentioned in [**?**]. Their example was used as the counterexample to

$$\varphi \text{ transitive } \not\Rightarrow \ \Phi \text{ transitive}, \tag{39}$$

but the map φ used in [**?**] is an irrational rotation on the unit circle, and it is obvious that any irrational rotation on the unit circle is totally transitive. □

Proposition 6 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is sensitively dependent, then φ is sensitively dependent.*

Proof. Let the assumptions be fulfilled. We shall show that φ is sensitively dependent for the same sensitivity constant δ as Φ. Take any $x \in X$ and any open neighborhood U of x. Then $\vartheta(U)$ is an open neighborhood of $\chi_{\{x\}}$ by Corollary **??**. Since Φ is sensitively dependent, there exist $A \in \vartheta(U)$ and $n \in \mathbb{N}$ such that

$$d_\infty(\Phi^n(\chi_{\{x\}}), \Phi^n(A)) > \delta. \tag{40}$$

ISBN: 978-989-95079-6-8

By Lemma **??**

$$d_\infty(\Phi^n(\chi_{\{x\}}), \Phi^n(A)) = d_\infty(\chi_{\{\varphi^n(x)\}}, \Phi^n(A)) > \delta. \quad (41)$$

Thus, by the definition of the metric d_∞, there exists $y_0 \in supp(A)$ such that $d_X(\varphi^n(x), \varphi^n(y_0)) > \delta$. Obviously, y_0 lies in U and this shows that φ is sensitively dependent. □

Proposition 7 *Let X be a compact metric space, $\varphi \in C(X)$, and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is exact, then φ is exact.*

Proof. Let the assumptions be fulfilled, and let $U \subseteq X$ be an open set. Then there exists a closed $U' \subseteq U$ with nonempty interior $int(U')$. By Corollary **??**, $\vartheta(int(U'))$ is an open subset of $\mathbb{F}^1(X)$. Since Φ is exact, there exists some $k \in \mathbb{N}$ for which $\Phi^k(\vartheta(int(U'))) = \mathbb{F}^1(X)$. Therefore, by the choice of U' and by Lemma **??**,

$$\mathbb{F}^1(X) = \Phi^k(\vartheta(int(U'))) \subseteq \Phi^k(\vartheta(U')) \quad (42)$$

and

$$\Phi^k(\vartheta(U')) = \vartheta(\varphi^k(U')) \subseteq \vartheta(\varphi^k(U)). \quad (43)$$

Thus, $\vartheta(\varphi^k(U))$ covers $\mathbb{F}^1(X)$ and hence, since $\mathbb{F}^1(X)$ also contains the characteristic functions of all singletons (see Lemma **??**), $\varphi^k(U) = X$. □

Remark 2 *The map Φ on $\mathbb{F}(X)$ cannot be exact since Φ cannot be transitive (see Lemma **??**). Moreover, the validity of the converse implication to Proposition **??** is still unknown.*

Proposition 8 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is strongly mixing then φ is strongly mixing.*

Proof. The proof of this proposition is a slight variation of the proof of Proposition **??**. □

Proposition 9 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is mildly mixing then φ is mildly mixing.*

Proof. Let the assumptions be fulfilled. Let Y be any compact metric space and (Y, ψ) be any transitive discrete dynamical system. According to the definition of the mild mixing property, we want to show that the product system $(X \times Y, \varphi \times \psi)$ is transitive, i.e. for any open sets $U, V \subseteq X \times Y$ there exists $k \in \mathbb{N}$ for which

$$(\varphi \times \psi)^k(U) \cap V \neq \emptyset. \quad (44)$$

So let $U, V \subseteq X \times Y$ be given. Then there are open subsets $U_1, V_1 \subseteq X$, $U_2, V_2 \subseteq Y$ for which $U_1 \times U_2 \subseteq U$ and $V_1 \times V_2 \subseteq V$. We also consider open subsets $U_1', V_1' \subseteq X$ such that $\overline{U_1'} \subseteq U_1$ and $\overline{V_1'} \subseteq V_1$. By Corollary **??**, $\vartheta(U_1')$ and $\vartheta(V_1')$ are open subsets of $\mathbb{F}^1(X)$. Thus, since Φ is mildly mixing, there exists $k \in \mathbb{N}$ for which

$$(\Phi \times \psi)^k(\vartheta(U_1') \times U_2) \cap (\vartheta(V_1') \times V_2) \neq \emptyset, \quad (45)$$

i.e.

$$\psi^k(U_2) \cap V_2 \neq \emptyset \quad (46)$$

and also

$$\Phi^k(\vartheta(U_1')) \cap \vartheta(V_1') \neq \emptyset. \quad (47)$$

As in the proof of Proposition **??** ($U_1' = U'$ and $V_1' = V'$), the last inequality implies that

$$\varphi^k(U_1) \cap V_1 \neq \emptyset. \quad (48)$$

Consequently, this together with (**??**) gives

$$(\varphi \times \psi)^k(U_1 \times U_2) \cap (V_1 \times V_2) \neq \emptyset. \quad (49)$$

Finally, by the choice of U_1, U_2, V_1, V_2, (**??**) is proven. □

Proposition 10 *Let X be a compact metric space, $\varphi \in C(X)$ and let Φ be the fuzzification of φ. If $\Phi : (\mathbb{F}^1(X), d_\infty) \to (\mathbb{F}^1(X), d_\infty)$ is weakly mixing, then φ is weakly mixing.*

Proof. The proof of this proposition is a slight variation of the proof of Proposition **??**. □

References

[1] Zadeh L. *Fuzzy sets and systems, Proc. Symp. on Systems Theory*. Polytechnic Institute Press, Brooklyn, New York, 1965.

[2] Diamond P. and Pokrovskii A. Chaos, entropy and a generalized extension principle. *Fuzzy Sets and Systems*, 61:277 – 283, 1994.

[3] Kloeden P. Chaotic iterations of fuzzy sets. *Fuzzy Sets and Systems*, 42:37 – 42, 1991.

[4] Román-Flores H. and Chalco-Cano Y. Some chaotic properties of zadeh's extension. *Chaos, Solitons and Fractals*, 35:452 – 459, 2008.

[5] Kupka J. On fuzzifications of discrete dynamical systems. *Fuzzy Sets and Systems*, submitted.

[6] García Guirao J. L., Kwietniak D., Lampart M., Oprocha P., and Peris A. Chaos on hyperspaces. *Nonlinear Analysis: Theory, Methods & Applications*, accepted manuscript.

[7] Schweizer B. and Smítal J. Measures of chaos and a spectral decomposition of dynamical systems on the interval. *Trans. Amer. Math. Soc.*, 344:737 – 754, 1994.

[8] Balibrea F., Smítal J., and Štefánková M. The three versions of ditributional chaos. *Chaos, Solitons and Fractals*, 23:1581 – 1583, 2005.

[9] Smítal J. and Štefánková M. Distributional chaos for triangular maps. *Chaos, Solitons and Fractals*, 21:1125 – 1128, 2004.

[10] Li S. ω-chaos and topological entropy. *Trans. Amer. Math. Soc.*, 339(1):243 – 249, 1993.

[11] Banks J., Brooks J., Cairns G., Davis G., and Stacey P. On devaney's definition of chaos. *Amer. Math. Monthly*, 99:332 – 334, 1992.

[12] Glasner E. and Weiss B. Sensitive dependence on initial conditions. *Nonlinearity*, 6:1067 – 1075, 1993.

[13] Bowen R. Entropy for group endomorphisms and homogeneous spaces. *Trans. Amer. Math. Soc.*, 153:401 – 414, 1971.

[14] Kuratowski K. *Topology*. Vol. II, Academic Press, London, New York, 1968.

[15] Diamond P. and Kloeden P. *Metric spaces of fuzzy sets: theory and applications*. World Scientific, Singapore, 1994.

[16] Kloeden P. Fuzzy dynamical systems. *Fuzzy Sets and Systems*, 7:275 – 296, 1982.

[17] Devaney R. *An introduction to chaotic dynamical systems*. 2nd ed. Boulder: Westview Press, 2003.

ISBN: 978-989-95079-6-8

Associativity of triangular norms in light of web geometry

Milan Petrík[1,2] Peter Sarkoci[3]

1. Institute of Computer Science, Academy of Sciences of the Czech Republic,
Prague, Czech Republic
2. Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Prague, Czech Republic
3. Department of Knowledge-Based Mathematical Systems, Johannes Kepler University,
Linz, Austria
E-mails: {petrik@cs.cas.cz|petrikm@cmp.felk.cvut.cz}, peter.sarkoci@{jku.at|gmail.com}

Abstract— The aim of this paper is to promote web geometry and, especially, the Reidemeister closure condition as a powerful and intuitive tool characterizing associativity of the Archimedean triangular norms. In order to demonstrate its possible applications, we provide the full solution to the problem of convex combinations of nilpotent triangular norms.

Keywords— Archimedean triangular norm, web geometry, Reidemeister closure condition.

1 Triangular norms

The notion of *triangular norm* was originally introduced within the framework of probabilistic metric spaces [12]. Since then, triangular norms have found diverse applications in the theory of fuzzy sets, fuzzy decision making, in models of certain many-valued logics or in multivariate statistical analysis; for a reference see the books by Alsina, Frank, and Schweizer [5] and by Klement, Mesiar, and Pap [8].

A conjunctor is a function $K\colon [0,1]^2 \to [0,1]$ which is nondecreasing in both arguments, commutative, and which satisfies the *boundary condition* $T(x,1) = x$ for all $x \in [0,1]$. A triangular norm (shortly a *t-norm*, usually denoted by T) is a conjunctor which satisfies the *associativity equation* $T(T(x,y),z) = T(x,T(y,z))$ for all $x,y,z \in [0,1]$. This paper deals mainly with *Archimedean t-norms*. Let us recall that for every t-norm T, a number $n \in \mathbb{N}\cup\{0\}$, and $x \in [0,1]$ a *natural power* of x, denoted $(x)_T^{(n)}$, is defined by:

$$(x)_T^{(n)} = \begin{cases} 1 & \text{if} \quad n = 0\,, \\ T\left(x,(x)_T^{(n-1)}\right) & \text{if} \quad n > 0\,. \end{cases} \tag{1}$$

A t-norm T is said to be *Archimedean* if and only if for every pair $x,y \in \,]0,1[$, $x < y$, there exists a natural number $n \in \mathbb{N}$ such that $(y)_T^{(n)} < x$. A t-norm which is continuous and strictly increasing on the half-open square $]0,1]^2$ is said to be *strict*. A continuous t-norm T is called *nilpotent* if and only if for every $x \in \,]0,1[$ there exists a natural number $n \in \mathbb{N}$ such that $(x)_T^{(n)} = 0$. A prototypical example of a strict and a nilpotent t-norm is the product t-norm, $T_{\mathbf{P}}(x,y) = x \cdot y$, and the Łukasiewicz t-norm $T_{\mathbf{L}}(x,y) = \max\{x+y-1,0\}$,

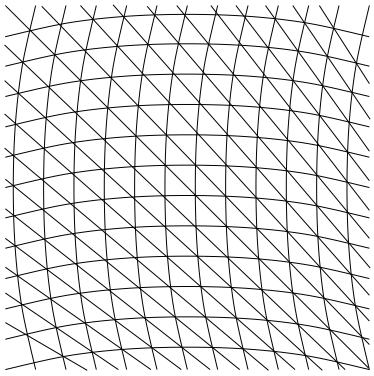


Figure 1: Example of a complete 3-web.

respectively. It is possible to show that a continuous Archimedean t-norm is either strict or nilpotent. A t-norm T is said to be *cancellative* if it satisfies $T(a,b) = T(a,c) \Rightarrow b = c$ for every $a,b,c \in [0,1]$, $a \neq 0$. A t-norm T is said to be *weakly cancellative* [9] if it satisfies $T(a,b) = T(a,c) \Rightarrow b = c$ for every $a,b,c \in [0,1]$ with $T(a,b) \neq 0$ and $T(a,c) \neq 0$. Every cancellative t-norm is weakly cancellative. Under the assumption of continuity, the set of cancelative t-norms and the set of strict t-norms coincide. Under the same assumption, the set of weakly cancelative t-norms coincides with the set of Archimedean t-norms.

2 Web geometry and local loops

In this section, *web geometry* is explained as a tool allowing to visualize algebraic identities. In particular, it is shown that it visualizes associativity of Archimedean t-norms. A detailed introduction to the subject is given in the monograph by Blashke and Bol [6]. Also the collection of papers by Aczél, Akivis, and Goldberg [1, 2, 3] can serve as an (English) introductory text.

Definition 2.1 *A* groupoid *(or a* magma*) is an algebraic structure $\mathcal{G} = (G,\circ)$ on a set G where $\circ\colon G\times G \to G$ is a binary operation. A* quasigroup *is a groupoid in which the*

ISBN: 978-989-95079-6-8

equations $a \circ x = b$ and $y \circ a = b$ have unique solutions for every a and b in G. Finally, a loop *is an algebraic structure $\mathcal{L} = (G, \circ, e)$ where $(G, \circ)$ is a quasigroup and e is an identity element, i.e. $x \circ e = x = e \circ x$ for every $x \in G$.*

The definition of a quasigroup, $\mathcal{G} = (G, \circ)$, allows to define the left and the right inverse. For this purpose we introduce a prefix notation $g(x, y) = x \circ y$.The left inverse is then defined for every $u \in G$ as ${}^{-1}g(u, y) = x$ such that $g(x, y) = u$. Similarly, the right inverse is, for every $v \in G$, $g^{-1}(x, v) = y$ such that $g(x, y) = v$. We say that g is *invertible* in x, resp. y. Defining a point $A = (a, b) \in G \times G$ we may introduce a new operation, $\bullet\colon G \times G \to G$, as

$$u \bullet v = g\left({}^{-1}g(u, b), g^{-1}(a, v)\right) . \qquad (2)$$

It is easy to show that $(G, \bullet)$ is a loop with a unit element $e = g(a, b)$; this loop is called a *local loop* of the quasigroup $(G, \circ)$ at the point $A = (a, b)$.

Definition 2.2 *Let M be a non-empty set and let $\lambda_1, \lambda_2, \lambda_3$ be three families of subsets of M. To elements of M we refer as to points and to elements of λ_α, $\alpha \in \{1, 2, 3\}$, as to lines. We say that the system $(M, \lambda_1, \lambda_2, \lambda_3)$ is a* complete 3-web *if and only if the following conditions hold:*

1. *Any point $p \in M$ is incident to just one line of each family λ_α, $\alpha \in \{1, 2, 3\}$.*
2. *Any two lines of different families are incident to exactly one point of M.*
3. *Two distinct lines from the same family λ_α are disjoint.*

Note that Item 3 of the definition is redundant and can be derived from Item 1. We kept the definition in this redundant form in order to keep some of the furher considerations simpler.

Usually M is equipped with a topology which turns the set into a manifold. In such a case the families λ_α are often required to be foliations. Figure 1 shows such an example of 3-web where M is a two dimmensional domain in plane and λ_α are foliations of co-dimension 1.

Notice that all the examples of 3-webs, shown in figures of this paper, are simplified. That means that only some particular lines from uncountable sets λ_α, $\alpha \in \{1, 2, 3\}$, are drawn.

Complete 3-webs are closely connected to quasigroups and, especially, loops. Every quasigroup defines a 3-web in the following way. Let $\mathcal{G} = (G, \circ)$ be a quasigroup defined on a manifold $M = G \times G$ and let $x, y \in G$. Let the lines of sets λ_1, λ_2, and λ_3 be given by the equations $x = a$, $y = b$, and $x \circ y = c$ respectively for fixed $a, b, c \in G$. The pair (M, λ_α), $\alpha \in \{1, 2, 3\}$, is then a 3-web.

Conversely, a 3-web (M, λ_α), $\alpha \in \{1, 2, 3\}$, on a manifold M, defines a binary operation which is invertible with respect to both operands; yet this time the result is not unique. Denote one set of the lines of the 3-web as λ_x, second as λ_y, and third as λ_z (the lines in the set λ_z are called *contour lines*), let $A \in M$ and let $g_x \in \lambda_x$ and $g_y \in \lambda_y$ be the lines passing through A; cf. Figure 2. Now we define an operation $*\colon \lambda_z \times \lambda_z \to \lambda_z$. Take $u, v \in \lambda_z$, define points $U, V \in M$ as $U = u \cap g_x$ and $V = v \cap g_y$, respectively. Let $g_u \in \lambda_y$ be the line passing through U and let $g_v \in \lambda_x$ be the line passing through V. Denote the intersection of these two lines as $W = g_u \cap g_v$; the line w passing through W is the result of the operation $u * v$.

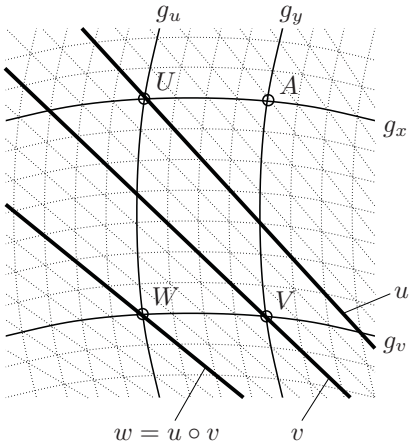


Figure 2: Operation defined on a complete 3-web.

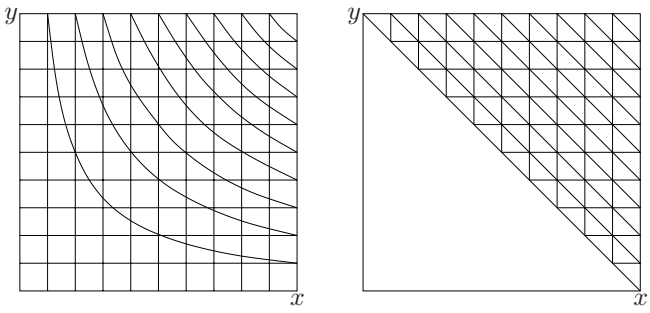


Figure 3: 3-webs given by the product t-norm, $T_{\mathbf{P}}$, and by the Łukasiewicz t-norm, $T_{\mathbf{L}}$.

It can be shown easily that the operation $*$ is invertible in both variables. Moreover, the line $e \in \lambda_z$, passing through the point A, behaves, with respect to the oparation $*$, as the unit element. Thus $(\lambda_z, *)$ is a loop. Moreover, this loop coincides, up to an isomorphism, with the local loop of $(G, \circ)$ at the point A.

T-norms, defined on the unit interval $[0, 1]$, form neither groupoids nor loops. Nevertheless, the 3-web given by a continuous Archimedean (i.e. strict or nilpotent) t-norm T satisfies all the requirements given by Definition 2.2 if the manifold M is defined as the subset of $[0, 1]^2$ where the t-norm attains non-zero values. A manifold, induced this way by a binary operation $T\colon [0, 1]^2 \to [0, 1]$, will be denoted as $\operatorname{Man} T$. Thus

$$M = \operatorname{Man} T = \left\{(x, y) \in [0, 1]^2 \mid T(x, y) > 0\right\} . \qquad (3)$$

Figure 3 shows 3-webs given by $T_{\mathbf{P}}$ and $T_{\mathbf{L}}$ as examples of a strict and a nilpotent t-norm respectively.

ISBN: 978-989-95079-6-8

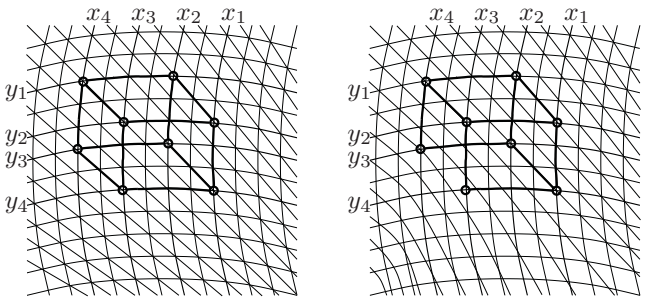


Figure 4: Example of a closed Reidemeister figure and a non-closed Reidemeister figure.

3 Reidemeister closure condition

Different types of 3-webs are characterized by closure conditions. These closure conditions have their counterparts in the related loops as algebraic properties of the loop operations. In this text we are interested in the associativity of t-norms since the associativity is the only property of a t-norm which cannot be intuitively interpreted from its graph. The 3-web counterpart of the associativity is the *Reidemeister closure condition*.

A 3-web satisfies the Reidemeister closure condition if and only if every Reidemeister figure in this web is closed; see Figure 4. Described in the terminology of contour lines, the Reidemeister closure condition is as follows. Let $x_1, x_2, x_3, x_4 \in \lambda_x$ and $y_1, y_2, y_3, y_4 \in \lambda_y$. If the points $(x_1 \cap y_2)$ and $(x_2 \cap y_1)$ lie on the same contour line (i.e. a line from the set λ_z), and if so do the pair of points $(x_1 \cap y_4)$ and $(x_2 \cap y_3)$, and the pair of points $(x_3 \cap y_2)$ and $(x_4 \cap y_1)$, then the points $(x_3 \cap y_4)$ and $(x_4 \cap y_3)$ also lie on the same contour line. The left part of Figure 4 shows a 3-web which satisfies the Reidemeister closure condition whereas the right part shows a 3-web where the condition is violated.

Let us now define a relation $\sim \subseteq M \times M$. We say that two points $A, B \in M$ are in the relation $\sim$, and we write $A \sim B$, if and only if they are both elements of the same line from the set λ_z. Using the language of $\sim$ the Reidemeister condition reads as:

$$\Big(\big((x_1 \cap y_2) \sim (x_2 \cap y_1)\big) \& \big((x_1 \cap y_4) \sim (x_2 \cap y_3)\big) \& \big((x_3 \cap y_2) \sim (x_4 \cap y_1)\big)\Big) \Rightarrow \big((x_3 \cap y_4) \sim (x_4 \cap y_3)\big). \quad (4)$$

4 Reidemeister closure condition and continuous Archimedean t-norms

We are going to show that the level set system of a continuous Archimedean t-norm always satisfies the Reidemeister closure condition. Moreover, the Reidemeister closure condition characterizes the associativity of a continuous Archimedean t-norm. For this purpose we recall that if we relax the requirement of the associativity in the definition of a t-norm, we obtain the definition of a conjunctor.

Let K be a continuous Archimedean conjunctor; for the sake of compactness we will use the infix notation: $K(x, y) = x \circ y$. Let (M, λ_α) be the corresponding 3-web defined on the manifold $M = \operatorname{Man} K = \{(x, y) \in [0, 1]^2 \mid x \circ y > 0\}$. Such a 3-web, as in the case of the continuous Archimedean t-norms, satisfies all the requirements given by Definition 2.2. In view of the previous section, the 3-web given by $\circ$ satisfies the Reidemeister closure condition if and only if the following condition holds for any $x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4 \in M$:

$$\Big(\big(x_1 \circ y_2 = x_2 \circ y_1\big) \& \big(x_1 \circ y_4 = x_2 \circ y_3\big) \& \big(x_3 \circ y_2 = x_4 \circ y_1\big)\Big) \Rightarrow \big(x_3 \circ y_4 = x_4 \circ y_3\big). \quad (5)$$

In the sequel, this condition will be denoted by $\mathbf{R}$.

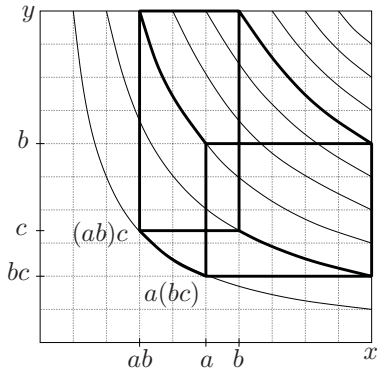


Figure 5: Reidemeister figure on the 3-web given by a continuous Archimedean t-norm.

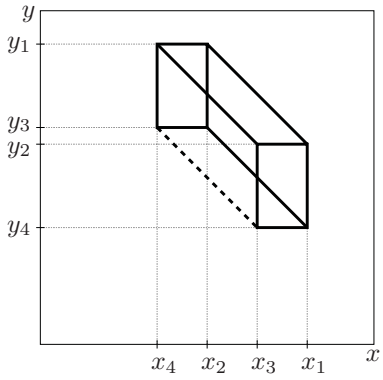


Figure 6: Illustration for the proof: an arbitrary Reidemeister figure.

Figure 5 shows that $\mathbf{R}$ implies the associativity of $\circ$ in a rather intuitive way. For any $a, b, c \in [0, 1]^2$, such that $(a \circ b) \circ c \in \operatorname{Man} \circ$ and $a \circ (b \circ c) \in \operatorname{Man} \circ$, there can be constructed a Reidemeister figure which, as can be seen in Figure 5, is closed if and only if $(a \circ b) \circ c = a \circ (b \circ c)$.

Figure 5 shows also that, in the other way round, if $\circ$ is associative then all the Reidemeister figures that "touch" the lines given by $x = 1$ and $y = 1$ are closed. In other words, $\mathbf{R}$ has to be satisfied for any $x_2, x_3, x_4, y_2, y_3, y_4 \in [0, 1]$ and, at least, for $x_1 = y_1 = 1$. Let us denote this weaker condition by $\mathbf{R_1}$. It is now clear that K is associative if and only if it

ISBN: 978-989-95079-6-8

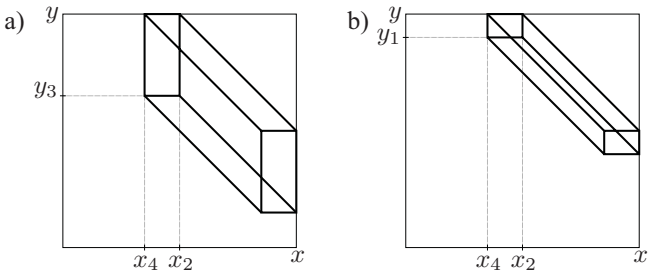


Figure 7: Illustration for the proof.

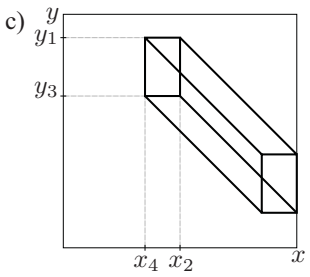


Figure 8: Illustration for the proof.

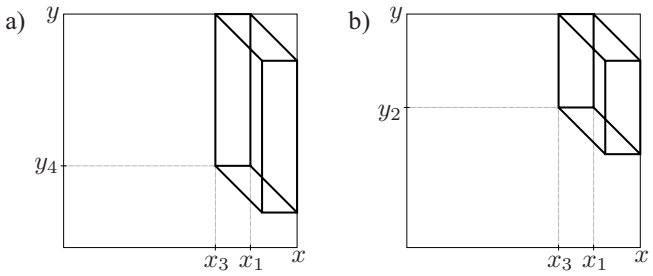


Figure 9: Illustration for the proof.

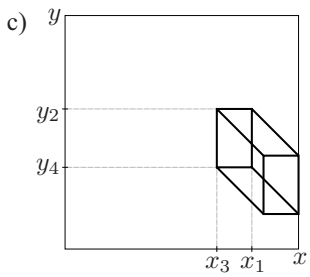


Figure 10: Illustration for the proof.

satisfies $\mathbf{R_1}$. In order to show that $\circ$ is associative if and only if it satisfies $\mathbf{R}$, we need to show that $\mathbf{R_1} \Leftrightarrow \mathbf{R}$. Obviously, $\mathbf{R} \Rightarrow \mathbf{R_1}$. The inverse implication, $\mathbf{R_1} \Rightarrow \mathbf{R}$, is given the following way.

Let us have a continuous Archimedean conjunctor K which satisfies $\mathbf{R_1}$. Let us have a Reidemeister figure drawn on its 3-web for arbitrary x_1, x_2, x_3, x_4, $y_1, y_2, y_3, y_4 \in [0, 1]$, see Figure 6. We are going to show that this figure shall be always closed. Thanks to $\mathbf{R_1}$, the Reidemeister figures, shown in Figure 7-a and Figure 7-b, are closed. Combining these two figures together it can be concluded that the Reidemeister figure in Figure 8-c is closed as well. By the same deduction, from the closedness of the Reidemeister figures in Figure 9-a and Figure 9-b it can be concluded that the Reidemeister figure in Figure 10-c is closed. Now, combining the closed Reidemeister figures in Figure 8-c and Figure 10-c, the closedness of the Reidemeister figure in Figure 6 is proven.

Corollary 4.1 *A continuous Archimedean conjunctor is associative (and, thus, a t-norm) if and only if it satisfies* $\mathbf{R}$.

5 Applications

5.1 Problem of convex combinations of t-norms

The question was introduced by Alsina, Frank, and Schweizer [4]. The approach of web geometry allows to give the following, recently published [11], answers:

Theorem 5.1 *Let T_1 and T_2 be two continuous Archimedean t-norms such that $\operatorname{Man} T_1 \neq \operatorname{Man} T_2$. Then no non-trivial convex combination F of T_1 and T_2 is a t-norm.*

Corollary 5.2 *Combining the result of Theorem 5.1 with the result given by Jenei [7], a non-trivial convex combination of two distinct nilpotent t-norms is never a t-norm.*

Corollary 5.3 *A non-trivial convex combination of a strict and a nilpotent t-norm is never a t-norm. This is an alternative proof of the result given earlier by Ouyang, Fang, and Li [10].*

Acknowledgment

The first author was supported by the Grant Agency of the Czech Republic under Project 401/09/H007.

References

[1] J. Aczél. Quasigroups, nets and nomograms. *Advances in Mathematics*, 1:383–450, 1965.

[2] M. A. Akivis and V. V. Goldberg. Algebraic aspects of web geometry. *Commentationes Mathematicae Universitatis Carolinae*, 41(2):205–236, 2000.

[3] M. A. Akivis and V. V. Goldberg. Local algebras of a differential quasigroup. *Bulletin of the American Mathematical Society*, 43(2):207–226, 2006.

[4] C. Alsina, M. J. Frank, and B. Schweizer. Problems on associative functions. *Aequationes Mathematicae*, 66(1–2):128–140, 2003.

[5] C. Alsina, M. J. Frank, and B. Schweizer. *Associative Functions: Triangular Norms and Copulas*. World Scientific, Singapore, 2006.

[6] W. Blaschke and G. Bol. *Geometrie der Gewebe, topologische Fragen der Differentialgeometrie*. Springer, Berlin, 1939.

[7] S. Jenei. On the convex combination of left-continuous t-norms. *Aequationes Mathematicae*, 72(1–2):47–59, 2006.

ISBN: 978-989-95079-6-8

[8] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*, vol. 8 of *Trends in Logic*. Kluwer Academic Publishers, Dordrecht, Netherlands, 2000.

[9] F.Montagna, C. Noguera, and R. Horčík. On Weakly Cancellative Fuzzy Logics. *Journal of Logic and Computation*, 16(4):423–450, August 2006.

[10] Y. Ouyang and J. Fang. Some observations about the convex combinations of continuous triangular norms. *Nonlinear Analysis*, 2007.

[11] M. Petrík, P. Sarkoci. Convex combinations of nilpotent triangular norms. *Journal of Mathematical Analysis and Applications*, 350(1):271–275, February 2009. DOI: 10.1016/j.jmaa.2008.09.060.

[12] B. Schweizer, A. Sklar. *Probabilistic Metric Spaces*. North-Holland, Amsterdam 1983; 2nd edition: Dover Publications, Mineola, NY, 2006.

ISBN: 978-989-95079-6-8

Fuzzy Arithmetic with Parametric LR Fuzzy Numbers

Luciano Stefanini, Laerte Sorini

DEMQ (Department of Economics and Quantitative Methods)
University of Urbino 'Carlo Bo', Italy.
Email: lucste@uniurb.it (L. Stefanini), laerte@uniurb.it (L. Sorini)

Abstract— *In this paper we suggest and describe a new family of parametric representations for LR fuzzy numbers and use them for fuzzy calculations and arithmetic.*

Keywords— LR Fuzzy Numbers, LU-Fuzzy Parametrization, Fuzzy Arithmetic.

1 Introduction

The arithmetical structure of fuzzy numbers have been developed in the 1980's and Dubois and Prade introduced the well known LR model and the corresponding formulas for the fuzzy operations (see the recent publication [1] and the references therein).

In general, the arithmetic operations on fuzzy numbers can be approached either by the direct use of the membership function (by the Zadeh extension principle) or by the equivalent use of the $\alpha - cuts$ representation. A solid result in fuzzy theory and practice is that calculations cannot be performed by using the same rules as in arithmetic with real numbers and in fact fuzzy calculus will not always satisfy the same properties (e.g. distributivity, invertibility and others).

A parametric representation of LR fuzzy numbers and the derivation of the corresponding arithmetic operator have been first introduced by Giachetti and Young (see [2], [3]). In this paper, we will see that it is possible to define a more flexible parametric representation of fuzzy numbers that allow a large variety of possible shapes and is very simple to implement (see [5] and [6] for a parametric $\alpha - cuts$ representation).

We suggest a parametrization for the LR fuzzy numbers, similar to the work done for the LU parametrization in [5] and [6]. In terms of the parameters representing the LR fuzzy numbers, it is possible to define the operators for the fuzzy arithmetic and we show that the errors of the approximations can be reduced to any small tolerance (by increasing the number of parameters); in fact, within the space of differentiable fuzzy numbers, the approximations form a dense subspace.

2 Fuzzy Numbers in LR Parametric Form

We will consider fuzzy quantities, i.e. fuzzy sets defined over the field $\mathbb{R}$ of real numbers and we will focus on fuzzy numbers, having a particular form of the membership function.

A general *fuzzy set* over a given set (or space) $\mathbb{X}$ of elements (the universe) is usually defined by its membership function $\mu : \mathbb{X} \longrightarrow \mathbb{T} \subseteq [0,1]$ and a fuzzy (sub)set u of $\mathbb{X}$ is uniquely characterized by the pairs $(x, \mu_u(x))$ for each $x \in \mathbb{X}$; the value $\mu_u(x) \in [0,1]$ is the membership grade of x to the fuzzy set u. So, a fuzzy set is given by

$$\begin{aligned} u &= \{(x, \mu_u(x)) | x \in \mathbb{X}\} \\ \text{given } \mu_u &: \mathbb{X} \longrightarrow [0,1] \end{aligned}$$

Fundamental concepts in fuzzy theory are the *support*, the *level-sets* (or *level-cuts*) and the *core* of a fuzzy set (or of its membership function): let μ_u be the membership function of a fuzzy set u over $\mathbb{X}$. The support of u is the (crisp) subset of points of $\mathbb{X}$ at which the membership grade $\mu_u(x)$ is positive:

$$supp(u) = \{x | x \in \mathbb{X}, \mu_u(x) > 0\}; \tag{1}$$

we always assume that $supp(u) \neq \emptyset$. For $\alpha \in]0,1]$, the $\alpha-$level cut of u (or simply the $\alpha - cut$) is defined by

$$[u]_\alpha = \{x | x \in \mathbb{X},\ \mu_u(x) \geq \alpha\} \tag{2}$$

and for $\alpha = 0$ (or $\alpha \to +0$) by the closure of the support

$$[u]_0 = cl\{x | x \in \mathbb{X},\ \mu_u(x) > 0\}.$$

The core of u is the set of elements of $\mathbb{X}$ having membership grade 1

$$core(u) = \{x | x \in \mathbb{X}, \mu_u(x) = 1\} \tag{3}$$

and we say that u is normal if $core(u) \neq \emptyset$.

Of our interest are fuzzy sets when the space $\mathbb{X}$ is $\mathbb{R}$ (unidimensional real fuzzy sets).

Well-known properties of the $level - cuts$ are:

$$[u]_\alpha \subseteq [u]_\beta \text{ for } \alpha > \beta, \tag{4}$$

$$[u]_\alpha = \bigcap_{\beta < \alpha} [u]_\beta \text{ for } \alpha \in]0,1] \tag{5}$$

and (if $x \in supp(u)$, otherwise $\mu_u(x) = 0$)

$$\mu_u(x) = \sup\{\alpha |\ \alpha \in]0,1] \text{ for which } x \in [u]_\alpha\}. \tag{6}$$

A particular class of fuzzy sets is when the support is a convex set (C is said *convex* if $(1-t)x' + tx'' \in C$ for every $x', x'' \in C$ and all $t \in [0,1]$) and the membership function is quasi-concave: we say that the membership function μ_u is quasi-concave if $\mu_u((1-t)x' + tx'') \geq \min\{\mu_u(x'), \mu_u(x'')\}$ for every $x', x'' \in supp(u)$ and $t \in [0,1]$. Equivalently, μ_u is quasi-concave if the level-cuts $[u]_\alpha$ are convex sets for all $\alpha \in [0,1]$.

A third property of the fuzzy quantities is related to the semi-continuity of the membership function and to the closedness of the level-cuts: μ_u is said to be upper semicontinuous if $\lim_{x \longrightarrow \widehat{x}} \sup \mu_u(x) = \mu_u(\widehat{x})$ at every $\widehat{x} \in supp(u)$ or, equivalently, if the level-cuts $[u]_\alpha$ are closed sets for all $\alpha \in [0,1]$.

Definition (*fuzzy number):* u is a fuzzy number if

(i.) μ_u has nonempty bounded support and is normal,

(ii.) μ_u is quasi-concave,

(iii.) μ_u is upper semicontinuous

ISBN: 978-989-95079-6-8

or, equivalently:

(a.) $[u]_\alpha$ are nonempty convex sets for all $\alpha \in [0,1]$

(b.) $[u]_\alpha$ are closed and bounded (compact) sets for all $\alpha \in [0,1]$

(c.) $[u]_\alpha$ satisfy conditions (4) and (5).

We will denote by $\mathcal{F}$ the set of fuzzy numbers. It can be structured by an addition and a scalar multiplication, defined either by the level sets or, equivalently, by the Zadeh extension principle. Let $u, v \in \mathcal{F}$ have membership functions μ_u, μ_v and $\alpha - cuts$ $[u]_\alpha$, $[v]_\alpha$, $\alpha \in [0,1]$ respectively. The addition $u+v \in \mathcal{F}$ and the scalar multiplication $ku \in \mathcal{F}$ for $k \in \mathbb{R}\backslash\{0\}$ have membership functions (extension principle)

$$\mu_{u+v}(z) = \sup\{\min\{\mu_u(x), \mu_v(y)\} | z = x+y\} \quad (7)$$

$$\mu_{ku}(x) = \mu_u(\frac{x}{k}) \quad (8)$$

and level cuts

$$[u+v]_\alpha = [u]_\alpha + [v]_\alpha \quad (9)$$
$$= \{x+y | x \in [u]_\alpha,\ y \in [v]_\alpha\}$$
$$[ku]_\alpha = k[u]_\alpha = \{kx | x \in [u]_\alpha\}. \quad (10)$$

We denote by $[u_\alpha^-, u_\alpha^+], \forall \alpha \in [0,1]$ the level-cuts of $u \in \mathcal{F}$ and by $[\widehat{u}^-, \widehat{u}^+]$ its core (level set at $\alpha = 1$). If $u_\alpha^- = \widehat{u}^-$ and $u_\alpha^+ = \widehat{u}^+, \forall \alpha \in [0,1]$ we have a crisp interval or a crisp number. We say that u is positive if $u_\alpha^- > 0$, $\forall \alpha \in [0,1]$ and that u is negative if $u_\alpha^+ < 0, \forall \alpha \in [0,1]$.

Definition (*LR-fuzzy number*): An LR-fuzzy number (or interval) u has membership function of the form

$$\mu_u(x) = \begin{cases} L(\frac{x-a}{b-a}) & \text{if } x \in [a,b] \\ 1 & \text{if } x \in [b,c] \\ R(\frac{d-x}{d-c}) & \text{if } x \in [c,d] \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where $L, R : [0,1] \longrightarrow [0,1]$ are two non-decreasing *shape functions* such that $R(0) = L(0) = 0$ and $R(1) = L(1) = 1$. If $b = c$ we obtain a fuzzy number. The support is the compact interval $[a,d]$ and the core is $[b,c]$.

If L and R are invertible functions, then the $\alpha - cuts$ are obtained by

$$[u]_\alpha = [u_\alpha^-, u_\alpha^+] \quad (12)$$
$$\text{with}$$
$$u_\alpha^- = a + (b-a)L^{-1}(\alpha)$$
$$u_\alpha^+ = d - (d-c)R^{-1}(\alpha)$$

The usual LR-fuzzy notation is $u = \langle a,b,c,d\rangle_{L,R}$ for an interval and $u = \langle a,b,c\rangle_{L,R}$ for a number. We refer to functions $L(.)$ and $R(.)$ as the left and right branches (shape functions) of u, respectively and to the functions $u_{(.)}^-$ and $u_{(.)}^+$ as the lower and upper branches on u, respectively. Note that, in the representation (12), we have $u_0^- = a$, $u_1^- = b$, $u_1^+ = c$ and $u_0^+ = d$.

2.1 Parametric LR-fuzzy numbers

In this section we present the basic elements of a parametric representation of the shape functions based on monotonic Hermite-type interpolation. We first introduce some models for differentiable monotonic shape functions $p : [0,1] \longrightarrow [0,1]$ such that $p(0) = 0$ and $p(1) = 1$ with $p(t)$ differentiable and increasing on $[0,1]$; for any given nonnegative parameters $\beta_i \geq 0$, $i = 0,1$ we consider functions $p : [0,1] \longrightarrow [0,1]$ satisfying the four Hermite interpolation conditions

$$p(0) = 0,\ p(1) = 1$$
$$p'(0) = \beta_0,\ p'(1) = \beta_1.$$

We obtain infinite many functions simply by fixing the two parameters β_i that give the slopes (first derivatives) of the function at $t = 0$ and $t = 1$. To explicit the slope parameters we denote the interpolating function p by

$$t \longrightarrow p(t; \beta_0, \beta_1) \text{ for } t \in [0,1].$$

We recall here two of the basic forms from [5]:

∘ (2,2)-rational spline:

$$p(t; \beta_0, \beta_1) = \frac{t^2 + \beta_0 t(1-t)}{1 + (\beta_0 + \beta_1 - 2)t(1-t)}; \quad (13)$$

∘ mixed exponential spline:

$$p(t; \beta_0, \beta_1) = \frac{1}{m}[t^2(3-2t) - \beta_0(1-t)^m + \beta_0 + \beta_1 t^m] \quad (14)$$
$$\text{where } m = 1 + \beta_0 + \beta_1.$$

Both functions p in (13) and (14) are strictly increasing on [0,1] and we obtain a linear $p(t) = t, \forall t \in [0,1]$ if $\beta_0 = \beta_1 = 1$ and a quadratic $p(t) = t^2 + \beta_0 t(1-t)$ if $\beta_0 + \beta_1 = 2$ with $\beta_0 \neq 1$.

The (parametric) monotonic functions like (13) and (14) can be used as models for the shape functions L and R; in fact, if $a \leq b \leq c \leq d$ and $\beta_{0,L}, \beta_{1,L} \geq 0$, $\beta_{0,R}, \beta_{1,R} \geq 0$ are given, we obtain an LR-fuzzy number with the membership function

$$\mu_u(x) = \begin{cases} p(\frac{x-a}{b-a}; \beta_{0,L}, \beta_{1,L}) & \text{if } x \in [a,b] \\ 1 & \text{if } x \in [b,c] \\ p(\frac{x-d}{c-d}; \beta_{0,R}, \beta_{1,R}) & \text{if } x \in [c,d] \\ 0 & \text{otherwise} \end{cases}. \quad (15)$$

To make a uniform notation,we denote a, b, c, d as $a = u_{0,L}$, $b = u_{1,L}$, $c = u_{1,R}$, $d = u_{0,R}$ and $\beta_{0,L}^* = \frac{\beta_{0,i}}{b-a}$, $\beta_{1,L}^* = \frac{\beta_{1,i}}{b-a}$, $\beta_{0,R}^* = \frac{\beta_{0,j}}{c-d}$, $\beta_{1,R}^* = \frac{\beta_{1,j}}{c-d}$; so, the eight parameters completely define u:

$$u = (u_{0,L}, \beta_{0,L}^*, u_{0,R}, \beta_{0,R}^*; u_{1,L}, \beta_{1,L}^*, u_{1,R}, \beta_{1,R}^*) \quad (16)$$

provided that $u_{0,L} \leq u_{1,L} \leq u_{1,R} \leq u_{0,R}$ and $\beta_{0,L}, \beta_{1,L} \geq 0$, $\beta_{0,R}, \beta_{1,R} \geq 0$.

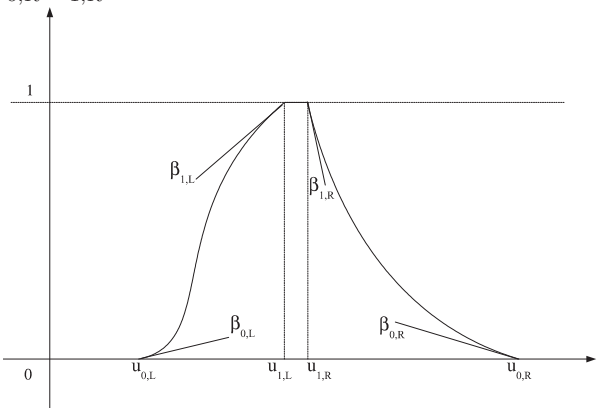


Figure 1: Typical LR parametric membership function

ISBN: 978-989-95079-6-8

Remark: As the nonnegative parameters $\beta_{i,L}$ and $\beta_{i,R}$ ($i = 1, 2$) are completely independent, each side of the membership function may have independent (monotonic) forms, e.g. concavity, convexity, inflextions etc. of left and right sides are independent, as illustrated in the figure. The shapes are simply "controlled" by the 4 parameters $\beta_{i,L}$ and $\beta_{i,R}$ ($i = 1, 2$).

Denote by $\mathbb{F}^{LR}$ the set of LR-fuzzy numbers defined by (16). The family of fuzzy numbers $\mathbb{F}^{LR}$, which include triangular and trapezoidal fuzzy numbers if all $\beta_{i,L} = \beta_{i,R} = 1$, are characterized by eight parameters and it appears that the inclusion of the slopes of left and right functions, even without generating piecewise monotonic approximations over subintervals is able to capture much more information than the linear approximation (see [5] and [6]).

The level sets included into the representation are the core and the support. Note that if we are interested to represent exactly the core, the support and the midpoint interval we require $N = 2$, $\alpha_0 = 0$, $\alpha_1 = 0.5$, $\alpha_2 = 1$, the support $[u_{0,L}, u_{0,R}]$ (corresponding to $\alpha = \alpha_0$), the midpoint interval $[u_{1,L}, u_{1,R}]$ (corresponding to $\alpha = \alpha_1$), the core $[u_{2,L}, u_{2,R}]$ (corresponding to $\alpha = \alpha_0$) and the six parameters $\beta_{i,L}, \beta_{i,R} \geq 0$ to model the first derivatives of $\mu_u(x)$ at the extremal point of the three intervals above.

More generally, we can design differentiable parametric shape functions $L()$ and $R()$ by fixing $N + 1$ distinct level sets $0 = \alpha_0 < \alpha_1 < ... < \alpha_N = 1$ and assigning, for each α_i ($i = 0, 1, ..., N$), the four parameters needed for each level.

In this general case, we need the values $u_{i,L}$, $u_{i,R}$ but, instead of the parameters $\beta_{i,L}, \beta_{i,R}$, it is convenient to give directly the first derivatives of $\mu_u(x)$ at $u_{i,L}$, $u_{i,R}$ and calculate the $\beta_{i,L}, \beta_{i,R} \geq 0$ to use in the monotonic functions like (13) and (14) and model the membership function like (15) piecewise on each subinterval $[u_{i-1,L}, u_{i,L}]$ and $[u_{i,R}, u_{i-1,R}]$.

For $i = 0, 1, ..., N$, denote directly by $\delta u_{i,L} \geq 0$ and by $\delta u_{i,R} \leq 0$ the first derivatives of $\mu_u(x)$ at the points $x = u_{i,L}$ and $x = u_{i,R}$, respectively and suppose they are given. The membership function $\mu_u(x)$ is obtained by the following simple procedure:

Assign first $\mu_u(x) = 0$ for all x;
for all $x \in [u_{N,L}, u_{N,R}]$ set $\mu_u(x) = 1$;
for $i = 1, 2, ..., N$
  for $x \in [u_{i-1,L}, u_{i,L}]$ set
    $\beta_0 = \frac{u_{ni} - u_{n-1,i}}{\alpha_n - \alpha_{n-1}} \delta u_{i-1,L}$
    $\beta_1 = \frac{u_{ni} - u_{n-1,i}}{\alpha_n - \alpha_{n-1}} \delta u_{i,L}$
    $\mu_u(x) = \alpha_{i-1} + (\alpha_i - \alpha_{i-1}) p(\frac{x - u_{n-1,i}}{u_{ni} - u_{n-1,i}}; \beta_0, \beta_1)$
  end
  for $x \in [u_{i,R}, u_{i-1,R}]$ set
    $\beta_0 = \frac{u_{nj} - u_{n-1,j}}{\alpha_n - \alpha_{n-1}} \delta u_{i-1,R}$
    $\beta_1 = \frac{u_{nj} - u_{n-1,j}}{\alpha_n - \alpha_{n-1}} \delta u_{i,R}$
    $\mu_u(x) = \alpha_{i-1} + (\alpha_i - \alpha_{i-1}) p(\frac{x - u_{n-1,j}}{u_{nj} - u_{n-1,j}}; \beta_0, \beta_1)$
  end
end

Denote by $\mathbb{F}_N^{LR}$ the fuzzy numbers obtained in the form above; as each shape function is monotonic, the left and right branches are monotonic increasing on $[u_{0,L}, u_{N,L}]$ (the left of μ_u) and decreasing on $[u_{N,R}, u_{0,R}]$ (the right of μ_u). The number of parameters is $4N + 4$ and they are simply constrained to be $u_{0,L} \leq ... \leq u_{N,L} \leq u_{N,R} \leq ... \leq u_{0,R}$ and $\delta u_{i,L} \geq 0$ and $\delta u_{i,R} \leq 0$ to ensure monotonicity. In general, we will have $u_{i-1,L} < u_{i,L}$ and $u_{i-1,R} > u_{i,R}$ but it is easy to consider case of equality so that the graph of μ_u is a vertical line (discontinuity) at $u_{i,.}$ if $u_{i-1,.} = u_{i,.}$.

In this general form, an LR fuzzy number is represented as follows:

$$u = (\alpha_i; u_{i,L}, \delta u_{i,L}, u_{i,R}, \delta u_{i,R})_{i=0,1,...,N}. \qquad (17)$$

Example:

Consider a quasi-Gaussian membership function ($m \in \mathbb{R}$, $k, \sigma \in \mathbb{R}^+$; if $k \to +\infty$ the support is unbounded)

$$\mu_{qG}(x) = \begin{cases} \exp(-\frac{(x-m)^2}{2\sigma^2}) & \text{if } m - k\sigma \leq x \leq m + k\sigma \\ 0 & \text{otherwise.} \end{cases} \qquad (18)$$

Its LR parametrization for $m = 0$, $\sigma = 2$, $k = 4$, approximated with $N = 4$ (five points), is

Table 1. LR parametrization of fuzzy number (18)

α_i	$u_{i,L}$	$\delta u_{i,L}$	$u_{i,R}$	$\delta u_{i,R}$
0.0	-8.0	0.00033	8.0	-0.00033
0.25	-3.33022	0.20814	3.33022	-0.20814
0.5	-2.35482	0.29435	2.35482	-0.29435
0.75	-1.51705	0.28445	1.51705	-0.28445
1.0	0.0	0.0	0.0	0.0

It is interesting to note that, in our parametric representation (17) of LR-fuzzy numbers, the membership function has differentiable left and right branches; if we require that $\mu_u(x)$ be differentiable in all the (internal) points of the support of u, in particular also at $x = u_{N,L}$ and $x = u_{N,R}$, we simply need to set $\delta u_{N,L} = 0$ and $\delta u_{N,R} = 0$ in the parametric representation (17). This seems to be an important advantage in all the applications (e.g. fuzzy intelligent systems, neuro-fuzzy learning models) where overall differentiability of μ_u is required.

The generated LR fuzzy numbers form a subspace of the space of fuzzy numbers. In the case of differentiable membership functions, it is immediate to understand that the union of all parametric fuzzy numbers having the form (17) for all integer $N \geq 1$, i.e.

$$\mathcal{F}^{LR} = \bigcup_{N \geq 1} \mathbb{F}_N^{LR},$$

is dense into the space of (differentiable) fuzzy numbers. In fact, each fuzzy number (17) has the property of interpolate (exactly) the values $u_{i,L}, u_{i,R}$ ($i = 0, 1, ..., N$) and it is sufficient to refine the points α_i to obtain any desired precision. In our experience, approximations with values of N from 5 to 20 have in general a very small error, of the order of $10^{-6} - 10^{-2}$.

Following [4], for $p \in [0, \infty]$, we can define a geometric distance $D_p(u, v)$ between fuzzy numbers $u, v \in \mathbb{F}_N^{LR}$, given by

$$D_p^{LR}(u, v) = \left(\sum_{i=0}^{N} |u_{i,L} - v_{i,L}|^p + |u_{i,R} - v_{i,R}|^p \right)^{1/p} + \\ + \left(\sum_{i=0}^{N} |\delta u_{i,L} - \delta v_{i,L}|^p + |\delta u_{i,R} - \delta v_{i,R}|^p \right)^{1/p}$$

and

ISBN: 978-989-95079-6-8

$$D_\infty^{LR}(u,v) = \max_{i=0,\dots,N}\{|u_{i,L}-v_{i,L}|, |u_{i,R}-v_{i,R}|\} + \\ + \max_{i=0,\dots,N}\{|\delta u_{i,L}-\delta v_{i,L}|, |\delta u_{i,R}-\delta v_{i,R}|\}.$$

If we model the LR-fuzzy numbers by a (2,2)-rational spline $p(\alpha;\beta_0,\beta_1)$ like (13) the inverse $p^{-1}(t;\beta_0,\beta_1)$ can be computed analytically as we have to solve the quadratic equation (with respect to α)

$$\alpha^2+\beta_0\alpha(1-\alpha) = t[1+(\beta_0+\beta_1-2)\alpha(1-\alpha)] \text{ i.e.}$$

$$\begin{aligned}(1+A(t))\alpha^2 - A(t)\alpha - t &= 0 \text{ where}\\ A(t) &= -\beta_0+\beta_0 t+\beta_1 t-2t.\end{aligned}$$

If $A(t)=-1$ then the equation is linear and the solution is $\alpha=t$. If $A(t)\neq -1$, then there exist two real solutions $\alpha_1 = \frac{2\sqrt{t}}{2+2A(t)}$, $\alpha_2 = \frac{2\sqrt{t}+2A(t)}{2+2A(t)}$ and we choose the one belonging to $[0,1]$.

3 Fuzzy arithmetic with LR parametrization

The fuzzy extension principle introduced by Zadeh is the basic tool for fuzzy calculus; it extends functions of real numbers to functions of fuzzy numbers and it allows the extension of arithmetic operations and calculus to fuzzy arguments. We have already defined the addition (7) and the scalar multiplication (8).

3.1 Basic arithmetic operators

Let $\circ \in \{+,-,\times,/\}$ one of the four arithmetic operations and let $u,v\in\mathbb{F}_{\mathbb{I}}$ be given fuzzy intervals (or numbers), having $\mu_u(.)$ and $\mu_v(.)$ as membership functions and level-cuts representations $u=(u^-,u^+)$, $v=(v^-,v^+)$; the extension principle for the extension of $\circ$ defines the membership function of $w=u\circ v$ by

$$\mu_{u\circ v}(z) = \sup\{\min\{\mu_u(x),\mu_v(y)\}|z=x\circ y\}. \qquad (19)$$

In terms of the $\alpha-cuts$, the four arithmetic operations and the scalar multiplication for $k\in\mathbb{R}$ are obtained by the well-known *interval arithmetic* (for all $\alpha\in[0,1]$)

Addition ($w=u+v$):
$[u+v]_\alpha = [u_\alpha^- + v_\alpha^-, u_\alpha^+ + v_\alpha^+]$,

Scalar multiplication ($w=ku$):
$[ku]_\alpha = [\min\{ku_\alpha^-, ku_\alpha^+\}, \max\{ku_\alpha^-, ku_\alpha^+\}]$,

Subtraction ($w=u-v$):
$[u-v]_\alpha = [u_\alpha^- - v_\alpha^+, u_\alpha^+ - v_\alpha^-]$,

Multiplication ($w=uv$):
$$\begin{cases}(uv)_\alpha^- = \min\{u_\alpha^-v_\alpha^-, u_\alpha^-v_\alpha^+, u_\alpha^+v_\alpha^-, u_\alpha^+v_\alpha^+\}\\ (uv)_\alpha^+ = \max\{u_\alpha^-v_\alpha^-, u_\alpha^-v_\alpha^+, u_\alpha^+v_\alpha^-, u_\alpha^+v_\alpha^+\}\end{cases},$$

Division ($w=u/v$): if $0\notin[v_0^-, v_0^+]$
$$\begin{cases}(u/v)_\alpha^- = \min\left\{\frac{u_I^-}{v_I^-}, \frac{u_I^-}{v_I^+}, \frac{u_I^+}{v_I^-}, \frac{u_I^+}{v_I^+}\right\}\\ (u/v)_\alpha^+ = \max\left\{\frac{u_I^-}{v_I^-}, \frac{u_I^-}{v_I^+}, \frac{u_I^+}{v_I^-}, \frac{u_I^+}{v_I^+}\right\}\end{cases}.$$

In recent work (see [7]) the following generalized subtraction has been defined and analyzed:

Definition: Let u,v be two fuzzy numbers; we define the generalized Hukuhara difference (gH-difference for short) of u and v as the fuzzy number w such that

$$u\ominus_g v = w \iff \begin{cases}(i) & u=v+w\\ \text{or} & (ii)\quad v=u-w\end{cases}. \qquad (20)$$

The α-cuts of both $u\ominus_g v$ can be expressed in terms of the α-cuts of u and v; but it is possible that $u\ominus_g v$ is not well defined; by the use of the parametrization (17) it is easy to test if the operation is well defined or not.

Consider two LR-fuzzy numbers u and v ($N=1$ for simplicity)

$$\begin{aligned}u &= (u_{0,L},\delta u_{0,L},u_{0,R},\delta u_{0,R}; \qquad (21)\\ &\quad u_{1,L},\delta u_{1,L},u_{1,R},\delta u_{1,R}),\\ v &= (v_{0,L},\delta v_{0,L},v_{0,R},\delta v_{0,R}; \qquad (22)\\ &\quad v_{1,L},\delta v_{1,L},v_{1,R},\delta v_{1,R}).\end{aligned}$$

Note that in the formulas below u and v are not restricted to have the same $L(.)$ and $R(.)$ shape functions and changing the slopes will change the form of the membership functions (in all the cases below, eventually assume $\frac{0}{0}=0$).

The values and the slopes at the considered α-cuts are computed exactly; in fact we use the exact standard formulae for the derivative of addition, multiplication, division and function composition. Corresponding to membership values $\alpha\neq\alpha_i$, $i=0,1,...,N$, the membership functions is approximated by using the rational (13) or the mixed (14) shape functions.

The approximate addition is the following

$$(u+v) = \left(\begin{array}{l} u_{0,L}+v_{0,L}, \frac{\delta u_{0,i}\,\delta v_{0,i}}{\delta u_{0,i}+\delta v_{0,i}},\\ u_{0,R}+v_{0,R}, \frac{\delta u_{0,j}\,\delta v_{0,j}}{\delta u_{0,j}+\delta v_{0,j}};\\ u_{1,L}+v_{1,L}, \frac{\delta u_{1,i}\,\delta v_{1,i}}{\delta u_{1,i}+\delta v_{1,i}},\\ u_{1,R}+v_{1,R}, \frac{\delta u_{1,j}\,\delta v_{1,j}}{\delta u_{1,j}+\delta v_{1,j}}\end{array}\right).$$

Nothe that if the left and right shapes of u and v are the same (e.g. linear, quadratic) then the addition is exact.

The general algorithm for approximate LR addition is
Algorithm (*LR addition*) $w=u+v$
for $i=0,1,...,N$
$w_{i,L}=u_{i,L}+v_{i,L}$, $w_{i,R}=u_{i,R}+v_{i,R}$
$\delta w_{i,L}=\frac{\delta u_{n_i}\,\delta v_{n_i}}{\delta u_{n_i}+\delta v_{n_i}}$, $\delta w_{i,R}=\frac{\delta u_{n_j}\,\delta v_{n_j}}{\delta u_{n_j}+\delta v_{n_j}}$
end

The difference $w=u-v$ is similar
Algorithm (*LR subtraction*) $w=u-v$
for $i=0,1,...,N$
$w_{i,L}=u_{i,L}-v_{i,R}$, $w_{i,R}=u_{i,R}-v_{i,L}$
$\delta w_{i,L}=\frac{\delta u_{n_i}\,\delta v_{n_j}}{\delta v_{n_j}-\delta u_{n_j}}$, $\delta w_{i,R}=\frac{\delta u_{n_j}\,\delta v_{n_i}}{\delta v_{n_i}-\delta u_{n_j}}$
end

Also the approximate multiplication $w=uv$ can be obtained easily; in the case of two positive LR-fuzzy numbers w is given by

$$w_{LR} = \left(\begin{array}{l} u_{0,L}v_{0,L}, \frac{\delta u_{0,i}\,\delta v_{0,i}}{v_{0,i}\,\delta v_{0,i}+u_{0,i}\,\delta u_{0,i}},\\ u_{0,R}v_{0,R}, \frac{\delta u_{0,j}\,\delta v_{0,j}}{v_{0,j}\,\delta v_{0,j}+u_{0,j}\,\delta u_{0,j}};\\ u_{1,L}v_{1,L}, \frac{\delta u_{1,i}\,\delta v_{1,i}}{v_{1,i}\,\delta v_{1,i}+u_{1,i}\,\delta u_{1,i}},\\ u_{1,R}v_{1,R}, \frac{\delta u_{1,j}\,\delta v_{1,j}}{v_{1,j}\,\delta v_{1,j}+u_{1,j}\,\delta u_{1,j}}\end{array}\right).$$

ISBN: 978-989-95079-6-8

The general algorithm for approximate LR multiplication is

Algorithm (*LR multiplication*) $w = uv$
(eventually, use $\frac{0}{0} = 0$)
for $i = 0, 1, ..., N$
$m_i = \min\{u_{i,L}v_{i,L}, u_{i,L}v_{i,R}, u_{i,R}v_{i,L}, u_{i,R}v_{i,R}\}$
$M_i = \max\{u_{i,L}v_{i,L}, u_{i,L}v_{i,R}, u_{i,R}v_{i,L}, u_{i,R}v_{i,R}\}$
$w_{i,L} = m_i \ , \ w_{i,R} = M_i$
if $u_{i,L}v_{i,L} = m_i$ **then** $\delta w_{i,L} = \frac{\delta u_{n_i}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} + u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,L}v_{i,R} = m_i$ **then** $\delta w_{i,L} = \frac{\delta u_{n_i}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} + u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,R}v_{i,L} = m_i$ **then** $\delta w_{i,L} = \frac{\delta u_{n_j}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} + u_{n_j}\, \delta u_{n_j}}$
elseif $u_{i,R}v_{i,R} = m_i$ **then** $\delta w_{i,L} = \frac{\delta u_{n_j}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} + u_{n_j}\, \delta u_{n_j}}$
endif
if $u_{i,L}v_{i,L} = M_i$ **then** $\delta w_{i,R} = \frac{\delta u_{n_i}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} + u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,L}v_{i,R} = M_i$ **then** $\delta w_{i,R} = \frac{\delta u_{n_i}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} + u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,R}v_{i,L} = M_i$ **then** $\delta w_{i,R} = \frac{\delta u_{n_j}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} + u_{n_j}\, \delta u_{n_j}}$
elseif $u_{i,R}v_{i,R} = M_i$ **then** $\delta w_{i,R} = \frac{\delta u_{n_j}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} + u_{n_j}\, \delta u_{n_j}}$
endif
end

The division is similar:

Algorithm (*LR division*) $w = u/v$, $0 \notin [v]_0$
(eventually, use $\frac{0}{0} = 0$)
for $i = 0, 1, ..., N$
$m_i = \min\{u_{i,L}/v_{i,L}, u_{i,L}/v_{i,R}, u_{i,R}/v_{i,L}, u_{i,R}/v_{i,R}\}$
$M_i = \max\{u_{i,L}/v_{i,L}, u_{i,L}/v_{i,R}, u_{i,R}/v_{i,L}, u_{i,R}/v_{i,R}\}$
$w_{i,L} = m_i \ , \ w_{i,R} = M_i$
if $u_{i,L}/v_{i,L} = m_i$ **then** $\delta w_{i,L} = \frac{v_{n_i}^2\, \delta u_{n_i}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} - u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,L}/v_{i,R} = m_i$ **then** $\delta w_{i,L} = \frac{v_{n_j}^2\, \delta u_{n_i}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} - u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,R}/v_{i,L} = m_i$ **then** $\delta w_{i,L} = \frac{v_{n_i}^2\, \delta u_{n_j}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} - u_{n_j}\, \delta u_{n_j}}$
elseif $u_{i,R}/v_{i,R} = m_i$ **then** $\delta w_{i,L} = \frac{v_{n_j}^2\, \delta u_{n_j}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} - u_{n_j}\, \delta u_{n_j}}$
endif
if $u_{i,L}/v_{i,L} = M_i$ **then** $\delta w_{i,R} = \frac{v_{n_i}^2\, \delta u_{n_i}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} - u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,L}/v_{i,R} = M_i$ **then** $\delta w_{i,R} = \frac{v_{n_j}^2\, \delta u_{n_i}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} - u_{n_i}\, \delta u_{n_i}}$
elseif $u_{i,R}/v_{i,L} = M_i$ **then** $\delta w_{i,R} = \frac{v_{n_i}^2\, \delta u_{n_j}\, \delta v_{n_i}}{v_{n_i}\, \delta v_{n_i} - u_{n_j}\, \delta u_{n_j}}$
elseif $u_{i,R}/v_{i,R} = M_i$ **then** $\delta w_{i,R} = \frac{v_{n_j}^2\, \delta u_{n_j}\, \delta v_{n_j}}{v_{n_j}\, \delta v_{n_j} - u_{n_j}\, \delta u_{n_j}}$
endif
end

The algorithm for the gH-difference $w = u \ominus_g v$ in LR form is also easy, if we recall that the α-cuts of w are the intervals (see [7]) $[\min\{u_\alpha^- - v_\alpha^-, u_\alpha^+ - v_\alpha^+\}, \max\{u_\alpha^- - v_\alpha^-, u_\alpha^+ - v_\alpha^+\}]$:

Algorithm (*LR gH-subtraction*) $w = u \ominus_g v$
(eventually, use $\frac{0}{0} = 0$)
for $i = 0, 1, ..., N$
define $d_{i,L} = u_{i,L} - v_{i,L}$ and $d_{i,R} = u_{i,R} - v_{i,R}$
if $d_{i,L} < d_{i,R}$ **then**
$w_{i,L} = d_{i,L}$, $w_{i,R} = d_{i,R}$
$\delta w_{i,L} = \frac{\delta u_{n_i}\, \delta v_{n_i}}{\delta v_{n_i} - \delta u_{n_i}}$, $\delta w_{i,R} = \frac{\delta u_{n_j}\, \delta v_{n_j}}{\delta v_{n_j} - \delta u_{n_j}}$
elseif $d_{i,L} > d_{i,R}$ **then**
$w_{i,R} = d_{i,L}$, $w_{i,L} = d_{i,R}$
$\delta w_{i,R} = \frac{\delta u_{n_i}\, \delta v_{n_i}}{\delta v_{n_i} - \delta u_{n_i}}$, $\delta w_{i,L} = \frac{\delta u_{n_j}\, \delta v_{n_j}}{\delta v_{n_j} - \delta u_{n_j}}$
else (here $d_{i,L} = d_{i,R} = d$)
$w_{i,L} = d$, $w_{i,R} = d$
$\delta w_{i,L} = 0$, $\delta w_{i,R} = 0$
endif
end
Test if the calculated $w_{i,L}, w_{i,R}$ are ordered correctly and $\delta w_{i,L} \geq 0, \delta w_{i,R} \leq 0$, then w is $u \ominus_g v$.

The operations above are exact at the nodes $\alpha = \alpha_i$, $i = 0, 1, ..., N$, and have very small global errors for all $\alpha \in [0, 1]$ (if N is sufficiently high, of the order of 5 to 20). Further, it is easy to control the error by using a sufficiently fine $\alpha-$decomposition and the results have shown that both the rational (13) and the mixed (14) models perform well.

Some parametric membership functions in the LR framework are present in many applications and the use of nonlinear shapes in increasing. Usually, one defines a given family, e.g. linear, quadratic, sigmoid, quasi gaussian, and the operations are performed within the same family.

Our proposed parametrization allows an extended set of flexible fuzzy numbers and is able to approximate all other forms with acceptable small errors and the additional advantage of producing good approximations to the results of the arithmetic operations even between LR fuzzy numbers having very different original shapes.

Furthermore, at the explicitly considered $\alpha-$cuts of the decomposition, all the values $w_{i,L}$, $w_{i,R}$ and the slopes $\delta w_{i,L}$, $\delta w_{i,R}$ $(i = 0, 1, ..., N)$ of the membership function are exact.

3.2 Computation of fuzzy-valued functions

For a continuous function $f : \mathbb{R}^n \to \mathbb{R}$, let $v = f(u_1, u_2, ..., u_n)$ denote its fuzzy extension, based on the application of the Zadeh's Extension Principle (EP for short). It is well known that the $\alpha - cuts$ $[v_\alpha^-, v_\alpha^+]$ of v are obtained by solving the following box-constrained global optimization problems ($\alpha \in [0, 1]$)

$$v_\alpha^- = \min\{f(x_1, ..., x_n) | x_k \in [u_k]_\alpha,\ k = 1, 2, ..., n\} \quad (23)$$

$$v_\alpha^+ = \max\{f(x_1, ..., x_n) | x_k \in [u_k]_\alpha,\ k = 1, 2, ..., n\} \quad (24)$$

where $[u_k]_\alpha = [u_{k,\alpha}^-, u_{k,\alpha}^+]$, $k = 1, 2, ..., n$, are the $\alpha - cuts$ of u_k.

The lower and upper values v_α^- and v_α^+ of v define equivalently (as f is assumed to be continuous) the image of the cartesian product $\underset{k=1}{\overset{n}{\times}} [u_k]_\alpha$ via f, i.e. $[v_\alpha^-, v_\alpha^+] = f([u_1]_\alpha, ..., [u_n]_\alpha)$.

Except for simple elementary cases for which the optimization problems above can be solved analytically, the direct application of (23) and (24) is difficult and computationally expensive (see [6]).

At least if f is differentiable, the advantages of the LR representation appear to be quite interesting, based on the fact that a small number of α points is in general sufficient to obtain good approximations (this is the essential gain in using the slopes to model fuzzy numbers), so reducing the number of constrained *min* (23) and *max* (24) problems to be solved directly. On the other hand, finding computationally efficient extension solvers is still an open research field in fuzzy calculations.

We now give the details of the fuzzy extension of general differentiable functions of only one variable, by the LR representation. The case of multidimensional differentiable functions can be approached in a similar way, by considering the

ISBN: 978-989-95079-6-8

partial derivatives of $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and the chain rules for the composition of multidimensional functions.

In all the computations we will adopt the EP method, but also if other approaches are adopted, the representation still remains valid.

Let $f : \mathbb{R} \to \mathbb{R}$; its *(EP)*-extension $v = f(u)$ to a fuzzy argument $u = (u^-, u^+)$ has $\alpha - cuts$

$$[v]_\alpha = [\min\{f(x) \mid x \in [u]_\alpha\}, \max\{f(x) \mid x \in [u]_\alpha\}]. \tag{25}$$

If f is monotonic increasing we obtain $[v]_\alpha = [f(u_\alpha^-), f(u_\alpha^+)]$ while, if f is monotonic decreasing, $[v]_\alpha = [f(u_\alpha^+), f(u_\alpha^-)]$; for simplicity, in the monotonic case we assume that the derivative of f is not null over the support of u, but it is possible to design the algorithm also in the case where, for some α, $f'(u_\alpha^+) = 0$ or $f'(u_\alpha^-) = 0$.

The LR representation of $v = (\alpha_i;\ v_{i,L}, \delta v_{i,L}, v_{i,R}, \delta v_{i,R})_{i=0,1,...,N}$ is obtained by the following algorithm:

Algorithm: (1-dim monotonic extension)

Let $u = (\alpha_i; u_{i,L}, \delta u_{i,L}, u_{i,R}, \delta u_{i,R})_{i=0,1,...,N}$ be given and $f : [u_{0,L}, u_{0,R}] \to \mathbb{R}$ be differentiable and monotonic; calculate $v = f(u)$.

for $i = 0, 1, ..., N$
  if (f is increasing) **then**
    $v_{i,L} = f(u_{i,L}),\ \delta v_{i,L} = \frac{\delta u_{n,i}}{f'(u_{n,i})}$
    $v_{i,R} = f(u_{i,R}),\ \delta v_{i,R} = \frac{\delta u_{n,j}}{f'(u_{n,j})}$
  elseif (f is decreasing) **then**
    $v_{i,L} = f(u_{i,R}),\ \delta v_{i,L} = \frac{\delta u_{n,j}}{f'(u_{n,j})}$
    $v_{i,R} = f(u_{i,R}),\ \delta v_{i,R} = \frac{\delta u_{n,j}}{f'(u_{n,j})}$
  endif
end

As an example, the monotonic exponential function $f(x) = \exp(x)$ has LR-fuzzy extension

$$\exp(u) = (\alpha_i; \exp(u_{i,L}), \exp(u_i^-)\,\delta u_{i,L}/\exp(u_{i,L}), \exp(u_{i,R}), \delta u_{i,R}/\exp(u_{i,R}))_{i=0,1,...N}$$

In the non monotonic (differentiable) case, we have to solve the optimization problems in (25) for each $\alpha = \alpha_i$, $i = 0, 1, ..., N$, i.e.

$$(EP_i): \begin{cases} v_{i,L} = \min\{f(x) | x \in [u_{i,L}, u_{i,R}]\} \\ v_{i,R} = \max\{f(x) | x \in [u_{i,L}, u_{i,R}]\}. \end{cases}$$

The $\min$ (or the $\max$) can occur either at a point which is coincident with one of the extremal values of $[u_{i,L}, u_{i,R}]$ or at a point which is internal; in the last case, the derivative of f is null and $\delta v_{i,L} = +\infty$ (or $\delta v_{i,R} = -\infty$).

In the calculations, we "approximate" $\pm\infty$ by $\pm BIG$ where BIG is a big positive number.

Algorithm: (1-dim non monotonic LR extension)

Let $u = (\alpha_i; u_{i,L}, \delta u_{i,L}, u_{i,R}, \delta u_{i,R})_{i=0,1,...,N}$ be given and $f : [u_{0,L}, u_{0,R}] \to \mathbb{R}$ be differentiable and monotonic; calculate $v = f(u)$.

for $i = 0, 1, ..., N$
  solve $\min\{f(x) | x \in [u_{i,L}, u_{i,R}]\}$
  let $\widehat{x}_i = \arg\min\{f(x) | x \in [u_{i,L}, u_{i,R}]\}$
  if $\widehat{x}_i = u_{i,L}$ **then** $v_{i,L} = f(u_{i,L}),\ \delta v_{i,L} = \frac{\delta u_{n,i}}{f'(u_{n,i})}$
  elseif $\widehat{x}_i = u_{i,R}$ **then** $v_{i,L} = f(u_{i,R}), \delta v_{i,L} = \frac{\delta u_{n,j}}{f'(u_{n,j})}$
  else $v_{i,L} = f(\widehat{x}_i),\ \delta v_{i,L} = BIG$
  endif
  solve $\max\{f(x) | x \in [u_{i,L}, u_{i,R}]\}$
  let $\widehat{x}_i = \arg\max\{f(x) | x \in [u_{i,L}, u_{i,R}]\}$
  if $\widehat{x}_i = u_{i,L}$ **then** $v_{i.R} = f(u_{i,L}), \delta v_{i,R} = \frac{\delta u_{n,i}}{f'(u_{n,i})}$
  elseif $\widehat{x}_i = u_{i,R}$ **then** $v_{i,R} = f(u_{i,R}), \delta v_{i,R} = \frac{\delta u_{n,j}}{f'(u_{n,j})}$
  else $v_{i,R} = f(\widehat{x}_i),\ \delta v_{i,R} = -BIG$
  endif
end

4 Conclusion and further work

We suggest a parametrization of LR fuzzy numbers u by the use of parametric monotonic (simple) functions to model the left and the right branches of u. The obtained parametrizations form a subspace of the space of fuzzy numbers and can be possibly be refined to become a dense subspace. Within the parametrizations, we define arithmetic operators for the basic arithmetic and for the fuzzy extension of functions by the application of Zadeh's extension principle.

References

[1] D. Dubois, H. Prade (ed.), *Fundamentals of Fuzzy Sets*, Kluwer, Boston, The Handbooks of Fuzzy Sets Series, 2000.

[2] R. E. Giachetti and R.E. Young: A Parametric Representation of Fuzzy Numbers and their Arithmetic Operators, *Fuzzy Sets and Systems,* 91, (1997) 185-202.

[3] R. E. Giachetti and R.E. Young: Analysis of the Error in the Standard Approximation used for Multiplication of Triangular and Trapezoidal fuzzy Numbers and the Development of a New Approximation, *Fuzzy Sets and Systems,* 91, (1997) 1-13.

[4] S. Heilpern, Representation and application of fuzzy numbers, *Fuzzy Sets and Systems*, 91, 1997, 259-278.

[5] L. Stefanini, L. Sorini, M. L. Guerra, Parametric representation of fuzzy numbers and application to fuzzy calculus, *Fuzzy Sets and Systems*, 157, 2006, 2423 – 2455

[6] L. Stefanini, L. Sorini, M.L. Guerra, Fuzzy Numbers and Fuzzy Arithmetic, in W. Pedrycz et al. (Eds) "Handbook of Granular Computing" , Chapter 12, J. Wiley & Sons, 2008.

[7] L. Stefanini, A Generalization of Hukuhara Difference, in D. Dubois et al. (Eds.), Soft Methods for Handling Variability and Imprecision, ASC 48, Springer Verlag, 2008, 203-210.

ISBN: 978-989-95079-6-8

Type-2 Fuzzy Arithmetic using Alpha-planes

Hussam Hamrawi Simon Coupland

Centre for Computational Intelligence, Faculty of Technology, De Montfort University
Leicester, UK
Email: hussamw@dmu.ac.uk, simonc@dmu.ac.uk

Abstract— This paper examines type-2 fuzzy arithmetic using interval analysis. It relies heavily on alpha-cuts and alpha-planes. Furthermore, we discuss the use of quasi type-2 fuzzy sets proposed by Mendel and Liu and define quasi type-2 fuzzy numbers. Arithmetic operations of such numbers are defined and a worked example is presented.

Keywords— Fuzzy Arithmetic, Type-2 Fuzzy Arithmetic, Type-2 Fuzzy Numbers, Quasi Type-2 Fuzzy Numbers.

1 Introduction

Zadeh introduced the concept of type-2 fuzzy sets in the first of a trio of papers in 1975 [1]. In these papers he also defined linguistic variables, interval-valued fuzzy sets and their representations using the resolution identity (later known as α-cuts). Type-2 fuzzy logic has gained much attention recently due to its ability to handle uncertainty [2], and many advances appeared in both theory and applications [3, 4, 5]. Type-2 fuzzy numbers and the associated arithmetic operations have not received the same attention, only two main contributions appear in the literature that specificly target arithmetic operations using the representation theorem and the extension principle [6, 7]. Recent work on fuzzy systems has proposed an extension to α-cuts for type-2 fuzzy sets through the notion of α-planes [8]. Further work by Mendel and Liu [9] defined a quasi type-2 fuzzy logic system as a restricted special case of a type-2 fuzzy logic system represented by its α-planes.

In this paper we examine type-2 fuzzy arithmetic through α-planes. We use arithmetic operations on type-1 and interval type-2 fuzzy sets through their α-cut representation, moreover, we define quasi type-2 fuzzy numbers and derive their arithmetic operations.

This paper is organised as follows: Section 2 reviews some basic type-2 fuzzy set definitions used in this paper; Section 3 defines type-2 fuzzy numbers and quasi type-2 fuzzy numbers; Section 4 derives the arithmetic operations for these numbers; Section 5 provides a worked example; finally Section 6 provides a conclusion.

2 Type-2 Fuzzy Sets

We present a review of the basic terminology used in this paper. A type-2 fuzzy set (**T2FS**) [2, 10] is defined by equation (1)

$$\widetilde{A} = \int_{\forall x \in X} \int_{\forall u \in J_x \subseteq [0,1]} \mu_{\widetilde{A}}(x,u)/(x,u) \qquad (1)$$

where $\int\int$ denotes the union over all admissible *domain values* x and *secondary domain values* u, and $\mu_{\widetilde{A}}(x,u)$ is a *type-2 membership function*, a T2FS is three dimensional (3D). The *Vertical Slice* (**VS**) is the two dimensional (2D) plane in the u and $\mu_{\widetilde{A}}(x,u)$ axes for a single value of $x = x^{'}$, then the VS is defined by equation (2).

$$VS(x^{'}) = \mu_{\widetilde{A}}(x^{'},u) \equiv \mu_{\widetilde{A}}(x^{'}) = \int_{u \in J_{x^{'}}} f_{x^{'}}(u)/u \qquad (2)$$

where $f_{x^{'}}(u) \in [0,1]$ is called the *secondary grade* and J_x represents the domain of the *secondary membership function* called the *secondary domain*, of course the VS is a *type-1 fuzzy set* (**T1FS**) in $[0,1]$. The *Vertical Slice Representation* (**VSR**) of T2FS is represented by the union of all the vertical slices

$$\widetilde{A} = \{(x, \mu_{\widetilde{A}}(x)) | \forall x \in X\} \qquad (3)$$

The *Footprint Of Uncertainty* (**FOU**) is derived from the union of all primary memberships

$$FOU(\widetilde{A}) = \int_{x \in X} J_x \qquad (4)$$

Another important notion is the *principal membership function* (**PrMF**) defined as the union of all the primary memberships having secondary grades equal to 1

$$Pr(\widetilde{A}) = \int_{x \in X} u/x | f_x(u) = 1 \qquad (5)$$

This notion allows us to view a T1FS as special case of T2FS in which its PrMF is obtained for only one primary membership having a secondary grade at unity [11, 12]. Based on [2] the FOU is described to be bounded by two membership functions a lower $\underline{\mu_{\widetilde{A}}(x)}$ and an upper $\overline{\mu_{\widetilde{A}}(x)}$. The FOU can be described in terms of its upper and lower membership functions (**MF**s)

$$FOU(\widetilde{A}) = \int_{x \in X} \left[\underline{\mu_{\widetilde{A}}(x)}, \overline{\mu_{\widetilde{A}}(x)}\right] \qquad (6)$$

Interval type-2 fuzzy set (**IT2FS**) is defined to be a T2FS where all its secondary grades are of unity $\forall f_x(u) = 1$. A IT2FS can be completely determined using its FOU given in equation (6). Recalling from Zadeh [1] an *α-level set*, A_α, that comprises of elements $x \in X$ of a type-1 fuzzy subset A of universe X

$$A_\alpha = \{x | \mu_A(x) \geq \alpha\} \qquad (7)$$

where $\alpha \in [0,1]$, the fuzzy set A can be represented (decomposed) as

$$A = \bigcup_{\alpha \in [0,1]} \alpha.A_\alpha \qquad (8)$$

ISBN: 978-989-95079-6-8

where $\mu_A(x) = \sup_{x \in A_\alpha} \alpha$ and (sup) denotes the supremum, this decomposition theorem [13] is called the *α-cut Representation*. The same analogy is used by Tahayori *et al.* [14]:

$$\widetilde{A}_{\widetilde{\alpha}} = \{(x,u) | f_x(u) \geq \widetilde{\alpha}\} \tag{9}$$

and the T2FS $\widetilde{A}$ can be represented (decomposed) as

$$\widetilde{A} = \bigcup_{\widetilde{\alpha} \in [0,1]} \widetilde{\alpha}.\widetilde{A}_{\widetilde{\alpha}} \tag{10}$$

where[1] $f_x(u) = \sup_{(x,u) \in \widetilde{A}_{\widetilde{\alpha}}} \widetilde{\alpha}$. Liu [8] calls equation (9) an *α-plane* due to its 2D nature and consequently equation (10) is called the *α-plane Representation*. Liu [8] and Wagner and Hagras [15] noted that when $\widetilde{\alpha} = 0$, $\widetilde{A}_{\widetilde{\alpha}}$ is actually the FOU and Liu and Mendel [9] generalise it to a footprint of uncertainty at plane $\widetilde{\alpha}$ denoted by $FOU(\widetilde{A}_{\widetilde{\alpha}})$. We are interested in the use of this representation as these α-planes can compute some of T2FS operations using IT2FS operations by applying Zadeh's extension principle [1] as follows

$$f(\widetilde{A}) = f(\bigcup_{\widetilde{\alpha} \in [0,1]} \widetilde{\alpha}.\widetilde{A}_{\widetilde{\alpha}}) = \bigcup_{\widetilde{\alpha} \in [0,1]} \widetilde{\alpha}.f(\widetilde{A}_{\widetilde{\alpha}}) \tag{11}$$

As Zadeh himself noted that this principle can be used with functions between sets, relations, etc. and it can extend any function from crisp sets to its fuzzy and obviously type-2 fuzzy counterparts. In [9] based on observations about the shape of the centroid Liu and Mendel defined the notion of *quasi type-2 fuzzy logic system* (QT2FLS). They propose approximating the T2FS using two α-planes ($\widetilde{A}_{\widetilde{\alpha}=0}$ and $\widetilde{A}_{\widetilde{\alpha}=1}$). Great use of these methods is made later in this paper.

3 Type-2 Fuzzy Numbers

A type-1 fuzzy number (T1FN) is defined as a fuzzy set that is both normal and convex [16, 17]. Normality is required in order to capture the concept of a fuzzy number being a set of real numbers close to a specific crisp number [13], in other words when all the uncertainty about a number disappears it reduces to a crisp number. Convexity is required as it allows meaningful arithmetic operations to be performed on fuzzy sets using the well established methods from interval analysis since α-level sets are closed intervals [13]. Coupland *et al.* [6, 7] define a type-2 fuzzy number (T2FN) as a type-2 fuzzy set having a numerical domain. Although no assumption for normality has been defined for these T2FN the examples used assumed normality. From another perspective within the framework of interval-valued fuzzy sets (IVFS) [18], which are equal to IT2FS [5, 19], interval-valued fuzzy numbers (IVFN) are defined to be convex and normal [20, 21]. Here it is desired to define a host of cases that may be conceived to qualify as T2FN. First normality is defined, recalling its definition within the context of T1FN.

Definition 3.1 (Normal T1FS) *A T1FS, A, is said to be normal if its height $h(A)$ is equal to 1 i.e. $\sup \mu_A(x) = 1$.*

Whenever a T1FS is not normal it is called *subnormal*. In IT2FS we differentiate between two cases of normality, one when both the upper and lower MFs are normal, and the other when only the upper T1FS is normal.

[1] Here $\widetilde{\alpha}$ is used instead of α to distinguish that these are α-cuts in the third dimension.

Definition 3.2 (Normal IT2FS) *A IT2FS, $\widetilde{A}$, is said to be normal if its upper MF is normal i.e. $\sup \overline{\mu_A(x)} = 1$.*

Note that there is a point in which a crisp number can be reached which is only depending on the upper MF, this can be seen as a less restrictive condition and is widely used in applications (e.g. Computing with words [22]).

Definition 3.3 (Perfectly Normal IT2FS) *A IT2FS, $\widetilde{A}$, is said to be perfectly[2] normal if both its upper and lower MFs are normal i.e. $\sup \overline{\mu_A(x)} = \sup \underline{\mu_A(x)} = 1$.*

Here the crisp number is reached when uncertainties about both the upper and lower MFs disappear, it is more restrictive, but conceptually more appealing since it generalises the concept of normal T1FS with only a single peak point where it is completely crisp. It can be seen as a special case of a normal IT2FS, this is clear in figure (1).

Definition 3.4 (Partially Normal T2FS) *A T2FS, $\widetilde{A}$, is said to be partially normal if its FOU is **a normal IT2FS.***

This is the weakest case in which a T2FS can qualify to be a number. There is an argument, this can not qualify to be a fuzzy number at all as the secondary membership function is clearly not about 1 at any point (i.e $f_x(u) \neq 1, \forall f_x(u)$).

Definition 3.5 (Normal T2FS) *A T2FS, $\widetilde{A}$, is said to be normal if its FOU is **a normal IT2FS** and it has a PrMF.*

Definition 3.6 (Perfectly Normal T2FS) *A T2FS, $\widetilde{A}$, is said to be perfectly normal if its FOU is a perfectly normal IT2FS and it has a PrMF which is normal (i.e. either a normal T1FS or a normal IT2FS).*

Second, recalling the general definition of a T1FN using piecewise-defined functions.

Definition 3.7 (T1FN[13]) *Let A be a fuzzy subset on real numbers. Then, A is a fuzzy number if and only if there exists a closed interval $[m_1, m_2] \neq \phi$ such that*

$$\mu_A(x) = \begin{cases} 1 & x \in [m_1, m_2] \\ l(x) & x \in [s, m_1) \\ r(x) & x \in (m_2, e] \\ 0 & x \in (-\infty, s); x \in (e, \infty) \end{cases} \tag{12}$$

where $l(x) \in [0,1]$ is monotonically increasing and continuous from the right; $r(x) \in [0,1]$ is monotonically decreasing and continuous from the left; $\langle s, m_1, m_2, e \rangle$ are the parameters that define the T1FN.

Observe that if A is subnormal it is not considered a number, but it is useful to define a *type-1 fuzzy sub-number* (T1FsN) which satisfies all the properties of a T1FN except it is subnormal, this can be defined for a fuzzy set A with height $h(A) = \sup_{\forall x} \mu_A(x) = h_A$ as follows

$$\mu_A(x) = \begin{cases} h_A & x \in [m_1, m_2] \\ l(x) & x \in [s, m_1) \\ r(x) & x \in (m_2, e] \\ 0 & x \in (-\infty, s); x \in (e, \infty) \end{cases} \tag{13}$$

[2] this term has been used by Kaufmann and Gupta [17] describing a perfectly triangular T2FN.

ISBN: 978-989-95079-6-8

Using equations (7) and (8) a T1FN can be represented as intervals using their α-cuts $A_\alpha = [a_1^\alpha, a_2^\alpha]$ as follows

$$A = \bigcup_{\forall\alpha} \alpha.[a_1^\alpha, a_2^\alpha] = \bigcup_{\forall\alpha} \alpha.[l^{-1}(\alpha), r^{-1}(\alpha)] \quad (14)$$

IT2FN can be represented by its lower and upper membership functions which themselves are T1FS. Here we use the terms above to define IT2FN. Let $FOU(\widetilde{A}) = [\underline{FOU(\widetilde{A})}, \overline{FOU(\widetilde{A})}]$ [3] to represent a perfectly normal IT2FS defined by its lower and upper membership functions, respectively. Then, using equation (14) individually for each of the memberships, it follows that

$$\underline{FOU(\widetilde{A})} = \bigcup_{\forall\alpha} \alpha.[\underline{a_1^\alpha}, \underline{a_2^\alpha}] = \bigcup_{\forall\alpha} \alpha.[\underline{l^{-1}}(\alpha), \underline{r^{-1}}(\alpha)]$$

$$\overline{FOU(\widetilde{A})} = \bigcup_{\forall\alpha} \alpha.[\overline{a_1^\alpha}, \overline{a_2^\alpha}] = \bigcup_{\forall\alpha} \alpha.[\overline{l^{-1}}(\alpha), \overline{r^{-1}}(\alpha)]$$

then

$$FOU(\widetilde{A}) = \bigcup_{\forall\alpha} \alpha.\left[[\overline{a_1^\alpha}, \underline{a_1^\alpha}], [\underline{a_2^\alpha}, \overline{a_2^\alpha}]\right] \quad (15)$$

for a normal IT2FS with lower membership function height, $h(FOU(\widetilde{A})) = h_l$, a IT2FN can be represented as

$$FOU(\widetilde{A}) = \begin{cases} \bigcup_{\forall\alpha} \alpha.\left[[\overline{a_1^\alpha}, \underline{a_1^\alpha}], [\underline{a_2^\alpha}, \overline{a_2^\alpha}]\right] & \alpha \le h_l \\ \bigcup_{\forall\alpha} \alpha.[\overline{a_1^\alpha}, \overline{a_2^\alpha}] & \alpha > h_l \end{cases} \quad (16)$$

Then a IT2FN can be defined by definition (3.8)

Definition 3.8 (IT2FN)
Let $FOU(\widetilde{A}) = [\underline{FOU(\widetilde{A})}, \overline{FOU(\widetilde{A})}]$ be an interval type-2 fuzzy subset on real numbers. Then, $FOU(\widetilde{A})$ is an interval type-2 fuzzy number if:

- *$\underline{FOU(\widetilde{A})}$ and $\overline{FOU(\widetilde{A})}$ are T1FNs, in this case it is called a perfectly normal IT2FN.*
- *$\underline{FOU(\widetilde{A})}$ is a T1FsN [4] and $\overline{FOU(\widetilde{A})}$ is a T1FN, in this case it is called a normal IT2FN.*

This definition can be seen in figure (1), observe that both IT2FSs are IT2FNs. We can also consider an interval type-2 fuzzy sub-number (IT2FsN) when bothe the upper and lower MFs are T1FsNs. Now, it is desired to define T2FN using the same terminology. From equation (10) a T2FS

$$\widetilde{A} = \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{A}_{\widetilde{\alpha}})$$

and using equation (15) a perfectly normal T2FS

$$\widetilde{A} = \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.(\bigcup_{\forall\alpha} \alpha.\left[[\overline{a_1^\alpha}, \underline{a_1^\alpha}], [\underline{a_2^\alpha}, \overline{a_2^\alpha}]\right]) \quad (17)$$

and using equation (16) a normal T2FS

$$\widetilde{A} = \begin{cases} \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.(\bigcup_{\forall\alpha} \alpha.\left[[\overline{a_1^\alpha}, \underline{a_1^\alpha}], [\underline{a_2^\alpha}, \overline{a_2^\alpha}]\right]) & \alpha \le h_l \\ \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.(\bigcup_{\forall\alpha} \alpha.[\overline{a_1^\alpha}, \overline{a_2^\alpha}]) & \alpha > h_l \end{cases} \quad (18)$$

A partially normal T2FS will only require $\forall\widetilde{\alpha}; \widetilde{\alpha} \in [0, \sup_{\forall(x,u)} f_x(u)]$.

[3] FOU is used here since the FOU fully describes the IT2FS.
[4] see equation (13).

Definition 3.9 (T2FN) *Let $\widetilde{A} = \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{A}_{\widetilde{\alpha}})$ be a type-2 fuzzy subset on real numbers. Then, $\widetilde{A}$ is a type-2 fuzzy number if:*

- *$FOU(\widetilde{A}_{\widetilde{\alpha}=0})$ is a perfectly normal IT2FN, and $FOU(\widetilde{A}_{\widetilde{\alpha}=1}) = Pr(\widetilde{A})$ is normal (i.e. either a T1FN or a normal IT2FN), in this case it is called a perfectly normal T2FN.*
- *$FOU(\widetilde{A}_{\widetilde{\alpha}=0})$ is a normal IT2FN, and $FOU(\widetilde{A}_{\widetilde{\alpha}=1})$ ($Pr(\widetilde{A})$) exist, in this case it is called a normal T2FN.*
- *$FOU(\widetilde{A}_{\widetilde{\alpha}=0})$ is a normal IT2FN, in this case it is called a partially normal T2FN.*

Note that a special case of $FOU(\widetilde{A}_{\widetilde{\alpha}=1}) = Pr(\widetilde{A})$ is when all the vertical slices constructing $Pr(\widetilde{A})$ are triangular T1FN then $Pr(\widetilde{A})$ is T1FS, this is depicted in figures (2) and (3). We can also define a type-2 fuzzy sub-number (T2FsN) when its FOU is an IT2FsN. In some applications one may need to restrict a fuzzy set to a specific form, e.g. in computing with words Klir *et al.* [23, 24] defined a procedure to convert any given convex fuzzy set to a fuzzy interval that is expressed in some standard form using some specific criteria. We now examine a special form of T2FN based on some ideas from [9]. Some observations about the shape of a centroid led to the proposition of a quasi type-2 fuzzy logic system, similirly, a definition of a quasi T2FN (QT2FN) can be proposed.

Assumption 3.10 (QT2FS) *Consider the following propositions about a T2FN*

A1 The T2FN is Normal (i.e FOU is normal IT2FN and PrMF is normal).

A2 The FOU upper and lower membership functions, the PrMF, and the vertical slices are characterised by piecewise functions.

A3 All the vertical slices that construct PrMF are T1FN.

A4 As for the rest of the vertical slices, they are T1FsN in which their supremum lay on piece-wise functions between the sides of the upper membership function of the FOU and the PrMF denoted $\widetilde{l}(x)$ on the left side, and $\widetilde{r}(x)$ on the right side.

A5 All the mid points of the FOU upper and lower membership functions, and the PrMF are at the same domain value.

these assumptions allow a T2FN to be completely determined by its FOU and PrMF, just like a T1FS based on certain assumptions can be completely determined by its core and support.

Definition 3.11 (QT2FN) *Let $\widetilde{A} = \bigcup_{\forall\widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{A}_{\widetilde{\alpha}})$ be a type-2 fuzzy subset on real numbers. Then, $\widetilde{A}$ is a quasi type-2 fuzzy number if it is completely determined by its FOU and PrMF.*

Figures (3) and (4) are both QT2FNs.

ISBN: 978-989-95079-6-8

4 Arithmetic using Alpha-planes

In Kaufmann and Gupta [17] a comprehensive discussion on fuzzy numbers and arithmetic operations which is formulated and mostly dependent upon the interval of confidence (α-level sets). Recalling operations on intervals, let $[a_1, a_2]$ and $[b_1, b_2]$ be two interval numbers, then

$$[a_1, a_2] \oplus [b_1, b_2] = [a_1 \oplus b_1, a_2 \oplus b_2]$$

$$[a_1, a_2] \otimes [b_1, b_2] = [\min(a_1 \otimes b_1, a_1 \otimes b_2, a_2 \otimes b_1, a_2 \otimes b_2) , \max(a_1 \otimes b_1, a_1 \otimes b_2, a_2 \otimes b_1, a_2 \otimes b_2)] \quad (19)$$

where $\oplus \in \{+, -\}$, $\otimes \in \{\times, \div\}$ and $0 \notin B$ if $\otimes = \div$. Interval operations are extended to T1FN. Let $A = \bigcup_{\forall \alpha} \alpha.[a_1^\alpha, a_2^\alpha]$ and $B = \bigcup_{\forall \alpha} \alpha.[b_1^\alpha, b_2^\alpha]$ be two T1FN, then

$$A \circ B = \bigcup_{\forall \alpha} \alpha.([a_1^\alpha, a_2^\alpha] \circ [b_1^\alpha, b_2^\alpha]) \quad (20)$$

where $\circ = \{+, -, \times, \div\}$. Also these operations are extended to IT2FN. Let $FOU(\widetilde{A}) = \bigcup_{\forall \alpha} \alpha. \left[\overline{[a_1^\alpha, a_1^\alpha]}, \underline{[a_2^\alpha, a_2^\alpha]}\right]$ and $FOU(\widetilde{B}) = \bigcup_{\forall \alpha} \alpha. \left[\overline{[b_1^\alpha, b_1^\alpha]}, \underline{[b_2^\alpha, b_2^\alpha]}\right]$ be two perfectly normal IT2FN, then [17]

$$FOU(\widetilde{A}) \circ FOU(\widetilde{B}) = \bigcup_{\forall \alpha} \alpha.(\left[\overline{[a_1^\alpha, a_1^\alpha]} \circ \overline{[b_1^\alpha, b_1^\alpha]}, \underline{[a_2^\alpha, a_2^\alpha]} \circ \underline{[b_2^\alpha, b_2^\alpha]}\right]) \quad (21)$$

If $FOU(\widetilde{A})$ and $FOU(\widetilde{B})$ are normal IT2FN with lower membership function heights $h_{\widetilde{A}}$ and $h_{\widetilde{B}}$ respectively, then the following changes are made to equation (21)

- if $0 \leq \alpha \leq \min(h_{\widetilde{A}}, h_{\widetilde{B}})$ no changes are made.
- if $h_{\widetilde{A}} < \alpha \leq h_{\widetilde{B}}$ then $\underline{a_1^\alpha} = \underline{a_1^{h_{\widetilde{A}}}}$; $\underline{a_2^\alpha} = \underline{a_2^{h_{\widetilde{A}}}}$, and if $h_{\widetilde{B}} < \alpha \leq h_{\widetilde{A}}$ then $\underline{b_1^\alpha} = \underline{b_1^{h_{\widetilde{B}}}}$; $\underline{b_2^\alpha} = \underline{b_2^{h_{\widetilde{B}}}}$.
- if $\max(h_{\widetilde{A}}, h_{\widetilde{B}}) < \alpha \leq 1$ then equation (21) becomes $FOU(\widetilde{A}) \circ FOU(\widetilde{B}) = \bigcup_{\forall \alpha} \alpha.([\overline{a_1^\alpha, a_2^\alpha}] \circ [\overline{b_1^\alpha, b_2^\alpha}])$.

Wu and Mendel [25] noticed that such methods that result in discontinuous or nonconvex sets are neither desirable nor technically correct. Similarly such results appear for normal IT2FN arithmetic operations. According to Wu and Mendel [25] the following changes are made to equation (21)

$$FOU(\widetilde{A}) \circ FOU(\widetilde{B}) = \begin{cases} \bigcup_{\forall \alpha} \alpha.(\left[\overline{[a_1^\alpha, a_1^\alpha]} \circ \overline{[b_1^\alpha, b_1^\alpha]}, \underline{[a_2^\alpha, a_2^\alpha]} \circ \underline{[b_2^\alpha, b_2^\alpha]}\right]) \\ \quad , \text{if } 0 \leq \alpha \leq \min(h_{\widetilde{A}}, h_{\widetilde{B}}) \\ \bigcup_{\forall \alpha} \alpha.([\overline{a_1^\alpha, a_2^\alpha}] \circ [\overline{b_1^\alpha, b_2^\alpha}]) \\ \quad , \text{if } \min(h_{\widetilde{A}}, h_{\widetilde{B}}) < \alpha \leq 1 \end{cases} \quad (22)$$

Clearly T2FN arithmetic can be extended, let $\widetilde{A} = \bigcup_{\forall \widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{A}_{\widetilde{\alpha}})$ and $\widetilde{B} = \bigcup_{\forall \widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{B}_{\widetilde{\alpha}})$ be two T2FNs, then

$$\widetilde{A} \circ \widetilde{B} = \bigcup_{\forall \widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{A}_{\widetilde{\alpha}}) \circ \bigcup_{\forall \widetilde{\alpha}} \widetilde{\alpha}.FOU(\widetilde{B}_{\widetilde{\alpha}})$$

using the extension principle in equation (11) it follows that

$$\widetilde{A} \circ \widetilde{B} = \bigcup_{\forall \widetilde{\alpha}} \widetilde{\alpha}.(FOU(\widetilde{A}_{\widetilde{\alpha}}) \circ FOU(\widetilde{B}_{\widetilde{\alpha}})) \quad (23)$$

then we can use equation (21) for perfectly normal T2FNs or equation (22) for normal T2FNs. Next we examine QT2FN, let $\widetilde{A}^Q$ and $\widetilde{B}^Q$ be QT2FN completely determined by the FOU and PrMF, $\langle FOU(\widetilde{A}^Q_{\widetilde{\alpha}=0}), Pr(\widetilde{A}^Q)\rangle$ and $\langle FOU(\widetilde{B}^Q_{\widetilde{\alpha}=0}), Pr(\widetilde{B}^Q)\rangle$, respectively. Then the result of basic arithmetic operations between them is a QT2FN

$$\widetilde{C}^Q = \widetilde{A}^Q \circ \widetilde{B}^Q \quad (24)$$

completely determined by $\langle FOU(\widetilde{C}^Q_{\widetilde{\alpha}=0}), Pr(\widetilde{C}^Q)\rangle$ where $FOU(\widetilde{C}^Q_{\widetilde{\alpha}=0}) = FOU(\widetilde{A}^Q_{\widetilde{\alpha}=0}) \circ FOU(\widetilde{B}^Q_{\widetilde{\alpha}=0})$ and $Pr(\widetilde{C}^Q) = Pr(\widetilde{A}^Q) \circ Pr(\widetilde{B}^Q)$. This operation can only be performed for addition and subtraction as the functions $\widetilde{l}(x)$ and $\widetilde{r}(x)$ are preserved. In the case of multiplication and devision an approximation for these functions should be used. It has to be mentioned that methods to approximate a T1FN to some standard form is used in the literature (see Grzegorzewski [26]). Providing approximation to QT2FN is still an open question.

5 Worked Example

In this example we only consider QT2FN as it gives sufficient insight on T2FN. Let us consider the following triangular QT2FN $\widetilde{3}^Q = \langle FOU(\widetilde{3}^Q_{\widetilde{\alpha}=0}), Pr(\widetilde{3}^Q)\rangle$ depicted in figure (5) with parameters[5] $FOU(\widetilde{3}^Q_{\widetilde{\alpha}=0}) = \langle 1.5, 2.25, 3, 3.45, 4.75\rangle$, $h_{\widetilde{3}} = 0.6$, and $Pr(\widetilde{3}^Q) = \langle 1.75, 3, 4.25\rangle$. And the QT2FN $\widetilde{12}^Q = \langle FOU(\widetilde{12}^Q_{\widetilde{\alpha}=0}), Pr(\widetilde{12}^Q)\rangle$ depicted in figure (6) with parameters $FOU(\widetilde{12}^Q_{\widetilde{\alpha}=0}) = \langle 10.25, 11.5, 12, 12.5, 14\rangle$, $h_{\widetilde{12}} = 0.7$, and $Pr(\widetilde{12}^Q) = \langle 10.75, 12, 13.5\rangle$. When computing the addition $\widetilde{3}^Q + \widetilde{12}^Q$, first, we determine a suitable number of α-cuts along u for both FOU and PrMF[6]. In our case and for the sake of clarity we discretised u into 25 α-cuts. Then, we apply equation (24) to the decomposed IT2FS. This gives the result depicted in figure (7) with parameters $FOU(\widetilde{15}^Q_{\widetilde{\alpha}=0}) = \langle 11.75, 13.75, 15, 15.95, 18.75\rangle$, $h_{\widetilde{15}} = 0.6$, and $Pr(\widetilde{15}^Q) = \langle 12.5, 15, 17.75\rangle$. This result we would expect, $\widetilde{3}^Q + \widetilde{12}^Q = \widetilde{15}^Q$.

6 Conclusions

In this paper we presented methods to perform type-2 fuzzy arithmetic operations using well known arithmetic operations on intervals. In order to better analyse our methods we defined a set of terms that describe different normality formations of IT2FS and T2FS, some of which, have already been used in the literature without a clear definition. Our methods relied heavily on interval analysis through α-cuts and α-planes that allow T2FSs be represented as a collection of intervals. Furthermore, we examined the use of quasi type-2 fuzzy sets proposed by Mendel and Liu and defined quasi type-2 fuzzy numbers as a special case of type-2 fuzzy numbers. Finally, we derived arithmetic operations for quasi type-2 fuzzy numbers and provided an illustration by a worked example.

[5] The parameters used here are $FOU(\widetilde{A}^Q_{\widetilde{\alpha}=0}) = \langle s_1, s_2, m_1, e_2, e_1\rangle$ derived from equation (12) where $\overline{FOU(\widetilde{A}^Q_{\widetilde{\alpha}=0})} = \langle s_1, m_1, e_1\rangle$ and $\underline{FOU(\widetilde{A}^Q_{\widetilde{\alpha}=0})} = \langle s_2, m_1, e_2\rangle$.

[6] In the case of T2FN we first discretise along $f_x(u)$ in order to determine a suitable number of α-planes, then, we discretise along u for each of the α-planes.

ISBN: 978-989-95079-6-8

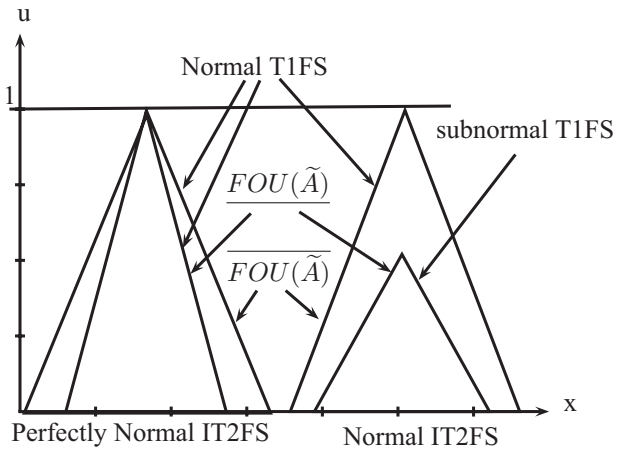


Figure 1: Different types of normal fuzzy sets

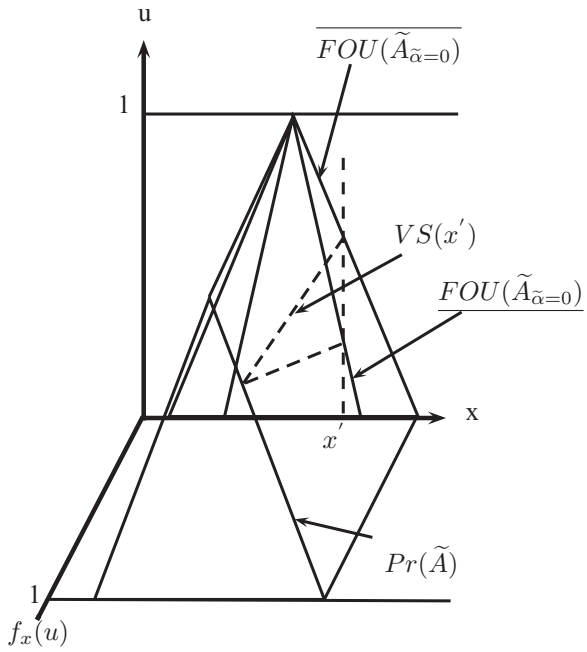


Figure 2: 3D representation of a perfectly normal T2FN, it is also a QT2FN.

We believe that this paper will form the basis for further work on type-2 fuzzy numbers and thier applications. Quasi type-2 fuzzy numbers may be viewed as a next step in the progress between IT2FN and T2FN. Arithmetic operations are already used in aggregation and averaging operations, measures that apply certain mathematical functions, and fusion functions between one system and another.

Further work with regards to the operations defined here will include approximation of standard forms of Quasi type-2 fuzzy numbers, and a thorough comparison between these methods and previous methods in terms of computational complexity, although, it is almost trivial that our methods are computationally sound as they are based on interval analysis.

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive comments which we believe have helped to improve this paper, and also would like to thank Mr. Simon Miller for his helpful discussions.

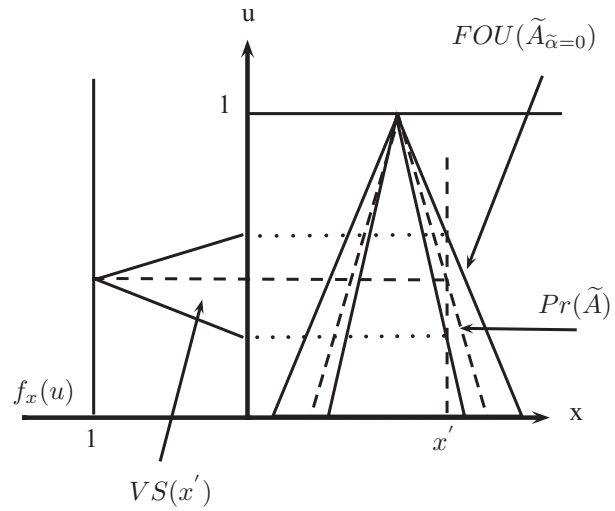


Figure 3: 2D representation of a perfectly normal T2FN of figure (2) with triangular vertical slices.

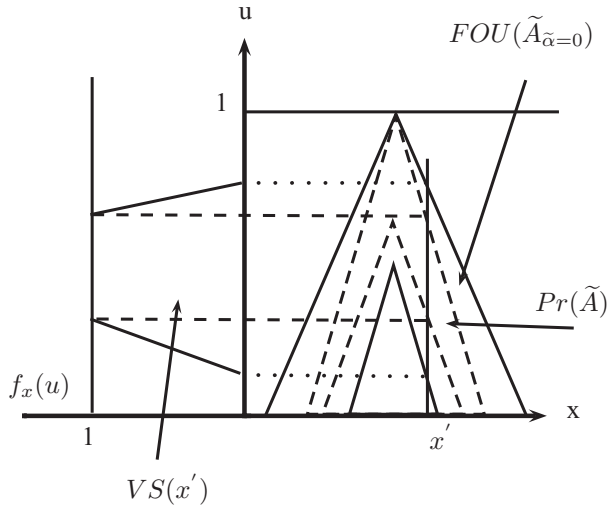


Figure 4: 2D representation of a normal T2FN with trapizoidal vertical slices, note that $Pr(\widetilde{A})$ is normal IT2FN.

References

[1] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-1. *Information Sciences*, 8:199 – 249, 1975.

[2] J.M. Mendel. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall, Upper Saddle River, NJ, 2001.

[3] R.I. John. Type-2 fuzzy sets:an appraisal of theory and applications. *Uncertainty, Fuzziness and Knowledge-Based Systems*, 6:563–576, 1998.

[4] R.I. John and S. Coupland. Type-2 fuzzy logic a historical view. *IEEE Computational Intelligence Magazine*, 2(1):57–62, February 2007.

[5] J. M. Mendel. Advances in type-2 fuzzy sets and systems. *Information Sciences*, 177(1):84–110, January 2007.

[6] S. Coupland and R.I. John. An approach to type-2 fuzzy arithmetic. In *Proc. UK Workshop on Computational Intelligence*, pages 107 – 114, 2003.

ISBN: 978-989-95079-6-8

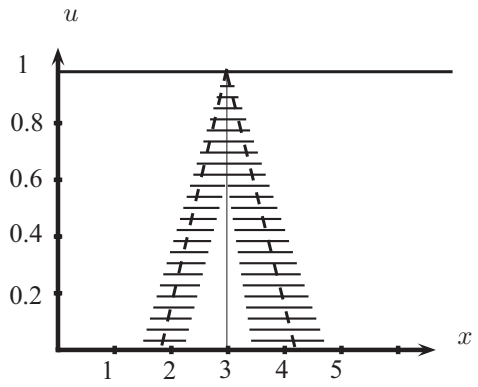


Figure 5: $\widetilde{3}^Q$ with discretised α-cuts, and the dashed line is $Pr(\widetilde{3}^Q)$

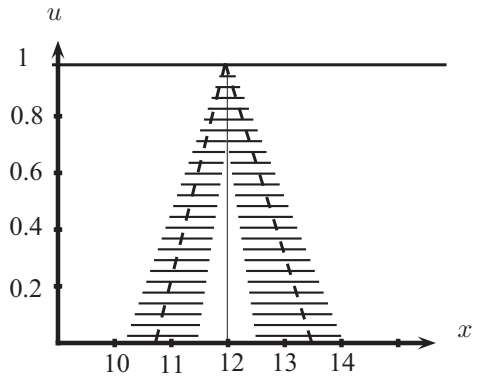


Figure 6: $\widetilde{12}^Q$ with discretised α-cuts, and the dashed line is $Pr(\widetilde{12}^Q)$

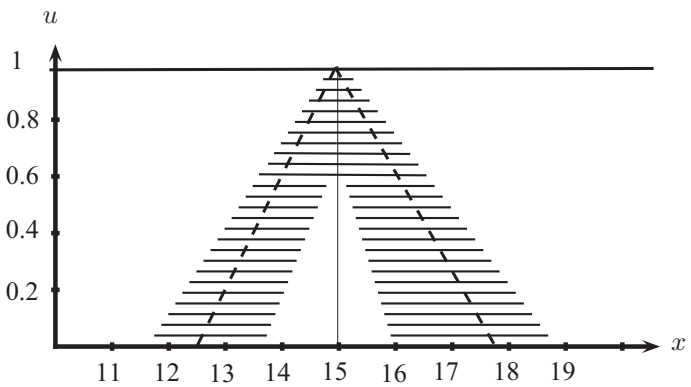


Figure 7: $\widetilde{15}^Q$ with discretised α-cuts, and the dashed line is $Pr(\widetilde{15}^Q)$

[7] W. Blewitt, S. Coupland, and S. M. Zhou. A novel approach to type-2 fuzzy addition. In *Proc. FUZZ-IEEE 2007*, pages 1456 – 1461, London, 2007.

[8] F. Liu. An efficient centroid type reduction strategy for general type-2 fuzzy logic systems. Walter J. Karplus Summer Research Grant Report, IEEE Computational Intelligence Society, 2006.

[9] J.M. Mendel and F. Liu. On new quasi-type-2 fuzzy logic systems. In *FUZZ-IEEE 2008*, pages 354–360, Hong Kong, 2008.

[10] J. M. Mendel and R.I. John. Type-2 fuzzy sets made simple. *IEEE Transaction on Fuzzy Systems*, 10(2):117–127, 2002.

[11] N.N. Karnik and J.M. Mendel. Operations on type-2 fuzzy set. *Fuzzy Sets and Systems*, 122(2):327–348, 2001.

[12] L.A. Lucas, T.M. Centeno, and M.R. Delgado. General type-2 fuzzy classifiers to land cover classification. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1743–1747. ACM New York, NY, USA, 2008.

[13] G.J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, Upper Saddle River, NJ, 1995.

[14] H. Tahayori, A. Tettamanzi, and G.D. Antoni. Approximated type-2 fuzzy set operations. In *Proc. FUZZ-IEEE 2006*, pages 9042 – 9049, Vancouver, Canada, July 2006.

[15] C. Wagner and H. Hagras. zSlices–Towards bridging the gap between Interval and General Type-2 Fuzzy Logic. In *FUZZ-IEEE 2008.*, pages 489–497, Hong Kong, 2008.

[16] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*, volume 144. Academic Press INC, NY, USA, 1980.

[17] A. Kaufmann and M. Gupta. *Introduction to Fuzzy Arithmetic Theory and Applications*. Van Nostran Reinhold Co. Inc., 1985.

[18] M.B. Gorzalczany. A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, 21:1–17, January 1987.

[19] J.M. Mendel, R.I. John, and F. Liu. Interval type-2 fuzzy logic systems made simple. *IEEE Transaction on Fuzzy Systems*, 14(6):808–821, December 2006.

[20] D. H. Hong and S. Lee. Some algebraic properties and a distance measure for interval-valued fuzzy numbers. *Information Sciences*, 148(1-4):1–10, December 2002.

[21] G. Wang and X. Li. The applications of interval-valued fuzzy numbers and interval-distribution numbers. *Fuzzy Sets and Systems*, 98(3):331–335, 1998.

[22] J.M. Mendel and D. Wu. Perceptual Reasoning: A New Computing with Words Engine. In *Granular Computing, 2007. GRC 2007. IEEE International Conference on*, pages 446–446, 2007.

[23] Olga Martin and George J. Klir. On the problem of retranslation in computing with preceptions. *International Journal of General Systems*, 35(6):655–674, 2006.

[24] G.J. Klir and K. Sentz. *Aspects of Automatic Text Analysis*, chapter On the issue of linguistic approximation. Springer, 2006.

[25] D. Wu and J. M. Mendel. Corrections to "Aggregation Using the Linguistic Weighted Average and Interval Type-2 Fuzzy Sets". *IEEE Transactions on Fuzzy Systems*, Accepted for future publication.

[26] P. Grzegorzewski. Trapezoidal approximations of fuzzy numbers preserving the expected interval–Algorithms and properties. *Fuzzy Sets and Systems*, 159:1354–1364, 2008.

ISBN: 978-989-95079-6-8

Operations for Real-Valued Bags and Bag Relations

Sadaaki Miyamoto
Department of Risk Engineering, University of Tsukuba
Email: miyamoto@risk.tuskuba.ac.jp

Abstract— *A generalization of bags is proposed and fundamental operations are described. That is, a real-valued bag is proposed which is a straightforward generalization of the traditional bags, but is necessary for complementation, and s-norm and t-norm operations. We mainly describe how complementation and s-norms are related. As a result a duality theorem between s-norm and t-norm is obtained. R-bag relations with max-s norm composition and max-t norm composition are also proposed. Further generalization that includes fuzzy bags and real-valued bags having membership of a region in a plane is moreover discussed.*

Keywords— bag relation, max-s composition, real-valued bag, set-valued bag, t-norm and conorm .

1 Introduction

Bags which are also called *multisets* have long been studied by computer scientists [5, 6]. In soft computing, Yager [19] have proposed fuzzy bags, and its theory and applications have been studied by several researchers [3, 4, 15, 16, 17, 18, 20]. In particular, the author [9, 10, 12] has redefined basic operations of the union and intersection for fuzzy bags. The definition anew by the authors is necessary in order to keep consistency between operations for ordinary fuzzy sets and those for fuzzy bags [12]. However, there are many problems that are left and should be studied.

One class of such problems is fundamental operations of bags including complementation and bag s-norms (and s-norms). Another class is bag relations that should have composition operations. In this paper we study these two problem classes. While studying these problems, we should generalize the traditional concept of bags, as described in this paper. We therefore introduce a generalization called R-bag which appears straightforward, but this generalization is necessary to consider complementation and bag t-norms. As a result, a duality theorem is derived using the complementation. Moreover a Minkwoski type bag s-norms are derived using a generating function.

Bag relations and their compositions for R-bags are then studied by introducing a max-s algebra and a max-t algebra. Here s and t stands for s-norm and t-norm.

Another generalization that includes R-bag and fuzzy bag is moreover mentioned and how the above consideration is extended to this generalized bag is suggested.

There are a number of propositions herein, but to prove propositions given in this paper is not difficult. Hence short notes on proofs are given in the appendix.

Although the purpose of this paper is to show theoretical properties of generalized bags, bags have many application possibilities, which are briefly mentioned in the conclusion.

2 Bags and R-Bags

The two terms of bags and multisets are used interchangeably, and we use *bags* throughout this paper.

2.1 Overview of crisp bags

Assume that the universal set $X = \{x_1, \ldots, x_n\}$ is finite for simplicity. A crisp bag M of $X = \{x_1, \ldots, x_n\}$ is characterized by a function $C_M(\cdot)$ (called count of M) whereby a natural number including zero $\boldsymbol{N} = \{0, 1, 2, \ldots\}$ corresponds to each $x \in X$ ($C_M : X \to \boldsymbol{N}$).

For a crisp bag, different expressions such as

$$M = \{k_1/x_1, \ldots, k_n/x_n\}$$

and

$$M = \{\overbrace{x_1, \ldots, x_1}^{k_1}, \ldots, \overbrace{x_n, \ldots, x_n}^{k_n}\}$$

are used. In such a way, an element of X may appear more than once in a bag. In the above example x_i appears k_i times in M ($i = 1, \ldots, n$).

Consider an example where $X = \{a, b, c, d\}$ and $C_M(a) = 2$, $C_M(b) = 1$, $C_M(c) = 3$, $C_M(d) = 0$. In other words, $M = \{a, a, b, c, c, c\}$. This means that a, b, c, and d are included 2, 1, 3, and 0 times, respectively, in M. We can write $M = \{2/a, 1/b, 3/c\}$ by ignoring an element of zero occurrence.

Basic relations and operations for crisp bags are as follows.

1. (inclusion):
   $M \subseteq N \Leftrightarrow C_M(x) \leq C_N(x), \quad \forall x \in X.$
2. (equality):
   $M = N \Leftrightarrow C_M(x) = C_N(x), \quad \forall x \in X.$
3. (union):
   $C_{M \cup N}(x) = \max\{C_M(x), C_N(x)\}.$
4. (intersection):
   $C_{M \cap N}(x) = \min\{C_M(x), C_N(x)\}.$
5. (addition or sum):
   $C_{M \oplus N}(x) = C_M(x) + C_N(x).$
6. (subtraction):
   $C_{M \ominus N}(x) = \max\{0, C_M(x) - C_N(x)\}.$
7. (scalar multiplication):
   $C_{\alpha M} = \alpha C_M(x)$, where α is a positive integer.
8. (Cartesian product):
   Let P is a bag of Y.
   $C_{M \times P}(x, y) = C_M(x) C_P(y).$

ISBN: 978-989-95079-6-8

We sometimes use $\vee$ and $\wedge$ for $\max$ and $\min$, respectively. We also note that the relations and operations are similar to those for fuzzy sets. However, bags have the addition which fuzzy sets do not have, and the Cartesian product for bags is different from that for fuzzy sets. Moreover, a complementation is not defined in general.

2.2 R-bags

An immediate generalization of crisp bag is to assume that $C_M(\cdot)$ has a nonnegative real-value instead of a natural number. For example, we admit $C_M(a) = 3.14$ or $C_N(a) = \pi$. Moreover we admit the value of infinity $+\infty$ for a technical reason. We thus assume $C_M : X \to [0, +\infty]$.

The above relations and operations 1–8 are used without any change except that α in the scalar multiplication can now be a positive real constant. For simplicity, a real-valued bag is called an R-bag here and the collection of all R-bags of X is denoted by $\mathcal{RB}(X)$.

Complementation of R-bags

A function $\mathcal{N} : [0, +\infty] \to [0, +\infty]$ with the next properties is used to define a complementation operation:

(i) $\mathcal{N}(0) = +\infty, \quad \mathcal{N}(+\infty) = 0.$

(ii) $\mathcal{N}(x)$ is strictly monotonically decreasing on $(0, +\infty)$.

A typical example is $\mathcal{N}(x) = const/x$ with $const > 0$.

Using this function, a natural operation for the complement is:

9.(complement):
$C_{\bar{M}}(x) = \mathcal{N}(C_M(x))$.

This operation justifies the generalization into R-bags. That is, even when we start crisp bags, the result of complementation is generally real-valued.

We immediately have the next two propositions of which the proof is easy and omitted.

Proposition 1 *For arbitrary $M, N \in \mathcal{RB}(X)$, the following properties hold:*

$$\overline{(\overline{M})} = M \tag{1}$$

$$\overline{M \cup N} = \bar{M} \cap \bar{N}, \qquad \overline{M \cap N} = \bar{M} \cup \bar{N}. \tag{2}$$

Proposition 2 *If we introduce an empty bag $\emptyset$ and the maximum bag* **Infinity** *in $\mathcal{RB}$ by*

$$C_{\emptyset}(x) = 0, \quad \forall x \in X, \tag{3}$$

$$C_{\mathbf{Infinity}}(x) = +\infty, \quad \forall x \in X. \tag{4}$$

Then we have

$$\bar{\emptyset} = \mathbf{Infinity}, \quad \overline{\mathbf{Infinity}} = \emptyset. \tag{5}$$

s-norms and t-norms of R-bags

There have been studies on t-norms for crisp bags [7, 1], but generalization into R-bags admits a broader class of s-norms and t-norms. For this purpose we introduce two functions $t(a, b)$ and $s(a, b)$ as those in fuzzy sets, but the definitions are different.

Definition 1 *Two functions $t : [0, +\infty] \times [0, +\infty] \to [0, +\infty]$ and $s : [0, +\infty] \times [0, +\infty] \to [0, +\infty]$ having the following properties (I)–(IV) are called a t-norm and an s-norm for bags, respectively. An s-norm is also called a t-conorm for bags.*

(I)[monotonicity] For $a \le c$, $b \le d$,

$$t(a, b) \le t(c, d),$$
$$s(a, b) \le s(c, d).$$

(II)[symmetry]

$$t(a, b) = t(c, d), \quad s(a, b) = s(b, a).$$

(III)[associativity]

$$t(t(a, b), c) = t(a, t(b, c)),$$
$$s(s(a, b), c) = s(a, s(b, c)).$$

(IV)[boundary condition]

$$t(0, 0) = 0, \quad t(a, +\infty) = t(+\infty, a) = a,$$
$$s(+\infty, +\infty) = +\infty, \quad s(a, 0) = s(0, a) = a.$$

A purpose to introduce such norms for bags is to generalize the intersection and union operations. First we note that $s(a, b) = a+b$ and $s(a, b) = \max\{a, b\}$ satisfy the above conditions. The norm $t(a, b) = \min\{a, b\}$ also satisfies (I)–(IV). Thus, s-norms and t-norms can represent the three operations of addition, union, and intersection.

We moreover introduce a generating function $g(x)$ for s-norm.

Definition 2 *A function $g : [0, +\infty] \to [0, +\infty]$ is called a generating function for s-norm if it satisfies the next (i)–(iii):*
(i) it is strictly monotonically increasing,
(ii) $g(0) = 0, \quad g(+\infty) = +\infty$,
(iii) $g(x + y) \ge g(x) + g(y), \quad \forall x, y \in [0, +\infty]$.
We have

Proposition 3 *Let*

$$s(a, b) = g^{-1}(g(a) + g(b)). \tag{6}$$

Then $s(a, b)$ is an s-norm.

An example of the generation function is

$$g(x) = x^p \quad (p \ge 1). \tag{7}$$

Moreover, note the following.

Proposition 4 *Let $s(a, b)$ is an s-norm and $\mathcal{N}$ is the complementation operator. Then*

$$t(a, b) = \mathcal{N}(s(\mathcal{N}(a), \mathcal{N}(b))) \tag{8}$$

is a t-norm. Suppose $t(a, b)$ is a t-norm, then

$$s(a, b) = \mathcal{N}(t(\mathcal{N}(a), \mathcal{N}(b))) \tag{9}$$

is an s-norm.

If a pair of t-norm and s-norm has the above property, we say (s, t) has the *duality* of norm and conorm. The duality has the next property.

ISBN: 978-989-95079-6-8

Proposition 5 *Suppose $s_0(a,b)$ is an s-norm and $t_0(a,b)$ is derived from $s_0(a,b)$ by the operation (8). Let*

$$s(a,b) = \mathcal{N}(t_0(\mathcal{N}(a),\mathcal{N}(b)))$$

Then $s(a,b) = s_0(a,b)$.

Suppose also that $t_0(a,b)$ is a t-norm and $s_0(a,b)$ is derived from $t_0(a,b)$ by the operation (8). Let

$$t(a,b) = \mathcal{N}(s_0(\mathcal{N}(a),\mathcal{N}(b)))$$

Then $t(a,b) = t_0(a,b)$. In summary, applying (8) and (9) (resp. (9) and (8)) successively, we have the original norm.

Note that s-norm and t-norm are applied to define bag operations $M\mathcal{S}N$ and $M\mathcal{T}N$, respectively, for $M, N \in \mathcal{RB}(X)$.

$$C_{M\mathcal{S}N}(x) = s(C_M(x), C_N(x)). \tag{10}$$
$$C_{M\mathcal{T}N}(x) = t(C_M(x), C_N(x)). \tag{11}$$

Let us consider examples of s-norms and t-norms.

Example 1 The standard operators

$$s(a,b) = \max\{a,b\} \tag{12}$$
$$t(a,b) = \min\{a,b\} \tag{13}$$

are an s-norm and a t-norm, respectively. This pair has the duality stated in Proposition 5. Note, however, that s-norm (12) does not have a generating function that satisfies (6).

Second example uses the generating function.

Example 2 Let $g(x)$ be given by (7). Then we have

$$s(a,b) = (a^p + b^p)^{\frac{1}{p}}, \tag{14}$$
$$t(a,b) = (a^{-p} + b^{-p})^{-\frac{1}{p}}. \tag{15}$$

are an s-norm and a t-norm, respectively. This pair has the duality stated in Proposition 5 when $\mathcal{N} = const/x$ is used.

The second example has interesting properties. First, $s(a,b) = a+b$ is a particular case of (14) for $p = 1$. Moreover $s(a,b) = \max\{a,b\}$ and $t(a,b) = \min\{a,b\}$ are obtained from (14) and (15) when $p \to +\infty$.

New findings are included in the above example. First, the dual t-norm derived from $s(a,b) = a + b$ is the harmonic mean:

$$t(a,b) = (a^{-1} + b^{-1})^{-1} = \frac{ab}{a+b}.$$

(Note that the arithmetic mean $\frac{a+b}{2}$ and the geometric mean $\sqrt{ab}$ are not t-norms). Second, $s(a,b) = (a^2 + b^2)^{\frac{1}{2}}$ with $p = 2$ which is similar to the Euclidean norm is an s-norm, and the dual t-norm is

$$t(a,b) = \frac{1}{\sqrt{\frac{1}{a^2} + \frac{1}{b^2}}}.$$

2.3 R-bags and fuzzy sets

Let us study what relations exist between fuzzy sets and R-bags. For this purpose we introduce another notation: the collection of all fuzzy sets of X is denoted by $\mathcal{FS}(X)$.

We introduce another function $h\colon [0,+\infty] \to [0,1]$ that is strictly monotonically increasing in $[0,+\infty)$ and

$$h(0) = 0, \quad h(+\infty) = 1.$$

An example of such a function is

$$h_0(x) = 1 - \exp(-K_0 x), \quad x \in [0,+\infty)$$

with constant $K_0 > 0$ and with the additional definition $h_0(+\infty) = 1$. Another example is

$$h_1(x) = 1 - \frac{K_1}{K_1 + K_2 x}, \quad x \in [0,+\infty)$$

with positive constants K_1, K_2 and with $h_1(+\infty) = 1$.

Since such a function has an inverse function $h^{-1}(y)$, we can define one to one correspondence between a fuzzy set and an R-bag:

$$\mu_{h(M)}(x) = h(C_M(x)). \tag{16}$$

For $M \in \mathcal{RB}(X)$, $h(M) \in \mathcal{FS}(X)$, and $h^{-1}(h(M)) = M$. Note that $h^{-1}(y)$ is also a strictly monotonically increasing.

We also have

$$h(\emptyset) = \emptyset, \quad h(\mathbf{Infinity}) = X;$$
$$h^{-1}(\emptyset) = \emptyset, \quad h^{-1}(X) = \mathbf{Infinity}.$$

Using these functions, we can generate operations for R-bags from those for fuzzy sets.

Proposition 6 *Let*

$$\mathcal{N}(x) = h^{-1}(1 - h(x)). \tag{17}$$

Then $\mathcal{N}(x)$ satisfies the properties of complementation operation. Hence for arbitrary $M \in \mathcal{RB}(X)$, $\bar{M}$ derived by

$$C_{\bar{M}}(x) = \mathcal{N}(C_M(x))$$

using (17) is a complement of M.

Note that this operation is derived from the complementation of fuzzy sets.

For the s-norm and t-norm, we have similar results. Let $s_F(a,b)$ and $t_F(a,b)$ are respectively an s-norm and a t-norm for *fuzzy sets*. We have the following.

Proposition 7 *For $a, b \in [0,+\infty]$, let*

$$s(a,b) = h^{-1}(s_F(h(a), h(b))) \tag{18}$$
$$t(a,b) = h^{-1}(t_F(h(a), h(b))) \tag{19}$$

Then $s(a,b)$ and $t(a,b)$ are an s-norm and a t-norm for R-bags.

we also have

Proposition 8 *Suppose $s(a,b)$ and $t(a,b)$ are an s-norm and a t-norm for R-bags. For $a, b \in [0,1]$, let*

$$s_F(a,b) = h(s(h^{-1}(a), h^{-1}(b))) \tag{20}$$
$$t_F(a,b) = h(t(h^{-1}(a), h^{-1}(b))) \tag{21}$$

Then $s_F(a,b)$ and $t_F(a,b)$ are an s-norm and a t-norm for fuzzy sets.

3 Bag Relations

What we consider in this section is a bag-relations and their composition. As a result we have max-s and max-t algebras for R-bags. The discussion in this section generalizes the max-plus algebra [2].

ISBN: 978-989-95079-6-8

3.1 max-s and max-t algebra

Let us use a particular notation of $\boxplus$ and $\boxdot$ for

$$a \boxplus b = \max\{a, b\}, \quad a \boxdot b = s(a, b) \tag{22}$$

where $s(a, b)$ is an s-norm for R-bags. We call this algebra as *max-s algebra*.

It is easy to see that the following properties hold.

$$a \boxplus b = b \boxplus a, \tag{23}$$
$$a \boxplus (b \boxplus c) = (a \boxplus b) \boxplus c, \tag{24}$$
$$a \boxplus 0 = a, \tag{25}$$
$$a \boxdot b = b \boxdot a, \tag{26}$$
$$a \boxdot (b \boxdot c) = (a \boxdot b) \boxdot c, \tag{27}$$
$$a \boxdot 0 = a. \tag{28}$$

Alternatively, we can define $\boxplus$ and $\boxdot$ for

$$a \boxplus b = \max\{a, b\}, \quad a \boxdot b = t(a, b) \tag{29}$$

where $t(a, b)$ is a t-norm for R-bags. We call this algebra as *max-t algebra*. We see that (23)–(27) hold, while (28) should be replaced by

$$a \boxdot +\infty = a. \tag{30}$$

We moreover note the next lemma.

Lemma 1 *Let a, b, c be real numbers. Then*

$$a \boxdot (b \boxplus c) = (a \boxdot b) \boxplus (a \boxdot c). \tag{31}$$

Thus the commutative, associative, and distributive properties hold for the max-s algebra. Noting Example 2, we observe that the max-t algebra includes the max-min algebra and the max-s algebra includes the max-plus algebra as particular cases.

3.2 Bag relations

We now define R-bag relations. The definition is simple. An R-bag relation R on $X \times Y$ is an R-bag R of $X \times Y$. The count function is denoted by $R(x, y)$ instead of $C_R(x, y)$ for simplicity. Sometimes an R-bag relation is also called bag relation for simplicity.

The reason why we call such a bag $R(x, y)$ a bag relation is that we can define composition operation.

Let X, Y, Z be three universes. Assume R is a bag relation of $X \times Y$ and S is a bag relation of $Y \times Z$. Then the max-s composition $R \circ S$ is defined as follows.

$$(R \circ S)(x, z) = \boxplus_{y \in Y}\{R(x, y) \boxdot S(y, z)\} \tag{32}$$

Note that

$$\boxplus_{y \in \{a_1, \ldots, a_L\}} = a_1 \boxplus a_2 \boxplus \cdots \boxplus a_L.$$

The max-t composition is defined by the same equation (32) except that $\boxdot$ uses a t-norm.

Note also that the addition is straightforward

$$(R_1 \boxplus R_2)(x, y) = R_1(x, y) \boxplus R_2(x, y), \tag{33}$$

for bag relations on $X \times Y$.

We have the following proposition.

Proposition 9 *The composition satisfies the associative property*

$$(R \circ S) \circ T = R \circ (S \circ T). \tag{34}$$

and the distributive property

$$(R_1 \boxplus R_2) \circ S = (R_1 \circ S) \boxplus (R_2 \circ S), \tag{35}$$
$$R \circ (S_1 \boxplus S_2) = (R \circ S_1) \boxplus (R \circ S_2). \tag{36}$$

We introduce the unit relations for the max-s and max-t compositions. Define O_X and Ω_X on $X \times X$:

$$O_X(x, y) = 0, \quad \forall x, y \in X, \tag{37}$$
$$\Omega_X(x, y) = +\infty, \quad \forall x, y \in X. \tag{38}$$

We define O_{XY} and Ω_{XY} on $X \times Y$ in the same way:

$$O_{XY}(x, y) = 0, \quad \forall (x, y) \in X \times Y, \tag{39}$$
$$\Omega_{XY}(x, y) = +\infty, \quad \forall (x, y) \in X \times Y. \tag{40}$$

Frequently we omit the subscripts like O and Ω when we have no ambiguity.

We then have

Proposition 10 *Assume that the max-s algebra is used. For arbitrary bag relation R on $X \times Y$,*

$$R \boxplus O = O \boxplus R = R, \tag{41}$$
$$R \circ O = O \circ R = R. \tag{42}$$

Proposition 11 *Assume that the max-t algebra is used. For arbitrary bag relation R on $X \times Y$,*

$$R \boxplus O = O \boxplus R = R, \tag{43}$$
$$R \circ \Omega = \Omega \circ R = R. \tag{44}$$

4 A Note on Further Generalization

The above generalization to R-bags does not include fuzzy bags [19, 10]. It appears that R-bags and fuzzy bags are inconsistent. However, there is further generalization discussed in [12] which uses a closed region in $\boldsymbol{R}^2$ as the count function for generalized bags. In this section we briefly note how the above discussion is extended into generalized bags which encompasses R-bags and fuzzy bags.

A generalized bag A has a count function as a region

$$\mathbf{N}(x, A) = C_A(x), \quad \forall x \in X \tag{45}$$

where $\mathbf{N}(x, A)$ is a closed set of $[0, +\infty)^2$. Let $\nu(y, z; x, A)$ is the characteristic function of the region $\mathbf{N}(x, A)$:

$$\nu(y, z; x, A) = \begin{cases} 1, & (y, z) \in \mathbf{N}(x, A), \\ 0, & (y, z) \notin \mathbf{N}(x, A). \end{cases}$$

This region should have the following properties:

(N1) For any $y \in [0, +\infty)$,

$$\eta(y) = \{z \in [0, +\infty) : \nu(y, z; x, A) = 1\}$$

is either an empty set or a closed interval $[0, \mu(y)]$ where $\mu(y) = \max\{y' : y' \in \eta(y)\}$.

ISBN: 978-989-95079-6-8

(N2) For all $y \in [0, +\infty)$, $\mu(y)$ is a monotonically decreasing function and $0 \leq \mu(y) \leq 1$.

(N3) For any $z \in [0, +\infty)$,

$$\theta(z) = \{y \in [0, +\infty) : \nu(y, z; x, A) = 1\}$$

is either an empty set or a closed interval $[0, \zeta(z)]$ where $\zeta(z) = \max\{z' : z' \in \theta(z)\}$.

(N4) For all $z \in [0, +\infty)$, $\zeta(z)$ is a monotonically decreasing function and $0 \leq \zeta(z) < +\infty$.

Roughly, a generalized bag is characterized by the pair $(\mu(y), \zeta(y))$ for each $x \in X$.

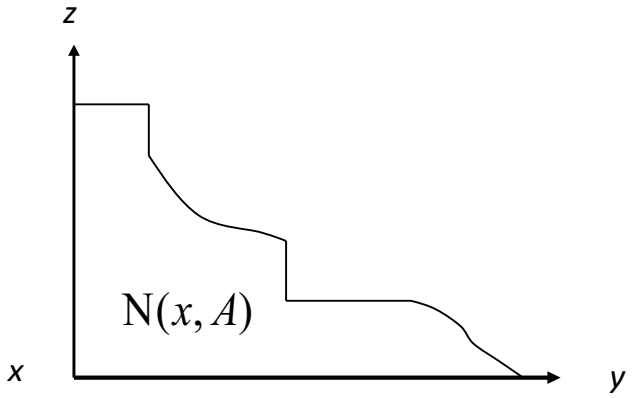


Figure 1: An illustration of $\mathbf{N}(x, A)$ as the count function for a generalized bag.

Two cuts which relate the generalized bag to R-bags and fuzzy sets are defined:

α-cut: Let $\alpha \in (0, 1]$. For the generalized bag A, $[A]_\alpha$ is an R-bag given by

$$C_{[A]_\alpha}(x) = \max\{z : \nu(\alpha, z; x, A) = 1\}.$$

ℓ-cut: Let $\ell \in (0, +\infty)$. For the generalized bag A, $\langle A \rangle_\ell$ is a fuzzy set given by

$$\mu_{\langle A \rangle_\ell}(x) = \max\{y : \nu(y, \ell; x, A) = 1\}.$$

It can be proved that the generalized bags are smallest generalization of fuzzy bags and R-bags [12]. Moreover, s-norms and t-norms can be used for the generalized bags using the α-cuts, while the complementation operation cannot be used for the generalized bags. Bag relations and their compositions can also be used for the generalized bags. The detailed consideration is too lengthy to give here and will be discussed elsewhere.

5 Conclusion

We have shown the operations of complementation and s-norm and t-norm for real-valued bags. Moreover R-bag relations and the max-s (or max-t) compositions are studied. Further generalization has moreover been proposed of which the details will be discussed in near future.

As seen herein, the theory of bags should be studied more deeply, since it has more ample mathematical structure than expected until now. More theoretical studies are thus needed. Relations between rough sets [14] and bags have been discussed [12], but there is much room for further studies. Application of bags such as those in information retrieval [8] and information mining [11, 13] will also be important.

Acknowledgment

This study has partly been supported by the Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science, No.19650052.

Appendix: Notes on Proofs

Since detailed proofs are too lengthy, we give short notes on how to prove the propositions given here. The proofs are mostly straightforward.

Notes on proofs of Propositions 1 and 2

These propositions are immediately proved by straightforward calculation using $\mathcal{N}$. The details are hence omitted.

Notes on proofs of Propositions 3 and 4

It is enough to check the conditions (I)–(IV) are satisfied.

Notes on proof of Proposition 5

$$\begin{aligned} s(a, b) &= \mathcal{N}(t_0(\mathcal{N}(a), \mathcal{N}(b))) \\ &= \mathcal{N}(\mathcal{N}(s_0(\mathcal{N}(\mathcal{N}(a)), \mathcal{N}(\mathcal{N}(b))))) \\ &= s_0(a, b). \end{aligned}$$

Notes on proof of Proposition 6

By a simple calculation, it is seen that $\mathcal{N}$ by (17) satisfies the conditions for the complementation.

Notes on proofs of Propositions 7 and 8

The conditions for s-norms and t-norms for fuzzy sets should be noted. Mostly it is easy to check that the conditions are satisfied, since the functions h and h^{-1} are monotone. The boundary conditions are satisfied from $h(0) = 0$ and $h(+\infty) = 1$.

Notes on proof of Lemma 1

Let $b \geq c$, then the left hand side is $s(a, b)$. The right hand side is

$$\max\{s(a, b), s(a, c)\} = s(a, b).$$

The case of $b < c$ is calculated likewise and we have the desired equality.

When we use a max-t algebra, we can prove the lemma in a similar way, since the monotone property is common to s-norm and t-norm. We omit the detail.

Notes on proof of Proposition 9

The detailed proof is lengthy, but the way to prove this proposition is just the same as the proof of the matrix properties for ordinary algebra, since those properties for ordinary algebra uses the commutative, associative, and distributive properties which are valid for the present algebra, since (23)–(28) (or (30)) and (31) are valid.

Notes on proofs of Propositions 10 and 11

Omitted as they are straightforward.

ISBN: 978-989-95079-6-8

References

[1] J. Casasnovas, G. Mayor, Discrete t-norms and operations on extended multisets, *Fuzzy Ses and Systems*, Vol.159, pp. 1165–1177, 2008.

[2] B. Heidergott, G.J. Olsder, J. van der Woude, *Max Plus at Work*, Princeton University Press, Princeton, NJ, 2006.

[3] B. Li, W. Peizhang, and L. Xihui, Fuzzy bags with set-valued statistics, *Comput. Math. Applic.*, Vol. 15, pp. 811–818, 1988.

[4] B. Li, Fuzzy bags and applications, *Fuzzy Sets and Systems*, Vol. 34, pp. 61–71, 1990.

[5] D. E. Knuth, *The Art of Computer Programming, Vol. 2 / Seminumerical Algorithms*, Addison-Wesley, Reading, Massachusetts, 1969.

[6] Z. Manna and R. Waldinger, *The Logical Basis for Computer Programming, Vol. 1: Deductive Reasoning*, Addison-Wesley, Reading, Massachusetts, 1985.

[7] G. Mayor, J. Torrens, Triangular norms on discrete settings, in: E.P. Klement, R. Mesiar, eds., *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*, Elsevier, Amsterdam, 2005, pp. 189–230.

[8] S. Miyamoto, *Fuzzy Sets in Information Retrieval and Cluster Analysis*, Kluwer, Dordrecht, 1990.

[9] S. Miyamoto, Basic operations of fuzzy multisets, *J. of Japan Society for Fuzzy Theory and Systems*, Vol. 8, No. 4, pp. 639–645, 1996 (in Japanese).

[10] S. Miyamoto, Fuzzy multisets with infinite collections of memberships, *Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97)*, June 25-30, 1997, Prague, Czech, Vol.1, pp.61-66, 1997.

[11] S. Miyamoto, Information clustering based on fuzzy multisets, *Information Processing and Management*, Vol.39, No.2, pp. 195–213, 2003.

[12] S. Miyamoto, Generalizations of multisets and rough approximations, *International Journal of Intelligent Systems*, Vol.19, No.7, pp. 639–652, 2004.

[13] S. Miyamoto, Y. Kawasaki, Text Analysis Based on Bag Neighborhood Model and Application to Incident Reports, Da Ruan et al., eds., Computational Intelligence in Decision and Control, Proc. of the 8th Intern. FLINS Conf., Madrid, Spain, 21-24 Sept. 2008, pp.1015-1020.

[14] Z. Pawlak, *Rough Sets*, Kluwer, Dordrecht, 1991.

[15] A. Ramer and C.-C. Wang, Fuzzy multisets, *Proc. of 1996 Asian Fuzzy Systems Symposium*, Dec. 11-14, 1996, Kenting, Taiwan, pp. 429–434.

[16] A. Rebai, Canonical fuzzy bags and bag fuzzy measures as a basis for MADM with mixed non cardinal data, *European J. of Operational Res.*, Vol. 78, pp. 34–48, 1994.

[17] A. Rebai, J.-M. Martel, A fuzzy bag approach to choosing the "best" multiattributed potential actions in a multiple judgment and non cardinal data context, *Fuzzy Sets and Systems*, Vol. 87, pp. 159–166, 1997.

[18] D. Rocacher, On fuzzy bags and their application to flexible querying, *Fuzzy Sets and Systems*, Vol. 140(1), pp.93-110, 2003.

[19] R. R. Yager, On the theory of bags, *Int. J. General Systems*, Vol. 13, pp. 23–37, 1986.

[20] R. R. Yager, Cardinality of fuzzy sets via bags, *Mathl. Modelling*, Vol. 9, No. 6, pp. 441–446, 1987.

ISBN: 978-989-95079-6-8

An Explicit Mapping for Kernel Data Analysis and Application to Text Analysis

Sadaaki Miyamoto[1] Yuichi Kawasaki[2] Keisuke Sawazaki[2]

1.Department of Risk Engineering, University of Tsukuba
Tsukuba, Japan
2.Graduate School of Systems and Information Engineering, University of Tsukuba
Tsukuba, Japan
Email: miyamoto@risk.tuskuba.ac.jp, {kawasaki,sawazaki}@soft.risk.tsukuba.ac.jp

Abstract— *Kernel data analysis is now becoming standard in every application of data analysis and mining. Kernels are used to represent a mapping into a high-dimensional feature space, where an explicit form of the mapping is unknown. Contrary to this common understanding, we introduce an explicit mapping which we consider standard. The reason why we use this mapping is as follows. (1) the use of this mapping does not lose any fundamental information in kernel data analysis and we have the same formulas in every kernel methods. (2) Usually the derivation becomes simpler by using this mapping. (3) New applications of the kernel methods become possible using this mapping. As an application we consider an example of text mining where we use fuzzy c-means clustering and cluster centers in the high-dimensional space and visualize the centers using kernel principal component analysis.*

Keywords— Kernel data analysis, fuzzy clustering, explicit mapping, text mining

1 Introduction

Kernel functions [9] which are not very new but was noted in support vector machines[10] and now is a well-known technique and becoming standard in many applications of data analysis and data mining. An important point in the use of kernel functions is that although we consider a mapping from a data space into a high-dimensional feature space, we need not to know its explicit form but we should know the inner product of the feature space. Generally, the feature space is not uniquely determined. Accordingly every formula in data analysis using kernel functions should be described in terms of the inner product.

Although kernel functions are really useful, but the derivation is sometimes complicated when original formulas should be rewritten by the inner product forms. A typical example is kernel fuzzy c-means clustering [3] and kernel SOM [4] in which the cluster centers in a feature space should be eliminated using inner product forms.

Here is a question: can we have a useful and explicit mapping and explicit representation of a high-dimensional feature space? The answer is YES, and we will show the mapping that is simple enough and useful in the sense that it leads to the same formulas when transformed into the inner product forms. To summarize, this explicit mapping and associated feature space have all information that is used in kernel functions for data analysis.

Another question arises: what is useful in this explicit mapping? We will show a real application of an example of text analysis. A result of kernel fuzzy c-means clustering and cluster centers will be shown using kernel principal component analysis.

Proofs of propositions are mostly omitted to save the space but brief notes are given as appendix.

2 Kernel Functions and an Explicit Mapping

2.1 Preliminary consideration

Although we consider kernel fuzzy c-means (FCM) clustering [3] and kernel principal component analysis (KPCA)[9] later, we describe FCM and omit KPCA for simplicity.

Assume that a set of data $X = \{x_1, \ldots, x_n\} \subset \boldsymbol{R}^p$ is given. Each data unit is also called an object or an individual and it is a point in the p-dimensional real space $x_k = (x_k^1, \ldots, x_k^p)^T \in \boldsymbol{R}^p$. We consider a mapping into a high-dimensional feature space $\Phi\colon \boldsymbol{R}^p \to H$ and associated kernel function

$$\boldsymbol{K}(x, y) = \langle \Phi(x), \Phi(y) \rangle \tag{1}$$

where $\langle \cdot, \cdot \rangle$ is an inner product of H. We also assume $\| \cdot \|_H$ is a norm of H. Thus H is an inner product space. In this section suppose we do not know function $\Phi(\cdot)$ explicitly but we know $\boldsymbol{K}(x, y)$. Specifically, the Gaussian kernel is used frequently:

$$\boldsymbol{K}(x, y) = \exp(-\lambda \|x - y\|^2) \tag{2}$$

where $\lambda > 0$ and $\|x\|$ is the norm of $\boldsymbol{R}^p$.

An objective function of fuzzy c-means using the feature space H is the following.

$$J_H(U, W) = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ki})^m \|\Phi(x_k) - W_i\|_H^2, \quad (m > 1) \tag{3}$$

where $U = (u_{ki})$ is $n \times c$ matrix representing membership of x_k to cluster i. U has the next constraint when it is optimized.

$$M = \{U = (u_{ki}) : \sum_{i=1}^{c} u_{ki} = 1,\ u_{ki} \geq 0, \forall k, i\}. \tag{4}$$

Moreover $W = (W_1, \ldots, W_c)$ shows cluster centers.

The iterative algorithm of fuzzy c-means clustering is basically an alternative minimization of $J_H(U, W)$ with respect to

ISBN: 978-989-95079-6-8

U and W until convergence. We have the next solutions.

$$u_{ki} = \left[\sum_{j=1}^{c} \left(\frac{D(x_k, W_i)}{D(x_k, W_j)} \right)^{\frac{1}{m-1}} \right]^{-1}, \tag{5}$$

$$W_i = \frac{\sum_{k=1}^{n} (u_{ki})^m \Phi(x_k)}{\sum_{k=1}^{n} (u_{ki})^m} \tag{6}$$

where we put

$$D(x_k, W_i) = \|\Phi(x_k) - W_i\|_H^2. \tag{7}$$

Equations (5) and (6) are repeated until convergence, but since $\Phi(x_k)$ is unknown, we should use another formula for kernel fuzzy c-means clustering.

The formula is derived by eliminating (6) from iteration, i.e., we substitute (6) into (7) to have an updating formula for $D(x_k, W_i)$[3]:

$$D(x_k, W_i) = \boldsymbol{K}(x_k, x_k) - \frac{2}{\sum_{k=1}^{n} (u_{ki})^m} \sum_{j=1}^{n} (u_{ji})^m \boldsymbol{K}(x_j, x_k) + \frac{1}{\{\sum_{k=1}^{n} (u_{ki})^m\}^2} \sum_{j=1}^{n} \sum_{\ell=1}^{n} (u_{ji} u_{\ell i})^m \boldsymbol{K}(x_j, x_\ell). \tag{8}$$

We hence repeat (5) and (8) until convergence when kernel fuzzy c-means clustering should be used.

2.2 An explicit mapping and its properties

We sometimes have applications in which we wish to have explicit cluster centers. Even if we cannot show them in a low-dimensional data space, there are ways to have approximate representations or visualizations, one of which is to show kernel principal components and another is to use SOM mapping.

To this end we use the following explicit mapping that is simple enough but seems unnoticed.

$$\Phi(x_k) = e_k \quad (k = 1, 2, \ldots, n) \tag{9}$$

where $H = \boldsymbol{R}^n$ and e_k is the k-th unit vector that has unity as ith component and all other components are zero. Note that $\Phi\colon X \to \boldsymbol{R}^n$, i.e., $\Phi(\cdot)$ is not defined on $\boldsymbol{R}^p$ but is limited to the finite set X. Moreover we assume that the inner product of $\boldsymbol{R}^n$ is

$$\langle e_k, e_\ell \rangle = \boldsymbol{K}(x_k, x_\ell) \tag{10}$$

instead of the usual $\langle e_k, e_\ell \rangle = \delta_{k\ell}$.

We have the next proposition.

Proposition 1. *If kernel $\boldsymbol{K}(x, y)$ is positive definite, $\langle e_k, e_\ell \rangle$ defined by (10) satisfy the axioms of the inner product of $\boldsymbol{R}^n$, that is, $\boldsymbol{R}^n$ with (10) is an inner product space.*

We now apply this mapping to fuzzy c-means clustering. It is sufficient to show the optimal solution of W_i. We then have the next proposition of which the proof is easy and omitted.

Proposition 2. *For all positive definite kernel $\boldsymbol{K}(x, y)$ and mapping Φ by (9), the cluster centers are the same and are given by*

$$W_i = \left(\frac{(u_{1i})^m}{\sum_{k=1}^{n} (u_{ki})^m}, \ldots, \frac{(u_{ni})^m}{\sum_{k=1}^{n} (u_{ki})^m} \right)^T, \quad i = 1, \ldots, c \tag{11}$$

The next proposition of which the proof is almost trivial shows an important result of the equivalence between the usual technique and the explicit mapping method.

Proposition 3. *Using the explicit mapping (11) to (7), we have (8).*

That is, (11) derived from the single mapping (9) has all necessary and sufficient information for kernel fuzzy c-means clustering.

As noted above, formulas in kernel principal component analysis are derived likewise. Since the derivation repeats what is described in textbooks [9], we omit the details. The mapping (9) is moreover useful for application to LVQ and SOM [2], where vectors for updating quantization vectors should be based on learning. The explicit mapping enables vector representations in $\boldsymbol{R}^n$, we can use every formulas in LVQ and SOM, while usual kernel methods should eliminate quantization vectors [6].

3 Analysis of Terms from a Set of Texts

We hereafter show an application in which mapping (9) has a natural interpretation. A typical example is analysis of terms from s set of texts. A well-known model for term analysis uses a vector-space model [8] which is also called a bag model. Mizutani and Miyamoto[6] generalize the model into fuzzy multiset model and use kernel based learning and clustering. Note that a term may occur many times in a text which we call here term occurrences, or simply occurrences. Term occurrences are handled as term frequencies or count in bags in the vector space model, or it is treated as fuzzy multisets in [6] when memberships are attached to each term occurrence. When c-means clustering is applied, cluster centers have implications as they are representatives of clusters.

In these models, however, a structure inside a text such as distances or topology between occurrences is not considered. Miyamoto and Kawasaki[5] propose a kernel-based model that handles fuzzy neighborhood. As a result, kernel-based data analysis can be applied but a drawback is that we cannot have cluster centers. We consider this model of fuzzy neighborhood here and uses the above explicit mapping and overcome the last drawback.

3.1 Fuzzy neighborhood model

Let us consider a model that consists of a quintuple:

$$< T, O, d, R, N >$$

where each element is as follows.

1. $T = \{t_1, \ldots, t_n\}$: a finite set of *terms* or *keywords*;
2. O and d: a metric space; O is a finite set of *term occurrences* where a metric d is defined;
3. R : a fuzzy relation of $T \times O$;
4. N : a fuzzy relation of $O \times O$ called *fuzzy neighborhood*.

ISBN: 978-989-95079-6-8

The relation $R(t,o)$ shows correspondence between a term and its occurrence. For simplicity we assume, for every $t \in T$, $R(t,\cdot) = \{o \in O : R(t,o) > 0\}$ is not empty. The most simple relation is

$$R(t,o) = \begin{cases} 1 & (o \text{ is an occurrence of } t), \\ 0 & (\text{otherwise}), \end{cases}$$

and this relation is assumed hereafter in this paper.

Moreover we impose the following conditions to N:

(i) $N(o,o) = 1$ for all $o \in O$;

(ii) $d(o,o') > d(o,o'') \Rightarrow N(o,o') \leq N(o,o'')$;

(iii) $N(o,o') \to 0$ as $d(o,o') \to \infty$;

(iv) $N(o,o') = N(o',o)$.

Our purpose is to define a natural inner product space on T using the structure of occurrence space. We define a proximity relation $p(t,t')$, $t,t' \in T$:

$$p(t,t') = \sum_{o\in O}\sum_{o'\in O} R(t,o)N(o,o')R(t',o'). \tag{12}$$

If the matrix $P = (p(t,t'))$ is positive-definite, we can define an inner product space. If P is positive definite, we can use a normalized relation

$$s(t,t') = \frac{p(t,t')}{\sqrt{p(t,t)p(t',t')}}. \tag{13}$$

which is also positive-definite.

Generally, the above equation (12) does not guarantee the positive definiteness. We hence consider conditions for the positive-definiteness. To obtain a general condition is not practically useful. We therefore describe a specific, but broad enough, application example.

3.2 A model of a text set

A set of texts where each term may occur many times is represented as a sequence. For example, let a term set is $T = \{a,b,c\}$. Then an example of O can be represented by

$$O = aabcbbcccabbccabab$$

or more precisely, we put suffixes:

$$O = a_1\, a_2\, b_1\, c_1\, b_2\, b_3\, c_2\, c_3\, c_4\, a_3\, b_4\, b_5\, c_5\, c_6\, a_4\, b_6\, a_5\, b_7 \tag{14}$$

in order to distinguish occurrences. Obviously, $R(a,\cdot) = \{a_1,\ldots,a_5\}$. What if two or more texts should be handled? We can connect those texts into a sequence using dummy terms. Hence a sequence is sufficient to represent a set of texts. A natural distance d on O is given by

$$d(o,o') = \{\text{the number of occurrences between } o \text{ and } o'\}+1 \tag{15}$$

Thus, $d(a_1,c_1) = 3$, $d(c_6,c_2) = 7$, *etc* in the above example.

Next, two specific examples of fuzzy neighborhood are considered.

Fuzzy neighborhood using a monotone function

Let us show and example of a fuzzy neighborhood. For this purpose we introduce a monotone-decreasing function $f\colon \boldsymbol{R} \to [0,1]$ such that

(I) $f(0) = 1$,

(II) $f(-x) = f(x)$,

(III) $\lim_{x\to\infty} f(x) = 0$.

We define

$$N(o,o') = f(d(o,o')). \tag{16}$$

We have

Proposition 4. *Fuzzy relation $N(o,o')$ defined by (16) satisfies the conditions (i)–(iv) for a fuzzy neighborhood.*

The next proposition is crucial to this model.

Proposition 5. *If function $f(x)$ is convex on $[0,\infty)$, then $P = (p(t,t'))$ is positive-definite. Namely, we can define an inner product space using P.*

Let us consider a simple example where a convex function

$$f(x) = \begin{cases} 1 - \frac{x}{3} & (0 \leq x \leq 3) \\ 0 & (x > 3) \end{cases}$$

is used and $f(-x) = f(x)$. For the example of (14), $N(a_1,b_1) = 1/3$, $N(a_2,b_1) = 2/3$, *etc*. we hence have

$$p(a,b) = \sum_{i=1}^{5}\sum_{j=1}^{7} N(a_i,b_j) = 4.$$

3.3 Fuzzy neighborhood using a hierarchical structure

Another class for fuzzy neighborhood uses a hierarchical structure of a text, that is, a text frequently has chapters, sections, subsections, and paragraphs. Hence the next distance can be induced using $0 < \alpha < \beta < \gamma$.

- If o and o' occurs in a same paragraph, $d(o,o') = \alpha$;
- else if o and o' occurs in a same subsection, $d(o,o') = \beta$;
- else if o and o' occurs in a same section, $d(o,o') = \gamma$;
- else $d(o,o') = \infty$.

Such a distance induced from a hierarchical classification satisfies

$$d(o,o'') \leq \max\{d(o,o'), d(o',o'')\}. \tag{17}$$

which is called ultra-metric property. It is also known that this property is equivalent to fuzzy equivalence that is also called fuzzy similarity. That is,

$$N(o,o') = f(d(o,o')) \tag{18}$$

is a fuzzy equivalence.

We have the next proposition.

Proposition 6. *If $d(o,o')$ satisfies (17) and $N(o,o')$ is defined by (18), then $P = (p(t,t'))$ is positive-definite. Namely, we can define an inner product space using this P.*

ISBN: 978-989-95079-6-8

Let us moreover consider a particular case of d when O consists of documents $D_1, \ldots, D_m$:

$$d(o, o') = \begin{cases} 0 & (o \text{ and } o' \text{ are in a same document}), \\ \infty & (o \text{ and } o' \text{ are in different documents}). \end{cases} \tag{19}$$

Suppose also that the frequency of occurrences of t in D_i is $F_i(t)$.

We have the next proposition.

Proposition 7. *If $d(o, o')$ is given by (19) and $N(o, o')$ is defined by (18). then*

$$p(t, t') = \sum_{i=1}^{m} F_i(t) F_i(t'), \tag{20}$$

which is used in the vector space model.

We thus note that the model in this section generalizes the vector space model.

Finally, we note that the above model can directly be applied to a document set $D = \{d_1, \ldots, d_m\}$ using

$$R_D(d, t) = \begin{cases} 1 & (\ t \text{ occurs in } d), \\ 0 & (\ t \text{ doe no not occur in } d). \end{cases}$$

Then we have a positive definite measure between two documents:

$$p_D(d, d') = \sum_{t, t' \in T} R_D(d, t) p(t, t') R_D(d', t'), \tag{21}$$

and a normalized measure

$$s_D(d, d') = \frac{p_D(d, d')}{\sqrt{p_D(d, d) p_D(d', d')}}, \tag{22}$$

which can be used as inner products of the document space.

4 Numerical Examples

We briefly show the results of kernel fuzzy c-means clustering (KFCM) with the kernel principal components (KPCA) applied to a set of documents extracted from a Japanese newspaper articles. Sixty articles from ASAHI Shinbun (ASAHI News) have been used. They are categorized by the publisher into three categories of 'Economics', 'Politics', and 'Social Affairs'; each category has 20 articles. Each of these categories are subdivided into subcategories of 'economic statistics', 'finance', and so on. Terms occurred more than twice were used and the number of terms is 556. The above measure (22) has been used throughout. The method of fuzzy c-means with $m = 1.2$ and with $c = 3$ has been applied. The following two neighborhood functions have been used.

- First type uses a convex function $f(x) = 1.2^{-d(o,o')}$.
- Second type uses a fuzzy equivalence: If two term occurrences are the same term, $N(o, o') = 1$; if they are in a same paragraph, $N(o, o') = 0.9$; if they are in a same document, $N(o, o') = 0.7$; if they are in a same subcategory: $N(o, o') = 0.5$; if they are in a same category: $N(o, o') = 0.3$.

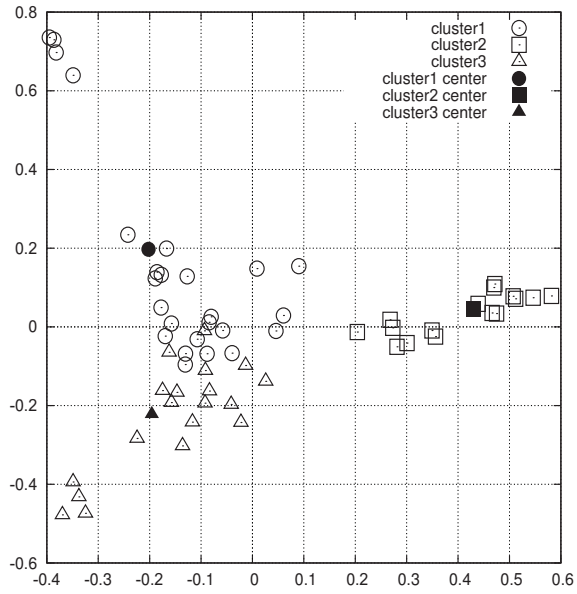


Figure 1: Two-dimensional display from KPCA with three clusters and cluster centers, where the first type of neighborhood function $f(x) = 1.2^{-d(o,o')}$ has been used.

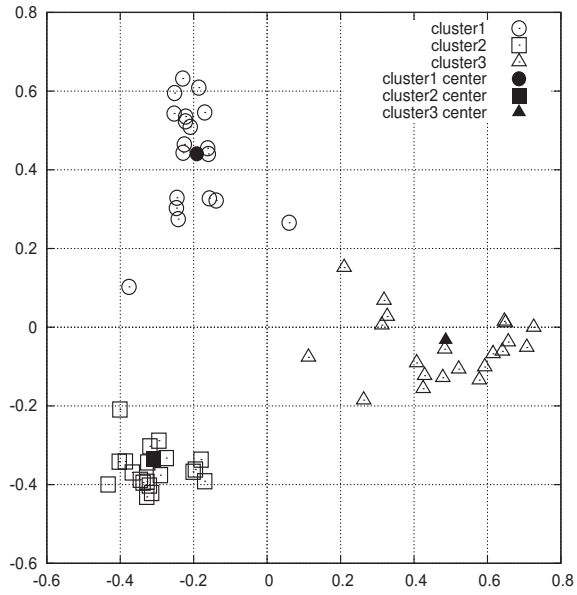


Figure 2: Two-dimensional display from KPCA with three clusters and cluster centers, where the second type of fuzzy equivalence has been used.

Figures 1 and 2 respectively show the two-dimensional figures of two-major axes from KPCA using the first and the second types of neighborhoods. The three symbols of white squares, triangles, and circles show the three clusters obtained from KFCM: fuzzy clusters have been made crisp by the maximum membership rule

$$x_k \to \textit{cluster } i \iff u_{ki} = \max_{1 \le j \le c} u_{kj}.$$

The black square, triangle, and circle are the cluster centers of the corresponding clusters. It seems that the second neigh-

ISBN: 978-989-95079-6-8

borhood of fuzzy equivalence divides more clearly the three clusters.

We have checked the correspondence of the obtained clusters with the actual classification provided from the publisher. The six correspondences have all been checked and the maximum correspondence percentages are given next. In the calculation 1000 trials with random initial values were used and the average numbers are shown below. That is, when we used the first type of neighborhood, the correct classification ratio was 73.5%, while the second type of fuzzy equivalence produced 97.9%.

We also tested a subclass of the above articles: 20 'Economics' and 20 'Politics' articles with the same conditions except that the number of clusters $c = 2$. The number of terms which occurred more than twice is 390. The results are shown as Figures 3 and 4. The separation was better than the previous case of three clusters. The correct classification ratio was 93.2% for the first type of neighborhood and 99.8 for the second neighborhood.

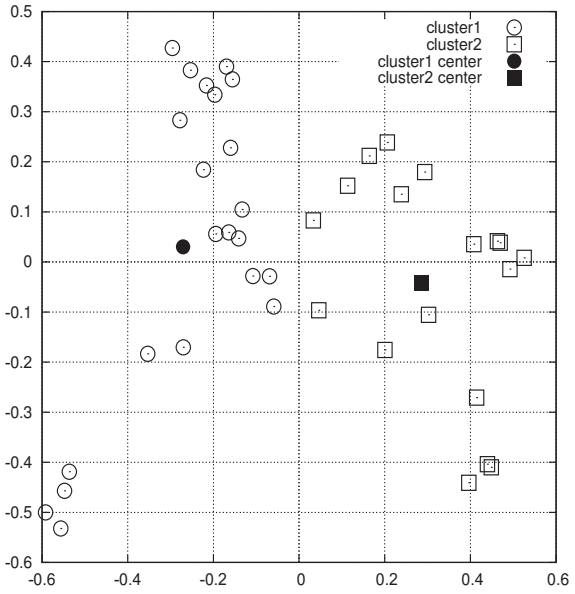


Figure 3: Two-dimensional display from KPCA with two clusters and cluster centers, where the first type of neighborhood function $f(x) = 1.2^{-d(o,o')}$ has been used.

5 Conclusion

We have proposed the use of an explicit mapping for kernel based data analysis. To summarize, we note the following advantages of the present method.

1. Using this mapping, we do not lose any fundamental information in kernel data analysis.
2. Generally the derivation becomes simpler using this mapping.
3. New applications of the kernel methods become easier using this mapping.

The last statement should be put into practice. In relation to fuzzy clustering, the method of fuzzy c-varieties should be

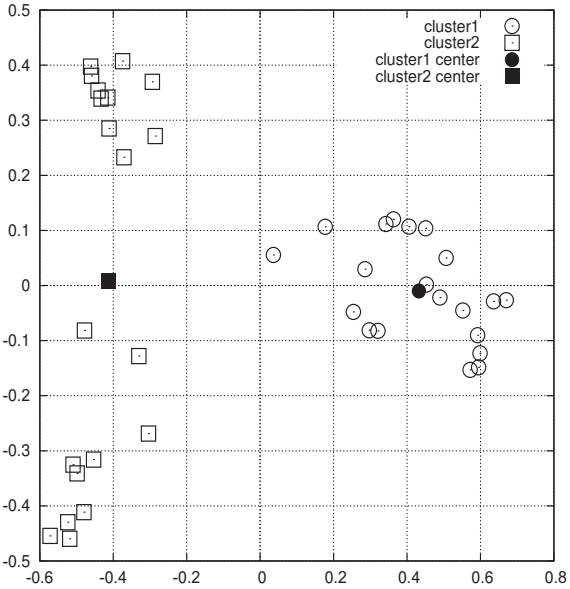


Figure 4: Two-dimensional display from KPCA with two clusters and cluster centers, where the second type of fuzzy equivalence has been used.

studied. We moreover have many research possibilities related to SOM.

This mapping invokes several problems to be solved. For example, when the number of objects are large, we have a problem of many dimensions which should be overcome using some handling large matrix techniques. We also should consider when and where such an explicit mapping become useless. Clarification of such a boundary between usefulness and uselessness is the ultimate objective of the present study.

Acknowledgment

This research has partly been supported by the Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science, Japan, No.19300074.

Appendix: Notes on Proofs

Since detailed proofs are too lengthy, we give notes on how to prove the propositions given here. The proofs are mostly straightforward.

Note on proof of Proposition 1

The detailed proof is given in standard textbooks [9]. As a rough sketch of the proof, note that the Mercer condition[10]

$$\int \int \boldsymbol{K}(x,y)\eta(x)\eta(y)dxdy \geq 0$$

for all $\eta(x)$ guarantees

$$\sum_{i,j} \boldsymbol{K}(x_i,x_j)\zeta_i\zeta_j \geq 0$$

$\forall \zeta_i \in \boldsymbol{R}$, by putting $\eta(x) = \sum_i \zeta_i \delta(x - x_i)$. Thus the matrix $K = (\boldsymbol{K}(x_i,x_j))$ is positive semi-definite. The kernel function generally does not distinguish positive semi-definiteness

ISBN: 978-989-95079-6-8

and positive definiteness, while positive-definiteness is required for the definition of an inner product. For this purpose a simple trick is to use a regularization which means that we take

$$K \to K + \epsilon I \quad (\epsilon > 0)$$

where ϵ is sufficiently small. The matrix $K + \epsilon I$ is positive definite and approximates K when K is positive semi-definite. In this paper we thus assume positive-definiteness throughout, by using such a regularization when needed.

Note on proof of Proposition 2

The proof of this proposition is not difficult by observing closely (6). Note that W_i given by (6) is the solution of

$$\min_{W_i} \sum_k (u_{ki})^m \|x_k - W_i\|^2$$

where the space can be an arbitrary inner product space, since the derivation of (6) uses a general variational principle valid for any Hilbert space. Note moreover that we substitute (11) into (5) to have the optimal solution of U. It should be noted that although optimal W_i is the same for all positive definite kernel, optimal U differs because (10) give different values for different kernels.

Note on proof of Proposition 3

A straightforward calculation shows this proposition is valid.

Note on proof of Proposition 4

The proof is immediate and omitted here.

Note on proof of Proposition 5

Take an arbitrary $c \in O$. The conclusion is immediately obtained from the Pólya's theorem [7] which states that

$$\sum_{a,b} z_a z_b f(|x(a) - x(b)|) \geq 0$$

when f is convex on $[0, +\infty)$ and $f(x) \to 0$ as $x \to +\infty$. Note that $x(a)$ is the real value defined by $x(a) = D(a, c)$ when a is the left hand of c; $x(a) = -D(a, c)$ when a is right hand of c.

Note on proof of Proposition 6

To prove this proposition, we consider a partition matrix $U = (u_{ij})$. Namely, an $n \times n$ real matrix U is called a partition matrix iff there exists a partition $K_1, \ldots, K_c$ of $\mathbf{n} = \{1, 2, \ldots, n\}$ ($\bigcup_j K_j = \mathbf{n}$ and $K_i \cap K_j = \emptyset$, for $i \neq j$) such that

$$u_{ij} = 1 \quad \forall i, j \in K_h$$

for some h and

$$u_{ij} = 0 \quad \forall i \in K_\ell, j \in K_h$$

for $h \neq \ell$.

We have

Lemma: *A partition matrix U is positive semi-definite. It is positive definite if and only if U is identity matrix ($U = I$).*

The proof of the lemma is immediate by observing

$$x^t U x = \sum_{i=1}^{c} \left(\sum_{x_j \in K_i} x_j \right)^2 .$$

The proof of this proposition is now straightforward. We assume that O is a finite set for simplicity. Then an equivalence relation is represented by a partition matrix. Moreover a fuzzy equivalence relation F is represented by a finite collection $U_1, \ldots, U_k$ of partition matrix and positive $\beta_1, \ldots, \beta_k$:

$$F = \sum_{j=1}^{k} \beta_j U_j.$$

Consequently we have

$$x^t F x = \sum_{j=1}^{k} \beta_j x^t U_j x \geq 0.$$

Hence $N(a, b)$ is positive definite. From Proposition 6, $p(t, t')$ is also positive definite. The proposition is thus proved.

Note on proof of Proposition 7

By straightforward calculation, we see this proposition holds.

References

[1] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Functions, Plenum Press, New York, 1981.

[2] T. Kohonen, Self-Organizing Maps, 2nd Ed., Springer, Berlin, 1995.

[3] S. Miyamoto, D. Suizu, Fuzzy c-means clustering using transformations into high dimensional spaces, Proc. of FSKD'02: 1st International Conference on Fuzzy Systems and Knowledge Discovery, Nov. 18-22, 2002, Singapore, Vol.2, pp. 656–660.

[4] S. Miyamoto, H. Ichihashi, K. Honda, Algorithms for Fuzzy Clustering, Springer, Berlin, 2008.

[5] S. Miyamoto, Y. Kawasaki, Kernel Space for Text Analysis Based on Fuzzy Neighborhoods, *WCCI 2008 Proceedings, 2008 IEEE World Congress on Computational Intelligence*, June 1-6, 2008, Hong Kong, pp.738-743, 2008.

[6] K. Mizutani, S. Miyamoto, Fuzzy multiset space and c-means clustering using kernels with application to information retrieval. in T.Bilgiç et al. Eds.: IFSA 2003, LNAI2715, Springer, Berlin, pp.387-395, 2003.

[7] G. Pólya, Remarks on characteristic functions, Proc. First Berkeley Conf. on Math. Stat. and Prob., pp.115–123, Univ. of Calf Press, Berkeley, 1949.

[8] G. Salton, M.J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, New York, 1985.

[9] B. Schölkopf, A.J. Smola, Learning with Kernels, the MIT Press, Cambridge, 2002.

[10] V.N. Vapnik, Statistical Learning Theory, Wiley, 1998.

ISBN: 978-989-95079-6-8

Distorted probabilities and m-separable fuzzy measures

Yasuo Narukawa[1] Vicenç Torra[2]

1.Toho Gakuen, 3-1-10, Naka, Kunitachi, Tokyo, 186-0004, Japan
1.The Institute for Research in Business Administration, Waseda University,
1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo,169-8050, Japan
2.IIIA - CSIC,
Campus UAB s/n, 08193 Bellaterra, Catalonia, Spain
Email: narukawa@d4.dion.ne.jp, vtorra@iiia.csic.es

***Abstract*— *Distorted probabilities are an important family of fuzzy measures. In a recent paper we introduced m-dimensional distorted probabilities, which generalize the former and permit us to have a smooth transition from distorted probabilities to unconstrained ones.*

In this paper we introduce the union condition and the strict union condition, and we show that when these conditions hold for a fuzzy measure, the fuzzy measure is a distorted probability. In addition, we present a few results that establish some relationships between other fuzzy measures.

Keywords: Fuzzy measures, distorted probabilities, m-dimensional distorted probabilities, m-symmetric fuzzy measures, union condition, strict union condition.

1 Introduction

Aggregation operators [28] are used in a large number of applications to combine information from different sources. Although the weighted mean is probably the most well known aggregation operator, other operators exist. E.g., the OWA, the WOWA and some fuzzy integrals. Choquet and Sugeno integrals are examples of such fuzzy integrals.

Fuzzy integrals permits the user to combine information when the sources supplying the information are not independent. To express this a priori knowledge about the sources, fuzzy integrals combine the input data with the information about the sources. Formally, the integrals integrate a function, which represents the data being aggregated, with respect to a fuzzy measure, which represents the a priori information about the sources.

A major difficulty for applying fuzzy integrals in real applications is that they are set functions, and thus, for any aggregation problem with n inputs, 2^n values should be defined. In fact, properly speaking, boundary conditions on the measure reduce this number to $2^n - 2$.

Real applications try to reduce the number of $2^n - 2$ required values using constrained measures. That is, measures that require less than $2^n - 2$ parameters. Sugeno λ-measures are probably the most used ones [22, 23, 20, 21]. Such measures solely require n values as well as an additional parameter λ, which can be deduced from the n values as [10] shows. k-order additive fuzzy measures are another family of measures with reduced complexity. This family, that has been extensively studied, is of special interest because the parameter k permits us to find a trade-off between expressiveness and complexity. In short, when $k = 1$ the measure has the lowest complexity (only $n - 1$ values are required) but it corresponds to a probability distribution. Instead, when $k = n$, any unconstrained fuzzy measure can be represented but at the cost of the highest complexity ($2^n - 2$ values should be defined).

In this paper we study m-dimensional distorted probabilities, another family of measures, introduced in [16], that generalizes distorted probabilities. Informally, distorted probabilities are measures that can be expressed in terms of a probability distribution and a function that distorts this distribution. Such measures, which were originated in psychology [19, 4, 5], have been extensively used. See, for example, the book by Aumann and Shapley (1974) [1] and their recent use in game theory. Recent research on such measures is on their determination either from examples (as in [27]) or from interviews [9]. Nevertheless, the modeling capabilities of distorted probabilities are limited. In [16] it was shown that the number of such measures with respect to the number of unconstrained ones is very small. Moreover, the larger is n, the smaller is the proportion of distorted probabilities. So, in most cases, fuzzy measures cannot be represented using distorted probabilities. To overcome this problem, we introduced in [16] m-dimensional distorted probabilities.

In this work we present some new results with respect to this family of measures. We show some conditions that, when fulfilled, imply distorted probabilities.

Other families of fuzzy measures have been studied in the literature. Two of them, that are relevant for the present paper, are the m-symmetric fuzzy measures [11, 12] and the hierarchically decomposable ones [26]. Some results establishing some relationships between these measures and distorted probabilities will be given.

The structure of the paper is as follows. In Sections 2, we review some concepts that are needed later on. In Section 3, we present the results establishing the connections among different kinds of fuzzy measures. The paper finishes with some conclusions.

2 Preliminaries

This section reviews some previous results in the literature that are needed in the rest of the paper. We start by defining fuzzy measures, and some of their families. Among them, we review m-dimensional distorted probabilities and a few results concerning these measures. The section finishes with a review of a few aggregation operators that are relevant for the purpose of this paper.

ISBN: 978-989-95079-6-8

2.1 Fuzzy measures

In this paper we will consider fuzzy measures on a finite universal set $X = \{x_1, \ldots, x_n\}$. For the sake of simplicity, when possible, we will consider $X := \{1, \ldots, n\}$. Now, we review the definition of fuzzy measure.

Definition 1 *A set function $\mu : 2^X \to [0, 1]$ is a fuzzy measure if it satisfies the following axioms:*

(i) $\mu(\emptyset) = 0$, $\mu(X) = 1$ *(boundary conditions)*

(ii) $A \subseteq B$ *implies* $\mu(A) \leq \mu(B)$ *(monotonicity)*

In order to distinguish measures satisfying (i) and (ii) with others that also satisfy some additional constraints (*e.g.* additivity $\mu(A \cup B) = \mu(A) + \mu(B)$ when $A \cap B = \emptyset$), we use the terms *unconstrained* fuzzy measures for the former ones and *constrained* fuzzy measures for the others.

2.2 m-symmetric fuzzy measures

The definition of these measures is based on the concept of *set of indifference*. Roughly speaking, a set of indifference is defined by elements that do not affect the value of the measure. That is, the elements of a set are indistinguishable with respect to the fuzzy measure.

Definition 2 *[11, 12] Given a subset A of X, we say that A is a set of indifference if and only if:*

$$\forall B_1, B_2 \subseteq A, |B_1| = |B_2|,$$

$$\forall C \subseteq X \setminus A \quad \mu(B_1 \cup C) = \mu(B_2 \cup C)$$

In this definition $|\cdot|$ corresponds to the cardinality of a set. We now consider m-symmetric fuzzy measures for the particular case of $m = 2$ and, then, we give the general definition.

Definition 3 *[11, 12] Given a fuzzy measure μ, we say that μ is an at most 2-symmetric fuzzy measure if and only if there exists a partition of the universal set $\{X_1, X_2\}$, with $X_1, X_2 \neq \emptyset$ such that both X_1 and X_2 are sets of indifference. An at most 2-symmetric fuzzy measure is 2-symmetric if X is not a set of indifference.*

Definition 4 *[11, 12] Given a fuzzy measure μ, we say that μ is an at most m-symmetric fuzzy measure if and only if there exists a partition of the universal set $\{X_1, \ldots, X_m\}$, with $X_1, \ldots, X_m \neq \emptyset$ such that $X_1, \ldots X_m$ are sets of indifference.*

The next proposition follows from this definition.

Proposition 1 *Every fuzzy measure μ is an at most n-symmetric fuzzy measure for $n = |X|$.*

So, all fuzzy measures can be considered as m-symmetric for a value of m large enough.

Definition 5 *[11, 12] Given two partitions $\{X_1, \ldots, X_p\}$ and $\{Y_1, \ldots, Y_r\}$ on the finite universal set X, we say that $\{X_1, \ldots, X_p\}$ is coarser than $\{Y_1, \ldots, Y_r\}$ if the following holds:*

$$\forall X_i \exists Y_j \text{ such that } Y_j \subseteq X_i$$

Definition 6 *Given a fuzzy measure μ, we say that μ is m-symmetric if and only if the coarsest partition of the universal set in sets of indifference contains m non empty sets. That is, the coarsest partition is of the form: $\{X_1, \ldots, X_m\}$, with $X_i \neq \emptyset$ for all $i \in \{1, \ldots, m\}$.*

Proposition 2 *[11, 12] Let μ be an m-symmetric measure with respect to the partition $\{X_1, \ldots, X_m\}$. Then, the number of values that are needed in order to determine μ is:*

$$\big[(|X_1| + 1) \cdots (|X_m| + 1)\big] - 2$$

An m-symmetric fuzzy measure can be represented in a $(|X_1| + 1) \cdots (|X_m| + 1)$ matrix M.

2.3 Hierarchically S-Decomposable Fuzzy Measures

[26] introduced Hierarchically S-Decomposable Fuzzy Measures. These measures (HDFM for short) can be seen as a generalization of S-decomposable measures. An important characteristic of S-decomposable fuzzy measures, is that the measure for any subset of X can be built from the measures on the singletons and a t-conorm S. When interactions among information sources are considered, such construction means that the interactions among pairs (or subsets) of sources can be expressed in a single and unique way. In particular, all interactions are modeled using the t-conorm S.

The so-called hierarchically S-decomposable fuzzy measures define a more general family of fuzzy measures as they permit us to express different kind of interactions between different subsets. This is achieved permitting us the use of different t-conorms for combining the measures of different singletons (and of different subsets).

This is obtained as follows: (i) the elements in X are structured in a hierarchy that gathers together elements that are *similar* (from the interactions point of view); (ii) each node of the hierarchy has associated a t-conorm to be used to *combine* the interactions. In this way, a richer variety of interactions can be expressed.

For example, if we have a fuzzy measure $\mu : 2^X \to [0, 1]$ with $X = \{x_1, x_2, x_3, x_4, \ldots, x_n\}$ such that $\mu(\{x_1\}) = 0.2$, $\mu(\{x_2\}) = 0.4$, $\mu(\{x_3\}) = 0.3$ and $\mu(\{x_4\}) = 0.3$. Then, we have a negative interaction between x_1 and x_2 defining $\mu(\{x_1, x_2\}) = \max(\mu(\{x_1\}), \mu(\{x_2\}))$. Instead, for a positive interaction between x_3 and x_4 we define $\mu(\{x_3, x_4\}) = \min(1, \mu(\{x_3\}) + \mu(\{x_4\}))$. Both situations can be modeled with the t-conorms $S_1(x, y) = \max(x, y)$ and $S_2(x, y) = \min(1, x + y)$.

We give below the definition for the particular case of 2-level HDFM. That is, a measure where the hierarchy has only two levels.

Definition 7 *[26] Given a fuzzy measure μ, we say that μ is a 2-level Hierarchically Decomposable Fuzzy Measure (2-level HDFM) if there is a partition $\{X_1, \ldots, X_m\}$ on X (we denote the elements in X_i by $X_i = \{x_{i,1}, \ldots, x_{i,m_i}\}$) and t-conorms S, S_1, …, S_m such that:*

$$\mu(A) = S(r_1(A), \ldots, r_m(A))$$

where

$$r_i(A) = S_i(\mu(\{x_{i,1}\} \cap A), \ldots, \mu(\{x_{1,m_i}\} \cap A))$$

ISBN: 978-989-95079-6-8

In the general case of HDFM, not presented here, a complete hierarchy is permitted and, then, the measure is defined recursively for each node using the t-conorm attached to the node, and the partition associated to the node.

2.4 Distorted probabilities

As briefly described in the introduction, distorted probabilities correspond to fuzzy measures that can be represented by a probability distribution and a distortion function. We formalize these measures as well as the required concepts below:

Definition 8 *Let $P : 2^X \to [0,1]$ be a probability measure. Then, we say that a function f is strictly increasing with respect to P if and only if*

$$P(A) > P(B) \text{ implies } f(P(A)) > f(P(B))$$

Remark: Since we suppose that X is a finite set, when there is no restriction on the function f, a strictly increasing function f with respect to P can be regarded as a strictly increasing function on $[0,1]$. Note that with respect to increasingness only the points in $\{P(A)|A \in 2^X\}$ are essential, the others are not considered by $f(P(A))$.

Definition 9 *[1, 2] Let μ be a fuzzy measure. We say that μ is a distorted probability if there exists a probability distribution P and a strictly increasing function f with respect to P such that $\mu = f \circ P$.*

The next theorem gives the necessary and sufficient condition for a fuzzy measure μ to be a distorted probability. The theorem is based on *Scott's condition*:

Definition 10 *[16] Let μ be a fuzzy measure, μ satisfies* Scott's condition *when for all $A_i, B_i \in 2^X$ such that $\sum_{i=1}^n 1_{A_i} = \sum_{i=1}^n 1_{B_i}$ the condition below holds:*
$\mu(A_i) \leq \mu(B_i)$ for $i = 2, 3, \ldots, n$ implies $\mu(A_1) \geq \mu(B_1)$.

Here 1_A represents the characteristic function of the set A. That is $1_A(x) = 1$ if and only if $x \in A$.

Using this condition, we can characterize distorted probabilities as follows:

Theorem 1 *[16] Let μ be a fuzzy measure; then, μ is a distorted probability if and only if Scott's condition holds.*

2.5 m-dimensional distorted probabilities

m-dimensional distorted probabilities were presented in [16] to overcome the limited expressiveness of distorted probabilities. They are defined as follows:

Definition 11 *[16] Let $\{X_1, X_2, \cdots, X_m\}$ be a partition of X; then, we say that μ is an at most m dimensional distorted probability if there exists a function f on $\mathbb{R}^m$ and probabilities P_i on $(X_i, 2^{X_i})$ such that:*

$$\mu(A) = f(P_1(A \cap X_1), P_2(A \cap X_2), \cdots, P_m(A \cap X_m)) \quad (1)$$

where f on $\mathbb{R}^m$ is strictly increasing with respect to each variable.

We say that an at most m dimensional distorted probability μ is an m dimensional distorted probability if μ is not an at most $m-1$ dimensional.

The next proposition follows from the definition above.

Proposition 3 *Every fuzzy measure is an at most n-dimensional distorted probability with $n = |X|$.*

Note that for $n = |X|$, we are considering the following partition of X: $\{X_1 = \{x_1\}, \ldots, X_n = \{x_n\}\}$. So, $f(a_1, \ldots, a_n) = \mu(A)$ when $a_i = 1$ if and only if $x_i \in A$.

Also, we can prove that m-dimensional distorted probabilities define a family of measures with increasing complexity with respect to m. This means that increasing the value of m, the number of measures representable increases. The following proposition establishes this property.

Proposition 4 *Let $\mathcal{M}_k$ be the set of all fuzzy measures that are k-dimensional distorted probabilities and let $\mathcal{M}_0$ be the empty set. Then $\mathcal{M}_{k-1} \subset \mathcal{M}_k$ for all $k = 1, 2, \ldots, |X|$.*

Corollary 1 *Given a fuzzy measure μ, there exists a $k \in \{1, 2, \ldots, |X|\}$ such that $\mu \in \mathcal{M}_k$ and $\mu \notin \mathcal{M}_{k-1}$.*

Therefore, the proposed family of fuzzy measures permits us to cover the whole set of fuzzy measures.

2.6 Aggregation operators

Now we define the OWA and the WOWA operators. They will be of relevance in this work. As explained in detail in [25], the OWA operator permits to give importance to the data (with respect to their position) while the WOWA permits to give importance to the data (as the OWA operator) and also to the information sources (as the weighted mean does).

Definition 12 *[29, 30] Let $\mathbf{w}$ be a weighting vector of dimension n (*i.e.*, $w_i \geq 0$ and $\sum_{i=1}^n w_i = 1$), then a mapping OWA: $\mathbb{R}^n \to \mathbb{R}$ is an* Ordered Weighted Averaging (OWA) operator *of dimension n if*

$$OWA_{\mathbf{w}}(a_1, ..., a_n) = \sum_{i=1}^{n} w_i a_{\sigma(i)}$$

where $\{\sigma(1), ..., \sigma(n)\}$ is a permutation of $\{1, ..., n\}$ such that $a_{\sigma(i-1)} \geq a_{\sigma(i)}$ for all $i = \{2, ..., n\}$ (i.e. $a_{\sigma(i)}$ is the i-th largest element in the collection $a_1, ..., a_n$).

Definition 13 *[25] Let $\mathbf{p}$ and $\mathbf{w}$ be two weighting vectors of dimension n, then a mapping WOWA: $\mathbb{R}^n \to \mathbb{R}$ is a* Weighted Ordered Weighted Averaging (WOWA) operator *of dimension n if*

$$WOWA_{\mathbf{p},\mathbf{w}}(a_1, ..., a_n) = \sum_{i=1}^{n} \omega_i a_{\sigma(i)}$$

where σ is defined as in the case of the OWA, and the weight ω_i is defined as:

$$\omega_i = w^*(\sum_{j \leq i} p_{\sigma(j)}) - w^*(\sum_{j<i} p_{\sigma(j)})$$

with w^* being a non-decreasing function that interpolates the points $\{(i/n, \sum_{j \leq i} w_j)\}_{i=1,\ldots,n}$ together with the point $(0,0)$. The function w^* is required to be a straight line when the points can be interpolated in this way.

ISBN: 978-989-95079-6-8

Alternatively, it is possible to define the WOWA operator directly using the function w^*. This will be denoted by $WOWA_{\mathbf{p},w^*}(a_1,\ldots,a_n)$ or $WOWA_{\mathbf{p},w^*}(f)$ when f is a function $f : X \to \mathbb{R}$ such that $f(x_i) = a_i$. In this latter case, we will read $WOWA_{\mathbf{p},w^*}(f)$ as the $WOWA$ of f with respect to $\mathbf{p}$ and w^*.

We finish the section with the definition of the Choquet integral.

Definition 14 *[3] Let μ be a fuzzy measure, then the* Choquet integral *of a function $f : X \to \mathbb{R}^+$ with respect to the fuzzy measure μ is defined by:*

$$(C)\int f d\mu (= C_\mu(f)) = \sum_{i=1}^{n} [f(x_{s(i)}) - f(x_{s(i-1)})]\mu(A_{s(i)})$$

where $x_i \in X$ and where $f(x_{s(i)})$ indicates that the indices have been permuted so that
$0 \le f(x_{s(1)}) \le \cdots \le f(x_{s(n)}) \le 1$,
$A_{s(i)} = \{x_{s(i)},\ldots,x_{s(n)}\}$ *and* $f(x_{s(0)}) = 0$.

3 Distorted probabilities and symmetric fuzzy measures

3.1 A sufficient condition for distorted probablity

We have discussed above Scott's condition: Definition 10 for distorted probability. Nevertheless, this condition is not easy to check. We present below another condition under which a fuzzy measure is a distorted probability. The new condition is easier to check.

Definition 15 *We say that a fuzzy measure μ satisfies a union condition (for short UC), if $A \cap C = \emptyset, B \cap D = \emptyset$,*

$$\mu(A) \ge \mu(B), \mu(C) \ge \mu(D) \Rightarrow \mu(A \cup C) \ge \mu(B \cup D).$$

The next proposition is obvious from the definition.

Proposition 5 *Suppose that a fuzzy measure μ satisfies UC. We have $A \cap C = \emptyset, B \cap D = \emptyset$,*

$$\mu(A) = \mu(B), \mu(C) = \mu(D) \Rightarrow \mu(A \cup C) = \mu(B \cup D).$$

Definition 16 *We say that a fuzzy measure μ satisfies a strict union condition (for short SUC), if $A \cap C = \emptyset, B \cap D = \emptyset$,*

$$\mu(A) > \mu(B), \mu(C) \ge \mu(D) \Rightarrow \mu(A \cup C) > \mu(B \cup D).$$

Proposition 6 *Suppose that a fuzzy measure μ satisfies SUC. We have $A \cap C = \emptyset, B \cap C = \emptyset$,*

$$\mu(A) > \mu(B) \Rightarrow \mu(A \cup C) > \mu(B \cup C).$$

Applying Proposition 5,6,we have the next proposition.

Proposition 7 *Suppose that a fuzzy measure μ satisfies UC. There exists a function F on*

$$\{(x,y) | x = \mu(A), y = \mu(B), A \cap B = \emptyset, A, B \in 2^X\}$$

such that

$$\mu(A \cup B) = F(\mu(A), \mu(B))$$

for $A, B \in 2^X, A \cap B = \emptyset$, and

$$F(x,0) = x, F(F(x,y),z) = F(x,F(y,z)).$$

Moreover if μ satisfies both UC and SUC, F is strictly monotone with respect to each variable.

Suppose that a fuzzy measure μ satisfies both UC and SUC. Since F is strictly monotone on

$$\{(x,y) | x = \mu(A), y = \mu(B), A \cap B = \emptyset, A, B \in 2^X\},$$

the domain of F can be extended to $[0,1] \times [0,1]$ and F is monotone with respect to each variable and F is continuous. Then F can be represented by strictly monotone function φ on [0,1] as

$$F(x,y) = \varphi^{-1}(\varphi(x) + \varphi(y)).$$

That is,

$$F(\mu(A), \mu(B)) = \varphi^{-1}(\varphi(\mu(A)) + \varphi(\mu(B)))$$

for $A \cap B = \emptyset$. Therefore we have

$$\varphi(\mu(A \cup B)) = \varphi(\mu(A)) + \varphi(\mu(B)).$$

Let $P(A) := \varphi(\mu(A))$. Then, P is a probability and we have $\mu(A) = \varphi^{-1}(P(A))$. Therefore we have the next theorem.

Theorem 2 *A fuzzy measure μ is distorted probability if μ satisfies both UC and SUC.*

3.2 Symmetric fuzzy measure

We start showing that a 1-Symmetric fuzzy measure is a special case of distorted probabilities.

Proposition 8 *Let $\mu = f \circ P$ be a distorted probability. Then, μ is a 1-symmetric fuzzy measure if and only if $P(A) = |A|/|X|$.*

Now we show that all m-symmetric fuzzy measures are m-dimensional distorted probabilities. This implies that 1-symmetric fuzzy measures are distorted probabilities.

Proposition 9 *Let μ be an m-symmetric fuzzy measure with respect to the partition $\{X_1,\ldots,X_m\}$. Then, μ is an m-dimensional distorted probability.*

Although the reversal of this proposition is not true, the next proposition characterizes one case in which m-dimensional distorted probabilities are m-symmetric fuzzy measures.

Proposition 10 *Let μ be an m-dimensional distorted probability. If, $p_i(x_j) = p_i(x_k)$ for all $x_j, x_k \in X_i$ and for all $i = 1,\ldots,m$, then μ is an m-symmetric fuzzy measure.*

It is known that OWA operators are equivalent to Choquet integrals with respect to symmetric fuzzy measures. Therefore, m-symmetric fuzzy measures permit us to define a generalization of OWA operators. The m-dimensional OWA is defined below:

Definition 17 *The m-dimensional OWA is defined as the Choquet integral with respect to an m-symmetric fuzzy measure.*

As proven in [24], a Weighted OWA (WOWA) operator is equivalent to a Choquet integral with respect to a distorted probability. Therefore, a Choquet integral with an m-dimensional probability can be seen as a generalization of the WOWA operator. We define an m-dimensional WOWA as follows:

ISBN: 978-989-95079-6-8

Definition 18 *The m-dimensional WOWA is defined as the Choquet integral with respect to an m-dimensional distorted probability.*

Then, considering Definitions 17 and 18 above, we have the following corollary from Proposition 9:

Corollary 2 *An m-dimensional OWA is a particular case of an m-dimensional WOWA. In other words, a Choquet integral with respect to an m-symmetric fuzzy measure is a particular case of a Choquet integral with respect to an m-dimensional distorted probability.*

3.3 m-separable fuzzy measure

As shown in Proposition 9, m-symmetric fuzzy measures are one special class of the m-dimensional distorted probabilities. We present another special class of fuzzy measure

Definition 19 *Let μ be a fuzzy measure. Then, we say that μ is a m-separable fuzzy measure if there exists a function g and a partition $\{X_1, \ldots, X_m\}$ of X such that*

$$\mu(A) = g(\mu(A \cap X_1), \ldots, \mu(A \cap X_m)) \tag{2}$$

where g is a m-dimensional function on $\mathbb{R}^m$. We say that g is a generating function for μ. We say that a generating function g is induced by h on $[0,1] \times [0,1]$ if $g(x_1, \ldots, x_m) = h(h(\ldots h(x_1, x_2), \ldots, x_{m-1}), x_m)$, $g(x_1, x_2, 0, \ldots, 0) = h(x_1, x_2)$.

Example 1 *Let $\{X_1, \ldots, X_m\}$ be a partition of X.*

1. *Suppose $g(x_1, \ldots, x_m) = x_1 + \cdots + x_m$, so g is induced by $h(x, y) = x + y$. Then we have*

$$\mu(A) = \mu(A \cap X_1) + \cdots + \mu(A \cap X_m).$$

This is an interadditivity defined in [14].

2. *Suppose $g(x_1, \ldots, x_m) = x_1 \vee \cdots \vee x_m$, so g is induced by $h(x, y) = x \vee y$. Then we have*

$$\mu(A) = \mu(A \cap X_1) \vee \cdots \vee \mu(A \cap X_m).$$

3. *Suppose $g(x_1, \ldots, x_m) = (x_1^2 + \cdots + x_m^2)^{1/2}$, so g is induced by $h(x, y) = (x^2 + y^2)^{1/2}$. Then we have*

$$\mu(A) = (\mu(A \cap X_1)^2 + \cdots + \mu(A \cap X_m)^2)^{1/2}.$$

Suppose that μ is a m-separable fuzzy measure generated by g and that g is induced by h. We say that h is associative if $h(h(x, y), z) = h(x, h(y, z))$. Suppose that h is strictly monotone and associative. Since g is symmetric, then h is symmetric, that is $h(x, y) = h(y, x)$. Then there exists a strictly monotone function φ such that $h(x, y) = \varphi^{-1}(\varphi(x) + \varphi(y))$. Define φ- Möbius inverse m^φ by

$$m^\varphi(A) := \sum_{B \subset A} (-1)^{|A \setminus B|} \varphi(\mu(B)).$$

Let $\mathcal{P}$ be a partition of X for a m-separable fuzzy measure. Applying theorem in [14], $m^\varphi(A) = 0 \Rightarrow A \not\subset C, C \in \mathcal{P}$. Let $M := \{A | m^\varphi(A) \neq 0\}$ We have $\mathcal{P} \subset M$. Define $A_l := \{A | A \in M, x_l \in A\}$ for $X = \{x_1, \ldots, x_l, \ldots, x_n\}$ and $M_l := \max\{|A| | A \in A_l\}$. Since $A \subset C, C \in \mathcal{P}$ for $A \in A_l$, we have the next proposition.

Proposition 11 *Let μ be a m - separable fuzzy measure generated by g, and g be induced by a strict monotone and associative h on $[0,1]$. Then we have*

$$m \times \min_{l \in |1,\ldots,n|} M_l \leq n.$$

We say that a fuzzy measure μ is a φ $k-$order additive if $max\{|A| | A \in M\} = k$. If $\varphi(x) = x$, a φ $k-$order additive fuzzy measure is a k-order additive fuzzy measure [6, 7].

Proposition 12 *Let μ be a m-separable fuzzy measure generated by g, and g be induced by strict monotone and associative h on $[0,1]$. If μ is k- additive, then we have $m \times k \geq n$.*

Now, we consider the relationship between the m-separable fuzzy measures and some other families of measures.

Proposition 13 *2-level HDFMs with S_i Archimedean t-conorms are a m-separable fuzzy measures.*

Theorem 3 *Let $\{X_1, \ldots, X_m\}$ be a partition of X and μ_i $i = 1, \ldots m$ be distorted probabilities represented by f_i and P_i (i.e., $\mu_i = f_i \circ P_i$). Then, there exists a m-separable fuzzy measure μ such that*

$$\sum_{i=1}^{m} ((C) \int f df_i \circ P_i) = (C) \int f d\mu \tag{3}$$

for all measurable function f.

As a corollary of this theorem, we have that the Choquet integral with respect to a m-separable fuzzy measure μ with $g(x_1, \ldots, x_m) = x_1 + \cdots + x_m$ can be represented as a two step Choquet integral.

Corollary 3 *Let μ be a m-separable fuzzy measure μ with $g(x_1, \ldots, x_m) = x_1 + \cdots + x_m$. Then the Choquet integral with respect to μ is represented as a two step Choquet integral of a 1st step integral with respect to a probability on $(1, \ldots, m)$. That is,*

$$(C) \int f d\mu = \int ((C) \int f df_i \circ P_i) dP(i). \tag{4}$$

4 Conclusions

In this paper we have introduced two new conditions, the Union Condition (UC) and the Strict Union Condition (SUC), and we have studied distorted probability under these conditions. We have shown that a fuzzy measure is a distorted probability when both UC and SUC conditions are satisfied.

Acknowledgements

Partial support by the Spanish MEC (projects ARES – CONSOLIDER INGENIO 2010 CSD2007-00004 – and eAEGIS – TSI2007-65406-C03-02) is acknowledged.

References

[1] Aumann, R. J., Shapley, L. S., (1974), Values of Non-atomic Games, Princeton Univ. Press.

[2] Chateauneuf, A., Decomposable Measures, (1996), Distorted Probabilities and Concave Capacities, Mathematical Social Sciences 31, 19-37.

[3] Choquet, G., (1953/54), Theory of capacities, Ann. Inst. Fourier, 5 131-295.

[4] Edwards, W., (1953), Probability-Preferences in Gambling, American Journal of Psychology 66 349-364.

[5] Edwards, W., (1962), Subjective probabilities inferred from decisions, Psychological review 69 109-135.

[6] Grabisch, M., (1996), k-order additive fuzzy measures, Proc. 6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU), 1345-1350, Granada, Spain.

[7] Grabisch,M. k-order additive discrete fuzzy measures and their representation. Fuzzy Sets and Systems 92 (1997), 167-189.

[8] Grabisch, M., Murofushi, T., Sugeno, M., (2000), Fuzzy Measures and Integrals: Theory and Applications, Physica-Verlag.

[9] Honda, A., Nakano, T., Okazaki, Y., (2002), Subjective evaluation based on distorted probability, Proc. of the SCIS/ISIS conference.

[10] Leszczyski, K., Penczek, P., Grochulski, W., (1985), Sugeno's fuzzy measure and fuzzy clustering, Fuzzy Sets and Systems, 15 147-158.

[11] Miranda, P., Grabisch, M., (2002), p-symmetric fuzzy measures, Proc. 8th IPMU, 545-552, Annecy, France.

[12] Miranda, P., Grabisch, M., P. Gil, (2002),p-symmetric fuzzy measures, Int. J. of Unc., Fuzz. and Knowledge-Based Systems, 10 (Supplement), 105-123.

[13] Murofushi, T., Sugeno, M., Machida, M., (1994), Non-monotonic fuzzy measures and the Choquet integral, Fuzzy Sets and Systems, 64:1 73-86.

[14] Murofushi, T., Sugeno, M., Fujimoto, K., (1997), Separated hierarchical decomposition of the Choquet integral, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 5:5 563-585.

[15] Narukawa, Y., Torra, V., (2004), On m-dimensional distorted probabilities and p-symmetric fuzzy measures, IPMU 2004 (ISBN 88-87242-54-2), vol. 2, 1279-1284, Perugia, Italy, July 4-9.

[16] Narukawa, Y., Torra, V., (2005), Fuzzy measure and probability distributions: distorted probabilities, IEEE Trans. on Fuzzy Systems, 13:5 617- 629.

[17] Narukawa, Y., Torra, V., (2005), On multidimensional distorted probability, IFSA 2005, Beijing, China.

[18] Ovchinnikov, S., (1998), An Analytic Characterization of Some Aggregation Operators, Int. J. of Intel. Systems, 13 59-68.

[19] Preston, M. G., Baratta, P., (1948), An Experimental Study of the Auction-Value of an Uncertain Outcome, American Journal of Psychology, 61 183-93.

[20] Soria-Frisch, A., (2003), Hybrid SOM and fuzzy integral frameworks for fuzzy classification, in Proc. IEEE Int. Conf. on Fuzzy Systems 840-845.

[21] Soria-Frisch, A., (2006), Unsupervised Construction of Fuzzy Measures through Self-Organizing Feature Maps and its Application in Color Image Segmentation, Int. J. of Approx. Reasoning, in press.

[22] Sugeno, M., (1974), Theory of fuzzy integrals and its applications, Ph. D. Dissertation, Tokyo Institute of Technology, Tokyo, Japan.

[23] Tahani, H., Keller, J. M., (1990), Information fusion in computer vision using the fuzzy integral, IEEE Trans. on Systems, Man and Cybernetics, 20:3 733-741.

[24] Torra, V., (1998), On some relationships between the WOWA operator and the Choquet integral, Proc. 7th IPMU 818-824, Paris, France.

[25] Torra, V., (1997), The Weighted OWA operator, Int. J. of Intel. Systems, 12, 153-166.

[26] Torra, V., (1999), On hierarchically S-decomposable fuzzy measures, Int. J. of Intel. Systems, 14:9, 923-934.

[27] Torra, V., (2000), Learning Weights for Weighted OWA Operators, Proc. IEEE Int. Conf. Industrial Elect., Control and Instrumentation (IECON 2000), Nagoya, Japan. (CD-ROM).

[28] Torra, V., Narukawa, Y., (2007), Modeling decisions: aggregation operators and information fusion, Springer, in press. ISBN: 978-3-540-68789-4.

[29] Yager, R. R., (1988), On ordered weighted averaging aggregation operators in multi-criteria decision making, IEEE Trans. Syst., Man, Cybern., 18 183-190.

[30] Yager, R. R., (1996), Quantifier Guided Aggregation Using OWA operators, Int. J. of Intel. Systems, 11, 49-73.

[31] Yager, R. R., (1999), Nonmonotonic OWA operators, Soft Computing 3 187-196.

ISBN: 978-989-95079-6-8

Image Analysis Applications of Morphological Operators based on Uninorms

Manuel González Hidalgo[1] Arnau Mir Torres[2] Daniel Ruiz Aguilera[2] Joan Torrens Sastre[2]

[1] Computer Vision, Graphics and Artificial Intelligence Group, Maths. and Computer Science Department.
[2] Fuzzy Logic and Information Fusion Group, Maths. and Computer Science Department,
Universitat de les Illes Balears. Palma de Mallorca, Spain.
Email: {manuel.gonzalez,arnau.mir,dmijts0,daniel.ruiz}@uib.es

***Abstract*—** This paper presents a continuation of the study on a mathematical morphology based on left-continuous conjunctive uninorms given in [1]. Experimental results are displayed using the morphological Top-Hat transformation, used to highlight certain components of the image, and on the reduction and elimination of noise using alternate filters that are generated with the operators of opening and closing associated with these morphological operators

***Keywords*—** Mathematical morphology, Top-Hat, alternate filters, uninorms, representable uninorms, idempotent uninorm.

1 Introduction

The fuzzy mathematical morphology is a generalization of binary morphology [2] using techniques of fuzzy sets [3, 4, 5, 6]. The basic tools of mathematical morphology are the so-called morphological operations that are applied to an image A, which is modified through a structural element B, whose size and shape are chosen in order to study the structure of the image A. The basic morphological operations are: erosion, dilation, opening and closing. Other fuzzy extensions have been introduced by others authors, see fo example [7, 6].

The fuzzy operators used to build a fuzzy morphology are conjunctors and implications. Among them, the most commonly used are the t-norms and their residual implications. Recently conjunctive uninorms, as a particular case of conjunctors, have also been used in this area [8, 9, 1]. In these two last recent works the authors use two kinds of conjunctive uninorms in order to obtain a mathematical morphology with "good" properties, including duality between the morphological operators.

This work can be seen as a continuation of [1] where the authors made a comparative study of the results obtained using the morphology based on uninorms with those obtained using t-norms and using the classical umbra approach. In that work, it is checked out that uninorms detect the edges of the images better than other approaches. Also, in [1], properties of idempotence and generalized idempotence for opening and closing operators based on uninorms are shown. Thus, the aim of this paper is to extend this study to " Top-Hat" operators (residual between openings and closings, a morphological gradient) based on uninorms, as well as to the so-called alternate filters (alternate compositions of openings and closings). The Top-Hat is used to highlight certain components of the image, while the alternate filters are used in elimination and reduction of noise.

The residuals and morphological gradients are still under study and have applications in image analysis. Thus, in [10], it is proposed a directional transition detection algorithm based on morphological residues, considering linear structurant elements. In 2002, T. Chen et al. in [11], design on the basis of mathematical morphology, a new detector of soft edges in dark regions called "Pseudo Top-Hat" which provides, in the case of the classical image of the video cameraman (see Figure 1), a better performance than those achieved by other edge detection methods used to compare the results. In [12], morphological gradients based on openings and closings were used to detect edges in CT medical images altered by noise. The results were compared with other contours detectors, improving the obtained edge images. More recently, in [13, 14] gradient operators are applied again to edge detection and image segmentation, respectively. Thus, in [13], an edge detector is developed for obtaining thin edges in regions or images with low-contrast. Tested with multiple images, even some of them disturbed by noise type salt and pepper and Gaussian noise of mean zero and variance 0.02, the method performs very well in comparison with the other four methods with which they work. In [14] operators were designed based on multiscale morphological gradients, addressing the segmentation problem of images to Computed Axial Tomography (CAT). All these works show that the morphological gradients remain relevant and useful in the analysis and image processing.

The removal, reduction and smoothing of noise is one of the key tasks of analysis and image processing as a preliminary step to artificial vision. Mainly because many techniques of interpretation, measurement, segmentation, detection of structures, among others, fail or diminish its effectiveness in presence of noise. In [12, 13] the designed gradient operators are robust in the presence of noise. Many efforts have been devoted in the literature, and also are currently devoted, to the smoothing, reduction and elimination of noise, from both points of view: crisp and fuzzy. Some examples can be seen in [15, 16, 17]. More specifically, we can find a good job on fuzzy filters in [18] and more recently, in [19], where noise reduction is handled. The design of basic and fast filters remains a subject of interest. In this work it will be studied the feasibility of alternate filters, from opening and closing of the fuzzy morphology based on uninorms, in order to use them in the elimination and reduction of noise.

This work is organized as follows. In the next section we will introduce the fuzzy morphological operators based on left-continuous conjunctive uninorms and their basic properties. Section 3 analyzes the Top-Hat and filtering based on this morphology and shows some experimental results together with their interpretations. Finally, Section 4 is devoted to conclusions and future work.

ISBN: 978-989-95079-6-8

2 Fuzzy morphological operators and its properties

We assume known the basics facts on uninorms used in this work which, in any case, can be found in [20, 21, 22]. We will use the following notation: $\mathcal{I}$ is an implication, $\mathcal{C}$ a conjunction, $\mathcal{N}$ a strong negation, U a conjunctive uninorm with a neutral element e, $\mathcal{I}_U$ its residual implication, and finally A a gray-scale image and B a gray-scale structural element.

We recall the definitions of fuzzy morphological operators following the ideas of De Baets in [3]. The method consists in fuzzify the logical operations, i.e. the Boolean conjunction and the Boolean implication, to obtain fuzzy operators. An n-dimensional gray-scale image is modeled as an $\mathbb{R}^n \longrightarrow [0,1]$ function. The values of an image must be in $[0,1]$ in order to consider it as a fuzzy object. Then, we will proceed to explain this method from the following definitions and propositions.

Definition. The *fuzzy dilation* $D_{\mathcal{C}}(A,B)$ and *fuzzy erosion* $E_{\mathcal{I}}(A,B)$ of A by B are the gray-scale images defined by

$$D_{\mathcal{C}}(A,B)(y) = \sup_x \mathcal{C}(B(x-y), A(x)) \quad (1)$$

$$E_{\mathcal{I}}(A,B)(y) = \inf_x \mathcal{I}(B(x-y), A(x)). \quad (2)$$

Note that the reflection $-B$ of a fuzzy image B is defined by $-B(y) = B(-y)$, for all $y \in \mathbb{R}^n$. Given two images B_1, B_2, we will say that $B_1 \subseteq B_2$ when $B_1(y) \leq B_2(y)$ for all $y \in \mathbb{R}^n$.

Definition. The *fuzzy closing* $C_{\mathcal{C},\mathcal{I}}(A,B)$ and *fuzzy opening* $O_{\mathcal{C},\mathcal{I}}(A,B)$ of A by B are the gray-scale images defined by

$$C_{\mathcal{C},\mathcal{I}}(A,B)(y) = E_{\mathcal{I}}(D_{\mathcal{C}}(A,B), -B)(y) \quad (3)$$

$$O_{\mathcal{C},\mathcal{I}}(A,B)(y) = D_{\mathcal{C}}(E_{\mathcal{I}}(A,B), -B)(y). \quad (4)$$

In this paper, we use as conjunction two types of left-continuous conjunctive uninorms and as implication, their residual implications. Specifically these two types of uninorms are the following.

- The representable uninorms: Let $e \in]0,1[$ and let $h : [0,1] \longrightarrow [-\infty, \infty]$ be a strictly increasing, continuous function with $h(0) = -\infty$, $h(e) = 0$ y $h(1) = \infty$. Then $U_h(x,y) =$

$$\begin{cases} h^{-1}(h(x)+h(y)), & \text{if } (x,y) \notin \{(1,0),(0,1)\}, \\ 0, & \text{in other case,} \end{cases}$$

  is a conjunctive representable uninorm with neutral element e, see [20], and its residual implication $\mathcal{I}_{U_h}$ is given by $\mathcal{I}_{U_h}(x,y) =$

$$\begin{cases} h^{-1}(h(y)-h(x)), & \text{if } (x,y) \notin \{(0,0),(1,1)\}, \\ 1, & \text{in other case.} \end{cases}$$

  Moreover, U_h satisfies self duality (except at the points (0,1) and (1,0)) with respect to the strong negation $\mathcal{N}(x) = h^{-1}(-h(x))$, see [21].

- A specific type of idempotent uninorms: Let $\mathcal{N}$ be a strong negation. The function given by

$$U^{\mathcal{N}}(x,y) = \begin{cases} \min(x,y), & \text{if } y \leq \mathcal{N}(x), \\ \max(x,y), & \text{in other case,} \end{cases}$$

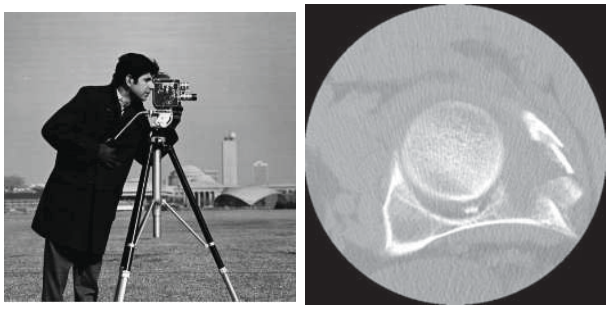


Figure 1: Original images used in experiments.

is a conjunctive idempotent uninorm. Its residual implication is given by (see [22])

$$\mathcal{I}_{U^{\mathcal{N}}}(x,y) = \begin{cases} \min(\mathcal{N}(x), y), & \text{if } y < x, \\ \max(\mathcal{N}(x), y), & \text{if } y \geq x. \end{cases}$$

These two types of conjunctive uninorms guarantees most of the good algebraic and morphological properties associated with the morphological operators obtained from them ([1, 9, 21, 22]). Among them, we highlight those described below. In all properties, U is a left-continuous conjunctive uninorm with neutral element $e \in]0,1[$, $\mathcal{I}_U$ its residual implication, A is a gray-level image and B a gray-scale structural element.

- The fuzzy dilation D_U is increasing in both arguments, the fuzzy erosion $E_{\mathcal{I}}$ is increasing in their first argument and decreasing in their second one, the fuzzy closing $C_{U,\mathcal{I}_U}$ and the fuzzy opening $O_{U,\mathcal{I}_U}$ are both increasing in their first argument.
- If $B(0) = e$ the fuzzy dilation is extensive and the fuzzy erosion is anti-extensive $E_{\mathcal{I}_U}(A,B) \subseteq A \subseteq D_U(A,B)$.
- Moreover, the fuzzy closing is extensive and the fuzzy opening is anti-extensive: $O_{U,\mathcal{I}_U} \subseteq A \subseteq C_{U,\mathcal{I}_U}(A,B)$.
- The fuzzy closing and the fuzzy opening are idempotent, i.e.: $C_{U,\mathcal{I}_U}(C_{U,\mathcal{I}_U}(A,B),B) = C_{U,\mathcal{I}_U}(A,B)$, and $O_{U,\mathcal{I}_U}(O_{U,\mathcal{I}_U}(A,B),B) = O_{U,\mathcal{I}_U}(A,B)$.
- If $B(0) = e$, then $E_{\mathcal{I}_U}(A,B) \subseteq O_{U,\mathcal{I}_U}(A,B) \subseteq A \subseteq C_{U,\mathcal{I}_U}(A,B) \subseteq D_U(A,B)$.
- For the two previous conjunctives uninorms of type U_h and $U^{\mathcal{N}}$, the duality between fuzzy morphological operators is guaranteed.

3 Residuals and basic filters

In the following experiments, the idempotent uninorm $U^{\mathcal{N}}$ with $\mathcal{N}(x) = 1 - x$ and the representable uninorm U_h with $h(x) = \ln\left(\frac{x}{1-x}\right)$ have been used. The obtained results are compared with the Łuckasiewicz t-norm and the classic development based on *umbra approach* (see [7]). In particular, the structural element used by the morphological operators is given by the following matrix:

$$B = e \cdot \begin{pmatrix} 0.86 & 0.86 & 0.86 \\ 0.86 & 1.00 & 0.86 \\ 0.86 & 0.86 & 0.86 \end{pmatrix}$$

ISBN: 978-989-95079-6-8

Figure 2: Left: closing. Right: opening. Top, an idempotent uninorm is used. Down, a representable uninorm is used.

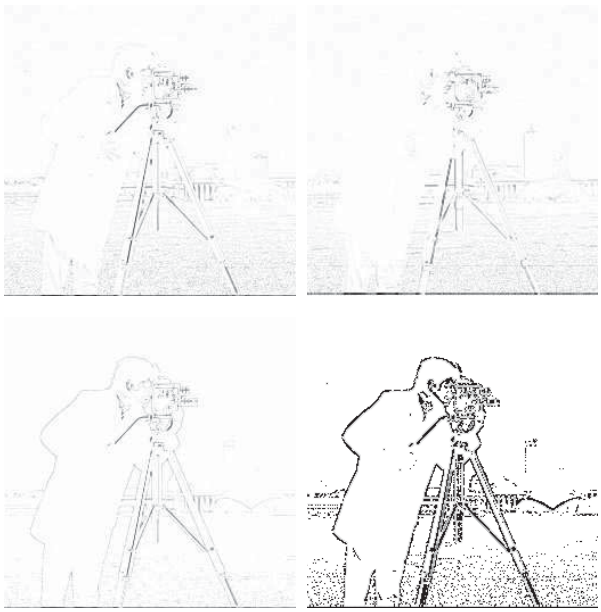


Figure 3: Left: Top-Hat. Right: Dual Top-Hat. Top: *umbra approach*. Down: Łuckasiewicz t-norm.

where e is the neutral element of the considered uninorm. The original images used in the experiments are shown in Figures 1 and 6 (they are represented by A in the fuzzy morphological operators). Two gray level original images are shown in Figure 1; one of them with size 256×256 is the classical image of a cameraman, and the other one of size 512×512 is the medical image of a hip.

The residual of two morphological operations or transformations is their difference. The first residual that can be defined in mathematical morphology is the morphological gradient, which is used as an edge detector and as a first approximation to a morphological segmentation. The morphological gradient, known in morphology as the Beucher gradient, is the difference between a dilation and an erosion, a dilation and the original image or the difference between the original image and its erosion. When we use the dilation and the erosion of a fuzzy set A by the structural element B, where the uninorm U is considered as a conjunctor and $\mathcal{I}_U$ as the residual implicator, the gradients are given by:

$$\nabla^{-}_{U,\mathcal{I}_U}(A,B) = A \setminus E_{\mathcal{I}_U}(A,B), \quad (5)$$

$$\nabla^{+}_{U,\mathcal{I}_U}(A,B) = D_U(A,B) \setminus A, \quad (6)$$

$$\nabla_{U,\mathcal{I}_U}(A,B) = D_U(A,B) \setminus E_{\mathcal{I}_U}(A,B), \quad (7)$$

where expression (5) represents the gradient by erosion, (6) the gradient by dilation and (7) the symmetrical gradient. From the point of view of the fuzzy inclusion, $\nabla^{-}_{U,\mathcal{I}_U}(A,B) \subseteq \nabla_{U,\mathcal{I}_U}(A,B)$ and $\nabla_{U,\mathcal{I}_U}(A,B) \supseteq \nabla^{+}_{U,\mathcal{I}_U}(A,B)$ are satisfied.

The goal of the gradients is the detection of the edges and/or the perimeter of the objects of the image. The correct choice of the structural element and the gradient will depend on the geometry of the objects. The application of the symmetrical gradient based on uninorms to the edge detection can be seen in [1], where the obtained results are better than those obtained by the gradients using t-norms or the classic approach.

The opening and closing defined in the previous section are the most elementary morphological filters, which are called *basic filters*. As we have seen in the previous section, the opening is an antiextensive morphological filter and the closing is an extensive morphological filter. So, in a first step, they can be used to remove non desired objects in an image. We can observe that the opening of a gray level image by a structural element removes the light zones of less size than the element and makes the light objects darker. The morphological closing helps to remove dark structures of less size than the structural element, toning down the dark objects. The size and shape of the used structurant element in the opening must agree with the image structures that we want to remove. Sometimes, structurant element of great size will remove the non desired shapes of an image, but also they can remove others shapes and the rest of the structures can be affected. Reduced sizes will be optimal when the images have small details. The effect of the closing and the opening on the cameraman's image can be observed in Figure 2, using the idempotent uninorm (top) and the representable uninorm (down).

The opening and the closing are morphological transformations and they are sensitive to compute the associated residuals, which are called *Top-Hat* transformations. The Top-Hat transformation, initiated by F. Meyer in 1977 (see [2]), finds structures of the images that have been removed by the opening or closing filter. If we choose the appropriated size and shape of the structural element, it is possible to filter and to remove some elements in the original image. The difference operation between the original map and the filtered map increases considerably the contrast of the removed zones.

The Top-Hat transformations are defined as a residual between the identity and the opening, which is called opening Top-Hat or white Top-Hat, or between the closing and the identity, which is called dual Top-Hat or black Top-Hat. They

ISBN: 978-989-95079-6-8

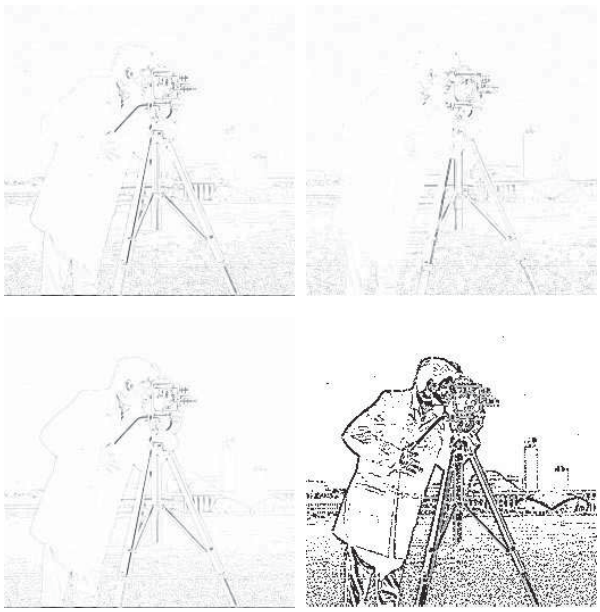


Figure 4: Left: Top-Hat. Right: Dual Top-Hat. Top, idempotent uninorm. Down, representable uninorm.

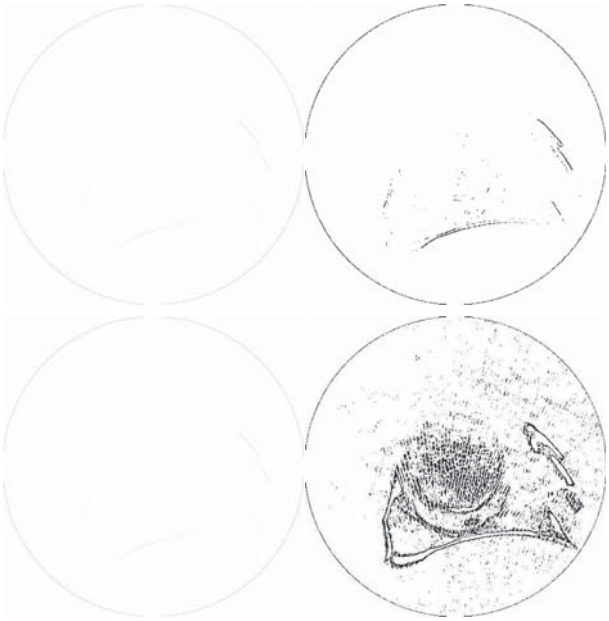


Figure 5: Left: Top-Hat. Right: Dual Top-Hat. Top, Łuckasiewicz t-norm. Down, representable uninorm.

are defined by:

$$\rho_{U,\mathcal{I}_U}(A,B) = A \setminus O_{U,\mathcal{I}_U}(A,B) \text{ (Top-Hat)}$$
$$\rho^d_{U,\mathcal{I}_U}(A,B) = C_{U,\mathcal{I}_U}(A,B) \setminus A \text{ (dual Top-Hat)}$$

The Top-Hat enhances the light objects that have been removed by the opening and it is used to extract contrasted components from the background, while the dual Top-Hat extracts the dark components that have been removed by the closing. Usually, the Top-Hat removes the soft trends, making thus an enhanced contrast. The detection of light objects is improved with an opening Top-Hat, in the same way with the dual Top-Hat does with the dark objects. The size of the structurant element is a deciding factor in order to apply the Top-Hat transformations.

The Top-Hat is idempotent (but it is not increasing) and it is antiextensive. The dual Top-Hat is neither increasing nor idempotent.

The obtained edges using the Top-Hat and the dual Top-Hat can be seen in Figures 3 and 4 for the cameraman image. The collected information using the idempotent uninorm is similar to the classical approach, and in the case of the representable uninorm, we can obtain (with the dual approach) more information than in the rest of the cases. Both top-hats are shown in Figure 5 for the hip image and it can be observed that the representable uninorm shows again more information.

It has been indicated that opening and closing are the basic filters of the fuzzy mathematical morphology. In order to build new mophological filters ([2]), we start with the basic filters and using composition or combination with other operations. New filters as sequential filters, center operators, contrast filters, etc. can be implemented. The basic applications of the morphological filters are the noise reduction and the selected extraction of image objects. Both applications are important in artificial vision since interpretation or measure techniques will fail in presence of noise, whereas the structure and the object selection are fundamental in segmentation processes.

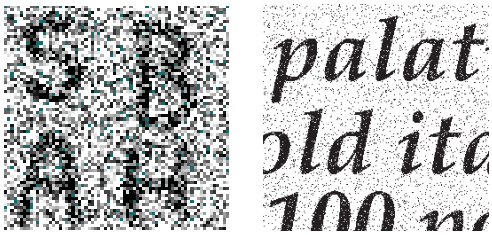


Figure 6: Original images with noise.

The first built filters from opening and closing are the so called *alternate filters*. Let $\xi(A,B)$ and $\psi(A,B)$ be the opening and closing, respectively, of a fuzzy object A by the structural element B using uninorm U as a conjuctor and $\mathcal{I}_U$ as a residual implicator. Using these two filters, four idempotent and growing filters can be generated: $\xi\,\psi$, $\psi\,\xi$, $\psi\,\xi\,\psi$ and $\xi\,\psi\,\xi$. These filters have similar properties to the classical morphological filters. Due to the idempotent property, the composition of more of three operators doesn't provide a new filter, except that we change the structural element.

In Figure 6, two binary images with added noise are shown. In the left image of size 70×74 the 60% of the points have been substituted by random graylevel noise[1]. In the right image of size 256×256, a gaussian noise has been added. Figure 7 shows the results obtained when we apply, from left to right, the alternate filters ψ, $\xi\,\psi$ and $\psi\,\xi\,\psi$, respectively. From top to bottom, the obtained results are shown using the different approaches. We can see that the uninorms remove the noise of the image in such a way that the *original* image is discernible. The same structure is shown in Figure 8.

The filtered image using the filter $\psi\,\xi$ and using the idempotent uninorm (very similar to the obtained image using the representable uninorm), Łuckasiewicz t-norm and *umbra approach* are shown in Figure 9. The filtered image with the

[1]This image has been downloaded of the web page of J. A. Sethian, http://math.berkeley.edu/~sethian/.

ISBN: 978-989-95079-6-8

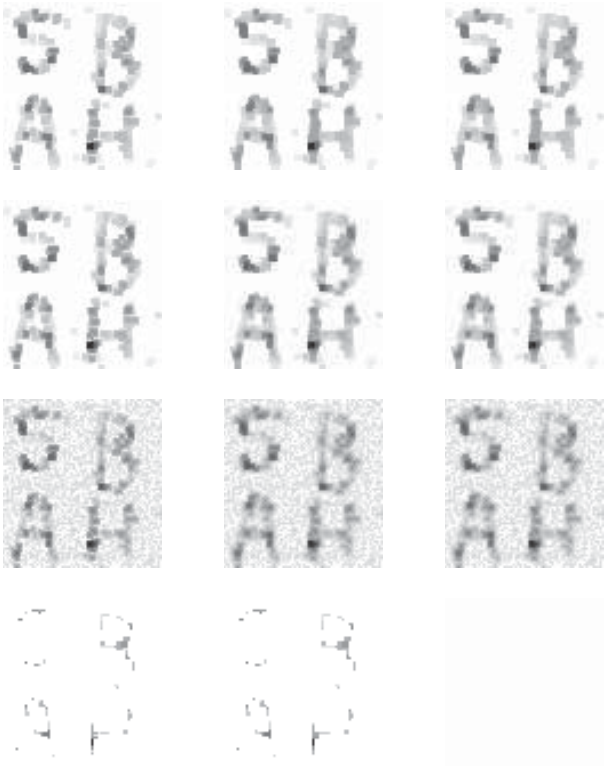


Figure 7: From left to right: filtered image using ξ, $\xi\,\psi$ and $\psi\,\xi\,\psi$, respectively. Top to bottom: idempotent uninorm, representable uninorm, Łuckasiewicz t-norm and *umbra approach*.

Figure 8: From left to right: filtered image using ξ, $\xi\,\psi$ and $\psi\,\xi\,\psi$, respectively. From top to bottom: idempotent uninorm, representable uninorm, Łuckasiewicz t-norm and *umbra approach*.

structure $\psi\,\xi\,\psi$ is very similar, except in the classical approach where the image gets smudged and in the case of the t-norm, the noise is softer.

A chest CT image of size 425×334 with "salt and pepper" added noise is shown in Figure 10 using 0.02 as a parameter. The results obtained using filter $\psi\,\xi\,\psi$ are shown in Figure 11. In the top left image, the Łuckasiewicz t-norm has been used and most of the noise remains. In the right image, the filtered image using the *umbra approach* is shown, a lot of "salt" noise is magnified. In the bottom images, the results obtained using uninorms are shown. The results using an idempotent uninorm are shown in the left image and the results using a representable uninorm are shown in the right image. The function generator of the representable uninorm is $h(x) = \frac{x-0.5}{x(1-x)}$. In the cases using uninorms the results are better than the ones using t-norms and umbra approach, because the noise is practically removed.

These type of filters have been used in the industrial control of pieces with partial occlusions or even in edge detection with noise elimination.

4 Conclusions and future work

In this work we have shown how the alternate filters based on opening and closing morphological operators using uninorms can remove the noise of an image preserving their structure. The obtained results improve the obtained results using t-norms and *umbra approach*. Also, we have obtained images very similar to the original ones without noise. So, the alternate filters based on uninorms are suitable to remove different

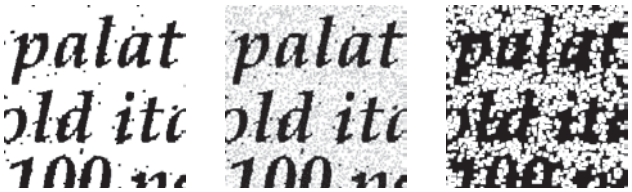


Figure 9: From left to right: filtered image using $\psi\,\xi$, an idempotent uninorm, Łuckasiewicz t-norm and *umbra approach*.

types of noise. The future work will consist on the construction of new filters. Also, we have shown the usefullness of the opening and closing filters in the edge detection. The future work includes the study of the effect of the choice of the structural element too. We have used isotropic structural elements in the experiments. That is, they have the same effect in all directions where they can be applied. We think that in some images, the use of anisotropic structural elements would improve the results.

We also want to apply the previous techniques to a more extended set of gray level images, even color images.

Acknowledgment

This work has been supported by the Government Spanish Grant MTM2006-05540, with FEDER support.

References

[1] M. González-Hidalgo, A. Mir-Torres, D. Ruiz-Aguilera, and J. Torrens. Edge-images using a uninorm-based fuzzy mathematical morphology: Opening and closing. In João Tavares

ISBN: 978-989-95079-6-8

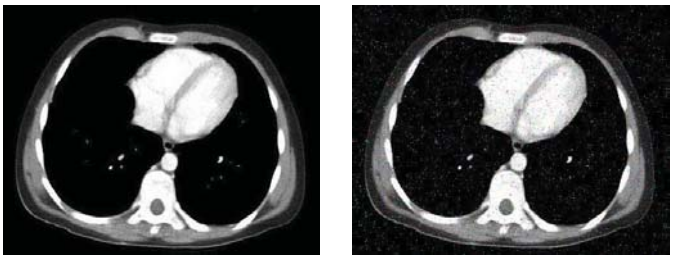


Figure 10: Left: Original image. Right: image with "salt and pepper" added noise.

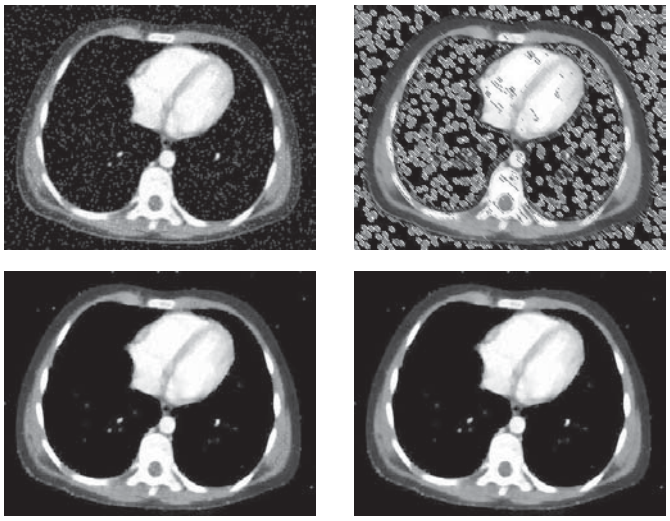


Figure 11: From left to right and from top to bottom, filtered image using $\psi\,\xi\,\psi$, and Łuckasiewicz t-norm, *umbra approach*, idempotent uninorm and representable uninorm.

and Natal Jorge, editors, *Advances in Computational Vision and Medical Image Processing*, number 13 in Computational Methods in Applied Sciences, chapter 8, pages 137–157. Springer, Netherlands, 2009.

[2] J. Serra. *Image analysis and mathematical morphology, vols. 1, 2*. Academic Press, London, 1982,1988.

[3] B. De Baets. Fuzzy morphology: A logical approach. In B. M. Ayyub and M. M. Gupta, editors, *Uncertainty Analysis in Engineering and Science: Fuzzy Logic, Statistics, and Neural Network Approach*, pages 53–68. Kluwer Academic Publishers, Norwell, 1997.

[4] B. De Baets, E.E. Kerre, and M. Gupta. The fundamentals of fuzzy mathematical morfologies part i: basics concepts. *International Journal of General Systems*, 23:155–171, 1995.

[5] B. De Baets, E.E. Kerre, and M. Gupta. The fundamentals of fuzzy mathematical morfologies part ii: idempotence, convexity and decomposition. *International Journal of General Systems*, 23:307–322, 1995.

[6] I. Bloch and H. Maître. Fuzzy mathematical morphologies: a comparative study. *Pattern Recognition*, 28:1341–1387, 1995.

[7] M. Nachtegael and E.E. Kerre. Classical and fuzzy approaches towards mathematical morphology. In Etienne E. Kerre and Mike Nachtegael, editors, *Fuzzy techniques in image processing*, number 52 in Studies in Fuzziness and Soft Computing, chapter 1, pages 3–57. Physica-Verlag, New York, 2000.

[8] B. De Baets, N. Kwasnikowska, and E. Kerre. Fuzzy morphology based on uninorms. In *Proceedings of the seventh IFSA World Congress*, pages 215–220, Prague, 1997.

[9] M. González, D. Ruiz-Aguilera, and J. Torrens. Algebraic properties of fuzzy morphological operators based on uninorms. In *Artificial Intelligence Research and Development*, volume 100 of *Frontiers in Artificial Intelligence and Applications*, pages 27–38, Amsterdam, 2003. IOS Press.

[10] S. Jamil Ferzoli Guimarães, N. Jerônimo Leite, M. Couprie, and A. de Albuquerque Araújo. A directional and parametrized transition detection algorithm based on morphological residues. In *Proceedings of the 15th Brazilian Symposium on Computer Graphics and Image Processing, SIBGRAPI'02*, pages 261–268, Washington, DC, USA, 2002. IEEE Computer Society.

[11] T. Chen, Q. H. Wu, R. Rahmani-Torkaman, and J. Hughes. A pseudo top-hat mathematical morphological approach to edge detection in dark regions. *Pattern Recognition*, 35(1):199–210, 2002.

[12] Zhao Yu-qian, Gui Wei-hua, Chen Zhen-cheng, Tang Jing-tian, and Li Ling-yun. Medical images edge detection based on mathematical morphology. In *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 6492–6495, Shanghai, China, 2005.

[13] J.-A. Jiang, C.-L. Chuang, Y.-L. Lu, and C.-S-Fahn. Mathematical-morphology-based edge detectors for detection of thin edges in low-contrast regions. *IET Image Processing*, 1(3):269–277, 2007.

[14] J. I. Pastore, E. Moler, and V. Ballarin. Multiscale morphological operators and geodesic distance applied to computer axial tomography segmentation. *IEEE Latin America Transactions*, 5(1):28–31, 2007.

[15] Hamid R. Tizhoosh. Fuzzy image enhancement: An overview. In Etienne E. Kerre and Mike Nachtegael, editors, *Fuzzy techniques in image processing*, number 52 in Studies in Fuzziness and Soft Computing, chapter 5, pages 137–171. Physica-Verlag, New York, 2000.

[16] Chang-Shing Lee and Yau-Hwang Kuo. Adaptative fuzzy filter and its application to image enhancement. In Etienne E. Kerre and Mike Nachtegael, editors, *Fuzzy techniques in image processing*, number 52 in Studies in Fuzziness and Soft Computing, chapter 6, pages 172–193. Physica-Verlag, New York, 2000.

[17] F. Farbiz and M. B. Menhaj. A fuzzy logic control based approach for image filtering. In Etienne E. Kerre and Mike Nachtegael, editors, *Fuzzy techniques in image processing*, number 52 in Studies in Fuzziness and Soft Computing, chapter 7, pages 194–221. Physica-Verlag, New York, 2000.

[18] M. Nachtegael, D. Van der Weken, D. Van De Ville, and E.E. Kerre, editors. *Fuzzy filters for image processing*, volume 122 of *Studies in Fuzziness and Soft Computing*. Springer-Verlag, 2003.

[19] D. Van De Ville, M. Nachtegael, D. Van der Weken, E.E. Kerre, and W. Philips. Noise reduction by fuzzy images filtering. *IEEE Transactions on Fuzzy Systems*, 11(4):429–436, 2003.

[20] J.C. Fodor, R.R. Yager, and A. Rybalov. Structure of uninorms. *Int. J. Uncertainty, Fuzziness, Knowledge-Based Systems*, 5:411–427, 1997.

[21] B. De Baets and J.C. Fodor. Residual operators of uninorms. *Soft Computing*, 3:89–100, 1999.

[22] D. Ruiz-Aguilera and J. Torrens. Residual implications and coimplications from idempotent uninorms. *Kybernetika*, 40:21–38, 2004.

ISBN: 978-989-95079-6-8

A value via posets induced by graph-restricted communication situations

Katsushige Fujimoto[1] Aoi Honda[2]
1.College of Symbiotic Systems Science, Fukushima University
1 Kanayagawa Fukushima 960-1296, Japan
2.Department of Systems Design and Informatics, Kyushu Institute of Technology
680-4 Kawazu Iizuka Fukuoka 820-8502, Japan
Email: fujimoto@sss.fukushima-u.ac.jp, aoi@ces.kyutech.ac.jp

Abstract— *This paper provides a new value (solution concept or allocation rule) of cooperative games via posets induced by graphs. Several values in a graph-restricted communication situation have been proposed or introduced by Myerson, Borm, and Hamiache... However, these values have been subjected to some criticisms in certain types of games. The value proposed in this paper withstands these criticisms. Moreover, these existing values have been defined only in situations represented by undirected graphs, while the notion of the value proposed in this paper can be extended to situations represented by directed graphs.*

Keywords— graph-restricted situations, communication situations, values, posets, cooperative games.

1 Introduction and Preliminaries

Throughout the paper, N denotes the universal set of n elements. For convenience, we often number the elements such that the universal set is $N = \{1, 2, \ldots, n\}$. A real-valued function $v : 2^N \to \mathbb{R}$ with $v(\emptyset) = 0$ is called a *game*. A monotone game (i.e., $v(A) \leq v(B)$ whenever $A \subseteq B \subseteq N$) is called a *capacity* or *a fuzzy measure*. We often call the pair (N, v), rather than v, a game or a capacity. The set of all games on N is denoted by $\mathcal{G}^N$. A real vector-valued function $\Phi : \mathcal{G}^N \to \mathbb{R}^{|N|}$ is called a *value*. In cooperative game theory, N is considered to be the set of all players. For every subset S of N, often called a *coalition*, $v(S)$ represents the (transferable) utility/profits that players in S can obtain if they decide to cooperate. For every game (N, v), the value $\Phi(N, v)$ represents an allocation rule, which provides an assessment of the benefits for each player from participating in a game v. For the sake of simplicity, we mainly discuss games in terms of various set functions (e.g., games, capacities, fuzzy measures, and so forth.) on N.

To avoid cumbersome notations, we often omit braces for singletons, e.g., by writing $v(i)$, $U \setminus i$ instead of $v(\{i\})$, $U \setminus \{i\}$. Similarly, for pairs, we write ij instead of $\{i, j\}$. Furthermore, cardinalities of subsets $S, T, \ldots$, are often denoted by the corresponding lower case letters $s, t, \ldots$, otherwise by the standard notation $|S|, |T|$,...

1.1 Games and capacities with graph restricted situations

In ordinary cooperative game theory it is implicitly assumed that all coalitions of N can be formed; however, this is generally not the case. For players to coordinate their actions, they must be able to communicate. The bilateral communication channels between players in N are described by a *communication network*. Such a network can be represented by an *undirected graph* (N, L), which has the set of players as its *nodes* $S \subseteq N$ and in which the players are connected by the set of *links* $L \subseteq \{ij \mid i, j \in N, i \neq j\}$; i.e., players i and j can communicate (directly) with each other if $ij \in L$. This paper will deals with only situations induced by communication networks described by undirected graphs. Many other approaches to the situations can be seen via the literatures [1, 2].

Definition 1.1 (communication situation)
The triple (N, v, L), which reflects a situation consisting of a game v on N and a communication network (N, L), is called a *communication situation*. We denote the set consisting of all communication situations on N by CS^N. For a coalition $T \subseteq N$, the restriction of (N, L) to T is denoted by $(T, L(T))$ and defined by $L(T) := \{ij \in L \mid ij \subseteq T\}$.

Definition 1.2 (feasible coalition) We say that players j and k are *connected in* $S \subseteq N$ if $j = k$ or there exists a subset $\{i_1, \cdots, i_m\} \subseteq S$ such that $j = i_1$, $k = i_m$, and $\{i_t, i_{t+1}\} \in L$ for all $t \in \{1, \cdots, m-1\}$. Then we denote $j \sim_S k$. Clearly, this relation $\sim_S$ is an equivalence relation. Hence, the notion of connectedness in S induces a partition $S/L := S/\sim_S$ of S. A coalition $S \subseteq N$ is said to be *feasible* in the communication network (N, L) if any two players, $j \in S$ and $k \in S$, are connected in S (i.e., $S/L = \{S\}$).

Example 1.1
Consider the communication situation (N_1, v, L_1) with $N_1 = \{1, 2, 3, 4, 5, 6, 7\}$ and $L_1 = \{12, 15, 26, 37, 47, 56\}$ (Fig.1). Then, all the players in $\{1, 2, 6\}$ can communicate with other;

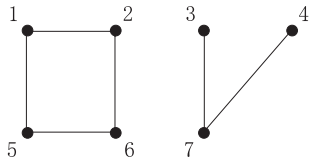


Figure 1: Communication network(N, L_1).

i.e., the coalition $\{1, 2, 6\}$ is feasible. Hence, they can fully coordinate their actions and obtain the value $v(\{1, 2, 6\})$. On the other hand, in the coalition $\{1, 2, 3, 4\}$, players 1 and 2 can communicate with each other, but players 3 and 4 cannot communicate with any other players in $\{1, 2, 3, 4\}$. Thus, feasible subcoalitions of $\{1, 2, 3, 4\}$ are $\{1, 2\}$, $\{3\}$, and $\{4\}$ (i.e., forming the coalition $\{1, 2, 3, 4\}$ is unfeasible). Hence, the value attainable by the players in $\{1, 2, 3, 4\}$ should be $v(\{1, 2\}) + v(\{3\}) + v(\{4\})$. In general, the value attainable by the players in $S \in N$ under a communication situation (N, v, L) is represented by

$$\sum_{T \in S/L} v(T). \tag{1}$$

ISBN: 978-989-95079-6-8

Definition 1.3 (network-restricted game [3]) The *network-restricted game* (N, v^L) associated with (N, v, L) is defined as

$$v^L(S) := \sum_{T \in S/L} v(T) \quad \text{for each } S \subseteq N. \tag{2}$$

Note that if (N, L) is the complete graph (i.e., $L = \{ij \mid i, j \in N, i \neq j\}$), the network-restricted game v^L is equal to the original game v.

The network-restricted game evaluates the possible gains from cooperation in a communication situation from the viewpoint of the players. The next example focuses on the importance of communication channels and links in a communication situation.

Example 1.2 In the communication situation L_1 depicted in Fig.1, the value obtainable by the players in the grand coalition N is

$$v^{L_1}(N) = v(\{1, 2, 5, 6\}) + v(\{3, 4, 7\}), \tag{3}$$

since $N/L_1 = \{\{1, 2, 5, 6\}, \{3, 4, 7\}\}$. If for some reason the communication link between players 4 and 7 is lost, the communication network L_1 becomes the new communication network $L_2 = \{12, 15, 26, 37, 56\}$. Then, $N/L_2 = \{\{1, 2, 5, 6\}, \{4\}, \{3, 7\}\}$ and the value obtainable by the players in the grand coalition N becomes

$$v^{L_2}(N) = v(\{1, 2, 5, 6\}) + v(\{4\}) + v(\{3, 7\}). \tag{4}$$

Then,

$$v^{L_1}(N) - v^{L_2}(N) \tag{5}$$

can be interpreted as a type of marginal contribution of the link $\{4, 7\} \in L_1$ to the communication network L_1.

Definition 1.4 (link game [4]) The *link game* (L, γ^v) associated with (N, v, L) consisting of a zero-normalized game v is a game on L defined by

$$\gamma^v(M) := v^M(N) = \sum_{T \in N/M} v(T) \quad \text{for each } M \subseteq L. \tag{6}$$

Note that, for an ordinary game v, γ^v is not a game on L since $\gamma^v(\emptyset) = \sum_{T \in N/\emptyset} v(T) = \sum_{i \in N} v(\{i\}) \neq 0$.

The link game $\gamma^v(M)$ represents the worth of the communication network $M \subseteq L$ as the worth of the grand coalition in the communication situation (N, v, M) through the network-restricted game v^M.

Definition 1.5 (Möbius Transform [5]**)** The *Möbius transform* of a game $v : 2^N \to \mathbb{R}$ (resp. $\gamma : 2^L \to \mathbb{R}$) is a game on N (resp. L) denoted by $\Delta^v : 2^N \to \mathbb{R}$ (resp. $\Delta^\gamma : 2^L \to \mathbb{R}$) and is defined by

$$\Delta^v(S) := \sum_{T \subseteq S} (-1)^{|S \setminus T|} v(T) \quad \text{for each } S \in 2^N. \tag{7}$$

$$(resp.\ \Delta^\gamma(M) := \sum_{K \subseteq M} (-1)^{|M \setminus K|} \gamma(K) \quad \text{for each } M \in 2^L). \tag{8}$$

Equivalently, we have that

$$v(S) = \sum_{T \subseteq S} \Delta^v(T) \quad \forall S \in 2^N. \tag{9}$$

$$(resp.\ \gamma(M) = \sum_{K \subseteq M} \Delta^\gamma(K) \quad \forall M \in 2^L). \tag{10}$$

Thus, the worth $v(S)$ (resp. $\gamma(M)$) of a coalition S (resp. communication network M) is equal to the sum of the Möbius transform of all its subcoalitions (subnetworks). This gives a recursive definition of the Möbius transform. The Möbius transform of every singleton is equal to its worth, while recursively, the Möbius transform of every coalition (resp. communication network) of at least two players (resp. links) is equal to its worth minus the sum of the Möbius transform of all its proper subcoalitions (resp. subnetworks). In this sense, the Möbius transform of a coalition S (resp. communication network M) can be interpreted as the extra contribution of the cooperation/synergy among the players in S (resp. links in M) that they did not already achieve by smaller coalitions (resp. networks). In fact, in the context of *interaction indices* (e.g.,[6, 7]), the Möbius transform $\Delta^v(S)$ is called the *internal interaction index* of S, which represents the magnitude of a type of interaction among the elements in S. The Möbius transform is also occasionally called the *Harsanyi dividends*[8].

Definition 1.6 (unanimity game) The *unanimity game* for a non-empty coalition $T \subseteq N$ is denoted by u_T and defined by

$$u_T(S) = \begin{cases} 1 & \text{if } S \supseteq T, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

For any game $v : 2^N \to \mathbb{R}$, v can be represented as

$$v(S) = \sum_{T(\neq \emptyset) \in 2^N} \Delta^v(T) \cdot u_T(S) \quad \forall S(\neq \emptyset) \in 2^N. \tag{12}$$

2 Values for communication situations

In this section, we briefly introduce the Shapley value for ordinary cooperative games and three existing values for communication situations that appear in the literatures [3, 4, 9], *the Myerson value, the position value, and the Hamiache value*.

Definition 2.1 (the Shapley value [10]) The *Shapley value* $\Phi : \mathcal{G}^N \to \mathbb{R}^{|N|}$ for a game $(N, v) \in \mathcal{G}^N$ is defined by

$$\Phi_i(N, v) := \sum_{T \ni i} \frac{1}{|T|} \Delta^v(T) \quad \text{for each } i \in N. \tag{13}$$

Definition 2.2 (the Myerson value [3]) The *Myerson value* $\Psi : CS^N \to \mathbb{R}^{|N|}$ for a communication situation $(N, v, L) \in CS^N$ is defined by

$$\Psi(N, v, L) := \Phi(N, v^L). \tag{14}$$

The Myerson value is the allocation rule that assigns to every communication situation (N, v, L) the Shapley value of the network-restricted game (N, v^L). Note that $\Psi(N, v, L) = \Phi(N, v)$ if (N, L) is the complete graph.

Definition 2.3 (position value [4]) The *position value* $\pi : CS^N \to \mathbb{R}^{|N|}$ for a communication situation $(N, v, L) \in CS^N$ is defined by

$$\pi_i(N, v, L) := \frac{1}{2} \sum_{\substack{l \in L \\ l \ni i}} \Phi_l(L, \gamma^v) \quad \text{for each } i \in N. \tag{15}$$

ISBN: 978-989-95079-6-8

The Shapley value $\Phi_l(L,\gamma^v)$ of a link $l \in L$, which is induced via (13) for the link game (L,γ^v), can be interpreted as a type of *expected marginal contribution* of the link l to all communication networks containing l. Then, the value is divided equally between the two players at the ends of the considered link $l \in L$. The position value of a given player $i \in N$ is obtained as the sum of all these shares.

We focus to a third value for communication situations, introduced by Hamiache [9]. Given a communication situation (N,v,L) and $S \subseteq N$, we denote by S^* the set of all nodes of the communication network (N,L) that are adjacent to at least one of the nodes of S,

$$S^* := \{i \in N \mid \exists j \in S \text{ such that } ij \in L\}. \tag{16}$$

Definition 2.4 (associated game [9]) For a value ϕ on CS^N (i.e., $\phi : CS^N \to \mathbb{R}^{|N|}$), the *associated game* v^*_ϕ of v with respect to ϕ is defined for $S \subseteq N$, by

$$v^*_\phi(S) := \begin{cases} v(S) + \displaystyle\sum_{j \in S^* \setminus S} \Big(\phi_i(S^{+j}, v|_{S^{+j}}, L(S^{+j})) - v(j)\Big) & \text{if } |S/L| = 1, \\ \displaystyle\sum_{T \in S/L} v^*_\phi(T) & \text{otherwise,} \end{cases} \tag{17}$$

where $S^{+j} := S \cup \{j\}$ and $v|_{S^{+j}}$ is the restriction of v to S^{+j}.

Hamiache [9] claims that there is a unique value ϕ, the so-called *Hamiache value*, for communication situations satisfying the following five properties, *component-efficiency*, *linearity w.r.t. games*, *independence of irrelevant players*, *positivity*, and *associated consistency*:

Component-efficiency :
For any (N,v,L) and any $S \in N/L$,

$$\sum_{i \in S} \phi_i(N,v,L) = v(S). \tag{18}$$

Linearity w.r.t. games :
For any $\alpha,\beta \in \mathbb{R}$ and $(N,v,L),(N,w,L) \in CS^N$,

$$\phi(N,\alpha v + \beta w, L) = \alpha\phi(N,v,L) + \beta\phi(N,w,L). \tag{19}$$

Independence of irrelevant players :
For any (N,L) and for any two feasible coalitions $R \subseteq T$,

$$\phi_i(N,u_R,L) = \phi_i(T,u_R,L(T)) \ \ \forall i \in T. \tag{20}$$

Positivity :
For any feasible coalition $T \subseteq N$,

$$\phi_i(T,u_T,L(T)) \geq 0 \ \ \forall i \in T. \tag{21}$$

Associated consistency:
For any $(N,v,L) \in CS^N$,

$$\phi(N,v,L) = \phi(N,v^*_\phi,L). \tag{22}$$

Note that $\phi(N,v,L) = \Phi(N,v)$ if (N,L) is the complete graph.

3 Posets induced by communication networks

3.1 Communication networks and posets

In this subsection, we consider and introduce a subposet of $B(n) := (2^N, \subseteq)$ induced by a communication network (N,L).

For a communication network (N,L), the set of all feasible coalitions in (N,L) is denoted by $P(N,L)$. i.e.,

$$P(N,L) := \{S \subseteq N \mid |S/L| = 1\}. \tag{23}$$

The set $P(N,L)$, together with set inclusion $\subseteq$ as an order on $P(N,L)$, is called the *poset induced by the communication network* (N,L).

Example 3.1 Let $N = \{1,2,3\}$, $L_a = \{12,13,23\}$, $L_b = \{13,23\}$, and $L_c = \{12\}$. Then the posets induced by communication networks (N,L_a), (N,L_b), and (N,L_c), as shown in (a) – (c) in Fig. 2, are represented as shown in (a) – (c) in Fig. 3, respectively.

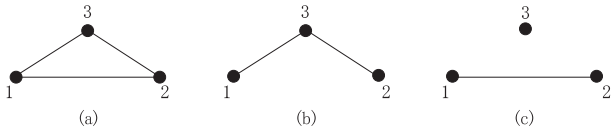


Figure 2: Communication networks on $N = \{1,2,3\}$.

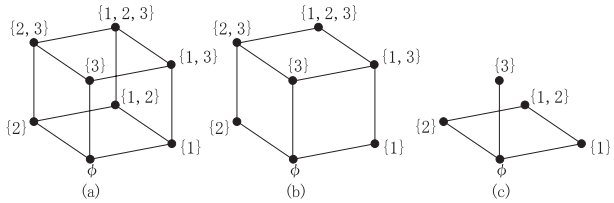


Figure 3: Posets corresponding to networks in Fig. 2.

Definition 3.1 (Möbius transform on posets)
Let $P := (N,\leq)$ be a poset. For a function $v : P \to \mathbb{R}$, the *Möbius transform* Δ^v of v is a function on P satisfying the following equation:

$$v(x) = \sum_{y \leq x} \Delta^v(y) \quad \forall x \in P. \tag{24}$$

Definition 3.2 (representation functions)
The *representation function of a communication situation* (N,v,L) is a function v^P on the poset $P(N,L)$ defined by

$$v^P(S) = v(S) \quad \text{for each } S \in P(N,L). \tag{25}$$

Then, the Möbius transform Δ^{v^P} of v^P is represented as

$$\Delta^{v^P}(S) := \sum_{\substack{T \in P(N,L) \\ T \subseteq S}} (-1)^{|S \setminus T|} v^P(T) \quad \forall S \in P(N,L). \tag{26}$$

Conversely,

$$v^P(S) := \sum_{\substack{T \in P(N,L) \\ T \subseteq S}} \Delta^{v^P}(T) \quad \forall S \in P(N,L). \tag{27}$$

Definition 3.3 (poset representation) The *poset representation of a communication situation* (N,v,L) is the pair $(P(N,L),\Delta^{v^P})$ of the poset induced by (N,v,L) and the Möbius transform Δ^{v^P} of representation function v^P of (N,v,L).

ISBN: 978-989-95079-6-8

4 A new value in communication situations

In this section, we introduce a new value for communication situations.

4.1 An interpretation of the Shapley value

Now, we consider the case $N = \{1,2\}$; the Shapley value $\Phi_1(N,v)$ of player 1 in a game v is obtained, from (13), as

$$\Phi_1(N,v) = \frac{1}{1}\Delta^v(\{1\}) + \frac{1}{2}\Delta^v(\{1,2\}). \tag{28}$$

This can be interpreted as an allocation rule of *Harsanyi dividends* (i.e., the Möbius transform) described as follows:

Allocation rule of Harsanyi dividends : We consider a process to form the coalition $\{1,2\}$. Then, there are two shortest paths from $\emptyset$ to $\{1,2\}$ in Fig. 2. One is the path $\emptyset \to \{1\} \to \{1,2\}$; another is the path $\emptyset \to \{2\} \to \{1,2\}$. The path $\emptyset \to \{1\} \to \{1,2\}$ can be interpreted as follows: *Player 1 makes an offer to player 2 for forming the coalition* $\{1,2\}$. *Player 2 accepts the offer and adds to the coalition* $\{1\}$ *to form the new coalition* $\{1,2\}$. Among these two paths, the only path that passes through $\{1\}$ is $\emptyset \to \{1\} \to \{1,2\}$. That is, the number of paths from $\emptyset$ to $\{1,2\}$ is 2, while of the number of paths via $\{1\}$ is 1. Then player 1 obtains $\frac{1 \text{ path}}{2 \text{ paths}}$ of the amount of the Harsanyi dividend $\Delta^v(\{1,2\})$ (i.e., $\frac{1}{2}\Delta^v(\{1,2\})$). In the same way, player 1 obtains $\frac{1}{1}\Delta^v(\{1\})$ and $\frac{0}{1}\Delta^v(\{2\})$. The Shapley value of player 1 is obtained as the sum of all these shares.

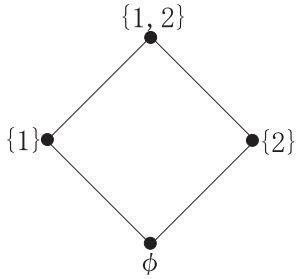


Figure 4: The Boolean lattice $B(2)$ on $N = \{1,2\}$.

This allocation rule can be extended to the case $N = \{1,2,3\}$ (Fig. 5).

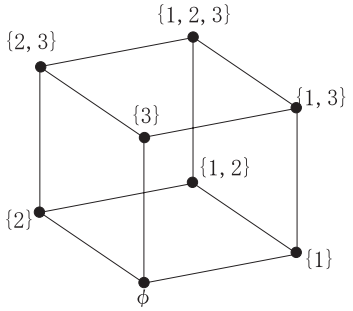


Figure 5: The Boolean lattice on $N = \{1,2,3\}$.

Indeed,

$$\begin{aligned}\Phi_1(N,v) &= \frac{1}{1}\Delta^v(\{1\}) + \frac{1}{2}\Delta^v(\{1,2\}) + \frac{1}{2}\Delta^v(\{1,3\}) \\ &+ \frac{0}{2}\Delta^v(\{2,3\}) + \frac{2}{6}\Delta^v(\{1,2,3\}).\end{aligned} \tag{29}$$

For instance, there are six shortest paths from $\emptyset$ to $\{1,2,3\}$. Among them, two paths pass through $\{1\}$, as shown in Fig. 6.

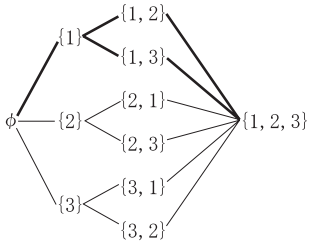


Figure 6: Shortest paths from $\emptyset$ to $\{1,2,3\}$.

4.2 An interpretation of the Myerson value

The Myerson value of (N,v,L) is the Shapley value of the network-restricted game (v^L,N). That is, the Myerson value is obtained by applying the above allocation rule to Harsanyi dividends $\{\Delta^{v^L}\}$ of v^L. Then Δ^{v^L} is given as follows:

Proposition 4.1 Let $(N,v,L) \in CS^N$ be a communication situation and $(B(n),\Delta^{v^L})$ the poset representation of the network-restricted game (N,v^L) associated with (N,v,L). Then,

$$\Delta^{v^L}(S) = \begin{cases} \Delta^v(S) & \text{if } S \in P(N,L), \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

4.3 Criticisms of existing values

Each of the existing values for communication situations, the Myerson value, the position value, and the Hamiache value, has been subject to criticisms, as follows.

The Myerson value :

$$\Psi_i(N,u_S,L) = \Psi_i(N,u_S,M) = \frac{1}{|S|} \quad \forall i \in N \tag{31}$$

whenever S is a feasible coalition in both (N,L) and (N,M). For example, consider the communication situation with $L = \{ij \subseteq N \mid j \in N \setminus i\}$ (i.e., L is a star with a central player i); then every player receives the same value (see Example $\Psi(N,v,L_e)$ in Example 5.3).

The position value :

Irrelevant null players often have positive values (see Example 5.2), where a null player $i \in N$ of the game (N,v) is a player satisfying $v(S \cup i) = v(S)$ for any $S \subseteq N$.

The Hamiache value :

It is very complex to compute the Hamiache value. Not only that, associated consistency is rather technical.

ISBN: 978-989-95079-6-8

4.4 A new value in communication situations

In this section, we propose a new value for communication situations that withstands all these criticisms.

Definition 4.1 (chain, saturated chain) A *chain* (or a *totally ordered set* or *linear ordered set*) is a poset in which any two elements are comparative. That is, a subset C of $P(N,L)$ is called a chain if $S \subseteq T$ or $T \subseteq S$ for any $S, T \in C$. The chain C of $P(N,L)$ is *saturated* (or *unrefinable*) if there does not exist $W \in P(N,L) \setminus C$ such that $S \subsetneq W \subsetneq T$ for some $S, T \in C$ and that $C \cup W$ is a chain.

Definition 4.2 (shortest path) For two feasible coalitions S, $T \in P(N,L)$, a saturated chain $\mathcal{P}$ of $P(N,L)$ is called a *shortest path* from S to T if $S, T \in \mathcal{P}$ and $S \subseteq W \subseteq T$ for any $W \in \mathcal{P}$. Then, we denote the set of all shortest paths from S to T by $\{S \to T\}$.

In the following, we propose a new value in communication situations, based on the interpretation of the Shapley value mentioned in Subsection 4.1.

Definition 4.3 We now propose a new value $\sigma(N,v,L)$ of a communication situation (N,v,L), as follows.

$$\sigma_i(N,v,L) := \sum_{S \in P(N,L)} \frac{|\{i \to S\}|}{|\{\emptyset \to S\}|} \Delta^{v^P}(S) \quad \text{for each } i \in N. \tag{32}$$

The number $|\{\emptyset \to S\}|$ of all shortest paths from $\emptyset$ to S indicates the number of all processes in which the feasible coalition S is formed. Also, $|\{i \to S\}|$ indicates the number of all processes in which the feasible coalition S is formed by the initiator $i \in N$. Then, the player $i \in N$ obtains $\frac{|\{i \to S\}|}{|\{\emptyset \to S\}|}$ of the amount of $\Delta^{v^P}(S)$ if $\Delta^{v^P}(S)$ is allocated in proportion to the frequency with which the player i initiates the formation of the feasible coalition S. The value proposed here of a given player $i \in N$ is obtained as the sum of all these shares.

Now we show an example that supports the naturalness of the definition of this value.

Example 4.1 We consider the communication situation (N,v,L) with $N = \{1,2,3\}$, $L = \{13,23\}$, and $\Delta^{v^P}(S) \geq 0$ for any $S \in P(N,L)$. The value $\sigma_i(N,v,L)$ proposed here of player $i \in N$ is represented as the values of the ammeters A_i in the electric circuit with current sources $I_S = \Delta^{v^P}(S)$, as shown in Fig.7.

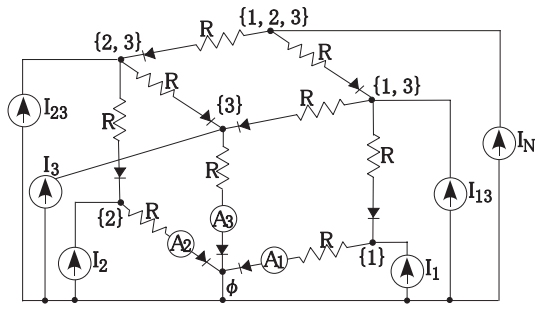


Figure 7: Electric circuit representing (N,v,L).

Property 1 The value σ proposed here satisfies *component-efficiency*, *linearity w.r.t. games*, *independence of irrelevant players*, and *positivity*.

Property 2 Let (N,u_N,L_c^*) be a communication situation with $L_c^* = \{cj \mid j \in N \setminus c\}$, $c \in N$. Then,

$$\sigma_i(N,u_N,L_c^*) = \begin{cases} \dfrac{1}{2} & \text{if } i = c, \\ \dfrac{1}{2(n-1)} & \text{otherwise.} \end{cases} \tag{33}$$

That is, if the communication network (N,L) is a star-graph with central player $c \in N$, in the unanimity game u_N, the central player obtains a half of the total amount of $u_N(N) = 1$ and the rest of the amount are shared out equally among the other players (see L_b, L_e in example 5.3).

However, we have not found any axiomatic characterization of the value proposed in this paper yet.

5 Comparison of existing values

In this section, we compare the existing four values (the Shapley, Myerson, position, and Hamiache values) and the value proposed in this paper. Examples 5.1 and 5.2 not only compare them but also illustrate the criticisms against the Shapley, Myerson, and position values, respectively.

Example 5.1 Consider the communication situation (N,v,L) with $N = \{1,2,3\}$, $L = \{13,23\}$ ((b) in Fig. 2), and

$$v(S) = \begin{cases} 0 & \text{if } |S| \leq 1 \\ 30 & \text{if } |S| = 2 \\ 36 & \text{if } S = N. \end{cases} \tag{34}$$

Then,

$$\Phi(N,v) = (12,12,12), \quad \Psi(N,v,L) = (7,7,22),$$
$$\pi(N,v,L) = (9,9,18), \quad \phi(N,v,L) = (9,9,18),$$
$$\sigma(N,v,L) = (9,9,18).$$

Example 5.2 Consider the communication situation (N,v,L) with $N = \{1,2,3\}$, $L = \{12,13,23\}$ (L_d in Fig. 8), and

$$v(S) = \begin{cases} 12 & \text{if } S \supseteq \{1,2\} \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

Then,

$$\Phi(N,v) = (6,6,0), \quad \Psi(N,v,L) = (6,6,0),$$
$$\pi(N,v,L) = (5,5,2), \quad \phi(N,v,L) = (6,6,0),$$
$$\sigma(N,v,L) = (6,6,0).$$

Example 5.3 Consider communication situations (N,u_N,L) with $2 \leq |N| \leq 4$, $|N/L| = 1$ (i.e., (N,L) is connected). Fig.8 shows all connected graphs (up to isomorphism) with $2 \leq n \leq 4$ nodes. Then, for any such communication situations (N,u_N,L),

$$\Phi_i(N,u_N,L) = \Psi_i(N,u_N,L) = \frac{1}{|N|} \quad \forall i \in N. \tag{36}$$

Table 1 shows comparisons of the remaining values (i.e., the position value π, the Hamiache value ϕ, and the value σ proposed in this paper), and illustrates that the value σ does not always coincide with the Hamiache value ϕ.

ISBN: 978-989-95079-6-8

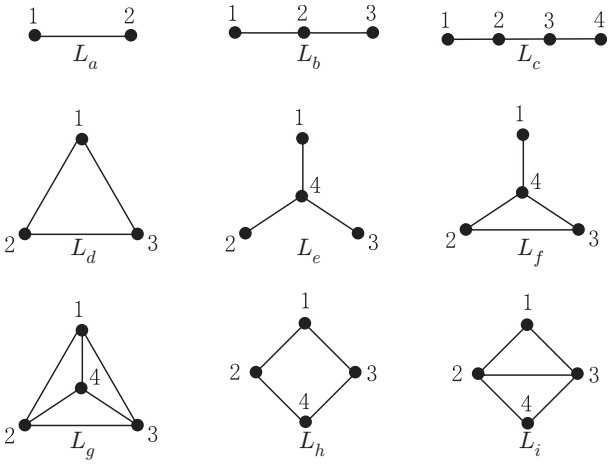


Figure 8: Graphs with at most four nodes.

Table 1: Comparison of existing values.

	π	ϕ	σ
L_a	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, \frac{1}{2})$
L_b	$(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$
L_c	$(\frac{1}{6}, \frac{2}{6}, \frac{2}{6}, \frac{1}{6})$	$(\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8})$	$(\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8})$
L_d	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
L_e	$(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{3}{6})$	$(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{3}{6})$	$(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{3}{6})$
L_f	$(\frac{3}{12}, \frac{2}{12}, \frac{2}{12}, \frac{5}{12})$	$(0.172, 0.190, 0.190, 0.448)$	$(\frac{2}{14}, \frac{3}{14}, \frac{3}{14}, \frac{6}{14})$
L_g	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$
L_h	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$
L_i	$(\frac{13}{60}, \frac{17}{60}, \frac{17}{60}, \frac{13}{60})$	$(\frac{3}{14}, \frac{4}{14}, \frac{4}{14}, \frac{3}{14})$	$(\frac{2}{10}, \frac{3}{10}, \frac{3}{10}, \frac{2}{10})$

Acknowledgment

This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology-Japan, Grant-in-Aid for Scientific Research (C), 19510136, 2007-2009.

References

[1] M. Grabisch and F. Lange. Games on latices, multichoice games and the shapley value: a new approach. *mathematical Methods of Operations Research*, 65(1):153–167, 2007.

[2] J. M. .Bilbao. *Cooperative games on combinatorial syructures*. Kluwer academic publishers, Boston, 2000.

[3] R. Myerson. Graphs and cooperation in games. *Mathematics of Operations Research*, 2:225–229, 1977.

[4] P. Borm, G. Owen, and S. Tijs. On the position value for communication situuations. *SIAM Journal on Discrete Mathematics*, 3:305–320, 1992.

[5] G.-C. Rota. On the foundations of combinatorial theory I. theory of Möbius functions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 2:340–368, 1964.

[6] K. Fujimoto. Interaction indices with respect to fuzzy measures (in japanese). *Journal of Japan Society of Fuzzy Theory and Intelligent Informatics*, 16(4):303–310, 2004.

[7] K. Fujimoto, I. Kojadinovic, and J.-L. Marichal. Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. *Games and Economic Behavior*, 55:72–99, 2006.

[8] J. Harsanyi. A simplified bargaining model for the n-person cooperative game. *International Economic Review*, 4:194–220, 1963.

[9] G. Hamiache. A value with incomplete information. *Games and Economic Behaviour*, 26:59–78, 1999.

[10] L. Shapley. A value for *n*-person games. *Contribution Theory of Games, II. Annals of Mathematics Studies*, 28:307–317, 1953.

ISBN: 978-989-95079-6-8

Ranking of fuzzy numbers, some recent and new formulas

Saeid Abbasbandy

Department of Mathematics, Science and Research Branch
Islamic Azad University, Tehran, 14778, Iran
Email: abbasbandy@yahoo.com

Abstract— *Ranking of fuzzy numbers plays a very important role in linguistic decision making and some other fuzzy application systems. Several strategies have been proposed for ranking of fuzzy numbers. Each of these techniques has been shown to produce non-intuitive results in certain cases. In this paper, some new approaches for ranking of trapezoidal fuzzy numbers are introduced.*

Keywords— Magnitude of fuzzy number, Parametric form of fuzzy number, Ranking of fuzzy numbers, Trapezoidal fuzzy number.

1 Introduction

Ranking of fuzzy numbers is an important component of the decision process in many applications. More than 30 fuzzy ranking indices have been proposed since 1976. In 1976 and 1977, Jain [1, 2] proposed a method using the concept of maximizing set to order the fuzzy numbers. Jain's method is that the decision maker considers only the right side membership function. A canonical way to extend the natural ordering of real numbers to fuzzy numbers was suggested by Bass and Kwakernaak [3] as early as 1977. Dubios and Prade 1978 [4], used maximizing sets to order fuzzy numbers. In 1979, Baldwin and Guild [5] indicated that these two methods have some disturbing disadvantages. Also, in 1980, Adamo [6] used the concept of α-level set in order to introduce α-preference rule. In 1981 Chang [7] introduced the concept of the preference function of an alternative. Yager in 1981 [8, 9] proposed four indices which may be employed for the purpose of ordering fuzzy quantities in $[0, 1]$. Bortolan and Degani have been compared and reviewed some of these ranking methods [10]. Chen and Hwang [11] thoroughly reviewed the existing approaches, and pointed out some illogical conditions that arise among them. Chen [12], Choobineh [13], Cheng [14] have presented some methods , and also more recently numerous ranking techniques have been proposed and investigated by Chu, Tsao [15] and Ma, Kandel and Friedman [16]. Nowadays many researchers have developed methods to compare and to rank fuzzy numbers. Some of those methods are counter-intuitive and non discriminating [18, 19, 20, 21, 22] and recently some methods based on different distance functions have been introduced for ranking of fuzzy numbers [23, 24, 25, 26, 27].

2 Preliminaries

Though there are a number of ways of defining fuzzy numbers, for the purposes of this paper we adopt the following definition, we will identify the name of the number with that of its membership function for simplicity. Throughout this paper, $\mathbb{R}$ stands for the set of all real numbers, E stands the set of fuzzy numbers, $u(x)$ for the membership function of every $u \in E$ and $x \in \mathbb{R}$.

Definition 2.1 *[28] A fuzzy number is a fuzzy set like $u : \mathbb{R} \to I = [0, 1]$ which satisfies:*

1. *u is upper semi-continuous,*
2. *$u(x) = 0$ outside some interval $[a, d]$,*
3. *There are real numbers $a,\ b$ such that $a \le b \le c \le d$ and*
   a. *$u(x)$ is monotonic increasing on $[a, b]$,*
   b. *$u(x)$ is monotonic decreasing on $[c, d]$,*
   c. *$u(x) = 1,\ b \le x \le c$.*

The membership function u can be expressed as

$$u(x) = \begin{cases} u_L(x), & a \le x \le b, \\ 1, & b \le x \le c, \\ u_R(x), & c \le x \le d, \\ 0, & otherwise, \end{cases}$$

where $u_L : [a, b] \longrightarrow [0, 1]$ and $u_R : [c, d] \longrightarrow [0, 1]$ are left and right membership functions of fuzzy number u. An equivalent parametric form is also given in [29] as follows:

Definition 2.2 *[29] A fuzzy number u in parametric form is a pair $(\underline{u}, \overline{u})$ of functions $\underline{u}(r), \overline{u}(r),\ 0 \le r \le 1$, which satisfy the following requirements:*

1. *$\underline{u}(r)$ is a bounded monotonic increasing left continuous function,*
2. *$\overline{u}(r)$ is a bounded monotonic decreasing left continuous function,*
3. *$\underline{u}(r) \le \overline{u}(r),\ 0 \le r \le 1$.*

The trapezoidal fuzzy number $u = (x_0, y_0, \sigma, \beta)$, with two defuzzifier x_0, y_0, and left fuzziness $\sigma > 0$ and right fuzziness $\beta > 0$ is a fuzzy set where the membership function is as

$$u(x) = \begin{cases} \dfrac{1}{\sigma}(x - x_0 + \sigma), & x_0 - \sigma \le x \le x_0, \\ 1 & x \in [x_0, y_0], \\ \dfrac{1}{\beta}(y_0 - x + \beta), & y_0 \le x \le y_0 + \beta, \\ 0, & otherwise, \end{cases}$$

and its parametric form is

$$\underline{u}(r) = x_0 - \sigma + \sigma r, \quad \overline{u}(r) = y_0 + \beta - \beta r.$$

ISBN: 978-989-95079-6-8

Let E_{TR} be the set of all trapezoidal fuzzy numbers on the real line. Provided that, $x_0 = y_0$ then u is a triangular fuzzy number, and we write $u = (x_0, \sigma, \beta)$. The support of fuzzy number u is defined as follows:

$$supp(u) = \overline{\{x \mid u(x) > 0\}},$$

where $\overline{\{x \mid u(x) > 0\}}$ is closure of set $\{x \mid u(x) > 0\}$.

The addition and scalar multiplication of fuzzy numbers are defined by the extension principle and can be equivalently represented in [30, 31, 32] as follows. For arbitrary $u = (\underline{u}, \overline{u})$, $v = (\underline{v}, \overline{v})$ we define addition $(u + v)$ and multiplication by scalar $k > 0$ as

$$(\underline{u+v})(r) = \underline{u}(r) + \underline{v}(r), \ \ (\overline{u+v})(r) = \overline{u}(r) + \overline{v}(r), \quad (1)$$

$$(\underline{ku})(r) = k\underline{u}(r), \ \ (\overline{ku})(r) = k\overline{u}(r). \quad (2)$$

To emphasis the collection of all fuzzy numbers with addition and multiplication as defined by (1) and (2) is denoted by E, which is a convex cone. The image (opposite) of $u = (x_0, y_0, \sigma, \beta)$, can be defined by $-u = (-y_0, -x_0, \beta, \sigma)$ (see [32, 33]).

Definition 2.3 *For arbitrary fuzzy numbers $u = (\underline{u}, \overline{u})$ and $v = (\underline{v}, \overline{v})$ the quantity*

$$D(u,v) = \left[\int_0^1 (\underline{u}(r) - \underline{v}(r))^2 \,\mathrm{dr} + \int_0^1 (\overline{u}(r) - \overline{v}(r))^2 \,\mathrm{dr}\right]^{1/2},$$

is the distance between u and v, [16, 17, 34]. The function $D(u, v)$ is a metric in E and (E, D) is a complete metric space.

The ordering indices are organized into three categories by Wang and Kerre [35] as follows:

1- **Defuzzification method**: Each index is associated with a mapping from the set of fuzzy quantities to the real line. In this case fuzzy quantities are compared according to the corresponding real numbers.

2- **Reference set method**: In this case a fuzzy set as a reference set is set up and all the fuzzy quantities to be ranked are compared with the reference set.

3- **Fuzzy relation method**: In this case a fuzzy relation is constructed to make pairwise comparisons between the fuzzy quantities involved.

Let M be an ordering method on E. The statement "two elements u and v in E satisfy that u has a higher ranking than v when M is applied" will be written as "$u \succ v$ by M". "$u \sim v$" and "$u \succeq v$" are similarly interpreted. The following reasonable properties for the ordering approaches are introduced by Wang and Kerre [35].

Reasonable properties (axioms)

A_1 : For an arbitrary finite subset Γ of E and $u \in \Gamma, u \succeq u$.

A_2 : For an arbitrary finite subset Γ of E and $(u, v) \in \Gamma^2, u \succeq v$ and $v \succeq u$, we should have $u \sim v$.

A_3 : For an arbitrary finite subset Γ of E and $(u, v, w) \in \Gamma^3, u \succeq v$ and $v \succeq w$, we should have $u \succeq w$.

A_4 : For an arbitrary finite subset Γ of E and $(u, v) \in \Gamma^2$, $\inf supp(u) \geq \sup supp(v)$, we should have $u \succeq v$.

A'_4 : For an arbitrary finite subset Γ of E and $(u, v) \in \Gamma^2$, $\inf supp(u) > \sup supp(v)$, we should have $u \succ v$.

A_5 : Let Γ and Γ' be two arbitrary finite subsets of E also u and v are in $\Gamma \cap \Gamma'$. We obtain the ranking order $u \succ v$ by M on Γ' if and only if $u \succ v$ by M on Γ.

A_6 : Let $u,\ v$, $u + w$ and $v + w$ be elements of E. If $u \succeq v$, then $u + w \succeq v + w$.

A'_6 : Let $u,\ v$, $u + w$ and $v + w$ be elements of E. If $u \succ v$, then $u + w \succ v + w$, when $w \neq 0$.

A_7 : Let $u,\ v,\ uw$, and vw be elements of E and $w \geq 0$. If $u \succeq v$ then $uw \succeq vw$.

3 Some new and recent methods

3.1 Method of D-distance

Let all of fuzzy numbers are positive or negative. Without less of generality assume that all of them are positive. The membership function of $a \in R$ is $u_a(x) = 1$, if $x = a$; and $u_a(x) = 0$, if $x \neq a$. Hence if $a = 0$ we have the following

$$u_0(x) = \begin{cases} 1 & x = 0, \\ 0 & x \neq 0. \end{cases}$$

Since $u_0(x) \in E$, left fuzziness σ and right fuzziness β are 0, so for each $u \in E$

$$D(u, u_0) = \left[\int_0^1 (\underline{u}(r)^2 + \overline{u}(r)^2)\mathrm{dr}\right]^{1/2}.$$

Thus we have the following definition.

Definition 3.1 *For u and $v \in E$, define the ranking of u and v by saying*

$$u > v \ \ iff \ \ d(u, u_0) > d(v, u_0),$$
$$u < v \ \ iff \ \ d(u, u_0) > d(v, u_0),$$
$$u \approx v \ \ iff \ \ d(u, u_0) = d(v, u_0).$$

Property 3.1. Suppose u and $v \in E$ are arbitrary there

(I) If $u = v$ then $u \approx v$.

(II) If $v \subseteq u$ and $\underline{u}(r)^2 + \overline{u}(r)^2 > \underline{v}(r)^2 + \overline{v}(r)^2$ for all $r \in [0, 1]$ then $v < u$.

Remark 3.1. (I) The distance triangular fuzzy number $u = (x_0, \sigma, \beta)$ of u_0 is defined as following

$$d(u, u_0) = \left[2x_0^2 + \sigma^2/3 + \beta^2/3 + x_0(\beta - \sigma)\right]^{1/2}.$$

(II) The distance trapezoidal fuzzy number $u = (x_0, y_0, \sigma, \beta)$ of u_0 is defined as following

$$d(u, u_0) = \left[2x_0^2 + \sigma^2/3 + \beta^2/3 - x_0\sigma + y_0\beta\right]^{1/2}.$$

(III) If $u \approx v$, it is not necessary that $u = v$. Since if $u \neq v$ and $(\underline{u}(r)^2 + \overline{u}(r)^2)^{1/2}$=$(\underline{v}(r)^2 + \overline{v}(r)^2)^{1/2}$ then $u \approx v$.

3.2 Method of min distance

In this subsection, we will propose the ranking of fuzzy numbers associated with the metric D in E.

Definition 3.2 *Let $\gamma = \{v_1, v_2, \ldots, v_n\} \subseteq E$ be is set of fuzzy numbers, we define a ranking on γ by min distance as*

$$v_i \succ v_j \ \text{ iff } \ D(v_i, \tilde{M}) > D(v_j, \tilde{M}),$$

ISBN: 978-989-95079-6-8

$$v_i \prec v_j \text{ iff } D(v_i, \tilde{M}) < D(v_j, \tilde{M}),$$
$$v_i \sim v_j \text{ iff } D(v_i, \tilde{M}) = D(v_j, \tilde{M}),$$

where $\tilde{M} = \tilde{min}\{v_1, \ldots, v_n\}$.

Dubois and Prade [30] present rules for computing $\tilde{min}$ and also comment on the properties of $\tilde{min}$.

Remark 3.2. The min distance, has the properties $A_1, A_2, \ldots, A_4'$.

3.3 Method of sign distance

Definition 3.3 *For arbitrary fuzzy numbers* $u = (\underline{u}, \overline{u})$ *and* $v = (\underline{v}, \overline{v})$*,the function*

$$D_p(u,v) = \left[\int_0^1 |\underline{u}(r) - \underline{v}(r)|^p \mathrm{dr} + \int_0^1 |\overline{u}(r) - \overline{v}(r)|^p \mathrm{dr}\right]^{1/p}, \quad (p \geq 1)$$

is the distance between u *and* v.

Definition 3.4 *Let* $\gamma : E \longrightarrow \{-1, 1\}$ *be a function that is defined as follows:*

$$\forall\, u \in E : \gamma(u) = sign\left[\int_0^1 (\underline{u}(r) + \overline{u}(r))\mathrm{dr}\right],$$

where

$$\gamma(u) = \begin{cases} 1 & \text{if } sign\left(\int_0^1 (\underline{u} + \overline{u})(r)\mathrm{dr}\right) \geq 0, \\ -1 & \text{if } sign\left(\int_0^1 (\underline{u} + \overline{u})(r)\mathrm{dr}\right) < 0. \end{cases}$$

Remark 3.3. (I) If $\inf supp(u) \geq 0$ or $\inf \underline{u}(r) \geq 0$ then $\gamma(u) = 1$.

(II) If $\sup supp(u) < 0$ or $\sup \overline{u}(r) < 0$ then $\gamma(u) = -1$.

Definition 3.5 *For* $u \in E$,

$$d_p(u, u_0) = \gamma(u) D_p(u, u_0),$$

is called sign distance.

Definition 3.6 *For* u *and* $v \in E$*, define the ranking of* u *and* v *by* d_p *on* E*, i.e.*

$$u \succ v \text{ iff } d_p(u, u_0) > d_p(v, u_0),$$
$$u \prec v \text{ iff } d_p(u, u_0) < d_p(v, u_0),$$
$$u \sim v \text{ iff } d_p(u, u_0) = d_p(v, u_0).$$

Remark 3.4. (I) The function d_p, sign distance, has the properties $A_1, A_2, \ldots, A_5$.

(II) The function d_p, sign distance, for $p = 1$ has the properties A_6, A_6' if

$$\inf\{supp(u), supp(v), supp(u+w), supp(v+w)\} \geq 0$$

or

$$\sup\{supp(u), supp(v), supp(u+w), supp(v+w)\} \leq 0.$$

(III) Suppose u and $v \in E$ are arbitrary, then

(a) If $u = v$ then $u \sim v$,

(b) If $v \subseteq u$ and $\gamma(u)\left(|\underline{u}(r)|^p + |\overline{u}(r)|^p\right) > \gamma(v)\left(|\underline{v}(r)|^p + |\overline{v}(r)|^p\right)$ for all $r \in [0,1]$ then $v \prec u$.

(IV) If $u \sim v$, it is not necessary that $u = v$. Since if $u \neq v$ and $\gamma(u)(|\underline{u}(r)|^p + |\overline{u}(r)|^p) = \gamma(v)(|\underline{v}(r)|^p + |\overline{v}(r)|^p)$ then $u \sim v$.

(V) If $u \preceq v$ then $-u \succeq -v$.

Therefore we can simply rank the fuzzy numbers by the defuzzification of $d_p(u, u_0)$. By Remark 3.4(V) we can logically infer ranking order of the images of the fuzzy numbers.

3.4 Method of H-distance

Definition 3.7 *A continuous function* $s : [0,1] \longrightarrow [0,1]$ *with the following properties is a* source *function*

1. $s(0) = 0$,
2. $s(1) = 1$,
3. $s(r)$ *is increasing.*
4. $\int_0^1 s(r)\mathrm{dr} = \frac{1}{2}$.

In fact, a reducing has the reflect of weighting the influence of the different r-cuts and diminishes the contribution of the lower r-levels. This is reasonable since these levels arises from values of membership function for which there is a considerable amount of uncertainty. For example, we can use $s(r) = r$.

Definition 3.8 *The Value and Ambiguity of a fuzzy number* $\widetilde{u}$ *are defined as follows, [36],*

$$Val_s(\widetilde{u}) = \int_0^1 s(r)[\overline{u}(r) + \underline{u}(r)]\mathrm{dr},$$
$$Amb_s(\widetilde{u}) = \int_0^1 s(r)[\overline{u}(r) - \underline{u}(r)]\mathrm{dr}.$$

Definition 3.9 *For* $\widetilde{u}, \widetilde{v} \in E$*, we define* ***H-distance*** *of* $\widetilde{u}$ *and* $\widetilde{v}$ *by*

$$D_H^s(\widetilde{u}, \widetilde{v}) = \frac{1}{2}\big\{\, |Val_s(\widetilde{u}) - Val_s(\widetilde{v})| + |Amb_s(\widetilde{u}) - Amb_s(\widetilde{v})| + d_H([\widetilde{u}]^1, [\widetilde{v}]^1)\big\},$$

where d_H *is the Hausdorff metric between intervals, and* $[.]^1$ *is the 1-cut representation of a fuzzy number.*

Property 3.2. The source distance, D_H^s, is a metric on E_{TR} and a pseudo-metric on E.

Remark 3.5. By the metric D_H^s and an arbitrary reference fuzzy set (like subsection 3.1 or 3.2), we can define a new ordering for fuzzy numbers.

3.5 Method of source distance

Definition 3.10 *For* $\widetilde{u}, \widetilde{v} \in E$*, we define* ***source distance*** *of* $\widetilde{u}$ *and* $\widetilde{v}$ *by*

$$D_s(\widetilde{u}, \widetilde{v}) = \frac{1}{2}\big\{\, |Val_s(\widetilde{u}) - Val_s(\widetilde{v})| + |Amb_s(\widetilde{u}) - Amb_s(\widetilde{v})| + \max\big\{\, |t_v - t_u|, |m_v - m_u|\big\}\big\},$$

where $[m_u, t_u]$ *and* $[m_v, t_v]$ *are the cores of fuzzy numbers* $\widetilde{u}$ *and* $\widetilde{v}$ *respectively.*

ISBN: 978-989-95079-6-8

Property 3.3. The source distance, D_s, is a metric on E_{TR} and a pseudo-metric on E.

Remark 3.6. By the metric D_s and an arbitrary reference fuzzy set (like subsection 3.1 or 3.2), we can define a new ordering for fuzzy numbers.

3.6 Method of magnitude

For an arbitrary trapezoidal fuzzy number $u = (x_0, y_0, \sigma, \beta)$, with parametric form $u = (\underline{u}(r), \overline{u}(r))$, we define the *magnitude* of the trapezoidal fuzzy number u as

$$Mag(u) = \frac{1}{2}\Big(\int_0^1 (\underline{u}(r) + \overline{u}(r) + x_0 + y_0) f(r) \mathrm{dr} \Big),$$

where the function $f(r)$ is a non-negative and increasing function on $[0, 1]$ with $f(0) = 0$, $f(1) = 1$ and $\int_0^1 f(r)\mathrm{dr} = \frac{1}{2}$. For example, we can use $f(r) = r$. The resulting scalar value is used to rank the fuzzy numbers. In the other words $Mag(u)$ is used to rank fuzzy numbers. The larger $Mag(u)$, the larger fuzzy number. Therefore for any two trapezoidal fuzzy numbers u and $v \in E$, we define the ranking of u and v by the $Mag(.)$ on E as follows:

1. $Mag(u) > Mag(v)$ if and only if $u \succ v$,
2. $Mag(u) < Mag(v)$ if and only if $u \prec v$,
3. $Mag(u) = Mag(v)$ if and only if $u \sim v$.

Then we formulate the order $\succeq$ and $\preceq$ as $u \succeq v$ if and only if $u \succ v$ or $u \sim v$, $u \preceq v$ if and only if $u \prec v$ or $u \sim v$. In the other words, this method is placed in the first class of Kerre's categories [35].

Remark 3.7. **(I)** If $\inf supp(u) \geq 0$ or $\inf \underline{u}(r) \geq 0$ then $Mag(u) \geq 0$.

(II) If $\sup supp(u) \leq 0$ or $\sup \overline{u}(r) \leq 0$ then $Mag(u) \leq 0$.

(III) For two arbitrary trapezoidal fuzzy numbers u and v, we have

$$Mag(u + v) = Mag(u) + Mag(v).$$

(IV) For all symmetric trapezoidal fuzzy numbers $u = (-x_0, x_0, \sigma, \sigma)$,

$$Mag(u) = 0.$$

(V) For any two symmetric trapezoidal fuzzy numbers $u = (x_0, y_0, \sigma, \sigma)$ and $v = (x_0, y_0, \beta, \beta)$,

$$Mag(u) = Mag(v).$$

Property 3.4. The function $Mag(.)$ has the properties $A_1, A_2, A_3, \ldots, A_6'$.

4 Conclusions

In spite of many ranking methods, no one can rank fuzzy numbers with human intuition consistently in all cases. The proposed methods can effectively rank various fuzzy numbers and their images. These methods have some mathematical properties. Moreover some pseudo metric on the set of fuzzy numbers and metric on trapezoidal fuzzy numbers are introduced. We may conclude that these ordering methods are relatively reasonable for fuzzy numbers based on the introduced axioms.

Acknowledgment

The author is thankful to the Islamic Azad University of Iran, Science and Research Branch, for the financial support.

References

[1] R. Jain. Decision-making in the presence of fuzzy variable. *IEEE Transactions on Systems, Man, and Cybernetics*, 6:698-703,1976.

[2] R. Jain. A procedure for multi-aspect decision making using fuzzy sets. *International Journal of Systems Science*, 8:1-7,1977.

[3] S. Bass and H. Kwakernaak. Rating and ranking of multiple-aspect alternatives using fuzzy sets. *Automatica*, 13:47-58,1977.

[4] D. Dubios and H. Prade. Oprations on fuzzy numbers. *International Journal of Systems Science*, 9:613-626,1978.

[5] J.F. Baldwin and N.C.F. Guild. Comparison of fuzzy numbers on the same decision space. *Fuzzy Sets and Systems*, 2:213-233,1979.

[6] M. Adamo. Fuzzy decision trees. *Fuzzy Sets and Systems*, 4:207-219,1980.

[7] W.K. Chang. Ranking of fuzzy utilities with triangular membership functions. *International Conference on Plicy Analysis and Informations Systems, Tamkang University*, 163-171,1981.

[8] R.R. Yager. On choosing between fuzzy subsets. *Kybernetes*, 9:151-154,1980.

[9] R.R. Yager. A procedure for ordering fuzzy subests of the unit interval. *Information Sciences*, 24:143-161,1981.

[10] G. Bortolan and R. Degani. A review of some methods for ranking fuzzy numbers. *Fuzzy Sets and Systems*, 15:1-19,1985.

[11] S.J. Chen and C.L. Hwang. *Fuzzy multiple Attribute Decision Making*. Spinger-Verlag, Berlin, 1992.

[12] S. Chen. Ranking fuzzy numbers with maximizing set and minimizing set. *Fuzzy Sets and Systems*, 17:113-129,1985.

[13] F. Choobineh and H. Li. An index for ordering fuzzy numbers. *Fuzzy Sets and Systems*, 54:287-294,1993.

[14] C.H. Cheng. A new approach for ranking fuzzy numbers by distance method. *Fuzzy Sets and Systems*, 95:307-317,1998.

[15] T. Chu and C.Tsao. Ranking fuzzy numbers with an area between the centroid point and orginal point. *Computers and Mathematics with Applications*, 43:111-117,2002.

[16] M. Ma and A. Kandel and M. Friedman. A new approach for defuzzification. *Fuzzy Sets and Systems*, 111:351-356,2000.

[17] S. Abbasbandy and B. Asady. A Note on "A new approach for defuzzification". *Fuzzy Sets and Systems*, 128:131-132,2002.

[18] S. Abbasbandy and M. Amirfakhrian. A new approach to universal approximation of fuzzy functions on a discrete set of points. *Applied Mathematical Modelling*, 30:1525-1534,2006.

[19] S. Abbasbandy and M. Amirfakhrian. The nearest trapezoidal form of a generalized left right fuzzy number. *International Journal of Approximate Reasoning*, 43:166-178,2006.

[20] S. Abbasbandy and B. Asady. Ranking of fuzzy numbers by sign distance. *Information Sciences*, 176:2405-2416,2006.

[21] L.H. Chen and H.W. Lu. An approximaite approach for ranking fuzzy numbers based on left and right dominance. *Computers and Mathematics with Applications*, 41:1589-1602,2001.

[22] J. Yao and K. Wu. Ranking fuzzy numbers based on decomposition principle and signed distance. *Fuzzy Sets and Systems*, 116:275-288,2000.

[23] S. Abbasbandy and C. Lucas and B. Asady. A new method for ranking of fuzzy numbers through using distance method. *Quarterly Journal of Science, Teacher Training University*, 3:83-88,2003.

[24] G. Facchinetti and R.G. Ricci. A characterization of a general class of ranking functions on triangular fuzzy numbers. *Fuzzy Sets and Systems*, 146:297-312,2004.

[25] S. Abbasbandy and M. Otadi and M. Mosleh. Ranking of fuzzy numbers by min distance. *Mathematics Scientific Journal, Islamic Azad University of Arak*, 2:1-14,2006-2007.

[26] S. Abbasbandy and M. Amirfakhrian. The nearest approximation of a Fuzzy quantity in parametric form. *Applied Mathematics and Computation*, 172:624-632,2006.

[27] S. Abbasbandy and T. Hajjari. A new approach for ranking of trapezoidal fuzzy numbers. *Computers and Mathematics with Applications*, 57:413-419,2009.

[28] R. Goetschel and W. Voxman. Elementary calculus. *Fuzzy Sets and Systems*, 18:31-43,1986.

[29] M. Ma and M. Friedman and A. Kandel. A new fuzzy arithmetic. *Fuzzy Sets and Systems*, 108:83-90,1999.

[30] D. Dubois and H. Prade. *Fuzzy Sets and System: Theory and Application*. Academic Press, New York, 1980.

[31] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338-353,1965.

[32] H.J. Zimmermann. *Fuzzy Sets Theory and its Application*. Kluwer Academic Press, Dordrecht, 1991.

[33] A. Kauffman and M.M. Gupta. *Introduction to fuzzy arithmetic: Theory and Application*. Van Nostrand Reinhold, New York, 1991.

[34] M. Ma and A. Kandel and M. Friedman. Correction to "A New approach for defuzzification". *Fuzzy Sets and Systems*, 128:133-134,2002.

[35] X. Wang and E.E. Kerre. Reasonable properties for the ordering of fuzzy quantities (I,II). *Fuzzy Sets and Systems*, 118:375-405,2001.

[36] W. Voxman. Some remarks on distance between fuzzy numbers. *Fuzzy Sets and Systems*, 100:353-365,1998.

ISBN: 978-989-95079-6-8

Lattice properties of discrete fuzzy numbers under extended min and max

Jaume Casasnovas Juan Vicente Riera
Dept. Mathematics and Computer Science ,University of the Balearic Islands
Palma de Mallorca, Spain
Email: {jaume.casasnovas,jvicente.riera}@uib.es

Abstract— *This paper proposes to study the lattice properties of two closed binary operations in the set of discrete fuzzy numbers. Using these operations to represent the meet and the join, we prove that the set of discrete fuzzy numbers whose support is a set of consecutive natural numbers is a distributive lattice. Finally, we demonstrate that the subsets of discrete fuzzy numbers, which have the same support, are distributive lattices too.*

Keywords— Fuzzy numbers, discrete fuzzy numbers, distributive lattice.

1 Introduction

It is possible to approach the theory of fuzzy numbers in different directions: theoretical[8, 9, 10, 13], geometric [1, 2], applications in engineering [13], social science [12], lattice theory[13, 20], etc.. Voxman [18] introduced the concept of a discrete fuzzy number as a fuzzy subset of real numbers with discrete support and analogous properties to a fuzzy number (convexity, normality). Also, like fuzzy numbers, it is possible to consider discrete fuzzy numbers from different points of view: theoretical [16, 18], applications in engineering [11, 19], social sciences [17], etc.

It is well known that, the arithmetic and lattice operations such as maximum and minimum on fuzzy numbers can be approached either by the direct use of the membership function (by the Zadeh's extension principle) or by the equivalent use of the *r-cuts* representation, for instance, [13, 14, 20].

Nevertheless, in the discrete case using the same methods, this process can yield a fuzzy subset that does not satisfy the conditions to be a discrete fuzzy number [3, 19]. In previous works [3, 4], we have presented a technique that allows us to obtain a closed addition on the set of discrete fuzzy numbers, DFN, and moreover, we focus on the addition of discrete fuzzy numbers whose support is an arithmetic sequence and even a subset of consecutive natural numbers. This type of numbers arise mainly when a fuzzy cardinality of a fuzzy set [5, 6] or a fuzzy multiset [15] is considered.

In [7] we define two closed binary operations in the set of discrete fuzzy numbers to obtain the maximum and the minimum of discrete fuzzy numbers. We prove as well that in the set $\mathcal{A}_1$, of discrete fuzzy numbers whose support is a set of consecutive natural numbers, these operations coincide with the function maximum and minimum obtained using the Zadeh's extension principle.

The aim of this paper is to continue studying the properties of these operations and if it is possible to obtain a structure of lattice using them. We will see that, in general, these binary operations only fulfill the associative, commutative and idempotent laws. We show as well that the triplets ($\mathcal{A}_1$,MAX_w,MIN_w) and (F_A,MIN_w,MAX_w) are distributive lattices, where F_A is the subset of discrete fuzzy numbers whose support is the support of A, with $A \in DFN$.

2 Preliminaries

In this section, we recall some definitions and the main results about discrete fuzzy numbers which will be used later.

Definition 2.1 *[13] A fuzzy subset u of $\mathbb{R}$ with membership mapping $u : \mathbb{R} \to [0,1]$ is called* fuzzy number *if its* support *is an interval $[a,b]$ and there exist real numbers s,t with $a \leq s \leq t \leq b$ and such that:*

1. *$u(x)$=1 with $s \leq x \leq t$*
2. *$u(x) \leq u(y)$ with $a \leq x \leq y \leq s$*
3. *$u(x) \geq u(y)$ with $t \leq x \leq y \leq b$*
4. *$u(x)$ is upper semi-continuous.*

We will denote the set of fuzzy numbers by FN.

Definition 2.2 *[18] A fuzzy subset u of $\mathbb{R}$ with membership mapping $u : \mathbb{R} \to [0,1]$ is called* discrete fuzzy number *if its support is finite, i.e., there are $x_1, ..., x_n \in \mathbb{R}$ with $x_1 < x_2 < ... < x_n$ such that $supp(u) = \{x_1, ..., x_n\}$, and there are natural numbers s,t with $1 \leq s \leq t \leq n$ such that:*

1. *$u(x_i)$=1 for any natural number and i with $s \leq i \leq t$ (*core*)*
2. *$u(x_i) \leq u(x_j)$ for each natural number i,j with $1 \leq i \leq j \leq s$*
3. *$u(x_i) \geq u(x_j)$ for each natural number i,j with $t \leq i \leq j \leq n$*

From now on, we will denote the set of discrete fuzzy numbers by DFN and $DFN(\mathbb{N})$ will stand for the set of discrete fuzzy numbers whose support is a subset of the set of Natural Numbers. Finally, a discrete fuzzy number will be denoted by *dfn*.

In general, the operations on fuzzy numbers f, g can be approached either by the direct use of their membership function, $\mu_f(x), \mu_g(x)$, as fuzzy subsets of $\mathbb{R}$ and the Zadeh's extension principle:

$$O(f,g)(z) = sup\{\mu_f(x) \wedge \mu_g(y) | O(x,y) = z\}$$

or by the equivalent use of the α-*cuts* representation[13]:

ISBN: 978-989-95079-6-8

$$O(f,g)^{\alpha} = O(f^{\alpha}, g^{\alpha}) = \{O(x,y) | x \in f^{\alpha}, y \in g^{\alpha}\}$$

and

$$O(f,g)(z) = sup\{\alpha \in [0,1] | z \in O(f,g)^{\alpha}\}$$

Nevertheless, in the discrete case, this process can yield a fuzzy subset that does not satisfy the conditions to be a discrete fuzzy number [3, 19]. For example, let $u = \{0.3/1, 1/3, 0.5/7\}$ and $v = \{0.4/2, 1/5, 1/6, 0.8/9\}$ be two discrete fuzzy numbers. If we use the Zadeh's extension principle to obtain their addition, it results the fuzzy subset $S = \{0.3/3, 0.4/5, 0.3/6, 0.3/7, 1/8, 1/9, 0.3/10, 0.8/12, 0.5/13, 0.5/16\}$ which doesn't fulfill the conditions to be a discrete fuzzy number, because the third property of the definition 2.2 fails. In a previous work [3, 4] we have presented an approach to a closed extended addition ($\oplus$) of discrete fuzzy numbers after associating suitable non-discrete fuzzy numbers, which can be used like a carrier to obtain the desired addition.
In a recent paper [4] we proved that a suitable carrier can be a discrete fuzzy number whose support is an arithmetic sequence and even a subset of consecutive natural numbers. Thus, we obtained the following results:

Proposition 2.3 *[4]*
Let $\mathcal{A}_r$ be the set $\{f \in DFN\ (\mathbb{N})$, such that supp(f) is the set of terms of an arithmetic sequence with r as common difference$\}$. If $f, g \in \mathcal{A}_r$. The following facts:

1. $f \oplus g \in DFN(\mathbb{N})$
2. $f \oplus g \in \mathcal{A}_r$

hold.

Remark 2.4 *[4] Note that the set $\mathcal{A}_1$ is the set of discrete fuzzy numbers whose support is a set of consecutive natural numbers.*

Finally, we will use a kind of representation in the study of discrete fuzzy numbers:

Theorem 2.5 *[19] Let u be a dfn and let u^r be the r-cut $=\{x \in \mathbb{R} | u(x) \geq r\}$ for any $r \in (0,1]$. Let's u^0 denote the support of u. Then the following statements (1)-(4) hold:*

1. *u^r is a nonempty finite subset of $\mathbb{R}$, for any $r \in [0,1]$*
2. *$u^{r_2} \subset u^{r_1}$ for any $r_1, r_2 \in [0,1]$ with $0 \leq r_1 \leq r_2 \leq 1$*
3. *For any $r_1, r_2 \in [0,1]$ with $0 \leq r_1 \leq r_2 \leq 1$, if $x \in u^{r_1} - u^{r_2}$ we have $x < y$ for all $y \in u^{r_2}$, or x>y for all $y \in u^{r_2}$*
4. *For any $r_0 \in [0,1]$, there exist some real numbers $r_0^{'}$ with $0 < r_0' < r_0$ such that $u^{r_0'} = u^{r_0}$ (i.e. $u^r = u^{r_0}$ for any $r \in [r_0', r_0]$).*

And conversely, if for any $r \in [0,1]$, there exist $A^r \subset \mathbb{R}$ satisfying the following conditions (1)-(4):

1. *A^r is a nonempty finite for any $r \in [0,1]$*
2. *$A^{r_2} \subset A^{r_1}$, for any $r \in [0,1]$ with $0 \leq r_1 \leq r_2 \leq 1$*
3. *For any $r_1, r_2 \in [0,1]$ with $0 \leq r_1 \leq r_2 \leq 1$, if $x \in A^{r_1} - A^{r_2}$ we have $x < y$ for all $y \in A^{r_2}$, or x>y for all $y \in A^{r_2}$*
4. *For any $r_0 \in [0,1]$, there exists a real number $r_0^{'}$ with $0 < r_0^{'} < r_0$ such that $A^{r_0'} = A^{r_0}$ (i.e. $A^r = A^{r_0}$, for any $r \in [r_0', r_0]$)*

then there exists a unique $u \in DFN$ such that $u^r = A^r$ for any $r \in [0,1]$.

3 Maximum and Minimum of discrete fuzzy numbers

It's well known, for example [13], that the set of fuzzy numbers is a distributive lattice using the following operations

$$MIN(u,v)(z) = \sup_{z=\min(x,y)} \min(u(x), v(y)), \forall z \in \mathbb{R} \quad (1)$$

$$MAX(u,v)(z) = \sup_{z=\max(x,y)} \min(u(x), v(y)), \forall z \in \mathbb{R} \quad (2)$$

for each couple $u, v \in FN$. If we use the same operations to obtain a similar structure in the set of discrete fuzzy numbers we see that this result is not possible. For instance, if we consider the discrete fuzzy numbers $u = \{0.3/1, 0.4/3, 1/4\}$ and $v = \{0.5/2, 1/5, 1/6\}$ and we use the previous definition to calculate the $MIN(u,v)$, we obtain the fuzzy subset $M = \{0.3/1, 0.5/2, 0.4/3, 1/4\}$ which doesn't satisfy the conditions of the definition 2.2. In [7], the authors study this drawback and we propose a new method to calculate them. Using this method, we will see later that, it is possible to provide different subsets of the set DFN with a structure of distributive lattice.

Definition 3.1 *[7] Let u, v be two dfn. For each $\alpha \in [0,1]$, let's consider the α-cut sets: $u^{\alpha} = \{x_1^{\alpha}, ..., x_p^{\alpha}\}, v^{\alpha} = \{y_1^{\alpha}, ..., y_k^{\alpha}\}$ for u and v respectively and the set $supp(u) \bigvee supp(v) = \{x \vee y | x \in supp(u), y \in supp(v)\}$. Let's define the set:*

$$A^{\alpha} = \{z \in supp(u) \bigvee supp(v) \text{ such that}$$

$$\min(u^{\alpha} \vee v^{\alpha}) \leq z \leq \max(u^{\alpha} \vee v^{\alpha})\} =$$

$$= \{z \in supp(u) \bigvee supp(v) \text{ such that}$$

$$(\min u^{\alpha} \vee \min v^{\alpha}) \leq z \leq (\max u^{\alpha} \vee \max v^{\alpha})\}$$

i.e.:

$$A^{\alpha} = \{z \in supp(u) \bigvee supp(v) | (x_1^{\alpha} \vee y_1^{\alpha}) \leq z \leq (x_p^{\alpha} \vee y_k^{\alpha})\}$$

Proposition 3.2 *[7] The finite set A^{α}, as defined above, satisfies the properties 1,2,3 and 4 of theorem 2.5 and a discrete fuzzy number, $MAX_w(u,v)$, whose α-cuts are the finite set A^{α} exists and $MAX_w(u,v)(z) = \sup\{\alpha \in [0,1]$ such that $z \in A^{\alpha}\}$.*

Proposition 3.3 *[7] If $u, v \in \mathcal{A}_1$, then $MAX(u,v)$, defined through the extension principle, coincides with $MAX_w(u,v)$. So, if $u, v \in \mathcal{A}_1$, $MAX(u,v)$ is a discrete fuzzy number and $MAX(u,v) \in \mathcal{A}_1$.*

ISBN: 978-989-95079-6-8

Definition 3.4 *[7] Let u, v be two dfn. For each $\alpha \in [0,1]$, let's consider the α-cut sets: $u^\alpha = \{x_1^\alpha, ..., x_p^\alpha\}$,$v^\alpha = \{y_1^\alpha, ..., y_k^\alpha\}$ for u and v respectively and the set $supp(u) \bigwedge supp(v) = \{x \wedge y | x \in supp(u), y \in supp(v)\}$. Let's define the set:*

$$B^\alpha = \{z \in supp(u) \bigwedge supp(v) \text{ such that}$$

$$\min(u^\alpha \wedge v^\alpha) \leq z \leq \max(u^\alpha \wedge v^\alpha)\} =$$

$$= \{z \in supp(u) \bigwedge supp(v) \text{ such that}$$

$$(\min u^\alpha \wedge \min v^\alpha) \leq z \leq (\max u^\alpha \wedge \max v^\alpha)\}$$

i.e.:

$$B^\alpha = \{z \in supp(u) \bigwedge supp(v) | (x_1^\alpha \wedge y_1^\alpha) \leq z \leq (x_p^\alpha \wedge y_k^\alpha)\}$$

Remark 3.5 *In general, if "$*$" is a binary operation the equalities*

- $\min(u^\alpha * v^\alpha) = \min u^\alpha * \min v^\alpha$
- $\max(u^\alpha * v^\alpha) = \max u^\alpha * \max v^\alpha$

do not hold. For example, if we consider the sets $u^\alpha = \{-10, -9, -8, -7, -6, -1\}$, $v^\alpha = \{0, 1, 2, 3, 4\}$ and the usual product of real numbers as binary operation then $\min(u^\alpha \cdot v^\alpha) = -40$ and $\min(u^\alpha) \cdot \min(v^\alpha) = -10 \cdot 0 = 0$, and $\max(u^\alpha \cdot v^\alpha) = 0$ and $\max(u^\alpha) \cdot \max(v^\alpha) = -1 \cdot 4 = -4$

Proposition 3.6 *[7] The finite set B^α, as defined above, satisfies the properties 1,2,3 and 4 of Proposition 2.5 and a discrete fuzzy number, $MIN_w(u, v)$, whose α-cuts are the finite set B^α exists and $MIN_w(u, v)(z) = \sup\{\alpha \in [0, 1]$ such that $z \in B^\alpha\}$.*

Proposition 3.7 *[7] If $u, v \in \mathcal{A}_1$, then $MIN(u, v)$, defined through the extension principle, coincides with $MIN_w(u, v)$. So, if $u, v \in \mathcal{A}_1$, $MIN(u, v)$ is a discrete fuzzy number and $MIN(u, v) \in \mathcal{A}_1$.*

We have seen, in the previous propositions, that the operations $MAX_w(u, v)$ and $MIN_w(u, v)$ are discrete fuzzy numbers.

Example 3.8 *If we use the method, as explained above, to calculate the minimum of $u = \{0.3/1, 0.4/3, 1/4\}$ and $v = \{0.5/2, 1/5, 1/6\}$, we obtain the following discrete fuzzy number $MIN_w(u, v) = \{0.3/1, 0.5/2, 0.5/3, 1/4\}$ where*

$$B^{0.3} = \{z \in \{1, 2, 3, 4\} \text{ such that } 1 \leq z \leq 4\} = \{1, 2, 3, 4\}$$

$$B^{0.4} = \{z \in \{1, 2, 3, 4\} \text{ such that } 2 \leq z \leq 4\} = \{2, 3, 4\}$$

$$B^{0.5} = \{z \in \{1, 2, 3, 4\} \text{ such that } 2 \leq z \leq 4\} = \{2, 3, 4\}$$

$$B^1 = \{z \in \{1, 2, 3, 4\} \text{ such that } 4 \leq z \leq 4\} = \{4\}$$

Proposition 3.9 *Let A, B and C be three dfn such that their supports are $supp(A)$,$supp(B)$ and $supp(C)$ respectively. The following properties hold:*

1. *Commutativity*

$$supp(A) \bigwedge supp(B) = supp(B) \bigwedge supp(A)$$

and

$$supp(A) \bigvee supp(B) = supp(B) \bigvee supp(A)$$

2. *Associativity*

$$(supp(A) \bigwedge supp(B)) \bigwedge supp(C) =$$

$$= supp(A) \bigwedge (supp(B) \bigwedge supp(C))$$

and

$$(supp(A) \bigvee supp(B)) \bigvee supp(C) =$$

$$= supp(A) \bigvee (supp(B) \bigvee supp(C))$$

3. *Idempotence*

$$supp(A) \bigwedge supp(A) = supp(A)$$

and

$$supp(A) \bigvee supp(A) = supp(A)$$

4. *If the supports of A and B are the same or $supp(B) \subseteq supp(A)$ or $A, B \in \mathcal{A}_1$ then the next absorption laws*

$$supp(A) \bigwedge (supp(A) \bigvee supp(B)) = supp(A)$$

$$supp(A) \bigvee (supp(A) \bigwedge supp(B)) = supp(A)$$

hold.

5. *If the supports of A, B and C are the same or $A, B, C \in \mathcal{A}_1$ then the following distributive properties*

$$supp(A) \bigwedge (supp(B) \bigvee supp(C)) =$$

$$= (supp(A) \bigwedge supp(B)) \bigvee (supp(A) \bigwedge supp(C))$$

and

$$supp(A) \bigvee (supp(B) \bigwedge supp(C)) =$$

$$= (supp(A) \bigvee supp(B)) \bigwedge (supp(A) \bigvee supp(C))$$

hold.

6. *For any $A, B \in DFN$ then*

$$supp(A) \bigwedge (supp(B) \bigvee supp(C))) \subseteq$$

$$\subseteq (supp(A) \bigwedge supp(B)) \bigvee (supp(A) \bigwedge supp(C))$$

and

$$supp(A) \bigvee (supp(B) \bigwedge supp(C))) \subseteq$$

$$\subseteq (supp(A) \bigvee supp(B)) \bigwedge (supp(A) \bigvee supp(C))$$

7. *For any $A, B \in DFN$ then*

$$supp(A) \subseteq supp(A) \bigwedge (supp(A) \bigvee supp(B))$$

$$supp(A) \subseteq supp(A) \bigvee (supp(A) \bigwedge supp(B))$$

ISBN: 978-989-95079-6-8

Proof From the definition 2.2, the support of a discrete fuzzy number is a finite linearly ordered subset of real numbers. Moreover, it is well known that the set of real numbers is a distributive lattice with the usual operations maximum(max) and minimum(min). Let X, Y and Z denote the supports of $A, B, C \in DFN$ respectively.

1. Using the commutative property of the real functions maximum and minimum the proof is trivial.

2. Associativity

   If $z \in (X \bigwedge Y) \bigwedge Z$ then $z = \min(x, c)$ where $x = \min(a, b)$, $a \in X$, $b \in Y$ and $c \in Z$. So $z = \min(\min(a, b), c)$. Using the associativity of the function minimum, $z = \min(\min(a, b), c) = \min(a, \min(b, c))$. Then $z \in X \bigwedge (Y \bigwedge Z)$. Therefore

   $$(X \bigwedge Y) \bigwedge Z \subseteq X \bigwedge (Y \bigwedge Z)$$

   If $z \in X \bigwedge (Y \bigwedge Z)$ then $z = \min(a, x)$ where $x = \min(b, c)$, $a \in X$, $b \in Y$ and $c \in Z$. So $z = \min(a, \min(b, c))$. Using the associativity of the function minimum, $z = \min(a, \min(b, c)) = \min(\min(a, b), c)$. Then $z \in (X \bigwedge Y) \bigwedge Z$. Therefore

   $$X \bigwedge (Y \bigwedge Z) \subseteq (X \bigwedge Y) \bigwedge Z$$

   The proof of the other associative law is analogous.

3. Idempotence

   If $z \in X \bigwedge X$ then $z = \min(a, a')$ with $a, a' \in X$. Therefore $z = a \in X$ or $z = a' \in X$. This means that $z \in X$ and so $X \bigwedge X \subseteq X$. On the other hand, it is evident that $X \subseteq X \bigwedge X$ because the function minimum is idempotent and then for each $z \in X$, $z = \min(z, z)$.

   The proof of the other idempotence law is similar.

4. Absorption

   If $z \in X \bigwedge (X \bigvee Y)$ then $z = \min(a, \max(a', b))$ where $a, a' \in X$ and $b \in Y$. Then if $z = a$ or $z = a'$ obviously $z \in X$. But if $z = b$ we have that $a' \leq b \leq a$. Using the hypothesis of the proposition we obtain that $b \in X$. Therefore $X \bigwedge (X \bigvee Y) \subseteq X$.

   If $z \in X$ then $z = \min(z, \max(z, b))$ for all $b \in Y$. Because if $\max(z, b) = b$ then $\min(z, b) = z$. And, if $\max(z, b) = z$ then $\min(z, z) = z$. So, $X \subseteq X \bigwedge (X \bigvee Y)$

5. Distributivity

   Let's $A, B, C \in \mathcal{A}_1$ where $supp(A) = \{x_1, \cdots, x_n\}$,$supp(B) = \{y_1, \cdots, y_m\}$ and $supp(C) = \{z_1, \cdots, z_k\}$ respectively. Then

   $$supp(A) \bigwedge (supp(B) \bigvee supp(C))$$

   is the set of natural numbers z such that they belong to the set

   $$\{x_1, \cdots, x_n\} \bigwedge \{\max(y_1, z_1), \cdots, \max(y_m, z_k)\}$$

   , i.e., it's the set of natural numbers z such that $z \in \{\min(x_1, \max(y_1, z_1)), \cdots, \min(x_n, \max(y_m, z_k))\} =$ (using the distributive property of natural numbers with the usual order)

   $$= \{z \in \{\max(\min(x_1, y_1), \min(x_1, z_1)), \cdots$$
   $$\cdots, \max(\min(x_n, y_m), \min(x_n, z_k))\}\}$$

   Where the previous set is an interval of consecutive natural numbers. On the other hand,

   $$(supp(A) \bigwedge supp(B)) \bigvee (supp(A) \bigwedge supp(C)) =$$
   $$= \{z \in \{\{\min(x_1, y_1), \cdots, \min(x_n, y_m)\} \bigvee$$
   $$\{\min(x_1, z_1), \cdots, \min(x_n, z_k)\}\} =$$
   $$= \{z \in \{\max(\min(x_1, y_1), \min(x_1, z_1)), \cdots$$
   $$\cdots, \max(\min(x_n, y_m), \min(x_n, z_k))\}.$$

   The proof of the second distributive law is analogous.

6. If $z \in X \bigwedge (Y \bigvee Z))$ then $z = \min(a, \max(b, c))$ where $a \in X$, $b \in Y$ and $c \in Z$. Using the distributive property of the minimum and the maximum of real numbers, $z = \max(\min(a, b), \min(a, c))$. Then we obtain that $z \in (X \bigwedge Y) \bigvee (X \bigwedge Z)$ and so $X \bigwedge (Y \bigvee Z)) \subseteq (X \bigwedge Y) \bigvee (X \bigwedge Z)$.

   Analogously for the other inclusion.

7. If $z \in X$ then $z = \max(z, \min(z, b))$ for all $b \in Y$. Because if $\min(z, b) = b$ then $\max(z, b) = z$. And, if $\min(z, b) = z$ then $\max(z, z) = z$. So, $X \subseteq X \bigvee (X \bigwedge Y)$

Remark 3.10 *In general, the absorption and the distributive laws of the supports do not hold. For example, if* $X = \{4, 7, 9\}$,$Y = \{4, 6, 7\}$ *and* $Z = \{8, 9, 10\}$ *represent the supports of the discrete fuzzy numbers* A, B *and* C *respectively, then*

a) $X \bigwedge (X \bigvee Y) = \{4, 6, 7, 9\}$ *but* $X = \{4, 7, 9\}$.

b) $X \bigwedge (Y \bigvee Z) = \{4, 7, 8, 9\}$, *but* $(X \bigwedge Y) \bigvee (X \bigwedge Z) = \{4, 6, 7, 8, 9\}$.

The following theorem establishes some properties of the binary operations MAX_w and MIN_w using the same conditions and results that the previous proposition 3.9:

Theorem 3.11 *Let* MAX_w *and* MIN_w *be the binary operations on* DFN *defined by the propositions 3.2 and 3.6, respectively. Then, for any,* $A, B, C \in DFN$ *the following properties hold:*

1. Commutativity: $MIN_w(A, B) = MIN_w(B, A)$, $MAX_w(A, B) = MAX_w(B, A)$

2. Associativity:

$$MIN_w(MIN_w(A, B), C) =$$
$$= MIN_w(A, MIN_w(B, C))$$

and

$$MAX_w(MAX_w(A, B), C) =$$
$$= MAX_w(A, MAX_w(B, C))$$

ISBN: 978-989-95079-6-8

3. Idempotence: $\begin{array}{l} MIN_w(A,A)=A \\ MAX_w(A,A)=A \end{array}$

If the supports of A and B are the same or $A,B,C \in \mathcal{A}_1$ then

4. Absorption: $\begin{array}{l} MIN_w(A, MAX_w(A,B))=A \\ MAX_w(A, MIN_w(A,B))=A \end{array}$

5. Distributivity:

$$MIN_w(A, MAX_w(B,C)) =$$
$$= MAX_w(MIN_w(A,B), MIN_w(A,C))$$

And

$$MAX_w(A, MIN_w(B,C)) =$$
$$= MIN_w(MAX_w(A,B), MAX_w(A,C))$$

Proof Let A, B and C be three dfn. Let's consider the α-cut sets: $A^\alpha = \{x_1^\alpha, ..., x_p^\alpha\}$, $B^\alpha = \{y_1^\alpha, ..., y_k^\alpha\}$, $C^\alpha = \{w_1^\alpha, ..., w_l^\alpha\}$ for A, B and C respectively.

1. We want to show that $MIN_w(A,B) = MIN_w(B,A)$

It is enough to prove that the discrete fuzzy numbers $MIN_w(A,B)$ and $MIN_w(B,A)$ are the same α-cut sets for each $\alpha \in [0,1]$.

By definition 3.4

$$MIN_w(A,B)^\alpha =$$
$$= \{z \in supp(A) \bigwedge supp(B) \text{ such that}$$
$$(\min A^\alpha \wedge \min B^\alpha) \le z \le (\max A^\alpha \wedge \max B^\alpha)\} =$$
$$= \{z \in supp(A) \bigwedge supp(B) | (x_1^\alpha \wedge y_1^\alpha) \le z \le (x_p^\alpha \wedge y_k^\alpha)\} =$$
$$= \{z \in supp(A) \bigwedge supp(B) | (y_1^\alpha \wedge x_1^\alpha) \le z \le (y_p^\alpha \wedge x_k^\alpha)\} =$$
$$= MIN_w(B,A)^\alpha$$

Analogously for the other commutative law.

2. We want to see that

$$MIN_w(MIN_w(A,B),C) = MIN_w(A, MIN_w(B,C))$$

By definition 3.4

$$MIN_w(MIN_w(A,B),C)^\alpha =$$
$$= \{z \in supp(MIN_w(A,B)) \bigwedge supp(C) \text{ such that}$$
$$\min MIN_w(A,B)^\alpha \wedge \min C^\alpha \le z \le \max MIN_w(A,B)^\alpha \wedge \max C^\alpha\} =$$
$$= \{z \in supp(MIN_w(A,B)) \bigwedge supp(C) \text{ such that}$$
$$(x_1^\alpha \wedge y_1^\alpha) \wedge w_1^\alpha \le z \le (x_p^\alpha \wedge y_k^\alpha) \wedge w_l^\alpha\} =$$
$$= \{z \in (supp(A) \bigwedge supp(B)) \bigwedge supp(C) \text{ such that}$$
$$(x_1^\alpha \wedge y_1^\alpha) \wedge w_1^\alpha \le z \le (x_p^\alpha \wedge y_k^\alpha) \wedge w_l^\alpha\} =$$
$$= \{z \in (supp(A) \bigwedge supp(B)) \bigwedge supp(C) \text{ such that}$$
$$x_1^\alpha \wedge (y_1^\alpha \wedge w_1^\alpha) \le z \le x_p^\alpha \wedge (y_k^\alpha \wedge w_l^\alpha\}) =$$
$$= \{z \in supp(A) \bigwedge (supp(B) \bigwedge supp(C)) \text{ such that}$$
$$x_1^\alpha \wedge (y_1^\alpha \wedge w_1^\alpha) \le z \le x_p^\alpha \wedge (y_k^\alpha \wedge w_l^\alpha\}) =$$
$$= MIN_w(A, MIN_w(B,C))^\alpha$$

The proof of the other associative law is similar.

3. We want to demonstrate that

$$MIN_w(A,A) = A$$

By definition 3.4

$$MIN_w(A,A)^\alpha =$$
$$= \{z \in supp(A) \bigwedge supp(A) \text{ such that}$$
$$(\min A^\alpha \wedge \min A^\alpha) \le z \le (\max A^\alpha \wedge \max A^\alpha)\} =$$
$$= \{z \in supp(A) | (x_1^\alpha \wedge x_1^\alpha) \le z \le (x_p^\alpha \wedge x_p^\alpha)\} =$$
$$= \{z \in supp(A) | x_1^\alpha \le z \le x_p^\alpha\} =$$
$$= A^\alpha$$

It is the same for the other idempotence law.

4. We want to demonstrate that

$$MIN_w(A, MAX_w(A,B)) = A$$

By definition 3.4

$$MIN_w(A, MAX_w(A,B))^\alpha =$$
$$= \{z \in supp(A) \bigwedge supp(MAX_w(A,B)) \text{ such that}$$
$$\min A^\alpha \wedge \min MAX_w(A,B)^\alpha \le z \le \max A^\alpha \wedge \max MAX_w(A,B)^\alpha\} =$$
$$= \{z \in supp(A) \bigwedge supp(MAX_w(A,B)) \text{ such that}$$
$$x_1^\alpha \wedge (x_1^\alpha \vee y_1^\alpha) \le z \le x_p^\alpha \wedge (x_p^\alpha \vee y_k^\alpha\}) =$$

(using the absorption law of real numbers)

$$= \{z \in supp(A) \bigwedge supp(MAX_w(A,B)) \text{ such that}$$
$$x_1^\alpha \le z \le x_p^\alpha\} = \{z \in supp(A) \text{ such that } x_1^\alpha \le z \le x_p^\alpha\} =$$
$$= A^\alpha.$$

(The previous equality is only possible if we use the same hypotheses and results of the proposition 3.9)

5. We want to show the first distributive law, ie,

$$MIN_w(A, MAX_w(B,C)) = MAX_w(MIN_w(A,B), MIN_w(A,C))$$

By definition 3.4

$$MIN_w(A, MAX_w(B,C))^\alpha =$$
$$= \{z \in supp(A) \bigwedge supp(MAX_w(B,C)) \text{ such that}$$
$$\min A^\alpha \wedge \min MAX_w(B,C)^\alpha \le z \le \max A^\alpha \wedge \max MAX_w(A,B)^\alpha\} =$$
$$= \{z \in supp(A) \bigwedge supp(MAX_w(B,C)) \text{ such that}$$
$$x_1 \wedge (y_1^\alpha \vee w_1^\alpha) \le z \le x_p^\alpha \wedge (y_k^\alpha \vee w_l^\alpha)\} =$$

(using the distributive law of real numbers)

$$= \{z \in supp(A) \bigwedge supp(MAX_w(B,C)) \text{ such that}$$
$$(x_1^\alpha \wedge y_1^\alpha) \vee ((x_1^\alpha \wedge w_1^\alpha) \le z \le (x_p^\alpha \wedge y_k^\alpha) \vee (x_p^\alpha \wedge w_l^\alpha)\} =$$
$$= \{z \in supp(A) \bigwedge (supp(B) \bigvee supp(C)) \text{ such that}$$
$$(x_1^\alpha \wedge y_1^\alpha) \vee ((x_1^\alpha \wedge w_1^\alpha) \le z \le (x_p^\alpha \wedge y_k^\alpha) \vee (x_p^\alpha \wedge w_l^\alpha)\} =$$
$$= \{z \in (supp(A) \bigwedge supp(B)) \bigvee (supp(A) \bigwedge supp(C))$$
$$\text{such that } (x_1^\alpha \wedge y_1^\alpha) \vee ((x_1^\alpha \wedge w_1^\alpha) \le z \le (x_p^\alpha \wedge y_k^\alpha) \vee (x_p^\alpha \wedge w_l^\alpha)\} =$$
$$= MAX_w(MIN_w(A,B), MIN_w(A,C))^\alpha$$

Analogously for the other distributive law. □

ISBN: 978-989-95079-6-8

Corollary 3.12 *The triplet ($\mathcal{A}_1$,MIN_w,MAX_w) is a distributive lattice, in which MIN_w and MAX_w represent the meet and join, respectively.*

Remark 3.13 *The lattice ($\mathcal{A}_1$,MIN_w,MAX_w) can also be expressed as the pair ($\mathcal{A}_1$,$\preceq$), where $\preceq$ is a partial ordering defined as:*

$A \preceq B$ if and only if $MIN_w(A, B) = A$
or, alternatively,
$A \preceq B$ if and only if $MAX_w(A, B) = B$
for any $A, B \in \mathcal{A}_1$.

Remark 3.14 *We can also define the partial ordering in terms of α-cuts:*

$A \preceq B$ if and only if $\min(A^\alpha, B^\alpha) = A^\alpha$
$A \preceq B$ if and only if $\max(A^\alpha, B^\alpha) = B^\alpha$
for any $A, B \in \mathcal{A}_1$ and $\alpha \in (0, 1]$, where A^α and B^α are finite sets (say , $A^\alpha = \{x_1^\alpha, \cdots, x_p^\alpha\}$,$B^\alpha = \{y_1^\alpha, \cdots, y_k^\alpha\}$). Then,

$$\min(A^\alpha, B^\alpha) = \{z \in supp(A) \bigwedge supp(B) \text{ such that}$$

$$\min(x_1^\alpha, y_1^\alpha) \leq z \leq \min(x_p^\alpha, y_k^\alpha)\} = A^\alpha \text{ and}$$

$$\max(A^\alpha, B^\alpha) = \{z \in supp(A) \bigvee supp(B) \text{ such that}$$

$$\max(x_1^\alpha, y_1^\alpha) \leq z \leq \max(x_p^\alpha, y_k^\alpha)\} = B^\alpha$$

Example 3.15
Let $u = \{0.4/1, 1/2, 0.8/3, 0.6/4, 0.5/5, 0.4/6, 0.3/7\}$ and $v = \{0.3/4, 0.6/5, 0.7/6, 0.8/7, 1/8, 0.8/9\}$ be two discrete fuzzy numbers. It is easy to prove that $u \preceq v$ because $MAX_w(u, v) = v$ or equivalently $MIN_w(u, v) = u$.

Corollary 3.16 *For each finite subset X of real numbers, let's consider the subset F_X of DFN such that any element of F_X has as support X. Then, the triplet (F_X,MIN_w,MAX_w) is a distributive lattice, in which MIN_w and MAX_w represent the meet and join, respectively.*

Example 3.17 *Let $u = \{0.3/3, 0.4/5, 1/7, 0.7/8\}$ and $v = \{0.3/3, 1/5, 0.7/7, 0.6/8\}$ be two discrete fuzzy numbers. It is easy to prove that $v \preceq u$ because $MAX_w(u, v) = u$ or equivalently $MIN_w(u, v) = v$.*

Remark 3.18 *It is possible to find two discrete fuzzy numbers $A, B \in \mathcal{A}_1$ such that $MAX_w(A, B) \neq A$ and $MAX_w(A, B) \neq B$ or equivalenty $MIN_w(A, B) \neq A$ and $MIN_w(A, B) \neq B$. This means that the two discrete fuzzy numbers, A and B, are not comparable. For example, let's consider $A = \{0.3/4, 0.5/5, 0.8/6, 1/7, 0.9/8, 0.7/9\}, B = \{0.5/6, 1/7, 0.9/8, 0.6/9, 0.5/10\} \in \mathcal{A}_1$. Then, $MIN_w(A, B) = \{0.3/4, 0.5/5, 0.8/6, 1/7, 0.9/8, 0.6/9\} \neq A, B$.*

Acknowledgment

We would like to express our thanks to anonymous reviewers who have contributed to improve this article.

References

[1] J.J. Buckley and Y. Qu. Solving systems of linear fuzzy equations. *Fuzzy Sets and Systems*, 43(2): 33–43, 1991.

[2] J.J. Buckley and E.Eslami. Fuzzy plane geometry I: Points and lines. *Fuzzy Sets and Systems*, 86(2): 179–188, 1997.

[3] J. Casasnovas and J. Vicente Riera. On the addition of discrete fuzzy numbers. *Wseas Transactions on Mathematics* 5(5): 549–554, 2006.

[4] J. Casasnovas and J. Vicente Riera. Discrete fuzzy numbers defined on a subset of natural numbers. *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing:Advances in Soft Computing* , Springer, 42: 573–582, 2007.

[5] J. Casasnovas and J.Torrens. An Axiomatic Approach to the fuzzy cardinality of finite fuzzy sets. *Fuzzy Sets and Systems*, 133: 193–209, 2003.

[6] J. Casasnovas and J.Torrens. Scalar cardinalities of finite fuzzy sets for t-norms and t-conorms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*,11 : 599–615, 2003.

[7] J. Casasnovas, J. Vicente Riera. "Maximum and Mininum of Discrete Fuzzy numbers". *Frontiers in Artificial Intelligence and Applications: artificial intelligence research and development*,IOS Press,163: 273–280, 2007.

[8] D. Dubois. A new definition of the fuzzy cardinality of finite sets preserving the classical additivity property. *Bull.Stud.Fuzziness Appl.(BUSEFAL)*, 5: 11–12, 1981.

[9] D.Dubois and H.Prade. *Fuzzy Sets and Systems:Theory and Applications*. Academic Press, New York, 1980.

[10] D.Dubois and H.Prade (Eds.).*Fundamentals of Fuzzy Sets*. The handbooks of Fuzzy Sets Series, Kluwer, Boston,2000.

[11] M.Hanns. On the implementation of fuzzy arithmetical operations for engineering problems.*Fuzzy Information Processing Society(NAFIPS)*:462–466, 1999.

[12] Wen-Ling Hung and Miin-Shen Yang, Fuzzy clustering on LR-type fuzzy numbers with an application in Taiwanese tea evaluation. *Fuzzy Sets and Systems* ,150: 561–577, 2005.

[13] George J. Klir and Yuan Bo. *Fuzzy sets and fuzzy logic (Theory and applications)*. Prentice Hall, 1995.

[14] G. Mayor et al. Multi-dimensional Aggregation of Fuzzy Numbers Through the Extension Principle. *Data Mining, Rought sets and Granular Computing*, Eds Lin, Yao, Zadeh, Physica-Verlag: 350-363, 2002.

[15] S. Miyamoto. Fuzzy Multisets and Their Generalizations. *Multisets Processing(LNCS)*, 2235 : 225–235, 2001.

[16] Daniel Rocacher, Patrick Bosc. *The set of fuzzy rational numbers and flexible querying*. Fuzzy Sets and Systems, 155(3): 317-339, 2005.

[17] D. Tadic. Fuzzy Multi-criteria approach to ordering policy ranking in a supply chain. *Yugoslav Journal of Operations Research*: 243–258, 2005.

[18] W. Voxman. Canonical representations of discrete fuzzy numbers. *Fuzzy Sets and Systems*, 54: 457–466, 2001.

[19] Guixiang Wang,Cong Wu, Chunhui Zhao. Representation and Operations of discrete fuzzy numbers. *Southeast Asian Bulletin of Mathematics* 28: 1003–1010, 2005.

[20] Kun-lun Zhang, Kaoru Hirota. On fuzzy number lattice $(\widetilde{R}, \leq)$. *Fuzzy Sets and Systems*, 92: 113–122, 1997.

ISBN: 978-989-95079-6-8

On Some Classes of Aggregation Functions that are Migrative

János Fodor Imre J. Rudas

Institute of Intelligent Systems, Budapest Tech
Budapest, Hungary
Email: {fodor,rudas}@bmf.hu

***Abstract*— *In this paper we introduce and describe two classes of aggregation functions with members that are migrative. First continuous triangular norms are studied that own the migrative property with respect to another fixed t-norm, in particular to the three prototypes minimum, product, and the Łukasiewicz t-norm. For classes of nilpotent and strict migrative t-norms the characterization and construction is carried out by solving functional equations for the generators. For the third case when the fixed t-norm is the minimum, an ordinal-sum-like construction is resulted. Then we formulate and study the migrative property for quasi-arithmetic means. The results are less permissive than for t-norms: we can hardly escape from the classical arithmetic mean.*

***Keywords*—** Triangular norms, quasi-arithmetic means, functional equations, migrative property.

1 Introduction

In this paper we study the migrative property for some classes of binary *aggregation functions*. These are increasing functions $F\colon [a,b]\times[a,b]\to[a,b]$ such that $F(a,a)=a$ and $F(b,b)=b$.

It is well-known that some operations in fuzzy logic (like t-norms, t-conorms, uninorms, nullnorms) can be considered aggregation functions on the closed unit interval in the above sense. In addition, these operations own a characteristic property of associativity that can be formulated as a well-known functional equation. A function $F\colon [a,b]\times[a,b]\to[a,b]$ is called *associative* if

$$F(x,F(y,z))=F(F(x,y),z),\quad x,y,z\in[a,b]. \tag{1}$$

Under some additional conditions, the above-mentioned logical operations are the general solutions of the associativity equation (1).

Another important class of aggregation functions consists of *means*. One of the first authors who introduced the term "mean" in a mathematical sense was Cauchy [1]. In that spirit, a function $F\colon [a,b]\times[a,b]\to[a,b]$ is called a *mean* in the interval $[a,b]\subset\mathbb{R}$ if

$$\min(x,y)\le F(x,y)\le\max(x,y)$$

for all $x,y\in[a,b]$. Well-known examples are the median, the arithmetic mean, the geometric mean, the root-power mean, the harmonic mean and the like.

Although these means and t-norms, t-conorms, uninorms, nullnorms have rather different properties, there is one particular common property they all share. This is called *bisymmetry*, introduced by Aczél [2], and is defined for a function $M\colon [a,b]\times[a,b]\to[a,b]$ by

$$F(F(x,y),F(u,v))=F(F(x,u),F(y,v)), \tag{2}$$

for all $x,y,u,v\in[a,b]$.

It is easy to see that if a function $F\colon [a,b]\times[a,b]\to[a,b]$ is symmetric (i.e., $F(x,y)=F(y,x)$ for all $x,y\in[a,b]$) and associative then F is bisymmetric. The converse statement is not true in general. If, however, F is symmetric and has a neutral element, then bisymmetry of F implies that F is associative [3].

We give the definition of α-migrative two-variable functions in Definition 1 just as it was proposed originally in [5]. Since that notion has something to do, in essence, with triangular norms, first we attempt to catch hold of the very meaning of it for t-norms. This is expressed by equation (6), and two more equivalent forms in Theorem 2. That is, for t-norms we can choose any of the three forms of (8), (9) or (10). Equation (8) is obtained by considering the classical associativity functional equation, then fixing the t-norm (T_0) inside the equation, and finally fixing one of the variables' value at α. Equation (9) can be obtained similarly: the only difference is that we fixed the t-norm outside the equation. Equation (10) is just equivalent to (8) and (9) for t-norms.

Therefore, when we realized that Definition 1 can never hold for t-conorms, and it is very restrictive for quasi-arithmetic means, we tried to develop new definitions that are context dependent. The idea is to use the characteristic composite functional equation of the function class under study (e.g., the bisymmetry equation for means), then fix the function inside (or outside), and finally fix one of the variables at some α. Or alternatively, consider the equation that corresponds to (10). While for t-norms these three approaches are equivalent, for quasi-arithmetic means this is not the case at all.

According to our sketched ideas and investigations, in this paper we introduce and describe two classes of aggregation functions (triangular norms and quasi-arithmetic means) with members that are migrative. The underlying compact interval is $[0,1]$ in the study. In our formulation the migrative property is context-dependent. First we show the original property and its extensions for triangular norms. Then migrativity is reformulated and studied for quasi-arithmetic means. In the conclusions we highlight the essential differences between the two classes.

2 The migrative property for t-norms

In this section we highlight some essential aspects and consequences of the original α-migrative property and its possible extensions.

A *triangular norm* (t-norm for short) $T:[0,1]^2\to[0,1]$ is an associative, commutative, non-decreasing function such that $T(1,x)=x$ for all $x\in[0,1]$. Prototypes of t-norms are

ISBN: 978-989-95079-6-8

the minimum $T_{\mathbf{M}}(x,y)=\min(x,y)$, the product $T_{\mathbf{P}}(x,y)=xy$, and the Łukasiewicz t-norm $T_{\mathbf{L}}(x,y)=\max(x+y-1,0)$. Obviously, the product t-norm $T_{\mathbf{P}}$ is α-migrative for any $\alpha\in]0,1[$.

As it is well-known, each continuous Archimedean t-norm T can be represented by means of a continuous additive generator (see e.g. [4]), i.e., a strictly decreasing continuous function $t\colon[0,1]\to[0,\infty]$ with $t(1)=0$ such that

$$T(x,y)=t^{(-1)}(t(x)+t(y)), \tag{3}$$

where $t^{(-1)}\colon[0,\infty]\to[0,1]$ is the pseudo-inverse of t, and is given by

$$t^{(-1)}(u)=t^{-1}(\min(u,t(0))).$$

2.1 The Class of α-Migrative Functions

In [5] the authors introduced a new term – α-migrative – for a class of two-variable functions (binary operations) as follows.

Definition 1. Let α be in $]0,1[$. A binary operation $T\colon[0,1]^2\to[0,1]$ is said to be *α-migrative* if we have

$$T(\alpha x,y)=T(x,\alpha y)\quad\text{for all } x,y\in[0,1]. \tag{4}$$

Although this definition seems to be rather general, the paper [5] deals with triangular norms and subnorms. Further progress can be seen in [6], where the authors studied a slightly modified definition of migrativity. They discussed also the associativity and bisymmetry of migrative aggregation functions.

Notice also that Definition 1 does not provide a meaningful notion for triangular conorms. Indeed, if S is a t-conorm then it is α-migrative if and only if $S(\alpha x,y)=S(x,\alpha y)$ holds for all $x,y\in[0,1]$. If we choose $y=0$ then we must have $\alpha x=x$ for all $x\in[0,1]$, because S is α-migrative. This is impossible when $\alpha\neq 1$. Similarly, if $y=1$ then we must have $S(x,\alpha)=1$ for all $x\in[0,1]$, which is again impossible unless $\alpha=1$. Therefore, even the correct definition of α-migrative t-conorms needs special care.

Obviously, the product t-norm $T_{\mathbf{P}}$ is α-migrative for any $\alpha\in]0,1[$. The following non-continuous t-norm T_β is also α-migrative:

$$T_\beta(x,y)=\begin{cases}\min(x,y) & \text{if } \max(x,y)=1,\\ \beta xy & \text{otherwise.}\end{cases} \tag{5}$$

It was provided in [8] as a counterexample to the question: Is a strictly monotone t-norm always continuous?

We provide all *continuous* solutions of the equation (4) in Theorem 1. More details and proofs can be found in our paper [7].

Theorem 1. *Let α be in $]0,1[$. Suppose T is a continuous t-norm. Then the following statements hold.*

(a) T is α-migrative if and only if we have

$$T(\alpha,y)=\alpha y\quad \textit{for all } y\in[0,1]. \tag{6}$$

(b) If T is α-migrative then T is strict.

(c) Suppose t is an additive generator of T. Then T is α-migrative if and only if there exists a continuous, strictly decreasing function t_0 from $[\alpha,1]$ to the non-negative reals with $t_0(0)<+\infty$ and $t_0(1)=0$ such that

$$t(x)=k\cdot t_0(\alpha)+t_0\left(\frac{x}{\alpha^k}\right)\quad \textit{if } x\in]\alpha^{k+1},\alpha^k], \tag{7}$$

where k is any non-negative integer.

In the next subsection we extend the above definition by allowing any fixed t-norm instead of the product in the defining equation.

2.2 The Class of (α,T_0)-Migrative Functions

Following our plan, the next definition extends the migrative property as follows. Some more details can be found in [9] and in [10] with proofs.

Definition 2. Let α be in $]0,1[$ and T_0 a fixed triangular norm. A binary operation $T\colon[0,1]^2\to[0,1]$ is said to be *α-migrative with respect to T_0* (shortly: (α,T_0)-migrative) if we have

$$T(T_0(\alpha,x),y)=T(x,T_0(\alpha,y)) \tag{8}$$

for all $x,y\in[0,1]$.

Obviously, Definition 1 is a particular case of Definition 2 with $T_0=T_{\mathbf{P}}$.

In the next theorem we give a simple but very useful characterization of (α,T_0)-migrative t-norms.

Theorem 2. *Let α be in $]0,1[$ and T_0 a fixed triangular norm. Then the following statements are equivalent for a t-norm $T\colon[0,1]^2\to[0,1]$.*

(i) T is α-migrative with respect to T_0;

(ii) T satisfies the following equation for all $x,y\in[0,1]$:

$$T_0(T(\alpha,x),y)=T_0(x,T(\alpha,y)); \tag{9}$$

(iii) T satisfies the following equation for all $x\in[0,1]$:

$$T(\alpha,x)=T_0(\alpha,x). \tag{10}$$

We do not recall the case when $T_0=T_{\mathbf{L}}$. The interested reader can find the related results in [9, 10].

For illustrating the non-representable case $T_0=T_{\mathbf{M}}$, we consider $(\alpha,T_{\mathbf{M}})$-migrative continuous triangular norms. In the present case the $(\alpha,T_{\mathbf{M}})$-migrative property is read as follows:

$$T(\min(\alpha,x),y)=T(x,\min(\alpha,y)),\ x,y\in[0,1]. \tag{11}$$

By Theorem 2 we know that the $(\alpha,T_{\mathbf{M}})$-migrative property is equivalent to the following functional equation:

$$T(\alpha,y)=\min(\alpha,y),\quad y\in[0,1]. \tag{12}$$

The description of all $(\alpha,T_{\mathbf{M}})$-migrative continuous triangular norms, as solutions to the equation (12), is given in the following theorem. For the proof see [10].

ISBN: 978-989-95079-6-8

Theorem 3. *A continuous t-norm T is $(\alpha, T_{\mathbf{M}})$-migrative if and only if there exist two continuous t-norms T_1 and T_2 such that T can be written in the following form:*

$$T(x,y) = \begin{cases} \alpha T_1\left(\frac{x}{\alpha}, \frac{y}{\alpha}\right) & x,y \in [0,\alpha], \\ \alpha + (1-\alpha)T_2\left(\frac{x-\alpha}{1-\alpha}, \frac{y-\alpha}{1-\alpha}\right) & x,y \in [\alpha,1], \\ \min(x,y) & \textit{otherwise}. \end{cases}$$

Notice that equations (8) and (10) has something to do with the *generalized associativity equation* that has also been studied and solved under some additional conditions, see [11, 12]. It can be written as follows:

$$F(G(x,y),z) = H(x,K(y,z)). \tag{13}$$

In this general framework the particular form of $H = F$ and $K = G$ in (13) correspond to (8).

3 The class of α-migrative quasi-arithmetic means

Let $[a,b]$ be a compact real interval. Aczél [2] proved that a function $M : [a,b]^2 \to [a,b]$ is continuous, symmetric, strictly increasing in each argument, idempotent and fulfils the bisymmetry equation (2) if and only if

$$M(x,y) = f^{-1}\left(\frac{f(x)+f(y)}{2}\right), \quad x,y \in [a,b] \tag{14}$$

with some continuous strictly monotonic function f. We also know that this result still holds for intervals of the form $]a,b]$, $[a,b[$, $]a,b]$ or even for any unbounded interval of the real line (see [13], pp 250–251, 280). A function M having a representation (14) is called a *quasi-arithmetic mean*.

When one tries to extend migrativity for quasi-arithmetic means, there are several options. One possibility is to choose a particular quasi-arithmetic mean M_0 (for instance, the arithmetic mean $M_{\mathbf{A}}(x,y) = \frac{x+y}{2}$), and consider the bisymmetry equation (2) extended in the following forms, with a fixed argument (say $y = \alpha$):

$$M(M_0(x,\alpha), M_0(u,v)) = M(M_0(x,u), M_0(\alpha,v)) \tag{15}$$
$$M_0(M(x,\alpha), M(u,v)) = M_0(M(x,u), M(\alpha,v)) \tag{16}$$

For motivations of these equations we refer to Theorem 2 and formulas (8)–(10) in case of t-norms.

We indeed chose the arithmetic mean, because of its key role in the representation of quasi-arithmetic means (14). In this case equations (15) and (16) are of the following respective forms:

$$M\left(\frac{x+\alpha}{2}, \frac{u+v}{2}\right) = M\left(\frac{x+u}{2}, \frac{\alpha+v}{2}\right), \tag{17}$$

$$\frac{M(x,\alpha)+M(u,v)}{2} = \frac{M(x,u)+M(\alpha,v)}{2}, \tag{18}$$

for a fixed $\alpha \in]0,1[$ and for all $x,u,v \in [0,1]$.

First we state that equation (18) is extremely restrictive: under the condition of idempotency, M must be the arithmetic mean.

Theorem 4. *Assume that an idempotent function $M\colon [0,1]^2 \to [0,1]$ satisfies equation (18). Then M is equal to the arithmetic mean.*

Therefore, if we want to define α-migrativity in a way that allows not only the arithmetic mean owning this property, we should try the other equation (17).

Lemma 1. *Let α be in $]0,1[$. Then*

$$M(\alpha,x) = \frac{\alpha+x}{2} \textit{ for all } x \in [0,1] \tag{19}$$

is a necessary condition for having (17) satisfied by an idempotent M.

This simple observation suggests that an appropriate definition of α-migrative quasi-arithmetic means might be the following one.

Definition 3. Let α be in $]0,1[$. A quasi-arithmetic mean $M\colon [0,1]^2 \to [0,1]$ is said to be *α-migrative* if (19) holds for all $x \in [0,1]$.

We solve equation (19) first, than we relate the solution to (17).

3.1 Quasi-Arithmetic Means that Satisfy (19)

Consider a quasi-arithmetic mean M in the form of (14) with a strictly increasing, continuous generator function f defined on $[0,1]$. Suppose M satisfies (19). That is, we have

$$f^{-1}\left(\frac{f(\alpha)+f(x)}{2}\right) = \frac{\alpha+x}{2} \quad \text{for all } x \in [0,1]. \tag{20}$$

Equation (20) implies that $f(0) = \lim_{x \downarrow 0} f(x)$ and $f(1) = \lim_{x \uparrow 1} f(x)$ are finite. Indeed, for instance, if $f(0) = -\infty$ then the left-hand side of (20) equals 0, while the right-hand side is $\frac{\alpha}{2}$, a contradiction. The other cases lead to a contradiction in a similar manner.

Therefore, (20) can be written in an equivalent form as

$$\frac{f(\alpha)+f(x)}{2} = f\left(\frac{\alpha+x}{2}\right) \quad \text{for all } x \in [0,1]. \tag{21}$$

The general solution of (21) is given in the next theorem.

Theorem 5. *Assume that $f : [0,1] \to \mathbb{R}$ is a strictly increasing, continuous function. Then f satisfies (21) if and only if there exist real numbers f_0, f_α, f_1 such that f can be written in the following form:*

$$f(x) = \begin{cases} \frac{f_\alpha - f_0}{\alpha} \cdot x + f_0 & \textit{if } x \in [0,\alpha] \\ \frac{f_1 - f_\alpha}{1-\alpha} \cdot x + \frac{f_\alpha - \alpha f_1}{1-\alpha} & \textit{if } x \in]\alpha,1] \end{cases}. \tag{22}$$

That is, the most general quasi-arithmetic means that satisfy (19) are generated by a piece-wise linear generator function. It is linear on $[0,\alpha]$ and on $[\alpha,1]$.

Next we show that the piecewise linear solution of (21) becomes linear when we feed it back into (17).

Theorem 6. *Let M be a quasi-arithmetic mean that satisfies (17). Then $M(x,y) = \frac{x+y}{2}$, i.e., M equals the arithmetic mean.*

Notice that this result is somewhat disappointing: we returned back again to quasi-arithmetic means.

ISBN: 978-989-95079-6-8

4 Conclusions

It is worth comparing results presented in this paper for t-norms and for quasi-arithmetic means. For t-norms a general form of (α, T_0)-migrativity resulted in two other equivalent formulations, and also infinite families of t-norms owning that property could be obtained. For quasi-arithmetic means idempotency alone implied the arithmetic mean as the only solution in (18). When we looked for quasi-arithmetic mean solutions of (17), it was also the arithmetic mean that satisfied the equation. The only exceptions are quasi-arithmetic mean solutions to (19) with piece-wise linear generator functions.

Notice that the general case when the fixed quasi-arithmetic mean $M_0(x,y) = f_0^{-1}\left(\frac{f_0(x)+f_0(y)}{2}\right)$ differs from the arithmetic mean can be handled in a similar way.

It might be worth studying other types of composite functional equations in the spirit of the present paper.

Acknowledgment

The authors thank the referees for their valuable suggestions.

The authors gratefully acknowledge the support obtained within the frames of Bilateral Romanian-Hungarian S&T Cooperation Project conducted by the Research and Technology Fund in Hungary. János Fodor has also been supported in part by OTKA K063405, and by COST Actions IC0602 and IC0702.

References

[1] Cauchy, A.L., *Cours d'analyse de l'Ecole Royale Polytechnique, 1ère partie; Analyse algébrique* (Paris) (1821).

[2] J. Aczél, On mean values, *Bulletin of the American Math. Society*, **54** (1948) 392–400.

[3] W. Sander, Some aspects of functional equations, in: E.P. Klement and R. Mesiar, Eds., Logical, algebraic, analytic and probabilistic aspects of triangular norms (Elsevier, Amsterdam-Boston-Heidelberg-London-New York-Oxford-Paris-San Diego-San Francisco-Sydney-Tokyo, 2005), pp. 143–187.

[4] E.P. Klement, R. Mesiar, E. Pap, *Triangular Norms*, Trends in Logic. Studia Logica Library, Vol. 8, Kluwer Academic Publishers, Dordrecht, 2000.

[5] F. Durante and P. Sarkoci, A note on the convex combinations of triangular norms, *Fuzzy Sets and Systems* 159 (2008) 77–80.

[6] H. Bustince, J. Montero and R. Mesiar, Migrativity of aggregation functions, Fuzzy Sets and Systems (2008), doi: 10.1016/j.fss.2008.09.018.

[7] J. Fodor and I.J. Rudas, On continuous triangular norms that are migrative, *Fuzzy Sets and Systems* 158 (2007) 1692 – 1697.

[8] M. Budinčević and M.S. Kurilić, A family of strict and discontinuous triangular norms, *Fuzzy Sets and Systems* 95 (1998) 381–384.

[9] J. Fodor and I.J. Rudas, Further results on migrative triangular norms, In: *Proc. 6th IEEE International Conference on Computational Cybernetics*, (November 27-29, 2008, Stará Lesná, Slovakia), pp. 123-126.

[10] J. Fodor and I.J. Rudas, On generalizations of the migrativity property for t-norms (2009, forthcoming).

[11] J. Aczél, V. D. Belousov and M. Hosszú, Generalized associativity and bisymmetry on quasigroups, Acta Math. Sci. Hung. 11 (1960) 127–136.

[12] Gy. Maksa, Quasisums and generalized associativity, Aequationes Math. 69 (2005) 6–27.

[13] J. Aczél, *Lectures on Functional Equations and Applications*, (Academic Press, New York, 1966).

ISBN: 978-989-95079-6-8

Intelligent Operation Support Method based on Time Change Fuzzy Sets

Seiji YASUNOBU[1] Mitsuru KOKUHUDA[1]

1.Department of Intelligent Interaction Technologies
University of Tsukuba, Tsukuba, Japan
Email: yasunobu@iit.tsukuba.ac.jp, kokuhuda_ ics@edu.esys.tsukuba.ac.jp

Abstract— *In the present paper, an effective support method for equipment operation under a dynamic environment is proposed. The proposed method presents operation instructions that the driver should be preparing to follow in the future. Using this method, the current and future states of the surrounding environment can be considered. The time change fuzzy set, which considers the state and time, to be universe of discourse. The time change fuzzy set is used for a computer as a buildup method with expert knowledge concerning the time change in the state. Experiments to support a lane change operation were conducted using a driving simulator. The obtained results indicated that the proposed method safely supports a lane change operation, and the time lag between the instruction being given by the system and operation being performed by the driver is decreased.*

Keywords— Dynamic environment, fuzzy inference, operation support, safety, time change fuzzy sets.

1 Introduction

When a human operates equipment in a dynamic environment, in which the surrounding conditions change with time, the operator should correctly understand complex changes in the surrounding environment. In addition, the operator must judge whether an operation is appropriate within a limited time. Therefore, an equipment operation support system is necessary in order to reduce the load on the operator and to achieve safe operation.

The equipment operation support system causes a time lag between the instruction by system and the operation by the driver. This is a significant problem in a dynamic environment, into which surrounding circumstances change with time.

Fuzzy sets can quantitatively manage human subjective assessments, such as ”He is *young*,” and thought processes, such as ”IF x is *small*, THEN y is *very large*” [1, 2]. Mamdani applied fuzzy control, which is an intelligent control, experimentally to a steam engine for the first time [3]. Yasunobu achieved the selection of the control order based on a multipurpose evaluation of an expert driver using the predictive fuzzy control and applied to the automatic train operation (ATO) system [4, 5, 6]. These fuzzy sets and control methods are applied to numerous intelligent cooperative control systems based on ”fuzzy instruction” [7, 9] and a ”fuzzy target”[8]. However, in fuzzy sets, it is difficult to build in expert knowledge concerning the time change. High-performance computers that have recently been developed have enabled equipment operation support, including expert knowledge, in real time.

In the present study, we construct an effective support method for use in a dynamic environment. When operating equipment in a dynamic environment, experts perform appropriate operations by considering not only the present state but also potential future states. In the present paper, time change fuzzy sets, which considers the state and time, to be universe of discourse. Time change fuzzy sets are used to provide expert knowledge to the computer regarding the time change of the states. Time change fuzzy sets can provide a sense of time change, such as ”Speed is continuous high” and ”Temperature change to cold from hot” to a computer.

The proposed method is applied to support a lane change operation using a driving simulator.

2 Support method in a dynamic environment

In this section, an effective support method for a dynamic environment is considered. The process of the proposed system is shown in Fig. 1.

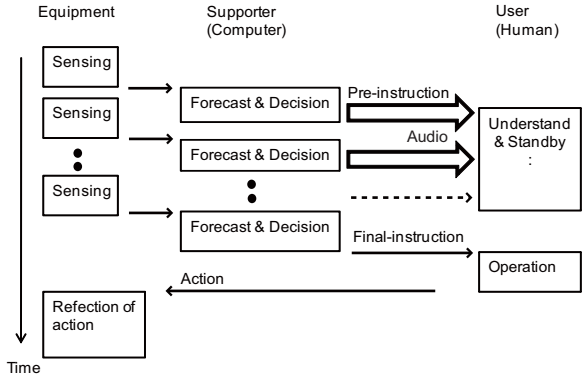


Figure 1: Process of the proposed operation support method.

The support (system) consists of 1) sensing equipment and surrounding states, 2) forecasting of future states, 3) instruction decision, and 4) presentation of instruction in advance and sending of a pre-instruction via audio. The driver (human) is 5) perceives the pre-instruction, and 6) interprets the intention of the system and waits for the final/pre-instruction. The support (system) continues 7) sensing the surrounding state and sends a final short signal. The driver (human) 8) starts an operation.

In the proposed method, the driver can confirm the operation instructions in advance. In addition, the time of 4) presentation of instruction in advance, 5) perception of instruction, 6) interpreting the intention of the system, can be sufficiently secured. The time lag can be related only to 8) operation.

The proposed method 2) forecasts future states and 3) decides pre-instruction or final-instruction, and the manner of action 2) and 3) is important.

ISBN: 978-989-95079-6-8

3 Fuzzy inference that considers future time

A computer can be made to emulate man's interpretive ability by building expert knowledge into a system using the fuzzy theory and fuzzy inference.

In a dynamic environment, an expert operates equipment by considering not only present states but also forecasting future states. It is necessary to define fuzzy sets concerning the change of states in order to build expert knowledge about the change of states and to perform inference by considering future states.

3.1 Time change fuzzy sets

In the present paper, we define a fuzzy set that changes with time. A fuzzy set to which a time axis has been added is referred to herein as a time change fuzzy set.

The total set of the state is assumed as R. The time change fuzzy set $\widetilde{X}fn$ in state $x(t)$ of the object is defined by the following expression (1):

$$\widetilde{X}_{fn(x,p)} = \int_{R\times P} \mu_{\widetilde{X}fn(x,p)}/(x,p), \quad x \in R, p \in (0,P). \tag{1}$$

Here, $\mu_{\widetilde{X}_{fn(x,p)}}$ is the membership value in universe of discourse the state x and the following time p, and P is the maximum following time.

An example of a time change fuzzy set is shown in Fig. 2. This fuzzy set describes a "change in distance from far to near". The axes of this figure indicate distance, x, the following time, p, and the membership degree of the fuzzy set μ.

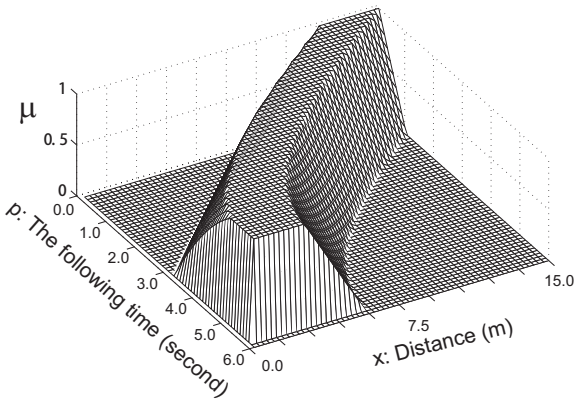


Figure 2: Example of a time change fuzzy set: change in distance from far to near.

Figure 3 shows the time change fuzzy set at $p = 0$ sec, $p = 3$ sec, and $p = 6$ sec.

Time change fuzzy sets indicate how a fuzzy set changes with time. Time change fuzzy sets can be used to build expert knowledge concerning the time change of states into a computer.

3.2 Fuzzy inference rules

Time change fuzzy sets are used to if-part and to then-part of fuzzy inference. To simplify explanation, three fuzzy rules for lane change support are described below.

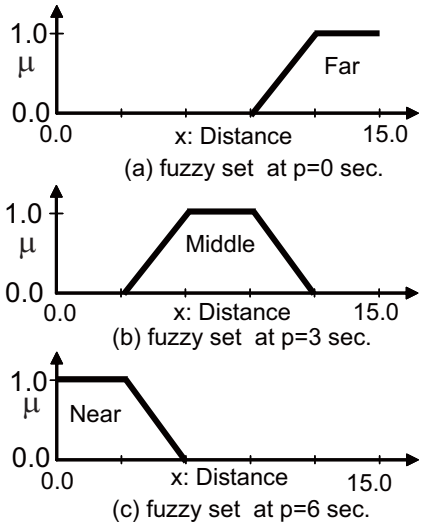


Figure 3: Time-sliced fuzzy sets of the time change fuzzy set: change in distance from far to near.

Rule 1: If forward vehicle (x) is *continuous far* and following vehicle (y) is *continuous far*, then lane change (z) is *continuous safe.*

Rule 2: If forward vehicle (x) is *continuous far* and following vehicle (y) is *change from far to near*, then lane change (z) is *change from safe to danger.*

Rule 3: If forward vehicle (x) is *change from near to far* and following vehicle (y) is *change from far to near*, then lane change (z) is *continuous middle.*

The terms x, y, and z in the above rules are defined by time change fuzzy sets, as shown in (1) . The above rules are then described as follows:

r1: $\widetilde{X}_{1(x,p)}$ *and* $\widetilde{Y}_{1(y,p)} \Rightarrow \widetilde{Z}_{1(z,p)}$,

r2: $\widetilde{X}_{2(x,p)}$ *and* $\widetilde{Y}_{2(y,p)} \Rightarrow \widetilde{Z}_{2(z,p)}$,

r3: $\widetilde{X}_{3(x,p)}$ *and* $\widetilde{Y}_{3(y,p)} \Rightarrow \widetilde{Z}_{3(z,p)}$.

3.3 Fuzzy inference process

Figure 4 shows an example of the fuzzy inference process. The if-part is given by time change fuzzy sets $\widetilde{X}_{(x,p)}$ and $\widetilde{Y}_{(x,p)}$, and the then-part is given by time change fuzzy set $\widetilde{Z}_{(x,p)}$. These fuzzy sets are in the form of time functions, as shown in Fig. 4. The input functions $\widehat{x}_{(p)}$ and $\widehat{y}_{(p)}$ are calculated by forecasting the future states from the present states.

The degree of membership of the input value to the time change fuzzy sets $\widetilde{X}'_{i(p)}$ and $\widetilde{Y}'_{i(p)}$ as shown in (2) and (3), respectively, is calculated by minimum operation, whereby time change fuzzy sets are removed by the input value. For example, when the time change fuzzy set $\widetilde{X}_{i(x,p)}$ is removed by the input value of $\widehat{x}_{(p)}$, the section area becomes $\widetilde{X}'_{i(p)}$.

$$\widetilde{X}'_{i(p)} = \widetilde{X}_{i(x,p)} \wedge \widehat{x}_{(p)} = \widetilde{X}_{i(\widehat{x}_{(p)},p)} \quad i = 1,3, \tag{2}$$

$$\widetilde{Y}'_{i(p)} = \widetilde{Y}_{i(y,p)} \wedge \widehat{y}_{(p)} = \widetilde{Y}_{i(\widehat{y}_{(p)},p)} \quad i = 1,3. \tag{3}$$

Each rule ri is evaluated to $\widetilde{Z}'_{i(z,p)}$, as shown in (4).

ISBN: 978-989-95079-6-8

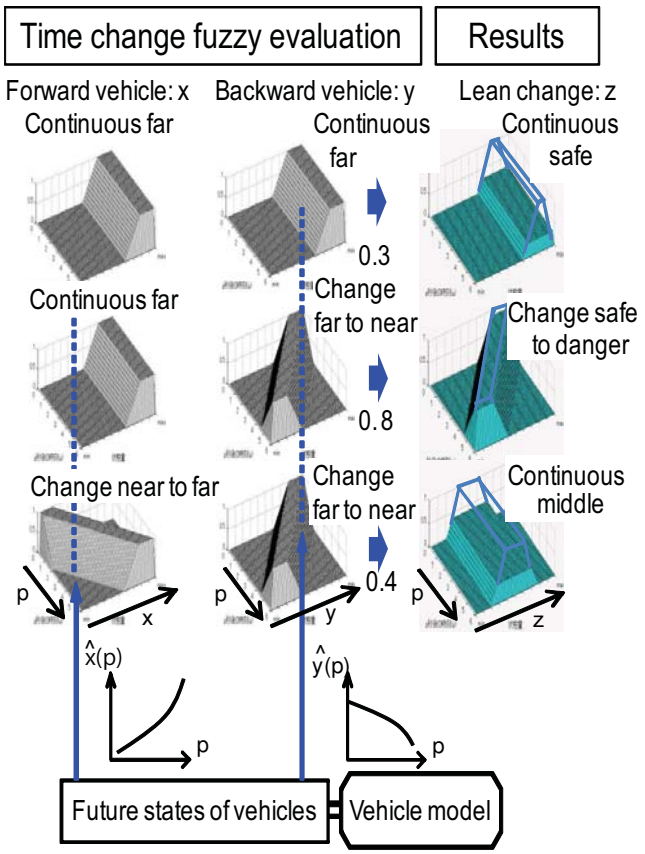


Figure 4: Fuzzy inference using time change fuzzy sets.

$$\widetilde{Z}'_{i(z,p)} = \widetilde{Z}_{i(z,p)} \wedge \widetilde{X}'_{i(p)} \wedge \widetilde{Y}'_{i(p)} \quad i = 1, 3. \tag{4}$$

3.4 *Result of fuzzy inference*

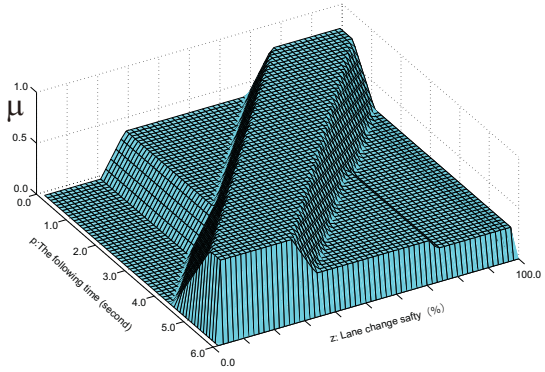
When the evaluation values of rules $\widetilde{Z}'_1$, $\widetilde{Z}'_2$, and $\widetilde{Z}'_3$ are 0.3, 0.8, and 0.4, respectively, as shown in Fig. 4, the result of fuzzy inference, i.e., $\widetilde{Z}_{ans(z,p)}$, as given by (5), is as shown in Fig. 5(a).

$$\widetilde{Z}_{ans(z,p)} = \widetilde{Z}'_{1(z,p)} \cup \widetilde{Z}'_{2(z,p)} \cup \widetilde{Z}'_{3(z,p)}. \tag{5}$$

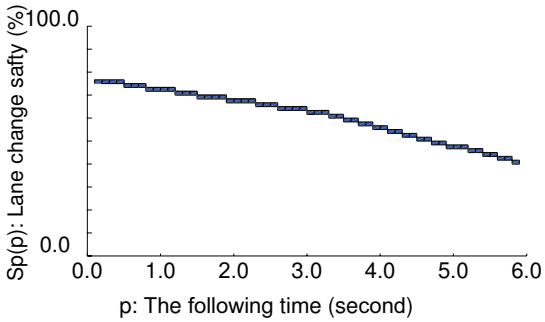
In the present study, $Sp_{(p)}$ is the lane change safety at time p obtained using the center of gravity method (defuzzification), as given by (6). Figure 5(b) shows the lane change safety at time p.

$$Sp_{(p)} = \frac{\int z \cdot \mu_{\widetilde{Z}ans(z,p)} dz}{\int \mu_{\widetilde{Z}ans(z,p)} dz} \tag{6}$$

Thus, the future output states are inferred from the future states of input, which are forecast from the present states. The system determines the operation that the driver should perform in the future based on the results of inference. The operation instruction is presented to the driver in advance, and the support method described in Section 2 is achieved. The shape of this fuzzy set looks like a type-2 fuzzy set[10]. The type-2 fuzzy set extends the membership grade to fuzzy value. Other side, this time change fuzzy set based on universe of discourse time and dynamic state, and simple membership grade from 0.0 to 1.0.



(a) Fuzzy set of lane change safety: $\widetilde{Z}_{ans}$



(b) Lane change safety at p: $Sp_{(p)}$

Figure 5: Result of fuzzy inference.

4 Application supporting a vehicle lane change

4.1 *Outline of the support system*

The support method proposed in Section 2 and fuzzy inference that considers the time change proposed in Section 3 are applied to support the lane change of a vehicle.

An expert driver can forecast the future states of other vehicles and judge intuitively whether a lane change can be performed safely.

In the present study, the knowledge of the expert driver is referred to as driving knowledge. A support system is constructed that forecasts future traffic states from present traffic states, judges whether a lane change can be performed safely and presents instructions to the driver.

It is preferable to maintain the velocity of the vehicle constant during a lane change so that the driver can operate the vehicle safely. Therefore, constant velocity is a requirement for safe lane changing. In addition, it is necessary to keep safety from the presentation of system's instruction to the operation of driver. This requirement is judged based on whether a sufficient distance between the driver's vehicle and the other vehicles will be maintained during a lane change.

The lane change support process is as follows. (a) Assume that the velocity of one's own vehicle is constant, (b) Forecast changes in the relative positions of the other vehicles, and (c) Present the instruction to change lanes if it appears that safety will be maintained.

4.2 *Forecast future states*

Other vehicles are assumed to maintain their velocities. The relative positions of the other vehicles are calculated as follows based on assumption (a) above.

ISBN: 978-989-95079-6-8

Predictive states after p seconds are as follows:

$$\widehat{x}_{(p)} = \widetilde{x}_{(t)} + v_x p \quad (7)$$

$$\widehat{y}_{(p)} = \widetilde{y}_{(t)} + v_y p, \quad (8)$$

where $\widetilde{x}_{(t)}$ and $\widetilde{y}_{(t)}$ denote the relative position of forward and backward vehicle at the present time, t, and v_x and v_y denote the relative velocity of these vehicles respectively.

4.3 Judge lane change appropriateness

In order to judge whether a lane change is possible based on the prediction of future states, expert knowledge concerning the future state of traffic is necessary. In the present study, the appropriateness of the lane change is referred to as the lane change appropriateness and is expressed as a number from 0 to 10. These numbers correspond to the following meanings: 0; dangerous(unsafe), 5; will be safe soon, and 10; safe.

Fuzzy inference is performed by using the predictive values of the relative positions of other vehicles as the input calculated in Section 4.2, and the lane change appropriateness is given as the output. The instruction to change lanes is presented to the driver based on the lane change appropriateness.

4.4 Presentation of audio instructions

In the proposed system, the driver is presented with audio instructions. The instruction is decided according to the lane change appropriateness and the change. In the present study, the instruction is output as follows.

Audio instruction:

· Present: Safety is high. Future: Safety is high.
⟶ Instruction: ”You can safely change lanes.”

· Present: Safety is high. Future: Safety is low.
⟶ Instruction: ”It’s getting dangerous.”

· Present: Safety is low. Future: Safety is high.
⟶ Instruction: ”You will be able to change lanes soon.”
After a few seconds.
⟶ Instruction: ”Change lanes now.” = Go signal

· Present: Safety is low. Future: Safety is low.
⟶ Instruction: ”You can’t change lanes.”

5 Driving simulator experiments

Experiments using the lane change support system are performed using proposed support method. The case in which the vehicle in the left-hand lane on a two-lane roadway enters the right-hand lane is assumed. The only vehicle in the left-hand lane is the driver’s own vehicle, and two or more other vehicles are driving in the right-hand lane. Fig. 6 shows a photograph of the driving simulator. The test drivers are three men in their twenties.

The following two situations are assumed in the experiments.

Situation 1) Approach from the rear: A vehicle approaches from the rear (Fig. 7(a)).

Figure 6: Driving simulator.

Situation 2) Nearby vehicle moves away: A nearby vehicle moves away gradually (Fig. 8(a)).

First, an experiment is performed using the support method in which only the present state is considered. Then, the experiment is performed using the proposed method (Section 2) in which the future states are considered. In the proposed method, time change fuzzy sets (Section 3) are used. Audio instructions are presented as described in Section 4.4.

5.1 Situation 1: Approach from the rear

5.1.1 Results obtained by considering only the present state

The results of an experiment conducted with one test driver in Situation 1 without considering future states is shown in Fig. 7(b). In this method, in which only the present state is considered, the experimental system can only judge whether the present state is ”safe” or ”unsafe”, and the safety of future states cannot be predicted. Fig. 7(b) indicates that when the driver started to operate the steering wheel, the system reported that the state changed from safe to unsafe. In this example, the time lag between the instruction of the system and the operation by the driver caused a problem.

5.1.2 Result obtained by considering the future states

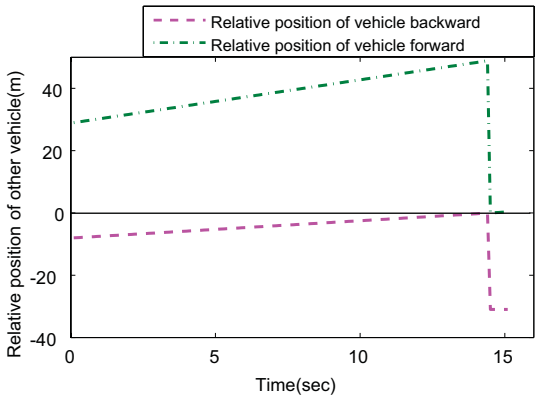
The results of an experiment conducted with one test driver in Situation 1 considering future states is shown in Fig. 7(c). As the results indicate, the experimental system did not indicate that the state became safe, and driver did not operate the steering wheel. In Situation 1, the experimental system can predict the safety of future states. In addition, the experimental system judges that indicating that the state will become safe is not appropriate considering the time of driver’s perception, understanding, and operation.

5.2 Situation 2: Nearby vehicle moves away

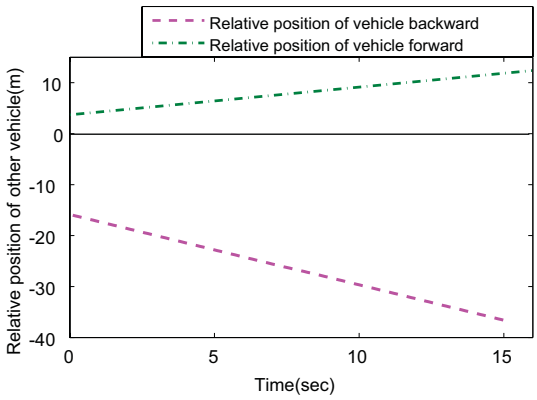
5.2.1 Results obtained by considering only the present state

The results of an experiment conducted with one test driver in Situation 2 are shown in Fig. 8(b). This figure shows the time lag between the time at which the “safe” message is provided by the system and the time at which the driver begins to operate the steering wheel. The average of time lag of the three testers is $1.70sec.$

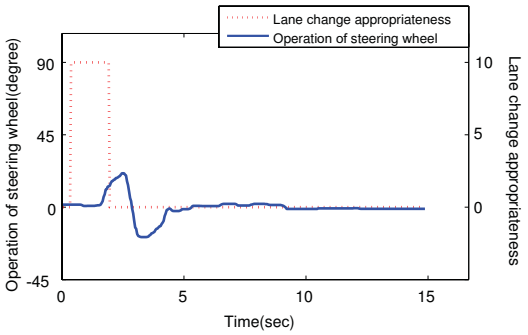
ISBN: 978-989-95079-6-8



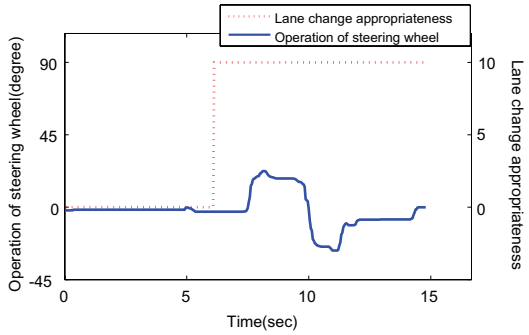
(a) Situation 1:Approach from the rear



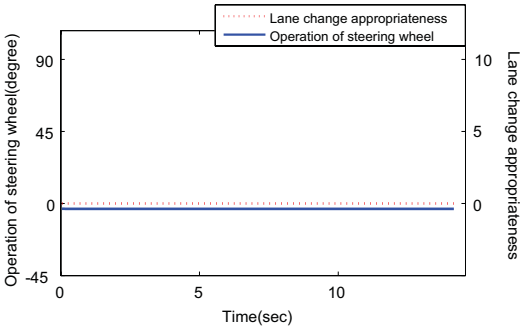
(a) Situation 2:Nearby vehicle moves away



(b) Results obtained by considering only the present state: Situation 1

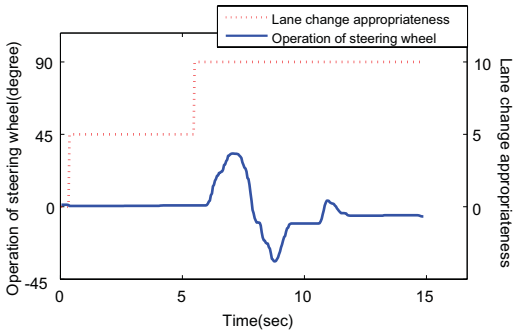


(b) Results obtained by considering only the present state: Situation 2



(c) Results obtained by considering future states: Situation 1

Figure 7: Experiments results obtained for Situation 1.



(c) Results obtained by considering future states: Situation 2

Figure 8: Experiments results obtained for Situation 2.

5.2.2 *Results obtained by considering future states*

The results of an experiment conducted with one test driver considering future states in Situation 2 are shown in Fig.8(c). The difference from Fig. 8(b) is that the message "It will be safe to change lanes soon" is presented to the driver, because this method considers not only the present state but also predicts future states. The "safe" message is then presented at the time when system judges that it will be safe to change lanes for a while. The final signal is then presented to the driver.

In this way, the driver is presented with the knowledge that he will be able change lanes soon, and can prepare to change lanes. In addition, the time required to provide the audio instruction can be shortened by using a signal to indicate that it is safe to change lanes.

As shown in Fig. 8(c), the time lag between the time at which the "safe" message is provided by the system and the time at which the driver begins to operate the steering wheel became shorter. The average time lag of three testers is $0.63sec$.

5.3 *Discussion*

The results of the experiments reveal the following.

- In Situation 1, the proposed support system can predict danger in future states, whereas the system that considers only the present state cannot predict danger in future states.
- In Situation 2, in the proposed support system, the time lag between the time at which the system provides an instruction and the time at which the driver performs an operation is shorter than that in the system that considers only the present state.

ISBN: 978-989-95079-6-8

Therefore, the effectiveness of the proposed system as a support system for use in a dynamic environment is demonstrated.

6 Conclusions

An effective support method for use in a dynamic environment is proposed in the present paper. The proposed method forecasts future states that are calculated from present states, and the driver is provided in advance with instructions to perform various operations. In order to achieve the proposed method, time change fuzzy sets are proposed as a method of building expert knowledge concerning the time change of the states into a computer.

The proposed method was applied to a lane change support system and experiments were performed. The obtained results indicate that it is possible to achieve a safety support system that can predict danger in future states and can reduce the time lag between the system providing an instruction and the driver performing an operation. In addition, the results indicate the effectiveness of the proposed support method.

References

[1] Lotfi A Zadeh. Fuzzy sets. *Information and Control*, 8: 338-353, 1965.

[2] Lotfi A Zadeh. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Trans. SMC*, 3-1: 28-44, 1973.

[3] E. H. Mamdani. Applications of Fuzzy Algorithms for Control of a Simple Dynamic Plant. *Proc. of IEEE*, 121-12: 1585-1588, 1974.

[4] S. Yasunobu, S. Miyamoto and H. Ihara. Fuzzy Control for Automatic Train Operation System. *4th IFAC/ IFIP/ IFRS Int. Conf. on Transportation Systems*, 39-45, 1983.

[5] S. Yasunobu and S. Miyamoto. Automatic Train Operation System by Predictive Fuzzy Control. *Industrial Applications of Fuzzy Control (M.Sugeno ed.)*, North-Holland, 1-18, 1985.

[6] S. Yasunobu, Fuzzy Engineering, SHOKODO, 1991(in Japanese).

[7] Y. Okamoto and S. Yasunobu. An Intelligent Cooperative Control System Based on Fuzzy Instruction. *Proc. of the 6th International Symposium on Advanced Intelligent Systems(ISIS2005)*, 341-344, 2005.

[8] Y. Chen and S. Yasunobu. Soft Target Based Obstacle Avoidance for Car-like Mobile Robot in Dynamic Environment. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2007)*, 1351-1356, 2007.

[9] S.H. Zhou and S. Yasunobu. An Intelligent Cooperative System using Fuzzy Instruction for Car-like Driving. *SICE Journal of Control, Measurement, and System Integration*, 1-4: 293-298, 2008.

[10] J.M.Mendel. Type-2 Fuzzy Sets and Systems: An Overview. *Computational Intelligence Magazine, IEEE*, 2-2: 20-29, 2007.

ISBN: 978-989-95079-6-8

Semantical evaluators

Antoon Bronselaer[1] Guy De Tré[1]
1. Department of Telecommunications and Information Processing, Ghent University
Ghent, Belgium
Email: {antoon.bronselaer,guy.detre}@ugent.be

Abstract— *In the context of a possibilistic framework for detection of object co-reference, evaluators have been defined as operators that compare two values and express the belief that such values are co-referent. Hereby, co-reference of two values means that these values describe the same entity in the real world. In this paper, a class of evaluators is investigated that determines the belief of co-reference based on semantical connections between values of the universe. These semantical connections are modeled by means of binary relations. In case these binary relations are not a-priori given, they can be (partially) learned in an iterative co-reference detection schema.*

Keywords— Possibility theory, co-reference, binary relations

Table 1: Sample of 'restaurant' dataset

name	address	city	type
campanile	624 s. la brea ave.	los angeles	american
campanile	624 s. la brea ave.	los angeles	californian
grill on the alley	9560 dayton way	los angeles	american
grill the	9560 dayton way	beverly hills	american

1 Introduction

In order to deal with the increasing amount of (partially) independent information sources, detection of duplicate information storage is of key importance to avoid inefficient storage management and inconsistencies. In the context of this paper, data are discussed in terms of *objects*. An object is a structured description of a real world entity. Hereby, structured refers to a complex user defined data structure that is used to model data. Such is the case with, among others, relational databases, XML-documents and Object Oriented environments. Objects that refer to the same entity are called *co-referent* objects. A typical property of an object is that it can be decomposed into sub-objects with a *well-defined* domain. These sub-objects are called *attributes*. Note that attributes are not bound to have an atomic data structure, i.e. any collection type is a valid datatype for an attribute. In case of relational databases, the attributes are the tuple fields.

As a running example throughout this paper, the 'restaurant' dataset [1] is used, which is a table from a relational database that describes restaurants. Each restaurant is represented by an object that has four attributes: name, address, city and type. Next to the attributes, the dataset contains a source field, which indicates the restaurant guide from which the information was taken ('fodor' or 'zagat') and an identifier field 'oid'. The goal of this dataset is to find the co-referent restaurants. The dataset is chosen as an example in this paper because syntactical evaluators are not suited for attributes 'city' and 'type'. It is shown in this paper how semantical evaluators can be used for these attributes. A sample of the dataset is shown in Table 1, where objects 1 and 3 have source 'fodor' and objects 2 and 4 have source 'zagat'. Objects 1 and 2 are co-referent objects and so are objects 3 and 4.

In a possibilistic setting, an evaluator E_U determines the possibility (or belief) of (non) co-reference about two values from a universe U. This universe can be the domain of complex objects or the domain of less complex attributes. The subject of this paper is to investigate a special class of evaluators called *semantical evaluators*. These evaluators use binary relations on the universe of interest to determine the belief that two values are co-referent. The transfer of knowledge from a relational level to a comparative level has already been studied in literature. The first notion of semantical connections between linguistic terms is pointed out by Quillian [2]. Resnik uses a network of semantical connections to construct a *semantical similarity measure* [3]. The key idea in this model is to compare the information that is held by two concepts in order to compute the similarity between two words. Further on, Resnik uses an existing taxonomy as knowledge base. This interesting approach has led to several further enhancements, which are all based on information theory [4, 5]. The knowledge base T reflects relations between elements of a universe of concepts. Each approach then uses T to construct a similarity measure. The research presented in this paper is in line of these approaches and provides a theoretical framework for semantical evaluators. The case in which the the taxonomy T is not a-priori known, will be treated in this paper, leading to an unsupervised schema for construction of binary relations. Unsupervised approaches for co-reference have been studied in [6, 7, 8]. The idea proposed here is in fact complementary with the key idea behind these approaches. In recent research, the term co-reference has become relevant in the field of computational linguistics [9, 10, 11]. The search for co-referent *non-structured* objects, i.e. flat text documents, has important applications in text clustering and information aggregation. In this recent field, semantics is of crucial importance to obtain good cluster accuracy. However, it is emphasized that semantical evaluators as introduced in this paper operate in a framework for *structured* objects, in contradiction to the frameworks for *non-structured* objects in [9, 10, 11]. The remainder of the paper is structured as follows. Section 2 summarizes some basic notations and preliminary knowledge. In Section 3, semantical evaluators are defined and properties are investigated. Section 4 introduces the iterative co-reference

ISBN: 978-989-95079-6-8

detection scheme, which is an unsupervised learning schema in which binary relations used by semantical evaluators are increasingly constructed. Finally, Section 5 summarizes the main contributions of this paper.

2 Preliminaries

Within the context of this paper, a (structured) object is a (structured) description of a real world entity that is decomposable into *attributes* and *co-referent* objects are objects that describe the same entity. The possibilistic model for detection of co-referent objects is based on the concept of *possibilistic truth values* [12, 13, 14]. A possibilistic truth value $\tilde{p}$ is a normalized possibility distribution over the domain $\mathbb{B} = \{T, F\}$ and represents uncertainty about the truth value of a proposition p. Hence $\tilde{p} = \{(T, \mu_{\tilde{p}}(T)), (F, \mu_{\tilde{p}}(F))\}$. A short notation thereof is $\tilde{p} = (\mu_{\tilde{p}}(T), \mu_{\tilde{p}}(F))$. The domain of all possibilistic truth values is denoted $\tilde{\wp}(\mathbb{B})$ which is complete lattice. Total order relations $\tilde{R}$ on this domain are defined as follows, with R a total order relation on the unit interval $[0, 1]$:

$$\tilde{p}_1 \ \tilde{R} \ \tilde{p}_2 \Leftrightarrow \begin{cases} \mu_{\tilde{p}_2}(F) \ R \ \mu_{\tilde{p}_1}(F), & \mu_{\tilde{p}_1}(T) = \mu_{\tilde{p}_2}(T) = 1 \\ \mu_{\tilde{p}_1}(T) \ R \ \mu_{\tilde{p}_2}(T), & \text{otherwise} \end{cases}$$

Possibilistic truth values describe available knowledge (or belief) about a proposition. It is emphasized that they do not describe a state of graded truth. The most important operators for knowledge aggregation in the domain of possibilistic truth values are generalized *conjunction*

$$\tilde{\wedge} : \tilde{\wp}(I)^2 \to \tilde{\wp}(I) : \tilde{p} \tilde{\wedge} \tilde{q} \mapsto \{ \ (T, \min\left(\mu_{\tilde{p}}(T), \mu_{\tilde{q}}(T)\right)), \\ (F, \max(\mu_{\tilde{p}}(F), \mu_{\tilde{q}}(F))) \ \}$$

and *disjunction*

$$\tilde{\vee} : \tilde{\wp}(I)^2 \to \tilde{\wp}(I) : \tilde{p} \tilde{\vee} \tilde{q} \mapsto \{ \ (T, \max(\mu_{\tilde{p}}(T), \mu_{\tilde{q}}(T))), \\ (F, \min(\mu_{\tilde{p}}(F), \mu_{\tilde{q}}(F))) \ \}$$

In the context of co-reference detection, possibilistic truth values are generated by *evaluators*. These are formally defined as follows:

Definition 1 (Evaluator)
Assume a universe U of values describing entities in the real world. An evaluator on U is a commutative function E_U such that:

$$E_U : U^2 \to \tilde{\wp}(\mathbb{B}) : (u, v) \mapsto E_U(u, v) = \tilde{p} \qquad (1)$$

with $p =$"u and v are co-referent", $\mu_{\tilde{p}}(T)$ is the possibility that p is true and $\mu_{\tilde{p}}(F)$ is the possibility that p is false.

An evaluator is *reflexive* if $\forall (u, u') \in U^2 : u = u' \Rightarrow E_U(u, u') = \{(T, 1)\}$ and is *strong reflexive* if $\forall (u, u') \in U^2 : u = u' \Leftrightarrow E_U(u, u') = \{(T, 1)\}$.

This paper deals with evaluators that extract knowledge from binary relations. A binary relation R on U is a subset of the Cartesian product $U \times U$. The elements of R are thus couples of elements from U. $(u, v) \in R$ is also denoted as $u \ R \ v$. An equivalence relation is a binary relation that is *reflexive* ($\forall u \in U : u \ R \ u$), *symmetrical* ($\forall (u, v) \in U^2 : u \ R \ v \Leftrightarrow v \ R \ u$) and *transitive* ($\forall (u, v, w) \in U^3 : (u \ R \ v \wedge v \ R \ w) \Rightarrow u \ R \ w$). A partial order relation is a binary relation that is reflexive, transitive and anti-symmetrical ($\forall (u, v) \in U^2 : (u \ R \ v) \wedge (v \ R \ u) \Rightarrow (u = v)$). A *strict* partial order relation is anti-reflexive ($\forall u \in U : \neg(u \ R \ u)$), anti-symmetric and transitive.

3 Semantical evaluators

3.1 Definitions

Evaluators have been introduced and investigated in [15]. In general, three classes of evaluators are distinguished: syntactical, semantical and hybrid evaluators. *Syntactical* evaluators determine possibility based on syntactical differences and are useful to deal with typographical errors in string data and inaccuracies with numerical data. *Semantical* evaluators are based on *binary relations* between elements of a universe U. The name 'semantical evaluator' reflects that in many cases, such binary relations model a semantical connection between elements of the universe. However, the use of the framework presented here is not restricted to semantical connections only. In fact, binary relations can very well express a syntactical connection, rather than a semantical one, for example the *equality* relation. Nevertheless, the applications of interest assume in most cases relations that are semantical in nature, which explains the name of this type of evaluators. *Hybrid* evaluators are a combination of several evaluators on the same domain. Syntactical evaluators are studied in [15], whereas semantical evaluators are the subject of this paper.

The rationale of a semantical evaluator emerges from the fact that co-reference of objects is nothing more than equality of the entities described by the objects. As such it can be seen that co-reference of objects can in fact be modeled by a symmetrical reflexive binary relation (say R_c for convenience) in the domain $O \times O$ with O the set of objects. Moreover, the domain O can be decomposed into sub-domains U_i which are the domains of attributes a_i. Hence, if a binary relation R_i is defined in $U_i \times U_i$, a key question is then how the information given by relation R_i can be translated into information about R_c. It is thus studied in this paper how a given binary relation R on a universe U can be used to generate the possibility that elements of U belong to co-referent objects. An important assumption hereby made is that R contains *positive* knowledge about co-reference, which means that connection of two elements u and v through a binary relation is regarded as evidence for their co-reference. If elements are not connected through R, it is concluded that u and v are not co-referent. This is formalized as follows.

Definition 2 (Semantical evaluator)
Assume a universe U and a binary relation R. A semantical evaluator $E_{U,R}$ is defined by means of a function f_R called the knowledge transfer function, such that $\forall (u, v) \in U^2 : E_{U,R}(u, v) = f_R(u, v)$ with:

$$f_R : U^2 \to \tilde{\wp}(\mathbb{B}) \qquad (2)$$

such that

$$f_R(u, v) = \begin{cases} \tilde{p}_{u,v} & u \ R \ v \vee v \ R \ u \\ (0, 1) & \neg(u \ R \ v) \wedge \neg(v \ R \ u) \end{cases} \qquad (3)$$

The condition $u \ R \ v \vee v \ R \ u$ in Definition 2 is a consequence of the assumption of commutativity in the definition of evaluators. As an example, let $\mathcal{S}$ be the domain of strings and consider the relation R='city-part-of'. If, for the sake of simplicity, it is assumed that $\tilde{p}_{u,v}$ is independent of u and v and equal to $(1, 0.3)$, then:

$$E_{\mathcal{S},R}(\text{"Manhattan"}, \text{"New York"}) = (1, 0.3)$$
$$E_{\mathcal{S},R}(\text{"New York"}, \text{"Chicago"}) = (0, 1)$$

ISBN: 978-989-95079-6-8

The assignment of possibilities by a semantical evaluator is constrained by the specificity of the relation R. Assume two binary relations R_1 and R_2 on universe U such that $R_1 \subset R_2$, then R_2 applies for more couples of elements than does R_1. In an extreme case, a relation R can apply for each couple of elements, which means that R is an equivalence relation on U with 1 equivalence class. Clearly such a relation offers no information about co-reference at all, which means that:

$$R = U^2 \Rightarrow \forall (u, v) \in U^2 : E_R(u, v) = (1, 1) \quad (4)$$

Thus, if more couples satisfy a relation R, R is less specific and less knowledge about co-reference can be derived from R. Hence, a knowledge transfer function must satisfy:

$$(R_1 \subseteq R_2) \Rightarrow \forall (u, v) \in R_1 : f_{R_1}(u, v) \tilde{\geq} f_{R_2}(u, v) \quad (5)$$

The principle of specificity can also be applied inside a relation R. Consider a binary relation R defined over the universe U. The projection of R over an element $u \in U$ is given by:

$$\text{Proj}_u(R) = \{(x, y) | x \; R \; y \wedge (x = u \vee y = u)\} \quad (6)$$

The knowledge transfer from R to $\tilde{\wp}(\mathbb{B})$ modeled by f_R is *monotone* if and only if for any triple $(u, v, w) \in U^3$ that satisfies $uRw \vee wRu$ and $vRw \vee wRv$, we have that:

$$|\text{Proj}_u(R)| \geq |\text{Proj}_v(R)| \Leftrightarrow f_R(u, w) \tilde{\leq} f_R(v, w) \quad (7)$$

It can be verified that monotone knowledge transfer is obtained if f_R is constructed as follows:

$$\begin{array}{l} u \; R \; v \vee v \; R \; u \Rightarrow \\ f_R(u, v) = \left(1, g\left(\frac{|\text{Proj}_u(R)|}{2|R|-1}, \frac{|\text{Proj}_v(R)|}{2|R|-1}\right)\right) \end{array} \quad (8)$$

with g a monotonic increasing, commutative function such that $g(1, 1) = 1$. Monotone transfer knowledge implies that it is *more* certain that u and v are co-referent if they are connected by R and if there is *less* evidence that u or v are co-referent with other elements.

A common problem with semantical evaluators is that relations are a-priori information. In other words, relations are constructed without observing the actual values on which the evaluator must operate. It is possible that the values on which R is constructed are not *equal* to values of an observed problem, but rather co-referent. This is caused by the fact that values considered for a-priori construction of R are typically standardized, which is not the case for values that are the subject of co-reference detection. For example, with R='city-part-of', it is known that "Manhattan" R "New York". Nevertheless:

$$E_{\mathcal{S},R}(\text{"Manhatan"}, \text{"New York"}) = (0, 1)$$

due to the spelling error in "Manhatan". To deal with such a situation, an evaluator can be used to detect co-references between values from an observed problem and values from the a-priori universe on which R is constructed. Typically, such an evaluator will be *syntactical*. A semantical evaluator is then used in a *chain* of evaluators.

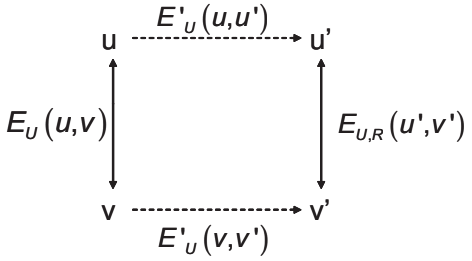


Figure 1: Evaluator chain

Definition 3 (Evaluator chain)
Assume a universe U equipped with a semantical evaluator $E_{U,R}$ and a reflexive evaluator E'_U. An evaluator chain is an evaluator E_U such that:

$$E_U(u, v) = \max_{(u', v')} E'_U(u, u') \tilde{\wedge} E_{U,R}(u', v') \tilde{\wedge} E'_U(v', v) \quad (9)$$

The principle of chaining is schematically depicted in Figure 1. Due to reflexivity of E'_U it can be seen that:

$$(u \; R \; v \vee v \; R \; u) \Rightarrow E_U(u, v) = E'_U(u, v) \quad (10)$$

Moreover, if E'_U is strong reflexive:

$$(u \; R \; v \vee v \; R \; u) \Leftrightarrow E_U(u, v) = E'_U(u, v) \quad (11)$$

In the context of iterative co-reference detection (Section 4), an important concept is that of a hybrid evaluator.

Definition 4 (Hybrid evaluator)
Assume a universe U and a set of evaluators $\mathcal{E} = \{E_{U,1}, ..., E_{U,n}\}$ on U. A hybrid evaluator is an evaluator on U such that:

$$E_U = c(E_{U,1}, ..., E_{U,n}) \quad (12)$$

with c a commutative, associative and monotonic increasing function called the combinatory function.

3.2 Examples

The most obvious example of a semantical evaluator is an evaluator that is based on the equality relation. In many cases, reflexivity of an evaluator is required which means that $\forall (u, v) \in U^2 : u = v \Rightarrow E_U(u, v) = (1, 0)$. Hence, in order to support this constraint, equality of elements must imply full impossibility of non co-reference. As mentioned in Section 3.1, this evaluator can be regarded as syntactical evaluator, because the binary relation is not semantical in nature. A related case is a semantical evaluator where R is relaxed to an *equivalence* relation, which has interesting applications such as detection of synonyms, i.e. words with the same meaning. Another example is the case where two words don't have the same meaning, but they both fully qualify for description of a given entity. This is because the entity can not be described exactly and more descriptions are possible. In that case, words are equivalent with respect to their appropriateness for description of an entity, despite the fact that these words do not have the same meaning. This is the case with the attribute 'type' of the 'restaurant' dataset, where for example 'steakhouse' and 'american' are possible (and thus equivalent) descriptions of the type of restaurant.

ISBN: 978-989-95079-6-8

Assume a universe U and assume a partition of U into n disjunct subsets U_i. Consider $n-1$ relations R_i such that each R_i is a surjective function between U_i and U_{i+1}. The set of relations $\{R_i\}_{i=1,2,...,n-1}$ is called a *hierarchy* on U. It can be seen that, by definition, each R_i from a hierarchy is anti-reflexive and anti-symmetrical. Due to the fact that $\forall i \in \{1,...,n-1\} : im(R_i) = dom(R_{i+1})$, it is possible to compose relations R_i and R_{i+1}.

$$\forall (u,w) \in U^2 : u\ R_1R_2\ w \Leftrightarrow (\exists v \in U : u\ R_1\ v \wedge v\ R_2 w) \tag{13}$$

Hence, $R_1R_2...R_n$ denotes the n-ary composition of n relations and thus, elements of non subsequent partitions are related through composition. For each hierarchy $\{R_i\}_{i=1,2,...,n-1}$, consider the set of relations $\{R_i^*\}$ such that:

$$u\ R_i^*\ v \Leftrightarrow \exists k \leq l \leq i : u\ R_k...R_l v \tag{14}$$

By definition:

$$\forall i,j \leq n : i < j \Rightarrow R_i^* \subset R_j^* \tag{15}$$

It is then possible to construct $n-1$ semantical evaluators with knowledge transfer function $f_{R_i^*}$. Due to the consistency constraint:

$$\forall i,j \leq n : i < j \Rightarrow f_{R_i^*} \tilde{\geq} f_{R_j^*} \tag{16}$$

It can be seen that composing all relations, leads to a strict partial order relation. Hence, each partial order relation can be decomposed into relations with higher granularity and thus higher belief of co-reference. This principle can be applied on the 'city' attribute of the restaurant dataset, because the values are connected by a strict partial order relation.

4 Iterative co-reference detection

4.1 Approach

In this Section, attention is given to the case where R is not a-priori given. This is the case when dealing with subjective information, for example the 'type' attribute of the restaurant dataset. There is no a-priori information about which types can be used to denote the same restaurant. In that case a semantical evaluator can be (partially) constructed based on a training set. However, in what follows a method is proposed to construct a semantical evaluator in cases where no training set is given. Assume objects consisting of n attributes a_i. For each attribute a_i, an evaluator $E_{dom(a_i)} = E_i$ must be constructed. The set of attributes can thus be partitioned into attributes that require a syntactical evaluator (A_{syn}) and attributes that require a semantical evaluator (A_{sem}), such that $A = A_{syn} \cup A_{sem}$ and $A_{syn} \cap A_{sem} = \emptyset$. For the 'restaurant' dataset, $A_{syn} = \{$'name', 'street'$\}$ and $A_{sem} = \{$'type', 'city'$\}$.

The key idea of iterative co-reference detection, is to find the co-referent objects in m iterations. Semantical evaluators for which no binary relation R is available are gradually constructed throughout the different iterations. An evaluator is called *unavailable* if it is a semantical evaluator for which the binary relation R is unknown. In each iteration j, a subset of the available evaluators is selected. With these evaluators, a candidate set C_j of co-referent object couples is generated. For each attribute a_i that requires a semantical evaluator, a binary relation R_j is constructed by projecting each object couple from C_j over the attribute a_i. In other words, if the object couple (o_1, o_2) is contained in the candidate set C_j and attribute a_i requires a semantical evaluator, then the couple $(\text{Proj}_{a_i}(o_1), \text{Proj}_{a_i}(o_2))$ is considered as a couple of co-referent values for attribute a_i. As a consequence, this couple of values is added to the binary relation R_j. For each attribute a_i that requires semantical evaluation, a novel relation is R_j is constructed in each iteration j. The reason to do this is explained next. Typically, the generation of the first candidate set is based on very stringent criteria. If this is not the case, many false positives are included in the candidate set, which implies that constructed binary relations contain a large amount of couples. As explained before, a binary relation with large cardinality implies inferred knowledge with much uncertainty (Eq. 4, Eq. 5). Throughout sequential iterations, the candidate set is expanded by (i) relaxation of syntactical evaluators and (ii) additional knowledge of semantical evaluators. As a consequence, there is a high probability that for $k \leq l$, $C_k \subset C_l$. If this is true, relations constructed in iteration k are subsets of the corresponding relations in iteration l. Due to the consistency constraint (Eq. 5), relations constructed in iteration l imply more uncertainty. Therefor, in order to preserve the level of (un)certainty in relations constructed in early iterations, each iteration constructs a new relation. Hence, for each attribute a_i that requires a semantical evaluator, m binary relations are constructed. Each of these relations $R_{i,j}$ yields a semantical evaluator $E_{i,R_{i,j}}$. The constructed evaluator for attribute a_i is a hybrid evaluator (Eq. 12) with $\tilde{\vee}$ as combinatory function:

$$E_i(u,v) = E_{i,R_{i,1}}(u,v)\tilde{\vee}...\tilde{\vee}E_{i,R_{i,m}}(u,v) \tag{17}$$

Hence, the constructed evaluator tries to find the most stringent binary relation by which the values are connected as this relation yields the most belief in co-reference. After iteration m, the last candidate set C_m of co-referent couples is generated which gives us the possible co-referent objects. It is noted that iterative co-reference detection is a stand-alone comparison process: it produces a set of possible co-referent couples. However, it can also be used to construct evaluators and their corresponding binary relations. In other words, the proposed method can be used to find co-referent objects, but it's purpose

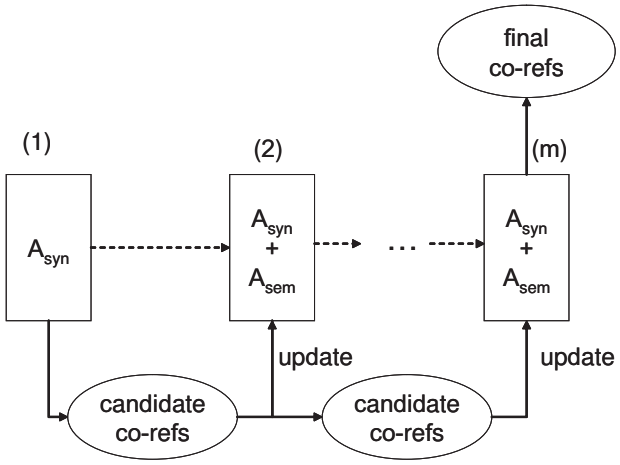


Figure 2: Iterative learning of semantical evaluators

ISBN: 978-989-95079-6-8

Table 2: Statistics of restaurant test

'type' evaluator	Co-refs	Non Co-refs
Equality	22	5018
Equality + R	100	27811

can also be limited to construction of binary relations, which makes it compatible with other (possibilistic) frameworks for co-reference detection. Figure 2 shows a schematical representation of iterative co-reference detection.

4.2 Example

The idea of iterative co-reference detection is illustrated on the 'restaurant' dataset. The focuss in this example lies on illustrating the construction of binary relations. The dataset consists of two lists ($|L_1| = 533$ and $|L_2| = 331$). Among the 176423 couples of restaurants, 112 couples are co-referent. For convenience, only two attributes are initially considered: 'name' and 'type'. It is assumed that a syntactical evaluator is used for 'name' and a semantical evaluator must be constructed for 'type'. For the first iteration, consider an evaluator for attribute 'name' based on equality (i.e. values are co-referent if they are equal). Among all possible couples, 83 couples have an equal name, of which 82 are identified as co-referent, based on ground truth (i.e. equal oid's). Hence, candidate set C_1 contains 83 couples of restaurants. Projecting this candidate set over the 'type' attribute and filtering identical couples out, yields 40 couples of possible co-referent restaurant types. These couples determine the binary relation R_1. Table 2 learns how R_1 allows for assignment of belief to 100 co-referent couples (89.3% of all co-referent couples). However, 27811 non-coreferent couples are also assigned belief (15.8% of all non co-referent couples). This is partially due to the type of data. As can be seen, 5018 non-coreferent couples have equal values for type, which is an indication that the 'type' attribute is prone to many false positives. It must be remembered that the goal is to construct an evaluator that is used in a larger object comparison system. Hence, the contribution of semantical evaluators lies in the assignment of belief to co-referent values, thereby trying to minimize the belief attached to non co-referent values. As can be seen in Table 2, the fraction of non-coreferent values that are assigned belief can still be large in absolute numbers. However, this fraction can be filtered out using an intelligent aggregation function to combine the belief assigned by the different evaluators. The constructed relation for 'type' assigns belief to 89.3% of the co-referent couples, after one iteration. Assume in a second iteration, that the 'street' attribute is involved, which requires a syntactical evaluator. For the sake of simplicity, assume an evaluator for 'street' based on equality. If the evaluations for 'name' and 'street' are combined through disjunction, i.e. two restaurants are co-referent if their names are equal *or* their streets are equal, then the candidate set C_2 contains 106 object couples of which 101 are actually co-referent. Suppose that iteration 2 would be the last iteration, then the model can be refined by adding semantical evaluation of the 'type' attribute, i.e. two restaurants are co-referent if their types are connected by R_1 *and* (names are equal *or* their streets are equal). Using this model results in a candidate set of 95 couples of which 94 are co-referent, which means that 83.9% of all co-referent couples are detected with a precision of 98.95%. If iteration 2 is not the last iteration, then C_2 can be used to construct a second binary relation R_2. Note that by construction $R_1 \subset R_2$.

An advantage of iterative co-reference detection is that the binary relations R are dynamically adapted throughout the different iterations. If an approach with a training set is used, the relation R can be partially constructed only on the observations in the training set. With iterative co-reference detection, binary relations are constructed by generating a candidate set from *all* known observations. Hence, the relations constructed in an iterative co-reference detection scheme will be more complete.

5 Conclusion

In the possibilistic framework for detection of complex co-referent objects, a central concept is that of an evaluator, which is a function that estimates belief/uncertainty about co-reference of (sub)-objects. This paper introduces a formal framework for *semantical* evaluators E_U that model the transfer of knowledge from binary relations on the universe, to knowledge about co-reference of elements. This transfer is modeled by means of a knowledge transfer function f_R. It is shown that, depending on the application, such a function must satisfy consistency constraints. It is also shown how semantical evaluators can be used in an evaluator chain and how to combine evaluators in a hybrid evaluator. Finally, when relations on U are unknown, it is investigated how to (partially) construct such relations without use of a pre-labeled training set. This construction process is called iterative co-reference detection.

References

[1] Sheila Tejada, Craig Knoblock, and Steven Minton. Learning object identification rules for information integration. *Information Systems*, 26(8):607–633, 2001.

[2] Ross Quillian. *Semantic Information Processing*, chapter Semantic memory. MIT Press, 1968.

[3] Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 448–453, 1995.

[4] Jay Jiang and David Conrath. Semantical similarity based on corpus statistics and lexical taxonomy. In *Proceedings of International Conference Research on Computational Linguistics*, pages 15–33, 1997.

[5] Yuhua Li. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):871–882, 2003.

[6] Matthew Jaro. Advances in record linking methodology as applied to the 1985 census of tampa florida. *Journal of the American Statistical Society*, 84(406):414–420, 1989.

[7] William Winkler. Improved decision rules in the fellegi-sunter model of record linkage. Technical Report RR93/12, Statistical Research Report Series, 1993.

[8] William Winkler. Methods for record linkage and bayesian networks. Technical Report RRS2002/05, Statistical Research Report Series, 2002.

[9] Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544, 2001.

ISBN: 978-989-95079-6-8

[10] Vincent Ng and Claire Cardie. Improving machine learning approaches to coreference resolution. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 104–111, Philadelphia, 2002.

[11] Xiaoqiang Luo and Imed Zitouni. Multi-lingual coreference resolution with syntactic features. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 660 – 667, 2005.

[12] Lotfi Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

[13] Henri Prade. Possibility sets, fuzzy sets and their relation to lukasiewicz logic. In *Proceedings of the International Symposium on Multiple-Valued Logic*, pages 223–227, 1982.

[14] Gert De Cooman. Towards a possibilistic logic. in: Fuzzy set theory and advanced mathematical applications, edited by da ruan, kluwer academic, pp. 89–133, boston., 1995.

[15] Antoon Bronselaer, Axel Hallez, and Guy De Tré. Evaluation in the possibilistic framework for object matching. In *Proceedings of the IPMU 08 Congres*, pages 1701–1708, Malaga, 2008.

ISBN: 978-989-95079-6-8

A Hybrid Meta-Heuristic to Solve the Portfolio Selection Problem

José M. Cadenas[1] Juan V. Carrillo[1] M. Carmen Garrido[1]
Carlos Ivorra[2] Teresa Lamata[3] Vicente Liern[2]

1.Dep. Ingeniería de la Información y las Comunicaciones. Universidad de Murcia.
2.Dep. Matemàtiques Economia i Empresa, Universidad de Valencia.
3.Dep. Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada.
Emails: jcadenas@um.es, jvcarrillo@um.es, carmengarrido@um.es,
carlos.ivorra@uv.es, mtl@decsai.ugr.es, vicente.liern@uv.es

Abstract— *We present a fuzzy version of the efficient portfolio selection problem which adds to the Markowitz's classical model the vagueness of the investor's preferences about the assumed risk. This is done by adapting some techniques previously applied by the authors in other logistic problems. We provide several procedures to solve it: an exact one and a hybrid meta-heuristic procedure more adequate for medium and large-sized problems. To show the usefullness of the model and the algorithms, we provide an example based on data from the Spanish index IBEX35.*

Keywords— Fuzzy programming, Heuristic strategies, Portfolio selection.

1 Introduction

The modern portfolio selection problem is a classical model for determining the optimal composition of a portfolio according to the preferences of an investor. The model is credited to Markowitz [17, 18], who is rightfully regarded as the founder of modern portfolio theory [6].

In a first approach, given n securities S_1, $S_2, \cdots, S_n$ in which we can invest non-negative quantities $x_1, x_2, \cdots, x_n$, a portfolio consists of a subset of securities in which all the capital should be invested satisfying investor's preferences. These preferences must take into account the risk that the investor is willing to assume, the expected return and, to a lesser extent, some diversification criteria.

One possibility is taking the return as objective function and including the risk as a constraint, i.e.

$$\begin{array}{ll} \text{Max.} & \text{Portfolio return} \\ \text{s.t.} & \text{Risk of portfolio} \leq R \\ & \sum_{i,=1}^{n} x_i = 1, \\ & l_i \leq x_i \leq u_i, \qquad i = 1, \ldots, n. \\ & l_i \geq 0 \qquad\qquad i = 1, \ldots, n. \end{array} \tag{1}$$

where R is the maximum risk accepted by the investor, and the last constraints express that all the capital must be invested as well as diversification conditions.

However, we could also consider the dual version (in an economic sense) in which the risk is minimized taking the expected return as a constraint.

Obviously, the portfolio selection, like most financial problems, is related with uncertainty because it consists in taking a decision about future events. Therefore, we do not have at our disposal more than historical data which usually are managed with statistical methods [18, 6]. Moreover, it is not easy to model the investor's preferences. After the seminal work by Markowitz, attention has been given in the study of alternative models [13, 14]. Most of these models are based on probability distributions, which are used to characterize risk and return. For instance, in the Markowitz model, variance and mean were generally deemed satisfactory measures of risk and return, respectively [18, 21, 7, 12]. However, according to the chosen modellization of the expected risk and return, different models coexist to select the best portfolio. Some of them propose to describe the risk by means of mean-absolute deviation ([11], [24]), other authors propose linear programming [22] or multiobjective [15] models.

Another way of dealing with uncertainty is working with models based on soft computing. This can be done by means of vague goals for the expected return rate and risk [23, 16], or by using possibility distributions to model the returns. This fact permits the incorporation of expert knowledge by means of a possibility grade, to reflect the degree of similarity between the future state of stock markets and the state of previous periods.

In this paper, we present several ways of dealing with the uncertainty associated to the portfolio selection problem. Namely, we are going to show that the techniques used by the authors in logistic problems[2, 3, 4] can be generalized to deal with the risk and return fuzziness in the portfolio selection problem. These techniques allows us to include in our fuzzy model the vagueness related with the fact that the investor fixes arbitrarily the risk that he/she will assume. The uncertainty concerning to the valuation of risk and returns is modellized by means of the usual variance and mean.

To solve our model, meta-heuristic techniques are often needed because of the size of the problem or because a quick solution is required in order to reflect real-time cases. Therefore, we propose a hybrid meta-heuristic procedure.

The paper is organized as follows: The next section introduces the framework for portfolio selection used in this paper. Section 3 examines the portfolio selection with flexible constraints, incorporating this flexibility in a fuzzy model. In section 4, we present a hybrid meta-heuristic to solve the portfolio selection problem with flexible constraints. Finally, Section 5 outlines the most important conclusions.

ISBN: 978-989-95079-6-8

2 Portfolio selection

We could consider the problem of an investor who intends to invest some money in securities in such a way that the rate of return is maximized and a given level of risk cannot be surpassed. If we have n securities, r_i is the return of the i-th security and x_i is the proportion of total investment funds devoted to the i-th security, then we have the constraint

$$\sum_{i=1}^{n} x_i = 1.$$

The risk of investment can be measured by the variance $x^t S x$, where $S = [s_{ij}]$ is the variance-covariance matrix. If R is the maximum risk accepted by the investor, we have the constraint

$$\sum_{i,j=1}^{n} s_{ij} x_i x_j \leq R.$$

Then, one of the simplest models for the portfolio selection problem is:

$$\begin{array}{rl} \text{Max.} & \displaystyle\sum_{i=1}^{n} r_i x_i \\ \text{s.t.} & \displaystyle\sum_{i=1}^{n} x_i = 1, \\ & \displaystyle\sum_{i,j=1}^{n} s_{ij} x_i x_j \leq R, \\ & x_i \geq 0, \;\; i = 1, \ldots, n. \end{array} \tag{2}$$

However, choosing the best investment options for a portfolio is a difficult task due to economic environment uncertainty and the problem of suitably reflecting decision maker desires in the model. Both stochastic and fuzzy programming provide different ways of handling the first kind of uncertainty [9]. This paper presents a fuzzy approach for the second kind, in which the portfolio problem includes the subjective criteria of the decision maker to determine the level of risk that he or she is able to support and the level of satisfaction to be assigned to a possible increase in return.

3 Flexible Portfolio selection

The portfolio selection problem has two data concerning decision maker preferences, namely the capital to be invested and the risk to be assumed. The investor can be assumed to know with certainty the capital that he/she would like to consider, and in fact, in the model this quantity has been normalized to the unit. However, determining the risk to be assumed could be more flexible. As a result, it is worth incorporating this flexibility in a fuzzy model.

The main idea is to consider partially feasible solutions involving slightly greater risk than that fixed by the decision maker, and to study the possibilities that they offer in order to improve the expected return. When compared with the logistic models (actually with the p-median case [2, 3, 4]), this problem happens to be more complicated, because the p-median problem is linear whereas the risk constraint in the portfolio model is quadratic. Moreover, in the p-median case, a small reduction in covered demand affected optimal cost in a simple linear way, whereas the way in which the maximum expected return depends on the accepted risk is rather more complicated.

A fuzzy set $\tilde{S}$ of partially feasible solutions is defined so that portfolio selection belongs to $\tilde{S}$ with a degree of membership that depends on how much it exceeds the risk R_0 fixed by the investor. On the other hand, a second fuzzy set $\tilde{G}$ is defined, whose membership function reflects the improvement of the return provided by a partially feasible solution with respect to the optimal crisp return z^*. In practice, we consider piecewise linear membership functions

$$\mu_{\tilde{S}}(x) = \begin{cases} 1 & \text{if } r \leq R_0, \\ 1 - \frac{r - R_0}{p_f} & \text{if } R_0 < r < R_0 + p_f, \\ 0 & \text{if } r \geq R_0 + p_f, \end{cases}$$

$$\mu_{\tilde{G}}(x) = \begin{cases} 0 & \text{if } z \leq z^*, \\ \frac{z - z^*}{p_g} & \text{if } z^* < z < z^* + p_g, \\ 1 & \text{if } z \geq z^* + p_g, \end{cases}$$

where r and z are the risk and the return provided by the portfolio x (which is assumed to satisfy the constraints of (2), except the second one); the parameter p_f is the maximum increment in the risk that the decision maker can accept, and p_g is the increment of the return that the decision maker would consider completely satisfactory. From this, we can define a global degree of satisfaction

$$\lambda(x) = \min\{\mu_{\tilde{G}}(x), \mu_{\tilde{S}}(x)\},$$

which is the membership degree to the fuzzy intersection of $\tilde{S} \cap \tilde{G}$. The fuzzy portfolio model becomes

$$\begin{array}{rl} \text{Max.} & \lambda(x) \\ \text{s.t.} & x \in \tilde{S}. \end{array} \tag{3}$$

In order to solve it, the optimal solution of (2) will be calculated for each risk level R. Therefore, we solve explicitly the Kuhn-Tucker conditions for the problem. To carry out the computations in a generic framework, we start making a change of variables to diagonalize the risk matrix. As we are interested in small variations of R, the variables x_i that are zero in the optimal crisp portfolio can be removed and hence assume that non-negative conditions are not active. We can also assume that the risk constraint is active.

Standard linear algebra theory ensures us that we can decompose

$$S = A^t D A, \tag{4}$$

where the matrix D is diagonal and A is regular. Then the change of variables $y = Ax$ transforms the problem to

$$\begin{array}{rl} \text{Max.} & \displaystyle\sum_{i=1}^{n} r'_i y_i \\ \text{s.t.} & \displaystyle\sum_{i=1}^{n} b_i y_i = 1, \\ & \displaystyle\sum_{i=1}^{n} d_i y_i^2 = R, \\ & A^{-1} y \geq 0, \end{array} \tag{5}$$

where $r' = r^t A^{-1}$, $b = (1, \ldots, 1) A^{-1}$.

ISBN: 978-989-95079-6-8

As nonnegative conditions are inactive, the optimal solution must satisfy

$$\sum_{i=1}^{n} b_i y_i = 1, \quad \sum_{i=1}^{n} d_i y_i^2 = R, \quad r'_i - b_i\lambda - 2d_i y_i \mu = 0,$$

where λ and μ are the Lagrange multipliers of the capital and risk constraints respectively. Hence:

$$y_i = \frac{r'_i - b_i\lambda}{2d_i\mu},$$

and, from the first constraint, we get:

$$\sum_{i=1}^{n} \frac{b_i r'_i}{2d_i} - \sum_{i=1}^{n} \frac{b_i^2}{2d_i}\lambda = \mu.$$

Call

$$K = \sum_{i=1}^{n} \frac{b_i r'_i}{2d_i}, \qquad L = \sum_{i=1}^{n} \frac{b_i^2}{2d_i},$$

so that $\mu = K - L\lambda$. Hence:

$$y_i(\lambda) = \frac{r'_i - b_i\lambda}{2d_i(K - L\lambda)}.$$

The second constraint gives us:

$$\sum_{i=1}^{n} \frac{(r'_i - b_i\lambda)^2}{4d_i} = R(K - L\lambda)^2$$

or

$$\sum_{i=1}^{n} \frac{r'^2_i}{4d_i} - \sum_{i=1}^{n} \frac{r'_i b_i}{2d_i}\lambda + \sum_{i=1}^{n} \frac{b_i^2}{4d_i}\lambda^2 = R(K^2 - 2KL\lambda + L^2\lambda^2),$$

$$\sum_{i=1}^{n} \frac{r'^2_i}{4d_i} - K\lambda + \frac{L}{2}\lambda^2 = R(K^2 - 2KL\lambda + L^2\lambda^2).$$

If

$$M = \sum_{i=1}^{n} \frac{r'^2_i}{4d_i} \tag{6}$$

we get

$$(RL^2 - L/2)\lambda^2 + (K - 2KLR)\lambda + RK^2 - M = 0.$$

This is a second degree equation on λ with discriminant

$$\Delta(R) = K^2 - 2ML + (4ML^2 - 2K^2L)R.$$

Hence,

$$\lambda(R) = \frac{2KLR - K \pm \sqrt{\Delta(R)}}{2L^2R - L},$$

and so

$$\mu(R) = K - L\lambda(R) = \frac{\pm\sqrt{\Delta(R)}}{1 - 2LR}.$$

The sign must be chosen so that $\mu(R) \geq 0$ (the Kuhn-Tucker sign condition). Once the right sign is chosen, this function $\lambda(R)$ allows us to calculate $y_i(R)$ and, from this, we get the optimal portfolio $x_i(R)$ for each R such that $x_i(R) > 0$. This reduces the range of R to a certain interval that can be computed. When we reach one end point of this interval, we need to start again with the problem corresponding to the new set of positive variables.

We can also compute the return

$$F(R) = \sum_{i=1}^{n} r'_i y_i(R).$$

Since

$$y_i(\lambda) = \frac{r'_i - b_i\lambda}{2d_i\mu} = \mp\frac{2LR-1}{2d_i\sqrt{\Delta(R)}}\left(r'_i - b_i\frac{2KLR - K \pm \sqrt{\Delta(R)}}{2L^2R - L}\right), \quad 1 \leq i \leq n,$$

we have

$$F(R) = \mp\frac{2LR-1}{2\sqrt{\Delta(R)}}\sum_{i=1}^{n}\left(\frac{r'^2_i}{d_i} - \frac{r'_i b_i}{d_i}\frac{2KLR - K \pm \sqrt{\Delta(R)}}{2L^2R - L}\right).$$

This expression allows us to calculate the degree of improvement of the goal $\mu_g(R)$ of the best portfolio with risk R, whereas its degree of feasibility $\mu_f(R)$ is trivially computed.

Now we can determine the risk R^* such that $\mu_f(R^*) = \mu_g(R^*)$, which is easily shown to be the risk of the portfolio maximizing λ. The portfolio $x(R^*)$ corresponding to $y(R^*)$ by the change of variables is the optimal solution of (3).

4 A hybrid meta-heuristic to solve the Portfolio Selection problem

When the size of the problem increases, traditional methods become useless as they would need too much time, due to the combinatorial explosion in the solution space. Genetic Algorithms [19, 5] are meta-heuristic methods that have already shown their kindness to solve optimization problems.

In this section we describe a hybrid meta-heuristic for the portfolio selection problem with flexible constraints, called GAFUZ-PF (Genetic algorithm + simulated Annealing + FUZzy PortFolio). This meta-heuristic uses a hybrid scheme, mixing the ideas of Simulated Annealing technique with the classic Genetic Algorithm. Note that we need the optimal crisp return z^* of the portfolio selection problem (2) because $\mu_{\tilde{G}}(x)$ depends on it. Hence, solving the crisp problem must be considered as a part of the process of solving the fuzzy problem (this means to solve two NP-hard problems). In order to realize this idea, we use three different populations on three stages of the genetic algorithm. First of all, it uses a population to achieve feasible solutions to the problem (with a risk less than or equal to R fixed) which will be part of the population of the algorithm. Once feasible solutions are found, in the second stage, the algorithm uses other population that evolves searching the best return. The third stage uses another population, which looks for the best fitness $\lambda(x)$.

Simulated Annealing [1, 10] is a meta-heuristic used for optimization problems. This heuristic replaces the current solution by a random "nearby" solution, chosen with a probability that depends on the difference between the corresponding

ISBN: 978-989-95079-6-8

function values and a global parameter T (called the temperature), that is gradually decreased during the process. The use of this idea with Genetic Algorithm is suggested by Mitchell [19], using it with the individual selection function. We apply Simulated Annealing to mutation function. Having a mutation function with a big mutation rate helps us to keep diversity in the population, but it introduces distortion on the population when the algorithm approaches to the optimum. So, it seems a good idea to use Simulated Annealing with a temperature depending on the number of generations and the fitness of the solution in order to make possible a big mutation rate at the beginning, and reduce it while increasing the number of generations and the fitness. Now we specify our options and we give a general outline of our approach. Figure 1 shows the scheme of `GAFUZ-PF` hybrid meta-heuristic.

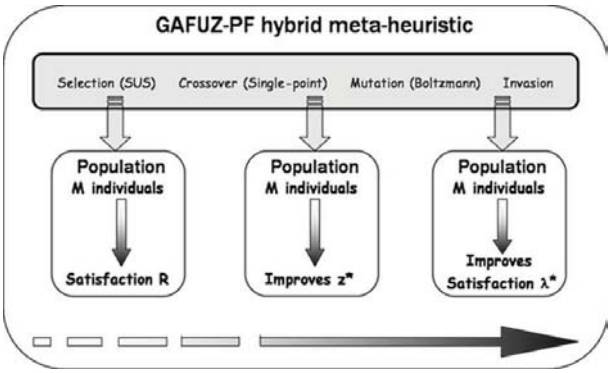


Figure 1: Scheme of hybrid meta-heuristic

4.1 Encoding, Fitness function and Population size

Each element of the population of the algorithm represents a solution to the described portfolio problem. The population encoding is made having an array of n-elements, where each gene represents the invest on one asset. The size of the population is fixed to 50 individuals, as it is known to be the best average size to avoid a too slow algorithm [19]. Furthermore, we use three different populations against time: the first one evolves searching for individuals to satisfy the second constraint of the problem 2. When it founds a suitable solution, it changes to the second stage, where it improves the return of the portfolio, taking account of the flexible constraint of R. Once the algorithm gets a state with $\lambda(x) > 0$, it changes the population and evolves taking into account p_f and p_g. The fitness function is the objective function of the problem (3).

The algorithm scheme is the classical one, with selection, crossover, mutation, invasion and having the number of generations as stop criteria.

4.2 Selecting the individuals

In order to select individuals for the crossover operator, we use the Stochastical Universal Sampling (SUS), as it has been shown to be better than Roulette Wheel [5]. Based on their fitness, SUS places the individuals on a Roulete and spuns it once with n selectors equally separated to select n individuals.

4.3 Crossover: generating new individuals

The crossover operator is the single-point crossover [19] adapted to the problem constraints: there is a fixed part from one parent that goes directly to the son; the genes from the second part (from the second parent) are added to the son (choosing them randomly) while the sum of its genes is ≤ 1.

4.4 Mutation

When we talk about mutation, there are some parameters used. The mutation rate (MUR) refers to the percentage of population individuals that will be mutated. On the other hand, we have the number of genes that will change (NR). As a mutation scheme we use Boltzmann mutation which is based on Simulated Annealing [5], with the following annealing function:

$$r = 1 - e^{-1/(n*max(\lambda(x)))}$$

It defines the mutation rate across the algorithm, reducing it as the algorithm evolves in order not to disturb the good solutions achieved when the algorithm is finishing. This rate specifies the percentage of individuals that would be mutated, since the number of mutations per individual is 1, as it is known to be the best choice [20]. A mutation consists in substract a random value from a gene and add this amount to other gene, both of them choose randomly.

4.5 Invasion

Finally, invasion is another genetic operator that helps to keep diversity in the population. It introduces a number of new randomly generated individuals, which replace some individuals at random in the population (except the best one). This helps to introduce new genes in the population, like mutation does. The invasion rate (IR) selected is 1% .

4.6 Stop criteria

The detention criterion is the number of generations (NG).

4.7 General outline of the `GAFUZ-PF` meta-heuristic

Now we have shown the configuration of the algorithm, we will see its general outline in algorithm 1.

4.8 Experiments

To test `GAFUZ-PF` hybrid meta-heuristic we have used various test problems. The PC used for the executions has the following features: Intel Pentium IV 3.00 GHz 2048 MB RAM.

• We have considered the returns on 20 assets from the Spanish index IBEX35. The set of assets included in the experiment represents Acesa (ACE), Arcelor (ACR), ACS, Altadis (ALT), BBVA, Bankinter (BKT), Dragados (DRC), Endesa (ELE), FCC, Iberdrola (IBE), Metrovacesa (MVC), NH Hoteles (NHH), Banco Popular (POP), Repsol (REP), SCH (SAN), Telefónica (TEF), Unión Fenosa (UNF), Vallermoso (VAL), Acerinox (ACX), Acciona (ANA) data respectively. We have considered the observations of the Wednesday prices as an estimate of the weekly prices. Hence, the return on the j- th asset during the k-th week is defined as $r_{kj} = (p_{(k+1)} - p_{kj})/p_{kj}$, where p_{kj} is the price of the j-th asset on the Wednesday of the k-th week. The used data base covers the period from January 1998 to March 2003. Table 1 contains the obtained results for two different risk levels R

ISBN: 978-989-95079-6-8

Algorithm 1 - GAFUZ-PF: Hybrid meta-heuristic for the portfolio selection

GAFUZ-PF( )
Require: r_i, s_{ij}, R, p_f, p_g
Ensure: z^*, $\overline{x}^*$, R^*, $\lambda(x)$, z, $\overline{x}$, R
begin
Generate initial population with 50 individuals at random.
Calculate the fitness of each individual.
while Number of generations $< NG$ **do**
*S*ELECT 50 individuals
if $\nexists$ individual | risk of the individual $\leq R$ **then**
select using risk
else
if the sum of individual fitness is = 0 **then**
select using z^*
else
select using fitness
end if
end if
*C*ROSS the 50 individuals to obtain 50 new individuals
Replace 49 individuals of the population with the new individuals, keeping the best (elitism)
MUR=Boltzmann(max.($\lambda(x)$), generation_number)
*M*UTATE $M \times MUR$ individuals, changing NR genes
Replace $50 \times IR$ individuals with random generated individuals
Calculate the new individuals' fitness, return and risk
end while
end

with the same parameters p_f and p_g. The first column contains the assets apearing either in the crisp or the fuzzy solution. The other four columns contain the crisp and the fuzzy solution in both cases. The first one has a medium satisfaction level ($\lambda = 0.45$) whereas the second one has a very low one ($\lambda = 0.14$). This means that, in the first case, the fuzzy solution is an alternative solution that the investor should take into account, whereas, in the second one, the crisp solution cannot be substantially improved. The last two rows of the table contain the crisp and fuzzy returns and risks for the two cases. These risks measure also the efficiency of our metaheuristic: the optimal crisp solution should have a risk equal to the given one, and the near-optimal crisp solutions given by our metaheuristic have a risk level differing by a 0.54% and 0.19%, respectively, from the optimal one. The time used for the algorithm to obtain the solution to this problem is 1.61 seconds on average.

5 Conclusions

In this paper we have presented the problem of portfolio selection, which is a difficult problem to solve due to economic environment uncertainty and the problem of suitably reflecting decision maker desires in the model. We present an approach where the problem of fuzzy portfolio includes the subjective criteria of the decision maker to determine the level of risk that he or she is able to support and the level of satisfaction to be assigned to a possible increase in return.

We have proposed an exact method to solve the fuzzy model of the portfolio problem with the main idea of finding partially

	$R = 0.000494$		$R = 0.000839$	
	$p_f = 0.0002,\ p_g = 0.002$		$p_f = 0.0002,\ p_g = 0.002$	
	Crisp	Fuzzy	Crisp	Fuzzy
ACE	0.316	0.200	0	0
ACR	0.011	0	0	0
ALT	0.168	0.199	0.207	0.162
DRC	0.019	0.205	0.353	0.461
IBE	0.165	0.014	0	0
MVC	0.050	0.102	0.157	0.098
NHH	0.011	0.033	0.056	0.065
POP	0.62	0	0	0
REP	0.031	0	0.001	0
UNF	0	0.077	0.0395	0.052
VAL	0.036	0	0	0
ACX	0.053	0.069	0.051	0.017
ANA	0.078	0.100	0.139	0.145
λ	0	0.454078	0	0.145928
Return	0.00202704	0.0029352	0.00353188	0.00382374
Risk	0.000491306	0.0006031840	0.00083643	0.0010098

Table 1: Obtained solutions by GAFUZ-PF

feasible solutions involving slightly greater risk than that fixed by the decision maker, and to study the possibilities that they offer in order to improve the expected return.

We have also proposed a hybrid meta-heuristic to solve the fuzzy model for problems with medium or big size where the traditional methods become useless as they would need too much time due to the combinatorial explosion in the solution space. The proposed meta-heuristic uses a hybrid scheme, which combines ideas from the Simulated Annealing technique and genetic algorithms. To test the proposed meta-heuristic we have used several test problems based on dates of IBEX35, the best-known index of Spanish Stock Markets. The results allow us to verify that with the small problems the solutions are very similar to those obtained with the exact model. Moreover, our proposal allows us to work with big problems (large time series, intra-day data, etc.) and the time taken to get the results is about 1.6 seconds, achieving the goal we wanted to get with that meta-heuristic.

Acknowledgements

This work has been partially supported by the Ministerio de Ciencia e Innovación of Spain, TIN2008-06872-C04-02, TIN2008-06872-C04-03, TIN2008-06872-C04-04.

References

[1] E. Aarts and J. Korst, "Simulated annealing and Boltzmann machines: a stichastic approach combinatorial to optimization and neural computing," John Wiley and Sons, New York, 1980.

[2] M. J. Canós, C. Ivorra, and V. Liern, *The fuzzy p-median problem: A global analysis of the solutions*, European Journal of Operational Research, **130** (2001), 430–436.

[3] M. J. Canós, C. Ivorra, and V. Liern, *Marginal analysis for the fuzzy p-median problem*, European Journal of Operational Research, **130** (2008), 430–436.

ISBN: 978-989-95079-6-8

[4] J.M. Cadenas, M.J. Canós, M.C. Garrido, C. Ivorra and V. Liern, *Soft-computing based heuristics for location on networks: The p-median problem*, Applied Soft Computing, (2008), doi:10.1016/j.asoc.2008.03.015

[5] J.M. Cadenas, M.J. Canós, J.V. Carrillo, M.C. Garrido, C. Ivorra and V. Liern *A hybrid algorithm for the fuzzy p-median problem.* IADIS International Conference Intelligent Systems and Agents 2008, Amsterdam (The Netherlands).

[6] E.F. Fama, "Foundations of Finance," Basil Blackwell, Oxford, 1977.

[7] A. Fernández-Carazo, *Un estudio holístico de la selección y planificación temporal de carteras de proyectos*, Rect@, **9** (2008), 5–44.

[8] P. Guo and H. Tanaka, *Decision analysis based on fused double exponential possibility distributions*, European Journal of Operational Research, **148** (2003), 467–479.

[9] M. Inuiguchi and J. Ramík, *Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem*, Fuzzy Sets and Systems, **111** (2000), 3–28.

[10] S. Kirkpatrick, C.D. Gelat and M.P. Jr Vecchi, *Optimization by Simulated Annealing*, Science, **220** (1983), 671–680.

[11] H. Konno and H. Yamazaki, *Mean-absolute deviation portfolio optimization model and its applications Tokyo Stock Market*, Management Science, **37** (1991), 519–531.

[12] H. Konno and A. Wijayanayake, *Portfolio optimization problem under concave transaction costs and minimal transactions unit constraints*, Mathematical Programming, **89** (2001), 233–250.

[13] Y. J. Lai and C. L. Hwang, "Fuzzy Mathematical Programming: Theory and Applications," Springer, Berlin, 1992.

[14] T. León, V. Liern and E. Vercher, *Viability of infeasible portfolio selection problems: A fuzzy approach*, European Journal of Operational Research, **139** (2002), 178–189.

[15] T. León, V. Liern and E. Vercher, *Two fuzzy approaches for solving multiobjective decision problems*, Computational Economics, **19** (2002), 273–286.

[16] T. León, V. Liern, P. Marco, J. V. Segura and E. Vercher, *A downside risk approach for the portfolio selection problem with fuzzy returns*, Fuzzy Economic Review, **9** (2004), 61–77.

[17] H.M. Markowitz, *Portfolio selection*, Journal of Finance, **7** (1952), 79–91.

[18] H.M. Markowitz, P. Kall and S. T. Wallace, "Portfolio selection: Efficient Diversification of Investments," John Wiley, New York, 1959.

[19] M. Mitchell, "An Introduction to Genetic Algorithms," The MIT Press, London, 1996.

[20] C.R. Reeves, Genetic Algorithms, In F. Glover and G.A. Kochenberger (editors)Handbook of Metaheuristics, Int. Series in Operations Research & Management Science 57, Kluwer Academic Publishers, 62–77, 2002.

[21] W.F. Sharpe, "Portfolio theory and capital market," McGraw-Hill, New York, 1970.

[22] M.G. Speranza, *Linear programming model for portfolio optimization*, Finance, **14** (1993), 107–123.

[23] J. Watada, *Fuzzy portfolio selection and its applications to decision making*, Tatra Mountains Mathematical Publications, **13** (1997), 219–248.

[24] S.A. Zenios and P. Kang, *Mean-absolute deviation portfolio optimization for mortgage-backed securities*, Annals of Operations Research **45** (1993), 433–450.

ISBN: 978-989-95079-6-8

Information System for the valuation of Universities in Spain

M. Socorro Garcia-Cascales[1], M. Teresa Lamata[2]

[1] Dpto de Electrónica, Tecnología de Computadoras y Proyectos.
Universidad de Politécnica de Cartagena. Murcia, España
[2] *Dpto. Ciencias de la Computación e Inteligencia Artificial.*
Universidad de Granada. 18071- Granada. España
Email: socorro.garcia@upct.es, mtl@decsai.ugr.es

Abstract— *The Spanish National Agency for Quality Assessment and Accreditation as part of its evaluation activities has established a procedure for evaluating both teaching and institutions, by means of the Institutional Assessment Programme. In this communication we shall focus on the external assessment phase for qualifications in the field of Industrial Engineering and specifically on the structures of the database for a Decision Support System on the universities' rankings. In particular, this paper will focus on obtaining the weight of the criteria and definition of the linguistic labels used in the external assessment phase.*

Keywords— Analytical Hierarchy Process (AHP), Linguistic Labels, Qualification evaluation, Weighting Criteria

1 Introduction

The university rankings which just 20 years ago were an innovation are today a normal characteristic in the majority of countries with extensive systems of higher education [7, 13, 16, 17, 18]. These lists have an increasing impact not only between the universities themselves, but also between different social sectors.

University rankings are lists of certain groups of institutions, classified in comparative form, and according to a common set of indicators, in descending order.

Likewise, the reorganisation at European scale of university studies as a result of the Bologna Process, will contribute in an active way to the harmonisation of basic European academic aspects. This aspect will allow, on the one hand, for a better connection between universities and on another hand, for an easier comparison between them; for which certain indicators of functioning will be necessary.

All the above leads us to think about the need for the existence of global rankings of universities as an instrument to measure their quality rigorously [7, 16]. In this context, the Decision Support Systems (DSS), seem to be useful for the evaluation of the qualifications in the universities' ranking.

In this paper, we shall focus on the area of industrial engineering within the Spanish university system and on the structure of the Database as a fundamental element of the DSS.

The aim of the paper is the use of the fuzzy AHP process to obtain the weight of the criteria within a DSS in order to facilitate the process of evaluation on the universities' rankings.

Another aim is to obtain the fuzzy numbers associated with the linguistic labels used to evaluate the ANECA in the external assessment phase.

The paper is organised as follows: The next section introduces the ANECA process. Section 3 introduces the linguistic variables and fuzzy sets. In section 4, the framework for the AHP method is defined. Section 5 describes the database. Section 6 shows the aggregation results and finally we outline the most important conclusions.

2 The ANECA Process

In this sense and within the framework of the Institutional Assessment Programme (*Programa de Evaluación Institucional*, "PEI"), the Spanish National Agency for Quality Assessment and Accreditation (*Agencia Nacional de Evaluación de la Calidad y Acreditación "*ANECA") presents this Guide [2] with the purpose of showing the steps in the Process of Institutional Assessment

The primary objective of the Institutional Assessment Programme is to facilitate an assessment process to officially improve the quality of education leading to obtain university degrees throughout the national territory, through self-diagnosis and the external view brought by experts.

The development of this programme is intended to promote assessment processes favouring the establishment or continuity of processes guaranteeing quality in teaching, as well as providing information to the students and their families, to society, to the governing bodies of the universities and to public administrations, regarding the quality of university teaching and their action plans.

This process is organised in three phases, see figure 1:

- *Self-assessment:* the "Self-assessment Report" describes and evaluates the situation of the assessed degree with respect to the criteria established, identifying strengths and weaknesses and enhancement proposals forming the basis of execution of the action plans that must be initiated on conclusion of the entire process. The report is written by the Self-assessment Committee.

ISBN: 978-989-95079-6-8

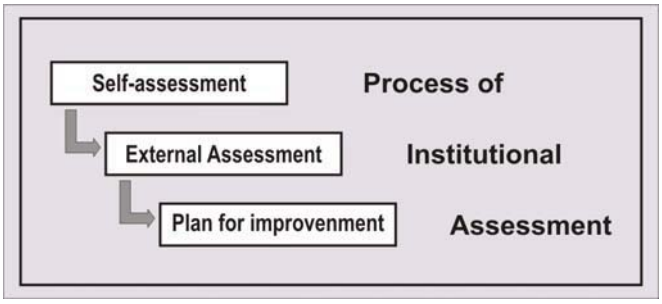


Figure 1: Phases of the Institutional Assessment Programme

- *External Assessment:* a group of external assessors to the teaching institution assessed, appointed by ANECA, and under its guidelines and supervision, analyzes the Self-Assessment Report, both through the study of documents and by means of a visit to the unit assessed, issues its recommendations, and proposes improvements. The result of this phase is the "External Assessment Report."

- *Plan for improvement*: it collects the main results of the Assessment process. This phase concludes with the plan for improvements of the degree, describing the proposed improvement actions in the Self-assessment and External Assessment phases, once their viability is analyzed. The tasks to be performed are determined according to their accomplishment, the persons responsible, resources involved, deadlines for their implementation, the indicators monitoring the actions proposed, and the benefits expected from them.

3 Linguistic variable and fuzzy sets

3.1 Linguistic variable

Most of the times, the decision-maker is not able to define the importance of the criteria or the goodness of the alternatives with respect to each criterion in a strict way. In many situations, we use measures or quantities which are not exact but approximate.

Since Zadeh [19] introduced the concept of fuzzy set and subsequently went on to extend the notion via the concept of linguistic variables, the popularity and use of fuzzy sets has been extraordinary. We are particularly interested in the role of linguistic variables as an ordinal scale and their associated terms, in this case triangular fuzzy number, as used in the multi-criteria decision making.

By a *linguistic variable,* [20,21,22], we mean a variable whose values are words or sentences in a natural or artificial language. For example *Age* is a linguistic variable if its values are linguistic rather than numerical, i.e., *young, not young, very young, quite young, old, not very old and not very young*, etc., rather than numbers as 20, 21,22, 23,.... .

Definition.1- A linguistic variable is characterized by a quintuple

$$\{X;T(X);U;G;M\}$$

in which

1. X is the name of the variable,
2. $T(X)$ is the term set of X, that is, the collection of its linguistic values
3. U is a universe of discourse,
4. G is a syntactic rule for generating the elements of $T(X)$ and
5. M is a semantic rule for associating meaning with the linguistic values of X.

In general for the decision-maker it is easier when he/she evaluates their judgments by means of linguistic terms. In those cases, the concept of fuzzy number is more adequate than that of real number.

3.2 Fuzzy sets

Then we have identified the linguistic variable with a fuzzy set [3,10,11]. The fuzzy set theory, introduced by Zadeh [19] to deal with vague, imprecise and uncertain problems has been used as a modelling tool for complex systems that can be controlled by humans but are hard to define precisely. A collection of objects (universe of discourse) X has a fuzzy set A described by a membership function f_A with values in the interval $[0,1]$.

$$f_A : X \to [0,1]$$

Thus A can be represented as $A = \{f_A(x) \mid x \in X\}$. The degree that u belongs to A is the membership function $f_A(x)$.

The basic theory of the triangular fuzzy number is described in Klir [12].

With regard to the fuzzy numbers, we will show only the mathematical operations that will be used throughout the development of the paper.

Definition 2. If T_1 and T_2 are two triangular fuzzy numbers defined by the triplets (a_1,b_1,c_1) and (a_2,b_2,c_2), respectively. Then, for this case, the necessary arithmetic operations with positive fuzzy numbers are:

a) Addition

$$T_1 \oplus T_2 = [a_1 + a_2, b_1 + b_2, c_1 + c_2] \quad (1)$$

b) Subtraction

$T_1 \Theta T_2 = T_1 + (-T_2)$ when the opposite

$-T_2 = (-c_2, -b_2, -a_2)$

$$\text{then} \quad T_1 \Theta T_2 = [a_1 - c_2, b_1 - b_2, c_1 - a_2] \quad (2)$$

c) Multiplication

$$T_1 \otimes T_2 = [a_1 \times a_2, b_1 \times b_2, c_1 \times c_2] \quad (3)$$

d) Division

$$T_1 \varnothing T_2 = [[a_1,b_1,c_1] \cdot [1/c_2, 1/b_2, 1/a_2]], \quad 0 \neq [a_2,b_2,c_2] \quad (4)$$

e) Scalar Multiplication

$$k \circ T_1 = (k \circ a_1, k \circ b_1, k \circ c_1) \quad (5)$$

ISBN: 978-989-95079-6-8

4 The Analytic Hierarchy Process Method (AHP)

The Analytic Hierarchy Process (AHP methodology [14,15] has been accepted by the international scientific community as a robust and flexible multi-criteria decision making tool for dealing with complex decision problems. AHP has been applied to numerous decision problems such as energy policy [1,7], project selection [6], measuring business performance [1], and evaluation of advanced manufacturing technology [4,5].

- Basically, AHP has three underlying concepts: Structuring the complex decision problem as a hierarchy of goal, criteria and alternatives,
- Pair-wise comparison of elements at each level of the hierarchy with respect to each criterion on the preceding level,
- and finally vertically synthesizing the judgements over the different levels of the hierarchy.

AHP attempts to estimate the impact of each one of the alternatives on the overall objective of the hierarchy. In this case, we only apply the method in order to obtain the criteria's weights.

We assume that the quantified judgements provided by the decision-maker on pairs of criteria (C_i, C_j) are represented in an *nxn* matrix as in the following:

$$C = \begin{array}{c} \\ C_1 \\ C_2 \\ . \\ . \\ C_n \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \cdots & C_n \end{array} \\ \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \ldots & c_{2n} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ c_{n1} & c_{n2} & \ldots & c_{nn} \end{bmatrix} \end{array} \quad (6)$$

The c_{12} value is supposed to be an approximation of the relative importance of C_1 to C_2, i.e., $c_{12} \approx (w_1/w_2)$. This can be generalized and the statements below can be concluded:

1. $c_{ij} \approx (w_i/w_j)$ $i,j=1, 2, ..., n$
2. $c_{ii}=1$, $i=1,2,...,n$
3. If $c_{ij}=\alpha$, $\alpha \neq 0$, then $c_{ji}=1/\alpha$, $i=1,2,...,n$
4. If C_i is more important than C_j then $c_{ij} \cong (w_i/w_j)>1$

This implies that matrix *A* should be a positive and reciprocal matrix with 1's in the main diagonal and hence the decision-maker needs only to provide value judgements in the upper triangle of the matrix. The values assigned to c_{ij} according to Saaty scale are usually in the interval of 1-9 or their reciprocals.

It can be shown that the number of judgements (*L*) needed in the upper triangle of the matrix are:

$$L = n(n-1)/2 \quad (7)$$

where *n* is the size of the matrix *C*.

The matrixes associated to the AHP approach are reciprocal, thus:

- The maximum eigenvalue (λ_{max}) is a positive real number and such that $\lambda_{max} \geq n$.
- Associated with this eigenvalue is a vector whose components are also positive. If this vector is normalized the vector of weights associated with the matrix is obtained.

Where the values are fuzzy, not crisp; instead of using lambda as an estimator to the weight, we will use the geometric normalized average, expressed by the following expression:

$$w_i = \frac{\prod_{j=1}^{n}(a_{ij}, b_{ij}, c_{ij})}{\sum_{i=1}^{m}\prod_{j=1}^{n}(a_{ij}, b_{ij}, c_{ij})} \quad (8)$$

where, (a_{ij}, b_{ij}, c_{ij}) is a fuzzy number.

5 The database

The Database has two parts. The first part consists of the summary of the different Reports of External Evaluation, published by ANECA. These reports are given by the members of the External Assessment Committee who are persons qualified by ANECA; the valuation given by the external assessors is impartial.

The second part of the database, which is the focus of this paper, corresponds with obtaining the weights of the criteria, as well as the numerical representation of the labels. See Figure 2.

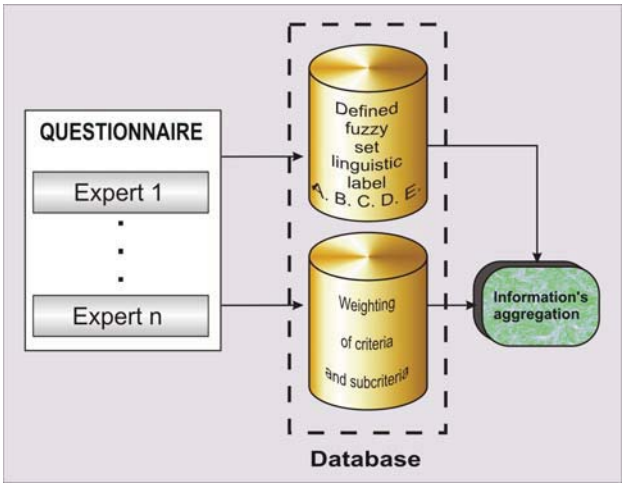


Figure 2: The database corresponding to this part of the work.

The study carried out has been based on a questionnaire designed to such effect. ANECA itself sent the questionnaire to the experts that it determined as proper assessors. The questionnaire was sent / answered by e-mail.

This questionnaire consists of two clearly differentiated parts;

- The first one has been produced on the basis of the hierarchic structure of the criteria and subcriteria (Table 1), using the AHP methodology to do so.
- The second part refers to the linguistic labels used in the ANECA survey for the external evaluation of university qualifications.

ISBN: 978-989-95079-6-8

Table 1. Weighting criteria/subcriteria

CRITERIA			General Weighting		
1st LEVEL	2nd LEVEL	3rd LEVEL	3th. LEVEL	2nd LEVEL	1st LEVEL
1. Educational programme (E.P)	1.1. Aims of the programme	Aims of the E.P.	(3.39,**5.26**,8.01)		
		Admission profile	(3.39,**5.26**,8.01)	(6.77,**10.52**,16.02)	
	1.2. Studies plan and its structure	Curricular content	(1.50,**2.32**,3.54)		
		Curricular coherence	(1.50,**2.32**,3.54)		
		Curricular consistency	(1.50,**2.32**,3.54)		
		Curricular updating	(1.50,**2.32**,3.54)		
		Aims of E.P.	(1.50,**2.32**,3.54)	(7.49,**11.61**,17.72)	(16.61,**22.13**,29.14)
2. Organization of teaching	2.1. Management and planning	Planning	(5.27,**7.94**,11.81)	(5.27,**7.94**,11.81)	
	2.2. Management and organization	Communication	(1.59,**2.39**,3.57)		
		Organisation of teaching	(1.59,**2.39**,3.57)		
		Improvement processes	(1.59,**2.39**,3.57)	(4.78,**7.16**,10.70)	(11.33,**15.10**,20.04)
3. Human Resources	3.1. Academic staff (AS)	Appropriateness AS	(5.11,**7.95**,12.15)		
		Implication AS	(5.11,**7.95**,12.15)	(10.21,**15.90**,24.29)	
	3.2. Administration and service staff (ASS)	Adaptation ASS	(2.50,**3.79**,5.94)	(2.50,**3.79**,5.94)	(14.44,**19.70**,26.70)
4. Material Resources	4.1 Classrooms	Appropriateness for numbers of students	(1.58,**2.50**,3.94)	(1.58,**2.50**,3.94)	
	4.2 Work Spaces	Appropriateness for numbers of students	(0.51,**0.81**,1.31)		
		Appropriateness (AS and ASS)	(0.51,**0.81**,1.31)		
		Infrastructures: practical	(0.51,**0.81**,1.31)	(1.53,**2.44**,3.93)	
	4.3. Laboratories, workshops and experimental spaces	Appropriateness for number of students	(1.82,**2.86**,4.53)	(1.82,**2.86**,4.53)	
	4.4. Library and document banks	Correctly furnished	(0.69,**1.09**,1.78)		
		Quality, quantity, …	(0.69,**1.09**,1.78)	(1.38,**2.19**,3.56)	(7.55,**9.99**,13.46)
5. Training Process	5.1. Student assistance and integral training	Capture	(0.54,**0.84**,1.33)		
		Student welcome actions	(0.54,**0.84**,1.33)		
		Support programmes	(0.54,**0.84**,1.33)		
		Professional orientation programmes	(0.54,**0.84**,1.33)		
		Tutorial action programmes	(0.54,**0.84**,1.33)		
		Integral training	(0.54,**0.84**,1.33)	(3.24,**5.06**,8.01)	
	5.2. Teaching-learning process	Methodology	(1.55,**2.46**,3.84)		
		Evaluation	(1.55,**2.46**,3.84)		
		External practical	(1.55,**2.46**,3.84)		
		Mobility	(1.55,**2.46**,3.84)	(6.21,**9.84**,15.34)	(11.04,**14.90**,20.10)
6. Results	6.1. Results of educational programme	Effectiveness of E.P.	(1.56,**2.73**,4.67)		
		Student satisfaction	(1.56,**2.73**,4.67)	(3.13,**5.47**,9.35)	
	6.2. Graduate results	Compliance with the graduate profiles	(3.39,**5.99**,10.34)	(3.39,**5.99**,10.34)	
	6.3. Academic staff results	Academic staff satisfaction	(1.82,**3.21**,5.70)	(1.82,**3.21**,5.70)	
	6.3. Results in society	Employers and other groups	(0.10,**1.76**,3.08)		
		Social link	(0.10,**1.76**,3.08)	(0.20,**3.52**,6.17)	(8.54,**18.18**,24.75)
Source: own production			**100.00**	**100.00**	**100.00**

Likewise, we will take into account the possible valuations of the experts, expressed in linguistic terms (A, B, C, D) These labels are arranged from largest to smallest as follows: A > B > C > D and with the semantics that we will see later.

It should be remembered that the results obtained correspond to a problem of group decision-making, formed by *n* experts, from whom we will obtain:

1.-The criteria weight as a result of the consensus

2.- The membership functions of the fuzzy numbers that represent the linguistic labels A,B,C and D.

5.1 Obtaining the weightings of the criteria/subcriteria

According to Table 1, the hierarchy structure has three levels. Considering expression (7), the large number of questions might lead to the survey not being answered, since L = 74, as we have 6 in 1st level criteria, subcriteria 16 in 2nd level and 37 in 3rd level indicators. This results in a questionnaire which is not feasible for experts to answer.

For this, and since the influence of the third level on the weight of the previous ones is practically null, a uniform distribution on them was supposed.

ISBN: 978-989-95079-6-8

For the first and second level, the procedure is as follows: firstly, we ask if all the criteria/subcriteria have the same weight, if so, we pass to the following level.

On the contrary, if they do not have the same weight, it will be continued by the following question of the questionnaire.

In this part, there is a question that asking about the order of importance of the criteria/subcritera; and finally, using the linguistic labels defined by Saaty [14], the criteria/subcriteria are compared at the same level, taking into account the order established before.

We chose to ask only for one row of the pairwise matrix and from here to generate the rest of the information matrix, which was carried out completely consistently. Thus, this part of the questionnaire had only 21 questions.

5.2 Obtaining the linguistic evaluations

The second part of the questionnaire is based on the semiquantitative survey of the Evaluation of the Education inside the protocol to elaborate the Report of External Evaluation developed by ANECA, in which the following labels are used: A: Excellent, B: Good, C: Average, D: Deficient. For the valuation of these labels we use the interval [0, 10].

These linguistic variables, by the uncertainty of their nature, justified the use of fuzzy numbers associated with each linguistic term.

6 The results of the aggregation for the group decision experts

By means of a primary group decision-making process, it is possible to see the results of the weighting of the criteria and sub-criteria for all the experts (Table 1).

Table 2. Fuzzy numbers associated with the labels A,B,C and D.

Semi quantitative labels	General
A: Excellent	(8.1354, 9.4054, 10.0000)
B: Good	(5.8108, 7.1081, 8.4054)
C: Average	(3.5090, 4.8108, 6.1126)
D: Deficient	(0.7355, 2.5135, 4.2916)

Now, taking these weightings as the base; and the definition of semiquantitative labels (Table 2), obtained by ANECA's own experts, it is possible for us, attending to these six criteria, to order the different Universities. We make reference, in this case, only to the area of industrial engineering.

For the evaluation of the alternatives the methodology used has been the fuzzy weighted sum model as:

$$FWSM = \sum_{j=1}^{n} \tilde{w}_j \cdot \tilde{a}_{ij} \qquad (9)$$

where, $\tilde{w}_j$ and $\tilde{a}_{ij}$ are fuzzy numbers.

The defuzzification method used is described in [8] when:

$$I_{\beta,\lambda}(A_i) = \beta S_M(A_i) + (1-\beta)\lambda S_R(A_i) + (1-\beta)(1-\lambda) S_L(A_i)$$

In this way, we have defined a fuzzy number as a function of the three integrals, $S_L(A_i)$, $S_M(A_i)$ and $S_R(A_i)$, where $S_R(A_i)$ represents the upper mean value associated with the inverse function of $f_A^R(x)$, $S_L(A_i)$ is the lower mean value of the $g_A^L(x)$ function and $S_M(A_i)$ is the area of the core of the fuzzy number, $\beta \in [0,1]$ is the index of modality that represents the importance of the central value against the extreme values, and $\lambda \in [0,1]$ is the degree of optimism of the decision maker.

We have considered the case in which the three areas have the same weight and it would correspond to the neutral decision maker, when $\alpha = 1/2$ y $\beta = 1/3$

Taking as an example (Table 3), the outcome results of the DSS, for the real case of the evaluation of five universities in the qualifications of industrial engineering, where we have obtained results both for the principal criteria and for the global evaluation.

Table 3: Ranking result for five Universities.

Criteria	C1 Educational Programme	C2 Teaching Organizatio	C3 Human Resources	C4 Material Resources	C5 Educational Process	C6 Results	IAP Gobal
U1	0.062	0.033	0.071	0.040	0.042	0.056	0.306
U2	0.100	0.067	0.106	0.049	0.055	0.089	0.466
U3	0.129	0.086	0.105	0.067	0.098	0.119	0.604
U4	0.118	0.080	0.088	0.038	0.075	0.081	0.482
U5	0.125	0.089	0.105	0.047	0.082	0.100	0.548

7 Conclusions

For the assessors/experts it is simpler to express their knowledge by means of linguistic labels, instead of having to do so by means of numerical values. For that reason, it is preferable to prepare a questionnaire to obtain the experts' knowledge, in which the answers are in the form of linguistic variables. These linguistic variables have been modelized by means of fuzzy triangular numbers and from a methodology widely accepted by the scientific community, since it is the Analytical Hierarchic Process; developed by Saaty in 1980.

Taking Table 1into account, we conclude that in general the most important criterion is the Educational Programme and that the lowest weighting is obtained for the Material Resources.

For future work, it would be interesting to carry out a study of the aggregation of the information as a secondary process. In this process the weight of the criteria and subcriteria are obtained for every expert. Later all this information would be aggregated.

Moreover, a comparative primary and secondary study of both types of processes would be desirable.

ISBN: 978-989-95079-6-8

Similarly, as future work, it would be interesting to make a comparative study of the evaluation criteria according to the area of the different experts. It is possible that the valuation of an expert from the area of engineering does not value the criteria in the same way as another from the humanities.

Acknowledgment

This work is partially supported by the DGICYT under project TIN2008-06872-C04-04 and Junta de Andalucía (P07-TIC02970), respectively.

References

[1] K. M. A. Al Harbi. Application of the AHP in project management. International Journal of Project Management;,19(1), 19-27, 2001 .

[2] Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) (2007). Programa de Evaluación Institucional. Anualidad 2006/2007. Guía de evaluación externa 2007. http://www.aneca.es/active/active_ense_pei0607.asp

[3] R.E. Bellman, L.A. Zadeh. Decision-making in a fuzzy environment. Management Sci., 17, 141-164.,1970.

[4] Chan F.T.S, Chan M.H, Tang N.K.H. Evaluation methodologies for technology selection. Journal of Materials Processing Technology 2000a ; 107: 330–337.

[5] F.T.S Chan, B.Jiang, N.K.H. Tang. The development of intelligent decision support tools to aid the design flexible manufacturing systems. International Journal of Production Economics, 65 (1): 73–84, 2000.

[6] C.-H. Cheng, K.-L. Yang, C.-L. Hwang. Evaluating attack helicopters by AHP based on linguistic variable weight. European Journal of Operation al Research 116 (2),423–435, 1999.

[7] G. Federkeil.. Rankings and Quality Assurance in Higher Education. Higher Education in Europe, 33, 2/3, 219-230, 2008.

[8] M.S, Garcia-Cascales,., M.T. Lamata,: A modification to the index of Liou and Wang for ranking fuzzy number, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 411-424, 2007.

[9] M. M. Kablan. Decision support for energy conservation promotion: an analytic hierarchy process approach. Energy Policy, 32 (10): 1151-1158, 2004.

[10] J. Kacprzyk, R.R. Yager. Linguistic summaries of data using fuzzy logic. International Journal of General Systems. 30, 2, 133-154, 2001.

[11] E.E. Kerre. The use of fuzzy set theory in electrocardiological diagnostics, in: M.M. Gupta and E. Sanchez, Eds., Approximate Reasoning in Decision Analysis Amsterdam; North-Holland 277-282, 1982.

[12] G.J. Klir and B. Yuan, Fuzzy sets and Fuzzy Logic, Prentice Hall PTR, New Jersey e d. 1995

[13] P. Montesinos, J. M.Carot, J. M. Martinez, F. Mora. Third Mission Ranking for World Class Universities: Beyond Teaching and Research. Higher Education in Europe, 33, 2 & 3, 259 – 27, 2008.

[14] T.L. Saaty. The Analytic Hierarchy Process. Macgraw Hill. 1980.

[15] T.L. Saaty. Group Decision Making and the AHP. Springer Verlag. New York. 1989.

[16] J. Sadlak J. Merisotis, N.C. Liu. University Rankings: Seeking Prestige, Raising Visibility and Embedding Quality. Higher Education in Europe, 33, 2 & 3, 195 – 199, 2008.

[17] M.E. Smith. Change and innovation in higher education: a role for corporate strategy?. Higher Education 16:37-52, 1987.

[18] E. de Weert. A macro-analysis of quality assessment in higher education. Higher Education 19:57-72, 1990.

[19] L. A. Zadeh . Fuzzy sets. *Information and Control*, 8:338-353, 1965.

[20] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning: Part 1. Information Sciences, 8, 199-249, 1975.

[21] L. A. Zadeh, J. Kacprzyk (eds). Computing with Words in Information / Intelligent Systems 1. Foundations. Studies in Fuzziness and Soft Computing, Physica-Verlag (Springer-Verlag), Heidelberg and New York. Vol. 34 1999.

[22] L. A. Zadeh, J. Kacprzyk (eds). Computing with Words in Information / Intelligent Systems 2. Applications. Studies in Fuzziness and Soft Computing. Physica-Verlag (Springer-Verlag), Heidelberg and New York. Vol. 33 1999.

ISBN: 978-989-95079-6-8

Multiple-Criteria Fuzzy Evaluation: The FuzzME Software Package

Jana Talašová[1] Pavel Holeček[2]

1., 2. Faculty of Science, Palacký University Olomouc
tř. Svobody 26, 771 46 Olomouc, Czech Republic
Email: talasova@inf.upol.cz, holecekp@inf.upol.cz

Abstract— *This paper introduces a new software product FuzzME. It was developed as a tool for creating fuzzy models of multiple-criteria evaluation and decision making. The type of evaluations employed in the fuzzy models fully corresponds with the paradigm of the fuzzy set theory; the evaluations express the (fuzzy) degrees of fulfillment of corresponding goals. The FuzzME software takes advantage of linguistic fuzzy modeling to the maximum extent. In the FuzzME software, both quantitative and qualitative criteria can be used. The basic structure of evaluation is described by a goals tree. Within the goals tree, aggregation of partial fuzzy evaluations is done either by one of fuzzified aggregation operators (fuzzy weighted average, ordered fuzzy weighted average) or by a fuzzy expert system.*

The paper contains an illustrative example of the software usage. The application concerns a soft-fact-rating problem that was solved in one of Austrian banks.

Keywords— Fuzzy expert system, Fuzzy OWA operator, Fuzzy weighted average, Multiple-criteria fuzzy evaluation, Normalized fuzzy weights, Software.

1 Introduction

In practice, sophisticated models of multiple-criteria evaluation are required (e.g. rating of clients in banks, evaluation of hospitals or universities, comparison of alternative solutions to ecological problems). For creating the evaluating models, setting some of their input data and interpreting their outputs, expert's knowledge is needed (e.g. evaluations of alternatives according to qualitative criteria, partial evaluating functions for quantitative criteria, a choice of a suitable type of aggregation, criteria weights, or eventually, a rule base describing the relation between criteria values, the overall evaluation and a linguistic description of obtained results). Because uncertainty is the typical feature of any expert information, the fuzzy set theory is a suitable mathematical tool for creating such models. For the practical use of the fuzzy models of multiple-criteria evaluation, their user-friendly software implementation is necessary. But a good theoretical basis of the used models is crucial, too. The clear and well-elaborated theory of multiple-criteria fuzzy evaluation makes it possible to create an understandable methodics for the software user. And a good methodics is essential for correct application of any software to solving real problems.

There is a large number of papers and books dealing with the theory and methods of multiple-criteria evaluation that make use of the fuzzy approach (e.g. [1], [2], [3], [4]). Multiple-criteria evaluation (as a basis of decision making) was even one of the earliest applications of fuzzy sets (see [1]). In more than 40 years of existence of the fuzzy sets theory, several software products for multiple-criteria decision making, which use the fuzzy modeling principles in different degrees and in different ways, have been developed. In practice, FuzzyTECH (see [5]) is probably the best-known of these. It enables to use the specific instruments of the fuzzy set theory for solving multiple-criteria evaluation and decision making problems. Generally, FuzzyTECH is a universal software product which makes it possible to create and use fuzzy expert systems (or fuzzy controllers). It also includes the possibility to derive fuzzy rule bases from given data by means of neural network algorithms. In the book [6], there were described many successful applications of this software to solving multiple-criteria evaluation and decision making problems in the areas of business and finance. Similarly, fuzzy toolboxes of general mathematical systems such as Matlab can be used for multiple-criteria decision making.

The FuzzME software (**Fuzz**y models of **M**ultiple-criteria **E**valuation), presented in this paper, is based on a theoretical concept of evaluation which is very close to the original Zadeh's ideas. Similarly to his paper [1], the evaluations of alternatives according to particular criteria represent their degrees of fulfillment of the corresponding partial goals. Besides evaluations expressed by real numbers in $[0, 1]$, fuzzy evaluations modeled by fuzzy numbers on the same interval are employed in the software. They represent, analogously, the fuzzy degrees of fulfillment of the partial goals which are connected to the criteria. Resulting fuzzy evaluations, which are obtained by aggregation, have a similar clear interpretation. This theoretical approach to (fuzzy) evaluation was published in the book [7] and in the paper [8].

The predecessor of the FuzzME software package in terms of the used theoretical basis was the NEFRIT software. This software for multiple-criteria evaluation and decision making, which is also based on fuzzy technologies, was developed in about 2000 by the Czech software company TESCO SW Inc. The fuzzy model of evaluation applied there is described in detail in [8] and in the book [7] (a demo version of NEFRIT is enclosed in the book). NEFRIT makes it possible to work with expert fuzzy evaluations of alternatives according to qualitative criteria. Values of the quantitative criteria can either be crisp or fuzzy. Evaluating functions for quantitative criteria represent membership functions of corresponding partial goals. The main evaluation structure is expressed by a goals tree. For aggregation of the partial fuzzy evaluations the weighted average method is used. The weights (crisp only, not fuzzy) express the shares of partial evaluations in the

ISBN: 978-989-95079-6-8

aggregated one. Fuzzy evaluations on all levels of aggregation express the fulfillment of the corresponding goals. The NEFRIT software was originally developed for the Czech National Bank (decision making about granting a credit). Further, it was used e.g. by the Czech Tennis Association, the Czech Basketball Association and in other institutions. Nowadays it is tested by the Supreme Audit Office of the Czech Republic.

In contrast to NEFRIT, the FuzzME software makes it possible to use also uncertain weights in the aggregation by means of the weighted average method. The theory of normalized fuzzy weights, procedures for their setting, and an effective algorithm for calculation of fuzzy weighted average are taken from [9], [10] and [11]. Another fuzzy aggregation operator, available in the FuzzME software, is a fuzzified OWA operator. Again, it works with normalized fuzzy weights. The fuzzy OWA operator and the used algorithm for its calculation are described in [12]. In the FuzzME software, multiple-criteria evaluating functions can also be defined by means of fuzzy rule bases. In accordance with [7], two algorithms are offered for the approximate reasoning - the standard Mamdani algorithm and a modified Sugeno algorithm.

There are also software tools for multiple-criteria decision making based on other mathematical methods. But they are usually designed for solving a particular assignment. Our investigation by means of Internet did not result software fully comparable to FuzzME. Its universality and comprehensiveness make it unique.

2 Preliminaries

A fuzzy set A on a universal set X is characterized by its membership function $A : X \to [0,1]$. $Ker\ A$ denotes a kernel of A, $Ker\ A = \{x \in X \mid A(x) = 1\}$. For any $\alpha \in [0,1]$, A_α denotes an α-cut of A, $A_\alpha = \{x \in X \mid A(x) \geq \alpha\}$. A support of A is defined as $Supp\ A = \{x \in X \mid A(x) > 0\}$. The symbol $hgt\ \ A$ denotes a height of the fuzzy set A, $hgt\ A = \sup\{A(x) \mid x \in X\}$. An intersection and a union of the fuzzy sets A and B on X are defined for all $x \in X$ by the following formulas: $(A \cap B)(x) = \min\{A(x), B(x)\}$, $(A \cup B)(x) = \max\{A(x), B(x)\}$.

A fuzzy number is a fuzzy set C on the set of all real numbers $\Re$ which satisfies the following conditions: a) the kernel of C, $Ker\ C$, is not empty, b) the α-cuts of C, C_α, are closed intervals for all $\alpha \in (0,1]$, c) the support of C, $Supp\ C$, is bounded. A fuzzy number C is called to be defined on $[a,b]$, if $Supp\ C \subseteq [a,b]$. Real numbers $c^1 \leq c^2 \leq c^3 \leq c^4$ are called significant values of the fuzzy number C if the following holds: $[c^1, c^4] = Cl(Supp\ C), [c^2, c^3] = Ker\ C$, where $Cl(Supp\ C)$ denotes a closure of $Supp\ C$.

Any fuzzy number C can be characterized by a pair of functions $\underline{c} : [0,1] \to \Re$, $\overline{c} : [0,1] \to \Re$ which are defined by the following formulas: $C_\alpha = [\underline{c}(\alpha), \overline{c}(\alpha)]$ for all $\alpha \in (0,1]$, and $Cl(Supp\ C) = [\underline{c}(0), \overline{c}(0)]$. The fuzzy number C is called to be linear if both the functions $\underline{c}$, $\overline{c}$ are linear. A linear fuzzy number is fully determined by its significant values because $\underline{c}(\alpha) = (c_2 - c_1) \cdot \alpha + c_1$, $\overline{c}(\alpha) = (c_3 - c_4) \cdot \alpha + c_4$. For that reason, we can denote it as $C = (c^1, c^2, c^3, c^4)$.

An ordering of fuzzy numbers is defined as follows: a fuzzy number C is greater than or equal to a fuzzy number D, if $C_\alpha \geq D_\alpha$ for all $\alpha \in (0,1]$.

A fuzzy scale makes it possible to represent a closed interval of real numbers by a finite set of fuzzy numbers. Let $T_1, T_2, ..., T_s$ be fuzzy numbers defined on $[a,b]$, forming a fuzzy partition on the interval, i.e., for all $x \in [a,b]$ the following holds

$$\sum_{i=1}^{s} T_i(x) = 1, \tag{1}$$

then the set of the fuzzy numbers can be linearly ordered (see [7]). If the fuzzy numbers $T_1, T_2, ..., T_s$ are defined on $[a,b]$, form a fuzzy partition on the interval and are numbered according to their linear ordering, then they are said to form a fuzzy scale on $[a,b]$.

An uncertain division of the whole into m parts can be modeled by normalized fuzzy weights. Fuzzy numbers $V_1, ..., V_m$ defined on $[0,1]$ are normalized fuzzy weights if for any $i \in \{1, ..., m\}$ and any $\alpha \in (0,1]$ it holds that for any $v_i \in V_{i\alpha}$ there exist $v_j \in V_{j\alpha},\ j = 1, ..., m,\ j \neq i$, such that

$$v_i + \sum_{j=1, j\neq i}^{m} v_j = 1. \tag{2}$$

3 The FuzzME software

The mathematical models of the FuzzME software are based primarily on the theory and methods of multiple-criteria evaluation that were published in [7] and [8]. The theory of normalized fuzzy weights, the definition of fuzzy weighted average, and the algorithm for its computation were taken from [9], [10] and [11]. The fuzzified OWA operator and the algorithm for its calculation published in [12] are also used in the software.

In the FuzzME software, the basic structure of the fuzzy model of multiple-criteria evaluation is expressed by a goals tree. The root of the tree represents the overall goal of evaluation and each branch corresponds to a partial goal. The goals at the ends of branches are connected either with quantitative or qualitative criteria.

When an alternative is evaluated, evaluations with respect to criteria connected with the terminal branches are calculated first. Independently of the criterion type, each of the evaluations is described by a fuzzy number defined on the interval $[0,1]$. It expresses the fuzzy degree of fulfillment of the corresponding partial goal.

These partial fuzzy evaluations are then aggregated according to the defined type of the tree node. Three types of aggregation are available: a fuzzy weighted average (fuzzy WA), an ordered fuzzy weighted average (fuzzy OWA) or aggregation by means of a fuzzy expert system. For aggregation by fuzzy weighted average or ordered fuzzy weighted average, normalized fuzzy weights must be set. The weights express uncertain shares of the partial evaluations in the aggregated one. For the fuzzy expert system, the fuzzy rule base must be defined and an inference algorithm must be chosen (either the Mamdani algorithm or the generalized Sugeno algorithm of approximate reasoning).

The overall evaluation reflects the degree of fulfillment of the main goal. A verbal description of the overall evaluation can be obtained by means of the implemented linguistic approximation algorithm.

ISBN: 978-989-95079-6-8

The overall evaluations can be compared within the frame of a given set of alternatives. By this comparison the best of the alternatives can be chosen. That is why the FuzzME software can be also used as a decision support system.

The import and export of data is supported by the software, too. The FuzzME software is available in the Czech and English versions.

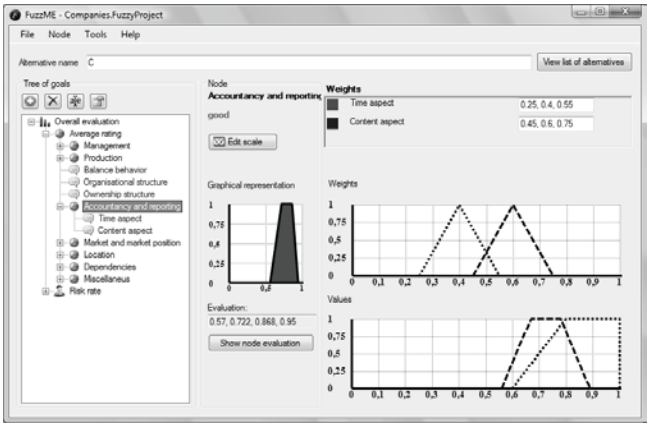


Figure 1: The main window of the software

3.1 Goals tree

Goals trees represent the basic structure of fuzzy models of multiple-criteria evaluation in the FuzzME software. When a goals tree is designed, the main goal is consecutively divided into goals of progressively lower levels. The process of division is stopped when such goals are reached whose fulfillment can be assessed by means of some known characteristics of alternatives (i.e. quantitative or qualitative criteria).

The design of a tree structure in the goals-tree editor is the first step in forming a fuzzy evaluation model in FuzzME. In the next step, the type of each node in the tree must be specified. For the nodes at the ends of tree branches the user defines if the node is connected with a quantitative or qualitative criterion. For the other nodes he/she sets the type of aggregation - fuzzy weighted average, ordered fuzzy weighted average or fuzzy expert system. An example of a goals tree is illustrated in Fig. 2.

3.2 Criteria of evaluation

In the models of evaluation created by the FuzzME software, qualitative and quantitative criteria can be combined arbitrarily.

3.2.1 Qualitative criteria

According to qualitative criteria, alternatives are evaluated verbally, by means of values of linguistic variables of special kinds - linguistic scales, extended linguistic scales and linguistic scales with intermediate values.

A linguistic variable is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is a set of its linguistic values, X is a universal set on which the meanings of the linguistic values are defined, G is a syntactic rule for generating values in $\mathcal{T}(\mathcal{V})$, and M is a semantic rule which maps each linguistic value $\mathcal{C} \in \mathcal{T}(\mathcal{V})$ to its mathematical meaning, $C = M(\mathcal{C})$, which is a fuzzy set on X.

A linguistic scale on $[a, b]$ is a special case of the linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $X = [a, b]$, $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_s\}$ and the meanings of the linguistic values $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_s$ are modeled by fuzzy numbers $T_1, T_2, \ldots, T_s$ which form a fuzzy scale on $[a, b]$. As the set of linguistic values of the scale is defined explicitly, it is not necessary to include the grammar G into the scale notation.

In the FuzzME software, the user defines a linguistic scale for each qualitative criterion in the fuzzy-scale editor. For example, the linguistic scale *quality of a product* can contain linguistic values *poor*, *substandard*, *standard*, *above standard* and *excellent*. The evaluating linguistic scale is usually defined on $[0, 1]$; in other cases, it has to be transformed to this interval.

The extended linguistic scale contains, besides elementary terms of the original scale, $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_s$, also derived terms in the form $\mathcal{T}_i$ *to* $\mathcal{T}_j$, where $i < j, i, j \in \{1, 2, \ldots, s\}$. For example, the user can evaluate *quality of a product* by the linguistic term *standard to excellent*. The meaning of the linguistic value $\mathcal{T}_i$ *to* $\mathcal{T}_j$ is modeled by $T_i \cup_L T_{i+1} \cup_L \cdots \cup_L T_j$, where $\cup_L$ denotes the union of fuzzy sets based on the Lukasiewicz disjunction; e.g. $(T_i \cup_L T_{i+1})(x) = \min\{1, T_i(x) + T_{i+1}(x)\}$ for all $x \in \Re$.

The linguistic scale with intermediate values is the original linguistic scale enriched with derived terms *between* $\mathcal{T}_i$ *and* $\mathcal{T}_{i+1}$, $i \in \{1, 2, \ldots, s-1\}$. The meaning of the derived term *between* $\mathcal{T}_i$ *and* $\mathcal{T}_{i+1}$ is modeled by the arithmetic average of the fuzzy numbers T_i and T_{i+1}.

In the FuzzME software, the user evaluates a given alternative according to a qualitative criterion by selecting a proper linguistic evaluation from a drop-down list box. He/she can choose the value from a standard linguistic scale, extended scale or scale with intermediate values.

The three mentioned structures of linguistic values are also applied when resulting fuzzy evaluations are approximated linguistically.

3.2.2 Quantitative criteria

The evaluation of an alternative with respect to a quantitative criterion is calculated from the measured value of the criterion by means of the evaluating function expertly defined for the criterion. The evaluating function is the membership function of the corresponding partial goal. The FuzzME software admits both crisp and fuzzy values of quantitative criteria. The fuzzy values represent inaccurate measurements or expert estimations of the criteria values. In the case of a fuzzy value, the corresponding partial fuzzy evaluation is calculated by the extension principle.

In the FuzzME software, the evaluating function of a quantitative criterion is formally set by means of a fuzzy number. For example, if the evaluating function is defined by a linear fuzzy number $F = (f_1, f_2, f_3, f_4)$, then f_1 is the lower limit of all at least partly acceptable values of the criterion, f_2 is the lower limit of its fully satisfactory values, f_3 is the upper limit of the fully satisfactory values, and f_4 is the upper limit of the acceptable values.

For example, when a bank evaluates expected profitability of projects, the evaluating function can be defined by a linear fuzzy number with significant values 10, 30, 500, 500. In that case, values lower than 10% are not satisfying at all (the client

ISBN: 978-989-95079-6-8

would not be able to pay the money back to the bank). For the values from 10% to 30% the satisfaction of the bank is growing linearly. Values greater than 30% are fully satisfactory from the bank's point of view. Values greater than 500% are not supposed to occur. This way we can define a monotonous evaluating function, which is the most common in the evaluating models, by a fuzzy number.

3.3 Methods of aggregation of partial evaluations

The calculated partial fuzzy evaluations are then consecutively aggregated according to the structure of the goals tree. With respect to the defined type of the tree node, the fuzzy weighted average method, the ordered fuzzy weighted average method or the fuzzy expert system method is used for the aggregation. Each of the aggregation methods is suitable for a different situation:

The fuzzy weighted average is used if the goal corresponding with the node of interest is fully decomposed into disjunctive goals of the lower level. The normalized fuzzy weights represent uncertain shares of these lower-level goals in the goal corresponding with the considered node.

Again, the ordered fuzzy weighted average requires that the goal corresponding with the given node is decomposed into disjunctive goals of the lower level. In contrast to the fuzzy weighted average, the usage of this aggregation operator supposes special user's requirements concerning the structure of partial fuzzy evaluations. The normalized fuzzy weights again represent uncertain shares of the partial evaluations in the aggregated one. But the normalized fuzzy weights are not linked to the individual partial goals; the correspondence between the weights and the partial evaluations is given by the ordering of partial evaluations of the alternative of interest. It means, evaluations with respect to the same partial goal can have different weights for different alternatives.

If the relationship between the evaluations of the lower level and the evaluation corresponding with the given node is more complex (if neither of the two previous methods can be used), and if expert knowledge about the relationship is available, then the aggregation function is described by a fuzzy rule base of a fuzzy expert system. The approximate reasoning is used to calculate the resulting evaluation. In particular, evaluating function described by a fuzzy expert system is used if the fulfillment of a goal at the end of a tree branch depends on several mutually dependent criteria (i.e., if combinations of criteria values bring synergic or disynergic effects to the resulting multiple-criteria evaluation).

3.3.1 Fuzzy weighted average

If the fuzzy weighted average is used for aggregation of partial fuzzy evaluations, then the uncertain weights of the corresponding partial goals, which express their shares in the superior goal, must be set. To define consistent uncertain weights, a special structure of fuzzy numbers, normalized fuzzy weights, must be used.

In the FuzzME software, both real and fuzzy normalized weights can be used. Normalized real weights, i.e., real numbers $v_1, ..., v_m$, $v_j \geq 0$, $j = 1, ..., m$, $\sum_{j=1}^{m} v_j = 1$, represent a special case of the normalized fuzzy weights.

The fuzzy weighted average of the partial fuzzy evaluations, i.e., of fuzzy numbers $U_1, ..., U_m$ defined on $[0, 1]$, with the normalized fuzzy weights $V_1, \ldots, V_m$, is a fuzzy number U on $[0, 1]$ whose membership function is defined for any $u \in [0, 1]$ as follows

$$U(u) = \max\{\min\{V_1(v_1), ..., V_m(v_m), U_1(u_1), ..., U_m(u_m)\} \mid \sum_{i=1}^{m} v_i u_i = u,\ \sum_{i=1}^{m} v_i = 1,\ v_i, u_i \in [0, 1], i = 1, ..., m\}. \quad (3)$$

For an expert who sets the fuzzy weights, it is not so easy to satisfy the condition of normality. That is why the FuzzME software allows to set only an approximation to the normalized fuzzy weights - fuzzy numbers $W_1, ..., W_m$ on $[0, 1]$ satisfying the following weaker condition

$$\exists w_i \in Ker\ W_i,\ i = 1, ..., n : \sum_{i=1}^{n} w_i = 1. \quad (4)$$

The software removes the potential inconsistence in $W_1, ...,$ W_m and derives the normalized fuzzy weights $V_1, ..., V_m$ from them.

The structure of normalized fuzzy weights and the fuzzy weighted average operation are studied in detail in [9], [10] and [11]. Conditions for verifying normality of fuzzy weights, an algorithm for normalization of fuzzy weights satisfying the condition (4), and an algorithm for calculating fuzzy weighted average, which are all used in the FuzzME software, can be found there. Let us notice, that the used algorithm of fuzzy weighted average calculation is very effective.

3.3.2 Ordered fuzzy weighted average

The fuzzy OWA operator is used in case that the evaluator has special requirements concerning the structure of the partial evaluation. For example, he/she does not want any partial goal to be satisfied poorly. Then the weight of the minimum partial evaluation of any alternative equals 1, and the weights of all its other partial evaluations equal 0. The aggregated fuzzy evaluations then represent the guaranteed fuzzy degrees of fulfillment of all the partial goals (the fuzzy MINIMAX method). Another example of the fuzzy OWA operator usage could be the evaluation of subjects who can choose in which of the three areas they will be mostly involved. The evaluation algorithm should take into account their right of choice. Then, e.g., the results in the area where the subject performs best contribute to the overall evaluation by about one half, results from the second area by one third and results from the area in which the subject was least involved contribute to the overall evaluation only by one sixth. A practical application of such a fuzzy evaluation model could be the overall evaluation of the academic staff with respect to their results in the areas of research, education, and management of education and science.

The ordered fuzzy weighted average represents a fuzzification of the crisp OWA operator by means of the extension principle. Uncertain weights are modeled by normalized fuzzy weights as in the case of fuzzy weighted average.

The following notation will be used to define the ordered fuzzy weighted average: if $(x_1, ..., x_m)$ is a vector of real numbers, then $(x^{(1)}, ..., x^{(m)})$ is a vector in which for all $j \in \{1, \ldots, m\}$, $x^{(j)}$ is the j-th greatest number of $x_1, ..., x_m$.

The ordered fuzzy weighted average of the partial fuzzy evaluations, i.e., of fuzzy numbers $U_1, ..., U_m$ defined on

ISBN: 978-989-95079-6-8

$[0,1]$, with the normalized fuzzy weights $V_1, \ldots, V_m$, is a fuzzy number U on $[0,1]$ whose membership function is defined for any $u \in [0,1]$ as follows

$$U(u) = \max\{\min\{V_1(v_1), ..., V_m(v_m), U_1(u_1), ..., U_m(u_m)\} \mid \sum_{i=1}^{m} v_i u^{(i)} = u,\ \sum_{i=1}^{m} v_i = 1,\ v_i, u_i \in [0,1], i = 1, ..., m\}. \quad (5)$$

The algorithm used to calculate the ordered fuzzy weighted average in the FuzzME software was taken from [12], where fuzzification of the OWA operator is described in detail. The used algorithm is an analogy to the one used for the fuzzy weighted average.

3.3.3 Fuzzy expert system

The fuzzy expert system is used if the relationship between the criteria (or the partial evaluations) and the overall evaluation is complicated. Theoretically, it is possible to model, with an arbitrary precision, any Borel measurable function by means of a fuzzy rule base (properties of Mamdani and Sugeno fuzzy controllers, see e.g. [13]) In reality, the quality of the approximation is limited by the expert's knowledge of the relationship.

If the fuzzy rule base models the relation between values of criteria and the fulfillment of the corresponding partial goal, then the evaluation function is of the following form

$$\begin{aligned} &\text{If } \mathcal{C}_1 \text{ is } \mathcal{A}_{1,1} \text{ and } \ldots \text{ and } \mathcal{C}_m \text{ is } \mathcal{A}_{1,m}, \text{then } \mathcal{E} \text{ is } \mathcal{U}_1 \\ &\text{If } \mathcal{C}_1 \text{ is } \mathcal{A}_{2,1} \text{ and } \ldots \text{ and } \mathcal{C}_m \text{ is } \mathcal{A}_{2,m}, \text{then } \mathcal{E} \text{ is } \mathcal{U}_2 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\text{If } \mathcal{C}_1 \text{ is } \mathcal{A}_{n,1} \text{ and } \ldots \text{ and } \mathcal{C}_m \text{ is } \mathcal{A}_{n,m}, \text{then } \mathcal{E} \text{ is } \mathcal{U}_n \end{aligned} \quad (6)$$

where for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, $(\mathcal{C}_j, \mathcal{T}(\mathcal{C}_j), V_j, M_j)$ are linguistic scales representing the criteria, $\mathcal{A}_{i,j} \in \mathcal{T}(\mathcal{C}_j)$ are their linguistic values, $(\mathcal{E}, \mathcal{T}(\mathcal{E}), [0,1], M_e)$ is a linguistic scale representing the evaluation of alternatives and $\mathcal{U}_i \in \mathcal{T}(\mathcal{E})$ are its linguistic values.

In the FuzzME software, rule bases are defined expertly. The user defines such a rule base by assigning a linguistic evaluation to each possible combination of linguistic values of criteria.

For given values of criteria, a resulting fuzzy evaluation is calculated either by the Mamdani fuzzy inference algorithm or by the generalized Sugeno inference.

In the case of the Mamdani fuzzy inference, the degree h_i of correspondence between the given m-tuple of fuzzy values $(A_1', A_2', \ldots, A_m')$ of criteria and the mathematical meaning of the left-hand side of the *i*-th rule is calculated for any $i = 1, \ldots, n$ in the following way

$$h_i = \min\{hgt(A_1' \cap A_{i,1}), \ldots, hgt(A_m' \cap A_{i,m})\}. \quad (7)$$

Then for each of the rules, the output fuzzy value U_i', $i = 1, \ldots, n$, corresponding to the given input fuzzy values, is calculated as follows

$$\forall y \in [0,1] : U_i'(y) = \min\{h_i, U_i(y)\}. \quad (8)$$

The final fuzzy evaluation of the alternative is given as the union of all the fuzzy evaluations that were calculated for the particular rules in the previous step, i.e.,

$$U' = \bigcup_{i=1}^{n} U_i'. \quad (9)$$

Generally, the result obtained by the Mamdani inference algorithm need not be a fuzzy number. So, for further calculations within the fuzzy model, it must be approximated by a fuzzy number.

The advantage of the generalized Sugeno inference algorithm (see [7]) is that the result is always a fuzzy number. In its first step, the degrees of correspondence h_i, $i = 1, \ldots, n$, are calculated in the same way as in the Mamdani fuzzy inference algorithm.

The resulting fuzzy evaluation U is then computed as a weighted average of the fuzzy evaluations U_i, $i = 1, 2, \ldots, n$, which model the mathematical meanings of linguistic evaluations on the right-hand sides of the rules, with the weights h_i. This is done by the following formula

$$U = \frac{\sum_{i=1}^{n} h_i . U_i}{\sum_{i=1}^{n} h_i}. \quad (10)$$

3.4 Overall fuzzy evaluations, the optimum alternative

The final result of the consecutive aggregation of the partial fuzzy evaluations is an overall fuzzy evaluation of the given alternative. The obtained overall fuzzy evaluations are fuzzy numbers on $[0,1]$. They express uncertain degrees of fulfillment of the main goal by the particular alternatives.

The FuzzME software compares alternatives according to the centers of gravity of their overall fuzzy evaluations. A center of gravity of a fuzzy number U on $[0,1]$ that is not a real number, is defined as follows

$$t_U = \frac{\int_0^1 U(x).x\ dx}{\int_0^1 U(x)\ dx}. \quad (11)$$

If $U = u$ and $u \in \Re$, then $t_U = u$. In the FuzzME software, the optimum alternative is the one whose overall fuzzy evaluation has the largest center of gravity.

At present, the FuzzME software is aimed above all at solving multiple-criteria evaluation problems. To ensure high performance in choosing the optimum alternative, it will be necessary to include in the software other methods of ordering of the fuzzy evaluations in the future. Some approaches are proposed in [7] and further research in this area is planned.

4 Example of a practical application of the FuzzME software

The FuzzME software was tested e.g. on a soft-fact-rating problem of one of the Austrian banks. The problem was solved in co-operation with the Technical University in Vienna (see [14]). The fuzzy model of evaluation represents a part of the creditability evaluation of companies carried out by the bank - the evaluation according to soft (qualitative) data, which complements the evaluation according to hard (quantitative) data. The previous practical experience of the bank

ISBN: 978-989-95079-6-8

showed that it is not good to restrict the evaluation to hard data only.

In total, 62 companies were evaluated by the fuzzy model. The goals tree of the model contained 27 qualitative criteria.

During the testing, two approaches were compared - the original soft-fact-rating model used by the bank and the fuzzy models created in the FuzzME software.

The original evaluation model used simple discrete numeric scales with intuitively set linguistic descriptors for the evaluation according to the particular qualitative criteria. The aggregation of partial evaluation was done by the standard weighted average.

In testing by the FuzzME software, the applied linguistically described numeric scales were analyzed. It was found that in some cases the correspondence between the linguistic and numerical values was not perfect. Two new fuzzy models were formed. The first one used uniform fuzzy scales representing a simple fuzzification of the original numeric scales. The other worked with fuzzy values which tried to model, as closely as possible, the linguistic descriptors used in the original evaluation model. The results of the two models were quite different. At the same time, the normalized crisp weights were replaced by normalized fuzzy weights which correspond better to the expert's knowledge about the importance of criteria.

The subsequent discussion on results of the soft-fact-rating showed that there exist criteria values and combinations of criteria values which signalize a substantial danger that the company will go bankrupt or at least will have problems acquitting the debt. That is why, besides the evaluation of companies based on fuzzy weighted average ("average rating"), a fuzzy expert system was applied to calculate another evaluation ("a risk rate of the company"). The particular rules of the base identified the dangerous combinations of criteria values and assigned to them the corresponding risk rates. The solely use of the original fuzzy model without the fuzzy expert system would have lead to a rating score, which may have underestimated the risk inherent to this company. The use of the fuzzy expert systems offers the possibility to visualize and calculate such additional risk combinations.

Finally, both evaluations were aggregated with the fuzzy MINIMAX method. This method is a special case of a fuzzy OWA operator.The resulting evaluation is the infimum of the fuzzy numbers representing the partial evaluations.

The obtained results showed that the solid theoretical basis of the evaluation fuzzy models formed in the FuzzME software improves the quality of evaluations. Positive experiences with such fuzzy models of evaluation could win over the present-day opponents to the soft-fact-rating in the future.

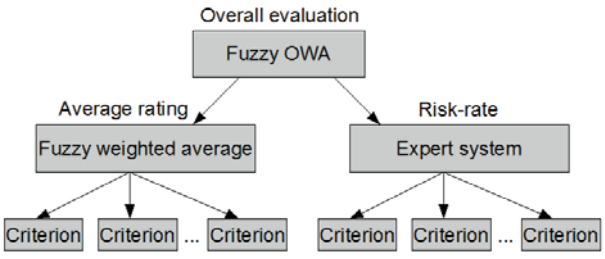


Figure 2: The simplified structure of the used goals tree

5 Conclusion

The software product FuzzME is a result of many years of research in the area of the theory and methods of multiple-criteria fuzzy evaluation. The type of evaluation consistently used in the software corresponds well to the fuzzy sets theory paradigm; the evaluations of alternatives express the fuzzy degrees of fulfillment of given goals. In the FuzzME software, several new methods, algorithms and tools of fuzzy modeling were implemented, e.g.: a structure of normalized fuzzy weights, fuzzy weighted average and ordered fuzzy weighted average operations and algorithms for their calculation, linguistic scales and linguistic variables derived from them. Well-elaborated theoretical basis of the FuzzME software provides a clear interpretation of all steps of the evaluation process and brings understanding of methodology to the user. The FuzzME software is user-friendly. The positive features of the software product proved themselves by solving the mentioned soft-fact-rating problem.

References

[1] R.E. Bellman and L.A Zadeh. Decision-making in fuzzy enviroment. *Management Sci.*, 17 (4):141–164, 2007.

[2] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans.Systems Man Cybernet*, 3 (1):183–190, 1988.

[3] Y. J. Lai and C. L. Hwang. *Multiple Objective Decision Making.* Springer - Verlag, Berlin, Heidelberg, 1994.

[4] H. Rommelfanger. *Fuzzy Decision Support Systeme.* Springer - Verlag, Berlin, Heidelberg, 1988.

[5] INFORM GmbH. FuzzyTECH home page. `http://www.fuzzytech.com/`.

[6] C. Von Altrock. *Fuzzy Logic and NeuroFuzzy Applications in Bussiness and Finance*. Prentice Hall, New Jersey, 1996.

[7] J. Talašová. *Fuzzy methods of multiple criteria evaluation and decision making (in Czech).* Publishing House of Palacký University, Olomouc, 2003.

[8] J. Talašová. NEFRIT - multicriteria decision making based on fuzzy approach. *Central European Journal of Operations Research*, 8(4):297–319, 2000.

[9] O. Pavlačka. *Fuzzy methods of decision making (in Czech).* PhD thesis, Faculty of Science, Palacký Univerzity, Olomouc, 2007.

[10] O. Pavlačka and J. Talašová. Application of the fuzzy weighted average of fuzzy numbers in decision-making models. *New Dimensions in Fuzzy Logic and related technologies, Proceedings of the 5th EUSFLAT Conference, Ostrava, Czech Republic, September 11-14 2007, (Eds. M. Štěpnička, V. Novák, U. Bodenhofer)*, II:455–462, 2007.

[11] O. Pavlačka and J. Talašová. The fuzzy weighted average operation in decision making models. *Proceedings of the 24th International Conference Mathematical Methods in Economics, 13th - 15th September 2006, Plzeň (Ed. L. Lukáš)*, pages 419–426, 2006.

[12] J. Talašová and I. Bebčáková. Fuzzification of aggregation operators based on Choquet integral. *Aplimat - Journal of Applied Mathematics*, 1(1):463–474, 2008. ISSN 1337-6365.

[13] B. Kosko. *Fuzzy thinking: The new science of fuzzy logic.* Hyperion, New York, 1993.

[14] K. Fürst. *Applying Fuzzy Models in Rating Systems.* Department of Statistics and Probability Theory, Vienna University of Technology, Vienna, 2008. Term paper.

ISBN: 978-989-95079-6-8

A Computing with Words path from fuzzy logic to natural language *

Claudio Moraga Enric Trillas

European Centre for Soft Computing, 33600 Mieres, Asturias, Spain
{claudio.moraga, enric.trillas}(at) softcomputing.es

Abstract *– Linguistic variables are reviewed and some aspects of the original definition revisited, showing that they provide a first formal frame to do computing with words in a "narrow sense". A two-levels hierarchy of languages is discussed in order to move to simple cases of computing with words in natural language, supported by linguistic variables.*

Keywords – Computing with words, linguistic variables, meaning modifiers

1 Introduction

In the late 90's, L.A. Zadeh introduced the idea of *computing with words* [1]. Much efforts have been done ever since to face the challenge and to develop the required proper concepts, models and methods to turn the bright initial idea into a sound area of research. Fuzzy logic seems predestined to be an appropriate formal frame to support the above efforts. Among other things, because fuzzy if-then rule bases may be understood as a specification of (the model of) a system in a language close to natural language and each rule, as the specification of a computation based on words, even though at the moment of doing real computations, this is done at the level of numbers. At this time, fuzzy if-then rule bases may be considered as a metaphor of what we would like to achieve by *really* computing with words. In this sense, what is currently known as fuzzy logic may be considered as a particular case of computing with words.

This paper shows, that in the context of fuzzy linguistic variables there are already processes that are representative of computing with words, albeit possibly in a "narrow sense". Furthermore it will be shown that it is possible to build a hierarchical relationship between words based on linguistic variables at the level of fuzzy logic and "words" – (actually short sentences)- at the level of natural language; to move from linguistic variables to "linguistic expressions" and from linguistic modifiers to "meaning modifiers". In analogy to the fact that linguistic modifiers, when applied to linguistic variables constitute possibly the simplest case of computing with words, it will be shown that "meaning modifiers", when applied to "linguistic expressions" may lead to non-trivial linguistic results of non-negligible complexity, in what may be thought of, as computing with "words".

The rest of the paper provides first a set of needed definitions to continue with a discussion about computing with words in fuzzy logic. An extension to computing with "words" in natural language is analyzed in the closing section.

2 Definitions

Definition 1 [1]:
Computing with words and perceptions, or CWP for short, is a mode of computing in which the objects of computation are words, propositions and perceptions described in a natural language.

Definition 2:
Let Ω be a set of (non-ambiguous) words that are used in a given context. The elements of Ω are considered to be pairs (representation, meaning). Representations may be taken from different syntactic domains, but meanings, from a single semantic domain. Moreover, let Γ be a finite set of functions $\{\gamma_1, \gamma_2, \ldots, \gamma_k\}$, with

$$\gamma_i : \Omega^n \rightarrow \Omega$$

The functions $\{\gamma_1, \gamma_2, \ldots, \gamma_k\}$ must be *designed* in such a way, that applied to words of Ω will produce reasonable words of Ω for the given context.

Definition 3:
In this paper when speaking of computing with words, the *objects* of computation are specified in definition 1. The *agents* of computation are functions taken from Γ, as stated in definition 2.

Examples: Let Ω be a subset of English.
If γ_1 is an appropriate "association" function. Then
γ_1(parent, children) = family
If γ_2 is a synonym function. Then
γ_2(pretty) = beautiful
If γ_3 is an antonym function. Then
γ_3(near) = far

These examples show that since the functions in Γ represent some "linguistic transformation", their interpretation should be taken take in account in the line of language precisiation (*à la* Zadeh).

At this stage, it should be mentioned that not under the scope of "computing with words", but simply with the goal of designing good fuzzy if-then rules, work has been done to select or build the most convenient operations connecting premises and the conclusion [2], [3]. In the context of this paper, the given references correspond to a basic contribution in the design of the most convenient functions for the corresponding Γ.

* This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias, Spain) and CICYT (Spain) under project TIN2008-06890-C02-01

ISBN: 978-989-95079-6-8

3 Computing with words in fuzzy logic

One of the key concepts in the development of fuzzy logic is that of a *linguistic variable*, introduced by L.A. Zadeh in 1975 [4]. The value set of a linguistic variable is a set of words –"linguistic terms"– denoting predicates or linguistic labels, and the syntactic domain of those words is constituted by appropriate convex fuzzy sets. "*… in general, a linguistic variable is associated with two rules: (1) a syntactic rule, which may have the form of a grammar for generating the names of the values of the variable; and (2) a semantic rule which defines an algorithmic procedure for computing the meaning of each value"* (See section 2 of Part II of [4])

In what follows, this idea is extended to include a formal grammar and a data base. The grammar will control possible transformations on the fuzzy sets representing the linguistic terms of a linguistic variable. The effect of linguistic modifiers (originally called "hedges of type I" [5]) as well as the generation of antonyms of linguistic terms (which was studied much later [6], [7], [8]) may, for instance, be formalized in terms of productions of a grammar. The data base, on the other hand, is meant to contain the meanings associated to the linguistic terms and their considered possible modifications through the grammar.

It is easy to see that the grammar and the data base have a functionality which is subsumed by that of the functions presented in Definition 2.

For the representation of linguistic terms as fuzzy sets, if no further information is available, a trapezoidal shape is appropriate since an agreement among users about the subset of the universe where the predicate is indeed valid, defines the core of the fuzzy set and the agreement about the subset of the universe where the predicate is not at all valid, defines its co-support. Our human experience with predicates indicates that the transitions between co-support and core are continuous and monotone. Therefore, if no further information is available, linear transitions leading to a piecewise linear convex normalized fuzzy set is an appropriate first choice. If the core contains only one point, then the trapezium reduces to a triangle. Notice that the above analysis does not require the trapezium (triangle) to be symmetric.

Trapezoidal fuzzy sets are simple to represent as an ordered quadruple of the ordinates of their four corners (t_1, t_2, t_3, t_4). The interval between the first and the fourth corner represents the support of the fuzzy set. (If non-normalized trapezoidal fuzzy sets are considered, then a preceding scaling factor is needed.)

Another basic assumption, normally associated to linguistic variables, is that all linguistic terms of the variable constitute a partition of unity, *i.e.*, at every point of the universe of discourse, the membership degrees of the linguistic terms add up to 1.

Definition 4:

A linguistic modifier is a unary function in the set of fuzzy sets representing linguistic terms. A linguistic modifier LM is *compressing* if for any fuzzy set A representing a linguistic term, $LM(A) \subset A$, and is *expanding* if $A \subset LM(A)$. There are however linguistic modifiers that are neither compressing nor expanding.

In what follows, to keep the notation as simple as possible, the label of a linguistic term will also be used to denote the membership function of the fuzzy set representing it.

Let $T := (t_1, t_2, t_3, t_4)$

$LM(T) := ((t_1+t_2)/2, t_2, t_3, (t_3+t_4)/2)$ is a compressing modifier and in what follows will be given the meaning *more_strictly*(T).

$LM(T) := (t_1, (t_1+t_2)/2, (t_3+t_4)/2, t_4)$ is an expanding modifier and in what follows will be given the meaning *roughly*(T).

Lemma 1:

Let $T := (t_1, t_2, t_3, t_4)$

Define $T_{left} := ½(t_1, (t_1+ t_2)/2, (t_1+ t_2)/2, t_2)$

$T_{right} := ½(t_3, (t_3+ t_4)/2, (t_3+ t_4)/2, t_4)$

Then:

$T = T_{left} + more\_strictly(T) + T_{right}$

Proof: By adding point-to-point the piecewise linear segments.

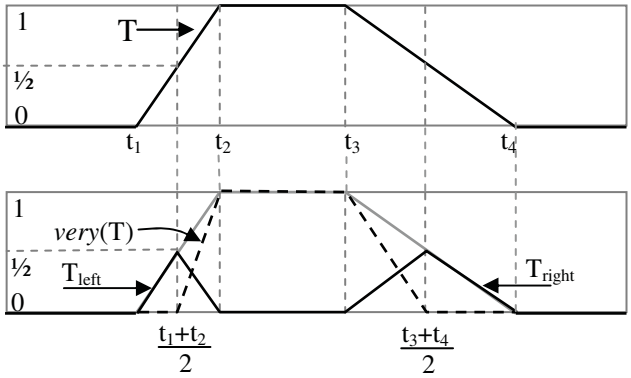


Figure 1**:** The relationship among the fuzzy sets T, *more_strictly*(T), T_{left} and T_{right}

It is easy to see that for all $x < t_3$, T_{right} takes the value 0. Moreover, since $(t_1 + t_2)/2$ is equidistant of t_1 and t_2, then at this point, T takes the value ½. (See Figure 1). Moreover between t_2 and t_3, both T_{left} and T_{right} have membership degree 0, meanwhile T and *more_strictly*(T) have membership degree 1, therefore in this interval the claim trivially holds. Table 1 summarizes the relevant membership values at t_1, $(t_1 + t_2)/2$ and t_2.

Table 1: relevant membership values

	t_1	$(t_1 + t_2)/2$	t_2
T	0	½	1
more_strictly(T)	0	0	1
T_{left}	0	½	0
more_strictly(T) + T_{left}	0	½	1

In the interval $[t_1, t_2]$, both *more_strictly*(T) and T_{left} are piecewise linear; therefore their sum will also be piecewise linear. In the subinterval $[t_1, (t_1+t_2)/2]$, since the membership value of *more_strictly*(T) is 0, the sum will be given by T_{left}, which has a representation as a straight line segment with slope $1/(t_2-t_1)$. In the subinterval $[(t_1+t_2)/2, t_2]$, both *more_strictly*(T) and T_{left} have a representation as straight line segments, therefore also their sum. It is simple to see

ISBN: 978-989-95079-6-8

from Figure 1, that the sum segment runs from the point $\langle(t_1+t_2)/2, 1/2\rangle$ to the point $\langle t_2, 1\rangle$, with slope $1/(t_2-t_1)$. Since both sum segments share a common point and have the same slope, they constitute a single straight line segment with slope $1/(t_2-t_1)$ in the whole interval $[t_1, t_2]$. This corresponds with the left part of T. A similar analysis will prove that in $[t_3, t_4]$ the sum of *more_strictly*(T) plus T_{right} returns the corresponding part of T. For reasons that will become clear below, no attempt will be done here to associate a meaning to T_{right} and T_{left}.

Corollary 1.1:
Assume that T is a rectangular trapezium. Without loss of generality let T := (t_1, t_2, t_3, t_3). Then :

$T = T_{left} + more\_strictly(T)$

since T_{right} has zero-support.

Example 1:
Consider a linguistic variable with three linguistic terms building a partition of unity, as shown in Figure 2(a). Furthermore, let it be assumed that some experiments done using this representation –(model)- of the variable, made advisable a refinement of the middle linguistic term, and that this is done by applying Lemma 1 (with $t_2 = t_3$). This is shown in Figure 2(b). Solving the superposition of the small triangles with the respective trapeziums leads to a new model of the linguistic variable, as shown in Figure 2(c).

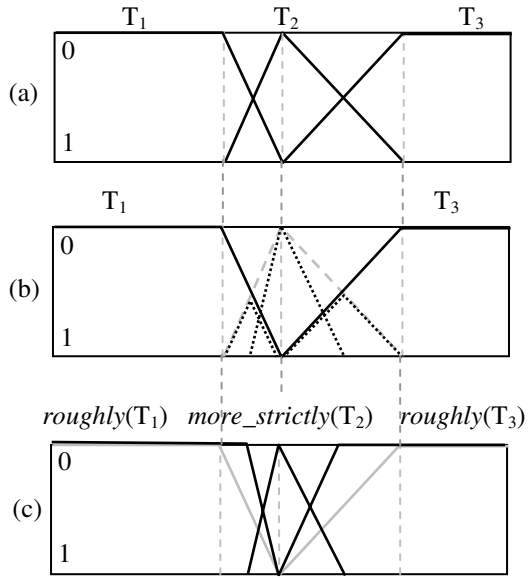


Figure 2**:** (a) Initial model of the linguistic variable
(b) Effect of applying Lemma 1 to T_2
(c) The resulting new model for the linguistic variable

An analysis of the process shows that T_2 changed into *more_strictly*(T_2). Furthermore, the core of T_1 was expanded by one half of the support corresponding to the right wing of the trapezoidal representation. From Definition 4 follows that it represents *roughly*(T_1). Similarly, T_3 turns into *roughly*(T_3). In summary, a new representation of the linguistic variable, which preserves the partition of unity, was obtained through a refinement process by using appropriate linguistic modifiers. This represents (possibly the simplest) elementary operations in the context of computing with words.

Case study 1:
In the context of linguistic variables with structure of a partition of unity, *two* consecutive linguistic terms T_j and T_{j+1} will be considered, which will be refined under Lemma 1.

Let T_j := (t_1, t_2, t_3, t_4)
and T_{j+1} := (t_3, t_4, t_5, t_6)
Lemma 1 generates:

$more\_strictly(T_j)$:= $((t_1+ t_2)/2, t_2, t_3, (t_3+ t_4)/2)$
$(T_j)_{right}$:= $½(t_3, (t_3+ t_4)/2, (t_3+ t_4)/2, t_4)$
$more\_strictly(T_{j+1})$:= $((t_3+ t_4)/2, t_4, t_5, (t_5+ t_6)/2)$
$(T_{j+1})_{left}$:= $½(t_3, (t_3+ t_4)/2, (t_3+ t_4)/2, t_4)$

(Lemma 1 also generates $(T_j)_{left}$ and $(T_{j+1})_{right}$, which are however not in the focus of this case study).
Notice that $(T_j)_{right} = (T_{j+1})_{left}$ therefore
if $T' = (T_j)_{right} + (T_{j+1})_{left}$ then
T' := $(t_3, (t_3+ t_4)/2, (t_3+ t_4)/2, t_4)$
It is easy to see that T' is an isosceles triangle. To T' the following *specification of meaning* can be associated:

$T' = between(more\_strictly(T_j), more\_strictly(T_{j+1}))$

and, the other way around, the operation *between* is defined as the structure of T'. See Figures 3(a) and 3(b).

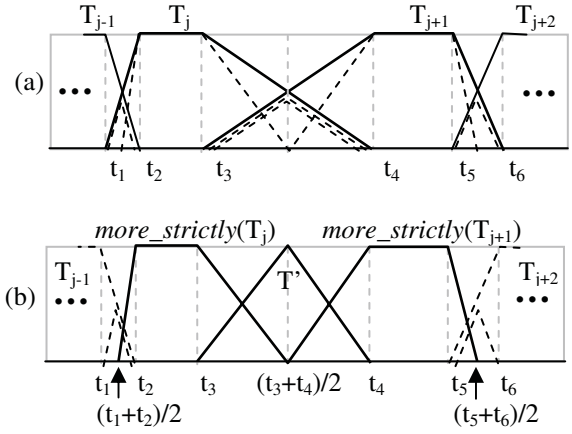


Figure 3**:** (a) Two neighbor linguistic terms and their refinement with Lemma 1 (dotted).
(b) Effect of the superposition of the overlapping side triangles generated by the refinement.

From $T' = between(more\_strictly(T_j), more\_strictly\ (T_{j+1}))$ it may be concluded that there exists a $\gamma \in \Gamma$ such that

$\gamma = between \circ (more\_strictly \times more\_strictly)$

and $T' = \gamma(T_j, T_{j+1})$

Notice that if all linguistic terms of a linguistic variable are refined with Lemma 1, then all "new" linguistic terms may be obtained with the composition of *between* and the Cartesian product *more_strictly* × *more_strictly*. A refinement of all linguistic terms after Lemma 1, may be related to wavelets [9], but almost duplicates the number of linguistic terms. Partial refinements, however, were studied in [10] under the scope of a metasemantics.

ISBN: 978-989-95079-6-8

Before studying a next case, two elementary operations will be introduced.

Definition 5:
Let $T_j := (t_1, t_2, t_3, t_4)$
Then, the operation *less than or equal to*, abbreviated "LE" is defined for all x in the universe of discourse, as:

$$(LE(T_j))(x) = \begin{cases} 1 & \text{if } x \le t_3 \\ T_j & \text{if } x > t_3 \end{cases}$$

Similarly, the operation *more than or equal to*, abbreviated "ME" is defined as:

$$(ME(T_j))(x) = \begin{cases} 1 & \text{if } x \ge t_2 \\ T_j & \text{if } x < t_2 \end{cases}$$

Case study 2:
Consider a linguistic variable with more than 3 linguistic terms and refine some linguistic term in the middle. What will be the effect upon its non-refined neighbors?

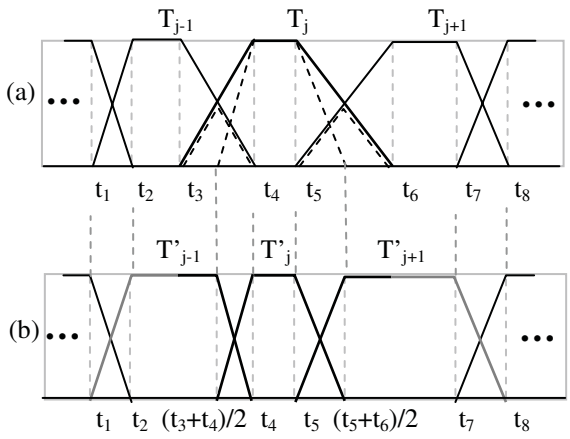


Figure 4 **:** (a) A linguistic variable with the prepared refinement of the j-th linguistic term.
(b) Effect of refining T_j upon its neighbors.

Figure 4 shows the linguistic terms before and after refining the j-th term. From Lemma 1 follows that:

$$T'_j = more\_strictly(T_j)$$

An interpretation of T'_{j+1} is not straight forward. Fig. 4(b) shows however that at the left hand side, the core of T'_{j+1} increased by one half of the support of the left wing, *i.e.*, by $\frac{1}{2}(t_6-t_5)$, with respect to the original core of T_{j+1}. Notice that this is the same (partial) effect caused by the operation $roughly(T_{j+1})$; however, the right hand side of T'_{j+1} does not change as compared to T_{j+1}. This may be given a formal expression as follows, for all x in the universe of discourse:

$$(T'_{j+1})(x) = minimum(roughly(T_{j+1}), LE(T_{j+1}))$$

or "in words",

$$T_{j+1} = roughly(T_{j+1})\ and\ LE(T_{j+1})$$

where *and* is realized with the operation *minimum*.

Similarly,

$$(T'_{j-1})(x) = minimum(roughly(T_{j-1}), ME(T_{j-1}))$$

or "in words",

$$T_{j-1} = roughly(T_{j-1})\ and\ ME(T_{j-1})$$

Example 1 and the two above cases show that at the lowest level, that of terms of linguistic variables, transformations may be formalized, which represent the possibly simplest cases of computing with words in narrow sense.

4 First steps towards natural language

When working with formal languages, "words" are build with "symbols" of an alphabet. Quotations are being used, since the "symbols" may be words of a reference language and, consequently the "words" will (normally) be (meaningful) sentences of the reference language. This model of a hierarchy of languages will be used in an attempt to transfer or imitate the elementary operations discussed above to short statements in natural language. These "words" will be written in square brackets. Furthermore the notation "linguistic statement" instead of linguistic variable, and "meaning modifier" instead of linguistic modifier will be used. Finally it becomes apparent that "words" will no longer be just pairs (representation, meaning), but at least triples (representation, structure, meaning). Representation is concerned with the rules of writing, meanwhile the structure refers to the grammatical ordering of the reference words to support a meaning.

Example 3:
Consider the "word" [*The weather is improving*], where the linguistic statement has the value set {hardly a bit, slightly, steadily, strongly}. How could the meaning of this "word" be emphasized?

i) *Use* the "word" consistently within the value subset {steadily, strongly}

ii) Design a meaning modifier MM_1 which integrates explicitly in the "word" the desired predicate from the value set, preserving the structure:
MM_1 (***steadily***, [*The weather is improving*]) =
[*The weather is* ***steadily*** *improving*]

It may be observed that this modifier changes the representation of the "word", preserves the grammar of the "word" and emphasizes its meaning.

iii) Design a meaning modifier MM_2 which integrates an additional reference word, taken from a pre-defined "subset of emphasizers" of the corresponding natural language, preserving the structure, to produce an enhancement:

MM_2 (***indeed***, [*The weather is improving*]) =
[*The weather is* ***indeed*** *improving*]

Notice that it is only claimed that both [*The weather is* ***steadily*** *improving*] and [*The weather is* ***indeed*** *improving*]

ISBN: 978-989-95079-6-8

emphasize the meaning of [*The weather is improving*], but not that they are synonyms.

Example 4:
Consider the "word" [*Yesterday it rained*]. A value set for the linguistic statement may be thought of as $D_1 \times D_2 \times D_3$, where D_1 = {in the morning, at midday, in the afternoon, in the evening, at night}, D_2 = {a couple of minutes, several hours, the whole day} and D_3 = {a few drops, moderately, cats-and-dogs}. Notice that D_1 and D_2 may be associated both to *yesterday* and to *rain*, meanwhile D_3 is clearly related only to *rain*.

i) It is requested that [*Yesterday it rained*] be emphasized. The *rained*-component may be emphasized in a way similar to that discussed in example 3. Can *yesterday* be emphasized? At the level of natural language, it is possible to design a 2-place meaning modifier MM_1 giving:

MM_1([*December the 1st*], [*Yesterday it rained*]) =
[*Yesterday December the 1st it rained*],

which may be considered as a possible way of emphasizing *yesterday*.

ii) In the context of [*Yesterday it rained*] D_1, D_2 and D_3 may be seen as value sets of the linguistic variables "when", "how_long" and "how_much" (at the reference level), respectively. As linguistic variables, they have grammars that, among other things, may change number, shape and distribution of the fuzzy sets representing the corresponding linguistic terms, as well as a database keeping track of their meanings. Assume that the corresponding grammars are G_1, G_2 and G_3. Let $r_{2,i}$ and $r_{3,j}$ represent the i-th and j-th rules of G_2 and G_3, respectively, such that:

$r_{2,i}(D_2)$ = {a couple of minutes, *half an hour*, several hours, the whole day}

$r_{3,j}(D_3)$ = {a few drops, moderately, *a lot*, cats-and-dogs}

Some elements of the new domains may have the same labels as in the original domains, but will possibly be different since at least some of them are shifted and compressed, because in both cases, a new linguistic term was introduced. (Notice here the importance of the data base with meanings).
Then it is possible to use linguistic modifiers at the low level to build meaning modifiers MM (at the upper level) to generate new, richer "words" as, for instance,

MM[*Yesterday it rained*] =
[*Yesterday it rained quite a lot about half an hour almost at midday*]

It may be observed that the meaning modifier reordered the domains as $r_{3,j}(D_3) \times r_{2,i}(D_2) \times D_1$, made use of the new generated linguistic terms "half an hour" and "a lot", applied the linguistic modifiers "quite" (to "a lot") "about" (to "half an hour") and "almost" (to "at midday"). Fuzzy logic offers sound methods to perform all these reference level operations, which put together, formally specify the meaning modifier.

Analysis of the meaning modification
[*quite a lot about half an hour almost at midday*]:

Recall that the value set for [*Yesterday in rained*] was specified as the Cartesian product of three domains. Therefore, its "value" is a triple, whose components are possibly modified linguistic terms within the prevailing versions of the corresponding domains. Therefore, each one of the linguistic terms may be analyzed separatedly and the required linguistic modifiers, *designed* to satisfy the requirements of the use of the language.

i) Design of *quite*, to produce the modified linguistic term *quite a lot*.
It should be noticed that *quite* is a rather complicated linguistic modifier, since it is very context dependent: in American English it seems to be used mostly to emphasize, in a way similar to *more_strictly*, meanwhile in British English it may also be ironically used to mean exactly the contrary. Furthermore, *quite* is also used as an idiom, like in "*quite* a few", possibly meaning the same as "*quite* a lot". In the case under analysis however, it seems that *quite* is intended to be used to suggest rather more precisely, what is expressed by "a lot" alone. If this is the case, then a compressing linguistic modifier should be used, but with possibly a smaller amount of compression than in the case of *more_strictly*.
A general analysis leading to a proposal for a prototype of *quite* is shown in Figure 5.

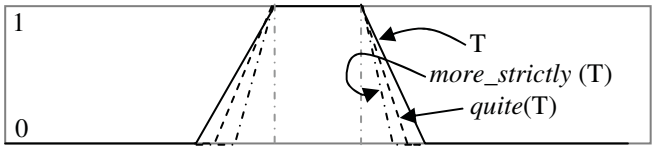


Figure 5**:** A possible prototype for *quite* as a linguistic modifier, which is less compressing than *more_strictly*

Observations on the *use* of *quite* in the context where it will be applied and in similar contexts will be needed to properly tune this modifier. Notice that it may well be the case that the contexts at the reference level and at the natural language level pose different constraints for the design of the modifier.

ii) Design of *about*, to produce the modified linguistic term *about half an hour*.
In the use of the language, *about* half an hour, appears to have a meaning close to that of *around* half an hour or *roughly* half an hour. Experiences with numerical approximation with fuzzy numbers indicate that *around* is an expanding modifier that preserves the core –(which in this case contains only one point, since the fuzzy set is triangular)– and increases the support. It seems reasonable to extend this modifier to trapezoidal fuzzy sets in the same way. On the other hand, *roughly*, as defined earlier, expands the core and preserves the support. If the meaning of *about* half an hour is indeed close to that of *around* half an hour and *roughly* half an hour, then the corresponding modified fuzzy sets should exhibit a high degree of "family resemblance" [11] and this can be used as an additional guiding constraint to design *about*.
A proposal for a prototype of *about* is illustrated in Figure 8 with respect to some abstract T (to more clearly show the effect of the (known) linguistic modifiers and of the proposed prototype).

ISBN: 978-989-95079-6-8

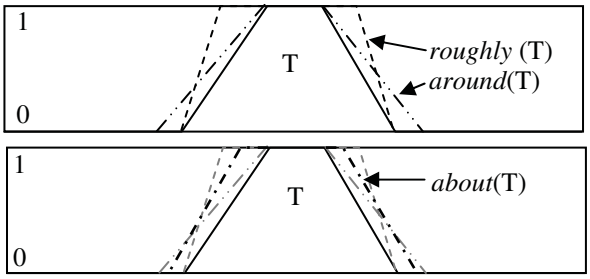


Figure 6: A possible prototype for *about*, for some abstract T, as related to *roughly* and *around*

As in the former case, observations on the *use* of the modifier in the target context will be needed to properly tune it.

iii) Design of *almost*, to produce the modified linguistic term *almost at midday*.
The use of the language, *e.g.* "the glass is *almost* empty" or "Peter was *almost* crazy with so many problems", indicates that this is a one-sided shifting and possibly expanding modifier, which applied to a fuzzy set representation of a linguistic term, displaces and possibly expands the fuzzy set towards the "less than" side. (See Figure 7). It is simple to see that *almost at midday* and *around midday* will have some "family resemblance", but to a clear weaker degree than the one discussed in the former paragraph.

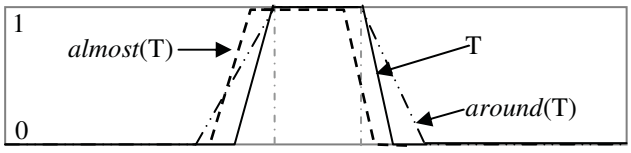


Figure 7: A possible prototype for *almost*, as compared to *around*, applied to some abstract T

Observations on the *use* of all above modifiers in the actual target context, including experimentation at the level of the language, will be needed to ensure their empirical correctness. See, for instance [12], [13], [14].

5 Conclusions

Within the formalism of linguistic variables in fuzzy logic, linguistic modifiers represent a very simple case of computing with words. Some new elementary operations have been defined, that allow new non trivial instances of low level computing with words. The paper shows that it is possible to extend these ideas to "linguistic statements" (short sentences) and "meaning modifiers" in natural language. The "linguistic statements" become the new "words" (as is the case of programming languages when considered as context free formal languages). This is done however keeping a solid base on linguistic terms and linguistic modifiers. The hierarchy of languages suggests that the more about computing with words is learnt in fuzzy logic, the more can be projected to natural languages to do computing with "words".

Acknowledgement

The authors thank the reviewers for their constructive criticism and the suggestions for the improvement of the paper.

References

[1] L.A. Zadeh, From computing with numbers to computing with words. From manipulation of measurements to manipulation of perceptions. *IEEE Trans. Circuits and Systems* **45** (1): 105-119, 1999

[2] C. Moraga, A. Pradera and E. Trillas, Evolutionary tuning of fuzzy if-then rules at the level of operations: A proposal. *Proc. III Congr. Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados MAEB'2003*, 530-537. Press University of Oviedo, 2003

[3] E. Trillas, On some features of fuzzy logic. *Proc. 27th Annual Meeting of the North American Fuzzy Information Processing Society NAFIPS.* 2008

[4] L.A. Zadeh, The concept of linguistic variable and its application to approximate reasoning. *Information Sciences* **8**: 199-249; 301-357; **9**: 43-80, 1975

[5] L.A. Zadeh, A fuzzy-set-theoretic interpretation of hedges. Memorandum M-335, Electronics Research Laboratory, University of California, Berkeley, 1972

[6] E. Trillas and T. Riera, Towards a representation of synonyms and antonyms by fuzzy sets. *Busefal* V:30-45, 1981

[7] E. Trillas, S. Cubillo and E. Castiñeira, On antonymy from fuzzy logic. *Proc. 9th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge Based Systems (IPMU'00)*, 79-84, 2000

[8] E. Trillas , C. Moraga, S. Guadarrama, S. Cubillo and E. Castiñeira, Computing with Antonyms. *Studies in Fuzziness and Soft Computing* **217**: 133-153, Springer, 2007

[9] A. Rodriguez de Soto, Building a hierarchical set of fuzzy rules using Wavelet Theory. *Proc. 9th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge Based Systems (IPMU'00)*, 1764-1769, Madrid, 2000.

[10] C. Moraga, A metasemantics to refine fuzzy if-then rules. *Proc. IEEE Int. Symp. on Multiple-valued Logic*, 148-153. IEEE-CS-Press, 2004

[11] E. Trillas, C. Moraga and A. Sobrino, On "family resemblances" with fuzzy sets. *Proc. IFSA'2009*, IFSA-Press, 2009

[12] J. Hellendorn, Reasoning with Fuzzy Logic. Dissertation, Delft University of Technology, The Netherlands, 1990

[13] V. Novak, A horizon shifting model of linguistic hedges for approximate reasoning, *Proceedings of the 5th IEEE International Conference on Fuzzy Systems,* vol. 1, 423–427, IEEE-CS-Press, 1996

[14] P.J. Macvicar-Whelen, Fuzzy sets, the concept of height, and the hedge 'very'. *IEEE Trans. On Systems, Man, and Cybernetics*, SMC-8, 507-511, 1978

ISBN: 978-989-95079-6-8

Approximation properties of inverse fuzzy transforms over a residuated lattice

Martina Daňková Martin Štěpnička
Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech republic
Email: {martina.dankova,martin.stepnicka}@osu.cz

Abstract— *This contribution focuses on inverse fuzzy transforms (shortly inverse F-transforms) over residuated lattices introduced by I. Perfilieva and their approximation properties. We will try to reduce some requirements used in the original work to prove Approximation theorem. Moreover, we show in which sense F-transforms are the best approximations.*

Keywords— Fuzzy transform, Approximation, Extensional relation.

1 Introduction

In [1], F-transforms were introduced basically in three forms: the first (ordinary) was constructed over the ordinary algebra of reals[1], while the other two were constructed over the residuated lattice. Note that up to now, most of the applications of F-transforms, e.g. ordinary/partial differential equations [2]/[3], image compression/fusion (see [4]/[5, 6]), data analysis [7] etc., employ only the ordinary one. The main reason follows from the character of the inverse F-transforms, where the ordinary F-transform average the function values. Whereas the other two approximate them from above and from below, respectively. And since these F-transforms (over the residuated lattice) can be viewed as formalizations of a collection of *graded fuzzy rules* [8], they are primarily suitable for applications, which use fuzzy "IF-THEN" rules.

In this contribution, we will focus on the inverse F-transforms over the residuated lattice and show their approximation properties. Moreover, we will provide a simpler way to get the approximation theorem than the one in [1] (Theorem 10 and 11). Thus, we would like to brush up this topic and make it more attractive for a broader community.

Note that we present mainly an excerpt of a larger manuscript [9] that has been submitted for publication. For background information and proof details, readers are referred to this upcoming article.

2 Basic definitions and overview of the known results.

2.1 Basic definitions

Definition 1 *A residuated lattice on L is an algebra*

$$\mathcal{L} = \langle L, \vee, \wedge, *, \to, \mathbf{0}, \mathbf{1} \rangle \tag{1}$$

with four binary operations and two constants such that

- $\mathcal{L} = \langle L, \vee, \wedge, \mathbf{0}, \mathbf{1} \rangle$ *is a lattice with the largest element* $\mathbf{1}$ *and the least element* $\mathbf{0}$ *w.r.t. the lattice ordering* $\leq$,
- $\mathcal{L} = \langle L, *, \mathbf{1} \rangle$ *is a commutative semigroup with the unit element* $\mathbf{1}$*, i.e.* $*$ *is commutative, associative, and* $\mathbf{1} * x = x$ *for all* $x \in L$,
- $*$ *and* $\to$ *form an adjoint pair, i.e.*

$$z \leq (x \to y) \quad \textit{iff} \quad x * z \leq y \quad \textit{for all } x, y, z \in L.$$

In the sequel, let us assume $\mathcal{L}$ be a residuated lattice of the form (1). Moreover, we define the biresidual operation (biresiduum)

$$x \leftrightarrow y = (x \to y) \wedge (y \to x).$$

Throughout the whole text, we will deal with fuzzy relations whose membership functions take values from the support of $\mathcal{L}$ and we denote this fact by $\subseteq$. Let $M_1, \ldots, M_n$ be some nonempty sets of objects, $\mathcal{R}_I = \{R_i \subseteq M_i^2\}_{i \in I}$ be a system of binary fuzzy relations ordered by $I = \{1, \ldots, n\}$, $f \subseteq M_1 \times \ldots \times M_n$ be an n-ary fuzzy relation.

Convention 2 *For the sake of brevity, we will denote* $M_1 \times \cdots \times M_n$ *by* $M^{(n)}$*. Moreover, we will write* $R(\bar{x}, \bar{y}) = R_1(x_1, y_1) * \cdots * R_n(x_n, y_n)$.

Definition 3 *We say that f is extensional w.r.t. $\mathcal{R}_I$ if*

$$R(\bar{x}, \bar{y}) * f(\bar{x}) \leq f(\bar{y}),$$

for each $\bar{x}, \bar{y} \in M^{(n)}$.

Let us recall that a fuzzy $*$-equivalence (also known as similarity or indistinguishability) $R \subseteq M^2$ (M be some nonempty set of objects) is reflexive, symmetric and transitive w.r.t. $*$. Observe that $\leftrightarrow$ is $*$-equivalence on L. In the following we will omit specification $*$ whenever it will be clear from the context.

Additionally, we recall notions used in the original approximation theorems in [1].

Definition 4 *A system of fuzzy sets* $\{A_1, \ldots, A_k\}$ *on a nonempty set M establishes a semi-partition if*

$$\bigvee_{x \in M} A_i(x) * A_j(x) \leq \bigwedge_{x \in M} A_i(x) \leftrightarrow A_j(x),$$

for each $i, j \in \{1, \ldots, k\}$.

[1] Due to a nature of the used operations, it can be reinterpreted over ŁΠ-algebra.

ISBN: 978-989-95079-6-8

2.2 $F^{\uparrow(\downarrow)}$-transforms

Let us recall inverse F-transforms from [1] generalized for the n-dimensional case[2]:

$$\begin{aligned}
&\text{Assume } f \subsetsim M^{(n)}, C \subseteq M^{(n)}, A_{j_i} \subsetsim M_i,\\
&\qquad \forall j \in J = \{1,\dots,k\}, \forall i \in I = \{1,\dots,n\},\\
&\qquad \text{and } \mathbf{j} \text{ be abbreviation for } [j_1,\dots,j_n],\\
&f^{\downarrow}_{F,k}(\bar{x}) = \bigvee_{\mathbf{j}\in J^n} A_{\mathbf{j}}(\bar{x}) * [\bigwedge_{\bar{c}\in C} A_{\mathbf{j}}(\bar{c}) \to f(\bar{c})],\\
&f^{\uparrow}_{F,k}(\bar{x}) = \bigwedge_{\mathbf{j}\in J^n} A_{\mathbf{j}}(\bar{x}) \to [\bigvee_{\bar{c}\in C} A_{\mathbf{j}}(\bar{c}) * f(\bar{c})], \text{ where}\\
&A_{\mathbf{j}}(\bar{x}) = A_{j_1}(x_1) * \ldots * A_{j_n}(x_n), \forall \mathbf{j} \in J^n.
\end{aligned}$$

In [1], $f^{\downarrow}_{F,k}$ is called inverse F$^{\downarrow}$-transform and $f^{\uparrow}_{F,k}$ inverse F$^{\uparrow}$-transform. Moreover, the expression in the square brackets above is denoted by $F^{\uparrow(\downarrow)}_{\mathbf{j}}$ in the case of $f^{\uparrow(\downarrow)}_{F,k}$ and known as $\mathbf{j}$-th component of F$^{\uparrow(\downarrow)}$-transform.

Below, we overview known results from [1] that are in the center of our interest:

1. From the inequality $a * (a \to b) \le b \le a \to (a * b)$, it follows that for an arbitrary f:
$$f^{\downarrow}_{F,k}(\bar{x}) \le f(\bar{x}) \le f^{\uparrow}_{F,k}(\bar{x}),\ \forall \bar{x} \in C. \tag{2}$$
2. The $\mathbf{j}$-th components of F$^{\uparrow}$-transform and F$^{\downarrow}$-transform are the least elements of the following respective sets:
$$S_{\mathbf{j}} = \{a \in M^{(n)} |\ A_{\mathbf{j}}(\bar{x}) \le f(\bar{x}) \to a,\ \forall \bar{x} \in C\}, \tag{3}$$
$$T_{\mathbf{j}} = \{a \in M^{(n)} |\ A_{\mathbf{j}}(\bar{x}) \le a \to f(\bar{x}),\ \forall \bar{x} \in C\}. \tag{4}$$
3. A core of the approximation theorems inheres in the following inequality:
$$\bigvee_{\mathbf{j}\in J^n} A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{x}) \le f^{\uparrow(\downarrow)}_{F,k}(\bar{x}) \leftrightarrow f(\bar{x}),\ \forall \bar{x} \in C. \tag{5}$$
Remark that a closeness of elements is expressed using $\leftrightarrow$, which is the dual concept to pseudo-metric as shown in [10], and naturally, the estimation goes there from below.

Below, we summarize the requirements that were used in [1] to prove (5), i.e., the estimation of the equivalence of $f^{\downarrow(\uparrow)}_{F,k}$ to the original function f:

	Requirements in [1] (Theorems 10 and 11)
1.	$\mathcal{L}$ be a BL-algebra,
2.	each $A \in \{A_{j_i}\}_{j\in J, i\in I}$ be normal,
3.	$\{A_{1_i},\dots,A_{k_i}\}$ forms a semi-partition $\forall i \in I$,
4.	f be extensional w.r.t. $\mathcal{P}_I = \{P_i\}_{i\in I}$ defined by $P_i(x,y) = \bigwedge_{j\in J} A_{j_i}(x) \leftrightarrow A_{j_i}(y)$, for all $i \in I$,
5.	$E = \{\bar{e}_j \in M^{(n)} |\ P_i(x, e_{j_i}) = A_{j_i}(x), i \in I, j \in J\}$ be subset of C (see p. 1013 in [1]).

[2] Note that in the original work, f was a mapping from a finite subset of L into L (support of the fixed residuated lattice $\mathcal{L}$), where $L = [0,1]$. Since we measure a precision of the approximation using values of L and we do not use specific properties of $[0,1]$, this determination is not needed in our work.

3 Approximation properties

3.0.1 Formulation of the problem and basic properties

Let $\mathcal{L}$ and $\mathcal{R}_I = \{R_i \subsetsim M_i^2\}_{i\in I}$ be as specified in Subsection 2.1, and $\varepsilon \in L$. Moreover, we assume to have only a partial information about $f \subsetsim M^{(n)}$ in the form of a set of samples $\{f(\bar{c}) |\ \bar{c} \in C \subseteq M^{(n)}\}$. The problem (ApproxF) that we are going to solve is formulated as follows:

$$\begin{aligned}
&\text{Find } D \subseteq M^{(n)} \text{ and } \tilde{f} \in \mathbb{M}^D_{\vee(\wedge)} \text{ such that}\\
&\qquad \varepsilon \le f(\bar{x}) \leftrightarrow \tilde{f}(\bar{x}),\ \forall \bar{x} \in C,
\end{aligned}$$

where

$$\mathbb{M}^D_{\vee} = \Big\{ \bigvee_{\bar{d}\in D} (R(\bar{d},\bar{x}) * g(\bar{d})) |\ g \subsetsim M^{(n)} \Big\} \quad \text{and}$$
$$\mathbb{M}^D_{\wedge} = \Big\{ \bigwedge_{\bar{d}\in D} (R(\bar{x},\bar{d}) \to g(\bar{d})) |\ g \subsetsim M^{(n)} \Big\}.$$

Due to terminology introduced in [11], we approach this problem from the view point of fuzzy relations of a special structure, the so-called *normal forms*:

$$\mathrm{DNF}^P_h(\bar{x}) = \bigvee_{\bar{p}\in P} (R(\bar{p},\bar{x}) * h(\bar{p})) \quad \text{and} \tag{6}$$
$$\mathrm{CNF}^P_h(\bar{x}) = \bigwedge_{\bar{p}\in P} (R(\bar{x},\bar{p}) \to h(\bar{p})), \tag{7}$$
$$\text{where } h \subsetsim M^{(n)}, P \subseteq M^{(n)}, \bar{x} \in M^{(n)}, \tag{8}$$

are called the disjunctive and conjunctive normal forms for h w.r.t. P, respectively. Hence,

$$\mathbb{M}^D_{\vee} = \Big\{ \mathrm{DNF}^D_g |\ g \subsetsim M^{(n)} \Big\} \quad \text{and}$$
$$\mathbb{M}^D_{\wedge} = \Big\{ \mathrm{CNF}^D_g |\ g \subsetsim M^{(n)} \Big\}.$$

It is possible to rewrite inverse F-transforms using normal forms if we take $\{R_i\}_I$ and D:

- Assume $M^{(n)}$ contains sufficiently many elements to define set D.
- Take $D = \{\bar{d}_j \in M^{(n)} |\ d_{j_i} \ne d_{j'_i},\ \forall j,j' \in J, \forall i \in I\}$.
- Define $R(\bar{x},\bar{d}_j) = R(\bar{d}_j,\bar{x}) = A_{\mathbf{j}}(\bar{x}), \forall j \in J, \mathbf{j} \in J^n$ and $R(\bar{x},\bar{y}) = \mathbf{0},\ \forall \bar{y} \in M^{(n)} \setminus D$.

In the case of lacking elements to construct set D, which can hypothetically arise because there is no restriction to the number of sets A_{j_i}, we can artificially extend $M^{(n)}$ to reach the required cardinality. Then we obtain that $\mathrm{DNF}^D_{\mathrm{CNF}^C_f}$ corresponds to an inverse $F^{\downarrow}$-transform $f^{\downarrow}_{F,k}$ and $\mathrm{CNF}^D_{\mathrm{DNF}^C_f}$ to an inverse $F^{\uparrow}$-transform $f^{\uparrow}_{F,k}$.

For an arbitrary $f \subsetsim M^{(n)}$, we can certainly show[3]

$$\bigvee_{\bar{d}\in D} (R(\bar{d},\bar{x}) * f(\bar{x})) \le f(\bar{x}) \le \bigwedge_{\bar{d}\in D} (R(\bar{x},\bar{d}) \to f(\bar{x})), \forall \bar{x} \in C.$$

[3] Considering $\bigvee_{\bar{d}\in D} R(\bar{d},\bar{x}) = \bigvee_{\bar{d}\in D} R(\bar{x},\bar{d}) = \mathbf{1}$ gives the following equality $f(\bar{x}) = \bigvee_{\bar{d}\in D} (R(\bar{d},\bar{x}) * f(\bar{x})) = \bigwedge_{\bar{d}\in D} (R(\bar{x},\bar{d}) \to f(\bar{x}))$, for all $\bar{x} \in C$, because in an arbitrary residuated lattice $\mathbf{1} * a = \mathbf{1} \to a = a$, for all $a \in L$. Hence, it is reasonable to apply this requirement for the partitioning of the domain of f in the practise.

ISBN: 978-989-95079-6-8

The task stands in the replacement of $f(\bar{x})$ by suitable values so that the approximation is the best possible w.r.t. the available knowledge. In the following proposition, we are going to show that the combinations of $\{\mathrm{DNF}_f^D, \mathrm{CNF}_f^C\}$ and $\{\mathrm{CNF}_f^D, \mathrm{DNF}_f^C\}$ have this property.

Theorem 5 – DNF_f^C *is the least element of the following set:*

$$S = \{g \subsetsim M^{(n)} |\ f(\bar{x}) \leq \bigwedge_{d\in D} (R(\bar{x},\bar{d}) \to g(\bar{d})),\ \forall \bar{x} \in C\}. \tag{9}$$

– CNF_f^C *is the greatest element of the following set:*

$$S' = \{g \subsetsim M^{(n)} |\ \bigvee_{d\in D} (R(\bar{d},\bar{x}) * g(\bar{d})) \leq f(\bar{x}),\ \forall \bar{x} \in C\}. \tag{10}$$

As it has been shown above, there cannot be better lower approximation of f from the set $\mathbb{M}_\vee^D$ than $\mathrm{DNF}^D_{\mathrm{CNF}_f^C} \in \mathbb{M}_\vee^D$. And analogously, $\mathrm{CNF}^D_{\mathrm{DNF}_f^C} \in \mathbb{M}_\wedge^D$ is the best upper approximation of f. In other words, the best choice for coefficients in the inverse formula of $\mathrm{F}^{\uparrow(\downarrow)}$-transform is to take just the coefficients of the direct $\mathrm{F}^{\uparrow(\downarrow)}$-transform. Both results do not consider any boundary for an $\varepsilon-$precision of the approximations. It is dependent on the suitable choice of $\mathcal{R}_I$ and the set of nodes D arising from the following estimation in terms of a $*$-equivalence:

Theorem 6 *Let $f \subsetsim M^{(n)}$ be extensional w.r.t. $\mathcal{S}_I = \{S_i \subsetsim M_i^2\}_{i\in I}$ as follows:*

$$S_i(x,y) = \bigvee_{\bar{d}\in D} R_i(x,d_i) * R_i(d_i,y). \tag{11}$$

Then we can prove

$$\bigvee_{\bar{d}\in D} R(\bar{x},\bar{d}) * R(\bar{d},\bar{x}) \leq \begin{array}{l}\mathrm{DNF}^D_{\mathrm{CNF}_f^C}(\bar{x}) \leftrightarrow f(\bar{x}),\\ \mathrm{CNF}^D_{\mathrm{DNF}_f^C}(\bar{x}) \leftrightarrow f(\bar{x}),\end{array} \quad \forall \bar{x} \in C. \tag{12}$$

Corollary 7 *Let $\overline{\mathcal{R}}_I = \{\bar{R}_i\}_{i\in I}$ be defined by*

$$\bar{R}_i(x,y) = \bigvee_{j\in J} A_{j_i}(x) * A_{j_i}(y), \tag{13}$$

for all $i \in I$. If $f \subsetsim M^{(n)}$ is extensional w.r.t. $\overline{\mathcal{R}}_I$ then (5) is valid.

Here, we provide an alternative proof of this assertion independent of the above theorem.

Proof: Due to extensionality, we have

$$A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{y}) \leq R(\bar{x},\bar{y}) = \bigvee_{\mathbf{j}\in J^n} A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{y}) \leq$$
$$\leq f(\bar{x}) \leftrightarrow f(\bar{y}), \text{ and hence}$$
$$A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{x}) * f(\bar{x}) \leq A_{\mathbf{j}}(\bar{x}) * (A_{\mathbf{j}}(\bar{y}) \to f(\bar{y})),$$
$$A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{x}) * f(\bar{x}) \leq A_{\mathbf{j}}(\bar{x}) * \bigwedge_{\bar{y}\in C} (A_{\mathbf{j}}(\bar{y}) \to f(\bar{y})),$$
$$f(\bar{x}) * \bigvee_{\mathbf{j}\in J^n} A_{\mathbf{j}}(\bar{x}) * A_{\mathbf{j}}(\bar{x}) \leq$$
$$\leq \bigvee_{\mathbf{j}\in J^n} (A_{\mathbf{j}}(\bar{x}) * \bigwedge_{\bar{y}\in C} (A_{\mathbf{j}}(\bar{y}) \to f(\bar{y}))) = f^{\downarrow}_{F,k}(\bar{x}),$$

for all $\bar{x} \in M^{(n)}$. Adding the left side of inequality (2), we receive the first required estimation (5). Analogously we proceed to prove the inequality for $f^{\downarrow}_{F,k}$. QED

Hence, to obtain the desired ε-approximation[4], it only remains to choose D w.r.t. $\mathcal{R}_I$ so that

$$(\forall \bar{x} \in C)(\varepsilon \leq \bigvee_{\bar{d}\in D} R(\bar{x},\bar{d}) * R(\bar{d},\bar{x})),$$

which completes the task.

Corollary 8 *Let $\varepsilon \in L, D \subseteq M^{(n)}$ and $f \subsetsim M^{(n)}$ be extensional w.r.t. $\mathcal{S}_I$ defined by (11). If for each $\bar{x} \in C$ there exists $d \in D:\ \varepsilon \leq R(\bar{x},d) * R(d,\bar{x})$ then*

$$\varepsilon \leq \begin{array}{l}\mathrm{DNF}^D_{\mathrm{CNF}_f^C}(\bar{x}) \leftrightarrow f(\bar{x}),\\ \mathrm{CNF}^D_{\mathrm{DNF}_f^C}(\bar{x}) \leftrightarrow f(\bar{x}),\end{array} \quad \forall \bar{x} \in C. \tag{14}$$

Corollary 9 *Let $\varepsilon \in L$ and $f \subsetsim M^{(n)}$ be extensional w.r.t. $\bar{\mathcal{R}}_I$ defined by (13). If for each $\bar{x} \in C$ there exists $\mathbf{j} \in J^n$: $\varepsilon \leq A_{\mathbf{j}}(x) * A_{\mathbf{j}}(x)$ then*

$$\varepsilon \leq f^{\downarrow(\uparrow)}_{F,k}(\bar{x}) \leftrightarrow f(\bar{x}),\ \forall \bar{x} \in C. \tag{15}$$

Corollary 10 *Let $L = [0,1]$, $*$ be a continuous t-norm, $\varepsilon \in L$, M be a pre-compact (totally bounded) subset over a standard metric space, $R \subsetsim M$ be reflexive and uniformly continuous. If $f \subsetsim M$ is extensional w.r.t. R then for every $\varepsilon < 1$ there exists a finite $D \subseteq M$ such that (14) is valid.*

Whenever we assume each $R \in \mathcal{R}$ to be a $*$-equivalence, we obtain an estimation by means of the following pseudometrics:

$$d_n(\bar{x},\bar{y}) = \varphi^{-1}R_1(x_1,y_1) \oplus_\varphi \ldots \oplus_\varphi \varphi^{-1}R_n(x_n,y_n)$$
$$= \bigoplus_{i\in I} \varphi^{-1}R_i(x_i,y_i),\ \forall \bar{x},\bar{y} \in M^{(n)},$$
$$d(x,y) = \varphi^{-1}(x \leftrightarrow y),\ \forall x,y \in L$$

where φ is an order-reversing bijection on L, $\oplus_\varphi$ is an operation defined by $x \oplus_\varphi y = \varphi^{-1}(\varphi(x) * \varphi(y))$.

Corollary 11 *Let each $R \in \mathcal{R}_I$ be a $*$-equivalence and $d_n : M^{(n)} \times M^{(n)} \to L, d : L \times L \to L$ be defined as above. If $f \subsetsim M^{(n)}$ is extensional w.r.t. $\mathcal{S}_I$ defined by (11) then f is 1-Lipschitz continuous w.r.t. (d,d_n), i.e.,*

$$d(f(\bar{x}),f(\bar{y})) \leq d_n(\bar{x},\bar{y}),\ \textit{for all } \bar{x},\bar{y} \in M^{(n)},$$

and

$$\begin{array}{l} d(\mathrm{DNF}^D_{\mathrm{CNF}_f^C}(\bar{x}), f(\bar{x}))\\ d(\mathrm{CNF}^D_{\mathrm{DNF}_f^C}(\bar{x}), f(\bar{x}))\end{array} \leq \bigwedge_{\bar{d}\in D} 2d_n(\bar{x},\bar{d})$$
$$= \bigwedge_{\bar{d}\in D} d_n(\bar{x},\bar{d}) \oplus d_n(\bar{d},\bar{x}),\ \forall \bar{x} \in C. \tag{16}$$

[4]Notice that the choice of $\mathcal{R}_I$ is the limitation for ε as well as for D, e.g. D might become very huge if ε is close to $\mathbf{1}$.

ISBN: 978-989-95079-6-8

Let us consider a simple problem to illustrate the way of approximation by F-transforms.

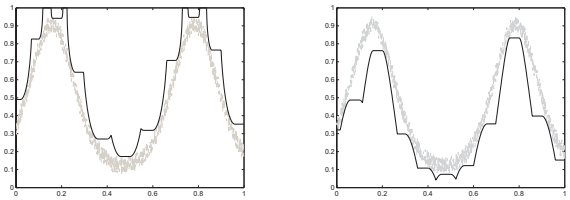
Example 12 *Let us take the following function*

$$f(x) = \frac{1}{3} \cdot e^{\sin(10x)} + Rand(x),\ \forall x \in M = [0,1],$$

where $Rand(x)$ is some additive random noise. Moreover, assume $*$ to be the t-norm generated by $g(x) = (1-x)^2$ and $\mathcal{L}$ be the associated residuated lattice. Take

$$R(x,y) = 1 \wedge (0 \vee (1.2 - 10 \cdot |x-y|)),$$

and create the approximation using 11 and 21 nodes and equidistant discretization $C = \{k \cdot 0.001 |\ k = 1, \ldots, 999\}$. The resulting approximations are depicted on Fig. 1, where f is represented by grey line and inverse $F^{\uparrow(\downarrow)}$-transforms by black lines, set D contains 11 nodes and set D' contains 21.



(a) $\mathrm{CNF}^{D}_{\mathrm{DNF}^{C}_{f}}$ (b) $\mathrm{DNF}^{D}_{\mathrm{CNF}^{C}_{f}}$

(c) $\mathrm{CNF}^{D'}_{\mathrm{DNF}^{C}_{f}}$ (d) $\mathrm{DNF}^{D'}_{\mathrm{CNF}^{C}_{f}}$

Figure 1: Approximation of f (grey line) from Example 12 using F-transforms (black line) with 11 (set D) and 21 (set D') nodes.

4 Conclusions

We have shown that the only one requirement is needed to show (5):

New requirements
$\mathcal{L}$ be a residuated lattice
f be extensional w.r.t. $\{\bar{R}_i\}_{i \in I}$

While in [1], we were limited to the stronger algebraic structure, a special type of partition, $E \subseteq C$, additionally requiring normality, and extensionality of f (see Subsection 2.2). Notice that all the additional requirements follow from the procedure of replacement of fuzzy sets A_{i_i} by P_i based on the special set E (such relations can be found in [12]). We have found a way how to get around this construction and to reach such simplification.

Moreover, our approach enabled to express direct and inverse $F^{\uparrow(\downarrow)}$-transforms using combinations of normal forms. This allowed us to show an optimality of choice of all coefficients of the direct $F^{\uparrow(\downarrow)}$-transform in the inverse formula at once (see Theorem 5). Additionally, we gave some interesting corollaries that contributes only to the theory of normal forms and they cannot be translated into the theory of $F^{\uparrow(\downarrow)}$-transforms, which shows that even though both theories have a lot in common, it is also important to study them separately.

Acknowledgment

We gratefully acknowledge partial support of the projects MSM6198898701 and 1M0572 of the MŠMT ČR.

References

[1] Irina Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy Sets and Systems*, 157:992–1023, 2006.

[2] Irina Perfilieva and Dagmar Plšková. Solving ode with fuzzy initial condition using fuzzy transform. In *Proc. of 10'th Czech-Japan Seminar on data Analysis and Decision Making under Uncertainty*, pages 130–139, Prague, 2007. Oeconomica.

[3] Martin Štěpnička and Radek Valášek. Numerical solution of partial differential equations with help of fuzzy transform. In *Proceedings of the FUZZ-IEEE 2005*, pages 1104–1109, Reno, Nevada, 2005. Fuzz-IEEE 2005.

[4] Irina Perfilieva and Radek Valášek. Fuzzy approach to data compression. In *Proc. of 8'th Czech-Japan Seminar on data Analysis and Decision Making under Uncertainty*, pages 91–100, Prague, 2005. Oeconomica.

[5] Irina Perfilieva and Martina Daňková. Image fusion on the basis of fuzzy transforms. In *Computational Intelligence in Decision and Control*, pages 471–476, New Jersey, 2008. World Scientific.

[6] Martina Daňková and Radek Valášek. Full fuzzy transform and the problem of image fusion. *Journal of Electrical Engineering*, 12:82–84, 2006.

[7] Vilém Novák, Irina Perfilieva, Antonín Dvořák, G. Chen, Q. Wei, and P. Yan. Fuzzy transform in the analysis of data. *International Journal of Approximate Reasoning*, 48:4–22, 2008.

[8] Martina Daňková. On approximate reasoning with graded rules. *Fuzzy Sets and Systems*, 158:652–673, 2007.

[9] Martina Daňková. Approximation of extensional fuzzy relations over residuated lattices. Submitted.

[10] Llorenç Valverde. On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, 17(3):313–328, 1985.

[11] Irina Perfilieva. Normal forms in BL-algebra of functions and their contribution to universal approximation of functions. *Fuzzy Sets and Systems*, 143:111–127, 2004.

[12] Frank Klawonn. Fuzzy points, fuzzy relations and fuzzy functions. In V. Novák and I. Perfilieva, editors, *Discovering the World with Fuzzy Logic*, pages 431–453. Physica-Verlag, Heidelberg, 2000.

ISBN: 978-989-95079-6-8

An Interval Type-2 Fuzzy Distribution Network

Simon M. Miller, Viara Popova, Robert John, Mario Gongora

Centre for Computational Intelligence, De Montfort University, Leicester, UK.
Email: {smiller,vpopova,rij,mgongora}@dmu.ac.uk

Abstract— *Planning resources for a supply chain is a major factor determining its success or failure. In this paper we introduce an* ***Interval Type-2 Fuzzy Logic model*** *of a distribution network. It is believed that the additional degree of uncertainty provided by Interval Type-2 Fuzzy Logic will allow for better representation of the uncertainty and vagueness present in resource planning models. First, the subject of* ***Supply Chain Management*** *is introduced, then some background is given on related work using Type-1 Fuzzy Logic. A description of the Interval Type-2 Fuzzy model is given, and a test scenario detailed. A* ***Genetic Algorithm*** *uses the model to search for a near-optimal plan for the scenario. A discussion of the results follows, along with conclusions and details of intended further work.*

Keywords— Distribution model, Evolutionary Computing, Interval Type-2 Fuzzy Logic, Resource Planning, Supply Chain Management

1 Introduction

There are a number of definitions of Supply Chain Management (SCM), each having minor variances but describing the same core idea. SCM is the management of material flow in and between facilities including vendors, manufacturing/assembly plants and distribution centres [1]. Planning the allocation of resources within a Supply Chain (SC) has been critical to the success of manufacturers, warehouses and retailers for many years. Mastering the flow of materials from their creation to the point of sale offers considerable advantages to those within a well managed SC. Poorly managed resources result in two main problems: stock outs and surplus stock.

The consequence of stock outs is lost sales, and potentially lost customers. Surplus stock causes additional holding cost and the possibility of stock losing value as it becomes obsolete. Holding some surplus stock is advantageous however; safety stock can be used in the event of an unexpected increase in demand or to cover lost productivity.

The problem has been addressed numerous times in the literature using different approaches. An overview of traditional and Computational Intelligence approaches to supply-chain resource planning can be found in [2]. Here a novel approach using Type-2 Fuzzy Logic (T2FL) [3] is reported together with optimisation by means of a Genetic Algorithm (GA)[4].

This research is part of a project on demand forecasting and resource planning (Data Storage, Management, Retrieval and Analysis: Improving Customer Demand and Cost Forecasting Methods funded by the Technology Strategy Board in the UK). The project aims to: (1) improve the forecasting of demand by using a variety of disparate sources and statistical and Machine Learning (ML) methods for analysis; (2) improve the allocation of resources in which the generated forecast is used as an input and the output is a (long- or short-term) plan of raw materials and resources within the supply chain in order to meet such demand. Various degrees of uncertainty are present in the different data sources used that are amplified in the generated forecast by applying methods of analysis with (again) varying degrees of inherent uncertainty. Furthermore, other data that is often used in resource planning such as transportation and other costs, customer satisfaction information, etc. is also uncertain. Therefore, FL and especially T2FL are particularly appropriate for this problem. While Type-1 FL (T1FL) has successfully been used many times for modelling SC operation (see Section 2), T2FL has been shown to offer a better representation of uncertainty on a number of problems (e.g., [5] and [6]). However, to date T2FL has not been applied to SCM in the literature. This paper presents a T2FL model of a SC problem which is optimised by a GA to find a near-optimal configuration.

The paper is organised as follows. Section 2 discusses Fuzzy Logic and its application to SCM. Section 3 introduces the model and the test case used for evaluation. The results from the experiments on the test case are presented in Section 4. Section 5 concludes the paper with a summary and future research directions.

2 Fuzzy Logic for Supply Chain Management

FL and GAs have been successfully used for supply chain modelling [2] and are particularly appropriate for this problem due to their capacity to tackle the inherent vagueness, uncertainty and incompleteness of the data used. A GA [4] is a heuristics search technique inspired by evolutionary biology. Selection, crossover and mutation are applied to a population of individuals representing solutions in order to find a near-optimal solution. FL is based on fuzzy set theory and provides methods for modelling and reasoning under uncertainty, a characteristic present in many problems, which makes FL a valuable approach. It allows data to be represented in intuitive linguistic categories instead of using precise (crisp) numbers which might not be known, necessary or in general may be too restrictive. For example, statements such as *'the cost is about n'*, *'the speed is high'* and *'the book is very old'* can be described. These categories are represented by means of a membership function which defines the degree to which a crisp number belongs to the category. In this research the aim is to allow linguistic terms to be used by Supply Chain Managers when describing their operation. For example, instead of asking for crisp numbers to describe the current stock level of a product, we may allow them to make statements like 'Warehouse 1 has *About 500* of product 1' or even more abstract 'Warehouse 1's stock level of product 1 is *low*'. By removing the need for exact information it is possible to produce a system that is much more usable when the information available is vague, uncertain or incomplete.

ISBN: 978-989-95079-6-8

T1FL has been applied to SC modelling numerous times with good results. Some of the research that is considered most relevant for this paper is discussed in the following paragraphs.

Petrovic et al. [7] use fuzzy sets to model vagueness and imprecision in customer demand, external supplier reliability and supply within a SC. The system demonstrates the effect of differing conditions and strategies on fill rate and holding costs of a SC. The results of the evaluation show that there is a slight improvement in the performance of the SC when the inventories are partially co-ordinated. In [8] Petrovic et al. employ two-level fuzzy optimisation to find ideal order-up-to quantities in a one warehouse-multiple retailer SC. The problem is decomposed for individual control of the warehouse and retailers; a co-ordination mechanism provides overall control of the SC. Customer demand, inventory levels, holding cost, and shortage cost are represented by fuzzy sets. A measure of satisfaction is derived from the cost incurred at each element. The solution produced by the system is the best compromise between members, though not necessarily the cheapest.

In [9] a fuzzy system is used with a GA to model a SC. The model uses a global policy of management with emphasis on integrating the production and distribution models. The GA searches for a near-optimal configuration; fuzzy sets are used to describe costs, returns, production capacities, storage capacities and forecasts. The proposed fuzzy method, a crisp method and a non-integrated method are compared. The crisp system is unable to produce a feasible configuration if actual demand is lower than the forecast. In contrast, the fuzzy model presented is robust and able to cope with fluctuation in demand and production capacity with little impact on profitability. The non-integrated model performed significantly worse than the fuzzy integrated model.

A similar approach is presented by Wang and Shu [10]. FL is used to represent customer demand, processing time and delivery reliability; a GA finds order-up-to levels. The system attempts to find the configuration that incurs the minimum cost. An optimism-pessimism index is set by the user and passed to the system. When optimistic, the model assumes the best case scenario for material response time. A pessimistic attitude produces the opposite effect. The results show that more pessimistic strategies increase the fill rate, reducing the sales lost through stock outs, and incur higher inventory cost as more stock is kept. More optimistic strategies result in a drop in fill rate and an increase in sales loss, though inventory cost is also reduced.

T2FL [3] provides the means to model an additional dimension of uncertainty as the membership functions are themselves modelled as fuzzy numbers, thus providing the means for a more accurate representation of uncertainty in a complex system and the potential for better performance. On a number of problems, T2FL has been shown to outperform T1FL (e.g., [5] and [6]), however it has not yet been applied to SC modelling in the literature. In this paper, Interval Type-2 FL (IT2FL) [11] is used which is computationally cheaper as it restricts the additional dimension, referred to as secondary membership function, to only take the values 0 or 1. We believe that the extra degree of freedom will allow a better representation of the uncertain and vague nature of data used in SCM. Section 3 describes the created model in detail as well as the test case used for evaluation.

3 Model

The model presented here describes a distribution network consisting of a set of manufacturing warehouses and customers. Each warehouse holds one or more products; when demand of an item is forecast for a customer, a warehouse is selected to satisfy the demand. Before the model is used it is configured to model the required network. This is done by telling the model how many warehouses, customers and products there are, the distances between the warehouses and customers, the costs associated with network operation and how many periods the model is to run for. As input, the model takes a forecast of demand over a given period and a suggested resource plan for the warehouses detailing how much of each item to hold at each warehouse in each period.

The purpose of the model is to calculate the cost of a given resource plan. This is calculated based on a number of factors:

Production Cost - Each product is assigned an individual production cost. The total production cost for each batch is calculated by multiplying the number of items by their production cost.

Holding Cost - A holding cost is charged if a product is kept in a warehouse for more than one period. The cost is calculated by taking a specified percentage of the purchase price of the goods held, for items carried over from one period to the next. The purpose of the charge is to represent the costs of storing items, depreciating value of stock and the losses incurred by tying up capital in unsold stock.

Transport Cost - The cost of transporting goods is produced using a matrix of distances between warehouses and customers, and a list of transport costs per mile for each product. The product of the relevant cost and mileage gives the overall transport cost for a batch of product.

Stock out cost - Stock out is the shortfall of a product in a particular period. In this model we make the assumption that the customer is always provided with an item. If it is not in the warehouse, it is purchased at full retail price from a competing producer. The stock out cost is the sum of the value of items that had to be purchased in this period. Cost of purchase is not the only penalty incurred when stock out occurs. The customer may cancel their order if they discover that it is not in stock, and will be ordered elsewhere. To represent this a multiplier is applied to the stock out cost, taking into account the cost of lost orders.

Batch cost - The cost of setting up an order is called the batch cost. This represents the cost of administration and setting up any machines that are required, and picking the items for dispatch. There is a flat fee for each batch which is charged once at each warehouse for the production of a particular item for a particular customer.

Using these costs a total cost is produced, allowing comparison of competing solutions.

3.1 Interval Type 2 Fuzzy Logic

For the experimental model IT2FL has been used to represent some of the values within the model. Previous examples (as

ISBN: 978-989-95079-6-8

discussed in Section 2) of research in this area have focused on the use of T1FL; we believe there exists an opportunity to exploit the extra degree of uncertainty provided by IT2FL in a model of this type.

As the model operates on fuzzy numbers, fuzzy arithmetic is used to calculate costs. This involves taking fuzzy sets, discretising them, performing the arithmetic operation, and then reconstructing the fuzzy set. In this model, fuzzy sets are represented using a series of α-cuts. Each set is an array of pairs of intervals. Each pair shows the area of values covered at a particular value of μ, the first interval is the left hand side of the set, and the second the right. Storing the sets in this way removes the need to discretise before fuzzy arithmetic is performed, and then reconstruct the result. Operations on fuzzy sets are performed at interval level, corresponding intervals (at the same μ) are taken from two sets, the operation performed and the result stored in a third fuzzy set.

The following values are represented by IT2 fuzzy numbers: forecast demand, inventory level, transportation distances, transportation cost, stock out level, stock out cost, carry over and holding cost. For each of these values we can use the linguistic term *'about n'*, e.g., forecast demand of product 1 for customer 1 in period 1 may be *'about 200'*. When the fuzzy total cost for a solution has been calculated, it is defuzzified to produce a crisp cost value. Fig. 1 shows how the set *'about 200'* may look with the α-cut representation used, where x is the scale of values being represented. The set is described using a collection of pairs of intervals. Each pairing represents the left and right hand intervals of the set for a given value of μ. As stated before, representing the set in this way considerably simplifies the arithmetic operators that are applied as simple interval arithmetic is used throughout, without discretisation or reconstruction.

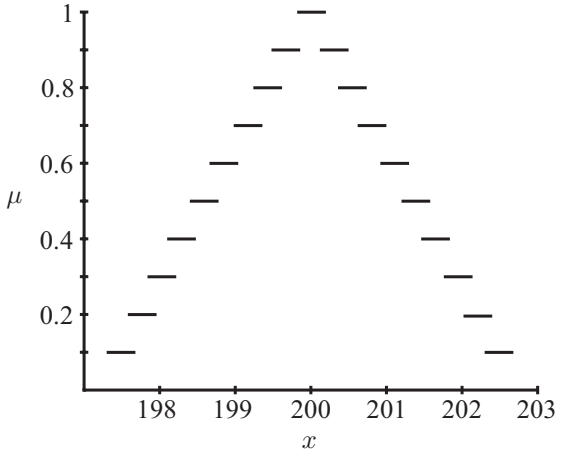


Figure 1: Interval representation of IT2 fuzzy set *'about 200'*

3.2 Defuzzification

The defuzzification process used is described in (1) where n is equal to the number of α-cuts, and l and r denote the left and right hand intervals respectively. To reference the left and right endpoints of the intervals $_1$ and $_2$ are used; so l_1 refers to the left bound of l and l_2 the right bound. M denotes the mid-point of an interval, in the case of M_i^x this means the centres of the intervals in x described by the i^{th} pair of intervals. For M^{α_i} we take the mid-point between the current and subsequent α-cut's μ values. The mid-point values are used along with the height (H) and length (L) to calculate the weighted area of each pair of intervals, giving more weight the those with higher μ values.

$$\frac{\sum_{i=0}^{n-1} M_i^x H_i L_i M^{\alpha_i}}{\sum_{i=0}^{n-1} H_i L_i M^{\alpha_i}} \tag{1}$$

where:

$$M_i^x = \begin{cases} \frac{x_{l_2}^{\alpha_i} + x_{l_1}^{\alpha_i}}{2} & \text{for left interval} \\ \frac{x_{r_2}^{\alpha_i} + x_{r_1}^{\alpha_i}}{2} & \text{for right interval} \end{cases} \tag{2}$$

$$H_i = \alpha_{i+1} - \alpha_i \tag{3}$$

$$L_i = \begin{cases} x_{l_2}^{\alpha_i} - x_{l_1}^{\alpha_i} & \text{for left interval} \\ x_{r_2}^{\alpha_i} - x_{r_1}^{\alpha_i} & \text{for right interval} \end{cases} \tag{4}$$

$$M^{\alpha_i} = \frac{\alpha_i + \alpha_{i+1}}{2} \tag{5}$$

In the test scenario, μ is discretised in a uniform manner. This removes the need to calculate the height of intervals as they are all the same, and therefore cancel out. The process used in the actual code is given in (6).

$$\frac{\sum_{i=0}^{n-1} M_i^x L_i M^{\alpha_i}}{\sum_{i=0}^{n-1} L_i M^{\alpha_i}} \tag{6}$$

It should be pointed out that while this method works well for the convex fuzzy numbers used in this system, it cannot be applied to non-convex fuzzy sets as the interval representation employed will cease to accurately represent the set. In this model, the fuzzy numbers will always be convex. In addition to this the method of representation being used ensures that the intervals are already present, no further discretisation is required. Using a method that requires re-discretisation would incur unnecessary computation. A consideration for future work is the possibility of using another approach such as the one proposed by Karnik and Mendel [12], however this method would involve rediscretising on the x axis, at present the model is discretised on the μ axis.

3.3 Optimisation

The focus of the experiment described here is the validation of an IT2FL model that has been constructed. This is to be achieved by using the model to find a good resource plan for a

ISBN: 978-989-95079-6-8

given forecast, confirming that the model can be used to evaluate potential solutions. A GA has been chosen for this purpose. GAs have been used successfully in previous work (e.g., [13] and [14]) to find good solutions with T1FL models. GAs are useful when a search area is too large to allow evaluation of every solution. In this case the GA is used to search for a resource plan that incurs the minimum cost. The GA has a population of 250 and is executed for 500 generations, in all 125000 solutions are evaluated. Even in this relatively small problem the GA only covers approximately $\frac{1}{10^{62}}$ of the total search space. New generations consist of: 1% individuals produced with elitism, 20% copied individuals, 20% individuals created with single point crossover and 59% of individuals created using mutation. A description of the chromosome, operators and processes employed follows.

Chromosome - The chromosome used to describe potential solutions is a 3-D matrix of inventory levels. Columns represent customers, and rows represent a product at a warehouse. If there are 3 warehouses and 5 products this will result in 15 rows. The first 5 rows will be products 1 to 5 for warehouse 1, and second 5 for warehouse 2 and so on. The third dimension represents the period, in each period there is a complete inventory plan for all warehouses and customers.

Initial population - The initial population is randomly generated. Each element of the resource plan matrix can be a number between 0 and 500 in steps of 100. This has been done to reflect the fact that in industry, products are usually manufactured in round quantities. If the model suggests that a warehouse should make 102 of product 1, this could lead to difficulties and extra expense. Limiting the valid inventory numbers also has the side effect of reducing the search space.

Fitness evaluation - Fitness is evaluated using the IT2FL model described. Candidate solutions are given to the model which evaluates them, and returns the cost. The cheaper a solution is, the fitter it is judged to be. In reality, cost may not be the only factor in deciding how much of a product to stock at each warehouse. Other criteria such as customer service level can also be used to prevent the system from choosing solutions that do not meet service requirements. Customer service level could be calculated by looking at the percentage of orders fulfilled completely, or orders fulfilled on time.

Selection - Selection is performed using a fitness ranking proportionate method similar to *roulette wheel selection*. First all solutions in the population are ranked by fitness. They are then given a number of elements of an array in proportion to their fitness ranking. For example, if we have a population of 250 the fittest individual would be allocated 250 elements in the array, the second fittest 249 and so on. An element of the array is then selected at random, and the identification number of the individual it contains is used to retrieve a parent. This tombola style approach ensures that it is possible for any individual to be selected, while weighting in favour of those with greater fitness.

Crossover - Crossover is achieved with a single point crossover. Two parents are selected using the method of selection described. Then, a new individual is created with the first half of the first parent and the second half of the second parent. Division is done by rows, if we have 5 nodes with 4 products we would have 20 rows; in this case the first 10 rows of the first parent and the final 10 of the second parent would be used to create a new individual. If the example were to span 6 periods, we would take the first 10 rows for all 6 periods of the first parent, and the final 10 rows for all 6 periods from the second to create our new individual.

Mutation - To create a mutated individual, a parent is selected, then one of the elements of its resource plan is randomly replaced with another valid value to create a new child.

3.4 Test Scenario

The test scenario is as follows:

The distribution network consists of 3 manufacturing warehouses located in Bristol (W1), Manchester (W2) and Edinburgh (W3) and 6 customers in Plymouth (C1), Cardiff (C2), Derby (C3), Liverpool (C4), Newcastle (C5) and Aberdeen (C6). Each manufacturing warehouse holds the same 5 products, and may supply any customer. The model will be executed over one period and holding costs are 10% of the purchase price of an item per period. Stock outs will be charged at 5 times the cost of purchasing an item, per item short. Each batch of product produced will incur a batch cost of £100. Table 1 details the costs associated with each product.

Table 1: Product costs (£s)

Product	1	2	3	4	5
Purchase price (per item)	3	4	2	3	3
Production cost (per item)	2	3	1	2	2
Transport cost (per item/per mile)	1	2	1	3	1

The forecast being used states that each customer would like 200 of each product. The location of the manufacturing warehouses and customers have been arranged so that a good solution is easily identifiable. Each warehouse has 2 customers that are best served by it. Table 2 shows how an ideal plan should look. In this case the problem is intentionally kept relatively simple, making it easy to evaluate the found solution and see if it is near the optimal solution. Clearly, simple problems do not require such a system in order to solve them. The work shown here is an initial experiment, future work will build upon this so that it can be applied to larger non-trivial problems where the benefit from the system will be greater.

The next section presents and interprets the results from the experiments with the GA.

4 Results

As stated previously, in each test the GA was applied to the model using 250 individuals over 500 generations. The test was repeated 10 times with differing random seeds, the best solution found in each test was recorded. Table 3 shows the costs incurred by the best solution found in each test. For reference, the cost of the ideal solution presented is £19178.26.

Using the IT2FL model the GA is able to find solutions that are close to the optimum. The best solution found is just £264.76 (1.38%) from the ideal solution, and the mean difference between solutions found and the ideal solution is £1344.79 (7.01%). Table 4 shows the best solution found (with random seed 1). Comparing this solution with the ideal

ISBN: 978-989-95079-6-8

Table 2: Ideal solution

Ware.	Prod.	C1	C2	C3	C4	C5	C6
W1	1	200	200	0	0	0	0
W1	2	200	200	0	0	0	0
W1	3	200	200	0	0	0	0
W1	4	200	200	0	0	0	0
W1	5	200	200	0	0	0	0
W2	1	0	0	200	200	0	0
W2	2	0	0	200	200	0	0
W2	3	0	0	200	200	0	0
W2	4	0	0	200	200	0	0
W2	5	0	0	200	200	0	0
W3	1	0	0	0	0	200	200
W3	2	0	0	0	0	200	200
W3	3	0	0	0	0	200	200
W3	4	0	0	0	0	200	200
W3	5	0	0	0	0	200	200

Table 3: Results of test - Best fitness

Seed	Best Fitness (£s)
0	19789.32
1	19443.02
2	20860.01
3	21048.27
4	20407.71
5	20854.52
6	20952.15
7	20950.17
8	20229.67
9	20695.66

solution we can see that not only are the plans found cost-effective, but also represent a sensible allocation of resources. The found solution is very close to the ideal solution, and where there are differences the found model is producing the correct amount of product, but a different warehouse has been chosen. This would not be the case if the model was inadequate for the problem and did not provide a clear distinction between good and bad solutions. All customers will receive their order, the only negative point is that some of customer 4 and 5's orders could be supplied by a closer warehouse.

When the program begins, the best individual in the first generation tends to cost around £70000. Fig. 2 shows a typical run, in this case for the test with a random seed of 1. Evolution can clearly be seen to be heading toward the ideal solution, we also know what the perfect solution looks like, and that the solutions found are close to it. This suggests that the model is working correctly. If the model were not correct, evolution may occur, but possibly in an unexpected direction. This could result in cheap solutions that are not sensible, or in an inability to find cost-effective solutions at all.

Out of interest, an extra test was run to see whether the GA would find a better solution if it were left to run over more generations. The best seed from the previous tests (random seed 1) was used as a starting point, and the GA was allowed to run for 2500 generations. In generation 2141 the ideal solution was found, proving that using only feedback from the model, the GA is able to find the ideal solution.

Table 4: Solution found with a random seed of 1

Ware.	Prod.	C1	C2	C3	C4	C5	C6
W1	1	200	200	0	200	0	0
W1	2	200	200	0	0	0	0
W1	3	200	200	0	0	0	0
W1	4	200	200	0	0	0	0
W1	5	200	200	0	0	0	0
W2	1	0	0	200	0	200	0
W2	2	0	0	200	200	0	0
W2	3	0	0	200	200	200	0
W2	4	0	0	200	200	0	0
W2	5	0	0	200	200	200	0
W3	1	0	0	0	0	0	200
W3	2	0	0	0	0	200	200
W3	3	0	0	0	0	0	200
W3	4	0	0	0	0	200	200
W3	5	0	0	0	0	0	200

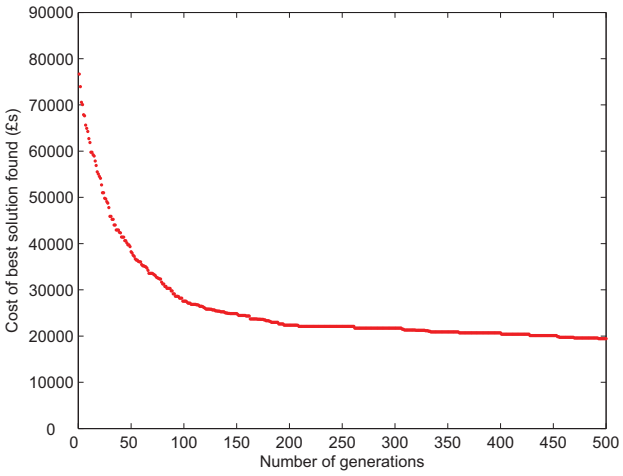


Figure 2: Evolution of best fitness

5 Conclusion

In this paper we have presented an IT2FL model that can be used to model a distribution network. The advantage of this model is that IT2FL accounts for the uncertain nature of the real world, allowing inputs to the model to be vague. IT2FL has been chosen as it is believed that the extra level of uncertainty over T1FL offered will benefit a model of this type, while avoiding the computational complexity of a Generalised Type-2 FL model.

Using the model, it was shown that a GA was able to find good resource plans that were both cost-effective and sensible. Over a longer duration, the GA was able to find the ideal solution. As a GA is solely guided by fitness and essentially 'blind' to the practicality of the solutions it finds, this was taken as indication of the model's validity.

There is much further work to be done with this model in order to make it useful in real world applications. Initially, the model will be extended for use in cases where an ideal result has been calculated by other means (e.g., a statistical model) so that performance comparison is possible. In these cases we would be looking for both accuracy of costings provided,

ISBN: 978-989-95079-6-8

and the ability of an optimisation algorithm to find good solutions. As well as being larger problems, these case studies may also call for more than just cost as a measure of fitness. As discussed in the paper, this will mean adding in a measure of customer satisfaction such as fill rate to ensure that good plans are those that are not only cost-effective, but also help to retain customers.

The current model does not take into account the temporal nature of manufacture and supply. Another dimension that is to be considered is how *lead times* affect the operation of the network. Lead times are another attribute that could benefit from an IT2FL representation, as we can never be certain when a product will arrive, or exactly how long it will take to produce. Part of this will include representing capacity, as at present only inventory is accounted for by the model. This will allow the optimising algorithm to suggest not only stock levels, but also capacity levels within a factory. Capacities can then be used to provide some indication of how long we should expect to wait for a given batch of product to be manufactured.

Acknowledgment

The research reported here has been funded by the Technology Strategy Board (Grant No. H0254E).

References

[1] D.J. Thomas and P.M. Griffin. Coordinated supply chain management. *European Journal of Operational Research*, 94(1):1–15, October 1996.

[2] S.M. Miller, V. Popova, R. John, and M. Gongora. Improving resource planning with soft computing techniques. In *Proceedings of UKCI 2008, De Montfort University, Leicester, UK.*, September 2008. available at: http://www.cci.dmu.ac.uk/preprintPDF/SimonUKCI(2).pdf.

[3] J.M. Mendel and R.I.B. John. Type-2 fuzzy sets made simple. *IEEE Transactions on Fuzzy Systems*, 10(2):117–127, April 2002.

[4] J.H. Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor: University of Michigan, 1975.

[5] H.A. Hagras. A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots. *IEEE Transactions on Fuzzy Systems*, 12(4):524–539, August 2004.

[6] N.N. Karnik and J.M. Mendel. Applications of type-2 fuzzy logic systems to forecasting of time-series. *Information Sciences*, 120:89–111, 1999.

[7] D. Petrovic, R. Roy, and R. Petrovic. Supply chain modelling using fuzzy sets. *International Journal of Production Economics*, 59:443–453, 1999.

[8] D. Petrovic, Y. Xie, K. Burnham, and R. Petrovic. Coordinated control of distribution supply chains in the presence of fuzzy customer demand. *European Journal of Operational Research*, 185:146–158, 2008.

[9] R.A. Aliev, B. Fazlollahi, B.G. Guirimov, and R. R. Aliev. Fuzzy-genetic approach to aggregate production-distribution planning in supply chain management. *Information Sciences*, 177:4241–4255, 2007.

[10] J. Wang and Y-F. Shu. Fuzzy decision modelling for supply chain management. *Fuzzy Sets and Systems*, 150(1):107–127, 2005.

[11] J.M. Mendel, R.I. John, and F. Liu. Interval type-2 fuzzy logic systems made simple. *IEEE Transactions on Fuzzy Systems*, 14(6):808–821, December 2006.

[12] N. Karnik and J. Mendel. Centroid of a type-2 fuzzy set. *Information Sciences*, 132:195–220, 2001.

[13] M. Sakawa and T. Mori. An efficient genetic algorithm for job-shop scheduling problems with fuzzy processing time and fuzzy duedate. *Computers & Industrial Engineering*, 36:325–341, 1999.

[14] C. Fayad and S. Petrovic. A fuzzy genetic algorithm for real-world job shop scheduling. In *Proceedings of the 18th international conference on Innovations in Applied Artificial Intelligence. Bari, Italy.*, pages 524–533, 2005.

ISBN: 978-989-95079-6-8

Fuzzy differential equation with π-derivative

Y. Chalco-Cano[1] H. Román-Flores[2] M.D. Jiménez-Gamero[3]

1.Departamento de Matemática, Universidad de Tarapacá
Arica, Chile
2.Instituto de Alta Investigación, Universidad de Tarapacá
Arica, Chile
3.Departamento de Estadística e Investigación Operativa Facultad de Matemáticas, Universidad de Sevilla
Sevilla, Spain
Email: {ychalco,hroman}@uta.cl, dolores@us.es

***Abstract*— *The π-derivative for fuzzy function is considered. Consequently, using this π-derivative we study fuzzy differentuial equations. In particular, we build a solution for a fuzzy differential equation with the help of a system of ordinary differential equations which is generates of the π-derivative.*

***Keywords*—** π-derivative for set valued functions, derivative for fuzzy functions, fuzzy differential equations.

1 Introduction

The spaces $\mathcal{K}$, of all closed and bounded intervals of $\mathbb{R}$ is not linear spaces since it do not contain inverse elements for the addition and therefore subtraction is not well defined. As a consequence, alternative formulations for subtraction have been suggested. One of those alternatives are the H-diference [3, 13] and by embedding the space $\mathcal{K}$ in a linear space [22].

The Radström's embedding theorem [22] tell us that there is a real normed linear space $\mathcal{B}$ and an isometric mapping $\pi : \mathcal{K} \to \mathcal{B}$. Then, taking advantage of this embedding theorem, a set-valued mappings F is π-differentiable at t_0 if $\pi \circ F$ is differentiable at t_0. In [3], we can find various properties of this derivative and its connection with other definitions of derivatives for set-valued mappings. On the other hand, in [20] each element of $\mathcal{B}$ can be represented in the form (x, δ). Thus, the π-derivative of $\pi \circ F$ is associated with the derivative of an ordered pair. In [9] the authors obtained a set $F'_{\pi}(t) \in \mathcal{K}$, which is said π-derivative of F at t. Some properties and the conection with generalized derivative are studied.

The aim of this paper is introduce the π-derivative for fuzzy interval valued functions based in the generalization of π-derivative for set-valued mappings given in [9]. Consequently, using the π-derivative for fuzzy functions, we study fuzzy differential equations. In particular, we build a solution for a fuzzy differential equation with the help of a system of ordinary differential equations which is generates of the π-derivative.

2 Preliminaries

Let $\mathbb{R}$ be the 1-dimensional Euclidian space. We denote by $\mathcal{K}$ the family of all bounded intervals

$$\mathcal{K} = \{A = [a, b] \subset \mathbb{R} \mid A \neq \emptyset \text{ is bounded } \}$$

The Hausdorff metric H is defined by

$$H(A, B) = \max \{|a - c|, |b - d|\}.$$

where $A = [a, b]$ and $B = [c, d]$.

Also, by using the Minkowski sum between two sets, we defined the following operations on $\mathcal{K}$

$$A + B = \{a + b \mid a \in A,\ b \in B\} \text{ and } \lambda A = \{\lambda a \mid a \in A\}. \tag{1}$$

The spaces $\mathcal{K}$ is not linear spaces since it do not contain inverse elements and therefore subtraction is not well defined [1, 2].

We have, from (1), that $A - B = A + (-1)B$. Also, if $A = B + C$, then the Hukuhara difference of A and B is denoted by $A -_H B$ and it is equal to C. The Hukuhara difference of A and B is also called the geometrical difference between the sets A and B [27].

In [3] the authors introduce subtraction in $\mathcal{K}$ by using the Radström's embedding theorem [22] which tell us that there is a real normed linear space $\mathcal{B}$ and an isometry $\pi : \mathcal{K} \to \mathcal{B}$ such that $\pi(\mathcal{K})$ is a convex cone in $\mathcal{B}$.

To construct $\mathcal{B}$, we consider the following equivalence relation in the space $\mathcal{K} \times \mathcal{K}$

$$(A, B) \sim (C, D) \quad \text{if and only if } \ A + D = B + C.$$

If $\langle A, B\rangle$ is the equivalence class of the pair (A, B), then $\mathcal{B}$ is the quotient space $\mathcal{K} \times \mathcal{K}/\sim$. Now, in $\mathcal{B}$ we introduce the operations of addition and scalar multiplication by means

$$\langle A, B\rangle + \langle C, D\rangle = \langle A + C, B + D\rangle,$$

$$\lambda \langle A, B\rangle = \begin{cases} \langle \lambda A, \lambda B\rangle & \lambda \geq 0, \\ \langle |\lambda| B, |\lambda| A\rangle & \lambda < 0 \end{cases}$$

Then $\mathcal{B}$ is a linear space.

One defines the embedding $\pi : \mathcal{K} \to \mathcal{B}$ as follows

$$\pi(A) = \langle A, 0\rangle, \ \ A \in \mathcal{K},$$

so that $\langle A, 0\rangle$ is the equivalence class
$\{(A + D, D) \mid A, D \in \mathcal{K}\}$.

The metric and the norm in $\mathcal{B}$ are defined by

$$\rho(\langle A, B\rangle \ , \ \langle C, D\rangle) = H(A + D \ , \ C + B)$$

$$\|\langle A, B\rangle\| = \rho(\langle A, B\rangle \ , \ \langle 0, 0\rangle)$$

If $A, B \in \mathcal{K}$, then the difference of the sets A and B is an alement of $\mathcal{B}$ equal to $\langle A, B\rangle$. Since $\mathcal{B}$ is a linear space, difference has all the properties of difference in linear spaces.

ISBN: 978-989-95079-6-8

In general, the difference of two sets in $\mathcal{K}$ is not necessarily an element of $\mathcal{K}$. On the other hand if the Hukuhara diference is well defined for two sets A, B in $\mathcal{K}$, then

$$\langle A, B\rangle = \langle A -_H B, 0\rangle .$$

3 π-derivative for set-valued mappings

In this Section we present the π-derivative for set-valued mappings. This concept has been studied by the authors in [9].

Elements of the space $\mathcal{K}$ are bounded and closed intervals, therefore each equivalence class can be represented in the form

$$\langle [x, x+\delta]\ ,\ 0\rangle\ ,\ \delta \geq 0 \quad \text{or} \quad \langle 0\ , [-x, -x-\delta]\rangle\ ,\ \delta < 0.$$

We will call this the canocical representation [20].

In fact, an arbitrary element $([a,b]\ ,\ [c,d])$ belongs to a class of the first kind if $b-a \geq d-c$ (then $x = a-c,\ \delta = (b-a)-(d-c)$) and to a class of the second kind if $b-a < d-c$ (then $x = a-c,\ \delta = (b-a)-(d-c)$).

For the canonical representation of an equivalence class we will use the notation (x, δ), which implies that

$$(x,\delta) = \begin{cases} \langle [x,\ x+\delta]\ ,\ 0\rangle & \text{if} \quad \delta \geq 0 \\ \langle 0,\ [-x,\ -x-\delta]\rangle & \text{if} \quad \delta < 0. \end{cases} \tag{2}$$

Given $A = [a,b] \in \mathcal{K}$ we have

$$\pi(A) = \pi([a,b]) = \langle [a,\ b]\ ,\ 0\rangle = \langle [x, x+\delta], 0\rangle ,$$

where $x = a$ and $\delta = b-a$.

Theorem 1 *([20]) The following results hold*

(a) $(x,\delta) = (y,\beta) \Longleftrightarrow x = y,\ \delta = \beta$;

(b) $(x,\delta) + (y,\beta) = (x+y, \delta+\beta)$;

(c) $a(x,\delta) = (ax, a\delta)$ *for each* $a \in \mathbb{R}$.

Theorem 2 *([20]) The equality*

$$D_\pi(x(t), \delta(t)) = (x^{'}(t), \delta^{'}(t))$$

holds, where $D_\pi(x(t),\delta(t))$ *is the* π*-derivative ([3]) of the pair* $(x(t),\delta(t))$ *and* $x^{'}$ *is the derivative of the real-valued function* $x(t)$.

Let $F : T \to \mathcal{K}$ be a set-valued mapping. If we denote by $F(t) = [f(t), g(t)]$, then

$$\pi(F(t)) = \langle [x(t), x(t)+\delta(t)], 0\rangle ,$$

where $x(t) = f(t)$ and $\delta(t) = g(t) - f(t)$.

From Theorem 2 $D_\pi(x(t),\delta(t)) = (x^{'}(t), \delta^{'}(t))$ and from canonical representation (2) we obtain the set $F_\pi^{'}(t) \in \mathcal{K}$ such that

$$F_\pi^{'}(t) = \begin{cases} \left[x^{'}(t), x^{'}(t) + \delta^{'}(t)\right] & \text{if} \quad \delta^{'}(t) \geq 0 \\ \left[x^{'}(t) + \delta^{'}(t), x^{'}(t)\right] & \text{if} \quad \delta^{'}(t) < 0. \end{cases} \tag{3}$$

We will say that $F_\pi^{'}(t_0)$ is the π-derivative of the set-valued mapping F at t_0.

Corollary 1 *([9]) Let* $F : T \to \mathcal{K}$ *be a set-valued mapping and we denote* $F(t) = [f(t), g(t)]$. *Then* F *is* π*-differentiable at* t_0 *if and only if* f *and* g *are differentiable functions at* t_0.

Example 1 *([9]) Consider the set-valued mapping* $F : (-1,1) \to \mathcal{K}$ *defined by* $F(t) = \left[-t^2, e^{-t}\right]$. *If* $f(t) = -t^2$ *and* $g(t) = e^{-t}$, *we have that* f *and* g *are differentiable functions and* $f^{'}(t) = -2t$ *and* $g^{'}(t) = -e^{-t}$. *In this case* $x^{'}(t) = -t^2$ *and* $\delta^{'}(t) = -e^{-t} + 2t$. *Therefore, from (3) we have*

$$F_\pi^{'}(t) = \begin{cases} [-e^{-t}\ ,\ -2t] & \text{if} \quad t \in (-1, 0.35173] \\ [-2t\ ,\ -e^{-t}] & \text{if} \quad t \in (0.35173, 1). \end{cases}$$

Note that the π-derivative (3) is coincident with the generalized derivative introduced in [26].

4 π-derivative for fuzzy functions

Let $\mathcal{F}$ be the set of all fuzzy intervals with bounded α-level intervals. This means that if $u \in \mathcal{F}$ then the α-level set is a closed bounded interval which we denote by $[u]^\alpha$, for all $\alpha \in [0,1]$.

Let T be a real interval. A mapping $X : T \to \mathcal{F}$ is called a fuzzy function. We denote

$$[X(t)]^\alpha = X_\alpha(t) = [f_\alpha(t), g_\alpha(t)],\ \ t \in T,\ \ 0 \leq \alpha \leq 1.$$

Note that X_α is a set-valued mapping for each $\alpha \in [0,1]$. The π-derivative $X_\pi^{'}(t)$ of a fuzzy function X is defined by

$$[X_\pi^{'}(t)]^\alpha = X_{\alpha\pi}^{'}(t),\ \ 0 \leq \alpha \leq 1, \tag{4}$$

provided that is equation defines a fuzzy interval $X_\pi^{'}(t) \in \mathcal{F}$. Note that, if the family $\{X_{\alpha\pi}^{'}(t)\}$ satisfies the conditions of the Representation Theorem [10], then there exists the π-derivative $X_\pi(t)^{'}$ of the fuzzy function X.

Example 2 *Consider the fuzzy function* $X : (0, +\infty) \to \mathcal{F}_C$ *defined by*

$$X(t)(s) = \begin{cases} \frac{s}{t} + 1 & \text{if} \quad -t \leq s \leq 0 \\ -\frac{s}{t^2} + 1 & \text{if} \quad 0 \leq s \leq t^2 \\ 0 & \text{if} \quad s \notin [-t, t^2]. \end{cases}$$

Then, for all $\alpha \in [0,1]$ *we have*

$$[X(t)]^\alpha = [f_\alpha(t), g_\alpha(t)] = [(\alpha-1)t, (1-\alpha)t^2],$$

and from (3)

$$X_{\alpha\pi}^{'}(t) = [(\alpha-1), 2(1-\alpha)t].$$

Now, the family $\{X_{\alpha\pi}^{'}(t)\}_{\alpha\in[0,1]}$ *satisfies the conditions from Representation Theorem [10] for each* $t > 0$. *Therefore, there exists the* π*-derivative* $X_\pi^{'}(t)$ *of the fuzzy function* X *for each* $t > 0$ *and*

$$\left[X_\pi^{'}(t)\right]^\alpha = [(\alpha-1), 2(1-\alpha)t].$$

ISBN: 978-989-95079-6-8

5 Fuzzy differential equation with π-derivative

We consider the initial value problem

$$X^{'}(t) = F(t, X(t)), \quad X(0) = X_0, \tag{5}$$

where $F : [0, T] \times \mathcal{F} \to \mathcal{F}$ is a continuous function and X_0 is a fuzzy interval.

Note that there are different interpretations for the problem (5). For example, in the problem (5), we can consider the H-derivative [15, 16, 17, 19, 23, 24, 25]; the Generalized derivative [4, 5, 6, 7, 8]. Also, the problem (5) can be rewrite as a family of differential inclusions [11, 12, 14]. In this section we will consider the π-derivative in the problem (5), i.e. we will study the following problem

$$X^{'}_{\pi}(t) = F(t, X(t)), \quad X(0) = X_0, \tag{6}$$

where $F : [0, T] \times \mathcal{F} \to \mathcal{F}$ is a continuous function and X_0 is a fuzzy interval. A solution of (6) is a continuous fuzzy function $X : T \to \mathcal{F}$ which verifies the equation (6) for each $t \in T$.

Taking in account the Theorem 1 and Theorem 2, we obtain a useful procedure to solve the fuzzy differential equation (6).

In fact, denote

$$[X(t)]^{\alpha} = [f_{\alpha}(t), g_{\alpha}(t)] \ , \quad [X_0]^{\alpha} = [u^0_{\alpha}, v^0_{\alpha}]$$

and

$$[F(t, X(t))]^{\alpha} = [U_{\alpha}(t, f_{\alpha}(t), g_{\alpha}(t)) \ , \ V_{\alpha}(t, f_{\alpha}(t), g_{\alpha}(t))].$$

Then, using the canocical representation (Section 3), we obtain

$$\pi\left([X(t)]^{\alpha}\right) = (x_{\alpha}(t) \ , \ \delta_{\alpha}(t)) \ , \ \pi\left([X_0]^{\alpha}\right) = \left(x^0_{\alpha} \ , \ \delta^0_{\alpha}\right)$$

where $x_{\alpha}(t) = f_{\alpha}(t)$, $\delta_{\alpha}(t) = g_{\alpha}(t) - f_{\alpha}(t)$, $x^0_{\alpha} = u^0_{\alpha}$ and $\delta^0_{\alpha} = v^0_{\alpha} - u^0_{\alpha}$. Also

$$\pi\left([F(t, X(t))]^{\alpha}\right) = \left(U^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t)), V^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t))\right)$$

Thus, we consider the following family of equations corresponding to the problem (6)

$$\begin{array}{l} D_{\pi}(x_{\alpha}(t), \delta_{\alpha}(t)) = \left(U^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t)), V^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t))\right) \\ (x_{\alpha}(0), \delta_{\alpha}(0)) = (x^0_{\alpha}, \delta^0_{\alpha}) \end{array} \tag{7}$$

with $\alpha \in [0, 1]$.

Now, using the Theorem 1 and 2 and the problem (7), we obtain a solution from problem (6). For this, we proceed as follows:

(i) Solve the differential system

$$\left\{ \begin{array}{ll} x^{'}_{\alpha}(t) = U^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t)), & x_{\alpha}(0) = x^0_{\alpha} \\ \delta^{'}_{\alpha}(t) = V^1_{\alpha}(t, x_{\alpha}(t), \delta_{\alpha}(t)), & \delta_{\alpha}(0) = \delta^0_{\alpha} \end{array} \right.$$

for x_{α} and δ_{α};

(ii) from class $(x_{\alpha}(t), \delta_{\alpha}(t))$ we obtain the interval

$$X_{\alpha}(t) = \left\{ \begin{array}{ll} [x_{\alpha}(t), x_{\alpha}(t) + \delta_{\alpha}(t)] & \text{if} \quad \delta_{\alpha}(t) \geq 0 \\ [x_{\alpha}(t) + \delta_{\alpha}(t), x_{\alpha}(t)] & \text{if} \quad \delta_{\alpha}(t) < 0 \end{array} \right.$$

(iii) By using the Representation Theorem [10], we build a fuzzy solution $X(t)$ such that $[X(t)]^{\alpha} = X_{\alpha}(t)$, for all $\alpha \in [0, 1]$.

Example 3 *Let us consider the fuzzy malthusian problem*

$$\left\{ \begin{array}{l} X'_{\pi}(t) = -\lambda X(t) \\ X(0) = \ X_0, \end{array} \right. \tag{8}$$

where $\lambda > 0$ and, as in [12], the initial condition X_0 is a symmetric triangular fuzzy number with support $[-a, a]$. That is,

$$[X_0]^{\alpha} = [-a(1 - \alpha), a(1 - \alpha)].$$

For obtain a solution of (8), we have solve the following differential system

$$\left\{ \begin{array}{ll} x^{'}_{\alpha}(t) = -\lambda x_{\alpha}(t), & x_{\alpha}(0) = -a(1 - \alpha) \\ \delta^{'}_{\alpha}(t) = -\lambda \delta_{\alpha}(t), & \delta_{\alpha}(0) = 2a(1 - \alpha). \end{array} \right.$$

The solutions of this system are

$$x_{\alpha}(t) = -a(1 - \alpha)e^{-\lambda t} \quad \textit{and} \quad \delta_{\alpha}(t) = 2a(1 - \alpha)e^{-\lambda t},$$

where $\delta_{\alpha}(t) \geq 0$ for all $t > 0$. Thus there exists a fuzzy solution from (8) $X(t)$ for all $t > 0$, such that

$$\begin{array}{rcl} [X(t)]^{\alpha} & = & [x_{\alpha}(t), x_{\alpha}(t) + \delta_{\alpha}(t)] \\ & = & [-a(1 - \alpha)e^{-\lambda t}, a(1 - \alpha)e^{-\lambda t}]. \end{array}$$

6 Conclusions

In this paper we present the concept of π-derivative for fuzzy function using the canonical representation of a class equivalence of two intervals. Those classes of equivalence are obtained by embedding the space of closed and bounded intervals in a Banach space using the embedding theorem [22].

In a forthcoming paper we will study the properties of the π-derivative fuzzy and the connection between this approach and the π-derifferentiabiliy in the fuzzy context introduced in [21]. Also, we will study the connection with others concept of fuzzy derivatives. For example, the connection with the G-differentiability introduced in [4]. Note that this concept of π-derivative is equivalent to generalized derivative in the case of set-valued function [9].

In Section 5 we study fuzzy differential equations considering π-derivative. In particular we present a form for obtain a fuzzy solution from FDEs. This approach have some advantages in relation to others interpretation, for example, an advantage that have the fuzzy solutions obtained considering π-derivative seem to be more intuitive than other solution using H-derivative (Example 3). It is worthy to stress also that the interpretation using extension principle [7, 18] and fuzzy differential inclusion [11, 12, 14] has a disadvantage because we cannot speak about the fuzzy derivative. In this issue, in a forthcoming paper, we will study the existence of solution of a fuzzy differential equation considering π-derivative.

Acknowledgment

The research in this paper has been partially supported by Ministerio de Ciencia e Innovación, Spain, through grant MTM2008-00018 and by Conicyt-Chile by project Fondecyt 1080438.

ISBN: 978-989-95079-6-8

References

[1] J.P. Aubin and A. Cellina *Differential Inclusions* Springer-Verleg, New york, 1984.

[2] S.M. Assev Quasilinear operators and their application in the theory of multivalued mappings *Proc. Steklov Inst. Math.* 2, 23–52, 1986.

[3] H.T. Banks and M.Q. Jacobs A differential calculus for multifunctions *J. Math.Anl. Appl.* 29, 246–272, 1970

[4] B. Bede and S. G. Gal Generalizations of the differentiability of fuzzy number valued functions with applications to fuzzy differential equation *Fuzzy Sets and Systems* 151, 581–599, 2005.

[5] B. Bede, I. J. Rudas and A. L. Bencsik First order linear fuzzy differential equations under generalized differentiability *Information Sciences* 177, 1648–1662, 2007.

[6] Y. Chalco-Cano and H. Román-Flores On the new solution of fuzzy differential equations *Chaos, Solitons & Fractals* 38, 112–119, 2008.

[7] Y. Chalco-Cano and H. Román-Flores Comparation between some approaches to solve fuzzy differential equations *Fuzzy Sets and Systems* 160, 1517–1527, 2009.

[8] Y. Chalco-Cano, M.A. Rojas-Medar and H. Román-Flores Sobre ecuaciones diferenciales difusas *Boletín de la Sociedad Española de Matemática Aplicada* 41, 91–99, 2007.

[9] Y. Chalco-Cano and H. Román-Flores Generalized derivative and π-derivative for set-valued mappings, Submitted to publication.

[10] P. Diamond and P. Kloeden *Metric Space of Fuzzy Sets: Theory and Application* World Scientific, Singapure, 1994.

[11] P. Diamond Brief note on the variation of constants formula for fuzzy differential equations *Fuzzy Sets and Systems* 129, 65–71, 2002.

[12] P. Diamond Time-dependent differential inclusions, cocycle attractors and fuzzy differential equations *IEEE Trans. Fuzzy System* 7, 734–740, 1999.

[13] M. Hukuhara Integration des applications mesurables dont la valeur est un compact convexe *Funkcial Ekvac.* 10, 205–223, 1967.

[14] E. Hüllermeier An approach to modeling and simulation of uncertain dynamical systems *International Journal Uncertainty, Fuzziness Knowledge-Bases Systems* 5, 117–137, 1997.

[15] O. Kaleva Fuzzy differential equations *Fuzzy Sets and Systems* 24, 301–317, 1987.

[16] O. Kaleva A note on fuzzy differential equations *Nonlinear Analysis* 64, 895–900, 2006.

[17] V. Lakshmikantham and R.N. Mohapatra *Theory of Fuzzy Differential Equation and Inclusions* Taylor & Francis, London, 2003.

[18] M. Misukoshi, Y. Chalco-Cano, H. Román-Flores and R.C. Bassanezi Fuzzy differential equations and the extension principle *Information Sciences* 177, 3627–3635, 2007.

[19] J.J. Nieto and R. Rodríguez-López Euler polygonal method for metric dynamical systems *Information Sciences* 177, 587–600, 2007.

[20] N.V. Plotnikova Systems of linear differential equations with π-derivative and linear differential inclusions, *Sbornik: Mathematics* 196, 1677–1691, 2000.

[21] M. Puri and D. Ralescu Differential and fuzzy functions *J. Math. Anal. Appl.* 91, 552–558, 1983.

[22] H. Radström An embedding theorem for spaces of convex sets *Proc. Amer. Math. Soc.* 3, 165–169, 1952.

[23] R. Rodriguez-López Comparison results for fuzzy differential equations *Information Sciences* 178, 1756–1779, 2008.

[24] L.J. Rodríguez-Muñis, M. López-Díaz, M.A. Gil The s-differentiability of a fuzzy-valued mapping *Information Sciences* 151, 283–299, 2003.

[25] H. Román-Flores and M. Rojas-Medar Embedding of level-continuous fuzzy sets on Banach spaces *Information Sciences* 144, 227–247, 2002.

[26] L. Stefanini and B. Bede Generalized Hukuhara differentiability of interval-valued functions and interval differential eqautions, *Nonlinear Analysis*, In Press.

[27] A. Tolstonogov. *Differential Inclusions in a Banach Space* Kluwer Academic Publishers, The Netherlands, 2000.

ISBN: 978-989-95079-6-8

Fuzzy logic approach to represent and propagate imprecision in agri-environmental indicator assessment

Assaghir Zainab[1,2] Girardin Philippe[1] Napoli Amedeo[2]

1.INRA: French National Institute for Agronomic Research
Colmar, France
2.LORIA: Lorraine Laboratory of Informatics Research and its Applications
Nancy, France
Email: {assaghiz,napoli}@loria.fr, {zassaghir,girardin}@colmar.inra.fr

Abstract— *The indicator of groundwater contamination developed and used in agriculture is calculated from data available in the field or data estimated by an expert. The modeling of this indicator generally requires a large number of parameters whose measure is imprecise. Several information sources provide information about the same imprecise quantities which have to be combined for defining what is called an "indicator of groundwater contamination" (I_{gro}). This indicator estimates the impact of cultivation practices on the groundwater contamination.*

In this paper, we explore a possibilistic information fusion method by using the notion of maximal coherent subsets to represent the imprecisions of multisource variables of the indicator. We also calculate the bounds of this indicator, and we propagate imprecision by using an interval analysis. Finally, we present the indicator's results for pesticides applied on different crops.

Keywords— Possibility theory, fuzzy set, fusion, maximal coherent subset, interval analysis, indicator

1 Introduction

Data used in application domains are often imperfect: imprecise, uncertain, incoherent... Some data come from information sources (database, expert) and others are approximative measurements. When multiple sources –such as a database, an expert, a website, a book– deliver information about some unknown quantity, aggregating this information can be a tedious task, especially when information are incoherent. Actually, there are many approaches to specify the impact of input parameter imprecision of a model, such as statistical sampling methods, the Monte-Carlo method [1], and bayesian method. These methods require either a significant number of data and a lot of computation time, or an exact definition of the statistical properties of the input parameters. There exists also alternative approaches based on fuzzy set and possibility theory [2, 3] to express, in a non-probabilistic sense, imprecision of parameters. Each approach represents parameters imprecision in a specific way. When, we need to aggregate data from multiple sources, fuzzy approaches show more flexibility in the treatment of incoherent information. Recently, Destercke et al [4] explored a possibilistic information fusion using the notion of maximal coherent subsets to synthesize information from several incoherent data. In this paper, we will use this method to synthesize an information from several information sources to calculate variables used in indicator assessment.

An indicator measures a certain aspect of the agrosystem. It is proposed to help farmers to improve the environmental sustainability of their agricultural practices and to reduce the pollution of environment such as indicator of groundwater contamination I_{gro}. It estimates the possibility for a pesticide to reach groundwater through leaching. The modeling and the assessment of an indicator generally require a large number of parameters whose measure is imprecise such as the variable "soil depth", that indicates how thick the soil cover is. Multiple sources provide information about the imprecise quantities which compose the indicator I_{gro}, for example, pesticide characteristics such as pesticide half-life $DT50$. The calculation of I_{gro} is based on a system of decision rules using fuzzy sets [5]. Authors in [5] do not take into account the imprecision in parameters of the indicator.

In this paper, we present an original work that, at our knowledge, has not been carried out until now. We propose here an analysis of imprecision of the indicator, especially variables used and their nature. We are interested in imprecision of parameters provided by several sources. Then, we use imprecision for evaluating the amplitude and the effect of imprecision on the value of the indicator I_{gro}. Firstly, we change the input values of variables in the process of calculation. Then, we use intervals of values in place of single values of input variables in the calculation process of the indicator I_{gro}. We also explore a fusion method, using the notion of maximal coherent subset, for summarizing information about the quantity given by several sources [4]. Finally, we calculate a variation interval for the indicator I_{gro} given by the input intervals of the variables.

The paper is divided as follows: the first section presents the indicator of groundwater contamination I_{gro}, its definition, structure and data used. Theoretical preliminaries are introduced in the third section. Then, the fourth section describes the fusion method used and the way how maximal coherent subsets can be used to obtain intervals of the imprecise variables; we also explore a way to compute the lower and upper bounds of the indicator. Then, the fifth section presents the indicator of groundwater contamination, its definition and its structure. Finally, the sixth section shows the results of the indicator for four pesticides applied in different crops.

2 Application description

In this section, we define the indicator I_{gro}, its structure and data used in the process of calculation.

ISBN: 978-989-95079-6-8

2.1 The risk of groundwater contamination I_{gro}

2.1.1 Definition and data used

The indicator module I_{gro} reflects the potential of a pesticide to reach groundwater through leaching and to affect its potential use as a source of drinking water. I_{gro} as proposed by Van der Werf and Zimmer [5] depends on four input variables : (1) pesticide leaching potential (GUS); (2) position ; (3) leaching risk, and (4) toxicity of the pesticide for humans (based on Acceptable Daily Intake). Table 1 shows variables used to calculate I_{gro}.

Table 1: All variables for the I_{gro} and their description.

Variables	Description
Pesticide characteristics (quantitative variables)	
GUS	pesticide leaching potential
DT50	pesticide half-life
koc	organic-carbon partition
ADI	Acceptable Daily Intake
Environmental characteristics (qualitative variables)	
Leaching risk	Quantity leaching
Application characteristics (quantitative variables)	
Position	Position of application

Variable "pesticide leaching potential" (GUS) provides an estimation of the risk of leaching of the compound. This potentiel is calculated by the formula:

$$GUS = \log(DT50) * (4 - \log(koc)) \quad (1)$$

where variable DT50 is the pesticide half-life and "koc" is the organic-carbone partition coefficient.
Variable DT50 is the time required for the pesticide concentration under defined conditions to decrease to 50% of the amount in application. Variable "koc" is the organic-carbon constant that describes the tendency of a pesticide to bind to soil particles. Variable "position" is the position of application of the pesticide (on the crop, on the soil, in the soil). This position is the interception rate of active ingredient per leaf area of studied culture. It is obtained from variable "soil cover". Variable "soil cover" is a value thats depends on the leaf surface of the culture. The variation of "soil cover" over time is fixed. Then, the position depends on the date of application by the farmer. The variable "leaching risk" depends on characteristics of the soil. In I_{gro}, the estimation of soil "leaching risk" is given by experts on a scale between 0 (minor leaching risk) and 1 (major leaching risk). The "position" and "leaching risk" take values between 0 and 1. The variable "Acceptable Daily Intake" (ADI) reflects chronic toxicity from humans. The pesticide properties (DT50, koc, and ADI) are given by several heterogeneous information sources such as Agritox (database, France), RIVM (database, Netherlands), Pesticide Manual (book, england), Dabene (expert, France), etc .... For each pesticide, these infomation sources give fuzzy intervals for each characteristic. Let us consider the following example: Three sources (Pesticide manuel: a book, Agritox: a database, Dabene: an expert) provide information, about the variable "DT50" (pesticide half-life) of "I_{gro}", in term of two intervals Int_1 and Int_2. The information is summarized in Table 2 and represented on Figure 1. The interval Int_2 represents the most plausible values for the variable DT50.

Table 2: Example information from sources

Source	Int_1	Int_2
Dabene	[10,42]	[20,35]
Agritox	[46,53]	50
Pesticide Manual	[21,56]	[35,40]

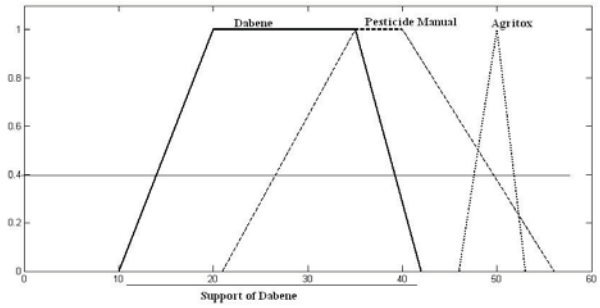


Figure 1: An example of values of DT50 given by sources, α-cut with $\alpha = 0.4$ and the support of data given by Dabene

2.1.2 The structure of the indicator of groundwater contamination I_{gro}

The indicator of groundwater contamination I_{gro} is calculated on a scale between 0 (maximal risk) and 10 (minimal risk). These values are calculated according to fuzzy if-then rules and to the degree of membership of the input variables to fuzzy subsets. The mechanism is explained below.

To calculate the indicator I_{gro}, we need to aggregate expertise knowledge and variables heterogeneous (qualitatives and quantitatives), authors in [5] are used fuzzy control [6] for achieving this two goals. For all input variables, two fuzzy subsets F (Favourable, i.e. the sets of values giving rise to acceptable environmental effect) and U (Unfavourable, i.e. the set of values giving rise to unacceptable environmental effect) are defined. Table 3 shows favourable and unfavourable limits for input variables of I_{gro}, which are extracted from literature or based on human expert knowledge.

Table 3: Favourable and unfavourable limits of input variables

Variable	Favourable limit	Unfavourable limit
DT50	1	30
GUS	1.8	2.8
ADI	1	10^{-4}
Position	1	0
Leaching risk	0	1

For each class (favourable, unfavourable, and transition interval which corresponds to values between favourable and unfavourable limits), we have defined a membership function on knowledge of experts.
The membership of values of input variables takes any value in the interval [0,1]. The value 0 represents complete non-membership and the value 1 represents complete membership. Values between 0 and 1 are used to represent partial membership. The membership function is defined in such a way that the value of an input variable either belongs fully to one of the

ISBN: 978-989-95079-6-8

two fuzzy subsets (favourable, unfavourable), or partially to both. In the latter case, the value is within a transition interval. We used membership functions that are sinus shaped in the transition interval, as they provide smoother variations of the output values than membership functions that are linear in the transition interval [5].
For example, experts classify the leaching potential GUS into three classes. They affect pesticides classified as "leacher" (i.e. GUS $>$ 2.8 or called unfavourable subset) a membership value of 1 for the fuzzy subset U and membership value of 0 for the fuzzy subset F. Pesticides classified as "no-leacher" (GUS $<$ 1.8 or called favourable subset) are given a membership value of 0 for the subset U and a membership value of 1 for the fuzzy subset F. The class of borderline compounds (1.8 $<$ GUS $<$ 2.8) falls within a transition interval where the membership value for F decreases from 1 (GUS = 1.8) to 0 (GUS= 2.8), and the membership value of U increases from 0 to 1 (see Fig. 2). The favourable (respectively unfavourable) membership function "F-Function" (respectively "U-Function") of GUS is the degree for GUS to be in the favourable subset (respectively unfavourable subset). We have also "U-Function" = 1 - "F-Function". The function of subset favourable of GUS is $0.5 + 0.5\cos(\pi(\text{GUS} - 1.8))$. If variable GUS has a value of 2.5, then "F-Function" = 0.16 and "U-Function" = 0.84.

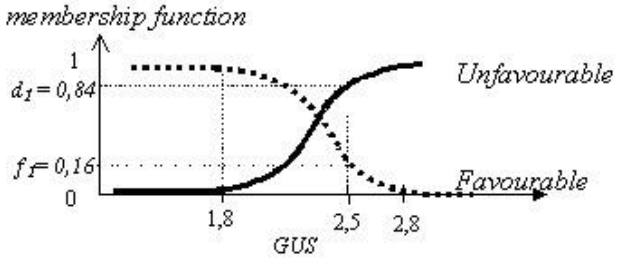


Figure 2: F-Function and U-Function function of variable GUS

The calculations are carried out according to a set of if-then rules. The experts attribute values between 0 and 10 for each rule.They consider firstly –what determines the leaching– the properties of organic matter (molecule), so they use variable GUS. If the variable GUS is favourable (GUS $<$ 1.8), then there is no problem. The molecule does not present the characteristics of leaching. If variable GUS is unfavourable (GUS $>$ 2.8), the experts are interested in the other variables of indicator. When a molecule is able to leach, it is able to reach the soil. Then, the experts use variable "position" of application. If the variable "position" is favourable, experts set a value to 9 for indicating the risk of leaching, but it is not too important. Then the environment is taken into account through the variable "leaching risk" (is the molecule leaching or not?, is the soil sensitive to the leaching?) and, finally they set on the variable "ADI".
Fig. 3 shows the if-then rules of I_{gro}. For example, if all input variables of I_{gro} are F then conclusion is 10. If all input variables of I_{gro} are U then conclusion is 0.

For each rule, we obtain the truth value by applying the minimum operator on the set of membership of the rule. For the first rule, if all input variables are F, the memberships of

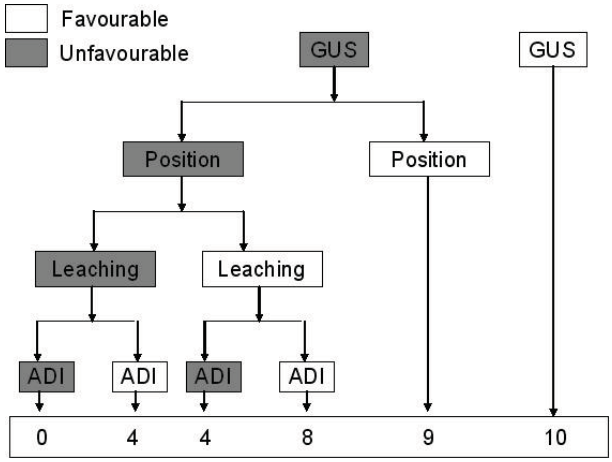


Figure 3: If-then rules of the indicator I_{gro} (I_{gro} has $2^4 = 16$ rules)

four input variables are f_i for each variable, then the truth value is $\omega_1 = \min_{i=1,\ldots,4}(f_i)$. The final score of the indicator is the average of rule conclusions weighted by their values of truth

$$I_{gro} = \frac{\sum_{i=1}^{16} \omega_i s_i}{\sum_{i=1}^{16} \omega_i}$$

3 Preliminaries

3.1 Problem statement

In this paper, we consider the indicator of groundwater contamination I_{gro}, which depends on parameters: GUS, DT50, "koc", ADI, "leaching risk" and "position". It is calculated by Van der Werf and others in [5] as $I_{gro} = \frac{\sum_{i=1}^{16} \omega_i s_i}{\sum_{i=1}^{16} \omega_i}$, where ω_i depends of variables of I_{gro}, and $s_i \in [0, 10]$.
We are interested here in multisource variable DT50, koc, ADI. Assume that the set of n information sources provide n fuzzy intervals about these variables (see Table 2). We have to search the most plausible values to calculate the indicator. We use here the support of fuzzy interval (α-cut with $\alpha = 0$ i.e. the interval Int_1 in Table 2). We use the notion of maximal coherent subset into a possibilistic fusion to synthesize a final result for multisource variables. These results combined with values of others variables will be used to compute the bounds (lower and upper) of the indicator I_{gro}.
In this paper, $\forall x \in \mathbb{R}$, x^- and x^+ represent the lower and upper bounds of interval where x can vary.

3.2 Possibility theory, fuzzy sets

In this section, we briefly summarize basic concepts of possibility theory, which are needed for understanding this paper. Fuzzy sets were firstly introduced by Zadeh [2] as a possible way to handle uncertainty in processing incertain or imprecise data, and to control expert knowledge. This theory allows the notion of graduation to express whether an element belongs to a set (For more details, see[2, 3, 7]).

3.2.1 Possibility theory

Possibility theory was introduced in 1978, in connection with fuzzy set theory, to allow reasoning to be carried out on imprecise or vague knowledge, making it possible to deal with uncertainties on knowledge [3, 8]. The basic tool of possibility

ISBN: 978-989-95079-6-8

theory is possibility distribution. A possibility distribution is equivalent to the definition of a normalized fuzzy membership function. We are interested in this theory, in this paper, because we need to synthesize information of a variables which come from various sources. When we want to synthetize the value of a variable (i.e. approximate the value of the variable depending a set of a different sources (This process is called "information fusion" (see here after). In such a situation, possibility theory can be used to handle this kind of problems. This theory presents many choices of fusion operators in different contexts.

3.3 Information fusion

Definition 1 *Information fusion consists of merging, or exploiting conjointly, several sources of information for answering questions of interest and make proper decisions.*

A large set of information fusion operators has been designed in the possibility theory framework [3]. They can be split into three subsets, according to the behavior of operators:

- Conjunctive type operators: it is the equivalent of a set intersection. It makes the assumption that all sources are reliable (i.e. all sources give information with high confidence), and usually results in very precise information. If there is an incoherence between variable values (i.e. detected by domain expert), then the result of the conjunction becomes poorly reliable, or even empty.
- Disjunctive type operators: it is the equivalent to a set union. It makes the assumption that at least one source is reliable. The result of a disjunctive operator can be considered as very reliable.
- Trade-off type operators: they are compromise operators between conjunctive and disjunctive operators. They are typically used when sources are partially conflicting (i.e. in contradiction). As its name indicates, such an operator tries to make a trade-off between disjunction and conjunction for achieving a good balance between informativeness and reliability.

4 Methodology

4.1 Fusion based on maximal coherent subsets to assess imprecision of multi-source variables

Information sources which supply the information about imprecise variables (DT50, koc, ADI) are heterogeneous (book, expert, database). They come from different places (France, USA, Netherlands), then the type of soil, weather and laboratory conditions are not similar. Furthermore, we ignore the reliability of sources (they do not supply a degree of confidence with the value). They provide data in term of a fuzzy interval (or fuzzy number). Also, no information about conflict and relations between sources is available. Then, we can not use a conjunctive rule (respectively disjunctive and compromise rule). We need a fusion method which take into account all information given by sources (i.e. without discarding any source). The notion of maximal coherent subset is a natural way for achieving this goal. The fusion method using the notion of maximal coherent subset consists firstly inn applying a conjunctive operator into each non-conflicting subset of sources, then a disjunctive operator between those partial results. With such a method, as much precision as possible is gained when we do not ignore any sources [9, 10]. We will explain in details how this approach applies to support of fuzzy intervals of variables of I_{gro} (see the section 4.3).

4.2 Computing maximal coherent subsets of intervals

In this section, we describe how we can obtain values of multisource variable using the fusion method based on maximal coherent subset of intervals, which are give by information sources. Let us consider n intervals, $I_i = [a_i, b_i], (i = 1, \ldots, n)$. This method allows to find every maximal subsets of sources. A maximal subset is obtained when $\bigcap_{i\in[n]} I_i \neq \emptyset$, where $[n]$ represents the set of 2^n subsets of the set $\{1, \ldots, n\}$. Then, we apply the union of these partial results (i.e. $\bigcup_j \bigcap_{i\in[n]} I_i$).

Algorithm 1: Maximal coherent subset of intervals

Input: n intervals
Output: List of m maximal coherent subsets I_j
List=$\emptyset$;$j = 1$;$I_j = \emptyset$;
Order in an increasing order
$\{a_i, i = 1, \ldots, n\} \cup \{b_i, i = 1, \ldots, n\}$;
Rename them $\{c_i, i = 1, \ldots, 2n\}$with $type(i) = a$
if $c_i = a_k$ and $type(i) = b$ if $c_i = b_k$;
for $i = 1$ to $2n - 1$ **do**
  if $type(i) = a$ **then**
    Add source k to I_j t.q. $c_i = a_k$;
    if $type(i + 1) = b$**then**
      Add k to List (I_j);
      $j = j + 1$;
  else
    Remove source k from I_j such as $c_i = b_k$;

Algorithm 1, that finds maximal coherent subsets, was introduced by Dubois and others in [11]. The algorithm 1 is linear in the number of intervals, and thus computationally efficient. The algorithm is based on an increasing sorting of the interval end-points into a sequence $(c_i)_{i=1,\ldots,2n}$. Each time and only then, an element c_i of type upper bound which is an upper bound of an interval, followed by an element c_{i+1} of type lower bound wich is a lower bound of an interval meet, a maximally coherent set is obtained.

4.3 Example of application

We consider the variable DT50 (pesticide half-life) provided by three information sources (see Table 2): a French expert (Dabene), a French database (Agritox), and an England book (Pesticide Manual).
Using algorithm 1 on the level $\alpha = 0$ (i.e. the support Int_1 of fuzzy intervals in Table 2), we find two maximal coherent subsets $K_1 = \{\text{Dabene, Pesticide Manual}\}$ and $K_2 = \{\text{Agritox, Pesticide Manual}\}$. After applying the maximal coherent subset method, the result of the variable DT50 is: (Dabene $\cap$ Pesticide Manual) $\cup$ (Agritox $\cap$ Pesticide Manual) = [21,42] $\cup$ [46,53].

4.4 Computation method to the fuzzy weighted average

The objective of this section is to compute the bounds of the indicator I_{gro}. The function $f(s_1, s_2, \ldots, s_n, \omega_1, \ldots, \omega_n) = \frac{\sum_{i=1}^{n} \omega_i s_i}{\sum_{i=1}^{n} \omega_i}$, where ω_i is a fuzzy interval and $s_i \in \mathbb{R}$, is called

ISBN: 978-989-95079-6-8

fuzzy weighted average by Dong and Wong [12]. Authors in [12] proposed an algorithm to compute the fuzzy weighted average. Their algorithm is based on the α-cut representation of fuzzy sets and combinatorial interval analysis. Subsequently, Liou and Wang [13] suggested an improved fuzzy weighted average algorithm to simplify the computational process. Afterwards, Lee and Park [14] have proposed an even more efficient algorithm for fuzzy weighted average by reducing the number of arithmetic operations to $O(n \log n)$. The main idea of Lee and Park in [14] is to sort the s_i variables such as for all $i < j$, we have $s_i \leq s_j$. In this order, we can find a rank k with $1 < k < n$ such as for all $i \leq k$, the function f is decreasing with respect to ω_i, and for all $i > k$, the function f is increasing with respect to ω_i. Then, we can compute the lower and the upper bounds of the fuzzy weighted average in accordance with [15],which are bounds of I_{gro}, as:

$$f_k^- = \frac{\sum_{j=1}^{k}(\omega_j)^+.s_j + \sum_{j=k+1}^{n}(\omega_j)^-.s_j}{\sum_{j=1}^{k}(\omega_j)^+ + \sum_{j=k+1}^{n}(\omega_j)^-} \quad (2)$$

$$f_k^+ = \frac{\sum_{j=1}^{k}(\omega_j)^-.s_j + \sum_{j=k+1}^{n}(\omega_j)^+.s_j}{\sum_{j=1}^{k}(\omega_j)^- + \sum_{j=i+1}^{n}(\omega_k)^+} \quad (3)$$

Where ω_i^- and ω_i^+ are respectively the lower and upper bounds of ω_i. The equation (2) (respectively (3)) represent the lower (upper) bound of the level k and $k = 1, \ldots, n$. Then, we determinate the fuzzy interval of indicator such as introduced by Fortin and others in [15]:

$$[\min_{k=1,\ldots,n}(f_k^-), \max_{k=1,\ldots,n}(f_k^+)] \quad (4)$$

5 Results

In this application when computing I_{gro}, about fifteen information sources are used such as ARS (USA), Agritox (France), Pesticide manual (England)... These sources provide information based on a fuzzy interval or on a fuzzy number. In this paper, we use the support of fuzzy intervals given by the sources. The decision rules do not change. For each variable, we replace the input value by an input interval. We propagate the imprecision into computation indicator to calculate a fuzzy interval of indicator. For a pesticide application, an indicator of groundwater contamination score is calculated.
The application considers the indicator of groundwater contamination I_{gro} in several sites with different crops. The positions of application of pesticides is shown in Table 4. Its value is calculated as: "position $= \frac{\text{soil coverd by the crop}}{100}$". The "soil covered" by the crop is estimated by the user of the systems [16]. The variable "leaching risk" set the value of 0.9 [16].

Table 4: Values of the variables of application condition

Pesticide name	Position	soil covered
Chloridazone	0.9	90
Glyphosate	0.75	75
Nicosulfuron	0.1	10
Isoproturon	0.1	10

Pesticide properties (DT50, koc, and ADI) given by several information source are calculated by fusion based on maximal coherent subsets. Table 5 shows the results of pesticide characteristics. The values shown in Table 5 are convexified to simplify the calculation. By using the interval analysis [17] and the equation 1, we compute the fuzzy interval of variable GUS, such as:

$$\text{GUS}^- = \log(\text{DT50}^-) \times (4 - \log(\text{koc}^+)) \quad (5)$$

$$\text{GUS}^+ = \log(\text{DT50}^+) \times (4 - \log(\text{koc}^-)) \quad (6)$$

where ([DT50$^-$,DT50$^+$]) and ([koc$^-$,koc$^+$]) are fuzzy intervals of variables DT50 and Koc.

Table 5: Results of pesticide characteristics

Pesticide name	GUS	DT50	koc	ADI
Chloridazone	[2.53,3.19]	[21,46]	120	0.025
Glyphosate	[0.73,2.97]	[18,47]	[167,2640]	0.3
Nicosulfuron	[2.78,5.39]	[15,43]	[5,43]	0.4
Isoproturon	[2.68,3.03]	[22,28]	[80,99.99]	0.4

Afterwards, we compute the membership intervals of input variables of indicator. For each input variable that varies in $[a, b]$ and has a membership function f, we compute the membership interval by using the function:

$$f([a,b]) = \begin{cases} [f(a), f(b)] & \text{if } f \text{ is increasing} \\ [f(b), f(a)] & \text{if } f \text{ is decreasing} \end{cases} \quad (7)$$

Then, we calculate the truth intervals of the sixteen decision rules of I_{gro}. These truth intervals can be computed by minimum extension applied on intervals such as $\min([a, b], [c, d]) = [\min(a, c), \min(b, d)]$. Then, we use the equations (2) and (3) to compute respectively the lower and upper bounds of indicator for the sixteen α-cuts established by the fuzzy input membership intervals. Finally, we obtain the final interval of indicator by the equation (4).
The Table 6 shows the results of I_{gro} for different pesticides. The fuzzy interval in Table 6 represent the exact interval when the indicator is varying. It represents also the support, of the fuzzy interval of the indicator I_{gro}, that can obtain from all fuzzy intervals of input variables.

Table 6: Results of indicator

Pesticide name	Interval of I_{gro}
Chloridazone	[4.3,9.82]
Glyphosate	[0,10]
Nicosulfuron	[4.42, 4.45]
Isoproturon	[4.42,5.27]

The results presented in Table 6 show a lot of imprecision in some cases. We take example of the organic matter "glyphosate": the indicator obtained by application of glyphosate on the soil varies between 0 and 10. This imprecision is due to the variable GUS –which varies in [0.73,2.97]– that reach limits fixed by experts in Table 3 (i.e (lower limit) $<$ (favourable limit), and (upper limit) $>$ (unfavourable limit)). We note that the indicator varies in [0,10], when we have one of the input variables has a lower (respectively upper) limit exceeds favourable (respectively unfavourable) limit. In the

ISBN: 978-989-95079-6-8

calculation of only one value of indicator I_{gro} [5], the role of the variable GUS is more important than the role of others variables. In this paper, all variables have the same influence in the final result.

6 Discussion and conclusions

In this paper, we introduced the calculus of the indicator of groundwater contamination I_{gro}. We try to control the imprecisions of input variables by using a possibilistic fusion method based on the maximal coherent subsets. Then, we use the notion of interval analysis to propagate the imprecision in the process of calculus of indicator. A fusion method for merging fuzzy subsets, based on the notion of maximal coherent subsets, is proposed to represent the imprecision of input parameters of the indicator of groundwater contamination. This notion appears as a very natural way to conciliate two objectives, gaining information and considering all the conflicts between sources. The method of fusion is simple: it can be applied without any additional information, and its computational complexity remains affordable. The way is summarizes information is conceptually attractive: maximal coherent subsets are the best we can do in the presence of conflict.
We have also described a way to calculate the bounds of indicator. The method of fuzzy interval arithmetic used in this paper is also simple. The support of the fuzzy number of intervals is calculated. Also, we can calculate the membership function of indicators by computing several values of α-cut of input variables. This method allows to obtain exact fuzzy profiles of indicator with reducing time complexity as in the sampling case. It still remains to validate the fusion method using the notion of maximal coherent subset in contrast with other fusion rules. We have to calculate the fuzzy interval for all α-cut to obtain the distribution of I_{gro}. We plane to use the methodology described in this paper (representing by fuzzy intervals the imprecision and propagating this imprecision by interval analysis) to calculate other indicators such as indicator of pesticide in environment which has a process of calculation similar to process of I_{gro}.

References

[1] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.

[2] L.A. Zadeh. Fuzzy sets. *Information and Control*, 3(8):338–353, 1965.

[3] D. Dubois and H. Prade. *Théorie des possibilités*. Masson, Paris, 1985.

[4] S. Destercke, D. Dubois, and E. Chojnacki. Fusion using maximal coherent subset. In *IEEE, International Conference on Fuzzy Systems*, pages 445–450, 2007.

[5] H. Van der Werf and C. Zimmer. An indicator of pesticide environmental impact based on a fuzzy expert system. *Chemosphere*, 36(10):2225–2249, 1998.

[6] D. Driankov, H. Hellendoorn, and M. Reinfrank. *An introduction to fuzzy control*. Springer-Verlag New York, Inc., New York, NY, USA, 1993.

[7] B. Bouchon-Meunier. *La logique floue*. Presses Universitaires de France, Etats-Unis, 1999.

[8] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1:3–28, 1978.

[9] N. Rescher and R. Manor. On inference from inconsistent premises. *Theory and Decision*, 1:179–217, 1970.

[10] S. Benferhat, D. Dubois, and H. Prade. Reasoning in inconsistent startified knowledge bases. *Proceedings of the 26th inter. Symp. in Multi-valued logic(ISMVL)*, pages 184–189, 1996.

[11] D. Dubois, P. Hajek, and H. Prade. Knowledge-driven versus data-driven logics. *J. of Logic, Language and Information*, 9(1):65–89, 2000.

[12] W. M. Dong and F. S. Wong. Fuzzy weighted averages and implementation of the extension principle. *Fuzzy Sets and Systems*, 21(2):183 – 199, 1987.

[13] T.S. Liou and M. J. Wang. Fuzzy weighted average: An improved algorithm. *Fuzzy Sets and Systems*, 49(3):307 – 315, 1992.

[14] D. H. Lee and D. Park. An efficient algorithm for fuzzy weighted average. *Fuzzy Sets and Systems*, 87(1):39 – 45, 1997.

[15] J. Fortin, D. Dubois, and H. Fargier. Gradual numbers and their application to fuzzy interval analysis. *IEEE Trans. on fuzzy systems*, 16(2):388–402, 2008.

[16] C. Bockstaller and P. Girardin. Mode de calcul des indicateurs agri-environnementaux de la méthode indigo. Rapprt technique, INRA, 2003.

[17] Ramon Edgar Moore. *Interval Analysis*. Prentice-Hall, New York, 1966.

ISBN: 978-989-95079-6-8

Representation of States on MV-algebras by Probabilities on R-generated Boolean Algebras

Brunella Gerla[1] Tomáš Kroupa[2,3]
1. Department of Informatics and Communication, University of Insubria,
Via Mazzini 5, 21100 Varese, Italy
2. Institute of Information Theory and Automation of the ASCR
Pod Vodárenskou věží 4, 182 08 Prague, Czech Republic
3. Department of Mathematics, Faculty of Electrical Engineering
Czech Technical University in Prague, 166 27 Prague
Email: brunella.gerla@uninsubria.it, kroupa@utia.cas.cz

***Abstract*— *Any MV-algebra M can be embedded as a lattice in the Boolean algebra $B(M)$ that is R-generated by M. We relate the study of states on an MV-algebra M to the study of finitely additive probabilities on $B(M)$. In particular, we show that each state on M can be uniquely extended to a finitely additive probability on $B(M)$. In case that M is a PMV-algebra, the conditional state $s(a|b)$ defined for $a, b \in M$ with $s(b) \neq 0$ is extended to the classical conditional probability $p(a \cdot b|b)$ on $B(M)$ of the a-proportion of the event b, given the event b.*

***Keywords*—** MV-algebra, R-generated Boolean algebra, state, probability, conditional probability

1 Introduction

States on MV-algebras were investigated by Mundici in [1] as $[0, 1]$-valued additive functionals on (equivalence classes of) formulas in Łukasiewicz propositional logic with the intention to capture the notion of "average" truth degree of a formula. In [2] a probability theory on MV-algebras is systematically developed. The study of states is frequently enhanced by looking directly at certain probabilities induced by the states: there is a one-to-one correspondence between the states on an MV-algebra and the Borel probability measures on the maximal spectrum of the MV-algebra [3, 4], where the bijection is the Lebesgue integral of a continuous function on the maximal spectrum with respect to a uniquely determined Borel probability measure. In this paper we introduce another way of representing a state within a Boolean probability theory. By Jenča's result [5], each MV-algebra M is embedded as a lattice in a uniquely determined Boolean algebra $B(M)$ that is said to be R-generated by M (see [6], [5]). We will show how to relate the study of states over M to the study of certain finitely additive probabilities on $B(M)$. In particular, our approach sheds a new light on the definition of conditional state appearing in [7].

In Section 2 we repeat the basic notion concerning MV-algebras and R-generated Boolean algebras. Section 3 contains the main results: Proposition 3, in which we show that there is only one probability on $B(M)$ extending a state on M, and Proposition 4, which gives a geometrical and topological description of such extensions in case M is a semisimple MV-algebra. In Section 4 we relate conditional states on M to conditional probabilities on $B(M)$ (Proposition 5).

2 Preliminary Notions

2.1 MV-algebras

MV-algebras [8] were introduced by Chang in 1958 as the algebraic counterpart of propositional Łukasiewicz logic.

Definition 1. An MV-algebra is a structure $\langle A, \oplus, \neg, 0\rangle$ with a binary operation $\oplus$, a unary operation $\neg$ and a constant 0 such that $\langle A, \oplus, 0\rangle$ is an abelian monoid and the following equations hold for every $x, y \in A$:

$$\neg\neg x = x$$
$$x \oplus \neg 0 = \neg 0$$
$$\neg\left(\neg x \oplus y\right) \oplus y = \neg\left(\neg y \oplus x\right) \oplus x$$

The most important example of an MV-algebra is the real unit interval $[0, 1]$ equipped with operations $x \oplus y = \min(1, x + y)$ and $\neg x = 1 - x$. Indeed, the class of MV-algebras form a variety that is generated by $[0, 1]$.
On each MV-algebra A we define $1 = \neg 0$, $x \odot y = \neg\left(\neg x \oplus \neg y\right)$, $x \ominus y = x \odot \neg y$. Note that in the MV-algebra $[0, 1]$ we have $x \odot y = \max\left(0, x + y - 1\right)$ and $x \ominus y = \max(0, x - y)$.

Let A be an MV-algebra. For any two elements x and y of A we write $x \leq y$ if and only if $\neg x \oplus y = 1$. It follows that $\leq$ is a partial order that in the case of the algebra $[0, 1]$ coincides with the natural order. Further, defined connectives

$$x \vee y = \neg\left(\neg x \oplus y\right) \oplus y \quad (1)$$
$$x \wedge y = \neg\left(\neg x \vee \neg y\right) \quad (2)$$

are such that the structure $\langle A, \wedge, \vee, 0, 1\rangle$ is a distributive lattice with bottom element 0 and top element 1. An *MV-chain* is an MV-algebra in which the order relation $\leq$ is total.

For any MV-algebra M, the set $\{x \in M \mid x \oplus x = x\}$ is the largest Boolean algebra contained in M.

Given an MV-algebra A and a set X, the set A^X of all functions $f : X \to A$ becomes an MV-algebra if the operations $\oplus$ and $\neg$ and the element 0 are defined pointwise. MV-algebras of functions taking values in $[0, 1]$ can be characterized by means of their ideals.

An *ideal* of an MV-algebra A is a subset I of A such that $0 \in I$, if $x, y \in I$ then $x \oplus y \in I$, and if $x \in I$, $y \in A$ and $y \leq x$ then $y \in I$. An ideal is *maximal* if it is not contained in any proper ideal. An MV-algebra A is said to be *semisimple* if

ISBN: 978-989-95079-6-8

and only if A is non trivial and the intersection of all maximal ideals of A is $\{0\}$. By the *maximal spectrum* we mean the (nonempty) set of all maximal ideals of A. This set can be made into a compact Hausdorff space.

It can be shown that an MV-algebra A is semisimple if and only if A is isomorphic to a separating MV-algebra of $[0,1]$-valued continuous functions on some nonempty compact Hausdorff space (actually the maximal spectrum of A), with pointwise operations. For other notions related to MV-algebras we refer the reader to [9].

A *product MV-algebra* (or PMV-algebra, for short, see [10], [11]) is a structure $\langle A, \oplus, \neg, \cdot, 0\rangle$, where $\langle A, \oplus, \neg, 0\rangle$ is an MV-algebra and $\cdot$ is a binary associative and commutative operation on A such that for any $x, y, z \in A$, $x \cdot 1 = x$ and $x \cdot (y \ominus z) = (x \cdot y) \ominus (x \cdot z)$.

Note that $[0,1]$ is a PMV-algebra where the operation $\cdot$ is the usual multiplication of real numbers. Further, a semisimple MV-algebra A is a PMV-algebra if and only if A is closed with respect to the pointwise real product of functions and, in this case, the unique product on A is the pointwise real product of functions.

2.2 Boolean algebra R-generated by an MV-algebra

Let M be an MV-algebra. There exists a unique (up to a Boolean isomorphism) Boolean algebra $B(M)$ such that the lattice reduct of M

(i) is a sublattice of $B(M)$ containing both the elements 0 and 1 of M,

(ii) generates $B(M)$ as a Boolean algebra.

The Boolean algebra $B(M)$ is called an *R-generated Boolean algebra*. See [6, Chapter II.4] for details. For every $a \in B(M)$ there exists $n \in \mathbb{N}$ and a finite chain $a_1 \le \cdots \le a_{2n}$ in M such that

$$a = \bigvee_{i=1}^{n} (a_{2i} \setminus a_{2i-1}) \tag{3}$$

holds true in $B(M)$ and $(a_{2i} \setminus a_{2i-1}) \wedge (a_{2j} \setminus a_{2j-1}) = 0$ for each $i, j \in \{1, \ldots, n\}$ with $i \neq j$, where $\setminus$ denote the symmetric difference in the Boolean algebra $B(M)$.

Theorem 1 (Jenča [5], Theorem 2)**.** *Let M be an MV-algebra. Then there exists a surjective mapping $\varphi_M : B(M) \to M$ such that*

$$\varphi_M(a) = \bigoplus_{i=1}^{n} (a_{2i} \ominus a_{2i-1}), \quad a \in B(M), \tag{4}$$

where $a_1, \ldots, a_{2n} \in M$ are as in (3) *and the value of $\varphi_M(a)$ is independent on the choice of the representation* (3)*. Moreover, the mapping φ_M satisfies*

(i) $\varphi_M(1) = 1$,

(ii) $\varphi(a) = a$, *for every* $a \in M$,

(iii) if $a, b \in B(M)$ *are such that* $a \wedge b = 0$, *then* $\varphi_M(a) \odot \varphi_M(b) = 0$ *and* $\varphi_M(a \vee b) = \varphi(a) \oplus \varphi(b)$.

The following two examples appear in [5].

Example 1. Let M be an MV-chain. Given two elements $a, b \in M$, put $[a, b) = \{x \in M \mid a \le x < b\}$. By [6], $B(M)$ is isomorphic to the Boolean algebra of all subsets of M having the form

$$[a_1, b_1) \cup \cdots \cup [a_n, b_n)$$

where $a_i, b_i \in M$ and $[a_i, b_i) \cap [a_j, b_j) = \emptyset$, for every $i, j \in \{1, \ldots, n\}$ with $i \neq j$. Note that each $a \in M$ can be identified with $[0, a) \in B(M)$. Then

$$\varphi_M([a_1, b_1) \cup \cdots \cup [a_n, b_n)) = (b_1 \ominus a_1) \oplus \cdots \oplus (b_n \ominus a_n).$$

Example 2. Suppose now M is a semisimple MV-algebra, then M can be viewed as an MV-algebra of continuous functions from the maximal spectrum X of M to the unit interval $[0,1]$. Let $\mathcal{B} = B([0,1])$ be the Boolean algebra of all subsets of $[0,1]$ of the form $[a_1, b_1) \cup \cdots \cup [a_n, b_n)$ for some $n \in \mathbb{N}$, where $[a_i, b_i) \cap [a_j, b_j) = \emptyset$ with $i, j \in \{1, \ldots, n\}$ and $i \neq j$. Just as in Example 1, each element a of the MV-algebra $[0,1]$ can be identified with the interval $[0, a)$ of $\mathcal{B}$.

Since lattice operations in $[0,1]^X$ (and hence in M) are componentwise, then $B(M)$ is a Boolean subalgebra of $\mathcal{B}^X$. Each element $f \in M \subseteq [0,1]^X$ can be identified with the function $f^* \in B(M)$ such that $f^* : x \in X \mapsto [0, f(x)) \in \mathcal{B}$. Then, according to (3), for every $g \in B(M)$ there is $n \in \mathbb{N}$ and $a_1 \le b_1 \le \ldots \le a_n \le b_n \in M$ such that

$$g = \bigvee_{i=1}^{n} b_i^* \setminus a_i^*. \tag{5}$$

The mapping φ_M is defined as follows. For every $g \in B(M)$ and any $x \in X$, we have, by (5):

$$g(x) = \bigvee_{i=1}^{n} [0, b_i(x)) \setminus [0, a_i(x)).$$

It suffices to let $\varphi_M(g) : x \in X \mapsto (b_1(x) \ominus a_1(x)) \oplus \cdots \oplus (b_n(x) \ominus a_n(x)) \in [0,1]$.

Observe that the length n of the chain representing g does not depend on the coordinate (maximal ideal) x.

3 States and Probabilities

A *state* on an MV-algebra M is a mapping $s : M \to [0,1]$ such that $s(1) = 1$ and $s(a \oplus b) = s(a) + s(b)$, for every $a, b \in M$ with $a \odot b = 0$. In case that M is a Boolean algebra, then we denote a state on it by p and call p a *(finitely additive) probability.* Let $\mathscr{S}(M)$ be the *state space* of M, that is, the (nonempty) set of all states on M. Analogously, by $\mathscr{S}(B(M))$ we denote the set of all probabilities on the R-generated Boolean algebra $B(M)$. In the sequel we will introduce and study the relation between the two state spaces.

Proposition 1. *If s is a state on an MV-algebra M, then there is a probability p on $B(M)$ such that $p(a) = s(a)$ for every $a \in M$.*

Proof. Put

$$p(a) = s(\varphi_M(a)), \quad a \in B(M), \tag{6}$$

where φ_M is the mapping from Theorem 1, and observe that this definition is correct since, by the same Theorem 1, it does not depend on the representation of a given in (3). Then $p(1) = s(\varphi_M(1)) = s(1) = 1$. Let $a, b \in B(M)$ be such that $a \wedge b = 0$. Then Theorem 1(iii) yields

$$\begin{aligned} p(a \vee b) &= s(\varphi_M(a \vee b)) = s(\varphi_M(a) \oplus \varphi_M(b)) \\ &= s(\varphi_M(a)) + s(\varphi_M(b)) = p(a) + p(b). \end{aligned}$$

Since every $a \in M$ is a fixed point of φ_M due to Theorem 1(ii), we get that p coincides with s over M. □

ISBN: 978-989-95079-6-8

The next proposition shows that every probability on $B(M)$ is uniquely determined already on the embedded MV-algebra M.

Proposition 2. *If p and p' are two probabilities on $B(M)$ and $p(b) = p'(b)$ for every $b \in M$, then $p = p'$.*

Proof. Let $a \in B(M)$. By equation (3), there exists a finite chain $a_1 \leq \cdots \leq a_{2n}$ in M such that

$$a = \bigvee_{i=1}^{n} (a_{2i} \setminus a_{2i-1}) \tag{7}$$

in $B(M)$ and $(a_{2i} \setminus a_{2i-1}) \wedge (a_{2j} \setminus a_{2j-1}) = 0$, for each $i, j \in \{1, \ldots, n\}$ with $i \neq j$. Hence

$$\begin{aligned} p(a) &= p\Big(\bigvee_{i=1}^{n} (a_{2i} \setminus a_{2i-1})\Big) = \sum_{i=1}^{n} p(a_{2i} \setminus a_{2i-1}) \\ &= \sum_{i=1}^{n} (p(a_{2i}) - p(a_{2i-1})) = \sum_{i=1}^{n} (p'(a_{2i}) - p'(a_{2i-1})) \\ &= \sum_{i=1}^{n} p'(a_{2i} \setminus a_{2i-1}) = p'\Big(\bigvee_{i=1}^{n} (a_{2i} \setminus a_{2i-1})\Big) \\ &= p'(a). \end{aligned}$$

□

Putting together Proposition 1 with (6) and Proposition 2, we get the following uniqueness result.

Proposition 3. *If s is a state on an MV-algebra M, then the probability $p = s \circ \varphi_M$ on $B(M)$ is the unique probability on $B(M)$ such that $p(a) = s(a)$ for every $a \in M$.*

The above introduced correspondence between the states in $\mathscr{S}(M)$ and the probabilities in $\mathscr{S}(B(M))$ works in general only in one direction. Define a mapping $\Phi : \mathscr{S}(M) \to \mathscr{S}(B(M))$ by

$$\Phi(s)(a) = s(\varphi_M(a)), \tag{8}$$

for every $s \in \mathscr{S}(M)$ and every $a \in B(M)$. Since every MV-algebra M is embedded into $B(M)$, it is natural to expect that the state space of the Boolean algebra $B(M)$ is much larger than that of M. Indeed, already in case that M is the standard MV-algebra $[0, 1]$, the set $\mathscr{S}([0, 1])$ contains only one element (the state s defined by $s(x) = x$, for every $x \in [0, 1]$). So the image of $\Phi(\mathscr{S}([0, 1]))$ is a singleton, but $\mathscr{S}(B([0, 1]))$ contains infinitely-many probabilities: for every $x \in [0, 1]$, the mapping

$$a \in B([0, 1]) \mapsto \begin{cases} 1, & x \in a, \\ 0, & \text{otherwise}, \end{cases}$$

is a two-valued probability and hence it belongs to $\mathscr{S}(B([0, 1]))$.

Properties of the operator Φ are summarized below. In particular, we will show that Φ preserves the geometrical-topological structure of $\mathscr{S}(M)$. In order to show this, the convex sets $\mathscr{S}(M)$ and $\mathscr{S}(B(M))$ are endowed with the subspace topology of the product spaces $[0, 1]^M$ and $[0, 1]^{B(M)}$, respectively, so that both $\mathscr{S}(M)$ and $\mathscr{S}(B(M))$ became compact spaces (see [1]).

Proposition 4. *Let M be a semisimple MV-algebra. Then:*

(i) $\Phi(s)(a) = s(a)$, for every $s \in \mathscr{S}(M)$ and every $a \in M$;

(ii) the mapping Φ is an affine homeomorphism of $\mathscr{S}(M)$ onto the compact convex set $\Phi(\mathscr{S}(M))$;

(iii) the set $\Phi(\mathscr{S}(M))$ is affinely homeomorphic to the set of all Borel probability measures on the maximal spectrum X of M;

(iv) if M is not a Boolean algebra, then $\Phi(\mathscr{S}(M)) \subsetneq \mathscr{S}(B(M))$.

Proof. (i) This follows directly from the definition (8) and Proposition 3.

(ii) First, we will show that the mapping Φ is *affine,* that is, for every $s, s' \in \mathscr{S}(M)$ and $\alpha \in [0, 1]$, we have

$$\Phi(\alpha s + (1 - \alpha)s') = \alpha\Phi(s) + (1 - \alpha)\Phi(s').$$

For every $a \in B(M)$, the definition (8) yields:

$$\begin{aligned} \Phi(\alpha s + (1 - \alpha)s')(a) &= (\alpha s + (1 - \alpha)s')(\varphi_M(a)) \\ &= \alpha s(\varphi_M(a)) + (1 - \alpha)s'(\varphi_M(a)) \\ &= \alpha\Phi(s)(a) + (1 - \alpha)\Phi(s')(a). \end{aligned}$$

The mapping Φ is injective. Given $s, s' \in \mathscr{S}(M)$ with $s \neq s'$, find $a \in M$ such that $s(a) \neq s'(a)$. Hence it follows from (i) that $\Phi(s) \neq \Phi(s')$. The mapping Φ is continuous: take a net s_γ of elements of $\mathscr{S}(M)$ such that $s_\gamma \to s$ for some $s \in \mathscr{S}(M)$. This means that $s_\gamma(b) \to s(b)$ for every $b \in M$. Thus for every $a \in B(M)$, we get

$$\Phi(s_\gamma)(a) = s_\gamma(\varphi_M(a)) \to s(\varphi_M(a)) = \Phi(s)(a).$$

The set $\Phi(\mathscr{S}(M))$ is compact convex as an affine continuous image of $\mathscr{S}(M)$. Since Φ is a continuous bijection $\mathscr{S}(M) \to \Phi(\mathscr{S}(M))$, it is actually a homeomorphism by compactness.

(iii) The set $\mathscr{B}(X)$ of all Borel (σ-additive) probability measures on the maximal spectrum X of M is compact and convex [12, Proposition 5.22]. In the topology of $\mathscr{B}(X)$ a net μ_γ converges to μ in $\mathscr{B}(X)$ if for every continuous function $f : X \to \mathbb{R}$ we have $\int f \, \mathrm{d}\mu_\gamma \to \int f \, \mathrm{d}\mu$. By Corollary 29 in [3] or [4, Proposition 1.1], there exists a unique mapping $\Psi : \mathscr{S}(M) \to \mathscr{B}(X)$ such that

$$s(a) = \int_X \hat{a} \, \mathrm{d}\Psi(s) \tag{9}$$

for every $s \in \mathscr{S}(M)$ and every $a \in M$, where $\hat{a}$ is the image of a via the isomorphism identifying the elements of M with the separating MV-algebra of continuous functions $X \to [0, 1]$. We claim that the mapping Ψ is an affine homeomorphism. It is onto since every integral (9) with respect to some Borel probability measure from $\mathscr{B}(X)$ determines a state. It is also injective as every two states $s, s' \in \mathscr{S}(M)$ coincide whenever $\Psi(s) = \Psi(s')$ by the representation (9). The mapping Ψ is affine since the functional $\mu \in \mathscr{B}(X) \mapsto \int_X f \, \mathrm{d}\mu$ is linear for every continuous function $f : X \to [0, 1]$. Finally, we will show that Ψ^{-1} is continuous. Take a net μ_γ and μ with $\mu_\gamma \to \mu$ in $\mathscr{B}(X)$. It results directly from the definition of convergence in $\mathscr{B}(X)$ and in $\mathscr{S}(M)$ that $\Psi^{-1}(\mu_\gamma) \to \Psi^{-1}(\mu)$ in $\mathscr{S}(M)$, so Ψ^{-1} is continuous. Since $\mathscr{S}(M)$ is compact, the mapping Ψ is a homeomorphism. As an inverse of an affine homeomorphism is an

ISBN: 978-989-95079-6-8

affine homeomorphism and a composition of affine homeomorphisms is again an affine homeomorphism, take the mapping $\Psi \circ \Phi^{-1} : \Phi(\mathscr{S}(M)) \to \mathscr{B}(X)$ to finish the proof.
(iv) The MV-algebra M is not a Boolean algebra, so there exists an element $f \in M$ and $x \in X$ with $0 < f(x) < 1$. Put $\beta = \frac{\min(f(x),1-f(x))}{2}$ and define a mapping $p : B(M) \to [0,1]$ by

$$p(g) = \begin{cases} 1, & \beta \in g(x), \\ 0, & \text{otherwise}, \end{cases} \qquad g \in B(M).$$

A routine check shows that p is a probability on $B(M)$. We will show that there is no state $s \in \mathscr{S}(M)$ such that $\Phi(s) = p$. By way of contradiction, assume that such a state s exists. Then Proposition 3 together with (8) imply that $s(g) = p(g)$ for every $g \in M$. Since $s(f) = p(f) = 1$ and $s(\neg f) = p(\neg f) = 1$, we get

$$s(f \oplus \neg f) = s(1) = 1 \neq 2 = s(f) + s(\neg f),$$

which is a contradiction and thus $p \in \mathscr{S}(B(M)) \setminus \Phi(\mathscr{S}(M))$. □

On the one hand, every state on a semisimple MV-algebra can be viewed as the integral (9) with respect to a uniquely determined Borel probability measure that is defined on the maximal spectrum X of M. On the other hand, the operator Φ maps a state to a finitely additive probability on a Boolean subalgebra of the direct sum $\mathcal{B}^X$ (see Example 2). While the transformation Φ enables only to embed $\mathscr{S}(M)$ into a (huge) state space $\mathscr{S}(B(M))$, we will show below that and how probabilities on direct sums of Boolean algebras appear naturally already in classical Kolmogorov model of probability.

Let $B_1, \dots, B_k$ be Boolean algebras. Let p_i be a probability on B_i for $i = 1, \dots, k$. By B we denote a Boolean algebra that is the direct sum of $B_1, \dots, B_k$ (see [13, §16]). Every element of $b \in B$ can be identified with an n-tuple $(b_1, \dots, b_k)$ such that $b_i \in B_i$ for each $i = 1, \dots, k$. The elements b of B are random events whose meaning is "precisely one of all the events $b_1, \dots, b_k$ occurs". These classes of random events capture *two-stage random experiments*, such as the random selection of a ball from a randomly selected box. In this interpretation, each number $i = 1, \dots, k$ denotes a box and the corresponding Boolean algebra B_i models all outcomes of a random selection of a ball from the box i. Precisely, for some nonnegative $\alpha_1, \dots, \alpha_k$ with $\sum_{i=1}^k \alpha_i = 1$, the procedure can be described as follows:

(i) select B_i with a probability α_i;
(ii) if B_j was selected in the previous stage, then perform the random experiment with outcomes in B_j and probabilities described by p_j.

A probability p on B should then express the above introduced meaning of the two-stage experiment. This means that for every $b \in B$, we get

$$p(b) = \sum_{i=1}^{k} \alpha_i p(b_i). \tag{10}$$

The interrelationship between states on MV-algebras and probabilities on direct sums as defined above is made possible by the mappings Φ and φ_M.

Example 3. Let M be the direct product of two standard MV-algebras, that is, $M = [0,1]^2$. According to (9) every state s on M is of the form

$$s(a) = s((a_1, a_2)) = \alpha a_1 + (1-\alpha)a_2, \quad a = (a_1, a_2) \in M, \tag{11}$$

for some $\alpha \in [0,1]$. The Boolean algebra $B(M)$ is precisely the direct sum of the two Boolean algebras $\mathcal{B}$, where $\mathcal{B}$ is as in Example 2. The MV-algebra M embeds into $B(M)$ by sending each $a = (a_1, a_2) \in M$ to the function $a^*(i) = [0, a_i)$, $i = 1, 2$. Let λ denotes the restriction of Lebesgue measure to $\mathcal{B}$. Then it follows from (11) that

$$\Phi(s)(g) = \alpha\lambda(g(1)) + (1-\alpha)\lambda(g(2)), \quad g \in B(M). \tag{12}$$

Indeed, as in Example 2, for every $g \in B(M)$ and $i = 1, 2$ find the representation $a_1^i, b_1^i, \dots, a_n^i, b_n^i \in [0,1]$ such that $g(i) = [a_1^i, b_1^i) \cup \dots \cup [a_n^i, b_n^i)$ with all the intervals disjoint. Then

$$\varphi_M(g)(i) = (b_1^i \ominus a_1^i) \oplus \dots \oplus (b_n^i \ominus a_n^i)$$

and $(\varphi_M(g)(1), \varphi_M(g)(2)) \in M$. As a consequence,

$$\begin{aligned} \Phi(s)(g) &= s(\varphi_M(g)) = \alpha\varphi_M(g)(1) + (1-\alpha)\varphi_M(g)(2) \\ &= \alpha(b_1^1 \ominus a_1^1) \oplus \dots \oplus (b_n^1 \ominus a_n^1) \\ &\quad + (1-\alpha)(b_1^2 \ominus a_1^2) \oplus \dots \oplus (b_n^2 \ominus a_n^2) \\ &= \alpha\lambda(g(1)) + (1-\alpha)\lambda(g(2)), \end{aligned}$$

which proves (12).

The model from Example 3 can be given a straightforward probabilistic interpretation. Suppose that some underground platform is accessible from a street either via a stairway or via an escalator. Mr. Smiley chooses the stairway with a probability $\alpha \in [0,1]$ and the escalator with a probability $1-\alpha$. Train arrivals are uniformly distributed over the interval $[0,1]$. What is a probability that Mr. Smiley's waiting time for the train will be smaller or equal to $a_1 \in [0,1]$ (in case he uses the stairway) or smaller or equal to $a_2 \in [0,1]$ (in case he uses the escalator)? The investigated event $a = ([0, a_1), [0, a_2))$ belongs to $B(M)$, while the many-valued event $a = (a_1, a_2)$ is an element of M. If $p(a)$ denotes the probability of a, then (10) together with uniformity of the waiting times yield

$$\begin{aligned} p(a) &= \alpha\lambda([0, a_1]) + (1-\alpha)\lambda([0, a_2]) = \alpha a_1 + (1-\alpha)a_2 \\ &= s(a). \end{aligned}$$

4 Conditional Probability

In this section we use the probabilistic operator Φ to study conditioning in the framework of MV-algebras and states. Conditional probability on MV-algebras was studied from a variety of perspectives. In [7] the conditional probability on MV-algebras is defined in a way that mimicks the classical (Boolean) approach of conditioning "an event by an event" or "an event by a subalgebra". This approach makes use of an additional operation (product) on an MV-algebra and it was employed by Montagna in order to prove de Finetti-style coherence theorem for conditional bets in many-valued logic [14]. A completely different definition of conditional probability (so-called "conditional") was proposed by Mundici in [15].

ISBN: 978-989-95079-6-8

The following approach to conditioning is taken from [7, Definition 5]. Let s be a state on a PMV-algebra M. Given a many-valued event $b \in M$ with $s(b) > 0$, put

$$s(a|b) = \frac{s(a \cdot b)}{s(b)}, \quad a \in M. \tag{13}$$

The number $s(a|b)$ is called a *conditional state of a given b*. If $s(b)$ vanishes, then we leave $s(a|b)$ undetermined. This makes the function $a \in M \mapsto s(a|b)$ defined for "almost all" $b \in M$. Since $(a_1 \oplus a_2) \cdot b = (a_1 \cdot b) \oplus (a_2 \cdot b)$ for every $a_1, a_2 \in M$ such that $a_1 \odot a_2 = 0$, it can be easily verified that the mapping $s(.|b) : M \to [0,1]$ is a state (so-called *conditional state*) on M whenever $b \in M$ is such that $s(b)$ is nonzero. If p is a probability on a Boolean algebra B, then a product operation $\cdot$ on B becomes $\wedge$ and the formula (13) for $p(b) > 0$ reduces to

$$p(a|b) = \frac{p(a \wedge b)}{p(b)}, \quad a \in B. \tag{14}$$

Analogously, the function $p(.|b) : B \to [0,1]$ is a probability on B whenever $p(b) > 0$. In this case we call $p(.|b)$ a *conditional probability*. The major difference between (13) and (14) is the consequence of non-idempotence of the product operation on M. For every $b \in M$ with $s(b) \neq 0$, we have in general only

$$s(b|b) \leq 1 \quad \text{and} \quad s(\neg b|b) \geq 0. \tag{15}$$

In contrast to this, in classical probability $p(b|b) = 1$ and $p(\neg b|b) = 0$ holds true for any $b \in B$ with $p(b) \neq 0$. Surprisingly, this concept of conditioning found a natural application in Montagna's framework [14]: conditional bets are updated in proportion to the truth value of the event in the condition. In that follows we make an effort to justify the definition (13) with the probabilistic interpretation of states via Φ and $B(M)$.

Proposition 5. *Let M be an MV-algebra with product and $s(.|b)$ be a conditional state on it for $b \in M$ with $s(b) \neq 0$. Then there exists a probability p on $B(M)$ such that*

$$\Phi(s(.|b))(a) = p(a \cdot b|b), \quad a \in M. \tag{16}$$

Proof. Put $p = \Phi(s)$. For every $a \in M$, we obtain

$$\begin{aligned} \Phi(s(.|b))(a) = s(a|b) &= \frac{s(a \cdot b)}{s(b)} = \frac{\Phi(s)(a \cdot b)}{\Phi(s)(b)} \\ &= \frac{p(a \cdot b)}{p(b)} = \frac{p((a \cdot b) \wedge b)}{p(b)} \\ &= p(a \cdot b|b), \end{aligned}$$

since $(a \cdot b) \wedge b = a \cdot b$. □

It is worth emphasizing that the probability $\Phi(s(.|b))$ is not a conditional probability on M since $\Phi(s(.|b))(b) = p(b \cdot b|b)$ fails to be equal to 1 in general. So the operator Φ does not map conditional states to conditional probabilities. The formula (16) nevertheless connects the values of conditional states of a given b on M to the values of the classical conditional probability of $a \cdot b$ given b on $B(M)$. Conditioning a by a many-valued event b means in terms of classical conditional probability this: given b, what is a probability $p(a \cdot b|b)$ that an observation of the a-proportion of the many-valued event b will be made? Thus the conditional state defined by the formula (13) leads in accordance with (16) to the conditional probability of $a \cdot b$ given b on the Boolean algebra $B(M)$, which focuses only on the event which is included in b. The use of product in place of infimum makes the difference: the equality $p(a \wedge b|b) = p(a|b)$ holds true in contrast to $p(a \cdot b|b) \neq p(a|b)$. Taking into account the proposed interpretation, the inequalities in (15) look no longer unnatural. The "probabilistic" meaning of $s(b|b)$ is the value of $p(b \cdot b|b)$ that corresponds to the occurrence of the event $b \cdot b$ (given b) rather than b alone as the notation $s(b|b)$ suggests.

Example 4. Consider the situation from Example 3. Assume now that we know the actual waiting time does not exceed $b = (b_1, b_2) \in M = [0,1]^2$. Then the value $s(a|b)$ is according to Proposition 5 the same as $p(a \cdot b|b)$. This probability is precisely the conditional probability that the waiting time will be at most the a-proportion of b given our knowledge that it is at most b.

Acknowledgment

The first author acknowledges the grant No.1M0572 of the Ministry of Education, Youth and Sports of the Czech Republic that supported her visit at the Institute of Information Theory and Automation of the ASCR in Prague. The work of Tomáš Kroupa was supported by the grant No.1M0572 of the Ministry of Education, Youth and Sports of the Czech Republic. The second author owes his gratitude to Mirko Navara for the explanation of a role of Boolean direct sums in classical probability theory.

References

[1] D. Mundici. Averaging the truth-value in Łukasiewicz logic. *Studia Logica*, 55(1):113–127, 1995.

[2] B. Riečan and D. Mundici. Probability on MV-algebras. In *Handbook of Measure Theory, Vol. I, II*, pages 869–909. North-Holland, Amsterdam, 2002.

[3] T. Kroupa. Every state on semisimple MV-algebra is integral. *Fuzzy Sets and Systems*, 157(20):2771–2782, 2006.

[4] G. Panti. Invariant measures in free MV-algebras. To appear in *Communications in Algebra,* arXiv:math/0508445v2, 2008.

[5] G. Jenča. Boolean algebras R-generated by MV-effect algebras. *Fuzzy Sets and Systems*, 145(2):279–285, 2004.

[6] G. Grätzer. *General lattice theory*. Birkhäuser Verlag, Basel, second edition, 1998.

[7] T. Kroupa. Conditional probability on MV-algebras. *Fuzzy Sets and Systems*, 149(2):369–381, 2005.

[8] C. C. Chang. Algebraic analysis of many valued logics. *Trans. Amer. Math. Soc.*, 88:467–490, 1958.

[9] R. L. O. Cignoli, I. M. L. D'Ottaviano, and D. Mundici. *Algebraic foundations of many-valued reasoning*, volume 7 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 2000.

[10] A. Di Nola and A. Dvurečenskij. Product MV-algebras. *Multiple-Valued Logics*, 6:193–215, 2001.

[11] F. Montagna. An algebraic approach to propositional fuzzy logic. *Journal of Logic, Language and Information*, 9:91–124, 2000.

[12] K. R. Goodearl. *Partially ordered abelian groups with interpolation*, volume 20 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1986.

ISBN: 978-989-95079-6-8

[13] R. Sikorski. *Boolean Algebras*. Springer-Verlag, Berlin, 1964.

[14] F. Montagna. A notion of coherence for books on conditional events in many-valued logic. Submitted to *Journal of Logic and Computation,* 2009.

[15] D. Mundici. Faithful and invariant conditional probability in Łukasiewicz logic. 2009.

ISBN: 978-989-95079-6-8

Contributions to Dynamic Multicriteria Decision Making Models

Tiago C. Pais[1] Rita A. Ribeiro[1]

1 UNINOVA, Campus FCT-Universidade Nova Lisboa,
Caparica 2829-516, Portugal
Email: {tpp, rar}@uninova.pt

Abstract— *In this work we start by presenting a possible architecture for MultiCriteria Dynamic Decision Making (MCDDM) problems. Then we focus on contributions for the rating process because it involves two important aspects of any dynamic spatial-temporal decision problem: how to deal with uncertainty in dynamically changing input data and how consider different criteria importance, depending on criteria satisfaction and respective phase (or iteration) of the decision process. To explain and clarify the contributions we use an illustrative example of site selection for spacecraft landing on planets.*

Keywords— multicriteria dynamic decision making, imprecision on input data, dynamic criteria weighting.

1 Introduction

The main objective of this work is to present some contributions for building general multicriteria dynamic decision making (MCDMM) models. We will focus on two important aspects of any dynamic spatial-temporal decision problem: a) how to deal with uncertainty in dynamically changing input data; b) and how to consider different criteria importance, depending on criteria satisfaction and on the decision process phase (iteration). These two aspects belong to the classical process of evaluating alternatives of MCDM [1] [2] [3].

In general, the aim of Multicriteria or Multi-attribute decision making problems is to find the best compromise solution from all feasible alternatives assessed by pre-defined criteria (or attributes). This type of problems is widespread in real life situations [2], [4]. There are many methods and approaches proposed in the literature to deal with this static decision process, from utility methods to scoring and ranking ones [2] [3]. However, when facing multicriteria dynamic decisions making problems (MCDDM), where feedback from step to step is required, there are few contributions in the literature (see for example [5] [6]) .

To better explain our proposed contributions to deal with uncertainty in dynamically changing input data and criteria weights changes, during the temporal decision process, we use an simplified illustrative case study of a site selection problem for spacecraft landing on planets [7] [8].

2 MultiCriteria dynamic decision making (MCDDM)

Usually, a multicriteria decision model [1] includes 2 main tasks, rating the alternatives by aggregating their classifications for each criterion and then ranking them. In this work we discuss a dynamic multicriteria decision model because we are addressing decision paradigms with several iterations, which need to take into account past/historic information, to reach a decision (or consensus).

According to Richardson and Pugh [9] system dynamics problems have two main features: "they are dynamic: they involve quantities which change over time"; "involves the notion of feedback." This is exactly the perspective in this work: dynamically changing data and feedback from historical data.

A general multicriteria dynamic decision making (MCDDM) architecture is depicted in Fig. 1 (adapted from [7] and using concepts from [9]).

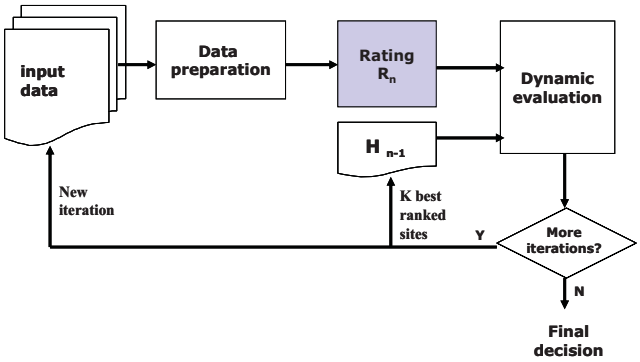


Figure 1: MCDDM architecture.

R_n represents the set of alternatives rating values from iteration n and H_{n-1} represents the historic set of rated alternatives from the previous iteration. The feedback is the set of ranked alternatives of iteration n (updated historic set of each iteration).

In summary, the main characteristics of processes/components, included in a MCDDM are:

a) Data preparation – this process deals with transformation of inputs for the decision model. It includes three main tasks: a.1) cleaning & filtering the data to extract the required criteria and the set of alternatives to be assessed (per iteration); a.2) normalization and representation of criteria for the decision model; a.3) Definition of weighting functions to express the relative importance for each criterion, within a phase or per iteration.

b) Rating process – This process refers to the classification of each alternative regarding all criteria, weighted with their relative importance. I.e. it refers to the aggregation of criteria values weighted with relative importance of criteria. Several methods exist in the literature [1] [2] [3], in our case

ISBN: 978-989-95079-6-8

we used an improved WSM method, as explained in section 3.2..

c) Dynamic evaluation process – This process refers to the aggregation of historical information about alternatives with their current rating to obtain the set of ranked alternatives [7] for the current iteration (n). When there are no more iterations the decision process stops and the best alternative is the one with the highest rate. For this evaluation we propose to use full-reinforcement operators as described in [7]

d) Historical process – This process determines a subset of good alternatives, from the ranked alternatives, to be considered as feedback information from iteration to iteration. Since this process takes places after ranking, it means historical information for a specific site "remembers" its past behaviour plus the actual rating. There are different sizes (K) for this set of historical information, depending on different phases of the decision process. For our case study we consider three phases according to expert advice.

The topic addressed in detail, in this work, will be the rating process (b), which is highlighted in grey in Figure 1.

3 Case Study: Site selection for spacecraft landing on planets

The case study used for discussing the rating process of multicriteria dynamic decision making (MCDDM) models is based on a real project [7], which objective was to determine the best target site for landing on planets for spacecrafts. The case study MCDDM model was developed for the final descent of the spacecraft (around 2 Km from surface), when hazard maps can be obtained. Some values used in the description are just indicative due to reasons of confidentiality.

3.1 Background

In this case study the criteria are hazard maps corresponding to images taken by the Spacecraft during the final descent (around 1000 m from surface) [8] [7]. We consider 7 input criteria, which correspond to the hazard maps obtained during the descent process: slope, texture, fuel, reachability, distance, shadow and scientific interest. The alternatives are the pixels of the combined search space (derived from merging 7 matrices of 512x512, corresponding to the input images/hazard maps), resulting from a data preparation process (not discussed in this work). The resulting set of alternatives is about 260.000 possible sites for each iteration (512*512= 262,144). The number of iterations is about 40 and for each one there is a evaluation process (called ranking process) including a combination of the *k* best alternatives from iteration *n-1* (historic set) with the current rated ones at the *n* iteration (for details about this process see [7]). We only consider the last iteration as historical information for two reasons: a) time constrains- we have less than 1 minute to perform 40 iterations hence an efficient process is required; b) the historic set implicitly includes all past ratings because we aggregate previous information with current one, i.e we update the alternative evaluation at each iteration and pass this information to the next iteration.

As mentioned in the abstract, in this work we focus on the rating process of MCDMM because it involves two important aspects of any dynamic spatial-temporal problem: how to deal with uncertainty in the input data of each iteration and how to consider different importance for each criteria, depending on the criteria satisfaction as well as on the phase of the decision process (iteration).

Before discussing the rating process we should explain that we assume 3 phases for this case study as follows. Phase 1 is executed until 1000 m of altitude and considers a historic of 250 set of sites, the rationale is that at higher attitudes there is a need for considering more alternatives (pixels) because the Spacecraft is still too far away from the surface and the information about the sites is still quite imprecise. Phase 2 is executed between the 1000 m and 300 m of altitude (parameters can be changed) and uses a historic set of 10 sites; since we are getting closer to the surface what we need is a small historic set (pixels) of "good" alternatives, ensuring sufficient distance between them (radius) to cover a reasonable search space. Phase 3 starts being executed at 300 m of altitude when the Spacecraft is quite close to the surface and the historic set considered was 5 sites.

Although it is out of scope to discuss the data preparation process for this case study we should mention that the results of this process are fuzzified variables (for references dealing with fuzzification see for example [4]), corresponding to each hazard map, adapted (i.e. re-calculated) at each iteration. Hence, a variable input includes a value and its corresponding membership value in the fuzzified variable.

3.2 Rating process

The rating process deals with aggregation of all classifications obtained by each alternative in each criterion. During the three phases considered, we compute these values for each alternative (per iteration). The general formulation for the weighted sum method (WSM), used as rating process in our dynamic decision process (MCDDM) is:

$$r_i = \oplus (W(ac_{i1}) \otimes ac_{i1}, ..., W(ac_{in}) \otimes ac_{in}) \quad (1)$$

Where,

$\oplus$ is the sum operator;

$\otimes$ is the product operator;

ac_{ij} is the accuracy & confidence membership value of j^{th} criterion for site *i*. (section 3.3)., and $W(ac_{ij}) = \dfrac{L(ac_{ij})}{\sum_{k=1}^{n} L(ac_{ik})}$, represents the normalized relative weight, considering a weighting function, $L(ac_{ij})$ as described in section 3.4.,

The basic principle of the aggregation method WSM (weighted sum method) is that "The overall score of an alternative is computed as the weighted sum of the attribute values" [8]. In our case study we tested other methods of aggregation besides this one (e.g. WPM, Compromise ratio [2]), however WSM combined with the proposed weighting functions, showed the best trade-of between computational efficiency versus correctness of results.

ISBN: 978-989-95079-6-8

3.3 Accuracy & confidence in input values

The accuracy and confidence parameters (a_{ij} and wc_j, respectively) will be taken into account in the decision model using the following expression:

$$ac_{ij} = wc_j * (1 - \max_{x \in [a;b]} \{ | \mu(x) - \mu(x_{ij}) | \}) * \mu(x_{ij}) \quad (2)$$

where:

wc_j is the confidence associated to criterion j;

x_{ij} is the value of jth criterion for site i;

$\mu(.)$ is the membership degree in a fuzzy set;

The interval [a,b] for x is defined by (a_{ij} is the accuracy associated to criterion j for site i):

$$a = \begin{cases} \min(D), & if\ x_{ij} - a_{ij} < \min(D) \\ x_{ij} - a_{ij}, & if\ x_{ij} - a_{ij} \geq \min(D) \end{cases}$$

$$b = \begin{cases} x_{ij} + a_{ij}, & if\ x_{ij} + a_{ij} < \max(D) \\ \max(D), & if\ x_{ij} - a_{ij} \geq \max(D) \end{cases} \quad (3)$$

D is the variable domain.

The accuracy is given as a percentage of criterion value and confidence belongs to the unit interval. For example, an accuracy of 90% for slope means that each slope value belongs to the interval $[a,b]$ where $a_{ij} = 0.9 \times x_{ij}$. On the other hand, a 0.9 confidence value means that we have a confidence on the $[a,b]$ interval of 0.9.

As can be observed, this formulation enables dealing with imprecision in the collected input hazard maps. Moreover, it allows dealing with two types of imprecision: lack of accuracy in the data and lack of confidence in the data. We believe this is a simple and easy way to tackle the problematic of how to handle imprecision in data in dynamic decision models. It is one of the contributions proposed in this work.

3.4 . Weighting functions

The second contribution is the proposal is to compute the weights using the following linear weighting function [10], [11]:

$$L(ac_{ij}) = \alpha \frac{1 + \beta ac_{ij}}{1 + \beta}, where\ 0 \leq \alpha \leq 1\ and\ 0 \leq \beta \leq 1 \quad (4)$$

Where: ac_{ij} is the accuracy and confidence membership value of j^{th} criterion for site *i*.

The logic of these weighting functions is that the satisfaction value of a criterion should influence its assigned relative importance. For example, if we are buying a car and the price is a "very important" criterion, if the car is quite expensive the final decision result should be less than the simple multiplication of weight*satisfaction value. We will not use quadratic weighting functions because it penalizes values near 1, contrarily to the linear function that presents a more smooth behaviour, which better fits our model [10], [11].

We consider the relative importance of criteria has different morphologies for each criterion, depending on each of the three phases defined. The definition of these weighting functions morphologies is given by parameters α and β. The α parameter provides the semantics for the weighting functions as follows:

- Very Important (VI =1);
- Important (I=0.8);
- Average importance (A=0.6);
- Low importance (L=0.4);
- Very Low importance (VL=0.2).

The β parameter provides the slope for the weighting functions, which will depend on the criterion at hand, with the logic that higher β means higher slope. For our case study this parameter has the following values:

- High (H =1);
- Medium (M=0.6667);
- Low (L=0.3333);
- Null (N=0);

As mentioned above the relative importance of criteria changes according to the altitude of the spacecraft, i.e. depends on the closeness to planet surface. We defined different weights for each sub-phase we are in, as described in the next two sub-sections, but these importances's can be modified by the decision maker anytime.

3.4.1 Discussion of Criteria Weighting in case study

Phase 1 (altitude: > 1000m, <2000; historic set size 250)

This phase refers to altitudes between 1000m to 2000m and a historic size of 250 candidate alternatives. In Table 1 we show the proposed values for α and β parameters for this phase, in Figure 2 we depict their respective plots, where Y-axis represents the weights and X-axis shows criteria satisfaction values within interval [0-1].

Table 1: α & β parameters (Phase 1).

	Fuel	Reach	Slope	Dist	Shad	Text	ScIn
α	1	1	1	0.6	0.4	0.8	0.2
β	0.25	0.667	0.111	0.333	0	0.333	0

For example, the rationale for "very important" is that although fuel, reachability and slope are all "very important" criteria for this phase, a lower satisfaction for criteria reachability should be quite penalized (e.g. X=0.25 corresponds to W=0.7, hence W(x)= 0.25*0.7=> 0.175). Further,, fuel and slope also consider high penalties for lower satisfaction values, visible on their plot (Figure 2).

ISBN: 978-989-95079-6-8

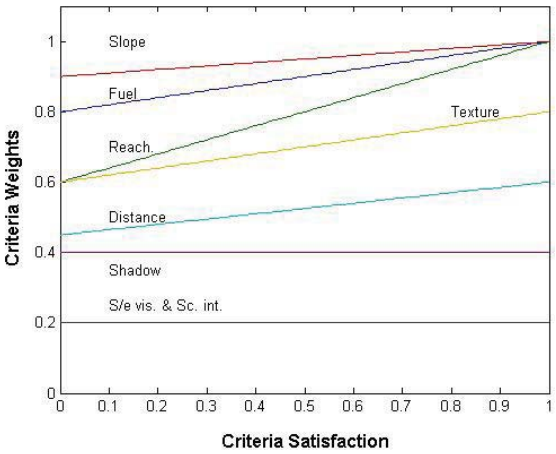


Figure 2: Weighting functions Phase I.

Phase I1 (altitude: ≤ 1000m, >300m; historic set size 10)
In Table 2 we depict the proposed values for α and β parameters for this phase, in Figure 3 their respective plots.

Table 2: α & β parameters (Phase II).

	Fuel	Reach	Slope	Dist	Shad	Text	ScIn
α	0.6	1	1	0.6	0.4	0.8	0.2
β	0.5	0.667	0.111	0.333	0	0.333	0

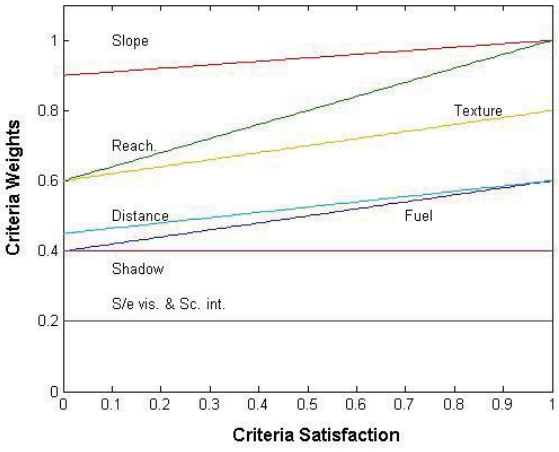


Figure 3: Weighting functions Phase II.

For example, in this phase the same rationale as in phase 1 applies for the two "very important" criteria, slope and reachability, althought slope is less penalized for lower satisfaction of criteria (e.g.0.8 weight for satisfaction 0.5 and weight 1 for satisfaction 1).

Phase III (altitude: ≤ 300m; historic set size 5)
In Table 3 we depict the proposed values for α and β parameters for this phase, in Figure 4 their respective plots.

Table 3: α & β parameters (Phase III).

	Fuel	Reach	Slope	Dist	Shad	Text	ScIn
α	0.2	1	1	0.2	0.4	0.8	0.2
β	0	0.667	0.111	0	0	0.333	0

For example, in this phase scientific interest and shadow do not play an important role because the alternatives are focused on a small area. Further, it almost does not require fuel due to the surface proximity. Hence their weights are small and constant, 0.2 for fuel & scientific interest, and 0.4 for shadow (all flat functions not dependent on satisfaction of criterion).

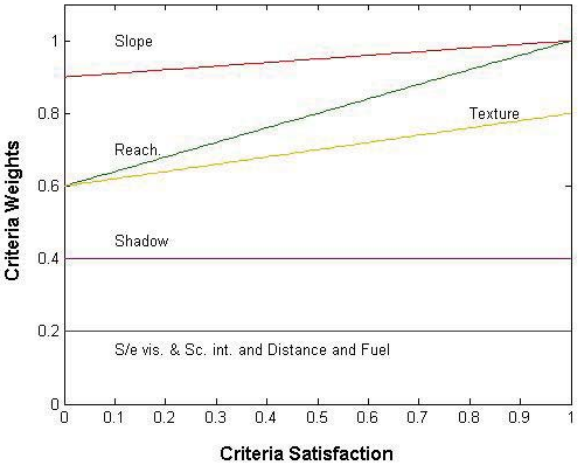


Figure 4: Weighting functions Phase III.

In summary, the relative importance proposed for each criterion followed the semantic rational:

- Slope relative importance (weight) in all phases is considered to be "very important";
- Reachability relative importance (weight) in all phases is considered to be "very important";
- Texture relative importance (weight) in all phases is considered to be "important";
- Shadow relative importance (weight) in all phases is considered to be "low";
- Distance relative importance (weight) in the first 2 phases is considered to be "average" and then it will change to "low" and "very low";
- Scientific interest (Sc int.) relative importance (weight) in all phases is considered to be "very low";
- Sun/Earth visibility (S/e vis.) relative importance (weight) in all phases is considered to be "very low";
- Fuel relative importance (weight) in phase 1 is considered to be "very important";
- Fuel relative importance (weight) in phase 2 is considered to be "medium";
- Fuel weight in phase 3 is considered to be "very low".

The rational above is expressed by the α parameter, while the β parameter provides more or less penalties for lower satisfaction values in the criteria, as depicted in the figures for each phase. A simplification of the weighting functions could be to define "a priori" a slope for each importance i.e. avoids having to define different β parameters for each weighting functions. This option simplifies the weight assignment process but it is less accurate.

ISBN: 978-989-95079-6-8

An important point to highlight is that the parameter values proposed for the phases (or iterations) can be tuned for any other dynamic decision application.
This section concludes the second proposed contribution of this work, i.e. contributions to deal with different criteria weights for steps of the dynamic decision process.

4 Discussion

In Table 4 are depicted the best three alternatives on the last iteration of each phase using default weights (described in previous section). For instance, the best alternative on the end of the 2nd phase (iteration 25) has 3D coordinates: (86 , -27 , 12) and rating value of 0.7807.

We can observe that from 16th to the 25th iteration the best site is completely different. However, the 2nd and 3rd best sites seem to be on the same region on both iterations (50 meters south). Furthermore, the best and 2nd best sites from iteration 25 changed positions with each other on the 40th iteration.

Table 4 – Rating values and 3D coordinates of the best sites on the last iteration of each phase (default weights).

Iter.	3D coord (meters)	R_n
16 (1st Phase)	(51 , 29 , 35)	0.7833
	(65 , -0.4 , 33)	0.7801
	(94 , 26 , 35)	0.7798
25 (2nd Phase)	(86 , -27 , 12)	0.7807
	(66 , -50 , 10)	0.7783
	(95 , -26 , 12)	0.7772
40 (3rd Phase)	(79 , -53 , -6)	0.9955
	(86 , -53 , -6)	0.5119
	-	-

In table 5 we present the results regarding the best sites on the last iteration of the 2nd and 3rd phase, only this time changing the Distance criterion weight to Very High on both phases. For illustrative purposes we considered the distance towards the centre of image for computing the Distance Criterion.
Comparing the results of Table 5 with Table 4, we can observe that at iteration 25th the first result is exactly the same and the 2nd and 3rd too, but on a different order. The best site maintained is position very close to the centre of the image and the distance criterion is now considered to be very high. By the same reason, site (95 , -26 , 12) went to the 2nd position because it is must closer to the image centre than site (66 , -50 , 10)..

On the other hand, the 2 best sites on the 40th iteration are completely different on both Tables. The sites on Table 4 are not close to the centre of the image and when the Distance criterion is set to very high the model tends to rate higher sites closer to the centre of the image. Hence, the best sites on Table 5 are closer to centre of the image (see Figure 5).
It should be noted that the experiences discussed here were adaptations of tests done within a research project [7] [8]. Furthermore, here we do not discuss the results rating values, because we are not addressing the aggregation methods used to evaluate the alternatives.

Table 5 – Rating and 3D coordinates of 3 best sites on last iterations of 2nd and 3rd phases (weight for Distance Criterion =Very high).

Iter.	3D coord (meters)	R_n
25 (2nd Phase)	(86 , -27 , 12)	0.7991
	(95 , -26 , 12)	0.7938
	(66 , -50 , 10)	0.7932
40 (3rd Phase)	(88 , -31 , -3)	0.9979
	(78 , -30 , -3)	0.9977
	-	-

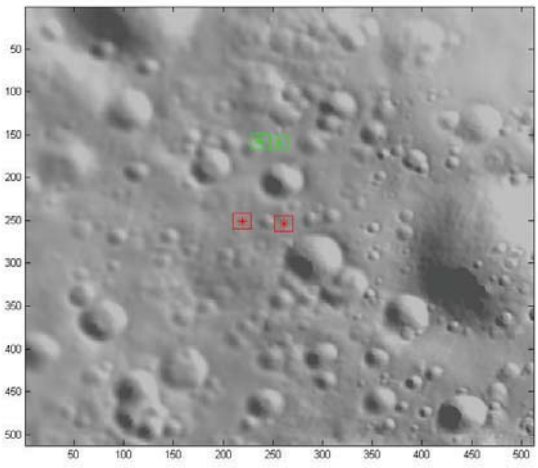


Figure 5 – Image of the surface on iteration 40. Green stars best sites from Table 4 and red stars from Table 5.

In Figure 5 we show the results, just discussed for the last iteration (40). It is quite notorious the influence of changing criterion weights for Distance, from average to very-high, on the last phase. Raising the importance of that criterion "forces" the final decision choice to be closer (shortest distance) to the initial target (center of image), even though the surface is more "bumpy" than the choices from Phase II.

5 Conclusions

In this work we presented a general architecture to cope with dynamic multi-attribute decision problems. We focused our attention on two spatial-temporal aspects of the rating process, .specifically proposing contributions for dealing with uncertainty in dynamically changing input data and also for dealing with dynamic changes in criteria importance.
Potential applications of this work could be on medical diagnose decision support and/ or fault, detection and isolation (FDI) problems.

Acknowledgment

The authors wish to thank Yannick Devassoux from Astrium EADS for his assistance with understanding the case study domain. This work was partially funded by EADS Astrium contract 4572019617.

ISBN: 978-989-95079-6-8

References

[1] Ribeiro, R.A., Fuzzy multiple attribute decision making: A review and new preference elicitation techniques. *Fuzzy Sets and Systems* 78(2): 155-181, 1996

[2] E. Triantaphyllou, *Multi criteria decision making method: a comparative study*, Kluwer academic publishers, Applied optimization series, Vol. 44, 2002.

[3] Chen, S.-J. and C.-L. Hwang. Fuzzy Multiple Attribute Decision Making, *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, 1992.

[4] Ross, T. (2004). *Fuzzy Logic with Engineering Applications*, John Wiley & Sons.

[5] Townsend, J. T. & Busmeyer, J. Dynamic representation of decision making. In: R. Port and T. Van Gelder (Eds.), *Mind in Motion: Dynamics, Behavior and Cognition*. 4th chapter, MIT Press, 1995.

[6] S. Khanmohammadi, R. A. Ribeiro, J. Jassbi. Multi criteria decision making using dynamics of criteria. Proceedings of the 11th Mediterranean Conference on Control and Automation (MED03), June 2003, med.ee.nd.edu/MED11/pdf/papers/t3-013.pdf

[7] T. C. Pais, R.A. Ribeiro, Y. Devouassoux, S. Reynaud Dynamic ranking algorithm for landing site selection. In: Proceedings of the Int. Conference on Information Processing and Management of Uncertainty (IPMU´08), Malaga, June 2008.

[8] Y. Devouassoux, S. Reynaud, G. Jonniaux, R. A. Ribeiro, T. C. Pais. Hazard avoidance developments for planetary exploration. In: Proceedings of the 7th International ESA Conference on Guidance, Navigation & Control Systems, Tralee, Ireland,2-5 June 2008

[9] Richardson, G. P. & Pugh III, A. L. *Introduction to System Dynamics Modeling with DYNAMO*. Portland, Oregon: Productivity Press, 1981.

[10] R. Marques-Pereira, Rita A. Ribeiro. Aggregation with generalized mixture operators using weighting functions. *Fuzzy Sets and Systems*, 137 (2003) 43-58.

[11] Rita A. Ribeiro, R. Marques-Pereira. Generalized mixture operators using weighting functions: a comparative study with WA and OWA. *European Journal of Operational Research*, 145 (2), March 1 (2003) 329-342.

ISBN: 978-989-95079-6-8

A Fuzzy Set Approach to Ecological Knowledge Discovery

Arkadiusz Salski

Institute of Computer Science, University of Kiel
Kiel, Germany
E-mail: asa@email.uni-kiel.de

Abstract— *Besides the problem of searching for effective methods for extracting knowledge from large databases (KDD) there are some additional problems with handling ecological data, namely heterogeneity and uncertainty of these data. A fuzzy set approach can be used to handle these problems at some stages of the knowledge discovery process. Ecological data can be defined as fuzzy sets without sharp boundaries, which reflect better the continuous nature of ecological parameters. The paper focuses on one of the important methods of data reduction, namely clustering, and on the data transformation and construction of a combining operator. Two support systems developed at the University of Kiel and their applications are presented, namely the Fuzzy Clustering System ECOFUCS and the Fuzzy Evaluation and Kriging System FUZZEKS.*

Keywords— ecological data, fuzzy clustering, fuzzy data transformation, uncertainty of ecological data.

1 Some properties of ecological data

Dealing with a very large data basis (long time series, spatial data with high resolution, etc.) is a typical problem in data mining and knowledge discovery, but there are some additional problems with handling the environmental data. The first one is the heterogeneity of ecological data sources (e.g. sources of quantitative and qualitative information). The next problem in ecological data mining is the large inherent uncertainty of these data that results not only from the presence of random variables but also from the difficult comparability of these data, approximate estimations, and imprecision and the subjectivity of the information obtained from an expert [12, 13]. This paper deals with processing both "subjective" data derived from experts and "objective" measurement data. Statistical or stochastic aspects of the uncertainty problem are not taken into account.

Some requirements for methods of searching for ecological knowledge arise from the properties of ecological data mentioned above. Special methods of data analysis should be used to handle the uncertainty and heterogeneity of these data. The fuzzy approach, as a possible way to handle uncertainty, is particularly useful for processing imprecise or uncertain data. Ecological data can be defined as fuzzy sets or fuzzy clusters without sharp boundaries. That reflects better the continuous nature of environmental parameters.

Knowledge discovery is a very complex process, which includes data cleaning, data integration, data reduction, data transformation, data mining, pattern evaluation and knowledge presentation [9]. Collaboration with an expert in the data domain can be very useful at some stages of this process. The fuzzy approach enables us to integrate the expert knowledge in the knowledge discovery process. The paper focuses on the applications of fuzzy sets and fuzzy operators at some stages of this process, namely data reduction (section 2) and constructing a combining attribute (section 3).

2 Data reduction: a fuzzy clustering approach

Clustering belongs to the most popular methods of numerosity reduction of data sets, i.e. "replacing" the data set by smaller representations such as clusters in order to reduce the size of the data set. The clustering methods are based on the principle of maximizing the intraclass similarity of objects and minimizing the interclass similarity of these objects, i.e. objects within a cluster have a high similarity, but are very dissimilar to objects in other clusters. Conventional clustering methods based on Boolean logic ignore the continuous nature of ecological parameters and the uncertainty of data. That can result in misclassification.

Fuzzy clustering methods provide additional information, namely the distribution of the membership values which can be interpreted as a similarity measure. The common fuzzy clustering methods, like the fuzzy c-means method, work only with crisp data, that means they provide the fuzzy partition only for crisp data (e.g. exact measurement data). In ecology we have often to deal with data with a semblance of accuracy. In such cases it may only be possible to obtain estimates of data scatter which can be treated in the context of fuzzy sets and used for defining fuzzy data in the form of fuzzy vectors in a high dimension, the so-called conical fuzzy vectors [5]. They are defined by the apex and the so-called panderance matrix which describes the accuracy of the data. This matrix contains spreads of data in each dimension on its diagonal. Yang [15] defined the distance between two conical fuzzy vectors, $\tilde{A}$ and $\tilde{B}$ in (1), as follows:

$$d^2(\tilde{A},\tilde{B}) = \left\|\bar{a}-\bar{b}\right\|^2 + tr\left((A-B)^T(A-B)\right) \qquad (1)$$

where:

A and B are the so-called panderance matrices of $\tilde{A}$ and $\tilde{B}$,
$\left\|\bar{a}-\bar{b}\right\|$ is the distance norm (metric) between the apexes $\bar{a}$ and $\bar{b}$, and the trace
$tr\left((A-B)^T(A-B)\right)$ is the diagonal sum of $\left((A-B)^T(A-B)\right)$.

Yang proved that $d^2(\tilde{A},\tilde{B})$ defined above is a complete metric, which is an assumption for the convergence of the

ISBN: 978-989-95079-6-8

fuzzy c-means clustering procedure by Bezdek [3]. That means, we can define the well known objective function of the fuzzy c-means procedure for conical fuzzy vectors by

$$F_{(c)} = \sum_{i=1}^{n} \sum_{j=1}^{c} \left(\mu_{ij}\right)^m d_c^2\left(\tilde{A}_i, \tilde{B}_j\right) \qquad (2)$$

where:

$\tilde{A}_i$ is the *i*th object and $\tilde{B}_i$ is the *j*th cluster, both defined as conical fuzzy vectors,

n is the number of objects, and

c is the number of clusters.

The clustering algorithm for conical vectors proposed by Yang has been extended for the diagonal norm using the so-called z-transformation of the Euclidian norm and implemented for the Fuzzy Clustering System EcoFucs v.5.1 developed at the University of Kiel [8]. The diagonal norm is a highly recommendable distance measure in the case of heterogeneous ecological data with different domain scales. In such cases we can transform data in a uniform manner before we start the fuzzy c-means procedure for conical vectors:

$$z - trans\left(\tilde{A}_i\right) = \frac{\tilde{A}_i - \tilde{V}}{\tilde{S}} \qquad (3)$$

where:

$\tilde{V} = \frac{1}{n}\sum_{j=1}^{n} \tilde{A}_j$ is the mean vector of all fuzzy conical vectors of the input data set, and

$\tilde{S}$ is the vector of spreads from the panderance matrix.

To obtain back the coordinates of the cluster centers in the real scale we have to apply the inverse transformation of the results of the fuzzy c-means procedure. EcoFucs works also with crisp data and offers four different distance norms as a measure of similarity between an object and a respective cluster (the Euclidean-, the Diagonal-, the Mahalonobis- and L1-norms) and a set of methods for calculating the start partition (WARD, conventional c-means, maximum-distance-algorithm, sharp or fuzzy random partitions). The choice of the distance norm depends on the data set. The partition efficiency indicators (entropy, partition coefficient, payoff and non-fuzziness index) available in EcoFucs can be very helpful in searching for the optimal partition and finding the objects which can serve as the representatives of each cluster (see the example below) [10].

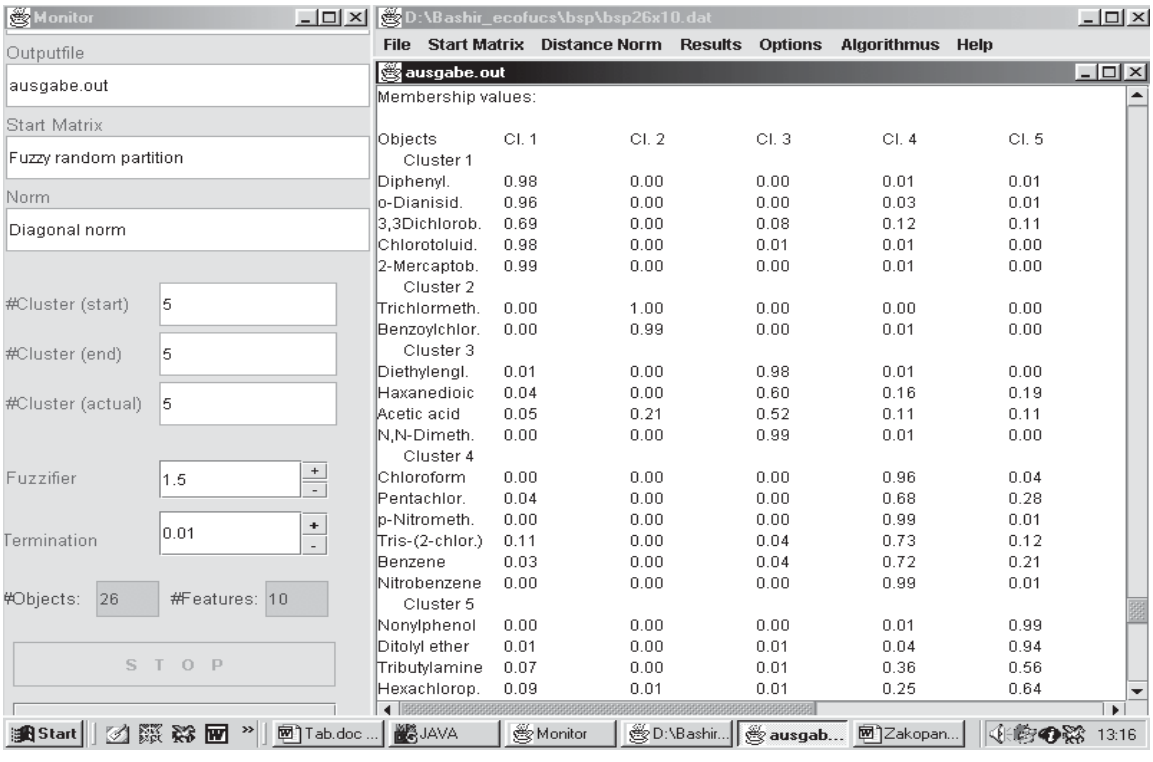


Figure 1: The main window of EcoFucs [8] with a small part of the results of the clustering of ecotoxicological data into 5 clusters.

The fuzzy clustering of chemicals according to their ecotoxicological properties [7] can be mentioned here as an example of data reduction. Both the large number of existing chemicals and the costs of ecotoxicological testing procedures make it necessary to select representative chemicals which faithfully reflect the relevant properties of possibly a major group of compounds. So the main tasks of this application were to find distinguishable clusters with characteristic properties and to find chemicals representative for each cluster. These representatives can be used for further data mining and other steps of the knowledge discovery process instead of all chemicals which belong to the cluster. Compared to conventional clustering methods a fuzzy clustering technique is more appropriate to handle the uncertainty of ecotoxicological data. The degree of uncertainty of these data is very high and can arise, for

ISBN: 978-989-95079-6-8

example, from a mixture of quantitative and qualitative data or from the difficult comparability of these data because of different measurement or test conditions (e.g. test results for different animals). The distribution of the membership values provides information from which the degree of similarity between the properties of a particular chemical and the properties characterising particular clusters can be deduced. This is particularly important since there are many chemicals with more or less overlapping properties. That would not be recognised by conventional clustering methods.

The analysis of the membership values helps us to find the representatives of each cluster. Their membership values are close (or equal) to 1. Figure 1 presents a small example of the results of the clustering of chemicals into 5 clusters for 10 ecotoxicological features (toxicity indicators and the potentials for biodegradability and hydrolysis). Trichlormethylbenzene and 2-Mercaptobenzothiazole, for example, can be taken as the representatives of clusters 2 and 1, respectively. We can also see some chemicals (e.g. Tributylamine and Hexachloropentadiene in cluster 5) with membership values strongly divided between different clusters. These values can be interpreted as the degree of similarity to the respective clusters.

3 Data transformation and the construction of a combining operator

One of the important stages of the knowledge discovery process is the data transformation needed to combine the data. In the case of heterogeneous data we have to normalise these data in a uniform manner before we construct the combining operator [4]. In order to do this we can scale data so that they fall within the same range. We can use the membership functions of fuzzy sets to transform the data into the interval 0.0 to 1.0. The definition of the membership functions of these fuzzy sets should express the evaluation criterion formulated by an expert.

We can consider an analysis of hydrogeological spatial data as an example. Figure 2 presents the membership function of the feature “large water table depth” used for the data normalization in the analysis of the suitability of a specified land unit as a waste disposal site [2, 14]. The shape of the defined membership function corresponds to the evaluation criterion. The values of the water table depth lower than 2 m are not suitable (the membership values equal 0); the values bigger than 5 m are very suitable for a waste disposal site (the membership values equal 1).

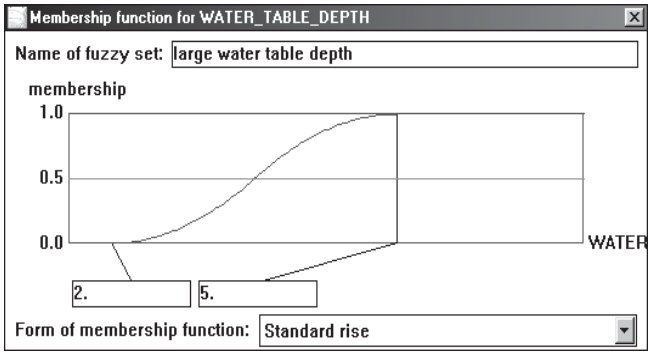


Figure 2: The membership function of the fuzzy set “large water table depth” used for the data transformation (the membership function window of FUZZEKS [2]).

Four land characteristics, namely water table depth, hydrologic conductivity, clay content and Cl concentration, are taken into account in this example. The values of these four features were transformed into a common scale 0.0 to 1.0 using suitable membership functions, like “large water table depth” or “high Cl concentration”. Now, we can combine these data by constructing a new combining attribute (Fig. 3), namely the joint degree of land suitability for a specified utilization (in this case, as a waste disposal site). Different logical and arithmetical operators can be used for the construction of such a combining attribute. Arithmetical operators (like the sum operator) can be weighted and that makes them particularly useful to express the degree of importance of a particular parameter. The distribution of weights is subjectively determined by a domain expert. In our example we combined the four land characteristics by means of a weighted sum operator and the “and”- operator using the Fuzzy Evaluation and Kriging System FUZZEKS developed at the University of Kiel (Fig. 3) [2].

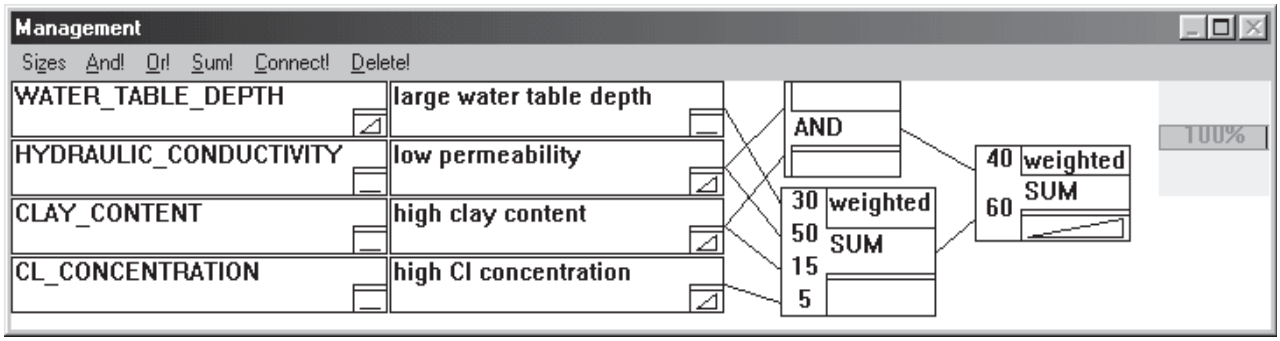


Figure 3: Constructing a combining operator (the management window of FUZZEKS [2]).

The calculated values of the joint attribute can be presented in the form of isolines using the fuzzy interpolation procedure, the so-called fuzzy kriging (Fig. 4). Fuzzy kriging is an extension of the conventional kriging procedure [1,6]. The application of the conventional methods of spatial interpolation is often restricted owing to insufficient amounts

ISBN: 978-989-95079-6-8

of data [11]. If the gathering of new data is too expensive or impossible, we can consider the use of additional imprecise data subjectively estimated by an expert. The fuzzy kriging procedure utilizes exact (crisp) measurement data as well as imprecise estimates obtained from an expert. These imprecise data can be defined as fuzzy numbers and taken as additional input data for the kriging procedure implemented in FUZZEKS. To simplify the preparation of the input data file a special ASCII-file format was implemented, combining both exact (crisp) and fuzzy data (fuzzy numbers) in one unified form [2]. Fuzzeks supports a user in the preparation of the so-called experimental variogram and in the interactive fitting of the crisp theoretical variogram, which is a basis for the interpolation procedure, to the fuzzy experimental variogram (see the small window left in Fig. 4).

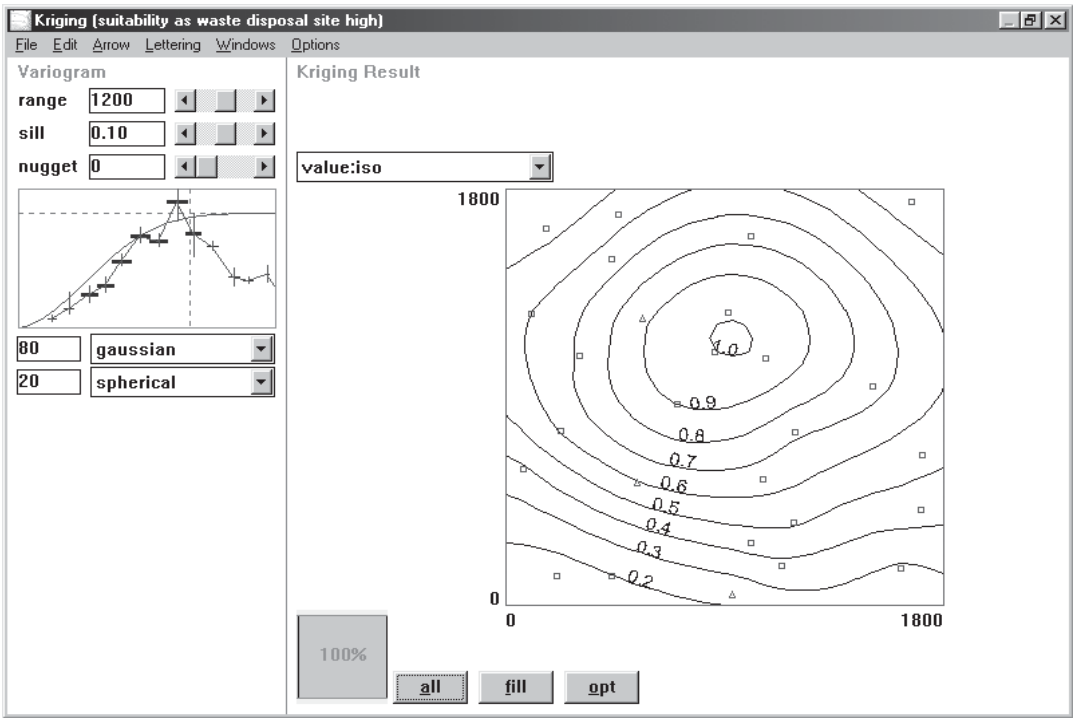


Figure 4: The presentation of the interpolated values of the joint attribute in the form of isolines (on the right in the main window of FUZZEKS,); interactive fitting of a theoretical variogram to an experimental variogram (small window left).

The logical structure of this fuzzy kriging procedure is shown in Fig. 5. The zigzag lines mark the stages with fuzzy data input in the form of fuzzy numbers. At two stages fuzziness is introduced into the calculation. First, fuzziness in the input values causes fuzziness in the experimental variogram. An expert takes the experimental variogram and its fuzziness into account when fitting the crisp theoretical variogram. Second, the fuzzy input values are used at the final step of kriging. Therefore, the kriging results are also fuzzy.

The main fuzzy kriging estimation is a linear combination of the input values and can be calculated using the extension principle and the α -cut-representation of fuzzy sets:

$$Z^*(x)_\alpha = \sum_{i=1}^{n} \delta_i(x) Z(x_i)_\alpha \qquad (4)$$

where:

$Z^*(x)_\alpha$ is is the α -cut of the interpolated value $Z^*(x)$ at the position *x*,

$Z(x_i)_\alpha$ are the α -cuts of the input values $Z(x_i)$, and

$\delta_i(x)$ are the crisp minimizing parameters.

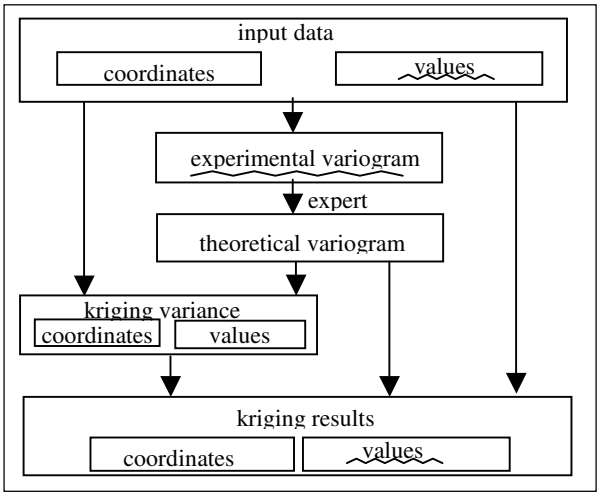


Fig. 5. Logical structure of fuzzy kriging (zigzag lines indicate the fuzziness of data).

ISBN: 978-989-95079-6-8

The kriging estimation (4) formulated by α-cut-representation of fuzzy sets has been used for the implementation of the fuzzy kriging procedure for the Fuzzy Kriging and Evaluation System FUZZEKS.

4 Final remarks

The fuzzy approach can support users in discovering interesting knowledge in uncertain ecological data at different stages of this process. The searching for representatives of groups of ecological objects by means of clustering methods can be very useful not only in numerosity reduction but also in dimensionality reduction of the data set. Fuzzy clusters of objects without sharp boundaries reflect better the continuous nature of ecological features. That enables a better interpretation of the data structure.

Heterogeneous and imprecise ecological data and vague expert knowledge can be integrated more effectively using a fuzzy approach. The data transformation by means of fuzzy sets can be used for this integration. The membership functions of fuzzy sets can describe the evaluation criteria (e.g. "high clay content") defined by an expert. The construction of a combining attribute in the next step reduces the dimensionality and the size of the data set.

And finally, the intention of the author was to draw the reader's attention to a large field for applications, namely ecology. The paper illustrates it briefly using some short application examples. The development of the easy to use tools (like EcoFucs and FUZZEKS developed at the University of Kiel) can be very helpful for the promotion of fuzzy methods in ecological applications.

References

[1] A. Bárdossy, I. Bogardi and W. E. Kelly, Geostatistics utilizing imprecise (fuzzy) information. *Fuzzy Sets and Systems*, 31/3: 311-327, 1989.

[2] F. Bartels, *Ein Fuzzy-Auswertungs- und Krigingsystem für raumbezogene Daten*. Diplomarbeit, Inst. für Informatik und Praktische Mathematik, Universität Kiel, 1997.

[3] J. C. Bezdek, A convergence theorem for the fuzzy c-means clustering algorithms. *IEEE Trans. PAMI*, PAMI-2(1): 1-8, 1980.

[4] Burrough P.A., Macmillan R.A. and Van Deursen, Fuzzy classification methods for determination land suitability from soil profile observations and topography. *Journal of Soil Science,* 43: 193-210, 1992.

[5] Celmiņš, A., Least Squares model fitting to fuzzy vector data. *Fuzzy Sets an Systems*, 22: 245-269, 1987.

[6] Diamond, P., Fuzzy kriging. *Fuzzy Sets and Systems*, 33/3: 315-332, 1989.

[7] Friderichs M., Fränzle O. and Salski A., Fuzzy clustering of existing chemicals according to their ecotoxicological properties. *Ecological Modelling*, 85/1: 27-40, 1996.

[8] Gömann, M., *Erweiterung des Fuzzy-Clustering-Systems EcoFucs*, Studienarbeit, Institut für Informatik, Christian-Albrechts-Universität zu Kiel, 2002.

[9] Han J. and Kamber M., *Data Mining. Concepts and Techniques*. Morgan Kaufman Publishers, San Francisco, 2001.

[10] Krenawi M., *Entwicklung eines plattformunabhängigen Systems zur Fuzzy-Clusteranalyse*. Diplomarbeit, Inst. für Informatik und Praktische Mathematik, Universität Kiel, 2001.

[11] Piotrowski, J.A., Bartels, F., Salski, A. and Schmidt, G., Geostatistical regionalization of glacial aquitard thicknessin northwestern Germany, based on fuzzy kriging., *Mathematical Geology*, 28/4: 437-452, 1996.

[12] Salski, A., Ecological modeling and data analysis. In: H.-J. Zimmermann: *Practical application of fuzzy technologies. The Handbooks of Fuzzy Sets*, Kluwer, , pp. 247-265, 1999.

[13] Salski, A., Ecological applications of fuzzy logic. In: F. Recknagel (ed): *Ecological Informatics,* Springer, pp. 3-14, 2002.

[14] Salski, A., Bartels, F., A fuzzy approach to land evaluation. *IASME Transactions,* 5/2: 774-780, 2005.

[15] Yang, M.-S. and Liu, H-H, Fuzzy clustering procedures for conical fuzzy vector data, *Fuzzy Sets and Systems*, 106: 189-200, 1999.

ISBN: 978-989-95079-6-8

Numerical Solution of Interval Differential Equations with generalized Hukuhara differentiability

Barnabas Bede, Luciano Stefanini
Department of Mathematics, University of Texas-Pan American, USA. email: bedeb@utpa.edu.
Department of Economics and Quantitative Methods, University of Urbino "Carlo Bo", Italy. email: lucste@uniurb.it

Abstract— *In the present paper we propose novel numerical methods for interval differential equations with generalized Hukuhara differentiability. The algorithms proposed here are based on characterization theorems of interval differential equations with ODEs. Usng the characterizations we can translate an interval problem into two ODE systems. We observe that an interval problem may have several solutions under the interpretation considered. Several examples are presented and discussed.*

Keywords— Generalized Hukuhara Differentiability, Interval Differential Equations, Interval Analysis.

1 Introduction

Interval Analysis was introduced as an attempt to handle interval (nonstatistical) uncertainty that appears in our mathematical or computer models of some deterministic real-world phenomena. The first monograph dealing with interval analysis is the book of R. Moore, [10]. Since then, Reliable Computing, Validated Numerics and Interval problems with Differential Equations are discussed in several monographs and research papers ([11], [2]).

Another major approach to a set of similar problems is that of differential inclusions and multivalued analysis ([1], [5]). This approach is also able to deal with discontinuous dynamical systems which do not fully fit into the Interval Analysis topic.

Hukuhara derivative of a set-valued mapping was first introduced by Hukuhara and it has been studied in several works. The paper of Hukuhara was the starting point for the topic of Set Differential Equations and later also for Fuzzy Differential Equations [8], [12], [6]. Recently, several works as e.g.[9], have brought back into the attention of the nonlinear analysis community, the topics of set differential equations and the Hukuhara derivative.

Recently a generalization of the Hukuhara difference for compact convex sets (gH-difference) was presented in [13] and the strongly and weakly generalized (Hukuhara) differentiability concepts have been proposed in [3]. Then, in [14] a combination (gH-derivative) was proposed and it was shown to be a useful concept in the interval setting. Our approach to interval differential equations is different from the approaches based on differential inclusions or interval analysis, however we can see that there is some connection among these approaches and the presented one, some connections with differential inclusions we discuss here.

After recalling some results of [14], we propose a numerical solution method for IDEs. Also, a numerical method for differential inclusions is discussed and several examples are proposed.

2 Generalized Hukuhara difference

Consider the space $\mathbb{I}$ of real intervals. Given two elements $A, B \in \mathbb{I}$ and $k \in \mathbb{R}$, the usual interval arithmetic operations, i.e. Minkowski addition and scalar multiplication, are defined by $A+B = \{a+b | a \in A, b \in B\}$ and $kA = \{ka | a \in A\}$. It is well known that addition is associative and commutative and with neutral element $\{0\}$. If $k = -1$, scalar multiplication gives the opposite $-A = (-1)A = \{-a | a \in A\}$ but, in general, $A + (-A) \neq \{0\}$, i.e. the opposite of A is not the inverse of A in Minkowski addition (unless $A = \{a\}$ is a singleton). A first implication of this fact is that, in general, additive simplification is not valid, i.e. $(A + C = B + C) \nRightarrow A = B$ or $(A + B) - B \neq A$ (the Minkowski difference is $A - B = A + (-1)B$).

To partially overcome this situation, the Hukuhara H-difference has been introduced as a set C for which $A \ominus B = C \iff A = B + C$ and an important property of $\ominus$ is that $A \ominus A = \{0\}\, \forall A \in \mathcal{K}_C^n$ and $(A+B) \ominus B = A, \forall A, B \in \mathcal{K}_C^n$.

The H-difference is unique, but it does not always exist (a necessary condition for $A \ominus B$ to exist is that A contains a translate $\{c\} + B$ of B).

A generalization of the Hukuhara difference proposed in [13] aims to overcome this situation.

Definition 1 *([13]) The generalized Hukuhara difference of two sets $A, B \in \mathbb{I}$ (gH-difference for short) is defined as follows*

$$A \ominus_g B = C \iff \begin{cases} (a) & A = B + C \\ \textit{or} \quad (b) & B = A + (-1)C \end{cases} \tag{1}$$

The following property ensures the existence of gH-differences.

Proposition 1 *([13]) The gH-difference $C = A \ominus_g B$ of two intervals $A = [a^-, a^+]$ and $B = [b^-, b^+]$ always exists and*

$$[a^-, a^+] \ominus_g [b^-, b^+] = [c^-, c^+] \tag{2}$$

with

$$c^- = \min\{a^- - b^-, a^+ - b^+\}$$
$$c^+ = \max\{a^- - b^-, a^+ - b^+\}.$$

Conditions (a) and (b) in (1) are satisfied simultaneously if and only if the two intervals have the same length and $c^- = c^+$.

3 Differentiation of interval valued functions

There exist several alternative definitions for the derivative of an interval-valued function. The first concept is a partic-

ISBN: 978-989-95079-6-8

ularization of the fuzzy concepts presented in [3] to the interval case, based on the generalized fuzzy differentiability concept, more general than one-sided derivatives, using the usual Hukuhara difference $\ominus$.

Definition 2 *([3]) Let $f :]a,b[\to \mathbb{I}$ and $x_0 \in]a,b[$. We say that f is strongly generalized (Hukuhara) differentiable at x_0, if there exists an element $f'(x_0) \in \mathbb{I}$, such that, for all $h > 0$ sufficiently small,*

(i) $\exists f(x_0+h) \ominus f(x_0)$, $f(x_0) \ominus f(x_0-h)$ and

$$\lim_{h \searrow 0} \frac{f(x_0+h) \ominus f(x_0)}{h} = \lim_{h \searrow 0} \frac{f(x_0) \ominus f(x_0-h)}{h} = f'(x_0),$$

or (ii) $\exists f(x_0) \ominus f(x_0+h)$, $f(x_0-h) \ominus f(x_0)$ and

$$\lim_{h \searrow 0} \frac{f(x_0) \ominus f(x_0+h)}{(-h)} = \lim_{h \searrow 0} \frac{f(x_0-h) \ominus f(x_0)}{(-h)} = f'(x_0),$$

or (iii) $\exists f(x_0+h) \ominus f(x_0)$, $f(x_0-h) \ominus f(x_0)$ and

$$\lim_{h \searrow 0} \frac{f(x_0+h) \ominus f(x_0)}{h} = \lim_{h \searrow 0} \frac{f(x_0-h) \ominus f(x_0)}{(-h)} = f'(x_0),$$

or (iv) $\exists f(x_0) \ominus f(x_0+h)$, $f(x_0) \ominus f(x_0-h)$ and

$$\lim_{h \searrow 0} \frac{f(x_0) \ominus f(x_0+h)}{(-h)} = \lim_{h \searrow 0} \frac{f(x_0) \ominus f(x_0-h)}{h} = f'(x_0).$$

Based on the gH-difference in [14] the following definition was proposed.

Definition 3 *Let $x_0 \in]a,b[$ and h be such that $x_0 + h \in]a,b[$, then the gH-derivative of a function $f :]a,b[\to \mathbb{I}$ at x_0 is defined as*

$$f'(x_0) = \lim_{h \to 0} \frac{1}{h}[f(x_0+h) \ominus_g f(x_0)]. \quad (3)$$

If $f'(x_0) \in \mathbb{I}$ satisfying (3) exists, we say that f is generalized Hukuhara differentiable (gH-differentiable for short) at x_0.

The next result is a characterization of the derivative concept proposed in (3) and it was proved in [14].

Theorem 2 *Let $f : [a,b] \longrightarrow \mathbb{I}$ be such that $f(x) = [f^-(x), f^+(x)]$. The function $f(x)$ is gH-differentiable if and only if $f^-(x)$ and $f^+(x)$ are differentiable real-valued functions and*

$$f'(x) = [\min\{(f^-)'(x), (f^+)'(x)\}, \quad (4)$$
$$\max\{(f^-)'(x), (f^+)'(x)\}]. \quad (5)$$

The following definition will be helpful in the discussion of the numerical methods for IDEs.

Definition 4 *Let $f : [a,b] \longrightarrow \mathbb{I}$ be gH-differentiable at $x_0 \in]a,b[$. We say that f is (i)-gH-differentiable at x_0 if*

$$\textit{(i.)} \qquad f'(x_0) = [(f^-)'(x_0), (f^+)'(x_0)] \quad (6)$$

and that f is (ii)-gH-differentiable at x_0 if

$$\textit{(ii.)} \qquad f'(x_0) = [(f^+)'(x_0), (f^-)'(x_0)]. \quad (7)$$

It is an interesting problem to see how the switch between the two cases (i.) and (ii.) in Definition 4 can occur.

Definition 5 *We say that a point $x_0 \in]a,b[$ is an l-critical point of f if it is a critical point for the length function $len(f(x)) = f^+(x) - f^-(x)$.*

Definition 6 *We say that a point $x_0 \in]a,b[$ is a switching point for the differentiability of f, if in any neighborhood V of x_0 there exist points $x_1 < x_0 < x_2$ such that*

type I) at x_1 (6) holds while (7) does not hold and at x_2 (7) holds and (6) does not hold, or

type II) at x_1 (7) holds while (6) does not hold and at x_2 (6) holds and (7) does not hold.

First, let us observe that any switching point is also an *l-critical* point. Indeed, if x_0 is a switching point then $[(f^-)'(x_0), (f^+)'(x_0)] = [(f^+)'(x_0), (f^-)'(x_0)]$ and so $(f^+)'(x_0) = (f^-)'(x_0)$ and $len(f(x_0))' = 0$. Clearly, not all *l-critical* points are also switching points.

The relation which exists between the two differentiability concepts was given in [14].

Theorem 3 *Let $f :]a,b[\to \mathbb{I}$ be a function $f(x) = [f^-(x), f^+(x)]$. The following affirmations are equivalent:*

(1) f is GH-differentiable

(2) f is gH-differentiable and the set of switching points is finite.

If the number of l-critical points is finite then the number of switching points is also finite. So, from now on we will use the GH-differentiability concept or equivalently the gH-differentiability with a finite number of switching points.

4 Interval differential equations

In this section we consider an interval valued differential equation

$$y' = f(x,y)\ , y(x_0) = y_0 \quad (8)$$
where
$$f : [a,b] \times \mathbb{I} \longrightarrow \mathbb{I}$$
with $f(x,y) = [f^-(x,y), f^+(x,y)]$ for $y \in \mathbb{I}$
$$y = [y^-, y^+],\ y_0 = [y_0^-, y_0^+].$$

We consider only GH-differentiable solutions, i.e. such that there exists $\delta > 0$ such that there are no switching points in $[x_0, x_0 + \delta]$.

An existence result for the solutions of an IDE was obtained in [14]. Let us consider $\overline{B}(y_0, q) \subset \mathbb{I}$ be a closed ball with center y_0 and radius q.

ISBN: 978-989-95079-6-8

Theorem 4 *Let $R_0 = [x_0, x_0 + p] \times \overline{B}(y_0, q)$, $y_0 \in \mathbb{I}$ nontrivial and $f : R_0 \to \mathbb{I}$ be continuous, nontrivial (i.e., y nontrivial interval gives $f(x, y)$ a nontrivial interval). If f satisfies the Lipschitz condition $D(f(x, y), f(x, z)) \leq L \cdot D(y, z)$, $\forall (x, y), (x, z) \in R_0$ then the interval problem*

$$\begin{cases} y' = f(x, y) \\ y(x_0) = y_0 \end{cases}$$

has two unique solutions $y^i, y^{ii} : [x_0, x_0 + r] \to \overline{B}(y_0, q)$ and the successive iterations in

$$y_0(x) = y_0$$

$$y_{n+1}(x) \ominus_g y_0 = \int_{x_0}^{x} f(t, y_n(t))dt,$$

or more precisely,

$$y^i_{n+1}(x) = y_0 + \int_{x_0}^{x} f(t, y^i_n(t))dt \qquad \textit{i.e.}$$

$$(a) \begin{cases} y^{i-}_{n+1}(x) = y_0^- + \int_{x_0}^{x} f^-(t, y^i_n(t))dt \\ y^{i+}_{n+1}(x) = y_0^+ + \int_{x_0}^{x} f^+(t, y^i_n(t))dt \end{cases}$$

and

$$y_0 = y^{ii}_{n+1}(x) - \int_{x_0}^{x} f(t, y^{ii}_n(t))dt \qquad \textit{i.e.}$$

$$(b) \begin{cases} y^{ii-}_{n+1}(x) = y_0^- + \int_{x_0}^{x} f^+(t, y^{ii}_n(t))dt \\ y^{ii+}_{n+1}(x) = y_0^+ + \int_{x_0}^{x} f^-(t, y^{ii}_n(t))dt \end{cases}$$

converge to these two solutions y^i and y^{ii} respectively.

Considering the interval problem

$$\begin{cases} y' = f(x, y) \\ y(x_0) = y_0 \end{cases} \tag{9}$$

we set

$$\begin{cases} \varphi^-(x, y^-, y^+) = f^-(x, y) \\ \varphi^+(x, y^-, y^+) = f^+(x, y) \end{cases} \tag{10}$$

with $\varphi^- \leq \varphi^+$ defined on a subset of $\mathbb{R}^3$.

Finally, we obtain two situations:

ODE(a): $y'^-(x) \leq y'^+(x)$; the differential equations are

$$\begin{cases} y'^-(x) = \varphi^-(x, y^-(x), y^+(x)) \\ y'^+(x) = \varphi^+(x, y^-(x), y^+(x)) \\ \text{s.t.} \\ y^-(x_0) = y_0^- \\ y^+(x_0) = y_0^+ \end{cases} \tag{11}$$

and if φ^- depends only on y^- and φ^+ depends only on y^+, the two equations are independent.

ODE(b): $y'^-(x) \geq y'^+(x)$; the differential equations are

$$\begin{cases} y'^-(x) = \varphi^+(x, y^-(x), y^+(x)) \\ y'^+(x) = \varphi^-(x, y^-(x), y^+(x)) \\ \text{s.t.} \\ y^-(x_0) = y_0^- \\ y^+(x_0) = y_0^+ \end{cases} \tag{12}$$

and if φ^+ depends only on y^- and φ^- depends only on y^+, the two equations are independent.

Now we present a characterization result obtained in [14]:

Theorem 5 *Let $R_0 = [x_0, x_0 + p] \times \overline{B}(y_0, q)$, $y_0 \in \mathbb{I}$ nontrivial and $f : R_0 \to \mathbb{I}$ be nontrivial and continuous. If f satisfies the Lipschitz condition $D(f(x, y), f(x, z)) \leq L \cdot D(y, z)$, $\forall (x, y), (x, z) \in R_0$ then the interval problem*

$$\begin{cases} y' = f(x, y) \\ y(x_0) = y_0 \end{cases} \tag{13}$$

is equivalent to the union of the ODEs (11) and (12) on some interval $[x_0, x_0 + q]$. Here, by equivalent, we understand that $y = [y^-, y^+] : [x_0, x_0 + q] \to \mathbb{I}$ is a solution of 13 if and only if $(y^-, y^+) : [x_0, x_0 + q] \to \mathbb{R}^2$ is a solution of one of the problems (11) or (12).

Remark 1 *We can write the IDE (13) in terms of the midpoint representations of y, y' and $f(x, y)$. Let $y = \langle \widehat{y}, \overline{y} \rangle$, $y_0 = \langle \widehat{y}_0, \overline{y}_0 \rangle$, $f(x, y) = \left\langle \widehat{f}(x, y), \overline{f}(x, y) \right\rangle$; then $y' = \langle \widehat{y}', |\overline{y}'| \rangle$ and (13) becomes*

$$\begin{cases} \widehat{y}' = \widehat{f}(x, y) \\ |\overline{y}'| = \overline{f}(x, y) \\ \widehat{y}(x_0) = \widehat{y}_0 \\ \overline{y}(x_0) = \overline{y}_0; \end{cases} \tag{14}$$

they represent a system of differential algebraic equations (DAEs) and can be investigated by the help of the corresponding theory.

All the preceding results hold for general interval-valued functions so possibly they depend on interval parameters, i.e. interval extensions of real-valued functions of the form $f(x, y, p_1, ..., p_n)$ with $p_i \in P_i \in \mathbb{I}$.

Let us analyze the case when $f(x, y)$ is the interval extension of a real valued function $f(x, p)$, i.e.

$$f(x, y) = [\inf_{p \in y} f(x, p), \sup_{p \in y} f(x, p)]. \tag{15}$$

Let us suppose further that f is monotonic with respect to p.

1) if $f(x, p)$ is increasing with respect to p,

$$case(a) : y'^-(x) \leq y'^+(x)$$
$$\begin{cases} y'^-(x) = f(x, y^-(x)) \\ y'^+(x) = f(x, y^+(x)), \\ \text{s.t. } y^-(x_0) = y_0^-, \\ y^+(x_0) = y_0^+ \end{cases} \tag{16}$$

$$case(b) : y'^-(x) \geq y'^+(x)$$
$$\begin{cases} y'^+(x) = f(x, y^-(x)) \\ y'^-(x) = f(x, y^+(x)), \\ \text{s.t. } y^-(x_0) = y_0^-, \\ y^+(x_0) = y_0^+ \end{cases} \tag{17}$$

2) if $f(x, p)$ is decreasing with respect to p

$$case(a) : y'^-(x) \leq y'^+(x)$$
$$\begin{cases} y'^-(x) = f(x, y^+(x)) \\ y'^+(x) = f(x, y^-(x)), \\ \text{s.t. } y^-(x_0) = y_0^-, \\ y^+(x_0) = y_0^+ \end{cases} \tag{18}$$

$$case(b) : y'^-(x) \geq y'^+(x)$$
$$\begin{cases} y'^+(x) = f(x, y^+(x)) \\ y'^-(x) = f(x, y^-(x)), \\ \text{s.t. } y^-(x_0) = y_0^-, \\ y^+(x_0) = y_0^+ \end{cases} \tag{19}$$

ISBN: 978-989-95079-6-8

Also, in this case, a very interesting result from [4] connects these cases to differential inclusions. In the followings we formulate a particularization of the result of Chalco-Cano and Roman-Flores. Namely we have:

Theorem 6 *[4] Let $f(x,p)$ be a monotonic function with respect to $p \in \mathbb{R}$ and let $y_0 \in \mathbb{I}$ nontrivial interval.*

a) If $f(x,y)$ is the interval extension of f and f is an increasing function, then the solution in case (a) of (16) and the attainable set of the differential inclusion $y' = f(x,y)$, $y(x_0) \in y_0$ coincide on some interval $[x_0, x_0+q]$.

b) If $f(x,y)$ is the interval extension of f and f is a decreasing function, then the solution in case (b) of (18) and the attainable set of the differential inclusion $y' = f(x,y)$, $y(x_0) \in y_0$ coincide on some interval $[x_0, x_0+q]$.

5 Solution methods for interval differential equations

From the results and discussion of previous sections, the interval differential equation concept presented in this paper does not coincide with the concept of a differential inclusion. It is a different, new approach to model interval uncertainty in dynamical systems. It is related (as it is seen from the above theorems) to differential inclusions but we do have in our case more than one solution. The existence of several solutions can be an advantage when we search for solutions with specific properties like e.g., periodic, almost periodic, asymptotically stable, etc. Also, it can be very useful when we have unknown correlations between the parameters. In those situations, the uncertainty about the correlation introduces supplementary uncertainty in our system, so the existence of several solutions appears to be natural in this case.

The above characterization theorems, together with the existence results, lead easily to a numerical algorithm to solve interval differential equations.

5.1 General description of the solution methods

First, let us remark that a switch between the cases (i) and (ii) of gH-differentiability is possible if and only if $y'(x)$ is a singleton as in fact possible switch-points are l-critical points, i.e., critical points of the length of $y(x)$. Let us remark that if at a point $y(x_0)$ is a singleton for some x_0, then this point is enforcing a switch to the case (i), because, according to the existence result in Theorem 4, the second solution does not exist in this case. All other l-critical points make possible a switch from (i) to (ii) case or vice versa, but they do not enforce a switch, so at each l-critical point x_0 with $y(x_0)$ nontrivial interval, two new local solutions arise. One is (i)-differentiable, the other is (ii)-differentiable.

These remarks, together with the characterization Theorem 5, lead to the following general approach to numerically solve interval differential equations. We solve basically ODEs (11) and (12) on subintervals of the time domain having the initial value y_0 updated at all the possible critical points. In this way on a bounded time interval we obtain a finite number of solutions. For the solution of (11) and (12) by the characterization theorem above, any efficient numerical method for ODEs can be used. This is an advantage of the method presented here, as we do not need to reinvent numerical methods for interval differential equations. Instead we can use the classical ones on the ODE translations of the interval differential equations.

This algorithm generates a tree structure for the solutions of the IDE. All the nodes of the tree will be critical points except the terminal nodes. All the branches represent local solutions of (i) or (ii) kind between two nodes. This is illustrated in Figure 1. Each node represents an l-critical point with a switching and each branch corresponds to one of the cases (i) or (ii) of differentiability.

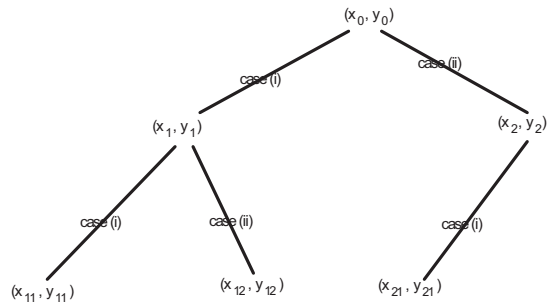


Figure 1. The structure of the tree of local solutions.

Our algorithm which generates all the solutions on a bounded interval is as follows.

Algorithm 1. (Find solutions of an interval differential equation - IDE)

Let us consider y_0 be any interval.

- Step 1. If y_0 is a singleton then we solve (11) and we obtain solution y_1.
- Step 2. Else we solve both (11) and (12) and we obtain solutions y_1, y_2.
- Step 3. Let $x_1 = \inf\{x > x_0 : y_1'(x) \in \mathbb{R}\}$ and $x_2 = \inf\{x > x_0 : y_2'(x) \in \mathbb{R}$ or $y_2(x) \in \mathbb{R}\}$ be the nearest l-critical point. (Let us remark here that since the algorithm applies to general interval-valued functions, it is possible that $y_2(x) \in \mathbb{R}$ however $y_2'(x) \notin \mathbb{R}$)
- Step 4. We insert the solutions which are found in a tree structure: y_1 in the left branch and y_2 in the right branch (the root will be simply (x_0, y_0)).
- Step 5. If $x_1 < X$ (a preset maximum value for x) then let $x_0 = x_1$ and $y_0 = y_1(x_1)$ and go to step 1.
- Step 6. If $x_2 < X$ then let $x_0 = x_2$ and $y_0 = y_2(x_2)$ and go to step 1.
- Step 7. Else Return.
- Step 8. Using a standard backtracking algorithm we generate all the solutions from the tree structure generated above.

The presented algorithm will generate a finite number of solutions on the interval $[x_0, X]$ provided that there are a finite number of critical points. Later we can extract those solutions which are closely reflecting the phenomenon which we have to model.

An approach to interval and fuzzy differential equations has been suggested in [7], based on differential inclusions.

ISBN: 978-989-95079-6-8

When $f(x,y)$ is the interval extension of a continuous function $f(x,p)$, $p \in \mathbb{R}$ and if we are interested in finding the *attainable set* for the differential inclusions $y' = f(x,y)$, $y(x_0) \in y_0$, $x \in [x_0, X]$, then we have a simpler algorithm based on the characterization Theorem 5 and the results shown in Theorem 6.

Let us consider y_0 be any interval. Let $f(x,y)$ be the interval extension of a real function $f(x,p)$. We use the same notation for both functions, and from the context we can identify them.

Algorithm 2. (Find the *attainable set* of a differential inclusion)

- Step 1. We find the points where the real function f changes its monotonicity w.r.t. p (if f is differentiable then we solve $\frac{\partial f}{\partial p} = 0$ and we find the critical points where monotonicity w.r.t. the second variable is changed). Let (x_1, y_1) be a critical point such that in $]x_0, x_1[$ there are no other critical points.
- Step 2. If f is increasing w.r.t. p on $[x_0, x_1] \times [y_0, y_1]$ we solve (11)
- Step 3. Else if f is decreasing w.r.t. p on $[x_0, x_1] \times [y_0, y_1]$ we solve (12)
- Step 4. If $x_1 < X$ then let $x_0 = x_1$ and $y_0 = y(x_1)$ and go to step 2.
- Else Stop.

This algorithm leads to the unique solution (*attainable set*) of the differential inclusion $y' = f(x,y)$, $y(0) \in y_0$.

It is easy to see that the proposed methods are very efficient from the numerical point of view, since for the local solutions we can use any standard algorithm.

5.2 Examples

The above algorithms were implemented in MATLAB. We have used MATLAB's standard ODE solver *ode45*, which is based on a Runge-Kutta (4,5) formula. Surely, any other solver could be used. Let us remark here that the critical points were in all the cases *a priori* determined. The critical points cannot be easily found on the run. The problem is that if we detect a critical point, due to the machine precision the algorithm finds further points which are close to the correct critical point as critical ones. Also, if we have a solution in the case (ii) with length decreasing asymptotically to zero, due to the machine precision the program detects them as critical points.

We will start with a simple example, which is easy to be solved analytically and we can compare the analytical solution to the numerical solution.

5.2.1 Example 1

Let us consider the interval differential equation

$$\begin{cases} y' = -y + [1,2]x \\ y(0) = [0,1] \end{cases} \quad x \in [0,4]. \tag{20}$$

We denote $y = [u, v]$, where u, v are real-valued functions. The systems (11) and (12) are respectively

$$\begin{cases} u' = -v + x \\ v' = -u + 2x \\ u(0) = 0 \\ v(0) = 1 \end{cases} \text{ and } \begin{cases} u' = -u + 2x \\ v' = -v + x \\ u(0) = 0 \\ v(0) = 1 \end{cases} \tag{21}$$

This equation (20) has exactly two solutions. One of them starts with the case (i) of differentiability

$$y_1(x) = [2x - e^x + 2e^{-x} - 1, x + e^x + 2e^{-x} - 2]$$

and there are no critical points in the trajectory.

If we start with case (ii) then we have a critical point of type II at $x = 1$ (i.e., $y(1)$ is a singleton). In this case we have to switch to case (i) of differentiability since the (ii)-differentiable solution does not exist. We obtain

$$u(x) = \begin{cases} 2x + 2e^{-x} - 2 & \text{if} \quad 0 \le x \le 1 \\ 2x - e^{x-1} + 2e^{-x} - 1 & \text{if} \quad 1 \le x \end{cases}$$

$$v(x) = \begin{cases} x + 2e^{-x} - 1 & \text{if} \quad 0 \le x \le 1 \\ x + e^{x-1} + 2e^{-x} - 2 & \text{if} \quad 1 \le x \end{cases}$$

The analytic solution and the numerical solutions obtained by the proposed algorithms are shown in Figs 2. and 3. respectively. We can see that the proposed method is very accurate. The error is controlled by the precision of *ode45* algorithm in MATLAB, and it is less than 10^{-6}.

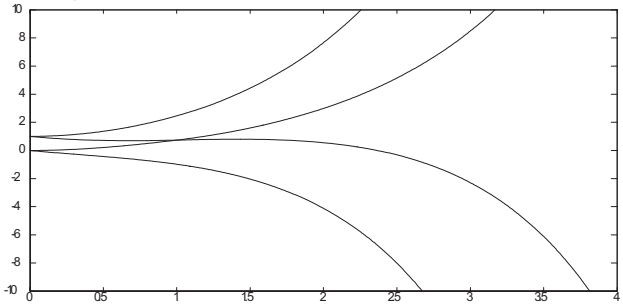


Figure 2. Analytic (exact) solutions of the IDE (20)

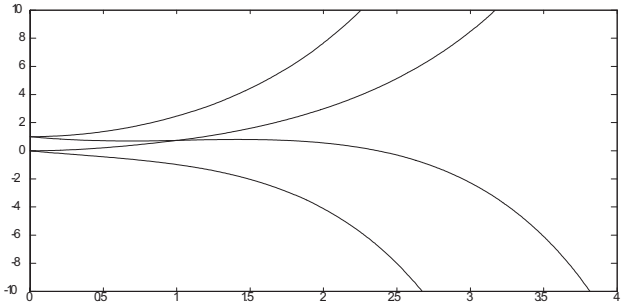


Figure 3. Numerical solutions of (20) by the proposed *Algorithm 1*.

5.2.2 Example 2

The structure of solution of an interval DE can be well illustrated on the problem

$$\begin{cases} y' = y \sin x + 2x, \\ y(0) = [1,3] \end{cases} \tag{22}$$

According to the characterization theorems it can be written as

$$\begin{cases} u'(x) = u(x)\sin(x) + 2x \\ v'(x) = v(x)\sin(x) + 2x \end{cases}$$

ISBN: 978-989-95079-6-8

or

$$\begin{cases} u'(x) = v(x)\sin(x) + 2x \\ v'(x) = u(x)\sin(x) + 2x \end{cases}.$$

The *l-critical points* are at $x = k\pi$, $k \in \mathbb{Z}$. If we solve the problem in the interval $[0, X]$ with $X < 2\pi$ we have only one *l-critical point* in the interval. Then we may have *four* solutions, as illustrated in Figure 4.

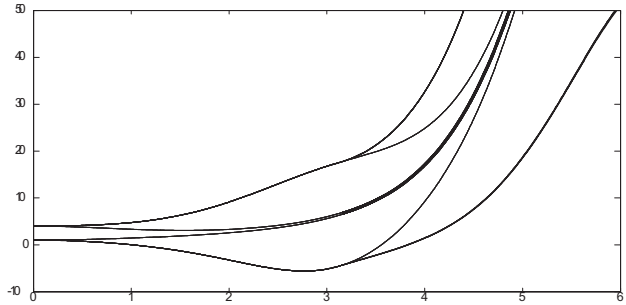


Figure 4. The four solutions of IDE (22) on the interval $[0, 6]$

If we consider time intervals containing more critical points, we may have more solutions. In Figure 5. we illustrate the solutions found by *Algorithm 1* to IDE (22) for $x \in [0, 11]$.

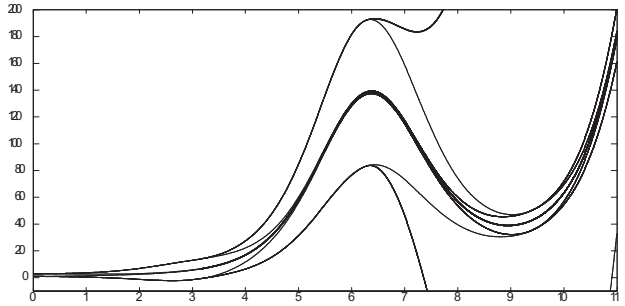


Figure 5. Existence of several solutions for IDE (22) on $[0, 11]$

5.2.3 Example 3

Regarding the second algorithm for solving differential inclusions when the function $f(x, y)$ is the interval extension of a crisp function we present the following problem

$$\begin{cases} y' = y\sin x + 2x, x \in [0, 6] \\ y(0) \in [1, 2]. \end{cases} \quad (23)$$

We observe that the function $y \sin x$ changes monotonicity at the points $k\pi$, $k \in \mathbb{Z}$. Also, we observe that in the interval $[0, \pi]$ it is increasing w.r.t. y. Using the *Algorithm 2* proposed for differential inclusions we obtain the solution presented in Figure 6.

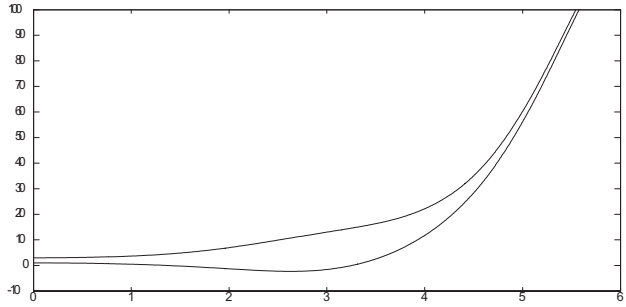


Figure 6. Solution of differential inclusion (23) using the proposed *Algorithm 2*

Remark: It is interesting to remark that the solution (attainable set) of the differential inclusion (23) is obtained by solving one of the ordinary differential equations (11) or (12) in steps 2. and 3. derived from the use of the generalized Hukuhara derivative.

References

[1] J.P. Aubin, A. Cellina, *Differential Inclusions*, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1984.

[2] R. Baker Kearfott, V. Kreinovich (Eds.) *Applications of interval computations*, Kluwer Academic Publishers, 1996.

[3] B. Bede, S.G. Gal, Generalizations of the differentiability of fuzzy-number-valued functions with applications to fuzzy differential equations, *Fuzzy Sets and Systems*, 151(2005), 581-599.

[4] Y. Chalco-Cano, H. Román-Flores, On new solutions of fuzzy differential equations, *Chaos, Solitons & Fractals*, 38(2008), 112-119.

[5] K. Deimling, *Multivalued Differential Equations*. W. De Gruyter, Berlin, 1992.

[6] P. Diamond, P. Kloeden, *Metric Spaces of Fuzzy Sets*, World Scientific, New Jersey, 1994.

[7] H. Hullermeier, An approach to modeling and simulation of uncertain dynamical systems, *Int. Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 5 (1997) 117-137.

[8] O. Kaleva, Fuzzy differential equations, *Fuzzy Sets and Systems*, 24 (1987) 301-317.

[9] V. Lakshmikantham, A.A. Tolstonogov, Existence and interrelation between set and fuzzy differential equations, *Nonlinear Analysis*, 55 (2003), no. 3, 255-268.

[10] R.E. Moore, *Interval Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1966.

[11] A. Neumaier, *Interval Methods for Systems of Equations*, Cambridge University Press (1990).

[12] M. Puri, D. Ralescu, Differentials of fuzzy functions, *J. Math. Anal. Appl.*, 91(1983), 552-558.

[13] L. Stefanini, A generalization of Hukuhara difference for interval and fuzzy arithmetic, in D. Dubois et Al. (Eds), *Soft Methods for Handling Variability and Imprecision*, Advances in Soft Computing (vol 48), Springer, 2008.

[14] L. Stefanini, B. Bede, Generalized Hukuhara Differentiability of Interval-valued Functions and Interval Differential Equations, *Nonlinear Analysis*, (2009) to appear. doi:10.1016/j.na.2008.12.005.

ISBN: 978-989-95079-6-8

Phi-Calculus fuzzy arithmetic in control: Application to model based control

Laurent Vermeiren[1,2,3], Thierry Marie Guerra[1,2,3], Hakim Lamara[1,2,3]

[1] Univ Lille Nord de France, F-59000 Lille, France
[2] UVHC, LAMIH, F-59313 Valenciennes, France
[3] CNRS, UMR 8530, F-59313 Valenciennes, France
Email: {laurent.vermeiren, guerra}@univ-valenciennes.fr

***Abstract*—*This paper aims at applying a fuzzy arithmetic of intervals calculus and fuzzy quantities to automatic control. This one called φ–calculus fuzzy arithmetic is more practical than the extension principle one and α-cut based methods. It comes from a different representation of fuzzy numbers. The present paper follows up work in introducing. The present paper is interested in its use for a fuzzy internal model control scheme based.*

Keywords— Fuzzy arithmetic, fuzzy model inversion, fuzzy number, weighted fuzzy fusion, model based control.

1 Introduction

The fuzzy set theory elaborated by L.A. Zadeh [1] has been shown to be used in the characterization of fuzziness and/or uncertainty using a coherent mathematical model. Various applications came at hand, in particular in the fuzzy control area [2, 3]. The uncertain calculus was undergoing quite a boom these last years. Several works exhibit interesting results, often referring to intervals theory [4, 5] or arithmetic based on fuzzy numbers such as Triangular or Trapezoidal Fuzzy Numbers [6, 7]. In the fuzzy arithmetic case, the different approaches are generally based either on Zadeh's extension principle [1, 6], either on fuzzy relation [8], or finally use the α-cuts [9]. However, no general approach allowing common arithmetic operations to be used on fuzzy numbers is available. This work proposes the use of a so-called φ–calculus fuzzy arithmetic based on a "different" modeling of fuzzy numbers [10, 11]. For this algebra, the modeling of fuzzy numbers is considered through the distribution function instead of the classical membership function. The first part presents some generalities on this φ–calculus arithmetic.

One of the main interests of this algebra is to provide some nice properties for what is called "exact calculus" with fuzzy numbers. These properties can be used in order to invert a fuzzy model. Therefore, the second part presents a possible application in control to fuzzy internal model control scheme.

2 Fuzzy arithmetic: φ-calculus

In the literature, many modeling approaches of imprecision, which is involved in many applications and domains, use fuzzy numbers and fuzzy arithmetic. In many works, the methods have led to the development of various membership functions for representing fuzzy numbers. Fuzzy numbers are often represented in applications by LR fuzzy sets and in particular, triangular and trapezoidal fuzzy sets. There exist two approaches for fuzzy arithmetic, on one hand the Zadeh's extension principle, on the other hand, the alpha-cuts and intervals arithmetic. Our works concern the first case, using a new modeling for fuzzy numbers [11], and it allows including a large part of the results existing in the domain of arithmetic [4, 6, 8, 12].

2.1 Fuzzy numbers modeling

Usually, a fuzzy number is modeled by its membership function $\mu_{\tilde{a}}$ not null on a bounded set $Supp(\tilde{a}) \subset \mathbb{R}$. For example, a triangular fuzzy number (TFN) $\tilde{a}$ can be represented by the shorthand symbol (b, m, c) with $\mu_{\tilde{a}}(x)=1$ for $x=m$ (mode) and the kernel defined by the interval $[m-b, m+c]$.

Herein, instead of the classical membership function, the modeling used for the representation of fuzzy numbers is based on the distribution function $\varphi_{\tilde{a}}$ defined by the following expression:

$$\varphi_{\tilde{a}}(x)=\frac{\int_{-\infty}^{x}\mu_a(t)\cdot dt}{\int_{-\infty}^{+\infty}\mu_a(t)\cdot dt} \qquad (1)$$

Convergence of $\int_{-\infty}^{+\infty}\mu_a(t)\cdot dt$ (finite cardinality of $\tilde{a}$) is assumed by considering only membership functions on a compact support $I \subset \mathbb{R}$. Thus, the major interest is that the distribution function for all fuzzy number $\tilde{a}$ is always an increasing monotone function, from $Supp(\tilde{a})$ to $[0,1]$. Thus, an inverse function $\varphi_{\tilde{a}}^{-1}$ from $[0,1]$ to $Supp(\tilde{a})=[\underline{a},\overline{a}]$ can always be defined and its definition is very important for the operations. The set of fuzzy numbers represented by a distribution function is noted Φ.

It should be noticed that the distribution function for a singleton is also a singleton and the one for an interval is a line between its bounds.

ISBN: 978-989-95079-6-8

2.2 Fuzzy realization and extension

Similarly to defuzzification and fuzzification concepts, it is necessary to define a relation between a crisp value and a fuzzy number. Therefore, an application from Φ to $\mathbb{R}$, which associates a crisp number to a fuzzy number, is called *fuzzy realization*. Conversely, an application from $\mathbb{R}$ to Φ is called *fuzzy extension*.

The choice of a realization depends obviously on the application. The most frequently encountered are:

- The median realisation noted $R_{med}(\tilde{a})$, it associates to a distribution function $\varphi_{\tilde{a}}$ the number a_0 such as $\varphi_{\tilde{a}}(a_0)=0.5$. Through this median realization, the following equivalence relation $\Re$ in Φ can be defined: the set of the functions $\varphi_{\tilde{a}}$ such as $\varphi_{\tilde{a}}(a_0)=0.5$ defines the class of equivalence of $a_0 \in \mathbb{R}$. Every element of this class will be called "fuzzy a_0", and denoted $\tilde{a}_0$.
- The modal realisation noted $R_{\mathrm{mod}}(\tilde{a})$ is defined as: $a_0 = R_{\mathrm{mod}}(\tilde{a}) \Leftrightarrow \varphi_{\tilde{a}}(a_0) = \underset{x\in\tilde{a}}{Max}\left[\mu_{\tilde{a}}(x)\right]=1$.
- The mean realization noted $R_{mean}(\tilde{a})$ is defined as: $a_0 = R_{mean}(\tilde{a}) = \int_0^1 \varphi_{\tilde{a}}^{-1}(y).dy$.

2.3 arithmetic operations

Consider two fuzzy numbers $\tilde{a}$ and $\tilde{b}$ using their distribution $\varphi_{\tilde{a}}$ and $\varphi_{\tilde{b}}$. For a fuzzy number $\tilde{a}$ whose support includes 0 let us also define its "negative" $\varphi_{\underline{\tilde{a}}}$ and "positive" $\varphi_{\overline{\tilde{a}}}$ parts as $\varphi_{\tilde{a}} = \varphi_{\underline{\tilde{a}}} + \varphi_{\overline{\tilde{a}}}$ with:

$$\varphi_{\underline{\tilde{a}}}(x)=\begin{cases}0 & \text{if } x<\underline{a}\\ \varphi_{\tilde{a}}(x) & \text{if } x\in[\underline{a},0]\\ 1 & \text{if } x>0\end{cases}, \qquad (2)$$

$$\varphi_{\overline{\tilde{a}}}(x)=\begin{cases}0 & \text{if } x<0\\ \varphi_{\tilde{a}}(x) & \text{if } x\in[0,\overline{a}]\\ 1 & \text{if } x>\overline{a}\end{cases} \qquad (3)$$

The classical arithmetic operations on fuzzy numbers (addition $\tilde{a}+\tilde{b}$, pseudo-opposite $-\tilde{a}$, subtraction $\tilde{a}-\tilde{b}$, multiplication $\tilde{a}\times\tilde{b}$, pseudo-inverse only for a non mixed-type fuzzy number $\tilde{b}^{-1}=1/\tilde{b}$ and division $\tilde{a}/\tilde{b}$) are defined hereinafter [9].

$$\begin{aligned}
&\varphi^{-1}_{\tilde{a}+\tilde{b}}(x)=\varphi^{-1}_{\tilde{a}}(x)+\varphi^{-1}_{\tilde{b}}(x)\\
&\varphi_{-\tilde{a}}(x)=1-\varphi_{\tilde{a}}(-x)\ \forall x\in\mathbb{R}\\
&\varphi^{-1}_{\tilde{a}-\tilde{b}}(x)=\varphi^{-1}_{\tilde{a}}(x)+\varphi^{-1}_{-\tilde{b}}(x)\\
&\varphi^{-1}_{\tilde{a}\times\tilde{b}}(x)=\varphi^{-1}_{\underline{\tilde{a}}\times\underline{\tilde{b}}}(x)+\varphi^{-1}_{\overline{\tilde{a}}\times\overline{\tilde{b}}}(x)\\
&\varphi_{\tilde{a}^{-1}}(x)=1-\varphi_{\tilde{a}}(1/x)\ \forall x\in\mathbb{R}^*\\
&\varphi^{-1}_{\tilde{a}/\tilde{b}}(x)=\varphi^{-1}_{\tilde{a}\times\tilde{b}^{-1}}(x)
\end{aligned} \qquad (4)$$

with $\varphi^{-1}_{\underline{\tilde{a}}\times\underline{\tilde{b}}}=\min\left[\varphi^{-1}_{\underline{\tilde{a}}}\times\varphi^{-1}_{\underline{\tilde{b}}},\varphi^{-1}_{\overline{\tilde{a}}}\times\varphi^{-1}_{\overline{\tilde{b}}}\right]$ and $\varphi^{-1}_{\overline{\tilde{a}}\times\overline{\tilde{b}}}=\max\left[\varphi^{-1}_{\underline{\tilde{a}}}\times\varphi^{-1}_{\underline{\tilde{b}}},\varphi^{-1}_{\overline{\tilde{a}}}\times\varphi^{-1}_{\overline{\tilde{b}}}\right]$.

Let us notice that these four classical operations are compatible with the Moore's interval operations.

2.4 Weighted fuzzy fusion operator

This part deals with a fusion operator introduced within the framework of this Φ-calculus algebra. This operator is called the weighted fuzzy fusion (WFF) [13] and is useful for the next sections.

Consider two fuzzy numbers $\tilde{a},\tilde{b}\in\Phi$ and their distribution $\varphi_{\tilde{a}},\varphi_{\tilde{b}}: x\in\mathbb{R}\rightarrow y\in[0,1]$. The weighted fuzzy fusion (WFF) of the two fuzzy numbers is defined as:

$$\varphi_{WFF}(x)=\frac{p_{\tilde{a}}\cdot\varphi_{\tilde{a}}(x)+p_{\tilde{b}}\cdot\varphi_{\tilde{b}}(x)}{p_{\tilde{a}}+p_{\tilde{b}}},\quad \forall x\in\mathbb{R} \qquad (5)$$

with $p_{\tilde{a}},p_{\tilde{b}}\in\mathbb{R}$ and $p_{\tilde{a}}+p_{\tilde{b}}=1$.

3 Internal model control

3.1 Principle of the structure

As with any open-loop control scheme, the internal model control (IMC) structure applies only on stable systems with a minimum phase behavior, according to Fig. 1.

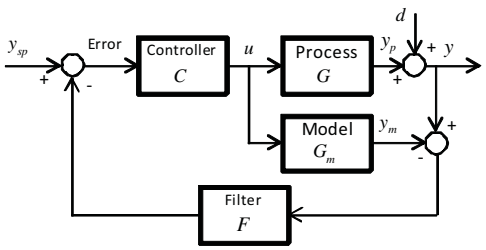


Figure 1: Internal model control scheme.

The filter F is introduced in order to filter out the measurement noise and to introduce some robustness in the loop. Its static gain is 1. Classically for linear models, it can be shown that the output equation y_k (Fig. 1) is given by:

$$y_k = F_{FB}(q^{-1})\cdot y_{sp_k} + S_{yd}(q^{-1})\cdot d_k \qquad (6)$$

with:

$$F_{FB}(q^{-1})=\frac{G(q^{-1})C(q^{-1})}{1+C(q^{-1})F(q^{-1})\left[G(q^{-1})-G_m(q^{-1})\right]}$$

$$S_{yd}(q^{-1})=\frac{1-G_m(q^{-1})C(q^{-1})F(q^{-1})}{1+C(q^{-1})F(q^{-1})\left[G(q^{-1})-G_m(q^{-1})\right]}$$

and the control law equation is given by:

ISBN: 978-989-95079-6-8

$$u_k = \frac{C\left(q^{-1}\right)}{1-C\left(q^{-1}\right)F\left(q^{-1}\right)G_m\left(q^{-1}\right)}\left(y_{c_k} - F\left(q^{-1}\right)y_k\right) \quad (7)$$

The minimum purpose of the internal model structure is to guarantee closed-loop stability and disturbance rejection, i.e. $F_{FB}(1)=1$ and $S_{yd}(1)=0$.

Recall also that the internal model control performs a kind of model simplification. Therefore its extension to fuzzy internal control needs the definition of an inverse model. The next part treats this issue.

3.2 Model inversion

Let us consider a stable time-invariant controllable SISO linear process model whose mathematical description is given in the form of recurrent equation:

$$y_k = -\sum_{i=1}^{N}\left(a_i \cdot y_{k-i} + b_i \cdot u_{k-i}\right) \quad (8)$$

where y_k and u_k correspond to the process output and input at sample k, $a_i, b_i, i \in \{1,\ldots,N\}$, denote the model parameters.

Suppose now that these parameters are fuzzy numbers, $\tilde{a}_i, \tilde{b}_i, i \in \{1,\ldots,N\}$ leading to:

$$y_k = -\sum_{i=1}^{N}\left(\tilde{a}_i \cdot y_{k-i} + \tilde{b}_i \cdot u_{k-i}\right) \quad (9)$$

For model inversion, the following question arises [14]:

knowing the description of uncertainties for the model, is it possible to synthesize a controller, based on the inverse model, able to maintain the model output within a tolerance envelope around the exact trajectory y_{sp} ?

Which means:

$$\text{Model output} \in \left[y_{sp} - \Delta,\ y_{sp} + \Delta\right] \quad (10)$$

where Δ is the accepted tolerance around the nominal trajectory.

To answer to this question, this work will use an approach developed in [14]. Notice that the "fuzzy" part of the uncertainties is not really taken into account, i.e. this problem could be solved using interval techniques, for example [15, 16, 17]. Nevertheless, the arithmetic provided therein gives an "easy" way to solve non exact calculus by means of probability distribution.

To compute the control law u_k at sample k, the terms defined at previous samples $\{y_{k-1}, i=1,\ldots,N\}$ and $\{u_{k-1}, i=1,\ldots,N\}$ are known. As $b_1 \neq 0$, equation (6) can be rewritten as:

$$u_k = \frac{1}{b_1}\left(y_{k+1} + \sum_{i=1}^{N}\left(a_i \cdot y_{k-i+1}\right) - \sum_{i=2}^{N}\left(b_i \cdot u_{k-i+1}\right)\right) \quad (11)$$

Or, considering the unknown term at sample k, y_{k+1}, expression (9) can be rewritten as follow:

$$u_k = \frac{1}{b_1}\left(y_{k+1} + \psi\left(z(k)\right)\right) \quad (12)$$

with

$$\psi\left(z(k)\right) = \sum_{i=1}^{N} a_i \cdot y_{k-i+1} - \sum_{i=2}^{N} b_i \cdot u_{k-i+1} \text{ and}$$

$$z(k) = \left[y_k \ \ \ldots \ \ y_{k-N+1} \ \ u_{k-1} \ \ \ldots \ \ u_{k-N+1}\right]^T \in \mathbb{R}^{2N-1}.$$

For a linear model without time delay, if y_{k+1} is replaced with the desired reference trajectory the ideal result is: $y_{k+1} = y_{sp_k}$. Therefore, with a perfect compensation, the model output follows the desired trajectory with a pure time delay corresponding to one sample.

Going back to the question (cf. (10)), we want to guarantee that $y_{k+1} \in \left[\underline{y}_{k+1}, \overline{y}_{k+1}\right]$. With equations (11) and (12), for a fuzzy model (9), we have straightforwardly:

$$\tilde{u}_k = \frac{1}{\tilde{b}_1}\left(\tilde{y}_{k+1} + \tilde{\psi}\left(\tilde{z}(k)\right)\right) \quad (13)$$

$$\tilde{\psi}\left(\tilde{z}(k)\right) = \sum_{i=1}^{N} \tilde{a}_i \cdot \tilde{y}_{k-i+1} - \sum_{i=2}^{N} \tilde{b}_i \cdot \tilde{u}_{k-i+1} \text{ and}$$

$$\tilde{z}(k) = \left[\tilde{y}_k \ \ \ldots \ \ \tilde{y}_{k-N+1} \ \ \tilde{u}_{k-1} \ \ \ldots \ \ \tilde{u}_{k-N+1}\right]^T \in \mathbb{R}^{2N-1}.$$

Equation (13) gives a model inversion description using fuzzy numbers. The term $\tilde{y}_{k+1}$ is replaced with a desired trajectory defined by a fuzzy number $\tilde{y}_{sp_k}$ to obtain a causal control law. For example, if $\tilde{y}_{sp_k}$ is a triangular fuzzy number (TFN) written as:

$$\tilde{y}_{sp_k} = triple\left(y_{sp} - \Delta,\ y_{sp},\ y_{sp} + \Delta\right) \quad (14)$$

the control law is:

$$\tilde{u}_k = \frac{1}{\tilde{b}_1}\left(\tilde{y}_{sp_k} + \tilde{\psi}\left(\tilde{z}(k)\right)\right) \quad (15)$$

$$\tilde{\psi}\left(\tilde{z}(k)\right) = \sum_{i=1}^{N} \tilde{a}_i \cdot \tilde{y}_{sp_k-i+1} - \sum_{i=2}^{N} \tilde{b}_i \cdot \tilde{u}_{k-i+1} \text{ and}$$

$$\tilde{z}(k) = \left[\tilde{y}_{sp_k} \ \ \ldots \ \ \tilde{y}_{sp_k-N+1} \ \ \tilde{u}_{k-1} \ \ \ldots \ \ \tilde{u}_{k-N+1}\right]^T \in \mathbb{R}^{2N-1}.$$

Of course, the vector $\tilde{z}(k)$ depends on the control part meaning that all $\tilde{y}_i$ are replaced with desired set point $\tilde{y}_{sp_i}$ which corresponds to the schematic diagram of Fig. 2.

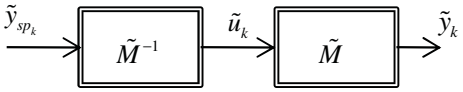


Figure 2: Fuzzy model inversion.

Naturally, perfect cancellation, i.e. $\tilde{y}_{k+1} = \tilde{y}_{sp_k}$ is not possible. Moreover using arithmetic of fuzzy numbers, or intervals or φ–calculus will introduce over-estimations. To illustrate this important point, an example issued from [14] is discussed hereinafter.

Let us consider the second-order:

$$\tilde{y}_k = -\tilde{a}_1 \cdot \tilde{y}_{k-1} - \tilde{a}_2 \cdot \tilde{y}_{k-2} + \tilde{b}_1 \cdot \tilde{u}_{k-1} + \tilde{b}_2 \cdot \tilde{u}_{k-2} \quad (16)$$

where $\tilde{a}_1$, $\tilde{a}_2$, $\tilde{b}_1$ and $\tilde{b}_2$ are fuzzy numbers defined as:

ISBN: 978-989-95079-6-8

$$\begin{aligned}\tilde{a}_1 &= triple(-0.6, -0.55, -0.5)\\ \tilde{a}_2 &= triple(0.05, 0.1, 0.15)\\ \tilde{b}_1 &= triple(0.6, 0.625, 0.65)\\ \tilde{b}_2 &= triple(0.15, 0.2, 0.25)\end{aligned} \quad (17)$$

The desired trajectory is:

$$y_{sp_k} = 0.5 \cdot \left(\sin\left(\frac{2\pi k}{50}\right) + \sin\left(\frac{2\pi k}{75}\right)\right) \quad (18)$$

and $\tilde{y}_{sp_k}$ is chosen as

$$\tilde{y}_{sp_k} = triple\left(y_{sp_k} - 1,\ y_{sp_k},\ y_{sp_k} + 1\right) \quad (19)$$

Using (15), the evolutions of the set point $\tilde{y}_{sp_k}$ envelopes (minimum $\underline{y}_{sp_k}$ and maximum $\overline{y}_{sp_k}$ of $Supp\left(\tilde{y}_{sp_k}\right)$) and the output model $\tilde{y}_k$ envelopes are illustrated in Fig. 3. Obviously, like interval calculus or α-cut fuzzy arithmetic, it clearly appears that the φ–calculus arithmetic generates overestimated results.

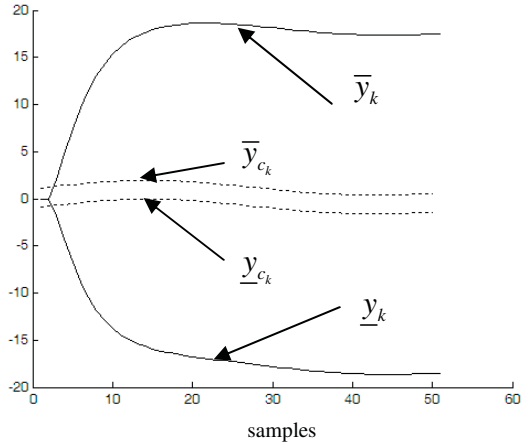


Figure 3: Set point $\tilde{y}_{sp_k}$ and output model $\tilde{y}_k$ envelopes.

We claim that this overestimation is not justified in such cases. Therefore, we need to introduce some extra constraints in order to take into account some knowledge and to reduce the pessimism of the results.

3.3 "Exact" inverse computation

Equation (15) corresponds to resolution of a fuzzy affine equation:

$$\tilde{b} \times \tilde{x} + \tilde{a} = \tilde{y} \quad (20)$$

The idea developed in [15] is to exactly solve this equation according to the two following steps.

Step1: Solve (20) with respect to $\tilde{d} = \tilde{b} \times \tilde{x}$.

The objective is to compute an "acceptable" solution $\tilde{d}$ of the equation $\tilde{d} + \tilde{a} = \tilde{y}$; i.e. the solution being exact when possible and approximated otherwise.

Notice that the exact solution can always be computed according to the arithmetic used:

$$\varphi_{\tilde{d}}^{-1}(u) = \varphi_{\tilde{y}}^{-1}(u) - \varphi_{\tilde{a}}^{-1}(u) \quad (21)$$

The problem arises when the result $\varphi_{\tilde{d}}^{-1}(u)$ is not an increasing function, resulting to $\tilde{d} \notin \Phi$. Nevertheless we can provide an approximate solution to the problem $\tilde{d}_{app} \in \Phi$. The fact of proposing an alternative to exact calculus is the key idea. Contrary to what generally authors accept, i.e. no solution for interval arithmetic or α-cut fuzzy arithmetic cases [14], we provide an approximate solution based on the initial $\varphi_{\tilde{d}}^{-1}(u)$.

For that purpose, we propose to use the weighted fuzzy fusion (WFF) operator on the set of pairs $\left(\varphi_{\tilde{d}}^{-1}(u_i), u_i\right)$, $i \in \{1, \ldots, N\}$. It results in a re-ordering of the pairs that constructs an increasing function $\tilde{d}_{app}$, thus $\tilde{d}_{app} \in \Phi$ is a feasible solution to the problem.

Step2: Finding a solution to equation $\tilde{d} = \tilde{b} \times \tilde{x}$ with $0 \notin Supp\left(\tilde{b}\right)$. If $\tilde{b} > 0$ and $\tilde{d} > 0$, the exact calculus uses a point by point division:

$$\varphi_{\tilde{x}}^{-1}(u) = \varphi_{\tilde{d}}^{-1}(u) \big/ \varphi_{\tilde{b}}^{-1}(u) \quad (22)$$

In the general case, we use a "positive-negative" decomposition (see section 2.3) for $\tilde{d}$ and $\tilde{x}$. We define:

$$\begin{cases}\varphi_{\tilde{d}}^{-1}(u) = \varphi_{\tilde{d}_-}^{-1}(u) + \varphi_{\tilde{d}_+}^{-1}(u)\\ \varphi_{\tilde{x}}^{-1}(u) = \varphi_{\tilde{x}_-}^{-1}(u) + \varphi_{\tilde{x}_+}^{-1}(u)\end{cases} \quad (23)$$

with $\varphi_{\tilde{d}_-}^{-1}(u)$, $\varphi_{\tilde{d}_+}^{-1}(u)$, $\varphi_{\tilde{x}_-}^{-1}(u)$ and $\varphi_{\tilde{x}_+}^{-1}(u)$ defined equations (3) and (4). At last, the exact solution is given by:

$$\begin{cases} if\ \tilde{b} > 0, \begin{cases}\tilde{x}_+ = \tilde{d}_+ \big/ \tilde{b},\ \tilde{x}_- = -\left(\left(-\tilde{d}_-\right) \big/ \tilde{b}\right),\\ \tilde{x} = \tilde{x}_+ + \tilde{x}_-\end{cases}\\ if\ \tilde{b} < 0, \begin{cases}\tilde{x}_+ = \tilde{d}_+ \big/ \left(-\tilde{b}\right),\ \tilde{x}_- = -\left(\left(-\tilde{d}_-\right) \big/ \left(-\tilde{b}\right)\right),\\ \tilde{x} = -\left(\tilde{x}_+ + \tilde{x}_-\right)\end{cases}\end{cases} \quad (24)$$

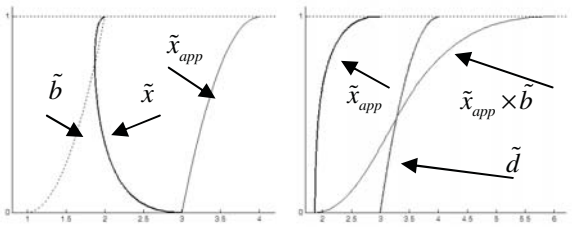
Once again the problem arises when the result $\varphi_{\tilde{x}}^{-1}(u)$ is not an increasing function, resulting to $\tilde{x} \notin \Phi$. Therefore approximate solution to the problem $\tilde{x}_{app} \in \Phi$ is generated in the same way using the WFF operator on the set of pairs $\left(\varphi_{\tilde{x}}^{-1}(u_i), u_i\right)$, $i \in \{1, \ldots, N\}$.

Consider an example to illustrate this second step with the fuzzy numbers $\tilde{b} = triple(1, 2, 2)$ and $\tilde{d} = triple(3, 3, 4)$. As shown Fig. 4(a), the exact solution $\tilde{x} \notin \Phi$, i.e. $\varphi_{\tilde{x}}^{-1}$ is not a monotone increasing function. Using the weighted fuzzy fusion (WFF) operator, an approximate solution $\tilde{x}_{app}$ is generated by re-ordering $\varphi_{\tilde{x}}^{-1}$. Fig. 4(b) shows the comparison between $\tilde{d}$ and $\tilde{b} \times \tilde{x}_{app}$. In a sense, the difference between these two fuzzy numbers exhibits the pessimism induced by the method when no exact solution in Φ is available. This pessimism can be shown comparing the supports of the numbers:

$$Supp\left(\tilde{d}\right) = [3, 4] \subset Supp\left(\tilde{b} \times \tilde{x}_{app}\right) = [1.87, 6].$$

ISBN: 978-989-95079-6-8

The next example demonstrates very clearly the interest of the complete algorithm used to find solutions to the problem $\tilde{b}\times\tilde{x}+\tilde{a}=\tilde{y}$. Consider again the inverse model of section 3.2 with the fuzzy numbers (17) and the desired trajectory (18), (19). Fig. 5 shows the same trial as for Fig. 3. For this test, we obtain the result considering, i.e. $\tilde{y}_{k+1}=\tilde{y}_{sp_k}$. This result indicates that an "exact solution" has been found at each time k.



a) « Exact » result b) φ–calculus result

Figure 4: Results with the numbers $\tilde{b}$ and $\tilde{d}$.

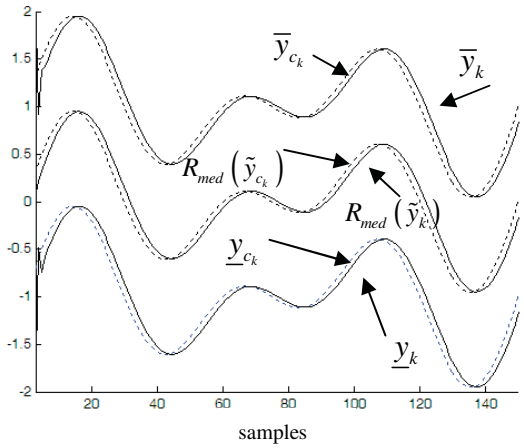


Figure 5: Set point $\tilde{y}_{sp_k}$ and output model $\tilde{y}_k$ envelopes.

In order to show the interest of the method, we will change the constraints level. To do so, the values of model uncertainties (17) are increased in order to deal with cases where an exact inversion is not always possible. New bounds for the fuzzy numbers (17) are defined. They correspond to the greatest possible values with pole-zero assignment located inside the unit circle – for evident stability purpose. Moreover, the fuzzy numbers are not more symmetric and correspond to:

$$\begin{aligned}\tilde{a}_1 &= triple(-0.825,-0.4,0.3)\\ \tilde{a}_2 &= triple(-0.175,0.25,0.95)\\ \tilde{b}_1 &= triple(0.415,0.625,0.835)\\ \tilde{b}_2 &= triple(-0.01,0.2,0.41)\end{aligned} \qquad (25)$$

Fig. 6 repeats the same trial with the desired trajectory (18), (19). The figure exhibits that an exact simplification, i.e. $\tilde{y}_{k+1}\neq\tilde{y}_{sp_k}$ is not always possible. Contrary to have no solution in such cases, the proposed approximate solutions seem perfectly adapted.

With the help of this inverse model control strategy based on constraints, justified in the control framework, it is now possible to consider a fuzzy internal model control.

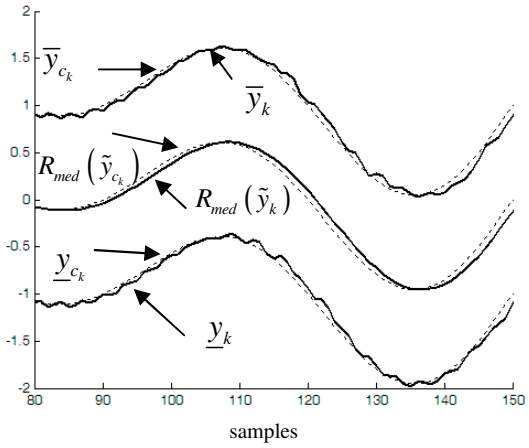


Figure 6: Set point $\tilde{y}_{sp_k}$ and output model $\tilde{y}_k$ envelopes.

3.4 Fuzzy internal model control

From Fig. 6 showing IMC scheme and different equations (6) and (7), it is possible to define several solutions.

First solution: Let $G_m(q^{-1})=q^{-r}G_m^+(q^{-1})$, $G_m^+(q^{-1})$ corresponds to function with an inverse (exactly proper and stable). We chose $C(q^{-1})=\dfrac{1}{G_m^+(q^{-1})}$ and propose the equivalent model given Fig. 7.

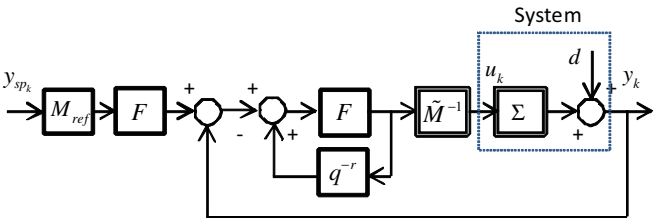


Figure 7: IMC1, equivalent scheme.

Indeed, equation (7) can be rewritten as follow:

$$u_k=\frac{1}{G_m^+(q^{-1})\left(1-q^{-r}\cdot F(q^{-1})\right)}\left(y_{c_k}-F(q^{-1})\,y_k\right) \qquad (26)$$

By analogy, we have the fuzzy inverse model $\left[\tilde{G}_m^+(q^{-1})\right]^{-1}$ multiplied by linear transfer function $\dfrac{F(q^{-1})}{1-q^{-r}\cdot F(q^{-1})}$.

For every case, it is possible to add a model reference control M_{ref} in order to attenuate the control signal.

ISBN: 978-989-95079-6-8

3.5 Example

Consider a nominal second-order transfer function:

$$\frac{y\left(q^{-1}\right)}{u\left(q^{-1}\right)}=\frac{B\left(q^{-1}\right)}{A\left(q^{-1}\right)}=\frac{0.0233\cdot q^{-1}+0.0197\cdot q^{-2}}{1-1.5729\cdot q^{-1}+0.6037\cdot q^{-2}} \quad (27)$$

and the following definition for the fuzzy numbers:

$$\begin{aligned}\tilde{a}_1 &= triple\left(-1.574,-1.573,-1.497\right)\\ \tilde{a}_2 &= triple\left(0.602,0.603,0.652\right)\\ \tilde{b}_1 &= triple\left(0.013,0.023,0.033\right)\\ \tilde{b}_2 &= triple\left(0.01,0.02,0.03\right)\end{aligned} \quad (28)$$

The control structure used is the IMC1 Fig. 7 with a reference model defined as:

$$\frac{B_m\left(q^{-1}\right)}{A_m\left(q^{-1}\right)}=\frac{0{,}1784\cdot q^{-1}+0{,}1071\cdot q^{-2}}{1-0{,}9315\cdot q^{-1}+0{,}2169\cdot q^{-2}} \quad (29)$$

Fig. 8 shows the results for different set points. The first set-point leads to results around the nominal model (27) (represented by the value 2). Secondly, two successive variations are made (set-point value of 3 and set-point value of 4), lastly a return to nominal position.

During both successive changes the model is changed in order to reach the bounds on $\tilde{a}_1$ and $\tilde{b}_2$. Two disturbances have been also added at samples 900 and 2300. The responses show a good robustness according to the parametric variations of the model.

This work presents a first possible track to use fuzzy arithmetic for control purpose. Next steps would be to prove stability and evaluate the robustness of these approaches. We can think to several possibilities including Lyapunov approach that are insightfully used for Takagi-Sugeno models [18] or Kharitonov polynomials [19].

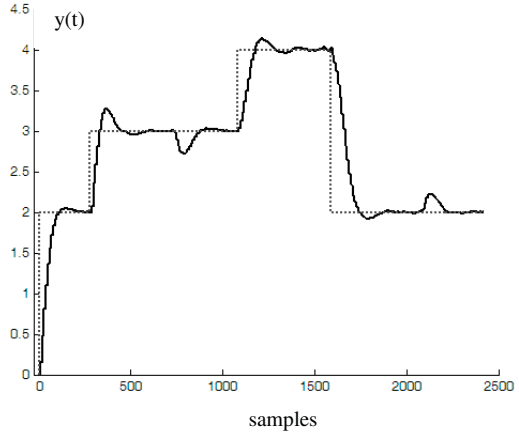


Figure 8: IMC1 results.

4 Conclusions

This paper attempts to show possible applications of the φ–calculus arithmetic for automatic control. As a first step, a new characterization of fuzzy numbers by the distribution function instead of the classical membership function has been presented. For this algebra, it is possible to use a set of arithmetic operators (addition, opposite and subtraction, multiplication, inverse and quotient) compatible with the classical algebra by using the median realization, or the Moore's calculus by using the fuzzy number support. A procedure allowing an "exact" calculus of an inverse model was proposed on the basis of φ–calculus. This approach was successfully implemented on internal model control structures.

Thus, the application of φ–calculus arithmetic to the stability of system seems to open some new prospects for other basic concepts (observation, stabilization…).

Acknowledgment

This work was supported in part by International Campus on Safety and Intermodality in Transportation, the European Community (through the FEDER European Funds for Regional Development), the Délégation Régionale à la Recherche et à la Technologie, the Ministère de l'Enseignement supérieur et de la Recherche, the région Nord Pas-de-Calais and the Centre National de la Recherche Scientifique (CNRS): the authors gratefully acknowledge the support of these institutions.

References

[1] L A. Zadeh, *Fuzzy sets*, Information & Control, 8:338-353, 1965.

[2] A. Sala, T. M. Guerra, R. Babuška. *Perspectives of fuzzy systems and control*, Fuzzy Sets and Systems, 156:432-444, 2005.

[3] L. Jaulin, M. Kieffer, O. Didrit, E. Walter. *Applied Interval Analysis*, Springer Edition, 2001.

[4] R.E. Moore. *Interval analysis*, Prentice-Hall-Englewood Cliffs, New Jersey, 1966.

[5] U.W. Kulish, W.L. Miranker. *Computer arithmetic in theory and practice*. Academic Press, 1981.

[6] D. Dubois, H. Fargier, J. Fortin. *A generalized vertex method for computing with fuzzy intervals*, Proc. of the International Conference on Fuzzy Systems (IEEE, ed.), 541-546, 2004.

[7] D. Dubois & H. Prade. *Fuzzy real algebra: some results*. Fuzzy Sets and Systems, 2:327-383, 1979.

[8] E. Sanchez. Solutions of fuzzy equations with extended operators. Fuzzy Sets and Systems, 12:237-248, 1984.

[9] Z. Zhao, R. Govind. Solutions of algebric equations involving generalised fuzzy numbers. Information sciences, 56:199-243, 1991.

[10] D. Roger, J.-M. Lecomte. *φ-calcul : une arithmétique pratique pour les nombres flous*. LFA'98, Rennes, 1998.

[11] H. Lamara, L. Vermeiren & D. Roger. « φ-calculus*: A new fuzzy arihtmetic* », 11th IEEE ETFA'06, Prague, Czech Republic, 2006.

[12] A. Kaufmann, M.M. Gupta. Introduction to fuzzy arithmetic theory and application. Van Nostrand Reinhold, New York, 1991.

[13] H. Lamara, L. Vermeiren, D. Roger. *Aggregation and fuzzification by weighted fuzzy fusion operator*. IEEE-FUZZ'07, London, United Kingdom, 2007.

[14] R. Boukezzoula, S. Galichet, L. Foulloy. *Nonlinear Internal Model Control: Application of inverse Model Based Fuzzy Control*, IEEE T. Fuzzy Systems, 11(6):814-829, 2003.

[15] R. Boukezzoula, L. Foulloy, S. Galichet. *Inverse Controller Design for Fuzzy Interval Systems*, IEEE T. Fuzzy Systems, 14(1):111-124, 2006.

[16] E. Gardeñes, M.A. Sainz, L. Jorba, R. Calm, R. Estela, H. Mielgo, A.A. Trepat. *Modal intervals*, Reliable Computing, 7:77-111, 2001, Kluwer Academic Publishers.

[17] J.E.F. Diaz. *Improvements in the Ray Tracing of Implicit Surfaces based on Interval Arithmetic*. Doctoral Thesis, Girona, Spain, 2008.

[18] T.M. Guerra, L. Vermeiren. LMI-based relaxed non quadratic stabilization conditions for non-linear systems in the Takagi-Sugeno's form, Automatica, 40(5) :823-829.

[19] C.W. Tao and J.S. Taur. Robust Fuzzy Control for a plant with fuzzy linear model, IEEE T. Fuzzy Systems, 13(1):20-41, 2005.

ISBN: 978-989-95079-6-8

Optimising the Fuzzy Granulation of Attribute Domains

Marcos E. Cintra[1] Heloisa A. Camargo[2] Trevor Martin[3]
1.Mathematics and Computer Science Institute, São Paulo University(USP)
P.O. Box 668, 13561-970, São Carlos-SP, Brazil
2.Computer Science Department, Federal University of São Carlos(UFSCar)
P.O. Box 676, 13565-905 São Carlos-SP
3.Mathematics Engeneering Department, University of Bristol
Queen's Building, University Walk, Bristol - UK - BS8 1TR
Email: cintra@icmc.usp.br, heloisa@dc.ufscar.br, Trevor.Martin@bristol.ac.uk

Abstract *— The definitions of the number of fuzzy sets and their proper distribution on their domains are fundamental issues for fuzzy systems since these basic parameters deeply affect the quality of the systems results, both in terms of performance rates and interpretability. Several methods have been proposed in the literature to define these parameters, although it is common to find works in which the number of fuzzy sets is defined empirically, distributing them equally in the domains. This paper presents a fast and easy method to estimate the number of fuzzy sets for each attribute and compares three methods for the distribution of fuzzy sets. Two of them are non-supervised methods, using same width and same frequency, the third one is an adaptation of the 1-R supervised method to discretize attributes. Experiments with 10 datasets for classification problems, 10-fold cross validation, using the Wang & Mendel method, and the classic and general fuzzy reasoning methods are presented and discussed.*

Keywords— Fuzzy systems, granulation of attributes, Wang & Mendel method, fuzzy data base generation.

1 Introduction

Fuzzy set theory and fuzzy logic [1] proposed by professor Loft A. Zadeh are the base for the fuzzy systems. Fuzzy set theory is used to represent and process information. The main characteristic of fuzzy sets, contrasting with crisp sets, is the progressive transition from one set to another. This natural characteristic of the fuzzy sets provides automatic mechanisms to deal with imprecision and uncertainty, which are inherent to real world knowledge. Moreover, the fuzzy logic theory prevents the creation of unnatural frontiers in the partitioning of attributes domains [2].

A fundamental issue, due to its direct impact on fuzzy systems, is the definition of the fuzzy sets that model the linguistic variables of a given domain, both in terms of shape (triangular, trapezoidal, S-function, etc) and partitioning of the attributes domains (number of fuzzy sets and their distribution). Unfortunately, there are no general rules or guidelines for these tasks that suit every domain [3]. In fact, many methods may have to be tested in order to find the appropriate definitions for a given application.

It is also common to find studies in which the definition of the membership functions is done empirically. In fact, the Gaussian, trapezoidal and triangular shapes are the most used in the literature, probably because they produce comparable results to other shapes and are easily interpretable [4]. Regarding the definition of the number of fuzzy sets and their distribution on the partitions, most of the papers in the literature use from 2 to 10 fuzzy sets for each attribute, equally distributed in their domains, *i.e.*, all fuzzy sets have same width.

Another aspect to be considered when defining fuzzy sets is personal interpretations, which make this task non trivial. For instance, let us consider a linguistic variable *temperature*; the interpretation of the linguist term *low temperature* is closely related to the region where one lives (consider a continental country such as Brazil, for example). Thus, personal interpretations may generate strong variations for the same concept: since each person may have different and personal interpretations on the meaning of a concept, it is natural that different membership functions may be created to define the same concept. This flexibility on the subjective selection of membership functions and, as previously stated, in the distribution of the fuzzy sets in their partitions shows the robustness of the fuzzy logic, which is closely related to the inherent characteristics of fuzzy sets [5].

Although this flexibility exists, the task of defining membership functions is largely ignored. The choice of the method, in fact, depends on the particular application and domain [6]. While the shape of the fuzzy sets may not present expressive differences in the results of fuzzy systems, the number of fuzzy sets and their distribution are relevant parameters that affect the system in terms of performance and in terms of interpretability. Thus, this paper presents a fast and easy method to estimate the number of fuzzy sets, in order to use this estimated number with more costly approaches, such as genetic algorithms. This paper also presents experiments carried out using three distinct methods for the definition of the distribution of fuzzy sets, in an attempt to provide further insight on the task, focusing specifically on classification problems. One of these methods uses same width distribution for the fuzzy sets, the second uses same frequency distribution, and the third takes the classes into consideration in the process.

This paper is organized as follows: Section 2 reviews methods for the automatic definition of fuzzy membership functions; Section 3 describes the three methods used in the experiments for the definition of fuzzy sets. Section 4 describes the heuristic method for the definition of the number of fuzzy sets proposed here. Section 5 presents the experiments and results, followed by the conclusions in Section 6.

ISBN: 978-989-95079-6-8

2 Methods for the Automatic Definition of Fuzzy Sets

There is a wide variety of methods applied for the definition of fuzzy sets, ranging from heuristic methods, genetic algorithms, artificial neural networks, clustering algorithms, the use of indexes, to the adaptation of classic machine learning methods, such as the K-NN algorithm. Next, we present some of these methods. More details on these and on other related methods can be found in [7] and on the provided references.

Several methods explore the use of special parameters for the distribution of fuzzy sets. In [2], for example, the authors use the concept of 2-tuple to define linguistic terms. The 2-tuple strategy is used to define both, the center of a fuzzy set (support), and a value for a symbolic translation, which enables the fuzzy sets to be moved sideways, always keeping the same base width. This symbolic translation of a linguistic term is a number within the interval $[-0.5, 0.5)$. A genetic algorithm is then used to define the best position for the fuzzy sets. The experiments used a fixed number of triangular shaped sets.

In [6] and [8] the authors use the Fuzzy C-Means algorithm for the definition of fuzzy membership functions. In the first method [6], a modified version of the Xie-Beni index [9] is used to define the number of clusters and, therefore, the number of fuzzy sets, for the Fuzzy C-Means algorithm. Once the number of fuzzy sets is defined, the centers of the sets are initialized based on two methods, the equalized universe method, which divides the partition equally by the number of sets, and the subtractive clustering method, which classifies the points present in the training data by their number of neighboring points (thus, a data point will have a high potential value if it has many neighboring points). After the initialization of the centers, the Fuzzy C-Means algorithm is employed to search for the best positions for centers. A hybrid learning algorithm for refining the system parameters based on the ANFIS method [10] is then presented. The second method [8] proposes the use of a Fuzzy C-Means variant for the generation of fuzzy term sets with ½ overlap. This method uses the mean squared error criterion to determine the number of fuzzy sets and the optimal shape of the membership functions associated with each term.

Genetic algorithms are used to tune the parameters of fuzzy sets in [11]. The genetic algorithm is used to find membership functions suitable for mining problems and then uses the final best set of membership functions to mine fuzzy association rules. The fitness of each chromosome is evaluated by the number of large 1-itemsets generated and by the suitability of the membership functions.

The use of artificial neural networks has also been explored for the definition of fuzzy membership functions. The authors in [12] and [13] present an overview of several neuro-fuzzy methods for the definition of fuzzy sets. These methods, although some present good results, have the problem of tunning parameters for the neural networks, and, in some cases, might need further tests.

There are also works focusing on the use of special indexes to generate membership functions. For instance, the concept of entropy is used in [14] and in [15] to determine membership functions. In [16] the authors propose a method for the definition of fuzzy sets based on two indexes, the fuzzy entropy and the fuzziness index, adopting the S shaped fuzzy set. These papers focus on image recognition applications. The Kappa measure was also used. In [17] the authors use the Kappa measure to calculate the *fuzzy Kappa* index in order to define a suitable segmentation of partitions.

Several other approaches can be found in the literature, such as heuristic methods [18], histograms [3], and even a fuzzy version of the classic machine learning K- nearest neighbor algorithm [19].

Although such a variety of methods exist, it is important to emphasize the fact that many studies define the number of fuzzy sets empirically, and employ the equalized universe method [6] to define the distribution of the fuzzy sets, which simply distributes the sets evenly in the partitions. Regarding the number of fuzzy sets, most researchers define this variable empirically using a range usually varying from 2 to 10 fuzzy sets per attribute. The main reasons why researchers decide to define such important variables without the aid of any formal method may include:

- The complexity of the methods available, which may require more time and effort to be implemented than the actual application in focus;
- The flexibility of the fuzzy logic for the definition of fuzzy variables and fuzzy terms, which allows the users to define their own fuzzy data base, since it will be the base for the generation of the fuzzy rules, which in turn will be adjusted to provide a suitable performance;
- The lack of consensus and/or guidelines on which of the available methods is the best for a given application;

The next section presents the methods used in the experiments conducted for this paper.

3 Evaluated Methods for Fuzzy Sets Distribution

The three methods tested and compared in this paper are the equalized universe method, same frequency method, and an adaptation of the 1-R method for attribute discretization. These methods are described next.

3.1 Equalized Universe Method

This method uses the same width for each fuzzy set [6]. Figure 1 presents an attribute described by five triangular fuzzy sets using the equalized universe method. Notice that, usually, the fuzzy sets at the extremes have half the width of the other fuzzy sets in the middle of the domain.

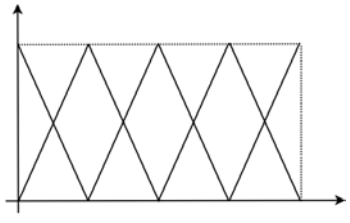


Figure 1: Attribute defined by 5 fuzzy sets using the equalized universe method.

This method is the most widely found in the literature. Some of the reasons for that may include the fact that it is

ISBN: 978-989-95079-6-8

a simple and easy method to use, and that it generates isosceles triangles, with ½ overlap, so no area of the domain will have a membership degree inferior to 0.5 [20]. Besides, they are easy to interpret.

3.2 Same Frequency Method

This method divides the domains of the attributes according to the frequency distribution of the examples. For our implementation, percentiles [21] were used to find the correct distribution for the sets. This way, if an attribute is described by 5 triangular fuzzy sets, the maximum membership degree for the first fuzzy set will be on the 16.7th percentile, on the 33.3th percentile for the second fuzzy set, on the 50.0th percentile for the third one, on the 66.7th percentile for the fourth one, and on the 83.3th percentile for the last fuzzy set. The triangles will not have the same base width, on the contrary, depending on the distribution of the examples, they are likely to be different. Notice that in the case of many examples with the same value, two or more fuzzy sets might be defined with exactly the same middle point. In order to avoid this problem, if the same example is chosen by two or more different percentile values, the neighboring examples of the left and right are chosen to replace the original value.

Figure 2 shows an attribute represented by 5 fuzzy sets distributed using the same frequency method, having 20 examples in the interval $[0, 10]$ (0, 0, 0, 1, 1, 2, 2, 2, 3, 5, 5, 6, 7, 7, 7, 9, 9, 9, 9, 9).

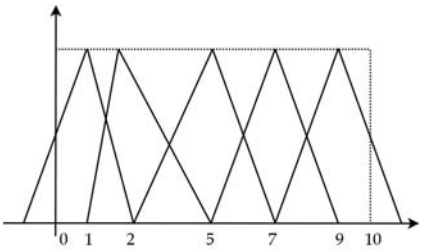


Figure 2: Attribute defined by 5 fuzzy sets using the same frequency method.

Notice that the fuzzy sets of the laterals are defined beyond the actual domain values. This strategy was used to avoid leaving any spaces of the domain with maximum membership degrees close to zero. The space between the actual beginning and end of the domain and the vertices of the triangles is calculated as the same space between the middle vertex of the triangle and the adjacent extremity of the domain.

3.3 1-R Supervised Method

The 1-R method is used for discretization of attributes. It basically divides the domain of an attribute in n intervals, in relation to the frequency of the examples, calculates the optimum class for each interval, and tries to find the best split point using the class. The original algorithm can be found in [22]. Algorithm 1 presents the adapted version of the 1-R algorithm used in the experiments.

Input: n = number of examples;
m = number of attributes;
i = number of fuzzy sets describing the attribute;
for a = *1* **to** m **do**
  Sort the examples by attribute a;
  Divide the examples in i intervals;
  for b = *1* **to** $i - 1$ **do**
    Calculate the optimal(majority) class for the interval;
    for c = *1* **to** *interval* $i - 2$ **do**
      for d = *index of first example of interval* $c + 1$ **to** *index of last example of interval* $c + 1$ **do**
        if *first example of interval* $c + 1$ *belongs to optimal class of interval* c **then** ;
        Adjust intervals so that the first example of interval $c + 1$ is transfered to interval c;
        else d = index of last example of interval $c + 1$;
        if d = *index of last example of interval* $c + 1$ **then** Split the examples of interval c between intervals c and $c + 1$;
      end
    end
  end
  For each interval, create a fuzzy set with core in the middle of the interval and laterals in the cores of the neighboring intervals;
end

Algorithm 1: 1-R adapted algoritghm.

The same strategy used for the same frequency method, to avoid leaving any spaces of the domain with maximum membership degrees close to zero, was used similarly for the 1-R adapted method.

Figure 3 shows an attribute represented by 5 fuzzy sets distributed by the same frequency method, having the following 20 examples (attribute value, class) in the interval $[0, 10]$: 0, 3; 0, 3; 0, 1; 1, 3; 1, 3; 2, 3; 2, 2; 2, 2; 3, 2; 5, 2; 5, 1; 6, 1; 7, 1; 7, 3; 7, 3; 9, 2; 9, 3; 9, 1; 9, 2; 9, 3.

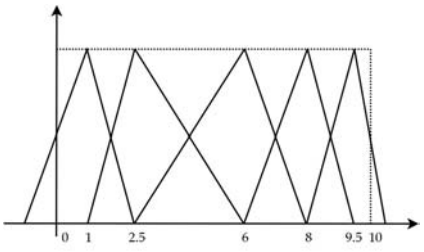


Figure 3: Attribute defined by 5 fuzzy sets using the 1-R adapted method.

4 Heuristic Method to Define the Number of Fuzzy Sets

The approaches used by the recently proposed methods to generate fuzzy rules usually demand a great computational effort, such as the genetic algorithms, that can take from minutes to days to converge to good solutions. Since the definition of the number of fuzzy sets is an open area of research, it is common to see researchers using a few different numbers of fuzzy sets for each attribute in order to validate their methods, usually ranging from 2 to 10, thus, repeating the complete round of experiments several times, according to the range of number of fuzzy sets used. If it were possible to establish an approximate optimal number of fuzzy sets, the total number of experiments would be greatly reduced. Thus, we propose a heuris-

ISBN: 978-989-95079-6-8

tic strategy to define this number using the Wang & Mendel method [23], which is able to generate fuzzy rule bases from data examples.

The Wang & Mendel method has been widely used due to its low complexity ($O(m \times n)$, m attributes, n examples) and the fact that it produces relatively small rule bases with good classification rates and no conflicting or redundant rules. However, nowadays it is possible to generate more precise fuzzy rule bases with even lower number of rules using other approaches, such as the genetic fuzzy systems or neural networks. The heuristic of the proposed method is based on the assumption that the Wang & Mendel method can be used as an indicator of the quality of the results that can be achieved by other methods, when using the same data base, *i.e.*, the same number, distribution, and shape of fuzzy sets.

Experiments shown in Section 5 were carried out using from 2 to 10 triangular shaped fuzzy sets defining each attribute of the domains. The idea is to run the experiments with the Wang & Mendel method before running the experiments with more time consuming approaches, and select the number of fuzzy sets that produced the best accuracy rates, or, in case of ties, the smaller number, due to interpretability reasons. Notice that a tolerance interval may be defined to allow the user to select the smaller number of fuzzy sets when a small difference between the accuracy rates is not significant, but the difference in the number of fuzzy sets is.

5 Experiments and Results

The experiments were conducted using 10-fold cross validation on 10 datasets. Table 1 shows a summary of the characteristics of the datasets used in the experiments, presenting the number of examples in each dataset (# Examples), number of attributes (# Attribs), number of classes (# Classes), and the majority class error (MCE).

Table 1: Characteristics of the datasets.

Dataset	# Examples	# Attribs	# Classes	MCE
Breast	684	10	3	34.99
Bupa	345	4	2	42.00
Credit	653	15	2	45.33
Diabetes	769	8	2	34.98
Glass	220	9	7	65.45
Heart	270	13	2	44.40
Iris	150	4	3	66.60
Segment	210	19	7	85.72
Vehicle	846	18	4	74.20
Wine	178	13	3	59.70

Two fuzzy reasoning methods, frequently employed in the fuzzy classification domain, were used, the classic and the general fuzzy reasoning methods. The classic fuzzy reasoning method uses the class of the rule with highest compatibility with an input pattern to classify it. The general fuzzy reasoning method, on the other hand, calculates the sum of compatibility degrees for each class and uses the class with highest sum to classify an input pattern. The accuracy was measured in terms of correct classification rates.

Table 2 presents the results for the experiments that were carried out using the classic fuzzy reasoning method. **Sets** describes the number of fuzzy sets used for each attribute in the experiments; **EU** represents the equalized universe method, which uses the same width for each fuzzy set, **F** represents the method that uses same frequency, and **1-R** represents the 1-R adapted method, which takes the classes into consideration when distributing the fuzzy sets.

Table 2: Classification rates - classic fuzzy reasoning method.

Sets	2			3			4		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.96	0.53	0.64	0.96	0.59	0.89	0.95	0.71	0.84
Bupa	0.49	0.46	0.43	0.52	0.51	0.55	0.55	0.53	0.57
Credit	0.54	0.58	0.60	0.54	0.58	0.73	0.53	0.55	0.63
Dia.	0.99	1.00	0.98	0.99	0.99	1.00	1.00	0.98	0.97
Glass	0.41	0.58	0.48	0.49	0.55	0.67	0.55	0.60	0.65
Heart	0.98	0.98	0.93	1.00	1.00	0.97	1.00	0.99	0.98
Iris	0.67	0.43	0.37	0.65	0.91	0.65	0.92	0.93	0.91
Seg.	0.47	0.14	0.26	0.74	0.15	0.54	0.71	0.15	0.50
Vehi.	0.99	0.95	1.00	1.00	0.98	1.00	1.00	0.97	1.00
Wine	0.94	0.99	0.99	0.99	0.98	0.97	1.00	1.00	0.98
Sets	**5**			**6**			**7**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.92	0.69	0.76	0.89	0.66	0.71	0.91	0.65	0.70
Bupa	0.56	0.52	0.54	0.55	0.48	0.52	0.55	0.46	0.48
Credit	0.53	0.55	0.58	0.53	0.55	0.56	0.54	0.55	0.56
Dia.	1.00	0.94	0.98	1.00	1.00	0.97	0.99	1.00	0.97
Glass	0.61	0.56	0.61	0.61	0.53	0.57	0.63	0.52	0.55
Heart	1.00	0.98	1.00	1.00	0.97	1.00	0.93	0.95	0.99
Iris	0.91	0.90	0.91	0.87	0.83	0.87	0.93	0.80	0.87
Seg.	0.67	0.15	0.39	0.66	0.14	0.30	0.62	0.15	0.23
Vehi.	0.98	0.95	1.00	0.98	1.00	0.98	0.97	1.00	0.99
Wine	1.00	0.99	0.99	0.99	0.98	0.99	0.99	0.93	1.00
Sets	**8**			**9**			**10**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.90	0.63	0.69	0.74	0.63	0.69	0.80	0.61	0.66
Bupa	0.61	0.45	0.42	0.58	0.43	0.42	0.55	0.43	0.45
Credit	0.54	0.55	0.55	0.60	0.55	0.55	0.57	0.55	0.55
Dia.	0.97	0.99	0.95	0.93	0.93	0.89	0.88	0.91	0.89
Glass	0.58	0.51	0.54	0.60	0.50	0.51	0.61	0.50	0.50
Heart	0.92	0.98	0.99	0.91	0.89	0.98	0.91	0.87	0.97
Iris	0.89	0.71	0.76	0.89	0.67	0.72	0.89	0.59	0.68
Seg.	0.54	0.14	0.21	0.53	0.15	0.18	0.45	0.14	0.16
Vehi.	0.95	0.95	0.99	0.94	0.97	0.97	0.93	0.95	0.98
Wine	0.93	0.88	0.95	0.81	0.85	0.93	0.73	0.81	0.91

Table 3 presents the results for the experiments that were carried out using the general fuzzy reasoning method. **Sets** describes the number of fuzzy sets used for each attribute; **EU** represents the equalized universe method, **F** represents the method that uses same frequency, and **1-R** represents the 1-R adapted method.

Table 4 shows the best estimated number of fuzzy sets, according to the proposed method, for the classic and general fuzzy reasoning methods, using the equalized universe method (**EU**), frequency method (**F**) and 1-R adapted method (**1-R**). The suffixes **C** and **G** were used to describe the classic and general fuzzy reasoning methods, respectively. A tolerance interval of 0.02 was used. Notice that the best classification rates used to define these values are light-gray shaded in Tables 2 and 3. The proposed method was able to estimate a small number of fuzzy sets for most of the datasets.

Figures 4 and 5 present the classification rates (vertical axis) for the classic and general fuzzy reasoning methods, respectively, for each of the 10 datasets, using from 2 to 10 fuzzy sets (horizontal axis) to define each of the attributes.

To test whether there is a significant difference among the methods, the Friedman test [24] was used with the null-hypothesis that the performance of the three methods, assessed in terms of the error rate, are comparable. As the null-hypothesis was rejected with a 95% confidence level, the Bonferroni-Dunn post-hoc test (to detect whether the differences among the methods are significant) was used [24]. Results showed that the equalized universe method is signifi-

ISBN: 978-989-95079-6-8

Table 3: Classification rates - general fuzzy reasoning method.

Sets	2			3			4		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.93	0.51	0.83	0.96	0.55	0.89	0.95	0.69	0.82
Bupa	0.57	0.55	0.44	0.53	0.59	0.57	0.55	0.61	0.57
Credit	0.49	0.48	0.53	0.49	0.48	0.65	0.48	0.47	0.51
Dia.	0.99	0.15	0.99	0.99	0.32	0.95	0.97	0.24	0.81
Glass	0.44	0.33	0.51	0.49	0.36	0.61	0.57	0.37	0.59
Heart	0.09	0.19	0.13	0.09	0.15	0.34	0.08	0.08	0.11
Iris	0.69	0.50	0.38	0.70	0.93	0.63	0.97	0.92	0.93
Seg.	0.41	0.14	0.20	0.70	0.15	0.56	0.71	0.16	0.51
Vehi.	0.98	0.44	1.00	0.98	0.44	0.95	0.97	0.28	0.65
Wine	0.93	0.52	0.99	0.89	0.22	0.81	0.73	0.05	0.36
Sets	**5**			**6**			**7**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.92	0.67	0.73	0.89	0.66	0.71	0.91	0.65	0.69
Bupa	0.59	0.63	0.61	0.57	0.58	0.62	0.57	0.57	0.63
Credit	0.48	0.46	0.47	0.48	0.46	0.46	0.49	0.45	0.46
Dia.	0.93	0.11	0.51	0.88	0.06	0.25	0.80	0.02	0.11
Glass	0.54	0.27	0.50	0.56	0.25	0.39	0.55	0.17	0.29
Heart	0.06	0.02	0.04	0.05	0.01	0.04	0.04	0.01	0.02
Iris	0.94	0.90	0.91	0.89	0.80	0.84	0.94	0.77	0.83
Seg.	0.70	0.16	0.42	0.70	0.16	0.32	0.63	0.15	0.24
Vehi.	0.93	0.11	0.30	0.79	0.05	0.17	0.69	0.02	0.06
Wine	0.39	0.03	0.12	0.18	0.00	0.02	0.07	0.02	0.00
Sets	**8**			**9**			**10**		
Meth.	EU	F	1-R	EU	F	1-R	EU	F	1-R
Bre.	0.90	0.63	0.69	0.74	0.63	0.69	0.80	0.61	0.66
Bupa	0.60	0.60	0.61	0.58	0.59	0.61	0.61	0.60	0.59
Credit	0.48	0.45	0.45	0.57	0.45	0.45	0.50	0.45	0.45
Dia.	0.69	0.01	0.06	0.58	0.00	0.02	0.47	0.01	0.01
Glass	0.51	0.15	0.21	0.53	0.10	0.14	0.52	0.07	0.11
Heart	0.03	0.00	0.01	0.03	0.01	0.00	0.02	0.00	0.00
Iris	0.91	0.73	0.77	0.94	0.69	0.73	0.91	0.57	0.61
Seg.	0.60	0.14	0.21	0.59	0.15	0.18	0.51	0.14	0.17
Vehi.	0.41	0.00	0.02	0.31	0.01	0.01	0.17	0.01	0.01
Wine	0.02	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.01

Table 4: Best estimated number of fuzzy sets for each dataset.

Method	EU - C	EU - G	F - C	F - G	1-R - C	1-R - G
Breast	2	3	4	4	3	3
Bupa	8	5	3	4	3	5
Credit	9	9	2	2	3	3
Diabetes	2	2	2	3	2	2
Glass	5	4	2	3	3	3
Heart	2	2	2	2	4	3
Iris	4	4	3	3	4	4
Seg.	3	3	2	2	3	3
Vehicle	2	2	3	2	2	2
Wine	3	2	2	2	2	2

cantly better than the 1-R method and the frequency method, with a 95% confidence level. The results of the statistical tests also showed that the 1-R method performed significantly better than the frequency method with a 95% confidence level.

Table 5 shows the average ranks for each of the methods tested for the classic and general fuzzy reasoning methods.

Table 5: Average ranks.

Methods	EU	Frequency	1-R
Classic	1.694	2.411	1.894
General	1.350	2.711	1.939

One of the reasons for the good performance of the equalized universe method may be the fact that it generates fuzzy sets with the same distance amongst them, promoting the creation of rules that will cover equal areas of the attributes partitions, rather than promoting the creation of fuzzy sets in areas with a concentration of examples but with less differences regarding the classes, which could have a deeper effect on the classification rates. The 1-R method, since it uses the classes to adjust the fuzzy sets defining each attribute, should perform

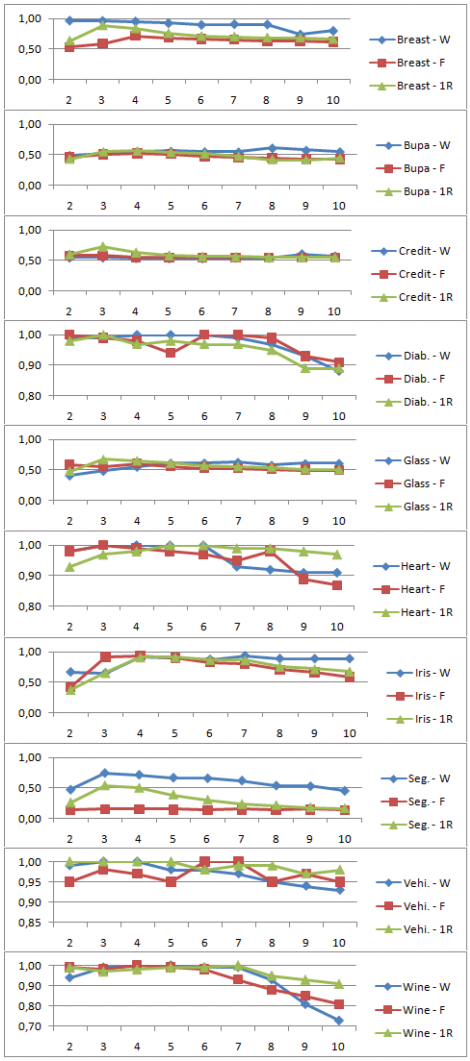


Figure 4: Results for the classic fuzzy reasoning method.

better than the frequency method.

6 Conclusions

In this paper we have tested and compared three different methods for the definition of fuzzy sets, the equalized universe method, the frequency method, and an adaptation of the 1-R supervised method. We have also proposed a heuristic method to estimate the number of fuzzy sets for each attribute using the Wang & Mendel method, to be used by other more costly approaches, able of finding smaller sets of rules with better accuracy. The experiments were carried out with 10 datasets for classification problems using 10-fold cross validation and the classic and general fuzzy reasoning methods. Each attribute of each dataset was defined using from 2 to 10 fuzzy triangular shaped sets. Statistical methods were carried out in order to find significant differences among the methods. The results show that the equalized universe method performed significantly better than the frequency and 1-R methods. The results also showed that the 1-R method performed better than the frequency method. Regarding the estimation of the number of fuzzy sets for the attributes, the proposed method provides support for the researcher to select small intervals of numbers of fuzzy sets that would be worthwhile using in more costly

ISBN: 978-989-95079-6-8

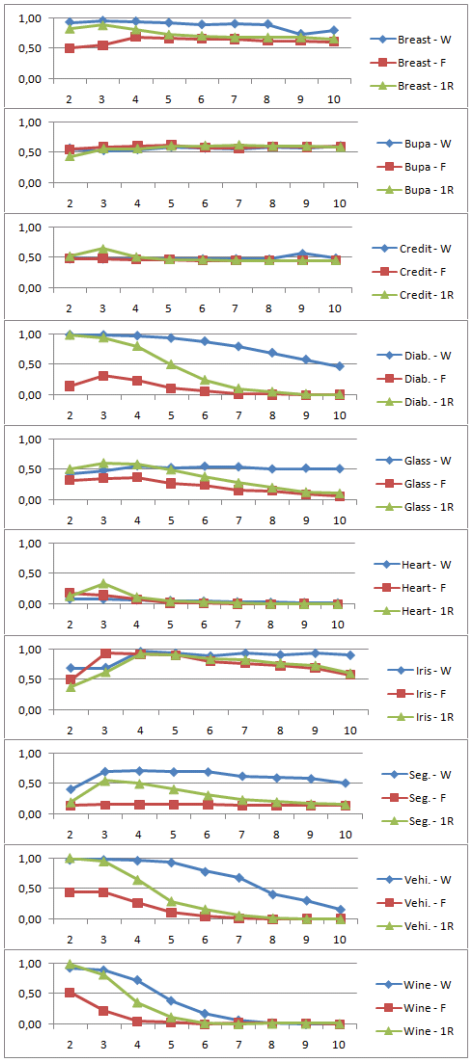


Figure 5: Results for the general fuzzy reasoning method.

approaches, based on the performance of the generated rule bases.

Acknowledgments

This research was supported by CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, and by the São Paulo State Research Foundation - FAPESP (Proc. No. 2007/05390-0).

References

[1] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] Rafael Alcalá, Jesús Alcalá-Fdez, M.J. Gacto, and Francisco Herrera. Genetic learning of membership functions for mining fuzzy association rules. *IEEE International Conference on Fuzzy Systems*, 1:1538–1543, 2007.

[3] Swarup Medasani, Jaeseok Kim, and Raghu Krishnapuram. An overview of membership function generation techniques for pattern recognition. *International Journal of Approxiamate Reasoning*, 19:391–417, 1998.

[4] M. Kaya and R. Alhajj. A clustering algorithm with genetically optimized membership functions for fuzzy association rules mining. *12th IEEE International Conference on Fuzzy Systems*, 2:881–886, 2003.

[5] Ling Zhang and Bo Zhang. The structure analysis of fuzzy sets. *Int. Journal of Approx. Reasoning*, 40:92–108, 2005.

[6] M. S. Chen and S. W. Wang. Fuzzy clustering analysis for optimizing membership functions. *Fuzzy Sets and Systems*, 103:239–254, 1999.

[7] M. E. Cintra, H. A. Camargo, and M. C. Monard. A study on techniques for the automatic generation of membership functions for pattern recognition. *C3N - Congresso da Academia Trinacional de Ciências*, 1:1–10, 2008.

[8] T. Warren Liao, Aivars K. Celmins, and Robert J. Hammell. A fuzzy c-means variant for the generation of fuzzy term sets. *Fuzzy Sets and Systems*, 135:241–257, 2003.

[9] Xuanli Lisa Xie and Gerardo Beni. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):841–847, 1991.

[10] J.S. Jang. Adaptive-network-based fuzzy inference systems. *IEEE Trans. on SMC*, 23:665–685, 1993.

[11] Tzung-Pei Hong, Chun-Hao Chen, Yu-Lung Wu, and Yeong-Chyi Lee. A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions. *Soft Computing*, 10(11):1091–1101, 2006.

[12] Ajith Abraham. Neuro fuzzy systems: State-of-the-art modeling techniques. *Lec. Notes in CS*, 2084:269–276, 2001.

[13] Detlef Nauck, Frank Klawonn, and Rudolf Kruse. *Foundations of Neuro-Fuzzy Systems*. John Wiley & Sons, Inc., 1997.

[14] H. D. Cheng and Jim-Rong Chen. Automatically determine the membership function based on the maximum entropy principle. *Information Sciences*, 96(3-4):163–182, 1997.

[15] Hisao Ishibuchi, Takashi Yamamoto, and Tomoharu Nakashima. Fuzzy data mining: Effect of fuzzy discretization. In *ICDM '01: Proc. of the 2001 IEEE Int. Conf. on Data Mining*, pages 241–248, USA, 2001. IEEE Computer Society.

[16] Grzegorz Nieradka and Bohdan Butkiewicz. A method for automatic membership function estimation based on fuzzy measures. *International Fuzzy Systems Association World Congress - IFSA 2007*, 4529:451–460, 2007.

[17] Weibei Dou, Yuan Ren, Qian Wu, Su Ruan, Yanping Chen, Daniel Bloyet, and Jean-Marc Constans. Fuzzy kappa for the agreement measure of fuzzy classifications. *Neurocomputing*, 70(4-6):726–734, 2007.

[18] Hisao Ishibuchi, Ken Nozaki, and Hideo Tanaka. Efficient fuzzy partition of pattern space for classification problems. *Fuzzy Sets and Systems*, 59(3):295–304, 1993.

[19] J.M. Keller, M.R. Gray, and J.A. Given Jr. A fuzzy k-nearest neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(4):580–585, 1985.

[20] Enrique H. Ruspini. A new approach to clustering. *Information and Control*, 15(1):22–32, 1969.

[21] Anthony J. Jakeman Jun and Michael McAleer Bai. *Effects of Misspecification in Estimating the Percentiles of Some Two- & Three-Parameter Distributions*. ANU, Australia, 1989.

[22] Robert C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11(1):63–90, 1993.

[23] L. Wang. The WM method completed: a flexible fuzzy system approach to data mining. *IEEE International Conference on Fuzzy Systems*, 11:768–782, 2003.

[24] Janez Demšar. Statistical comparison of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7(1):1–30, 2006.

ISBN: 978-989-95079-6-8

Discriminatory Components for Pattern Classification

Nick J. Pizzi[1,2,3] Witold Pedrycz[4]

1 Institute for Bidiagnostics, National Research Council
Winnipeg, Canada
2 Department of Computer Science, University of Manitoba
Winnipeg, Canada
3 Department of Applied Computer Science, University of Winnipeg
Winnipeg, Canada
4 Electrical and Computer Engineering, University of Alberta
Edmonton, Canada
Email: pizzi@cs.umanitoba.ca, pedrycz@ece.ualberta.ca

Abstract—*The pattern recognition literature is replete with the use of principal component analysis in the interpretation and analysis of data. However, in the specific case of classification, especially of biomedical patterns, this pre-processing method, which transforms possibly correlated features into a new set of uncorrelated variables, must be used with caution since a principal component, which may account for significant variance in the data, is not necessarily discriminatory. To compensate for this deficiency, we present a novel classification method using an adaptive network of fuzzy logic connectives to select the most discriminatory principal components. We empirically demonstrate the effectiveness of this method using a benchmark combination of a conventional classifier and principal component analysis.*

Keywords— fuzzy logic network; principal component analysis; biomedical data; pattern classification.

1 Introduction

Today's biomedical instrumentation provides the acquisition of complex data rich in information content; however, its analysis and interpretation is often difficult and requires the latest pattern analysis methodologies [1,2,3]. This is particularly true for the domain of biomedical pattern classification, the prediction by a classifier of the class (for example, normal or abnormal) to which a pattern (for instance, an infrared spectrum of a biofluid) belongs. This prediction is validated against a gold standard, an external reference test such as a pathologist's expert assessment of the same biofluid. Although many classifier methods exist [4,5,6], the most successful approaches combine them with well designed pre-processing techniques, which simplify, in some sense, the feature space prior to presentation to a classifier, and a sound validation protocol to ensure realistic and clinically useful results.

A standard approach to biomedical pattern classification is to use simple linear classifiers coupled with pre-processing transformations that create new features (coordinates, parameters) ordered by the cumulative variance of the original features. A common pairing is linear discriminant analysis coupled with principal component analysis. The ordering allows for the reduction of the feature space by using only the first few components that account for the bulk of the data variance. While this combination is often successful when used in the classification of patterns, its success lies more in the feature reduction aspect rather than the exploitation of the data variance. This is because the feature variance is not concomitant with the discriminatory power of individual features.

We present a new classification method that uses an adaptive network of fuzzy logic connectives that operate on the entire set of principal components. This method exploits the discriminatory power of any principal component regardless of the amount of variance for which it accounts. We empirically demonstrate that this method produces superior classification performance compared to a benchmark using the conventional approach of linear discriminant analysis operating on the first few principal components of the original feature space. Section 2 presents a general discussion on classification including classifier validation, principal component analysis, linear discriminant analysis, and a recent fuzzy adaptive logic network on which the current method is based. Details of our novel approach are presented in Section 3. The synthetic and biomedical datasets, experiment design, and results are discussed in Section 4 followed by some concluding remarks and areas of future investigations.

2 Pattern classification

2.1 Classifier Validation

We begin by defining formal pattern classification notation: N is the number of patterns (samples, vectors, individuals, or cases); n is the number of features (dimensions, attributes, coordinates, or measurements); c is the number of classes (groups); and $\boldsymbol{X}=\{(\boldsymbol{x}_k,\omega_k),\ k=1,2,\ldots,N\}$ is a set of N labeled patterns where $\boldsymbol{x}_k\in\Re^n$ and $\omega_k\in\Omega=\{1,2,\ldots,c\}$. A classifier is a system that determines a mapping $f:\boldsymbol{X}\rightarrow\Omega$. If a classifier predicts that the class label for pattern $\boldsymbol{x}_i$ is ω_p then: a correct classification occurs if $\omega_p=\omega_i$; otherwise it is considered a misclassification.

Many investigations involving data classification are biased as they use the entire dataset to determine the mapping. This approach leads to unrealistic classification results that do not take into account the possibility of overfitting, wherein the

ISBN: 978-989-95079-6-8

mapping becomes a simple table lookup, between the patterns and class labels (that is, it possesses no generalized predictive power for new patterns). To compensate for this bias, it is essential to perform some type of validation. For instance, patterns in $\boldsymbol{X}$ may be randomly allocated to a design subset, $\boldsymbol{X}^D$ containing N^D patterns, or a validation subset, $\boldsymbol{X}^V$ containing N^V patterns ($N^D+N^V=N$). Now, a mapping is determined using only design patterns, $f':\boldsymbol{X}^D\rightarrow\Omega$, but the classification performance is measured using f' with the validation patterns.

Performance of a classification system is measured using the $c\times c$ confusion matrix, $\boldsymbol{C}$, of the desired class labels (as determined by an external reference test or "gold standard") versus the predicted class labels (as generated by the classifier). If the class prediction for $\boldsymbol{x}_i$ is ω_p then element $[\omega_p,\omega_i]$ of $\boldsymbol{C}$ is incremented by one. The conventional performance measure is the ratio of correctly classified patterns to the total number of patterns, $P_O=(\sum_i n_{ii})/N^V$ (i=1, 2,…, c), where n_{ij} is the number of class i validation patterns that are predicted to belong to class j.

A significant problem with P_O occurs when there is a large disparity between class sizes. For instance, if a validation set has many more patterns in one class than another, the classification accuracy relating to the former will outweigh any effects from the latter. For instance, a high value for P_O may occur due to few misclassifications with the larger class even though many misclassifications occur with the smaller class. This is especially problematic for biomedical patterns where it is relatively easy to acquire patterns from "normal" samples (tissues, biofluids, and so on) but clinically difficult to obtain "abnormal" samples. As a result, it is often better to forgo the use of P_O for these types of datasets and use the class-wise average accuracy, P_A

$$P_A = c^{-1}\sum_i^N \frac{n_{ii}}{\sum_j^N n_{ij}} (i,j=1,\ldots,c) \qquad (1)$$

2.2 Principal Component Analysis

The motivation behind principal component analysis, first described by Pearson [7] (with a practical computing method described by Hotelling [8]), is to find a set of directions (coordinates) that explain as much of the variability of the original data as possible. In other words, the principal components are a new set of orthogonal linear coordinates such that the variances of the original features with respect to these derived coordinates are in decreasing orders of magnitude [9]. As a result, each principal component is uncorrelated with all other principal components (in a normal distribution, they are statistically independent). Moreover, it can be shown [10] that no other set of m coordinates can account for more of the variability in the original data than the first m principal components.

The first principal component, $\boldsymbol{Y}_1$, of the original features x_1, x_2, …, x_n, is the linear combination

$$\boldsymbol{Y}_1 = \sum_i^n a_{1i}x_i, \sum_i^n a_{1i}^2 = 1 \qquad (2)$$

The constraint on the coefficients is necessary; otherwise the variance of $\boldsymbol{Y}_1$ can be increased simply by increasing the value of any coefficient. The second principal component, $\boldsymbol{Y}_2$, would be computed in a similar fashion to (1). Fig. 1 is a plot of some bivariate data and their two principal components. It is clear from this figure that an additional constraint, orthogonality to the first principal component, is required to compute the second principal component, otherwise it would simply be driven to the first principal component. Orthogonality is ensured by restricting the variables of the second principal component to those that are uncorrelated with the first principal component. As a result of this orthogonality constraint, if there are n features then there can be up to n principal components [11]. In fact, if the original features are completely uncorrelated, then all n principal components must be used to take into account the variance in the original features. In this case, principal component analysis serves little purpose, with respect to pattern classification, since the motivation behind the technique is to reduce the dimensionality of the original input (feature) space. However, in "real-world" high-dimensional biomedical patterns, the converse is usually true; features are highly correlated and hence only $1\leq m<<n$ principal components are required to account for all (or nearly all) of the variation.

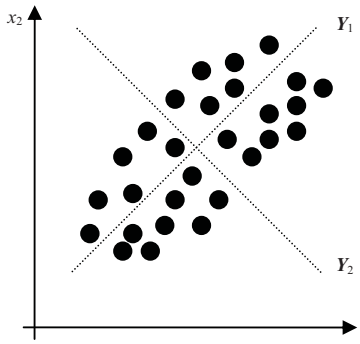


Figure 1: Principal components, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, for some bivariate data.

Determining the principal components is a straightforward process involving the computation of the eigensystem of the original data's covariance matrix, $\boldsymbol{V}$, whose element v_{lm} is the sample covariance between features x_l and x_m

$$v_{lm} = \frac{1}{N-1}\sum\nolimits_N \left(x_{il}-\mu_l\right)\left(x_{im}-\mu_m\right) \qquad (3)$$

where μ_j is the mean for feature x_j (cf. [12] for a derivation of the proof). The variances of the principal components are the eigenvalues of $\boldsymbol{V}$, $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n\geq 0$ (the covariance matrix is quadratic and hence admits no negative eigenvalues). The variance of a principal component, $\boldsymbol{Y}_i$, is λ_i and its constants a_{i1}, a_{i2}, …, a_{in} are the elements of the corresponding eigenvector.

A standard strategy employed in biomedical pattern classification is to take the first m principal components whose cumulative variance exceeds some pre-defined threshold. This reduction is often significant ($m<<n$): for instance in some high-dimensional infrared spectra more than 80% of the cumulative variance may be accounted for by only the first one or two principal components [13].

ISBN: 978-989-95079-6-8

While this pre-processing strategy may be effective, it must be noted that the components are ordered by maximal variance. Unfortunately, this does not necessarily translate into maximal discriminatory power [14]. For instance, if the method used to acquire values for a particular feature is extremely prone to measurement error, then this feature will have a high variance. Now, assuming this variance is greater than other features, the first principal component will be approximately equal to this suspect feature, and hence, this principal component will be useless in discriminating between classes. Conversely, a highly discriminatory feature may have an extremely small variance and hence will not contribute to the first few principal components. In summary, maximal discriminatory power is not equivalent to maximal variance.

2.3 Linear Discriminant Analysis

Linear discriminant analysis (LDA) [15] is a standard pattern classification strategy that determines linear boundaries between c classes while taking into account between-class and within-class variances. If the error distributions for the classes are the same (identical covariance matrices), LDA constructs the optimal linear boundary between the classes. In real-world situations, this optimality is seldom achieved since different classes typically give rise to different distributions.

LDA assigns a pattern, $\boldsymbol{x}$, to class i for which the probability distribution, $p_i(\boldsymbol{x})$, is greatest. That is, $\boldsymbol{x}$ is allocated to class i, if $q_i p_i(\boldsymbol{x}) \geq q_j p_j(\boldsymbol{x})$ $(\forall j=1,2,\ldots,c\ [j\neq i])$, where q_i is the class' prior (or proportional) probability. The discriminant function for class i is

$$D_i(\boldsymbol{x}) = \log q_i + \boldsymbol{\mu}_i^{\mathrm{T}} \boldsymbol{W}^{-1} \left(\boldsymbol{x} - \tfrac{1}{2} \boldsymbol{\mu}_i \right) \qquad (4)$$

where $\boldsymbol{\mu}_i$ is the mean for class i and $\boldsymbol{W}$ is the covariance matrix of the patterns in $\boldsymbol{X}$. The feature space hyperplane separating class i from j is defined by $F_{ij}(\boldsymbol{x})=D_i(\boldsymbol{x})-D_j(\boldsymbol{x})=0$. As a classification system, LDA is often used with the first m principal components of a dataset rather than the original features. This combination will serve as the classification benchmark against which our novel pattern classification method will be evaluated.

2.4 Fuzzy Adaptive Logic Network

Our novel approach is partly based upon a modification to the fuzzy adaptive logic network (cf. [16] for a thorough description). This network, which can be used for pattern classification, combines two different subsystems within a general architecture. A neurocomputing subsystem uses a set of perceptrons to construct class boundaries. Through a set of weights and respective inputs, a perceptron is defined as $P(\boldsymbol{x},\boldsymbol{w})=f(\sum_i w_i x_i + w_0)$ (f is a sigmoidal transfer function) describing an n-dimensional hyperplane. This geometric information is presented to the logic processing subsystem composed of a layer of fuzzy conjunctions ("and" elements) and disjunctions ("or" elements). The intent is to use these fuzzy logic connectives to combine the hyperplanes from the neurocomputing subsystem to form convex hull-like topologies. For instance, a convex region delineated by p perceptrons may be represented by the compound logic predicate, $Q = P_1(\boldsymbol{x},\boldsymbol{w}_1)$ *and* $P_2(\boldsymbol{x},\boldsymbol{w}_2)$ *and* … *and* $P_p(\boldsymbol{x},\boldsymbol{w}_p)$, which produces values close to one (in other words, approaching the crisp notion of *true*) when all contributing predicates are *true* (that is, the respective perceptrons produce high outputs).

To capture the geometric notion of disjoint regions one may take a union (in the fuzzy set theoretic sense) of the individual regions described by the Q's, $R = Q_1$ *or* Q_2 *or* … *or* Q_q. To implement these fuzzy predicates, one uses t-norms to model the *and* logic connectives and s-norms to model the *or* logic connectives. A t-norm, $\wedge$, is a function $[0,1]^2 \rightarrow [0,1]$ that is commutative, symmetric, monotonic, and satisfies the boundary conditions $x\mathrm{t}0=0$ and $x\mathrm{t}1=x$, while the boundary conditions for the s-norm, $\vee$, are $x\mathrm{s}0=x$ and $x\mathrm{s}1=1$. The fuzzy *or* and *and* connectives may now be defined as

$$\begin{aligned} OR(\boldsymbol{x};\boldsymbol{w}) &= \wedge_i \left(w_i \vee x_i \right) \\ AND(\boldsymbol{x};\boldsymbol{w}) &= \vee_i \left(w_i \wedge x_i \right) \end{aligned} \qquad (5)$$

where $\boldsymbol{x}$ are the inputs and $\boldsymbol{w}$ are the corresponding adjustable weights (connections) confined to the unit interval. In the case of $OR(\boldsymbol{x};\boldsymbol{w})$, the greater the weight value the more relevant the respective input (if all weights are 1, it becomes a standard *or* gate). In the case of $AND(\boldsymbol{x};\boldsymbol{w})$, the greater the weight value the less relevant the respective input (if all weights are 0, it becomes a standard *and* gate). If we restrict ourselves to differentiable t- and s-norms, a gradient descent strategy can be used to train a fuzzy adaptive logic network (cf. [16] for details).

3 Fuzzy logic classification network using principal components

Building upon the concepts described in Section 2, we now describe a novel pattern classification algorithm, COAP (Classification using a fuzzy Or/And network with Principal components), a component of which extends the fuzzy logic network architecture developed by the authors [17]. There are three major steps to the COAP algorithm: (i) apply principal component analysis on the original features to find all n principal components; (ii) Use a genetic algorithm to determine the optimal weights for the fuzzy logic network given the principal component values found in (i); (iii) use the patterns from the validation subset to assess the classification performance using the selected feature regions and principal component values. Fig. 2 illustrates the architecture of the COAP system.

Let us now look at each algorithmic step in more detail. After applying principal component analysis to the feature space, we replace pattern, $\boldsymbol{x}_i=[x_1,x_2,\ldots,x_n]$, with its respective principal component values, $\boldsymbol{y}_i=[\boldsymbol{Y}_1(\boldsymbol{x}_i),\boldsymbol{Y}_2(\boldsymbol{x}_i),\ldots,\boldsymbol{Y}_n(\boldsymbol{x}_i)]$. These principal component values for all design set patterns are subsequently presented to the fuzzy logic network component.

COAP's fuzzy logic network uses the product $(x_1 \times x_2)$ and probabilistic sum $(x_1+x_2-x_1 \times x_2)$ for the t- and s-norms, respectively, with p (user selected) *AND* connectives and c *OR* connectives. There are two issues with this network that

ISBN: 978-989-95079-6-8

do not exist with the fuzzy adaptive logic network described in Section 2.4. First, while output from a perceptron maps onto the unit interval (due to the sigmoidal nature of its transfer function), which is necessary for input into a fuzzy logic *AND* connective, principal component values map onto $\Re$. This can be dealt with by rescaling the principal component values prior to presentation to the fuzzy logic network ((*x*–*min*)÷(*max*–*min*), where *min* and *max* are the respective minimum and maximum for all principal component values).

The second, more serious, issue is that a gradient descent strategy cannot be used to minimize the fuzzy logic network error (optimize the weights) since the weight adjustments are now based on sets of principal components rather than differentiable perceptron output. We deal with this issue by using a straightforward implementation of a genetic algorithm [18,19,20] to perform the structural optimization of the network. While much slower than gradient descent, this solution still provides adequate computational performance. In this study, we implemented a conventional genetic algorithm as described in [21], but other more sophisticated genetic algorithm variants could certainly be explored.

As mentioned in Section 2.1, $\Omega=\{1,2,\ldots,c\}$; however, it is often beneficial [13] to use 1-of-*c* encoding for the class labels for iterative classifiers such as artificial neural networks or fuzzy logic networks, namely, $\Omega=\{\gamma_1,\gamma_2,\ldots,\gamma_c\}$ where, for $\boldsymbol{x}_i$, $\gamma_{\omega_i}=1$ and $\gamma_{\omega_j}=0$ $(\forall\omega_i\neq\omega_j)$. Instead of using one output element to represent all *c* classes, we use *c* output elements.

Finally, all performance results using P_A are based on the class predictions of COAP using the biomedical patterns from the validation subset. That is, the principal component values are obtained for the validation patterns and presented to the fuzzy logic network component. Subsequently, the predictions are compared against the desired class labels (if the output element with the maximum value corresponds to the desired label's non-zero component then we have a successful classification, otherwise it is a misclassification).

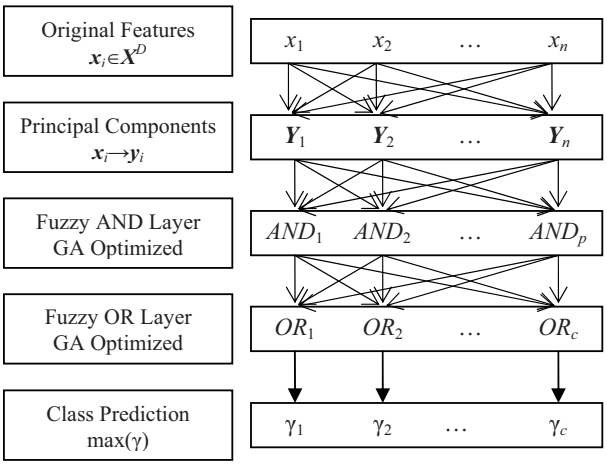


Figure 2: General architecture of COAP with the GA (genetic algorithm) optimized fuzzy logic network.

4 Experiments and discussion

4.1 Synthetic Datasets

We begin our experiments with the "exclusive or" dataset (*n*=2, *c*=2, *N*=4): patterns {{0,0},{1,1}} and {{0,1},{1,0}}; and respective principal component values of {{0,0},{1,–1}} and {{1,0},{0,–1}}. Intuitively, one expects that LDA would perform poorly in this case as no hyperplane can act as a class boundary to perfectly separate the patterns. Using LDA with the principal components, this is actually the case with P_A=0.5 (one misclassification for each class). [As this is a strictly pedagogical experiment, we skip validation.] Setting the initial genetic algorithm population to 50, the number of iterations to 5, and the number of *AND* connectives to 2, we now get perfect accuracy, P_A=1.0. The weights for the two *AND* connectives are {0.95,0.05} and {0.0,1.0}. The weights for the two *OR* connectives are {0.53,0.05} and {0.07,0.35}.

This next dataset is a variant of the exclusive or dataset described above (*n*=10, *c*=2, *N*=400, $\boldsymbol{x}\in[0,1]^n$). A pattern belongs to the first class, if all of its features are identical; otherwise it belongs to the second class. Fig. 3 is a plot of the first two features of this dataset. The initial genetic algorithm population is 100, the number of iterations is 5, and the number of *AND* connectives is 10. In this case, LDA with principal components also performed poorly, P_A=0.5, while COAP produced a higher classification accuracy, P_A=0.81.

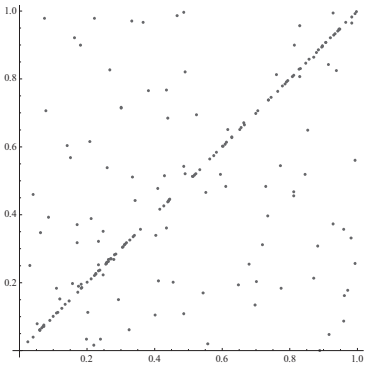


Figure 3: Plot of, x_1 and x_2, for the second synthetic dataset.

4.2 Biomedical Dataset

We used a biomedical dataset from the Machine Learning Repository at the University of California, Irvine [22]. This archive is used to evaluate machine learning algorithms. This dataset [23,24] comprises voice measurements acquired from two classes of subjects: those with Parkinson's disease (abnormal); and those without (normal). There are *N*=195 patterns from one of two classes: N_a=48 abnormal patterns; N_n=147 normal patterns. The *n*=22 features include: the maximum, minimum, and average fundamental vocal frequency (3 features); several measures of variation in fundamental frequency (5 features); several measures of variation in amplitude (6 features); measures of the ratio of noise to tonal components in the voice (2 features); nonlinear dynamical complexity measures (2 features); a

ISBN: 978-989-95079-6-8

signal fractal scaling exponent (1 feature); and nonlinear measures of fundamental frequency variation (3 features).

For this dataset, the following COAP parameters were used: p=7 (the number of *AND* connectives); crossover rate, 0.07; mutation rate, 0.005; size of genetic algorithm population, 200; and 20 iterations of the genetic algorithm. The patterns were randomly assigned to either a design subset (N^D=64 with 32 normal patterns and 32 abnormal patterns) or a validation subset (N^V=131 with 115 normal patterns and 16 abnormal patterns). Finally, given the class size disparity between the abnormal and normal patterns within the validation set, P_A was used to assess classification accuracy.

Table 1 lists the confusion matrices for COAP for the design patterns and validation patterns. For the design patterns, P_A=0.91, while P_A=0.85 for the validation patterns. While only 70% of the normal validation patterns were correctly classified, all abnormal validation patterns were correctly classified. Table 2 lists the confusion matrices for LDA using all n principal components. In the case of this benchmark, P_A=0.94 for the design patterns and P_A=0.70 for the validation patterns. The disparity between the design and validation results is a classic sign of potential overfitting. While 75% of the abnormal validation patterns were correctly classified, only 65% of the normal patterns were correctly classified.

Table 1: COAP confusion matrices for $\boldsymbol{X}^D$ and $\boldsymbol{X}^V$.

Desired vs	Design Set (P_A=0.91)		Validation Set (P_A=0.85)	
Predicted	Abnormal	Normal	Abnormal	Normal
Abnormal	29	3	16	0
Normal	3	29	35	80

Table 2: Benchmark confusion matrices for $\boldsymbol{X}^D$ and $\boldsymbol{X}^V$.

Desired vs	Design Set (P_A=0.94)		Validation Set (P_A=0.70)	
Predicted	Abnormal	Normal	Abnormal	Normal
Abnormal	31	1	12	4
Normal	3	29	40	75

Table 3 lists the confusion matrices (desired versus predicted class labels) using successive combinations of principal component in turn (ordered by variance) for both the design and validation sets. Note that: P_A is listed for the design set followed by the validation set; $\sum^{\alpha}\lambda_i$ refers to the cumulative variance of the first α principal components used to produce the corresponding results; and the last entry is the same as that described in Table 2. The best result, P_A=0.781, occurred when using the first 13 principal components. It is clear with this dataset that feature variance is not well correlated with discriminatory features (for example, the first three principal components, which account for 0.998 of the cumulative variance, produce a low accuracy score, P_A=0.60).

Fig. 3 summarizes these results by plotting the COAP validation result against the validation results using all successive combinations of principal components. It should also be noted that the last few entries in Table 3 clearly demonstrate a problem with overfitting with a disparity of approximately 0.24 between the design and validation results.

Table 3: Results using successive principal components.

	D vs P	$\boldsymbol{X}^D$		$\boldsymbol{X}^V$	
		A	N	A	N
$\sum^{1}\lambda_i$=0.729	A	24	8	8	8
P_A=0.77/0.63	N	7	25	27	88
$\sum^{2}\lambda_i$=0.947	A	22	10	7	9
P_A=0.69/0.61	N	10	22	25	90
$\sum^{3}\lambda_i$=0.998	A	23	9	7	9
P_A=0.70/0.60	N	10	22	27	88
$\sum^{4}\lambda_i$=1.000	A	22	10	9	7
P_A=0.73/0.62	N	7	25	36	79
$\sum^{5}\lambda_i$=1.000	A	28	4	10	6
P_A=0.83/0.69	N	7	25	29	86
$\sum^{6}\lambda_i$=1.000	A	28	4	10	6
P_A=0.83/0.69	N	7	25	29	86
$\sum^{7}\lambda_i$=1.000	A	28	4	10	6
P_A=0.84/0.69	N	7	25	29	86
$\sum^{8}\lambda_i$=1.000	A	28	4	11	5
P_A=0.83/0.72	N	7	25	32	83
$\sum^{9}\lambda_i$=1.000	A	27	5	11	5
P_A=0.81/0.70	N	7	25	32	83
$\sum^{10}\lambda_i$=1.000	A	27	5	12	4
P_A=0.81/0.74	N	7	25	32	83
$\sum^{11}\lambda_i$=1.000	A	29	3	13	3
P_A=0.92/0.74	N	2	30	38	77
$\sum^{12}\lambda_i$=1.000	A	29	3	13	3
P_A=0.92/0.75	N	2	30	37	78
$\sum^{13}\lambda$=1.000	A	30	2	14	2
P_A=0.92/0.78	N	3	29	36	79
$\sum^{14}\lambda$=1.000	A	30	2	14	2
P_A=0.92/0.78	N	3	29	37	78
$\sum^{15}\lambda_i$=1.000	A	30	2	14	2
P_A=0.92/0.78	N	3	29	37	78
$\sum^{16}\lambda_i$=1.000	A	32	0	12	4
P_A=0.94/0.71	N	4	28	39	76
$\sum^{17}\lambda_i$=1.000	A	32	0	14	2
P_A=0.95/0.76	N	3	29	41	74
$\sum^{18}\lambda_i$=1.000	A	31	1	12	4
P_A=0.94/0.70	N	3	29	40	75
$\sum^{19}\lambda_i$=1.000	A	31	1	12	4
P_A=0.94/0.70	N	3	29	40	75
$\sum^{20}\lambda_i$=1.000	A	31	1	12	4
P_A=0.94/0.70	N	3	29	40	75
$\sum^{21}\lambda_i$=1.000	A	31	1	12	4
P_A=0.94/0.70	N	3	29	40	75
$\sum^{22}\lambda_i$=1.000	A	31	1	12	4
P_A=0.94/0.70	N	3	29	40	75

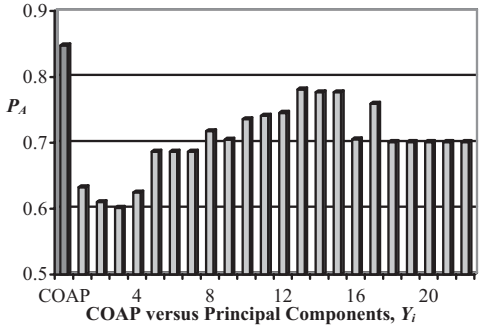


Figure 3: Summary of validation set accuracies.

ISBN: 978-989-95079-6-8

Finally, after examination of the weights of the fuzzy logic network, six principal components tended to significantly contribute to the COAP training, $\boldsymbol{Y}_1$, $\boldsymbol{Y}_4$, $\boldsymbol{Y}_5$, $\boldsymbol{Y}_{10}$, $\boldsymbol{Y}_{12}$, and $\boldsymbol{Y}_{16}$. This is further evidence that care must be taken when using principal component analysis as one may overlook discriminatory components simply because they account for little feature variance (for example, the variance for $\boldsymbol{Y}_{10}$, $\boldsymbol{Y}_{12}$, and $\boldsymbol{Y}_{16}$ is zero).

5 Conclusions

We have empirically demonstrated the effectiveness of a novel classification method that uses an adaptive network of fuzzy logic connectives to combine new features generated using principal component analysis. Using a "real world" biomedical dataset, COAP correctly classified significantly more patterns from a validation set compared to the benchmark.

While this novel classification method has demonstrated the utility of merging fuzzy logic connectives with multivariate statistical discrimination, the investigation has also led to the identification of future areas of research to potentially improve its overall effectiveness and computational performance. First, rather than setting the number of fuzzy *and* connectives by the user *a priori*, it would be worthwhile to investigate a cascade approach to determining an optimal number of *and* connections that would be completely data-driven. Second, alternative structural optimizations to the fuzzy logic network need to be examined beginning with more sophisticated evolutionary computational approaches or exploiting recent advances in stochastic optimization techniques. A final area of investigation is using more sophisticated component analysis methods for feature pre-processing (non-linear and local principal component methods, other kernel based approaches, and fuzzy set based feature selection/extraction approaches).

Acknowledgment

We are grateful to Max Little of the University of Oxford who, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, recorded the speech signals for the contribution of the biomedical dataset within the publicly available machine learning repository. We thank Drs. Asuncion and Newman of the University of California, Irvine who maintain the highly useful and publicly available data repository.

Conrad Wiebe and Aleksander Demko are gratefully acknowledged for the implementation and testing of some of the software framework components used in the implementation of this pattern classification method and validation protocol.

This investigation was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] D.L. Pavia, G.M. Lampman, G.S. Kriz and J.A. Vyvyan. *Introduction to Spectroscopy*. Fort Worth: Harcourt Brace College Publishers, 2008.

[2] H. Friebolin. *Basic One- and Two-Dimensional NMR Spectroscopy*. New York: Wiley and Sons, 1998.

[3] M. Akay. *Nonlinear Biomedical Signal Processing, Volume I: Fuzzy Logic, Neural Networks, and New Algorithms*. New York: IEEE Press, 2000.

[4] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. New York: Wiley Interscience, 2000.

[5] B. Bouchon-Meunier, G. Coletti, R.R. Yager. *Modern Information Processing: From Theory to Applications*. Amsterdam: Elsevier, 2006.

[6] I.H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques (2nd edition)*. San Francisco: Morgan Kaufmann Publishers, 2005.

[7] K. Pearson. On lines and planes of closest fit to a system of points in space. *Philosophical Magazine*, 2:557–572, 1901.

[8] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J Educational Psychology*, 24:417–441, 1933.

[9] I.H. Bernstein, C.P. Garbin, and G.K. Teng. *Applied Multivariate Analysis*. New York: Springer-Verlag, 1988.

[10] W. Krzanowski. *Principles of Multivariate Analysis*. New York: Oxford University Press, 1988.

[11] B.F.J. Manly. *Multivariate Statistical Methods: A Primer*. New York: Chapman & Hall, 1986.

[12] R. Gnanadesikan. *Methods for Statistical Data Analysis of Multivariate Observations*. New York: John Wiley, 1977.

[13] N.J. Pizzi, L.-P. Choo, J. Mansfield, M. Jackson, W.C. Halliday, H.H. Mantsch, R.L. Somorjai. Neural network classification of infrared spectra of control and Alzheimer's diseased tissue. *Artificial Intelligence in Medicine*, 7:67–79 (1995).

[14] J.D. Jobson. *Applied Multivariate Data Analysis: Categorical and Multivariate Methods*. New York: Springer–Verlag, 1992.

[15] G.A.F. Seber. *Multivariate Observations*. Hoboken: Wiley and Sons, 1984.

[16] W. Pedrycz, A. Breuer, and N.J. Pizzi. Fuzzy adaptive logic networks as hybrid models of quantitative software engineering. *Intelligent Automation and Soft Computing*, 12:189–209, 2005.

[17] N.J. Pizzi and W. Pedrycz. A fuzzy logic network for pattern classification. *Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society*, June 14–17, Cincinnati, USA, 2009 (submitted for review).

[18] K.A. De Jong. *Evolutionary Computation: A Unified Approach*. Cambridge: MIT Press, 2006.

[19] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading: Addison-Wesley, 1989.

[20] R.L. Haupt and S.E. Haupt. *Practical Genetic Algorithms*. Hoboken: John Wiley & Sons, 2004.

[21] C. Jacob. *Illustrating Evolutionary Computation with Mathematica*. San Diego: Academic Press, 2001.

[22] A. Asuncion and D.J. Newman. *UCI Machine Learning Repository* [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, 2007.

[23] M.A. Little, P.E. McSharry, E.J. Hunter, and L.O. Ramig. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 2008 (to appear).

[24] M.A. Little, P.E. McSharry, S.J. Roberts, D.A.E. Costello, and I.M. Moroz. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomedical Engineering Online*, 6:23, 2007.

ISBN: 978-989-95079-6-8

Image reduction with interval-valued fuzzy sets and OWA operators

D. Paternain, H. Bustince, J. Sanz, M. Galar, C. Guerra
Departamento de Automática y Computación, Universidad Pública de Navarra
Pamplona, Spain
Email: {daniel.paternain,bustince,joseantonio.sanz,mikel.galar,carlos.guerra}@unavarra.es

Abstract— *In this paper we propose a generalization of Atanassov's operators and we prove that these generalized operators and OWA operators of dimension* 2 *provide the same numerical results. We apply Atanassov's operators to image compression and use different families of OWA operators in order to calculate the coefficient* α *of the Atanassov's operator.*

Keywords— Image compressing, OWA operators, Atanassov's operator, Interval-valued fuzzy set.

1 Introduction

Image compression consists of reducing the resolution of an image in order to display it, print it, or even speed up some other computations (see [10, 11, 18]). The target of image reduction algorithms is to reduce the size of the image with the smallest possible quality loss.

Atanassov's operator was given in 1983 ([1, 2]) allowing to associate a fuzzy set to each Atanassov's intuitionistic fuzzy set or interval-valued fuzzy set ([9, 12, 14]). Later, in 1988, Yager presented the Ordered Weighted Averaging (OWA) operators ([16])). Our paper is based on the idea that, under certain conditions, the numerical results of both operators are the same. For this reason, we define an operator generalizing Atanassov's operators, being an OWA operator of dimension 2([3, 4, 5]).

We apply this operator in image compression. In this way, we divide the image in blocks. We associate each block with an interval, getting an interval-valued fuzzy set associated with the image. We use the Atanassov's operator to obtain a single point from each interval. The result is a new (fuzzy) set representing the reduced image.

In order to calculate the α parameter of the Atanassov's operator, we focus on several families of OWA operators of dimension 2. We calculate the α coefficient according to the type of OWA operator, with the aim of getting different reductions from a single image.

This paper is organized in the following way. First, we start recalling some concepts related to interval-valued fuzzy sets and Atanassov's operator. In Section 3, we see the relation between OWA operators of dimension 2 and generalized Atannasov's operators. In Section 4 we present different methods to calculate α, based on different families of OWA operators. In Section 5 we propose an algorithm to compress images and in Section 6 we see different experimental results. We finish with some conclusions and references .

2 Preliminary definitions

We denote by L([0,1]) the set of all closed subintervals of the unit interval [0,1], that is,

$$L([0,1]) = \{\mathbf{x} = [\underline{x}, \overline{x}] | (\underline{x}, \overline{x}) \in [0,1]^2 \text{ and } \underline{x} \leq \overline{x}\}.$$

L([0,1]) is a partially ordered set with respect to the order relationship $\leq_L$ defined in the following way: given **x**,**y** $\in$ L([0,1])

$$\mathbf{x} \leq_L \mathbf{y} \text{ if and only if } \underline{x} \leq \underline{y} \text{ and } \overline{x} \leq \overline{y}.$$

With this order relationship, (L([0,1]),$\leq_L$) is a complete lattice ([7, 8, 3, 4]), where the smallest element is $0_L = [0,0]$ and the largest is $1_L = [1,1]$.

Definition 1. *An interval-valued fuzzy set (IVFS) A on the universe* $U \neq \emptyset$ *is a mapping* $A : U \to L([0,1])$.

M_A(u)=[$\underline{A}$(u),$\overline{A}$(u)] $\in$ L([0,1]) is the membership degree of u $\in$ U, with $\underline{A}$(u), $\overline{A}$(u) $\in$ [0,1] denoting the lower bound and the upper bound respectively of the membership associated to u.

In 1983, Atanassov proposed an operator to associate a fuzzy set with each interval-valued fuzzy set (see [1, 2]). This operator associates each interval with a point.

Definition 2. *Let* $\alpha \in [0,1]$. *The Atanassov's operator* K_α *is a mapping* $K_\alpha : L([0,1]) \to [0,1]$ *defined by*

1. $K_0(x) = \underline{x}$ *for all* $x \in L([0,1])$,
2. $K_1(x) = \overline{x}$ *for all* $x \in L([0,1])$,
3. $K_\alpha(x) = K_\alpha([K_0(x), K_1(x)]) = K_0(x) + \alpha(K_1(x) - K_0(x)) = \underline{x} + \alpha(\overline{x} - \underline{x})$ *for all* $x \in L([0,1])$.

To generalize this operator, the following definition was proposed in [4]:

Definition 3. *Let* $\alpha \in$ *[0,1]. An operator* D_α *is a mapping* $D_\alpha : L([0,1]) \to [0,1]$ *such that it satisfies the following conditions:*

1. *If* $\underline{x} = \overline{x}$, *then* $D_\alpha(\boldsymbol{x}) = \underline{x}$,
2. $D_0(\boldsymbol{x}) = \underline{x}$, $D_1(\boldsymbol{x}) = \overline{x}$ *for all* $\boldsymbol{x} \in L([0,1])$,
3. *If* $\boldsymbol{x} \leq_L \boldsymbol{y}$ *with* $\boldsymbol{x}, \boldsymbol{y} \in L([0,1])$, *then* $D_\alpha(\boldsymbol{x}) \leq D_\alpha(\boldsymbol{y})$,
4. $D_\alpha([0,1]) = \alpha$ *for any* $\alpha \in [0,1]$.

Example 1.

$$D_\alpha([\underline{x}, \overline{x}]) = \begin{cases} \overline{x} & \text{if } \overline{x} \leq \alpha \\ \underline{x} & \text{if } \underline{x} \geq \alpha \\ \alpha & otherwise \end{cases}$$

Proposition 1. *For any* $\alpha \in [0,1]$, K_α *is a* D_α *operator.*

Next, we propose a theorem to construct D_α operators using one-variable real functions.

ISBN: 978-989-95079-6-8

Theorem 1. *Let $\alpha \in [0,1]$ and let $f : [0,1] \to [0,1]$ be a continuous and strictly increasing function. Then the operator*

$$D_\alpha : L([0,1]) \to [0,1] \text{ given by}$$
$$D_\alpha(\boldsymbol{x}) = f^{-1}\Big(pf(\underline{x}) + (1-p)f(\overline{x})\Big)$$

with $p = \frac{f(1)-f(\alpha)}{f(1)-f(0)}$, is a D_α operator in the sense of Definition 3.

Proof:

1. If $\underline{x} = \overline{x}$, then $D_\alpha(\mathbf{x}) = f^{-1}(pf(\underline{x}) + (1-p)f(\underline{x})) = = f^{-1}(f(\underline{x})) = \underline{x}$.
2. If $\alpha = 0$, then $p = 1$. In these conditions $D_0([\underline{x}, \overline{x}]) = = f^{-1}(f(\underline{x})) = \underline{x}$.
   If $\alpha = 1$, then $p = 0$ and therefore $D_1([\underline{x}, \overline{x}]) = \overline{x}$.
3. If $[\underline{x}, \overline{x}] \leq_L [\underline{y}, \overline{y}]$, as f is continuous and strict, we have that $D_\alpha([\underline{x}, \overline{x}]) = f^{-1}(pf(\underline{x}) + (1-p)f(\overline{x})) \leq f^{-1}(pf(\underline{y}) + (1-p)f(\overline{y})) = D_\alpha([\underline{y}, \overline{y}])$.
4. If $\mathbf{x} = [0,1]$, then $D_\alpha([0,1]) = f^{-1}(p(f(0) + (1-p)f(1)) = f^{-1}(\frac{f(0)(f(1)-f(\alpha))-f(1)(f(1)-f(\alpha))+f(1)(f(1)-f(0))}{f(1)-f(0)}) = f^{-1}(\frac{f(\alpha)(f(1)-f(0))}{f(1)-f(0)}) = \alpha$.

Remark: In this paper, we take $f(x) = x$. Under this condition, by Theorem 1, we have that

$$D_\alpha([\underline{x}, \overline{x}]) = \underline{x} + \alpha(\overline{x} - \underline{x}) = K_\alpha([\underline{x}, \overline{x}])$$

3 K_α and OWA operators

In [16], Yager introduced the Ordered Weighted Aggregation Operator (OWA operator) in the following way:

Definition 4. *A mapping $F : [0,1]^n \to [0,1]$ is called an OWA operator of dimension n if there exists a weighting vector W,$W = (w_1, w_2, \ldots, w_n) \in [0,1]^n$ with $\sum_i w_i = 1$ and such that*

$$F(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j$$

with b_j the j-th largest of the a_i.

In the original definition, Yager considered the OWA operators as a mapping from the whole euclidean space $\mathbb{R}^n$ to $\mathbb{R}$. However, for us it is better to reduce the domain only to $[0,1]^2$.

There exist three important special cases of OWA operators that coincide with well known aggregation functions, as we can see in [17].

1. F^*: The weighting vector, denoted as W^*, is defined as $w_1 = 1$ and $w_j = 0$ for all $j \neq 1$.
   $F^*(a_1, \ldots, a_n) = max\{a_1, \ldots, a_n\}$
2. F_*: The weighting vector, denoted as W_*, is defined as $w_n = 1$ and $w_j = 0$ for all $j \neq n$.
   $F_*(a_1, \ldots, a_n) = min\{a_1, \ldots, a_n\}$.
3. F_A: The weighting vector, denoted as W_A, is defined as $w_j = 1/n$ for all $j \in 1, \ldots, n$.
   $F_A(a_1, \ldots, a_n) = \frac{1}{n}\sum_{i=1}^{n} a_i$.

If we focus on two-dimensional OWA operators using as weighting vector $W = (\alpha, 1-\alpha)$, we can think of applying it to the bounds of an interval. In this case, the numerical result of applying the OWA operator on the bounds of an interval and the operator D_α acting on that interval is the same. While D_α operator acts over elements in L([0,1]), the domain of OWA operators is [0,1]x[0,1]. For this reason, OWA operators acting on the unit square require an ordering operation to ensure the elements to be the extremes of an interval defined in L([0,1]). That's the reason why we need the following theorems to study the relationship between these operators.

We define a new operator $\mathbb{D}_\alpha$ by composing the D_α operator with the map

$$\begin{array}{rcl} i : [0,1]^2 & \to & L([0,1]) \\ (x,y) & \to & [min(x,y), max(x,y)] \end{array}$$

Theorem 2. *1. Let $\alpha \in [0,1]$ and $\mathbb{D} = D_\alpha \circ i$ where D_α is the operator given in Definition 3. Then, if F is the OWA operator (of dimension 2) defined by the weighting vector $W = (\alpha, 1-\alpha)$, we have that*
$$\mathbb{D}_\alpha(x,y) = F(x,y) \text{ for all } x, y \in [0,1].$$

2. Let F be an OWA operator (of dimension 2) with weighting vector $W = (w_1, w_2)$. Then for any $(x,y) \in [0,1]^2$ we have that
$$F(x,y) = \mathbb{D}_\alpha(x,y), \text{ with } \alpha = w_1.$$

4 Calculation methods for α coefficient

In this section, we study how to calculate the α coefficient basing on different families of OWA operators. First, we study two measures defined in [17] and associated with any OWA operator: the orness and the dispersion.

Definition 5. *Let F be an OWA operator and W its weighting vector. The orness measure is defined as*

$$orness(F) = \frac{1}{(n-1)}\sum_{i=1}^{n}(n-i)w_i.$$

With this definition it is easy to see that $orness(F^*) = 1$, $orness(F_*) = 0$ and $orness(F_A) = 0.5$.

Proposition 2. $orness(\mathbb{D}_\alpha) = \alpha$

Proof:

$$orness(\mathbb{D}_\alpha(x)) = \frac{1}{2-1}\sum_{i=1}^{2}(2-i)w_i = w_1 = \alpha$$

Definition 6. *Let F be an OWA operator and W its weighting vector. The dispersion measure is defined as*

$$Disp(F) = -\sum_{i=1}^{n} w_i ln w_i.$$

Proposition 3. $Disp(\mathbb{D}_\alpha) = \alpha ln(\frac{1-\alpha}{\alpha}) + ln(1-\alpha)$

Proof:

$$Disp(\mathbb{D}_\alpha) = -\textstyle\sum_{i=1}^{2} w_i ln w_i =$$

$$= \alpha(ln(1-\alpha) - ln(\alpha)) - ln(1-\alpha) =$$

$$= \alpha ln(\tfrac{1-\alpha}{\alpha}) + ln(1-\alpha)$$

ISBN: 978-989-95079-6-8

Next, it is shown the relation between the coefficient α of the operator $\mathbb{D}_\alpha$ and some families of OWA operators.

Definition 7. *F is called a ME-OWA operator if F is an OWA operator such that given a desired value of orness β, it maximizes the dispersion (entropy). In particular we solve the following problem:*

$$\begin{array}{ll} \textit{Max} & -\sum_{i=1}^{n} w_i ln w_i \\ \textit{subject to} & \beta = 1/(n-1)\sum_{i=1}^{n}(n-i)w_i \\ \textit{where} & \sum_{i=1}^{n} w_i = 1,\ w_i \in [0,1]. \end{array}$$

Proposition 4. *The operator $\mathbb{D}_\alpha$ is a ME-OWA for all $\alpha \in [0,1]$, with $orness(\mathbb{D}_\alpha) = \alpha$.*

Definition 8. *F is called a generalized S-OWA operator if F is an OWA operator where*

$$\begin{array}{lcll} w_1 & = & a + \frac{1}{n}(1-(a+b)), & \\ w_i & = & \frac{1}{n}(1-(a+b)), & \textit{with } i = 2,\ldots,n-1, \\ w_n & = & b + \frac{1}{n}(1-(a+b)). & \end{array}$$

with $a, b \in [0,1]$ and $a+b \leq 1$.

Proposition 5. *Let $a, b \in [0,1]$ such that $a+b \leq 1$. If we take*

$$\alpha = \frac{1+a-b}{2}$$

then $\mathbb{D}_\alpha$ is a generalized S-OWA operator.

Definition 9. *F is called a BADD-OWA operator if F is an OWA operator where*

$$w_i = \frac{b_i^\beta}{\sum_{j=1}^{n} b_j^\beta}$$

with $\beta \geq 0$ and being b_i as in Definition 4.

Proposition 6. *Let $\beta \geq 0$. If we take*

$$\alpha = \frac{(\mathbb{D}_1(x,y))^\beta}{(\mathbb{D}_0(x,y))^\beta + (\mathbb{D}_1(x,y))^\beta}$$

with $x, y \in [0,1]$, then $\mathbb{D}_\alpha$ is a BADD-OWA operator.

Definition 10. *Let $\beta \geq 0$. F is called a modified BADD-OWA operator if F is an OWA operator where*

1. $w_i = \frac{(1/b_i)^\beta}{\sum_{j=1}^{n}(1/b_j)^\beta}$ *or*
2. $w_i = \frac{(1-b_i)^\beta}{\sum_{j=1}^{n}(1-b_j)^\beta}$ *or*
3. $w_i = \frac{1}{n-1}(1 - \frac{b_i^\beta}{\sum_{j=1}^{n} b_j^\beta})$ *or*
4. $w_i = \frac{(b_{n-i+1})^\beta}{\sum_{j=1}^{n} b_j^\beta}$

being b_i as in Definition 4.

Proposition 7. *Let $\beta \geq 0$. The following items are satisfied:*

1. *If we take*

$$\alpha = \frac{(1/\mathbb{D}_1(x,y))^\beta}{(1/\mathbb{D}_1(x,y))^\beta + (1/\mathbb{D}_0(x,y))^\beta} =$$

$$= 1 - \frac{(\mathbb{D}_1(x,y))^\beta}{(\mathbb{D}_1(x,y))^\beta + (\mathbb{D}_0(x,y))^\beta} = \frac{(\mathbb{D}_0(x,y))^\beta}{(\mathbb{D}_1(x,y))^\beta + (\mathbb{D}_0(x,y))^\beta}$$

then $\mathbb{D}_\alpha$ is a modified BADD-OWA operator in the sense of items 1, 3 y 4 of Definition 10.

2. *If we take*

$$\alpha = \frac{(1-\mathbb{D}_1(x,y))^\beta}{(1-\mathbb{D}_1(x,y))^\beta + (1-\mathbb{D}_0(x,y))^\beta}$$

then $\mathbb{D}_\alpha$ is a modified BADD-OWA operator in the sense of item 2 of Definition 10.

Proof:

1. Obviously, taking into account items 1,3 and 4 of Definition 10 and that $\mathbb{D}_\alpha$ is an OWA operator of dimension 2.
2. Direct $\square$.

5 Image reduction

We consider an image Q as a $N \times M$ matrix. Each coordinate of the pixels in the image Q is denoted as (i,j). The intensity or gray level of the pixel located in (i,j) is represented as q_{ij}, with $0 \leq q_{ij} \leq L-1$ for each $(i,j) \in Q$.

In our approach to image reduction we use interval-valued fuzzy sets and D_α operator. It has been shown (see [13, 6, 15]) that interval-valued fuzzy sets used in images allow the development of algorithms in several topics as edge detecion, contrast or thresholding with very good results. We propose the following algorithm:

1. Divide the image Q in blocks of size $n \times n$. If M or N are not multiple of n, we delete the minimum number of rows/columns in the boundary of the image until the new size of the image satisfies the property.
2. Associate each block with an interval in the following way: the lower bound of the interval is given by the minimum of the intensities in the block and the upper bound by the maximum.
3. Choose the α parameter in the operator D_α.
4. Associate each interval with the number obtained after applying the operator D_α.

Example: Let Q be a matrix of dimension 6×6 and let $n = 3$

$$\begin{pmatrix} q_{1,1} & q_{1,2} & q_{1,3} & q_{1,4} & q_{1,5} & q_{1,6} \\ q_{2,1} & q_{2,2} & q_{2,3} & q_{2,4} & q_{2,5} & q_{2,6} \\ q_{3,1} & q_{3,2} & q_{3,3} & q_{3,4} & q_{3,5} & q_{3,6} \\ q_{4,1} & q_{4,2} & q_{4,3} & q_{4,4} & q_{4,5} & q_{4,6} \\ q_{5,1} & q_{5,2} & q_{5,3} & q_{5,4} & q_{5,5} & q_{5,6} \\ q_{6,1} & q_{6,2} & q_{6,3} & q_{6,4} & q_{6,5} & q_{6,6} \end{pmatrix}$$

Then, the interval-valued fuzzy set associated with Q is formed by 4 elements:

$$\left(\begin{bmatrix} \bigwedge_{\substack{i=1,2,3 \\ j=1,2,3}} q_{i,j} & \bigvee_{\substack{i=1,2,3 \\ j=1,2,3}} q_{i,j} \\ \bigwedge_{\substack{i=4,5,6 \\ j=1,2,3}} q_{i,j} & \bigvee_{\substack{i=4,5,6 \\ j=1,2,3}} q_{i,j} \end{bmatrix} \begin{bmatrix} \bigwedge_{\substack{i=1,2,3 \\ j=4,5,6}} q_{i,j} & \bigvee_{\substack{i=1,2,3 \\ j=4,5,6}} q_{i,j} \\ \bigwedge_{\substack{i=4,5,6 \\ j=4,5,6}} q_{i,j} & \bigvee_{\substack{i=4,5,6 \\ j=4,5,6}} q_{i,j} \end{bmatrix} \right)$$

Remark: Notice that the symbols $\wedge$ and $\vee$ stand for minimum and maximum respectively.

Once the interval-valued fuzzy set associated with the image has been obtained, we build a fuzzy set. Applying the operator D_α to each interval we get the reduced image of size 2×2.

ISBN: 978-989-95079-6-8

Figure 1: Original Image Lena

Figure 2: Original image Cameraman

6 Experimental results

In this section we use the algorithm proposed in Section 5. Using the relation between the operator D_α and OWA operators of dimension 2, we propose several methods of calculating the α coefficient based on families of OWA operators.

We get two different situations according to the OWA operator used. In the first one, the value of α is constant for the whole image. This situation happens with ME-OWA operators and generalized S-OWA operators, due to their definition. In the second one, the value of α depends on the bounds of each interval, so its value varies for each block. This happens when we take BADD-OWA operators and modified BADD-OWA operators.

The tests of this section have been made with the image Lena (Figure 1) and Cameraman (Figure 2). The size of the block is $n = 3$ (submatrices 3×3). In this way, the size of the reduced image is 9 times smaller than the original.

6.1 Calculation with constant α

In this section we use ME-OWA operators and generalized S-OWA to calculate a constant value of α for the whole image. Once the value has been calculated, we apply the D_α operator to each interval in the image.

6.1.1 Reduction with ME-OWA operators

As we have seen in Section 4, the construction of ME-OWA operators is direct. For this reason, we analyze three specific cases: $\alpha = 0$, $\alpha = 0.5$ and $\alpha = 1$.

With $\alpha = 0$, we associate the lower bound of the interval to each block. With $\alpha = 0.5$, we take the mean point of the interval. Finally, with $\alpha = 1$, we associate the upper bound of the interval.

Obviously, the higher the value of α, the higher the membership degree and therefore the intensity of the image. The image is darker with $\alpha = 0$ than with $\alpha = 0.5$, which is also darker than with $\alpha = 1$.

$\alpha = 0$ $\alpha = 0.5$ $\alpha = 1$

Figure 3: Reduction using ME-OWA operators

6.1.2 Reduction with generalized S-OWA operator

To construct a generalized S-OWA operator, it is necessary, as we have seen in Definition 8, the election of two parameters $a, b \in [0, 1]$ such that $a + b \leq 1$. If we focus on the contribution of the parameters in the S-OWA operator, we can study several possibilities:

1. If $a = b$, then we have that $\alpha = 0.5$, getting the average of the bounds of the interval.
2. If $a = 1$ and $b = 0$, we get $\alpha = 1$.
3. If $a = 0$ and $b = 1$, we get $\alpha = 0$.
4. If $a + b < 1$ and $a \neq b$, when $a > b$, the upper bound of the interval is predominant over the lower bound, while when $a < b$, the importance is given to the lower.

Cases 1, 2 and 3 have been already analyzed in Section 6.1.1, so we focus in case 4. For this, we start taking $a = 0.5, b = 0.25$. As $a > b$, the importance is given to the upper bound of the interval, getting a value of $\alpha = 0.625$. If we increase the value of a, then α tends to 1. For example, if we take $a = 0.9, b = 0$, then $\alpha = 0.95$. With this value of α we get a reduced image very similar to the images studied in the last column of Figure 3. On the other side, if we take $a = 0.25, b = 0.5$, we get a value of $\alpha = 0.375$. If we increase the value of b, $\alpha \to 0$. We can see this if we take $a = 0, b = 0.9$ ($\alpha = 0.05$). Notice that as α increase, the image becomes lighter.

6.2 Calculation with variable α

In this section, the value of α is calculated by means of the bounds of the interval. This means that the value of α is variable. For this reason we use BADD-OWA and modified BADD-OWA operators.

ISBN: 978-989-95079-6-8

a=0,b=0.9 $\alpha = 0.05$ | a=0.25,b=0.5 $\alpha = 0.375$ | a=0.5,b=0.25 $\alpha = 0.625$ | a=0.9,b=0 $\alpha = 0.95$

Figure 4: Reduction using generalized S-OWA operators

6.2.1 Reduction with BADD-OWA operators

As we see in Definition 9, to construct a BADD-OWA operator it is necessary to select a value of $\beta \geq 0$. We know that if we take $\beta = 0$, then we get $\alpha = 0.5$, already studied in Section 6.1.1. In other case, if we take $\beta = 1$, the value of α is calculated as follows:

$$\alpha = \frac{\overline{x}}{\overline{x} + \underline{x}}$$

being $[\underline{x}, \overline{x}] \in L([0,1])$ the interval representing each block. We also know that $\alpha \geq 0.5$, and if we increase the value of β, $\alpha \to 1$. That is, the result tends to the upper bound of the interval. For this reason, with a high value of β, we get reduced images similar as the images analyzed in third column of Figure 3.

$\beta = 1$ | $\beta = 10$ | $\beta = 20$ | $\beta = 100$

Figure 5: Reduction using BADD-OWA operators

6.3 Reduction with modified BADD-OWA operators

If we use modified BADD-OWA operators, bearing in mind items 1 and 2 of Proposition 7, we can take two different expressions for α. As we have analyzed in BADD-OWA operators, the value of α depends on the interval and on the β parameter. As in Section 6.2.1, if $\beta = 0$, then $\alpha = 0.5$.

Under conditions of item 1 of Proposition 7 and taking $\beta = 1$, the value of α is as follows:

$$\alpha = \frac{\underline{x}}{\underline{x} + \overline{x}}$$

being $[\underline{x}, \overline{x}] \in L([0,1])$ the interval representing each block. In this case, $\alpha \leq 0.5$ and as β increases, $\alpha \to 0$ when $\beta \to \infty$ (notice that the image becomes darker as β increases and α decreases).

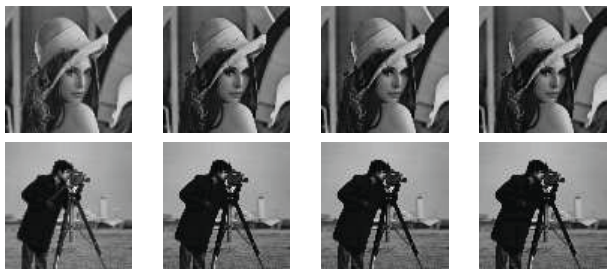
$\beta = 1$ | $\beta = 10$ | $\beta = 20$ | $\beta = 100$

Figure 6: Reduction using modified BADD-OWA operators (item 1)

If we base on item 2 of Proposition 7, for $\beta = 1$, α is as follows:

$$\alpha = \frac{1 - \overline{x}}{(1 - \overline{x}) + (1 - \underline{x})}$$

being $[\underline{x}, \overline{x}] \in L([0,1])$ the interval representing each block. In this case, the value of α also decreases when β increases.



$\beta = 1$ | $\beta = 10$ | $\beta = 20$ | $\beta = 100$

Figure 7: Reduction using modified BADD-OWA operators (item 2)

7 Conclusions and future research

In this paper, we have proved that we can obtain OWA operators of dimension 2 from Atanassov's operators. This fact has allowed us to use several families of OWA operators to get D_α operators for image reduction.

Some future lines of research can be:

1. Compare the different reductions obtained according to the value of α.
2. Compare our results with other reduction algorithms.
3. Consider different methods of reconstruction of the original image from the reduced one. Analyze which of the reduced images leads to the best reconstructed image.

Acknowledgment - This paper has been partially supported by the National Science Foundation of Spain, Reference TIN2007-65981.

ISBN: 978-989-95079-6-8

References

[1] K. Atanassov, Intuitionistic fuzzy sets, In: VIIth ITKR Session, Deposited in the Central Science and Technology Library of the Bulgarian Academy of Sciences, Sofia, Bulgaria, 1983, pp. 1684-1697.

[2] K. Atanassov, Intuitionistic fuzzy sets, Fuzzy Sets and Systems 20 (1986) 87-96.

[3] H. Bustince, E. Barrenechea, M. Pagola, Generation Of Interval-Valued Fuzzy And Atanassov's Intuitionistic Fuzzy Connectives From Fuzzy Connectives And From K(Alpha) Operators. Laws For Conjunctions And Disjunctions. Amplitude, International Journal of Intelligent Systems (2008) In Press

[4] H. Bustince, J. Montero, M. Pagola, E. Barrenechea, D. Gmez: A Survey On Interval-Valued Fuzzy Sets. In W. Pedrycz, A. Skowron, V. Kreinovichedrycz (Eds.): *Handbook of Granular Computing* John Wiley and sons, NewYork, (2008), Chapter 22.

[5] H. Bustince, J. Montero, E. Barrenechea, M. Pagola, Laws of Conjunctions and Disjunctions in Interval type 2 fuzzy sets. In: Proc. IEEE World Congress on Computational Intelligence, WCCI2008 , Hong Kong, 2008.

[6] T. Chaira, A.K. Ray, A new measure using intuitionistic fuzzy set theory and its application to edge detection, Applied Soft Computing 8(2) (2008) 919-927.

[7] C. Cornelis, G. Deschrijver, E.E. Kerre, Advances and challenges in interval-valued fuzzy logic, Fuzzy Sets and Systems 157 (2006) 622-627.

[8] G. Deschrijver,C. Cornelis, E.E. Kerre, On the representation of intuitionistic fuzzy T-norms and T-conorms, IEEE Transactions on Fuzzy Systems 12(1) (2004) 45-61.

[9] M.B. Gorzalczany, A method of inference in approximate reasoning based on interval-valued fuzzy sets, Fuzzy Sets and Systems 21 (1987) 1-17.

[10] E. Karabasssis, M. E. Spektsakis, An analysis of image interpolation, differentiation and reduction using local polynomial fits, Graphical models and image processing, 57 (3), (1995) 183-1995.

[11] H. Nobuhara, W. Pedrycz, S. Sessa, K. Hirota, A motion compression/reconstruction method based on max t-norm composite fuzzy relational equations, Information Sciences 176 (2006) 2526-2552.

[12] R. Sambuc, Function Φ-Flous, Application a l'aide au Diagnostic en Pathologie Thyroidienne, These de Doctorat en Medicine, University of Marseille (1975).

[13] H.R. Tizhoosh, Image thresholding using type-2 fuzzy sets, Pattern Recognition 38 (2005) 2363-2372.

[14] I. B. Turksen, Interval valued fuzzy sets based on normal forms, Fuzzy Sets and Systems 20(2) (1986) 191-210.

[15] I.K. Vlachos, G.D. Sergiadis, Intuitionistic fuzzy information - Applications to pattern recognition, Pattern Recognition Letters 28(2) (2007) 197-206.

[16] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decisionmaking, IEEE Trans. Syst. Man Cybern, 18 (1988) 183-190.

[17] R.R. Yager, Families of OWA operators, Fuzzy Sets and Systems, 59 (1993) 125-148.

[18] G. Y. Yang, H.Z. Shu, C. Toumoulin, G.N. Han, L.M. Luo, Efficient Legendre moment computation for grey level images, Pattern Recognition, 39 (2006) 74-80.

ISBN: 978-989-95079-6-8

Fuzzy Transforms, Korovkin Theorems and the Durrmeyer Operator

Barnabas Bede, Imre J. Rudas
Department of Mathematics, University of Texas-Pan American,
Edinburg, Texas, USA. email: bedeb@utpa.edu.
Department of Intelligent Engineering Systems, John von Neumann Faculty of Informatics,
Budapest Tech, Budapest, Hungary. email: rudas@bmf.hu

***Abstract*— *In the present paper a Korovkin-type theorem is proposed for the approximation operators defined by the inverse F-transforms. These results allow us to choose between a variety of shapes to be used as atoms of the fuzzy partitions used within the F-transform's framework. In this way we can enlarge considerably the class of F-transforms proposed recently by I. Perfilieva. The new fuzzy partitions are shown to include, for example the Bernstein basis polynomials. The F-transform considered with the Bernstein basis polynomials as atoms of a fuzzy partition, is shown to be the classical Durrmeyer operator.*

***Keywords*—** F-transform, Korovkin Theorems, Bernstein Polynomial, Durrmeyer operator.

1 Introduction

The F-transform was proposed by I. Perfilieva in [4] and studied in several papers [5], [7], etc. As starting point for the F-transforms, one considers first a fuzzy partition [6]. In the present paper we enlarge the class of F-transforms by considering arbitrary shapes and not necessarily small support for the atoms of the fuzzy partition used in the F-transform as an approximation method. To study the possibility of such extension we exploit the classical Korovkin theorems, [1] which are particularized in the present paper for the F-transform. The Korovkin Theorems are very important results in classical Approximation Theory, stating that the convergence of a sequence of positive linear approximation operators on a finite set of test functions implies the convergence of the operator for any function. There are two versions of the Korovkin type results. A qualitative and a quantitative version. The quantitative versions provide error estimates, beyond proving the convergence results of qualitative type. We will adopt the quantitative approach. Further-on in the present paper we consider the F-transforms based on Bernstein basis polynomials. It turns out that the F-transform that uses Bernstein basis polynomials as atoms for the fuzzy partition under discussion, leads to the classical Durrmeyer operators [2]. In this way, the theory of fuzzy approximation and especially approximation by inverse F-transforms can be embedded and studied, as a refined part of the classical theory of approximation.

In the present paper a fuzzy partition is defined as a finite sequence of fuzzy sets $A_i : [a,b] \to [0,1]$, $i = 1, ..., k$, such that $\sum_{i=1}^{k} A_i(x) = 1$ for any $x \in [a,b]$. The following is a generalized version of the Fuzzy Transform (F-transform) proposed in [4], since the condition regarding the small support of the atoms of the fuzzy partitions occurring in the F-transform is released.

According to [4], the F-transform is given by

$$f_i = \frac{\int_a^b A_i(x) f(x) dx}{\int_a^b A_i(x) dx}, i = 1, ..., k \tag{1}$$

and the discrete F-transform is given by

$$f_i = \frac{\sum_{j=1}^{n} A_i(x_j) f(x_j)}{\sum_{j=1}^{n} A_i(x_j)}, i = 1, ..., k \tag{2}$$

where $x_j \in [a,b]$, $j = 1, ..., n$ are given data. The inverse discrete F-transform is

$$F(x) = \sum_{i=1}^{k} A_i(x) f_i. \tag{3}$$

We denote by $C(K)$ the space of continuous functions $f : K \to \mathbb{R}$ with K a compact metric space. Then the space $C(K)$ is a Banach space with the uniform norm

$$\|f\| = \sup_{x \in K} |f(x)|.$$

If $K = [a,b]$ then we use the notation

$$C[a,b] = \{f : [a,b] \to \mathbb{R}, f \text{ continuous}\}.$$

We will consider inverse F-transform as an approximation operator. Let $\mathcal{L}(C(K))$ denote the space of linear operators of the form $T : C(K) \to C(K), T(\alpha f + \beta g) = \alpha T(f) + \beta T(g)$, $\forall \alpha, \beta \in \mathbb{R}$ and $f, g \in C(K)$. An operator is said to be positive if $T(f(x)) \geq 0 \; \forall x \in K$ whenever $f(x) \geq 0, \forall x \in K$. In the present paper positive linear operators play a very important role.

The inverse F-transform will be regarded in the present paper as a sequence of positive linear operators. Indeed, let us consider $F_{n,k} : C[a,b] \to C[a,b]$, given by $F_{n,k}(f)(x) = F(x)$, where $F(x)$ is given by (3) and f_i given by (1) or (2). Then, it is easy to see that $F_{n,k}$ is a linear operator, i.e., $F_{n,k} \in \mathcal{L}(C[a,b])$. Moreover, if we assume $f(x) \geq 0$ $\forall x \in [a,b]$, we can see that both the continuous and discrete F-transforms $f_i, i = 1, ..., k$ satisfy $f_i \geq 0$, and as a consequence the inverse F-transform $F_{n,k}(f)(x) = F(x) \geq 0$ and as a conclusion $F_{n,k}$ is a positive linear operator.

In order to obtain an error estimate and the uniform convergency of the F-transforms that we intend to study, seen as approximation operators via a particularization of the Korovkin Theory related to positive linear operators. For this aim we need the modulus of continuity. Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then $\omega(f, \cdot) : [0, \infty) \to [0, \infty)$, defined by

$$\omega(f, \delta) = \bigvee \{|f(x) - f(y)|; x, y \in [a,b], \; |x - y| \leq \delta\}$$

ISBN: 978-989-95079-6-8

is called the first order modulus of smoothness (modulus of continuity) of f.

2 Korovkin type theorem for the F-transform

Let us recall Korovkin's Theorem in what follows. We can see that it is directly applicable for the F-transform, since the F-transform is a positive linear operator, however, we will deduce later particularizations of the Korovkin results for the F-transform to better understand the F-transform method and to be able to provide error estimates for them.

Theorem 1 *(Korovkin, [3]) Let $L_n \in \mathcal{L}(C[a,b])$, $n = 1,2,...$ be a sequence of positive linear operators. Then there exists a finite set of test functions S, such that $\lim_{n\to\infty} L_n(s) = s$ for any $s \in S$ implies $\lim_{n\to\infty} L_n(f) = f$, for any $f \in C[a,b]$. The finite set S is called a Korovkin set.*

If we consider a sequence of positive linear operators on $C[0,1]$, then a Korovkin set (it is not unique) consists of $e_0, e_1, e_2,$ where $e_i(x) = x^i$, $i = 0,1,2$. If we consider $\mathcal{L}(C[-\pi,\pi])$ a Korovkin set consists of e.g., $e_0(x) = 1$, $e_1(x) = \sin x$ and $e_2(x) = \cos x$.

The adaptation of Korovkin Theorem for the F-transform is as follows.

Theorem 2 *(Korovkin Theorem for the discrete inverse F-transform) Let $F_{n,k} \in \mathcal{L}(C[a,b])$, $n,k = 1,2,...$ be a sequence of inverse F-transforms. The following general error estimate holds true for any $\delta > 0$:*

$$\|F_{n,k}(f) - f\| \leq \left(1 + \frac{1}{\delta}\sqrt{e_2 - 2e_1F_{n,k}(e_1) + F_{n,k}(e_2)}\right)\omega(f,\delta),$$

where $e_1(x) = x, e_2(x) = x^2$.

Proof. The inverse discrete F-transform can be expressed as

$$F_{n,k}(f)(x) = \sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)f(x_j)}{\sum_{j=1}^{n} A_i(x_j)}.$$

Let us observe that for the inverse F-transform we have $F_{n,k}(e_0) = e_0$, where $e_0(x) = 1\ \forall x \in [a,b]$, so $f(x)$ can be written as

$$f(x) = \sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)f(x)}{\sum_{j=1}^{n} A_i(x_j)}.$$

Then we get

$$|F(x) - f(x)| \leq \sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)\,|f(x) - f(x_j)|}{\sum_{j=1}^{n} A_i(x_j)}.$$

Further, using the properties of the modulus of continuity, by standard reasoning we have

$$|F(x) - f(x)| \leq \left(1 + \frac{1}{\delta}\sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)|x - x_j|}{\sum_{j=1}^{n} A_i(x_j)}\right)\omega(f,\delta),$$

for any fixed $\delta > 0$. The error is controlled by the ratio

$$R_{n,k}(x) = \sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)|x - x_j|}{\sum_{j=1}^{n} A_i(x_j)}.$$

Using Cauchy-Schwarz inequality we have

$$\sum_{j=1}^{n} A_i(x_j)|x - x_i| \leq \left(\sum_{j=1}^{n} A_i(x_j)|x - x_j|^2\right)^{\frac{1}{2}} \cdot \left(\sum_{j=1}^{n} A_i(x_j)\right)^{\frac{1}{2}},$$

and we obtain

$$R_{n,k}(x) \leq \sum_{i=1}^{k} A_i(x)\frac{\left(\sum_{j=1}^{n} A_i(x_j)|x - x_j|^2\right)^{\frac{1}{2}}}{\left(\sum_{j=1}^{n} A_i(x_j)\right)^{\frac{1}{2}}} = \sum_{i=1}^{k} \sqrt{A_i(x)}\frac{\left(A_i(x)\sum_{j=1}^{n} A_i(x_j)|x - x_j|^2\right)^{\frac{1}{2}}}{\left(\sum_{j=1}^{n} A_i(x_j)\right)^{\frac{1}{2}}}.$$

Using Cauchy-Schwarz inequality again we obtain

$$\sum_{i=1}^{k} \sqrt{A_i(x)}\frac{\left(A_i(x)\sum_{j=1}^{n} A_i(x_j)|x - x_j|^2\right)^{\frac{1}{2}}}{\left(\sum_{j=1}^{n} A_i(x_j)\right)^{\frac{1}{2}}} \leq \left(\sum_{i=1}^{k} A_i(x)\right)^{\frac{1}{2}}\left(\sum_{i=1}^{k}\frac{A_i(x)\sum_{j=1}^{n} A_i(x_j)|x - x_j|^2}{\sum_{j=1}^{n} A_i(x_j)}\right)^{\frac{1}{2}}.$$

Taking into account that $A_i(x)\ is$ a fuzzy partition we have

$$R_{n,k}(x) \leq \left(\sum_{i=1}^{k} A_i(x)\frac{\sum_{j=1}^{n} A_i(x_j)(x^2 - 2xx_j + x_j^2)}{\sum_{j=1}^{n} A_i(x_j)}\right)^{\frac{1}{2}} = \left(x^2 - 2xF_{n,k}(e_1)(x) + F_{n,k}(e_2)(x)\right)^{\frac{1}{2}}$$

which completes the proof. ■

Corollary 3 *Let $F_{n,k} \in \mathcal{L}(C[a,b])$, $n,k = 1,2,...$ be a sequence of inverse F-transforms. If $\lim_{n,k\to\infty} F_{n,k}(e_i) = e_i$, $i = 1,2$, then $\lim_{n\to\infty} L_n(f) = f$, for any $f \in C[a,b]$.*

For the continuous case we have a similar result.

Theorem 4 *(Korovkin Theorem for the inverse F-transform) Let $F_k \in \mathcal{L}(C[a,b])$, $k = 1,2,...$ be a sequence of inverse F-transforms. The following general error estimate holds true for any $\delta > 0$:*

$$\|F_k(f) - f\| \leq \left(1 + \frac{1}{\delta}\sqrt{e_2 - 2e_1F_k(e_1) + F_k(e_2)}\right)\omega(f,\delta),$$

where $e_1(x) = x, e_2(x) = x^2$.

Corollary 5 *Let $F_k \in \mathcal{L}(C[a,b])$, $k = 1,2,...$ be a sequence of continuous inverse F-transforms. If $\lim_{k\to\infty} F_k(e_i) = e_i$, $i = 1,2$, then $\lim_{n\to\infty} L_n(f) = f$, for any $f \in C[a,b]$.*

ISBN: 978-989-95079-6-8

We observe that the inverse F-transform can be expressed as

$$F_k(f)(x) = \sum_{i=1}^{k} A_i(x) \frac{\int_a^b A_i(x) f(x) dx}{\int_a^b A_i(x) dx}$$

so, similar reasoning to the previous Theorem can be followed.

The Korovkin type Theorem above shows that if we are able to control $\sqrt{e_2 - 2e_1 F_{n,k}(e_1) + F_{n,k}(e_2)}$ for the discrete case, or $\sqrt{e_2 - 2e_1 F_k(e_1) + F_k(e_2)}$ for the continuous case, for a given sequence of fuzzy partitions A_i, $i = 1, ..., k$, then we are able to control the error for any continuous function.

3 Example

Using the Korovkin-type result shown in the previous section we consider the F-transforms, with Bernstein basis polynomials used as a fuzzy partition this time. The proposed construction leads to the well-known Durrmeyer operator. Let us recall that the Bernstein basis polynomials are

$$p_{k,i}(x) = \binom{k}{i} x^i (1-x)^{k-i},$$

$k = 1, 2, ...$ $i = 0, 1, ..., k$ $x \in [0, 1]$.

The inverse F-transform is given by

$$F_k(f)(x) = \sum_{i=0}^{k} p_{k,i}(x) \frac{\int_0^1 p_{k,i}(x) f(x) dx}{\int_0^1 p_{k,i}(x) dx},$$

that is the classical Durrmeyer operator [2]. Let us calculate $F_k(e_1)$ and $F_k(e_2)$ in this case. It is well known that $p'_{k,.i}(x) = k(p_{k-1,i-1}(x) - p_{k-1,i}(x))$. Integrating we obtain $\int_0^1 p_{k-1,i-1}(x)dx = \int_0^1 p_{k-1,i}(x)dx$, i.e., for fixed k integrals of all Bernstein basis polynomials are the same. Since $\sum_{i=0}^{k} p_{n,i}(x) = 1$ we get $\int_0^1 p_{k,i}(x)dx = \frac{1}{k+1}$. By direct calculation

$$\int_0^1 p_{k,i}(x) x dx = \int_0^1 \binom{k}{i} x^{i+1} (1-x)^{k-i} dx$$

$$= \frac{i+1}{k+1} \int_0^1 p_{k+1,i+1} dx = \frac{i+1}{(k+1)(k+2)}.$$

Also,

$$\int_0^1 p_{k,i}(x) x^2 dx = \int_0^1 \binom{k}{i} x^{i+2} (1-x)^{k-i} dx$$

$$= \frac{i+1}{k+1} \frac{i+2}{k+2} \int_0^1 p_{k+2,i+2} dx = \frac{(i+1)(i+2)}{(k+1)(k+2)(k+3)}.$$

Then we obtain

$$f_i(e_1) = \frac{\int_0^1 p_{k,i}(x) x dx}{\int_0^1 p_{k,i}(x) dx}$$

$$= \frac{i+1}{k+2} \sim \mathcal{O}\left(\frac{i}{k}\right)$$

and

$$f_i(e_2) = \frac{\int_0^1 p_{k,i}(x) x^2 dx}{\int_0^1 p_{k,i}(x) dx}$$

$$= \frac{(i+1)(i+2)}{(k+2)(k+3)} \sim \mathcal{O}\left(\frac{i^2}{k^2}\right).$$

$$F_k(e_1)(x) \sim \sum_{i=1}^{k} p_{k,i}(x) \frac{i}{k} = B_k(e_1)(x) = x,$$

where B_k denotes the usual Bernstein operator

$$B_k(f)(x) = \sum_{i=1}^{k} p_{k,i}(x) f\left(\frac{i}{k}\right)$$

and it is well known that $B_k(e_1)(x) = x$. We also have

$$F_k(e_2)(x) \sim \sum_{i=1}^{k} p_{k,i}(x) \left(\frac{i}{k}\right)^2$$

$$= B_k(e_2)(x) = x^2 + \frac{x(1-x)}{k},$$

and finally we obtain: $\sqrt{e_2 - 2e_1 F_k(e_1) + F_k(e_2)} = \sqrt{\frac{x(1-x)}{k}}$ and by the above Korovkin type results we have

$$\|F_k(f) - f\| \leq C\omega\left(f, \sqrt{\frac{x(1-x)}{k}}\right).$$

As a conclusion we observe that the F-transform is a generalization of the Durrmeyer operators. For the discrete F-transform with Bernstein basis polynomials a similar result can be proven.

4 Conclusion

Korovkin-type theorems for the discrete and continuous inverse F-transforms regarded as positive linear approximation operators were provided. As an application we have studied the F-transform with Bernstein basis polynomials as atoms of the underlying fuzzy partition. It turns out that the F-transform with Bernstein basis polynomials is the classical Durrmeyer operator. Surely one can imagine different fuzzy partitions as well, so the F-transforms can be seen as generalizations of the Durrmeyer operator.

References

[1] R.A. DeVore, G.G. Lorentz, Constructive Approximation, Polynomials and Splines Approximation, Springer-Verlag, Berlin, Heidelberg, 1993.

[2] J.L. Durrmeyer, Une formule d'inversion de la transformée de Laplace: Applications à la théorie des moments. Thèse de 3e cycle (1967) Paris.

[3] P.P. Korovkin, Linear operators and approximation theory, Hindustan Pub. Corp. 1960.

[4] I. Perfilieva, Fuzzy Transforms: theory and applications, Fuzzy Sets and Systems, 157(2006) 993-1023.

[5] I. Perfilieva, Fuzzy Transforms, in Transactions on Rough Sets II (J.F. Peters et al. Eds.) LNCS 3135 (2004), 63-81.

[6] H. Ruspini, E. (1969) A new approach to clustering. Inform. and Control, Vol. 15, 22–32.

[7] M. Stepnicka, O. Polakovi, A neural network approach to the fuzzy transform, Fuzzy Sets and Systems, to appear.

ISBN: 978-989-95079-6-8

Fuzzy Logic Applications in Wireless Communications

Maria Erman[1] Abbas Mohammed[2] Elisabeth Rakus-Andersson[3]

[1,2]Department of Signal Processing, Blekinge Institute of Technology
S-37225 Ronneby, Sweden
[3]Department of Mathematics and Science, Blekinge Institute of Technology
S-37179 Karlskrona, Sweden
Email: maria.erman@bth.se, abbas.mohammed@bth.se, elisabeth.andersson@bth.se

Abstract— *A survey of fuzzy logic applications and principles in wireless communications is presented, with the aim of highlighting successful usage of fuzzy logic techniques in applied telecommunications and signal processing. To the best of our knowledge, this is the first such study of its kind. This paper will focus firstly on discerning prevalent fuzzy logic or fuzzy-hybrid approaches in the areas of channel estimation, channel equalization and decoding, and secondly outlining what conditions and situations for which fuzzy logic techniques are most suited for these approaches. Furthermore, after insights gained from isolating fuzzy logic techniques applied to real problems, this paper proposes areas for further research targeted to practice-oriented researchers.*

Keywords— Channel Equalization, Channel Estimation, Decoding, Fuzzy-hybrid systems, Fuzzy Logic, Wireless Communications

1 Introduction

Fuzzy Logic has been successfully applied in various areas pertaining to wireless communication systems. As fuzzy logic is used to model systems and situations, taking into consideration uncertainty and ambiguity, it can be an efficient tool to be utilized in problems for which knowledge of all factors is insufficient or impossible to obtain. Methods furnished with fuzzy logic have been shown to be useful in difficult conditions with respect to non-linear and time-variant systems. Additionally, the often mentioned advantages of using fuzzy logic in practical applications is to reduce complexity as well as to add robustness to the system under study.

Fuzzy logic and, more specifically, fuzzy control traditionally incorporates human expert knowledge into a rule-based framework. It may, however, be further expanded with learning algorithms to derive the fuzzy control parameters from sample data. These parameters may be obtained by combining fuzzy logic with related soft computing disciplines such as, e.g., neural networks, evolutionary computation techniques etc. On the other hand, a method developed by Wang and Mendel [1] derives the fuzzy rule base by using a combination of human experience and numerical data.

Wireless communications is a rapidly evolving industry, constantly challenging researchers for new techniques in order to meet the demands of ever higher performance and efficiency. The most obvious products of wireless communications, e.g., the worldwide adoption of the mobile telephone, wireless local area networks etc., exert a strong influence on many people's lives today.

In a wireless communication system, the *channel* is the medium by which information-bearing signals are transferred from a transmitter to a receiver. The characteristics of the channel are generally unknown, and, barring any distortion imposed by the channel, i.e., in ideal conditions, the transmitted data will be received without any errors.

In practice, however, it is unavoidable for the channel to be affected by distortion, hence degrading the performance of the receiver, severely limiting the throughput of the system. The wireless channel poses tough challenges for achieving reliable and fast transfers. While interference typically is not a major concern in wired transmission, i.e., in predicting the behavior of the signal in the transmission channel, it poses a great challenge in wireless transmission.

When there is no line-of-sight between the transmitter and the receiver, distortions to the signal in the form of effects such as scattering and reflections, etc. will follow, all of them resulting in a phenomenon called multipath propagation. Due to multipath propagation, the receiver encounters many signal paths from the transmitter, where each of these paths is delayed by an arbitrary amount and attenuated by various factors. As a result there will be a superposition of the different copies of the signal being in different phases, hence causing an amplification or attenuation of the signal power – also referred to as fading. Multipath propagation will cause previously sent data bits to smear into current data bits, referred to as intersymbol interference. The aim of the receiver in the communication system is to overcome the disturbances of the channel, intersymbol interference and noise, and correctly decode the data having been transmitted.

2 Contemporary Uses of Fuzzy Logic in Wireless Communications

In this section we will focus on fuzzy logic applications in channel estimation, channel equalization and decoding. The purpose of channel estimation is to accurately describe the channel and track its variations, and with the aid of channel equalization and decoding recover the original transmitted data. In the case of time-varying channels, adaptive techniques have to be employed and it is in this area that fuzzy techniques and/or neural networks find their main uses.

2.1 Channel Estimation

In [2] channel estimation is performed by tracking the channel coefficients, applying a fuzzy tracking method in a multipath fading Code Division Multiple Access (CDMA) [3] channel. CDMA is a spread-spectrum technology that makes it possible for transmitters to share the same frequency range. The

ISBN: 978-989-95079-6-8

fuzzy tracking method is based on Kosko's fuzzy associative memory models [4]. The fuzzy associative memory models combines fuzzy logic and a single-layer feed-forward neural network that saves the fuzzy logic rule-base in matrix form. The tracking used in [2] is iterative with the estimated symbol being used in the prediction of the coefficient. There are two inputs to the fuzzy tracker: the difference between the measured and predicted coefficients, and the change of difference between current and previous differences between the measured and predicted coefficients. The output of the fuzzy tracker yields a correction term for the next coefficient.

The motivation in [2] for using fuzzy channel estimation is due to fuzzy estimation not needing exact process models. Comparisons are made between the fuzzy tracker and a non-fuzzy tracker, a.k.a. the alpha tracker, where it is shown that the fuzzy tracker performs better under noisier multipath conditions.

Channel estimation using a fuzzy approach has also been performed with a multi-carrier modulation technique called Orthogonal Frequency Division Multiplexing (OFDM) [5]. In OFDM, multiple orthogonal subcarriers are used for the same channel. The data stream is divided into lower bit-rate data streams, each modulating a separate subcarrier.

There are two different methods for estimating the channel parameters at each subcarrier: blind channel estimation techniques and pilot assisted channel estimation, with the pilot being a reference signal used by the transmitter and the receiver. The blind channel estimation techniques do not use pilot samples and are thus more spectrally efficient, but at the cost of higher computational complexity and slower convergence rate. Pilot assisted channel estimation has typically been based either on the Minimum Mean-Square-Error (MMSE), the LS (Least Square) or the LMS (Least Mean-Square) algorithms [6], with the MMSE algorithm being more robust and performing better in time-varying channels.

In [7] a Takagi-Sugeno-Kang (TSK) [8] model is used for the pilot assisted channel estimation in an OFDM system. The TSK method is similar to that of the traditional Mamdani method – the difference being that the output of the TSK model is a linear function of the input variables instead of fuzzy sets. Pilot symbols are used in order to train the TSK fuzzy model as well updating the fuzzy model in order to track the channel. A Gaussian membership function is chosen, as it has been shown that a fuzzy system with bell-shaped Gaussians can approximate any continuous functions on compact sets to any degree of accuracy [9]. The TSK learning algorithm consists of defining the center and width of the rules dependent on the number of rules chosen, after which an adjustable parameter of the TSK model is trained from the first snapshot of the pilot subcarriers. Finally the channel transfer function is estimated and the adjustable parameter updated to track the channel. Simulation results show that the proposed TSK channel estimation model performs closely to the ideal MMSE, but with lower computational complexity.

The TSK fuzzy modeling technique used in [7] has been updated for a multiple-input multiple-output (MIMO) OFDM system in [10] with two transmit and two receive antennas. MIMO [11] is a way to increase system throughput without the need for higher transmit power or bandwidth, and has hence become a highly popular research topic. MIMO achieves increased throughput by using multiple antennas at the transmitter and the receiver. It is shown in [10] with computer simulations that the Word Error Rate (WER) is close to the MMSE method with lower computational complexity.

In [12] an MMSE linear receiver was proposed in which a fuzzy inference system was inserted into the LMS algorithm. The motivation to use fuzzy logic was for convergence and stability reasons. The LMS algorithm was modified with a fuzzy logic controlled adaptive step size and partial update. This modified algorithm was then used in simulations of noise cancellation in a space-time joint direct-sequence (DS) CDMA system in a dynamic fading multipath channel. It was shown that the performance of the authors' modified LMS algorithm was superior to that of the LMS algorithm.

The proposed algorithm in [12] was further used in [13] for channel estimation and tracking in OFDM systems for a time-variant channel. Simulation results indicated that the modified LMS algorithm had lower steady-state Mean-Square Error (MSE) and faster convergence speed compared to the ordinary LMS algorithm.

In [14], an adaptive neuro-fuzzy inference system (ANFIS) was evaluated for channel estimation in OFDM systems. The ANFIS uses a hybrid learning algorithm based on the LS and the gradient descent methods in order to train the parameters of the membership functions of a TSK fuzzy inference system. Clustering is used in order to group data and from this generate the TSK fuzzy rule-base. From the results of the computer simulations performed in [14] it can be seen that the ANFIS performs very closely to that of the MMSE algorithm albeit with less computational complexity.

2.2 Channel Equalization

Channel equalization is the process of removing the degradation caused by the channel with the aim of reconstructing the transmitted data. The wireless channel is time-variant, and in this kind of channel, non-linear distortion is usually encountered. Due to linear equalization not performing very well in these channels, efficient equalization should be both adaptive and non-linear.

The first results that seem to appear in the literature with regards to fuzzy logic applications in channel equalization are [15, 16]. Here fuzzy adaptive filters based on both the LMS and the Recursive Least Squares (RLS) algorithms are constructed and applied to channel equalization. Initially the fuzzy sets are defined over the filter input space, after which linguistic information from human experts and numerical data are combined and incorporated into the filter. The algorithms are then used to update the free parameters. The objective of using fuzzy adaptive filters is to improve the adaptation speed of the algorithms with the extra help of linguistic inference rules. The results indicate that the bit error rates of the fuzzy equalizer is close to that of the optimal equalizer.

Whereas Wang and Mendel in [15, 16] assumed a fixed delay, in [17] the delay and the membership functions are derived from training data. Two membership functions and two rules are defined that correspond to the input of the channel, which is binary, i.e. it takes on two different values. Together they form an output grounded on the conclusion of a fuzzy rule, from which an independent decision results. Test symbols are transmitted and the correlation between the de-

ISBN: 978-989-95079-6-8

sired outputs and the received signals are determined and form weights assigned to fuzzy outputs. A weighted sum of all the rules is used in the fuzzy inference, after which the defuzzified value is fed through a thresholding device for the final decision of the equalizer. The results from the simulations performed in [17] show that the fuzzy logic equalizer outperforms the LMS in non-linear channels as well as a neural network equalizer using the backpropagation algorithm. The fuzzy logic equalizer also needs fewer training samples compared to the LMS in linear channels for the same error performance.

The Wang-Mendel RLS Fuzzy Adaptive Filter [16] is extended in [18] to a complex fuzzy filter that can handle complex channel models and signals. In [19] human expert knowledge and heuristic reasoning are replaced altogether by a Multilayer Perceptron (MLP) preprocessor unit. The MLP unit consists of a 3-layer network, the training of which provides information to the fuzzy logic system. The LMS algorithm, less computationally complex than the RLS algorithm, is then used to update the free parameter of the system.

In [20] a Bayesian equalization architecture has been developed by using a fuzzy adaptive filter construction as in [15]. The adaptive equalization is visualized as a classification problem in which an observation vector is mapped to signal constellations. In contrast to a channel equalizer such as the maximum likelihood sequence estimation (MLSE) there is no need to include a channel estimator, thus making the equalization process less computationally complex. The derived fuzzy filter function in [20] using fuzzy basis functions [21], product inference, a center of gravity (COG) defuzzifier, and Gaussian membership functions, is able to properly represent the Bayesian decision solution. The performance of the fuzzy equalizer is close to the Bayesian, with the advantage of reduced computational complexity.

In [22] a further development of the fuzzy adaptive filters is presented: the type-2 fuzzy adaptive filter, which based on an unnormalized type-2 TSK fuzzy logic system using a training sequence. This is used to implement the Bayesian equalizer with a decision feedback structure, reducing the complexity of the equalizer compared to that of the transversal equalizer (TE). In contrast to a traditional transversal equalizer or filter, the decision feedback equalizer uses previous detector decisions to cancel intersymbol interference. The type-2 fuzzy sets [23] is an extension of ordinary fuzzy sets in that the membership grades are fuzzy as well. It is shown that an unnormalized output type-1 TSK fuzzy logic system is able to implement a Bayesian equalizer for a time-invariant channel, albeit being model free and not based on a Gaussian probability model. This is further developed into a more generalized form with the type-2 fuzzy adaptive filter to accommodate a time-varying channel.

To a lesser extent, work has also been conducted on blind methods for channel equalization. Blind methods are distinguished by only using information contained in the received signal, thus making both channel estimation and training data unnecessary, with the advantage of higher spectral efficiency. However, this also means that they are strongly dependent on the obtained statistical data.

In [24] the fuzzy-C-means (FCM) algorithm is used to perform joint equalization and demodulation of a signal modulated with the Quadrature Amplitude Modulation (QAM) scheme. The receiver mapping the signal onto a set of symbols can be reduced to a classification problem, rendering a clustering analysis useful. The aim of clustering analysis is to classify objects into groups or classes (clusters) with the objects in the same group having similarities. The FCM algorithm is an unsupervised algorithm, i.e., no external information outside the data itself is needed on which the algorithm operates. The membership functions are used as a measure of what degree the data is connected to the clusters, which in this particular application depends on the amplitude and the phase distances between the received symbols, i.e. the signal constellation points. The results, for a test environment with a static channel, indicate that the algorithm converges very quickly, is robust and has lower computational complexity than that of conventional MLSE receivers. One of the problems with the FCM algorithm however is that it forces points seriously degraded by noise to belong to one or more clusters with some degree, instead of giving it low or no membership in any cluster.

The FCM algorithm in [24] is improved on in [25] by introducing what is referred to as a fuzzy possibilistic C-means (FPCM) assisted blind channel equalization scheme for time-varying channels. The FPCM algorithm solves the problem with the FCM algorithm above by making it less sensitive toward the highly noisy symbol samples. The scheme also allows the receiver to take into account cluster center information in previous data, thereby improving the accuracy of the cluster centers with more data samples. However, it is also suggested that a forgetting factor might be taken into consideration to reduce the significance of cluster centers in previous data in a time-varying channel. Due to a rather large coherence time in high speed wireless transmission systems, it is stated in [25], that the accumulation of cluster centers is feasible even under time-variant conditions. The FPCM outperforms the FCM algorithm due to the former's capability of rejecting the interference of data seriously degraded by noise. It is also shown that the performance of both algorithms depend on the amount of data involved, as can be expected.

In [26] and [27] a blind equalization algorithm based on a fuzzy neural network is outlined. Equalization is performed with a combination of channel estimation and a fuzzy neural network classifier. The algorithm first blindly estimates the channel by using the fourth cumulants of the received sequences [28]. Afterward, an approximate deconvolution is carried out. The output from the deconvolution is then fed into a fuzzy neural network classifier. Simulations undertaken in time-invariant channels with 64-QAM, indicate that the convergence speed as well as BER are improved compared to that of a feedforward neural network blind equalization algorithm.

Similarly, in [29] an improvement of a feedforward neural network blind equalization algorithm is proposed by using a fuzzy neural network consisting of an input layer, a fuzzification layer, a rule layer, a normalization layer and a defuzzification layer with the aim of improving the convergence rate. Simulation results for 16-QAM shows that the fuzzy neural network has faster convergence speed and lower BER compared to that of the feedforward neural network blind algorithm.

ISBN: 978-989-95079-6-8

2.3 Decoding and Equalization

In contrast to conventional communication systems where encoding/decoding and channel equalization are performed separately, turbo equalization schemes combine the two mechanisms. This combination is carried out by iterating the equalizer and the channel decoder on the same set of received data. Since turbo equalizers have been implemented with the Bayesian algorithm, and it was shown that a TSK fuzzy logic system is able to implement a Bayesian model in [22], a turbo-fuzzy equalization approach should be feasible. Such an approach was introduced in [30], where a turbo equalizer using fuzzy filters is proposed. The fuzzy turbo equalizer is introduced with the motivation that fuzzy filters could deal with uncertainty characterized by impulse noise, and also has lower computational complexity compared to the Bayesian equalizer. To adapt the parameters of the fuzzy equalizer, the back-propagation algorithm is used.

A critique against the turbo-fuzzy equalizer in [30] can be found in [31] where it is stated that the turbo-fuzzy equalizer in [30] is unable to use the *a priori* information provided by the decoder, hence not having an iterative extrinsic information exchange between the fuzzy system and the decoder. This is improved on in [31] and [32] where fuzzy turbo equalization schemes with low complexity are proposed.

In [32] a turbo equalization scheme, based on the radial basis functions (RBF), is proposed by using an extended FCM algorithm. An emphasis is made on the low computational complexity this scheme provides compared to a turbo equalization scheme based on the Jacobian RBF in the context of binary phase-shift keying (BPSK) modulation in a Rayleigh-fading channel. Simulation results show that the scheme proposed performs closely to that of the Jacobian RBF based turbo equalization scheme but with a significant reduction in computational complexity.

In [31] the Jacobian RBF turbo equalization scheme is modified by using the same Bayesian equalization architecture based on a fuzzy adaptive filter structure as introduced in [20]. The simulations are performed for BPSK and QAM in a Rayleigh fading channel, and indicate that the proposed fuzzy adaptive filter TEQ scheme considerably reduces the computational complexity with only a slight degradation in performance compared to the Jacobian RBF turbo equalization scheme, providing a trade-off for highly low complexity-oriented circuit implementation.

3 Conclusions

In this paper we have traced the research being conducted over the last two decades, leading up to current research, in which the usage of fuzzy logic in wireless communications has yielded successful results. To the best of our knowledge, this is the first such study of its kind. The three areas in wireless communications focused on in this paper have been: channel estimation, channel equalization and decoding.

In channel estimation, the fuzzy based methods to have been applied have ranged from fuzzy tracking based on Kosko's fuzzy associative memory models and the TSK model, as well as fuzzy logic used in combination with adaptive algorithms such as the LMS and RMS algorithms or a neuro-fuzzy inference system. The neural network and adaptive algorithms are commonly used in order to train the parameters of membership functions in a fuzzy inference system.

In channel equalization the research in fuzzy adaptive filters, both type-1 and more recently type-2 TSK fuzzy logic systems, from Wang and Mendel have been highly influential. These fuzzy adaptive filters are able to use input from both human experts and/or training data. Using as a foundation the adaptive fuzzy filters, a Bayesian architecture has been developed which incorporates fuzzy basis functions and Gaussian membership functions, being able to properly represent the Bayesian decision solution. Another category in channel equalization is the blind methods which uses variants of the fuzzy-C-means algorithm or a fuzzy neural network. Turbo equalization has either been based on the Bayesian equalization architecture or the clustering approach with the fuzzy-C-means algorithm.

Conclusions that can be drawn from the research collated and presented in this paper with regards to the main benefits of using fuzzy logic based methods are:

- Fuzzy logic based methods particularly perform well under non-linear and time-variant conditions, where adaptive techniques have to be employed.
- When dealing with complex models that are not completely known and varying with time, fuzzy logic based methods can be used for faster convergence and reduced complexity with a slight degradation in performance compared to that of standard methods.
- When human expert knowledge is available, a fuzzy approach is highly suitable to incorporate this knowledge to complement available numerical data.

Building on the study we have presented in this paper, there are a few research areas that we consider merit further attention. These are: fuzzy adaptive equalization techniques for time-varying MIMO-channels, and fuzzy power control in MIMO-OFDM systems. Furthermore, another interesting area in which research is being conducted, is in cognitive radio [33]. Cognitive radio is an intelligent wireless communication system that adapts to its environment with the purpose of improving the spectrum efficiency. Both signal processing and machine learning techniques are of interest in cognitive radio, with, e.g., game theory being a commonly used method to model the transmit-power control problem. Research has also been conducted in fuzzy based game theory with the aim of application in cognitive radio in [34].

The purpose of this paper has been to give a background to common problems in wireless communication systems. Moreover, a survey of relevant research has been presented in wireless communication systems in which fuzzy methods have been used successfully. The aim has been to give a condensed and clear overview of conducted research as well as highlighting the common features of the problems in which fuzzy logic has been used in order to discern future areas to be investigated further by the ambitious researcher.

References

[1] L-X Wang and Jerry M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427, Nov-Dec 1992.

ISBN: 978-989-95079-6-8

[2] Antti Niemi, Jyrki Joutsensalo, and Tapani Ristaniemi. Fuzzy channel estimation in multipath fading CDMA channel. *The 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2:1131–1135, 2000.

[3] Andrew J Viterbi. *CDMA: Principles of Spread Spectrum Communication*. Prentice Hall PTR., 1995.

[4] Bart Kosko. *Neural Networks and Fuzzy Systems. A Dynamic Systems Approach to Machine Intelligence.* Prentice Hall, 1992.

[5] A. R. S. Bahai, B. R. Saltzberg, and M. Ergen. *Multi Carrier Digital Communications: Theory and Applications of OFDM.* Springer, 2004.

[6] M. Michele and M. Umberto. A comparison of pilot-aided channel estimation methods for OFDM systems. *IEEE Transactions on Signal Processing*, 49(12):3065–3073, December 2001.

[7] Jian Zhang, Zhi ming He, Xue gang Wang, and Yuan yuan Huang. A TSK fuzzy approach to channel estimation for OFDM systems. *Journal of Electronic Science and Technology in China*, 4(2), June 2006.

[8] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, Jan-Feb 1985.

[9] L-X Wang. Fuzzy systems are universal approximators. *IEEE International Conference on Fuzzy Systems*, pages 1163–1170, March 1992.

[10] Jian Zhang, Zhi ming He, Xue gang Wang, and Yuan yuan Huang. TSK fuzzy approach to channel estimation for MIMO-OFDM systems. *IEEE Signal Processing Letters*, 14(6), June 2007.

[11] D. Gesbert, M. Shafi, Shiu Da-shan, P.J. Smith, and A. Naguib. From theory to practice: an overview of MIMO space-time coded wireless systems. *IEEE Journal on Selected Areas in Communications*, 21(3):281–302, April 2003.

[12] Hsuan-Yu Lin, Chia-Chang Hu, Yu-Fan Chen, and Jyh-Horng Wen. An adaptive robust LMS employing fuzzy step size and partial update. *IEEE Signal Processing Letters*, 12(8):545–548, August 2005.

[13] Jyh-Horng Wen, Cheih-Yao Chang, Gwo-Ruey Lee, and Ching-Yao Huang. Ofdm channel prediction using fuzzy update lms algorithm in time-variant mobile channel. *IEEE 64th Vehicular Technology Conference*, pages 1–5, September 2006.

[14] M. Nuri Seyman and Necmi Taspinar. Channel estimation based on adaptive neuro-fuzzy inference system in OFDM. *IEICE Trans. Commun.*, E91-B(7):2426–2430, July 2008.

[15] L-X Wang and Jerry M. Mendel. Fuzzy adaptive filters, with application to nonlinear channel equalization. *IEEE Transactions on Fuzzy Systems*, 1(3):161–170, August 1993.

[16] L-X Wang and Jerry M. Mendel. An RLS fuzzy adaptive filter with applications to nonlinear channel equalization. *Second IEEE International Conference on Fuzzy Systems*, 2:895–900, Mar-Apr 1993.

[17] P. Sarwal and M. D. Srinath. A fuzzy logic system for channel equalization. *IEEE Transactions on Fuzzy Systems*, 3(2):246–249, May 1995.

[18] K.Y. Lee. Complex RLS fuzzy adaptive filter and its application to channel equalization. *Electronics Letters*, 30(19):1572–1573, September 1994.

[19] Chee Seng Tang and Chin Leonard. Using multilayer perceptron fuzzy adaptive filter in non-linear channel equalization. *IEEE Global Telecommunications Conference, Globecom '95*, 2:884–887, November 1995.

[20] Sarat Kumar Patra and Bernard Mulgrew. Efficient architecture for bayesian equalization using fuzzy filters. *IEEE Transactions on Circuits and Systems*, 45(7):812–820, July 1998.

[21] L-X Wang and Jerry M. Mendel. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Transactions on Neural Networks*, 3(5):807–814, September 1992.

[22] Qilian Liang and Jerry M. Mendel. Equalization of nonlinear time-varying channels using type-2 fuzzy adaptive filters. *IEEE Transactions on Fuzzy Systems*, 8(5):551–563, October 2000.

[23] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Inform. Sci.*, 8:199–249, 1975.

[24] Eugenio Costamagna, Lorenzo Favalli, and Pietro Savazzi. Blind recovery of M-QAM signals for fading channels using the fuzzy-C-means algorithm. *IEEE 49th Vehicular Technology Conference*, 1:516–520, May 1999.

[25] Jenn-Kaie Lain and Chia-Tai Huang. FPCM-Assisted blind channel equalization of M-QAM signals for time-varying channels. *IEEE International Conference on Networking, Sensing and Control*, pages 787–791, April 2007.

[26] Shen Fang, Sun Yunshan, Zhang Liyi, Li Yanqin, and Li He. Blind equalization algorithm based on fuzzy neural network in QAM system. *2nd IEEE Conference on Industrial Electronics and Applications*, pages 484–487, May 2007.

[27] Sun Yunshan, Zhang Liyi, Li Yanqin, Li He, and Yan Junwei. A QAM blind equalization algorithm based on fuzzy neural network. *IEEE International Conference on Control and Automation*, pages 1420–1423, May-June 2007.

[28] D. Gansawat and T. Stathaki. Blind identification of volterra models using minimax optimization with higher order cumulants. *Proceedings. 7th International Conference on Signal Processing*, 1:164–167, Aug-Sep 2004.

[29] Yongzing Hou and Qinghua Bu. A new blind equalization algorithm based on the fuzzy neural network controller. *11th IEEE International Conference on Communication Technology Proceedings*, pages 466–469, Nov 2008.

[30] Murilo Bellezoni Loiola, Moiss Ribeiro, and Joao Marcos T. Romano. A turbo equalizer using fuzzy filters. *14th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, pages 695–704, Sep-Oct 2004.

[31] Jenn-Kaie Lain, Kun-Huang Kuo, and Yuan-Kai Zhang Liao. Low-complexity fuzzy filter-assisted turbo equalization for dispersive rayleigh-fading channels. *European Transactions on Telecommunications*, 19(2):137–148, 2008.

[32] Kun-Huang Kuo, Jenn-Kaie Lain, and Chia-Tia Huang. Reduced-complexity radial basis function-assisted turbo equalization for dispersive rayleigh-fading channels. *IEEE International Conference on Systems, Man, and Cybernetics*, 5:3691–3696, October 2006.

[33] Simon Haykin. Cognitive radio: brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2):201–220, February 2005.

[34] Elisabeth Rakus-Andersson, Maria Salomonsson, and Hang Zettervall. Ranking of weighted strategies in the two-player games with fuzzy entries of the payoff matrix. *Eighth International Conference on Hybrid Intelligent Systems*, pages 186–191, September 2008.

ISBN: 978-989-95079-6-8

Learning Fuzzy Rule Based Classifier in High Performance Computing Environment

Vinícius da F. Vieira[1], Alexandre G. Evsukoff[1], Beatriz S. L. P. de Lima[1] and Sylvie Galichet[2]

[1]COPPE/Federal University of Rio de Janeiro, Brazil
Rio de Janeiro, Brazil
[2]LISTIC/Polytech'Savoie, University of Savoie,
BP 80439, 74944 Annecy le Vieux, France

Email: vfvieira@gmail.com, alexandre.evsukoff@coc.ufrj.br, bia@coc.ufrj.br, sylvie.galichet@univ-savoie.fr

Abstract *— An approach to estimate the number of rules by spectral analysis of the training dataset has been recently proposed [1]. This work presents an analysis of such a method in high performance computing environment. Two approaches for parallel implementation of the method were studied considering the structure selection genetic algorithm and the spectral decomposition. The results show that both approaches have allowed to reduce considerably the overall processing time.*

Keywords — Fuzzy rule-based classifier; genetic algorithm; spectral clustering; high performance computing.

1 Introduction

Fuzzy rule-based classifiers have been widely studied and many approaches have been reported in literature [2]. One of the main problems in the design of interpretable fuzzy rule-based models is the determination of an adequate number of rules, which is closely related to the complexity of the problem. An approach to estimate the number of rules by spectral analysis of the training dataset has been recently proposed [1]. The method employs also a genetic algorithm for structure selection, which selects the input variables and defines the number of fuzzy membership functions in the domain of each selected input variable. Although this method has obtained interpretable fuzzy rules, with good accuracy, it is very time consuming due to the numerical complexity of both the spectral decomposition and the genetic algorithm.

This work presents an analysis of the method proposed in [1] in high performance computing environment. Two approaches for parallel implementation of the method were studied considering the structure selection by the genetic algorithm and the spectral decomposition. In the first approach, only the genetic algorithm is implemented in parallel while the spectral decomposition is computed in serial. In the second approach both the genetic algorithm and the spectral decomposition are run in parallel in a two levels parallelization scheme.

This paper is organized as follows. Next section introduces the notation and parameters of the fuzzy modeling approach. In section 3 the fuzzy rule induction algorithm [1] is reviewed. In section 4 the parallel implementation is described. In section 5 the results are discussed and the conclusions are drawn in last section.

2 Fuzzy Symbolic Modeling

Consider a dataset composed of N input-output observed samples $T = \{(\mathbf{x}(t), y(t)),\ t = 1...N\}$, where $\mathbf{x} \in \mathcal{R}^p$ is a vector of input variables and $y \in \mathcal{N}$ is the output variable where a value $y = j$ refers to a class $B_j \in \mathcal{B}$.

The fuzzy symbolic model (FSM) relates a set of input symbols $\mathcal{A} = \{A_i, i = 1...n\}$ to the set of classes $\mathcal{B} = \{B_j, j = 1...m\}$ as rules in the form $A_i \rightarrow B_j$ read as:

$$\textit{if } \mathbf{x}(t) \textit{ is } A_i \textit{ then } y(t) \textit{ is } B_j. \quad (1)$$

As usual, the fuzzy model computes an approximation $\hat{y}(t) = \hat{f}(\mathbf{x}(t))$ in three steps, which are implemented as the operators:

Fuzzification:

$$\hat{\mathbf{u}}(t) = F(\mathbf{x}(t), \Xi). \quad (2)$$

Inference:

$$\hat{\mathbf{v}}(t) = I(\hat{\mathbf{u}}(t), \Phi). \quad (3)$$

Defuzzification:

$$\hat{y}(t) = D(\mathbf{v}(t)). \quad (4)$$

The operators and their parameters, which should be computed by the learning algorithm, are presented in the following [1].

2.1 Fuzzification

The compound symbol A_i in rule (1) represents a region on the input variables domain. The membership function $\mu_{A_i}(\mathbf{x}(t))$ is computed by the conjunction of p elementary symbols $\mu_{A_k^i}(x_k(t)), k = 1...p$ as:

$$u_i(t) = \mu_{A_i}(\mathbf{x}(t)) = \prod_{k=1...p} \mu_{A_k^i}(x_k(t)) \quad (5)$$

The component A_k^i in (5) is associated to an elementary fuzzy set defined in the one dimensional domain of the variable x_k. The fuzzy membership functions are supposed to be equally spaced Gaussian functions, and only the

ISBN: 978-989-95079-6-8

number of fuzzy sets in each input variable fuzzy partition is necessary to define fuzzy membership functions [1].

The table $\Xi = \{\xi_{ik}, i = 1\ldots n, k = 1\ldots p\}$ is the output of the rule induction algorithm, presented in Section 3.2. An element $\xi_{ik} \in \Xi$ defines which symbol in the partition of the variable x_k should be associated to the component A_k^i in the rule i. A value $\xi_{ik} = j$ indicates that the fuzzy set A_{kj} should be used as the component A_k^i in the rule A_i (cf. (5)).

2.2 Inference

Rule weights have been shown to allow more flexibility to the fuzzy model [3][4]. The inference operator is defined by the rule weights matrix $\mathbf{\Phi} \in [0,1]^{n \times m}$, of which each component $\varphi_{ij} = \mu_\Phi(A_i, B_j)$ represents the confidence of the rule $A_i \rightarrow B_j$. The fuzzy inference is computed by the sum-product composition operator:

$$\hat{\mathbf{v}}(t) = \hat{\mathbf{u}}(t).\mathbf{\Phi} \tag{6}$$

The rule weights matrix are computed as the solution of a bounded quadratic optimization problem, as discussed in section 3.3.

2.3 Defuzzification

In classification problems, defuzzification computes the class output index. Generally the maximum rule is used, such that the class index is computed as the component with the greatest membership value:

$$\hat{y}(t) = j : v_j(t) = \max(\hat{\mathbf{v}}(t)) \tag{7}$$

The design of the FSM is presented next.

3 Induction of Fuzzy Rules

The rule induction algorithm is based on the spectral analysis of the training dataset to determine the number of fuzzy rules. The rule-base weights are then computed by solving a bounded quadratic optimization problem.

3.1 Spectral Analysis

Consider a graph associated to the training dataset, of which the set of nodes represent the input variable samples and the set of edges is defined by the $N \times N$ adjacency (or affinity) matrix $\mathbf{A}$. The graph is undirected and weighted, hence the affinity matrix is symmetric, real valued and its elements represent the similarity between two input variable samples. In this work, the similarity metric is computed by the Gaussian function, such that each element of affinity matrix is computed as:

$$a_{ij} = \begin{cases} \exp\left(-\dfrac{\|\mathbf{x}(i) - \mathbf{x}(j)\|^2}{2\sigma^2}\right), & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where σ is a dispersion parameter that controls the spread of the similarity function, fixed in this work at $\sigma = 1$. The $N \times N$ matrix $\mathbf{D}$ is a diagonal matrix whose elements are the degree of the nodes of the graph, computed by the sum of similarities of neighbors of each node:

$$d_{ii} = \sum_{j=1..N} a_{ij} \tag{9}$$

The spectral analysis of the graph is based on the normalized Laplacian matrix, computed as [5][6]:

$$\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} \tag{10}$$

The rule induction algorithm is based on the spectral decomposition of the Laplacian matrix as:

$$\mathbf{L} = \mathbf{Z}\mathbf{\Lambda}\mathbf{Z}^{\mathrm{T}} \tag{11}$$

The number of positive eigenvalues of the Laplacian is used as an estimation of the number of rules [1].

3.2 The Rule Induction Algorithm

The combination of one-dimensional fuzzy sets is easier to understand than a multi-dimensional fuzzy set. Thus, the idea of the rule induction algorithm is to run a clustering algorithm on the dataset and then to associate a rule to each cluster. For each cluster, the nearest combination to the cluster center, among all possible combinations of fuzzy sets is chosen as the rule to be included in the model.

In the spectral clustering algorithm, the standard k-means algorithm is applied over the matrix $\widehat{\mathbf{Z}}$, computed as the first K columns of the eigenvector matrix $\mathbf{Z}$. Recall that the columns of $\mathbf{Z}$ are ordered according to the corresponding eigenvalues, such that the first columns correspond to the largest eigenvalues.

The rule induction algorithm is run within an outer loop for structure selection, which defines both the variables that should be used by the model and the number of fuzzy sets in each variable's domain. At each iteration, the structure selection algorithm provides the candidate structure represented by the vector $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_p]$, where each component $2 \leq \gamma_k \leq \gamma_{\max}$ is the number of fuzzy sets in the fuzzy partition of variable x_k and a value $\gamma_k = 1$ indicates that the variable x_k should not be considered by the model. The rule induction algorithm is sketched out in Algorithm 1.

Algorithm 1: Rule induction

input: $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_p]$; $T = \{(\mathbf{x}(t), \mathbf{y}(t)), t = 1...N\}$

output: $\Xi = \{\xi_{ik}, i = 1\ldots n, k = 1\ldots p\}$;

```
01 begin
02    compute the matrix A    //Equation (8)
03    compute the matrix D    //Equation (9)
04    compute L = D^{-1/2} A D^{-1/2}
05    compute L = Z Λ Z^T
06    n ← K : λ_i ≥ δ, i = 1...K  //estimate n rules
07    W ← clustering(n)       //assign rules to clusters
08    for i = 1...n
09       for k = 1...p
10          if (γ_k > 1)
11             ξ_ik = j* : α_kj* = arg min(α_kj − w_ik);
12 end
```

ISBN: 978-989-95079-6-8

The set of rules issued by the rule induction algorithm defines the premises of rules (1). The same set of rules is used for all classes, with the classification being assigned by the rule weights, which are computed as described next.

3.3 Rule Weights Optimization

The rule weight matrix can be written as $\mathbf{\Phi} = [\boldsymbol{\varphi}_1 | \ldots | \boldsymbol{\varphi}_m]$, where the vectors $\boldsymbol{\varphi}_j$ are the rule confidence weights that relates the rule premise symbol A_i to the rule consequent (class) B_j. Optimal rule weights $\boldsymbol{\varphi}_j$ can be computed in the least squares sense, by solving the following bounded quadratic programming problem for each class B_j [1][7]:

$$\underset{\boldsymbol{\varphi}_j}{minimize}\ \tfrac{1}{2}\boldsymbol{\varphi}_j^{\mathrm{T}}\mathbf{K}\boldsymbol{\varphi}_j - \mathbf{C}^{\mathrm{T}}\boldsymbol{\varphi}_j \qquad (12)$$

$$subject\ to\ 0 \le \varphi_{ij} \le 1,\ i = 1...n$$

where $\mathbf{K} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$ is a strictly positive definite matrix and $\mathbf{C}^{\mathrm{T}} = \mathbf{V}_j^{\mathrm{T}}\mathbf{U}$. The matrix $\mathbf{U} = [\mathbf{u}(1) | \ldots | \mathbf{u}(N)]^{\mathrm{T}}$ is the matrix of which each line is the fuzzy vector computed for each data sample, whose components are computed as (5). The vector $\mathbf{V}_j = [v_j(1) | \ldots | v_j(N)]^{\mathrm{T}}$ contains the memberships values of class B_j for each sample in the training set, such that $v_j(t) = \mu_{B_j}(y(t))$. The bounds attempt to constrain the weights within the $[0,1]$ interval, such that they can be interpreted as fuzzy rule confidence. The bound constrains also avoid high values for the rule weight parameters, providing the smoothness of the solution.

The quadratic optimization problem defined by (12) is widely known and the solution can be computed by efficient numerical algorithms [8].

For a fixed model structure, the more the solution values $\varphi_{ij} \approx 1$, the more certain is the classifier and also the greater is the norm of the weight vector, defined by:

$$\left\|\boldsymbol{\varphi}_j\right\|^2 = \boldsymbol{\varphi}_j^{\mathrm{T}}.\boldsymbol{\varphi}_j,\ j = 1...m \qquad (13)$$

The sum $\sum_{j=1..m}\left\|\boldsymbol{\varphi}_j\right\|^2$ is thus a measure of the amount of certain rules and also can be interpreted as a measure of the classifier quality.

3.4 Structure Selection

The structure selection is implemented by a genetic algorithm (GA). Each individual in the population represents an estimation of the structure vector $\boldsymbol{\gamma} = [\gamma_1,\ldots,\gamma_p]$, which defines the variables that must be included in the model and also the number of fuzzy sets in the domain of each variable.

The GA fitness function is computed as a trade-off between the model's accuracy and capacity according to the required complexity of the problem [1]:

$$R(\boldsymbol{\gamma}) = \frac{1}{2N}\sum_{j=1...m}\left\|\mathbf{U}(\boldsymbol{\gamma})\boldsymbol{\varphi}_j(\boldsymbol{\gamma}) - \mathbf{V}_j\right\|^2 + \frac{1}{2N}\sum_{j=1...m}\left\|\boldsymbol{\varphi}_j(\boldsymbol{\gamma})\right\|^2 \qquad (14)$$

where the firs term represent the model accuracy and the second accounts for model quality according to (13). The factor $\frac{1}{2N}$ is used to rescale the fitness function to the $[0.1]$ interval.

The GA performs simultaneously the selection of input variables to be considered in the model and the number of fuzzy sets in the domain of each (selected) variable. During the GA execution, it is frequently necessary to evaluate an individual that has been previously evaluated. The number of fuzzy rules is computed from the spectral analysis, which is a function only of the selected input variables. It is thus possible that two different individuals select the same input variables such that they will result in the same number of rules. In order to avoid the unnecessary reevaluation of the spectral analysis for individuals related to the same subset of variables, the number of rules computed by the solution of (11) is stored and related to the corresponding subset of variables. Consequently, the eigendecomposition (11) for a new individual is computed only if the corresponding subset of variables has not been evaluated before. The fitness functions for all visited individuals are also stored such that for a new individual, identical to another already evaluated, the fitness function (14) is not reevaluated.

The GA parameters for all tests described in section 5 were the same. The population was fixed at 32 individuals. An uniform crossover operator with 85% crossover rate was used. The mutation rate was set in 5%. The maximum number of iterations was fixed at 150. A value $\gamma_{\max} = 8$ was used to limit the number of fuzzy sets per variable.

All the implementation was made by free source software available on the web. The GA was implemented in Python programming language. The most computationally expensive tasks (spectral analysis and the rule weights optimization) were implemented in C using standard libraries.

4 Parallel Implementation

The structure selection GA is a very intensive computational task, since each candidate evaluation requires the solution of equation (11) and the GA itself requires the evaluation of a number of individuals during several iterations. The modern multi-core hardware platforms allow the reduction of the global processing time in complex application through parallel implementation.

Depending on the application and the available hardware resources, the parallelism can be exploited in different ways, through the partitioning of the problem with respect to data and/or tasks. In data parallelism, the same task is allocated to different processors, where each one of them works with a different portion of the data. This approach applies to problems whose solution involves a great mass of data. In task parallelism, different tasks are allocated to different

ISBN: 978-989-95079-6-8

processors. This approach applies to problems where there is interdependency among the tasks.

When designing a parallel program, the communication among the processors involved in the execution must be considered carefully. There are two main communication paradigms in parallel applications: memory sharing and message passing. The memory sharing parallel programming model is based in the same principle as the serial programming; essentially, a single memory addressing space is used during the program execution flow. The message passing programming is conceptually independent of the hardware platform, operating system and programming language. As the memory addressing spaces are different in each process the concept of message is used. A message is a set of information that can be sent from process to another through a communication channel.

One of the most known and complete libraries for message passing is the Message Passing Interface (MPI). The MPI is a standardized and portable message passing library designed for distributed memory environments, parallel machines and heterogeneous network. In MPI programs, all the parallelism is explicit, i.e., the programmer must code the synchronization points and the message passing calling the functions. Each process, receives a copy of the entire parallel program code.

In this work two parallel strategies are evaluated: in the first one only the GA is implemented in parallel; in the second one, both the GA and the spectral analysis for each individual are processed in parallel in a two levels strategy.

4.1 Parallel Genetic Algorithm

Several features of GAs allow its parallelization in a very natural way and many approaches have been reported in literature [9]. The main approaches are divided in three types: fine-grained single population GA, coarse-grained multiple population GA and single global population master-slave GA.

Fine-grained parallel genetic algorithms use just one population, but there is a special structure that limits the interaction among the individuals. An individual can only compete and operate with individuals that belong to his neighborhood, which is assigned by a predefined topology.

Coarse-grained parallelization model, also known as island model, uses several populations that occasionally change individuals through migration operator. In island GAs, each sub-population evolves as in traditional GAs, excepting for the migration operator among different islands.

The master-slave parallel genetic algorithms, also known as parallel global population GAs, use a single population and the parallelization is done in the evaluation of the individuals. As in a traditional serial GA, each individual competes and operates with any other individuals in the population. In master-slave genetic algorithms, a master process is responsible by the application of the genetic operators to the population (selection, crossing, mutation) while the slave processors are responsible by the individuals' evaluation [10]. The communication among the individuals occurs when the master process sends the individuals to the slave processes and when the slave processes return their fitness values to the master.

In this work, a master-slave parallel implementation was used for the structure selection GA. The adoption of such GA parallel model is useful, since its implementation is independent of the machine architecture and of the communication model among the processes.

4.2 Two Levels Approach: Parallel GA with Parallel SA

The complexity of a full eigenproblem like (11) is $O(N^3)$ in processing time and $O(N^2)$ in storage requirements. In Algorithm 1, the solution of (11) requires the solution of an eigenproblem of the size of the number of records of the dataset N. The two-level strategy considers thus the parallelization of the spectral analysis (SA) for each individual, within the parallel GA.

The schematic representation of the two-level strategy is shown in Fig. 1. In the higher level there is the GA parallelism, where the master performs the operations over the population and sends the individuals to the slaves. In the lower level is the SA parallelism, where each slave shares the eigenproblem solution with other processors.

The two levels strategy is possible due to the concept of communicators groups in the MPI library that allows the definition of modules that encapsulate the communication operations within a group. The communicator identifies the group of processes and the context in which a given operation must be applied. The MPI library provides a standard communicator, called MPI COMM WORLD, which identifies every processes involved in the computation through global identifiers.

The execution of the GA is performed by the processors that belong to Group 0 (cf. Fig. 1.). Each slave processor in Group 0, also belongs to another group, which is responsible for computing the SA. Thus, new sub-communicators must be created, allowing the communication within each group of slave processors.

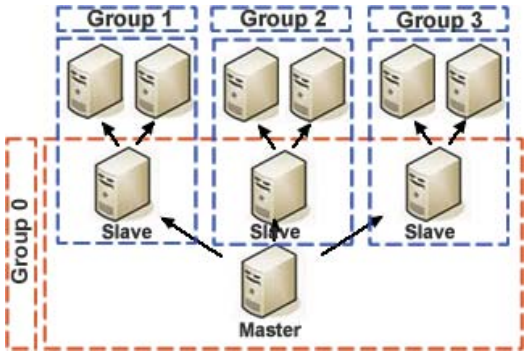


Fig. 1: Two-level Parallel Architecture

There are numbers of software libraries that implements parallel eigenproblem solution using different numerical methods. In this work, SLEPc (Scalable Library for Eingenvalue Problem Computation [11]) was adopted. SLEPc is a software package for solution of large scale problems built over PETSc (Portable, Extensible Toolkit for

ISBN: 978-989-95079-6-8

Scientific Computation), which is a set of libraries of data structure and routines for parallel scientific applications.

The Krylov-Schur solver, which implements the Arnoldi method for eigenvalues problem solution, was adopted in this work. It is probably the most robust solver in SLEPc [11].

5 Results and Discussion

5.1 Performance Results

Four benchmark datasets from the UCI Machine Learning Repository [12] were considered for the performance evaluation. Some synthetic datasets, with 18 variable and different numbers of records, were also considered. These datasets were considered in order to exploit a greater problem size.

Each benchmark evaluation was performed in 10-fold cross validation analysis. The evaluation of each benchmark was repeated 10 times to ensure the GA generalization. The average results for each dataset are shown in Table 1, where Acc. is the average accuracy (computed for the 10-fold cross validation) and Time is the average global time (in seconds) of the serial implementation. For the synthetic datasets just the time results were considered and the time values correspond to the execution time for just one GA generation.

All tests were run an 8-core Linux cluster composed of two machines, each one equipped with two dual-core Intel Xeon processors (1.6GHz) and 4GB DRAM. The serial processing time was computed using only one processor.

Table 1: Performance Results.

Dataset	p	m	N	rules	Acc.	Time
Diabetes	8	2	768	12.5	78.67	1795.72
Ionosphere	32	2	344	26.7	91.47	4752.24
Balance	4	3	625	8.3	81.41	145.69
Cancer	9	2	683	8.4	98.13	1665.62
Synthetic 1	18	-	1000	-	-	99.68
Synthetic 2	18	-	2000	-	-	580.18
Synthetic 3	18	-	3000	-	-	1952.7
Synthetic 4	18	-	4000	-	-	4339.75

As seen in Table 1, the modeling approach has obtained good accuracy with small number of rules results. The accuracy results obtained for the same datasets by other algorithms can be found in [1]. The focus of the analysis in this work is on the processing time and parallelization efficiency.

5.2 Parallel GA Results

The processing time (in seconds) of the parallel GA (with serial SA) implementation are shown in Table 2. The parallel processing time as a percentage of the serial processing time is also shown in Table 2. It can be seen that the parallel GA reduces the global execution time considerably.

Table 2: Execution time for parallel GA (with serial SA).

	2 Proc.		4 Proc.		8 Proc.	
	Time	%	Time	%	Time	%
Diabetes	1194.8	66.5	738.28	41.11	550.75	30.67
Ionosphere	2517.57	52.97	1354.1	28.49	853.56	17.96
Balance	107.29	73.64	88.35	60.64	69.51	47.71
Cancer	1070.35	64.26	696.54	41.81	476.76	28.62
Synthetic1	54.79	54.96	47.66	47.81	54.16	54.33
Synthetic2	317.14	54.66	210.73	36.32	230.89	39.79
Synthetic3	1050.2	53.78	700.89	35.89	620.05	31.75
Synthetic4	2380.84	54.86	1765.83	40.68	1486.28	34.24

The parallel efficiency of the parallel GA algorithm is shown in Fig. 2. The parallel efficiency was computed in the standard way as [13]:

$$\varepsilon_P = 100.\frac{1}{P}\frac{T_S}{T_P} \tag{15}$$

where P is the number of processors, T_S the serial processing time (in seconds) and T_P is the parallel processing time using P processors.

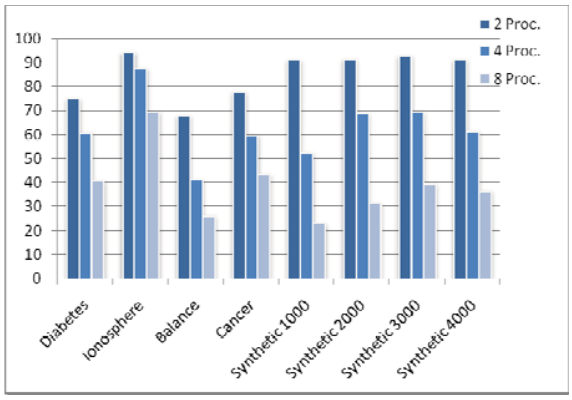


Fig. 2: Parallel Efficiency for Parallel GA With Serial SA.

The parallel efficiency decreases as the number of processors increases, due communication overhead and the load imbalance introduced by the reevaluation avoiding scheme, since the number of evaluated individuals is not the same for all the slave processes. A processor with fewer individual's evaluations will become idle waiting the other processors finish their work to synchronize in a new generation. The resulting load imbalance reflects in a great damage to the overall parallel efficiency.

The results show that this approach is not good for large datasets, since each GA individual evaluation allocates an instance of the Laplacian matrix in memory, making a bad use of the memory resources available.

5.3 Parallel GA with Parallel SA Results

The execution times (in seconds) of the two levels approach for the studied datasets are shown in Table 3 as so as the percentage of the serial processing time.

Table 3: Execution times for the two levels approach.

	2 Proc.		4 Proc.		8 Proc.	
	Time	%	Time	%	Time	%
Diabetes	2185.32	121.6	1303.46	72.58	833.62	46.42
Ionosphere	5704.34	120.0	3401.24	71.57	2949.77	62.07
Balance	185.67	127.4	128.28	88.05	116.71	80.10
Cancer	1628.91	97.79	994.00	59.67	745.43	44.75
Synthetic1	76.01	76.25	40.03	40.15	35.41	35.52
Synthetic2	368.81	63.56	195.59	33.71	160.38	7.64
Synthetic3	1182.08	60.53	615.06	31.49	416.85	21.34
Synthetic4	2537.81	58.47	1300.7	29.97	805.05	18.55

ISBN: 978-989-95079-6-8

The parallel efficiency of the two-level approach for the same datasets are shown in Fig. 3.

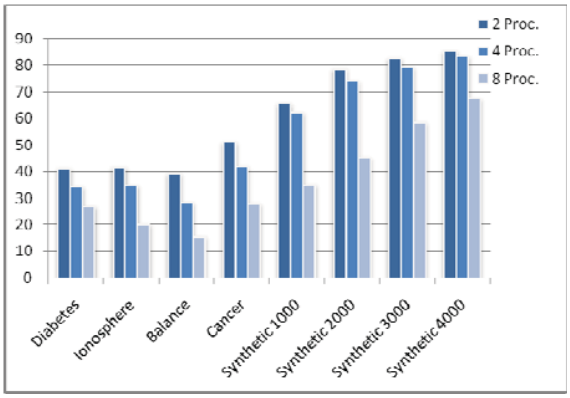


Fig. 3: Parallel Efficiency for the Two-Level Approach.

The results in Table 3 and Figure 3 show that for the four benchmark datasets studied, the global execution time is greater than the results presented for the parallel GA with serial SA. Consequently, the parallel efficiency is lower when compared to the results presented by those approach. This indicates that for small size datasets, the parallelization of the eigenproblem solution is not efficient. For larger datasets, the two levels approach shows more interesting results since the synthetic datasets present better results for the two levels approach when compared to the parallel GA with serial SA. This comparison indicates that, for datasets with a great number of records, the SA parallelization can be a good idea since it reduces the load imbalance introduced by the reevaluation avoiding strategy and makes a better use of the memory available for the method execution, since the Laplacian matrix used in the spectral analysis is sliced for the involved processes.

6 Conclusions

The design of fuzzy symbolic models [1] is focused on the interpretability and the resulting model's structure is selected by a wrapper genetic algorithm that defines the variables that should be included into the model and the number of fuzzy sets in the partition of each variable. The number of rules is determined by a spectral analysis method and rules' weights are optimized.

This work has evaluated two approaches for the parallel implementations of the fuzzy symbolic models. Both approaches have reduced the global execution time of the algorithm when compared to the serial times, which was the main drawback of the previous implementations.

The parallel efficiency results indicate that, even with the load imbalance introduced by the reevaluation avoiding strategy, the GA parallelization is an interesting approach for smaller datasets and the two-level approach is a good approach for larger datasets. Further studies may be done in order to increase the load balance among the processors and evaluate the implementation behavior for larger datasets.

Acknowledgement

The authors are grateful to the Laboratory of Computational Physiology (FISIOCOMP) at Federal University of Juiz de Fora (UFJF), where all the parallel experiments were run. This work has been supported by the Brazilian Research Council CNPq and the Brazilian Agency CAPES under the agreement CAPES/Cofecub.

References

[1] A.G. Evsukoff, S. Galichet, B. S. L. P. de Lima and N. F. F. Ebecken, "Design of interpretable fuzzy rule-based classifiers using spectral analysis with structure and parameters optimization". Fuzzy Sets and Systems (2008), doi: 10.1016/j.fss.2008.08.010.

[2] E. Hüllermeier, "Fuzzy methods in machine learning and data mining: status and prospects", Fuzzy Sets and Systems, vol. 156, no. 3, pp. 387-406, 2005.

[3] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems" IEEE Trans. Fuzzy Systems, vol. 9, no. 4, pp. 506–515, 2001.

[4] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems", IEEE Trans. on Fuzzy Systems, vol. 13, no. 4, pp. 428-435, 2005.

[5] M. Filippone, F. Camastra, F. Masulli and S. Rovetta, "A survey of kernel and spectral methods for clustering". Pattern Recognition, vol. 41, pp. 176-190, 2008.

[6] F. R. K. Chung, Spectral Graph Theory, CBMS Regional Conf. Series in Mathematics., no. 92. American Mathematic Society, 1997.

[7] A. G. Evsukoff, "Learning fuzzy rule-based classifier with rule weights optimization and structure selection by a genetic algorithm", Proceedings of the 2007 IEEE International Conference on Fuzzy Systems, London, 23-26 July, 2007.

[8] B. Schölkopf and A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, 2001.

[9] E. Cantú-Paz, "A Survey of Parallel Genetic Algorithms", Calculateurs Paralleles, vol. 10, 1998.

[10] C. R. Xavier, V. F. Vieira, D. M. S. Martins, R. W. Santos, "Comparing Two Parallel Genetic Algorithms for the Inverse Problem Associated to the Cardiac Bidomain Equations", IEEE 18th International Symposium on Computer Architecture and High Performance Computing, Ouro Preto, 17-20 October, 2006.

[11] V. Hernandez , J. E. Roman, V. Vidal, "SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems", ACM Transactions on Mathematical Software, v.31 n.3, p.351-362, September 2005.

[12] A. Asuncion, and D. J. Newman, "UCI Machine Learning Repository" [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, 2007.

[13] V. Kumar, A. Grama, A. Gupta and G. Karypis, Introduction to Parallel Computing - Design and Analysis of Algorithms, The Benjamin/Cummings Publishing Company, Inc., CA, 1994

ISBN: 978-989-95079-6-8

The Use of Interval Type-2 Fuzzy Logic as a General Method for Edge Detection

Olivia Mendoza[1] Patricia Melin[2] Juan R. Castro[1]

1 School of Engineering, University of Baja California
Tijuana, Mexico
2 Department of Research and Graduate Studies, Tijuana Institute of Technology
Tijuana, Mexico
Email: omendoza@uabc.mx, epmelin@hafsamx.org, jrcastror@uabc.mx

Abstract—*We describe a method for edge detection in digital images based on the morphological gradient and fuzzy logic. The goal is to improve one of the basic methods for edge detection in order to obtain a better result even without applying any filter to the image. The tests were made with a type-1 fuzzy inference system (T1FIS) and with an interval type-2 fuzzy inference system (IT2FIS). We show that the images obtained with fuzzy logic are better than the ones obtained with only the morphological gradient method. In particular the IT2FIS achieved the best results, because of the flexibility to model the uncertainty in the gradient values and the gray ranges for the edge images. In both TIFIS and IT2FIS the membership function parameters were obtained directly from the images; this allows application of the proposed method to images with different gray scales.*

Keywords— digital images, edge detection, image processing, interval type-2 fuzzy logic.

1 Introduction

Edge detection is an intermediate step in image pattern recognition in digital images and has the opposite effect of noise elimination; it consists on emphasizing pixels with gray tones that are different than their neighbours [1]. Some of the existing methods for edge detection have been enhanced with type-1 fuzzy logic [2][3], fractal theory [4][5], neural networks [6][7][8] and recently with interval-valued fuzzy operators [9].

In previous works proposed by the authors, the improvement of the Sobel edge detector using an IT2FIS was achieved [10][11]. The images resulting from these experiments were used as training data for image recognition with modular neural networks, obtaining good results [12][13].

In this paper, an improvement of a traditional edge detection method based on the morphological gradient is presented. This method does not need any filter to obtain the edges, but it is necessary to perform calculations in order to find the relation of each pixel of the image with the eight neighbours pixels around it.

In section two we present a brief description of an edge detection method based in morphological gradient, in section three we explain our contribution consisting on the development and tests with a type-1 fuzzy inference system and an interval type-2 fuzzy inference system; in both cases we apply the same criteria and conditions, for a valid comparison of the results. Section four shows the obtained results with the three methods. The main topic in this paper is the comparison of the results obtained with the IT2FIS, applying different footprint of uncertainty (FOU).

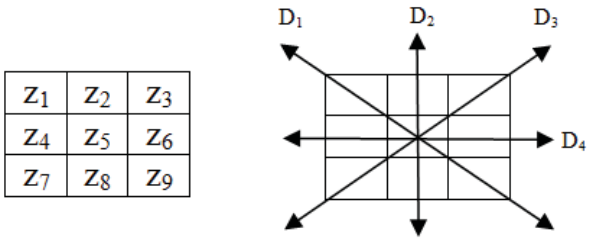
2 Edge detection using morphological gradient

The morphological gradient of a grayscale image can be viewed as the greatest absolute intensity difference between any two pixels within the structuring element, and can be defined with Equation (1) [14].

$$\left| \nabla(f) = \delta_g(f) - \varepsilon_g(f) \right| \tag{1}$$

For the discrete case, we use D_i instead $\nabla(f)$, then if we apply (1) to a 3x3 sample matrix as shown in Fig. 1, we can to obtain coefficients z_i with Equation (2), the possible edge directions D_i with (3). An approximation of (3) can be calculated without the root square using the absolute values of the differences instead the square values (4). The edges E can be calculated with Equation (5) [15][16].

$$\begin{aligned} z_1 &= f(x-1, y-1) & z_2 &= f(x, y-1) \\ z_3 &= f(x+1, y-1) & z_4 &= f(x-1, y) \\ z_5 &= f(x, y) & z_6 &= f(x+1, y) \\ z_7 &= f(x-1, y+1) & z_8 &= f(x, y+1) \\ z_9 &= f(x+1, y+1) \end{aligned} \tag{2}$$



Z1	Z2	Z3
Z4	Z5	Z6
Z7	Z8	Z9

Figure 1: A 3x3 sample matrix with coefficients and edge directions.

ISBN: 978-989-95079-6-8

$$D_1 = \sqrt{(z_5 - z_1)^2 + (z_9 - z_5)^2} \quad (3)$$
$$D_2 = \sqrt{(z_5 - z_2)^2 + (z_8 - z_5)^2}$$
$$D_3 = \sqrt{(z_5 - z_3)^2 + (z_7 - z_5)^2}$$
$$D_4 = \sqrt{(z_5 - z_4)^2 + (z_6 - z_5)^2}$$

$$D_1 = |z_5 - z_1| + |z_9 - z_5| \quad (4)$$
$$D_2 = |z_5 - z_2| + |z_8 - z_5|$$
$$D_3 = |z_5 - z_3| + |z_7 - z_5|$$
$$D_4 = |z_5 - z_4| + |z_6 - z_5|$$

$$E = D_1 + D_2 + D_3 + D_4 \quad (5)$$

The edge image obtained with Equation (5) does not consider any parameters to prevent noise or regions with very low or high contrast. Then some very dark or bright images could be a problem for this edge detector.

3 Edge detection using a fuzzy inference system

The fuzzy approach for edge detection using morphological gradient consists of using Equation (3) or (4) to obtain the gradient in the four directions, and use them as inputs for a fuzzy inference system (FIS) instead the Equation (5). This approach allows us to analyze the gray tones range of any particular image and use them as parameters to the membership functions in the FIS.

3.1 Type-1 fuzzy linguistic variables

The inputs to consider for the T1FIS are the morphological gradients D_i for each pixel with respect to a 3x3 neighbourhood matrix. The membership functions must represent the magnitude of the gradient. In order to adapt the membership functions to the range of gray tones depending on the image, we obtain the maximum, minimum and middle values of D_i, and we use these values for the centre of the membership functions, as shown in the example of Fig. 3.

The σ value for the particular example plotted in Fig. 3 is 20, the Gaussian membership functions for each D input were obtained with (9)(10) and (11), and the center of each function were obtained with (6)(7) and (8). This is the method that we propose to adapt the parameters of the membership functions depending on the contrast level of each image.

$$dlow_i = \min(D_i) \quad (6)$$
$$dhigh_i = \max(D_i) \quad (7)$$
$$dmiddle_i = dlow_i + (dhigh_i - dlow_i)/2 \quad (8)$$

$$\mu(dlow_i) = e^{\frac{-(x-dlow_i)^2}{2(\sigma_i)^2}} \quad (9)$$
$$\mu(dhigh_i) = e^{\frac{-(x-dhigh_i)^2}{2(\sigma_i)^2}} \quad (10)$$
$$\mu(dmiddle_i) = e^{\frac{-(x-dmiddle_i)^2}{2(\sigma_i)^2}} \quad (11)$$

where $\sigma_i = dhigh_i / 8$

For the outputs E (the edges), we select the range [0,255] as the gray tones scale, then we can obtain these membership functions directly with (12)(13) and (14). If some application needs a different gray tones scale, is necessary to adjust the values of *black*, *gray* and *white* of (12), (13) and (14) to the desire scale.

$$\mu_{black} = e^{\frac{-(x-black)^2}{2\sigma^2}} \quad (12)$$

where black=0

$$\mu_{gray} = e^{\frac{-(x-gray)^2}{2\sigma^2}} \quad (13)$$

where gray=255/2

$$\mu_{white} = e^{\frac{-(x-white)^2}{2\sigma^2}} \quad (14)$$

where white=255

where $\sigma = white/8$

In order to plot the membership functions for the inputs D_i of the FIS, as particular example we used the image in Fig. 2. First the image was converted to gray scale, then the following values were calculated, for each variable D.

$dlow_1$=0, $dmiddle_1$=94.7, $dhigh_1$=189.4, σ_1=23.6
$dlow_1$=0, $dmiddle_2$=87.9, $dhigh_2$=175.8, σ_2=21.9
$dlow_1$=0, $dmiddle_3$=89.8, $dhigh_3$=179.6, σ_3=22.4
$dlow_1$=0, $dmiddle_4$=75.6, $dhigh_4$=151.3, σ_4=18.9

Figure 2. Particular image to explain the parameters for the FIS variables.

ISBN: 978-989-95079-6-8

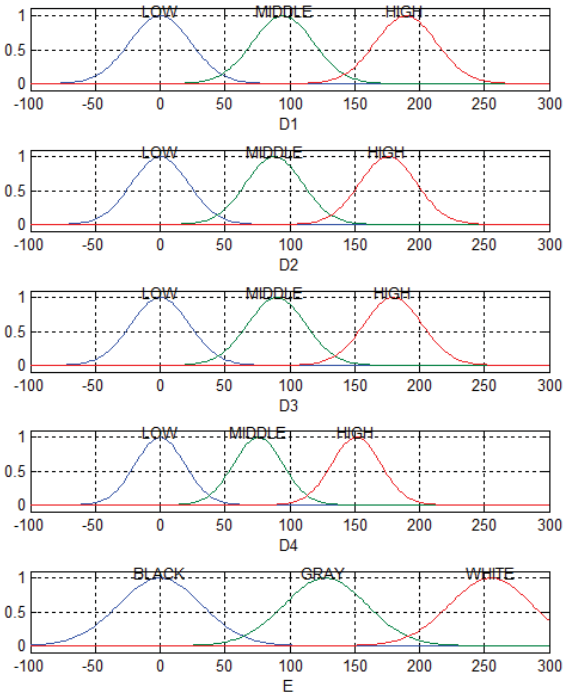


Figure 3: Linguistic variables for the TIFIS corresponding to the image in Fig. 2.

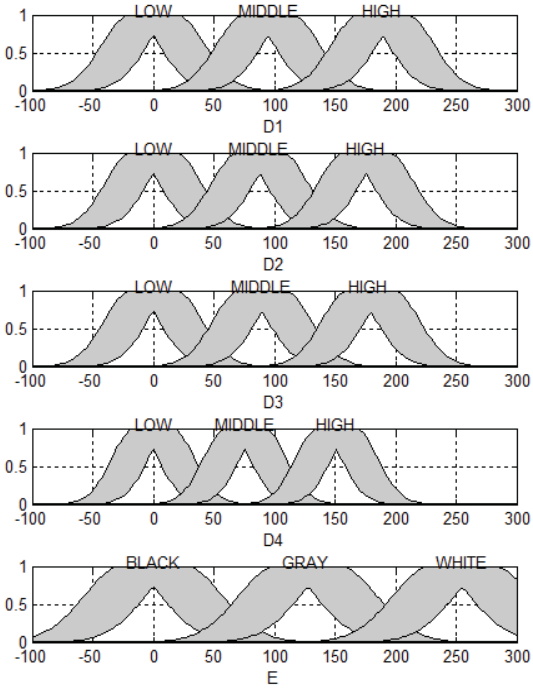


Figure 4: Linguistic variables for the IT2FIS corresponding to the image in Fig. 2, with $FOUd_i=0.4*dmiddle_i$

3.2 *Interval type-2 linguistic variables*

As we previously mentioned, the edge detection method proposed in this paper can be adapted to images with different contrast levels, but particularly the use of interval type-2 fuzzy sets enables us to handle the uncertainty by adding some FOU to the membership functions [17].

To compare the improvement of the edge detector using interval type-2 fuzzy sets, we build an IT2FIS using the same parameters than the TIFS [18]. The membership functions were obtained with Equations (6)-(14), but adding different sizes of FOU.

We made tests using different sizes of the FOU for the input variables D_i, using $dmiddle_i$ as parameter to obtain the values respect to the image contrast, in a proportional manner. The FOU sizes for D_i, then can be obtained using (15).

$$FOUd_i = \varepsilon * dmiddle_i \tag{15}$$

where ε is in (0,1)

The FOU for the output variable E, was calculated in a similar way to the inputs variables. In this case, we obtain the FOU values using *egray* as parameter (16).

$$FOUe = \varepsilon * egray \tag{16}$$

where ε is in (0,1)

The plot in Fig. 4, shows the linguistic variables with interval type-2 membership functions where the value for ε is *0.4*, adding $FOUd_i/2$ for inputs D_i and $FOUe_i/2$ for the output E, at each side of the center of the membership functions; then all the functions are symmetric.

3.3 *The rules*

The task of the fuzzy inference systems is to estimate a value for each pixel, which better describes the relation with its neighborhood.

1. If (D1 is HIGH) or (D2 is HIGH) or (D3 is HIGH) or (D4 is HIGH) then (E is WHITE)
2. If (D1 is MIDDLE) or (D2 is MIDDLE) or (D3 is MIDDLE) or (D4 is MIDDLE) then (E is WHITE)
3. If (D1 is LOW) and (D2 is LOW) and (D3 is LOW) and (D4 is LOW) then (E is BLACK)

The first rule establishes that a high gradient in any direction means an edge, the second rule is similar to the first one, because after many experiments with different images, a medium magnitude of the gradient in any direction results also in an edge, and the third rule is only to confirm the first two, because all derivatives with low magnitude means a homogeneous region in the image, that means there are not edge in this pixel. In Fig. 5 the solution surfaces for the TIFIS and IT2FIS are shown.

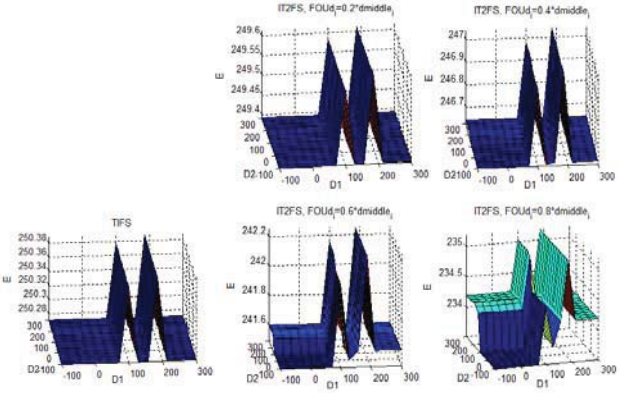


Figure 5: The solution surfaces for the image in Fig. 2, for TIFIS and IT2FIS with different FOU sizes.

ISBN: 978-989-95079-6-8

4 Experimental results

For a better appreciation of the results, we developed some programs in Matlab to obtain the edge image, under the same conditions. In all the images the morphological gradients were obtained with (2) and (3), then using the morphological gradient (MG) the edges were calculated with (5). For the fuzzy methods T1FIS and IT2FIS we use a fuzzy inference system instead of the gradient (5).

We choose Gaussian membership functions for all the variables of the both FIS (TIFIS and ITFIS), only for an easy configuration, because as we explained in section three we used the values obtained with the morphological gradient of each particular image to estimate the parameters for the membership functions.

For the first test, we used a sample of the ORL databases of faces, and selected arbitrary values for the FOU, only to demonstrate that the images obtained using IT2FIS are better that the images obtained with TIFIS. In the example of Fig. 6, we see the changes on the images with different values of the FOU. The best image is the one obtained with FOU=40, while the image obtained with T1FIS looses some details. For all the tests the use of the FIS's improve the results of the MG traditional method.

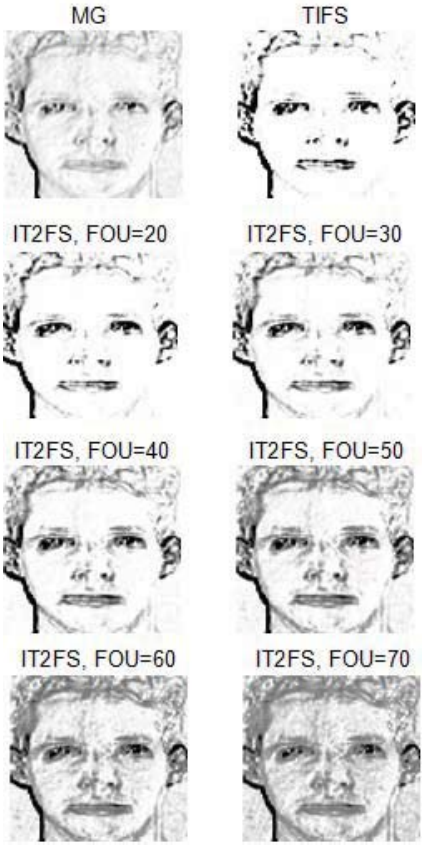


Figure 6: Edges images obtained using Morphological Gradient (MG), and the same method improved with a type-1 fuzzy system (TIFIS), and an interval type-2 fuzzy system (IT2FIS), with arbitrary FOU values.

For a better analysis of the results we plot frequency histograms for each result in Fig. 6. The histograms show that the images obtained with the IT2FIS when the FOU is increased, between 30 and 40, found more pixels corresponding to the edges preserving more details than T1FIS, but when the FOU increases more than 40, the pixels near white decreases, showing more noise in the background.

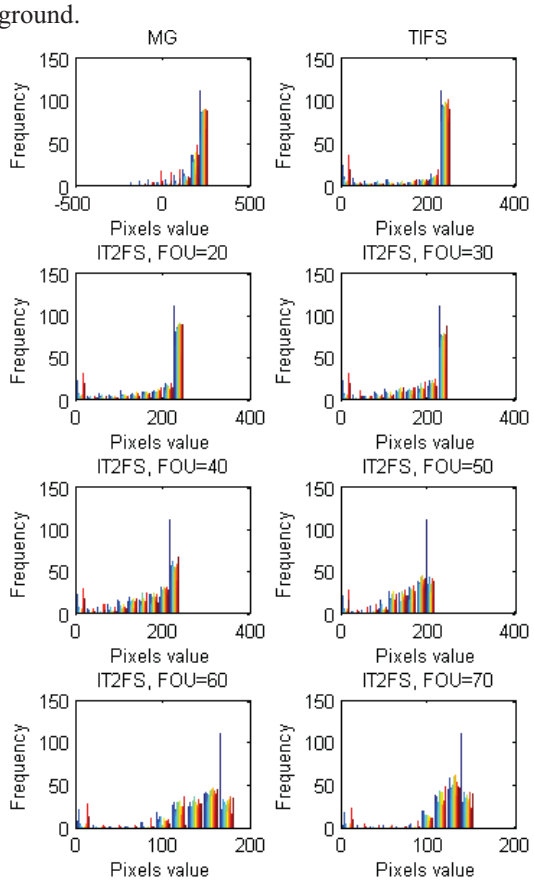


Figure 7: The frequency histograms correspond of the images shown in Fig. 6.

The second test, shown in Fig. 8, was made with an image used as benchmark for edge detection algorithms, and confirms the results obtained in Fig. 6.

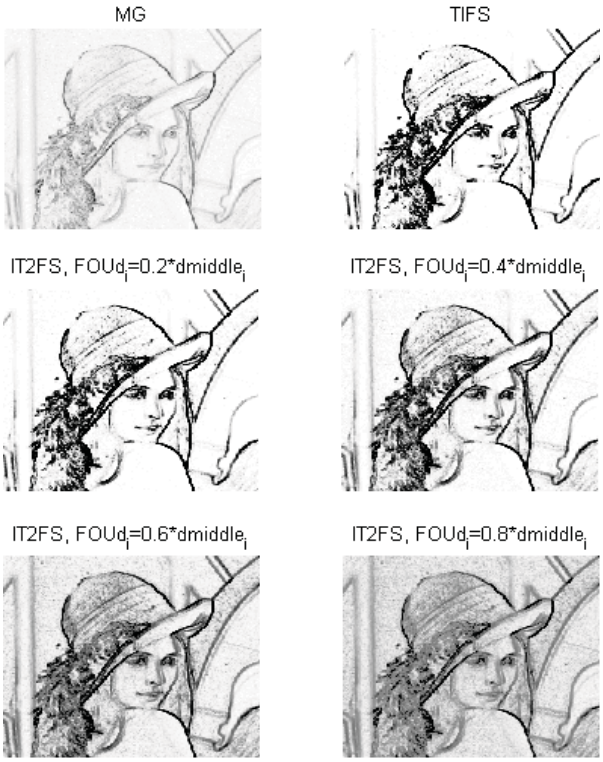


Figure 8: Edges achieved with FOU values obtained directly of the original image with equations (15) and (16).

ISBN: 978-989-95079-6-8

As we can see in Fig. 8 and Fig. 9, the edges can be seen better with the TIFIS method than de morphological gradient method, but more details of the image can be seen with the IT2FIS method when $FOUd_i=0.4*dmiddle_i$. Obviously the images obtained with a FOU near 0 are similar to the images obtained with the T1FIS. The image with $FOUd_i \geq 0.6*dmiddle_i$ show more details, but includes some textures that are not desired for edge detection.

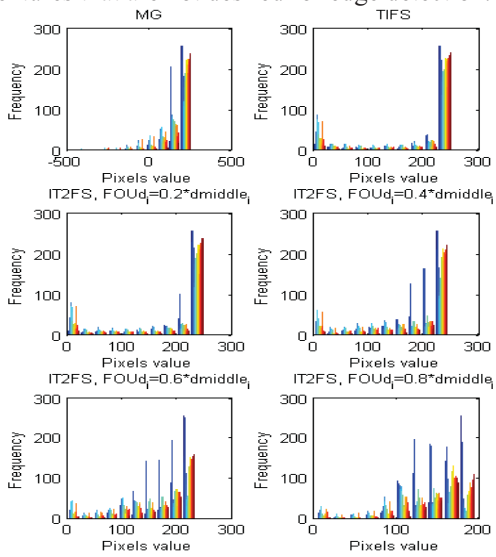


Figure 9: The frequency histograms correspond of the images shown in Fig. 8.

The parameters used for the T1FIS and IT2FIS in the tests of Fig. 8 and Fig. 9, were calculated with equations (6)-(16) and have the following values:

$dlow_1=0,\ dmiddle_1=105.9,\ dhigh_1=211.9,\ \sigma_1=26.4$
$dlow_1=0,\ dmiddle_2=109.6,\ dhigh_2=219.2,\ \sigma_2=27.4$
$dlow_1=0,\ dmiddle_3=107.5,\ dhigh_3=215.0,\ \sigma_3=26.8$
$dlow_1=0,\ dmiddle_4=74.6,\ dhigh_4=149.2,\ \sigma_4=18.6$
$\sigma e=31.8$

In Table I we show the numeric values of the FOUs used in the experiment below.

TABLE I: FOU values for the tests shown in Fig. 8.

FACTOR	$FOUd_1$	$FOUd_2$	$FOUd_3$	$FOUd_4$	$FOUE$
0.2	21.1	21.9	21.5	14.9	25.5
0.4	42.3	43.8	43.0	29.8	51.0
0.6	63.5	65.7	64.5	44.7	76.5
0.8	84.7	87.6	86.0	59.6	102

Figure 10: The original images used for the tests shown in Table II.

To finish the tests in this work, we repeat the same experiment with the images shown in Fig. 10, and compare the results with traditional algorithms [19] like the Canny and Sobel methods, as we can see in Table II.

TABLE II: Tests with Canny and Sobel methods using the Matlab defaults, and the best result using MG+IT2FIS.

Canny	Sobel	MG+IT2FIS
		

For all the tests we used the default parameters of Matlab. One of the images is a sample of the ORL database of faces

ISBN: 978-989-95079-6-8

[20], which is a set of images that are usually considered to compare the behavior of different image processing methods [21].
The other images are some of the typical images used for evaluation of edge detection algorithms, and were obtained them in the USC-SIPI Image Database [21].

In order to compare the results with the Canny and Sobel algorithms, we made modifications on the rules consequents, to obtain white edges and black background as the Matlab algorithms do it.

In all the tests we can observe that the images obtained with our method preserve more details of the original images than the traditional methods; that is good for the human observation. In the other hand, the images obtained with MG+IT2FIS are better to use in an image recognition application, because the training algorithm have more information to learn the image pattern. We can conclude that using interval type-2 fuzzy logic can improve edge detection in benchmark images, and for this reason it can be considered a good alternative for image processing.

5 Conclusions

The experiments presented in this paper show that the use of fuzzy inference systems can improve traditional methods for image processing. In particular, the morphological gradient method for edge detection, considered as one of the basic ones, when enhanced with an interval type-2 fuzzy inference system, results in a combination with satisfactory results. Especially the capability of the IT2FIS to model uncertainty in the morphological gradient and gray tone values for the edges achieved better images than the obtained with the T1FIS, because preserves more details of the original image.

References

[1] M. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer, 1997, A Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 12, pp. 1338-1359.

[2] L. Hua, H.D. Cheng, M. Zhanga, 2007, A High Performance Edge Detector Based on Fuzzy Inference Rules, Information Sciences: An International Journal, Elsevier Science Inc. New York, NY, USA, Volume 177 , Issue 21, pp. 4768-4784.

[3] Y.-H. Kuo, C.-S. Lee, C.-C. Liu, 1997, A New Fuzzy Edge Detection Method for Image Enhancement , Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, Spain , Vol. 2, Issue 1-5, pp. 1069-1074.

[4] K. Revathy, S. Lekshmi,,S. R. Prabhakaran Nayar, 2005, Fractal-Based Fuzzy Technique for Detection of Active Regions From Solar, Journal of Solar Physics Vol. 228, pp. 43-53.

[5] S. E. El-Khamy, M. Lotfy, and N. El-Yamany, 2000, A Modified Fuzzy Sobel Edge Detector, Radio Science Conference, 2000. 17th NRSC '2000. Seventeenth National, Minufiya, Egypt, pp. C32/1-C32/9.

[6] R. Wang, L.-Q. Gao, S. Yang and Y.-H. Chai, 2006, An Edge Detection Method by Combining Fuzzy Logic and Neural Network, Advances in Machine Learning and Cybernetics , Springer , Berlin, pp. 930-937

[7] T.Q. Pham and L.J. van Vliet, 2005, Blocking artifacts removal by a hybrid filter method, Proceedings of the Eleventh Annual Conference of the Advanced School for Computing and Imaging, E.U.A, p.p. 372-377.

[8] K. Suzuki, I. Horiba, N. Sugie, M. Nanki, 2003, Contour extraction of left ventricular cavity from digital subtraction angiograms using a neural edge detector, Systems and Computers Japan, Wiley Japan, p.p. 55-69

[9] H. Bustince, E. Berrenechea, M. Pagola, J.Fernandez, 2008, Interval-valued fuzzy sets constructed from matrices: Application to edge detection, Fuzzy Sets and Systems, Elsevier, available in: www.sciencedirect.com

[10] O. Mendoza, P. Melin, and G. Licea, 2007, Fuzzy Inference Systems Type-1 And Type-2 for Digital Images Edges Detection, Engineering Letters, International Association of Engineers, E.U.A., Volume 15 Issue 1 (Available on line: http://www.engineeringletters.com/issues_v15/issue_1/EL_15_1_7.pdf)

[11] O. Mendoza, P. Melin, G. Licea, 2007, A New Method for Edge Detection in Image Processing Using Interval Type-2 Fuzzy Logic, IEEE International Conference on Granular Computing (GRC 2007), Silicon Valley. Ca, U.S.A.

[12] O. Mendoza, P. Melin, G. Licea, 2008, A hybrid approach for image recognition combining type-2 fuzzy logic, modular neural networks and the Sugeno integral, Information Sciences, Elsevier, available in: www.sciencedirect.com

[13] O. Mendoza, P. Melin, O. Castillo, G. Licea, 2007, Type-2 Fuzzy Logic For Improving Training Data And Response Integration In Modular Neural Networks For Image Recognition, Foundations of Fuzzy Logic And Soft Computing (LNCC), Springer, Germany, pp. 604-612

[14] A. N. Evans, X.U., Liu, 2006, Morphological gradient operators for colour edges detection, IEEE Transactions on Image Processing, Vol. 15, Issue 6, pp. 1454-1463.

[15] Y. Becerikli, T. M. Karam, 2005, A New Fuzzy Approach for Edge Detection, in Computational Intelligence and Bioinspired Systems, LNCS, Springer, Berlin, pp. 943-951.

[16] F. Russo and G. Ramponi,1994, Edge extraction by FIRE operators Fuzzy Systems, IEEE World Congress on Computational Intelligence., pp, 249 – 253.

[17] J. Mendel, 2001, Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions, Prentice-Hall, U.S.A.

[18] J. R. Castro, O. Castillo, P. Melin, A. Rodriguez-Diaz, 2008, Building fuzzy inference systems with a new interval type-2 fuzzy logic toolbox: In Transactions on Computational Science Book Chapter, Vol. 4750, Springer, p.p. 104-114

[19] The USC-SIPI Image Database, available in http://sipi.usc.edu/database/

[20] O. Mendoza, P. Melin, 2007 The Fuzzy Sugeno Integral as a Decision Operator in the Recognition of Images with Modular Neural Networks, In: Hybrid Intelligent Systems Design and Analysis, Springer-Verlag, Germany, pp. 299-310.

[21] AT&T Laboratories Cambridge, The ORL database of faces, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

ISBN: 978-989-95079-6-8

Comparison of two versions of the Ferrers property of fuzzy interval orders

Susana Díaz[1] Bernard De Baets[2] Susana Montes[1]

1.Dept. Statistics and O. R., University of Oviedo
2.Dept. Appl. Math., Biometrics and Process Control; Ghent University
Email: {diazsusana,montes}@uniovi.es; bernard.debaets@ugent.be

Abstract— *We focus on the Ferrers property of fuzzy preference relations. We study the connection between the Ferrers property and fuzzy interval orders. A crisp total interval order is characterized by the Ferrers property of its strict preference relation. Also, a crisp preference structure is a total interval order if and only if its large preference relation satisfies the Ferrers property. For fuzzy relations the Ferrers property admits two non equivalent expressions. Here we compare both conditions by means of completeness. We also study if they characterize a fuzzy total interval order.*

Keywords— completeness, Ferrers property, partial interval order, total interval order, fuzzy relation

1 Introduction

Preference structures are the basis of the preference modelling theory. They formalize the answers of a decision maker over a set of alternatives. Formally, they are triplets of binary relations. For each pair of alternatives, any decision maker provides one (and only one) of the following answers: preference (by one of the alternatives), indifference (between both alternatives), inability to compare (the alternatives). These three situations lead to the three relations that make up a preference structure. Associated to a preference structure we find a relation called large preference relation that characterizes the preference structure from which it comes. Those four relations are the basis of the preference modelling ([9]).

Any preference structure can contain non-coherent relations if the decision maker gives conflicting answers. Different properties were formalized to model coherence. The Ferrers property is one of the most important ones. It has been proved to be a more realist condition than transitivity since the Ferrers property of the large preference relation does not imply the transitivity of the associated indifference relation (a non-realistic condition for some authors). However it does imply the transitivity of the strict preference relation. In a preference structure without incomparability, the large preference relation satisfies the Ferrers property if and only if the associated strict preference relation does.

A class of coherent preference structures are (total) interval orders: the alternatives can be identified with intervals not necessarily of the same length on the real line. A crisp preference structure without incomparabilities is a (total) interval order if and only if its associated large preference relation or its strict preference relation satisfy the Ferrers property.

Since several years ago, crisp relations are considered not flexible enough to model human decisions. In order to model real life situations in a more accurate way, fuzzy relations were introduced [10] and they have received a wide attention (see, e.g., [2, 3, 4, 6]). The notion of fuzzy preference structure was a topic of debate for several years. In this work we present a similar study for the notion of fuzzy interval order. This notion admits many different equivalent definitions in the classical context. We focus on the characterizations based on the Ferrers property. We first discuss the Ferrers property for fuzzy relations. This property admits two different ways to be formalized in our wider context. We compare them and we study their connection to fuzzy interval orders. In this context, a t-norm is needed to define fuzzy interval orders and the Ferrers property. Although we aim at considering any operator in our study, we prove that only two particular families have an appropriate behavior. Thus, we pay special attention to those operators. Since interval order is an important concept for the analysis of temporal events, each of which occurs over some time span (for instance, the time spans over which animal species are found or the occurrence of styles of pottery in archaeological strata), this kind of events in a fuzzy environment could be a potential application area for this study.

This work is organized as follows. In Section 2 we recall crisp relations. In Section 3 we formalize the two versions of the Ferrers property for fuzzy relations and we study their relationship. Section 4 is devoted to the completeness property. For crisp relations every reflexive Ferrers relation is complete. We study the implication for fuzzy relations. Section 5 contains the results on the connection between the first type of Ferrers property and fuzzy interval orders. Section 6 includes an analogous study for the second version of the Ferrers property. Section 7 contains some conclusions.

2 Crisp relations

2.1 Crisp preference structures

Let us consider a set of alternatives A whose elements we want to compare. In preference modeling the comparison is always carried out by pairs. When a decision maker consider two of those alternatives, the answer is always among the following ones: preference for one of the alternatives, indifference among both possibilities, inability to compare them. In the classical setting, each one of the three previous situations is identified with an $A \times A \to \{0,1\}$ application or relation. The first case is identified with the *strict preference relation*, denoted by P. We state $P(a,b) = 1$ if the decision maker prefers a to b and 0 otherwise. The second case is identified with the *indifference relation* denoted I. We assign $I(a,b) = 1$ if a and b are indifferent and $I(a,b) = 0$ if not. The *incomparability relation* is denoted J. If the alternatives a and b cannot be compared by the decision maker, $J(a,b) = 1$. Otherwise, $J(a,b) = 0$. P, I and J make up a preference structure. Let A^2 be the set of ordered pairs

ISBN: 978-989-95079-6-8

((a,b) and (b,a) are different pairs) of alternatives in A. The converse or transpose of Q is defined as $Q^t(a,b) = Q(b,a)$, its complement as $Q^c(a,b) = 1 - Q(a,b)$ and its dual as $Q^d = (Q^t)^c$. Every relation is also identified with a subset of A^2: $Q(a,b) = 1 \Leftrightarrow (a,b) \in Q$. One easily verifies that if we think of P, I and J as subsets of A^2, the quadruplet (P, P^t, I, J) establishes a particular partition of A^2. Formally [9]:

Definition 1 *A preference structure on A is a triplet (P, I, J) of relations on A that satisfy:*

(i) *P is irreflexive, I is reflexive and J is irreflexive;*

(ii) *P is asymmetrical, I and J are symmetrical;*

(iii) *$P \cap I = \emptyset$, $P \cap J = \emptyset$ and $I \cap J = \emptyset$;*

(iv) *$P \cup P^t \cup I \cup J = A^2$.*

Every preference structure is identified with a unique reflexive relation called *large preference relation* $R = P \cup I$. It connects a to b if the decision maker considers a preferred or indifferent to b. This relation leads back to the preference structure in the following way:

$$(P, I, J) = (R \cap R^d, R \cap R^t, R^c \cap R^d). \qquad (1)$$

We say that a relation R is *complete* if aRb or bRa for all a, b in the set of alternatives A. In [8] it was proven that R is complete if and only if the associated J does not connect any pair of alternatives (a,b),

$$R \text{ complete} \quad \Leftrightarrow \quad J = \emptyset.$$

A relation R is a Ferrers relation (see among others [8, 9]) if for any four alternatives a, b, c, d in A, it holds that if aRb and cRd then aRd or cRb. Another way of expressing this property is the following one: if aRb and cRd and $c \not R b$ then aRd. This property is also called biorder by some authors.

Let us remark also that the Ferrers and the completeness properties are connected.

Lemma 1 *Every reflexive Ferrers relation is complete.*

Let us denote the composition of two binary relations Q_1 and Q_2 by $Q_1 \circ Q_2$. That is, $a(Q_1 \circ Q_2)b$ if and only if there exists c such that $aQ_1c \wedge cQ_2b$. Then the equivalent compositional definition of Ferrers property is $R \circ R^d \circ R \subseteq R$.

2.2 Crisp interval orders

Interval orders are one of the most important types of "coherent" preference structures. Formally, we distinguish between partial and total interval order, depending on the possibility or impossibility of incomparable elements.

Definition 2 *A* partial interval order *is a preference structure (P, I, J) such that $P \circ I \circ P \subseteq P$. In the particular case $J = \emptyset$, (P, I, J) is said to be a* total interval order.

As commented in the introduction, this original definition admits different ways of being expressed. The following properties (among others) were shown to be equivalent in [8]:

Proposition 1 *Let $(P, I, \emptyset)$ be a preference structure without incomparability and R its large preference relation. The following conditions are equivalent.*

a) *(P, I, J) is an interval order.*

b) *P satisfies the Ferrers property.*

c) *R satisfies the Ferrers property.*

These definitions are equivalent only if the decision maker is able to compare all the pairs of alternatives, i. e., if $J = \emptyset$, equivalently if R is complete. If we do not impose this condition, we can still consider the previous properties and study their relationship. In what remains of this section we carry out this study.

It is very easy to prove that whenever R satisfies the Ferrers property, P satisfies the Ferrers property and $P \circ I \circ P \subseteq P$. It also holds that for any preference structure (P, I, J), if P satisfies the Ferrers property then $P \circ I \circ P \subseteq P$, but the converse implications do not hold.

The following proposition summarizes the relationship among the conditions of Proposition 1 when we consider any preference structure.

Proposition 2 *Let (P, I, J) be a (crisp) preference structure and let R be its large preference relation. Then,*

i) *(P, I, J) is a total interval order and R satisfies the Ferrers property are equivalent properties.*

ii) *(P, I, J) is a total interval order implies that P satisfies the Ferrers property. In addition to this, if P satisfies the Ferrers property then $P \circ I \circ P \subseteq P$. The converse implications to the ones showed in this item do not hold.*

Propositions 1 and 2 can be outlined as we show in Figure 1, where the missing implications do not hold.

$$\begin{array}{ccc} (P,I,J) \text{ total int. ord.} & \Leftrightarrow & R \text{ Ferrers} \\ & & \Downarrow \ \Updownarrow (J = \emptyset) \\ & & P \text{ Ferrers} \\ & & \Downarrow \ \Updownarrow (J = \emptyset) \\ & & (P,I,J) \text{ partial int. ord.} \end{array}$$

Figure 1: Connection among the characterizations of a crisp interval order.

3 Fuzzy preference structures

All the relations involved in a (crisp) preference structure are crisp. Then they are not always appropriate to model human decisions. They do not include intensity of preference. The lack of flexibility in crisp set theory led to introduce fuzzy relations. They allow the decision maker to encode degrees of preference rather than just its absence or presence. Fuzzy relations were introduced as a natural extension of the concept of crisp binary relation. They have been widely studied. For a complete review about fuzzy relations see [6].

Formally, a fuzzy relation is defined as a function Q from the cartesian product $A \times A$ into the interval $[0, 1]$. For every pair $(a,b) \in A \times A$, the value $Q(a,b)$ shows the degree of truth of the fact that aQb in the crisp sense. The complementary, transpose (or converse) and dual of the binary relation Q are defined for any $(a,b) \in A \times A$ as $Q^c(a,b) = 1 - Q(a,b)$, $Q^t(a,b) = Q(b,a)$ and $Q^d(a,b) = 1 - Q(b,a)$.

ISBN: 978-989-95079-6-8

The intersection of fuzzy relations is usually defined pointwisely based on some *t-norm*. We recall that the binary operator $T : [0,1] \times [0,1] \to [0,1]$ is a *triangular norm* or *t-norm* for short if it is commutative, associative, monotone and has 1 as neutral element. The three most important t-norms are the minimum operator $T_{\mathbf{M}}(x,y) = \min(x,y)$, the algebraic product $T_{\mathbf{P}}(x,y) = xy$ and the Łukasiewicz t-norm $T_{\mathbf{L}}(x,y) = \max(x+y-1,0)$. The minimum operator is the greatest t-norm; the smallest t-norm is the drastic product $T_{\mathbf{D}}$ defined by

$$T_{\mathbf{D}}(x,y) = \begin{cases} 0 & \text{if } \max(x,y) < 1\,, \\ \min(x,y) & \text{otherwise}\,. \end{cases}$$

On the other hand, we say that a value $x \in (0,1)$ is a zero-divisor of a t-norm T if there exists a value $y \in (0,1)$ such that $T(x,y) = 0$. The minimum and the product are t-norms that do not admit zero-divisors, while $T_{\mathbf{L}}$ and $T_{\mathbf{D}}$ do admit them.
An important family of t-norms that admit zero-divisors are the rotation invariant t-norms. A *rotation invariant t-norm* is a t-norm T that verifies:

$$T(x,y) \le z \;\Leftrightarrow\; T(x,1-z) \le 1-y, \quad \forall x,y,z \in [0,1].$$

The rotation invariance property for T means that the part of the space $[0,1]^3$ which is exactly below the graph of T remains invariant under an order three transformation. This transformation is indeed a rotation of $[0,1]^3$ with angle $2\pi/3$ around the axis that passes through the points $(0,0,1)$ and $(1,1,0)$. Rotation invariant t-norms satisfy in particular that $T(x,y) > 0 \Leftrightarrow x + y > 1$, i.e. the lower-left triangle of the unit square constitutes the zero-divisors of T. This excludes t-norms without zero-divisors (such as $T_{\mathbf{P}}$), nilpotent t-norms other than $T_{\mathbf{L}}$ and the drastic product $T_{\mathbf{D}}$.
As extreme members of the family we find the Łukasiewicz t-norm $T_{\mathbf{L}}$ and the nilpotent minimum $T_{\mathbf{nM}}$. This t-norm is defined as:

$$T_{\mathbf{nM}}(x,y) = \begin{cases} 0 & \text{if } x+y \le 1\,, \\ \min(x,y) & \text{otherwise}\,. \end{cases}$$

Similarly to the intersection, the union of fuzzy relations is based on a *t-conorm*, *i.e.* an increasing, commutative and associative binary operation on $[0,1]$ with neutral element 0. t-norms and t-conorms come in dual pairs: to any t-norm T there corresponds a t-conorm S through the relationship

$$S(x,y) = 1 - T(1-x, 1-y).$$

For the main three t-norms this yields the maximum operator $S_{\mathbf{M}}(x,y) = \max(x,y)$, the probabilistic sum $S_{\mathbf{P}}(x,y) = x + y - xy$ and the Łukasiewicz t-conorm (bounded sum) $S_{\mathbf{L}}(x,y) = \min(x+y,1)$. For more information on t-norms and t-conorms, we refer to [7].

The definition of a fuzzy preference structure has been discussed for many years [6]. The *assignment principle*, expressing that for any pair of alternatives (a,b) the decision maker is allowed to assign at least one of the degrees $P(a,b)$, $P(b,a)$, $I(a,b)$ and $J(a,b)$ freely in the unit interval, leads to a graduation of Definition 1 with intersection based on the Łukasiewicz t-norm and union based on the Łukasiewicz t-conorm. This definition admits the same short formulation as the classical one: a triplet (P,I,J) of fuzzy relations on A is an *additive fuzzy preference structure* on A if and only if I is reflexive ($I(a,a) = 1$ for any $a \in A$) and symmetrical, and for any $(a,b) \in A^2$ it holds that $P(a,b) + P^t(a,b) + I(a,b) + J(a,b) = 1$. This expression justifies the adjective *additive*. Note that P is irreflexive, and that J is irreflexive and symmetrical.

Another difficult point has been how to construct such a structure from a reflexive fuzzy relation. The most recent and most successful approach is that of De Baets and Fodor based on (indifference) *generators* [1].

Definition 3 *A generator i is a commutative $[0,1]^2 \to [0,1]$ mapping that satisfies $T_{\mathbf{L}} \le i \le T_{\mathbf{M}}$.*

Note that a generator always has neutral element 1. For any reflexive fuzzy relation R on A it holds that the triplet (P,I,J) of fuzzy relations on A defined by:

$$\begin{aligned} P(a,b) &= R(a,b) - i(R(a,b),R(b,a))\,, \\ I(a,b) &= i(R(a,b),R(b,a))\,, \\ J(a,b) &= i(R(a,b),R(b,a)) - (R(a,b)+R(b,a)-1)\,, \end{aligned}$$

is an additive fuzzy preference structure on A such that R is the union by the Łukasiewicz t-conorm of P and I, this is, $R(a,b) = P(a,b) + I(a,b)$. The fuzzy relation R is again called the *large preference relation*.
A family of generators that present important properties is the family of Frank t-norms (see, e.g., [6]). In the following section we will see that a particular Frank t-norm is a specially good generator: the Łukasiewicz t-norm.

Definition 4 *Given a t-norm T and a t-conorm S, a fuzzy relation Q is a type 1 T-S-*Ferrers *relation if*

$$T(Q(a,b),Q(c,d)) \le S(Q(a,d),Q(c,b)),\ \forall\, a,b,c,d, \in A.$$

The t-norm T and the t-conorm S are not necessarily connected. However, given T, the most usual t-conorm considered is the dual of T. In this case, the t-conorm is not specified and we just talk about type 1 T-Ferrers property. Along this work, this will be the setting: we will always consider as t-conorm the dual of the fixed t-norm.

Definition 5 *Given a t-norm T, a fuzzy relation Q is a* type 2 T-Ferrers *relation if*

$$T(Q(a,b),Q^d(b,c),Q(c,d)) \le Q(a,d),\ \forall\, a,b,c,d \in A.$$

The composition of two fuzzy relations Q_1 and Q_2 defined on A also involves t-norms. Given a t-norm T, the T-composition of Q_1 and Q_2 is the fuzzy relation $Q_1 \circ_T Q_2$ defined as follows

$$(Q_1 \circ_T Q_2)(a,b) = \sup\{T(Q_1(a,c),Q_2(c,b)) | c \in A\}.$$

With this notation we can also say that a fuzzy relation Q is a type 2 T-Ferrers relation if it holds that $Q \circ_T Q^d \circ_T Q \subseteq Q$.

Definition 6 [2] *An additive fuzzy preference structure is a T-*partial interval order *if it holds that*

$$P \circ_T I \circ_T P \subseteq P\,,$$

*or, equivalently, $T(T(P(a,b),I(b,c)),P(c,d)) \le P(a,d)$, $\forall\, a,b,c,d \in A$. It is a T-*total interval order *if it is a T-partial interval order and $J = \emptyset$.*

ISBN: 978-989-95079-6-8

Clearly, if $T_1 \leq T_2$, every type 1 T_2-Ferrers relation is also a type 1 T_1-Ferrers relation, every type 2 T_2-Ferrers relation is also a type 2 T_1-Ferrers relation and every T_2-interval order is also a T_1-interval order.

The equivalence between the notions of type 1 and type 2 T-Ferrers relations has already been studied in some particular cases ([5]). The following theorem further generalizes these results and characterizes all t-norms for which both notions coincide.

Theorem 1 *Consider a t-norm T. Then the following statements are equivalent:*

(i) *Any type 1 T-Ferrers relation is type 2 T-Ferrers.*

(ii) *Any type 2 T-Ferrers relation is type 1 T-Ferrers.*

(iii) *T is rotation-invariant.*

It follows from this theorem that in particular, for any reflexive relation R and for any strict preference relation P, the equivalence between the type 1 T-Ferrers property and the type 2 T-Ferrers property holds if and only if T is a rotation invariant t-norm. This justifies the importance we give to this family of t-norms in this work.

The theorem above does not apply to the minimum t-norm. However it is an important operator and the greatest t-norm. What can we say about it? Since $T_{\mathbf{nM}} \leq T_{\mathbf{M}}$, it follows immediately that any type 1 $T_{\mathbf{M}}$-Ferrers relation is a type 2 $T_{\mathbf{nM}}$-Ferrers relation and that any type 2 $T_{\mathbf{M}}$-Ferrers relation satisfies the type 1 $T_{\mathbf{nM}}$-Ferrers property. This result can be strengthened considerably.

Proposition 3

(i) *$T_{\mathbf{nM}}$ is the largest t-norm T such that any type 1 $T_{\mathbf{M}}$-Ferrers relation is also type 2 T-Ferrers.*

(ii) *$T_{\mathbf{nM}}$ is the largest t-norm T such that any type 2 $T_{\mathbf{M}}$-Ferrers relation is also type 1 T-Ferrers.*

4 Completeness of the large preference relation

We have seen that for crisp relations the completeness of the large preference relation follows from the Ferrers property. In this section we consider both types of the Ferrers property. We investigate the kind of completeness exhibited by a fuzzy reflexive relation satisfying one of the two properties.

Definition 7 *Let us consider a t-conorm S and a fuzzy relation Q defined on a set of alternatives A. Q is S-complete if it satisfies*

$$S(Q(a,b), Q(b,a)) = 1, \qquad \forall\, a, b \in A.$$

The two most important notions are based on the minimum and the Łukasiewicz t-norms. Given a relation Q we say that it is

- *Strongly complete:* if $Q(a,b) = 1$ or $Q(b,a) = 1, \forall a, b \in A$.
- *weakly complete:* if $Q(a,b) + Q(b,a) \geq 1, \forall a, b \in A$.

The weakest S-completeness condition is based on $S_{\mathbf{D}}$. Given Q, it is $S_{\mathbf{D}}$-complete if $\max(Q(a,b), Q(b,a)) > 0$ for all a, b in A.

No one of the "fuzzy notions" of completeness of R is equivalent to the condition $J = \emptyset$, but we can establish the following equivalence.

Lemma 2 [4] *Let R be a reflexive fuzzy relation and let J be the incomparability relation associated to R by means of any generator i. Then the following equivalence holds*

$$J = \emptyset \Leftrightarrow \begin{cases} R \text{ is weakly complete} \\ i|_S = T_{\mathbf{L}} \end{cases}$$

where $S = \{(u,v) \in [0,1]^2 : \exists (x,y) \in A^2$ with $R(x,y) = u, R(y,x) = v\}$.

Let us recall that when $i = T_{\mathbf{L}}$, the fuzzy preference structure obtained from a weakly complete reflexive relation R is the triplet $(P_{\mathbf{L}}, I_{\mathbf{L}}, J_{\mathbf{L}}) = (R^d, \min(R, R^t), \emptyset)$. We would like to stress the fact that the strict preference relation becomes the dual of the large preference relation.

We next focus on the completeness assured for reflexive type 1 and type 2 T-Ferrers relations.

Proposition 4 *Let T be a t-norm and S its dual t-conorm. Every type 1 T-Ferrers reflexive relation is S-complete.*

As an immediate consequence we get the following: Every type 1 T-Ferrers reflexive relation satisfies the weakest kind of completeness ($S_{\mathbf{D}}$-completeness).

Since strong and weak completeness are in fact $S_{\mathbf{M}}$-completeness and $S_{\mathbf{L}}$-completeness, it also follows from Proposition 4 that every type 1 $T_{\mathbf{L}}$-Ferrers relation is weakly complete and every type 1 $T_{\mathbf{M}}$-Ferrers relation is strongly complete. We can say more.

Theorem 2 *Consider a t-norm T. Every reflexive type 1 T-Ferrers relation is weakly complete if and only if $T(x,y) > 0$ for any pair $(x,y) \in [0,1]^2$ such that $x + y > 1$.*

This theorem applies to all rotation-invariant t-norms. In particular, it applies to $T_{\mathbf{L}}$, and hence to any $T \geq T_{\mathbf{L}}$. Theorem 2 also applies to t-norms without zero-divisors, although in that case the following much stronger result holds.

Theorem 3 *Any reflexive type 1 T-Ferrers relation is strongly complete if and only if the t-norm T has no zero-divisors.*

This theorem applies in particular to $T_{\mathbf{M}}$.

Next we repeat the same study for reflexive type 2 T-Ferrers relations.

Theorem 4 *Any reflexive type 2 T-Ferrers relation is weakly complete.*

A first comparison between types 1 and 2 of the T-Ferrers property is in favour of the type 2 definition as the previous theorem holds for any t-norm T. However, our enthusiasm is tempered by the following negative result.

Proposition 5 *Not all reflexive type 2 $T_{\mathbf{M}}$-Ferrers relations are strongly complete.*

ISBN: 978-989-95079-6-8

Despite this negative result, t-norms without zero-divisors have good properties.

Theorem 5 *Let $(P, I, \emptyset)$ be an additive fuzzy preference structure without incomparability and let R be its associated large preference relation. Let T be a t-norm. The following assertions are equivalent.*
(i) *If $(P, I, \emptyset)$ is a T-total interval order, then R is strongly complete.*
(ii) *If P is type 1 T-Ferrers, then R is strongly complete.*
(iii) *If P is type 2 T-Ferrers, then R is strongly complete.*
(iv) *If R is type 1 T-Ferrers, then R is strongly complete.*
(v) *T does not admit zero-divisors.*

The results presented in this section show that there are two important families of t-norms in the study of the Ferrers property: rotation invariant t-norms and t-norms without zero-divisors. In the two following sections we pay special attention to those two families [8].

5 Fuzzy interval orders and the type 1 Ferrers property

In this section we study the implications of Figure 1 that remain true for fuzzy relations when the Ferrers property of a crisp relation is generalized by means of the type 1 Ferrers property.

Lemma 3 [2] *Not every $T_{\mathbf{M}}$-total interval order verifies that its strict preference relation is type 1 $T_{\mathbf{M}}$-Ferrers. Not every $T_{\mathbf{L}}$-total interval order verifies that its strict preference relation is type 1 $T_{\mathbf{L}}$-Ferrers.*

Therefore, we cannot assure that for every total interval order its strict preference relation satisfies the type 1 Ferrers property.

Let us focus on the connection between the large and the strict preference relations.

Let us recall first of all that the results we present in this contribution concern a fixed generator $i = T_{\mathbf{L}}$, as it has shown an appropriate behavior (see Lemma 2).

Theorem 6 *The following statements are equivalent:*

(i) The strict preference relation obtained from a type 1 T-Ferrers reflexive relation also satisfies the type 1 T-Ferrers property.

(ii) The t-norm T satisfies $T(x, y) > 0$ for any pair $(x, y) \in [0, 1]^2$ such that $x + y > 1$.

In particular, the previous result can be applied to rotation-invariant t-norms and t-norms without zero-divisors.

The result presented in Theorem 6 can be strengthened under completeness conditions.

Theorem 7 *Consider a fuzzy reflexive relation R with corresponding fuzzy preference structure $(P_{\mathbf{L}}, I_{\mathbf{L}}, J_{\mathbf{L}})$ generated by means of $i = T_{\mathbf{L}}$. For any t-norm T, the following statements are equivalent:*

- *R weakly complete and type 1 T-Ferrers*
- *$J_{\mathbf{L}} = \emptyset$ and $P_{\mathbf{L}}$ type 1 T-Ferrers*

Proposition 6 *Let (P, I, J) be an additive fuzzy preference structure and let T be a rotation invariant t-norm. If P is a type 1 T-Ferrers relation, then (P, I, J) is a T-partial interval order.*

The previous implication cannot be extended to t-norms without zero-divisors. In [2] it was proven that not every type 1 $T_{\mathbf{M}}$-Ferrers reflexive relation satisfies that the fuzzy preference structure obtained from it (by the Łukasiewicz t-norm) is a $T_{\mathbf{M}}$-total interval order. It follows from here and Theorem 6 that the implication of Proposition 6 does not hold for the family of t-norms without zero-divisors.

Now we focus on the connection between a total interval order and the type 1 T-Ferrers property of the associated large preference relation.

Corollary 1 *Let T denote the minimum or the Łukasiewicz t-norms. Not every T-total interval order verifies that its associated large preference relation satisfies the type 1 T-Ferrers property.*

The converse implication neither holds for $T_{\mathbf{M}}$ [2]. However the implication does hold for rotation invariant t-norms.

Proposition 7 *Let T be a rotation invariant t-norm. The additive fuzzy preference structure obtained (by the Łukasiewicz t-norm) from a type 1 T-Ferrers reflexive relation is a T-total interval order.*

In Figures 2 and 3 we summarize the results obtained for the type 1 Ferrers property both for the family of rotation invariant t-norms and for the family of t-norms without zero divisors. We have proven that the implications missing do not hold.

$$\begin{array}{ccccc} (P,I,J) & T\text{-total int. ord.} & \Leftarrow & R & \text{type 1 } T\text{-Ferrers} \\ & (J=\emptyset)\nwarrow & & \not\swarrow \quad \not\nearrow & (J=\emptyset) \\ & & & P & \text{type 1 } T\text{-Ferrers} \\ & & & \Downarrow & \\ & & & (P,I,J) & T\text{-partial int. ord.} \end{array}$$

Figure 2: Connection among the type 1 T-Ferrers property and the notion of interval order for T a rotation invariant t-norm.

$$\begin{array}{ccccc} (P,I,J) & T\text{-total int. ord.} & & R & \text{type 1 } T\text{-Ferrers} \\ & (J=\emptyset) & & \not\swarrow \quad \not\nearrow & (J=\emptyset) \\ & & & P & \text{type 1 } T\text{-Ferrers} \\ & & & & \\ & & & (P,I,J) & T\text{-partial int. ord.} \end{array}$$

Figure 3: Connection among the type 1 T-Ferrers property and the notion of interval order for T a t-norm without zero-divisors.

6 Fuzzy interval orders and the type 2 Ferrers property

We already know the behaviour of the type 1 Ferrers property when dealing with interval orders. Now we focus on the type 2 Ferrers property.

ISBN: 978-989-95079-6-8

Proposition 8 *Let T be a t-norm without zero-divisors and let $(P, I, \emptyset)$ be a T-total interval order, then P satisfies the type 2 T-Ferrers property.*

However the implication cannot be extended to any t-norm.

Proposition 9 *Let T be a t-norm such that $T(x, 1-x) = 0$ for some $x \in (0,1)$. Not every T-total interval order satisfies that its strict preference relation is type 2 T-Ferrers.*

A particular type of the t-norms described in the previous proposition are the rotation invariant t-norms.

Next we study how the type 2 Ferrers property propagates between R and P.

Theorem 8 *Consider a t-norm T. Then the following statements are equivalent:*

(i) *For any reflexive relation R, if R satisfies the type 2 T-Ferrers property, then $J_{\mathbf{L}} = \emptyset$ and $P_{\mathbf{L}}$ satisfies the type 2 T-Ferrers property.*

(ii) *For any reflexive relation R, if $J_{\mathbf{L}} = \emptyset$ and $P_{\mathbf{L}}$ satisfies the type 2 T-Ferrers property, then R satisfies the type 2 T-Ferrers property.*

(iii) *T is rotation-invariant.*

We now study the relationship between the type 2 T-Ferrers property of P and a T-partial interval order.

Proposition 10 *Let (P, I, J) be an additive fuzzy preference structure and let T be a t-norm. If P is type 2 T-Ferrers then (P, I, J) is a T-partial interval order.*

It follows from the crisp case that the converse implication does not hold for any t-norm.

The characterization of total interval orders based on the Ferrers property of the associated large preference relation gets lost for fuzzy relations when considering the type 2 T-Ferrers condition.

Proposition 11 *It does not hold that for every $T_{\mathbf{M}}$-total interval order the associated large preference relation is type 2 $T_{\mathbf{M}}$-Ferrers. It neither holds that for any type 2 $T_{\mathbf{M}}$-Ferrers reflexive relation the associated by the Łukasiewicz t-norm preference structure is a $T_{\mathbf{M}}$-total interval order.*

Proposition 12 *For any type 2 $T_{\mathbf{L}}$-Ferrers reflexive relation R the associated fuzzy preference structure $(P_{\mathbf{L}}, I_{\mathbf{L}}, J_{\mathbf{L}})$ is a $T_{\mathbf{L}}$-total interval order.*
The converse implication cannot be assured for any type 2 $T_{\mathbf{L}}$-total interval order.

We draw in Figures 4 and 5 the results obtained for the type 2 Ferrers property.

7 Conclusions

We have studied the connection between the fuzzy Ferrers property and the definition of fuzzy interval orders. Since the Ferrers property admits two different ways to be formalized for fuzzy relations, we have compared those two versions by its connection to fuzzy interval orders. We have seen that we cannot identify a best definition since each property behaves better in different cases. We have seen that both properties are equivalent only when dealing with rotation invariant t-norms. For these operators, Ferrers property is a stronger condition than the notion of total interval order. For t-norms without zero-divisors, no one of them is good. The type 2 condition seems to assure the equivalence between total interval orders and the Ferrers property of the associated strict preference relation. However, type 1 behaves better if we want to propagate the Ferrers property between the large and the strict preference relations.

$$\begin{array}{ccc} (P,I,J)\ T\text{-total int. ord.} & \Leftarrow & R \text{ type 2 } T\text{-Ferrers} \\ (J=\emptyset)\nwarrow & & \swarrow\ \nearrow (J=\emptyset) \\ & P \text{ type 2 } T\text{-Ferrers} & \\ & \Downarrow & \\ & (P,I,J)\quad T\text{-partial int. ord.} & \end{array}$$

Figure 4: Connection between the type 2 T-Ferrers property and interval orders for T a rotation invariant t-norm.

$$\begin{array}{ccc} (P,I,J)\ T\text{-total int. ord.} & & R \text{ type 2 } T\text{-Ferrers} \\ (J=\emptyset)\swarrow\ \searrow & & (J=\emptyset) \\ & P \text{ type 2 } T\text{-Ferrers} & \\ & \Downarrow & \\ & (P,I,J)\quad T\text{-partial int. ord.} & \end{array}$$

Figure 5: Connection between the type 2 T-Ferrers property and interval orders for T a t-norm without zero-divisors.

Acknowledgment

The research reported on in this paper has been partially supported by Project FEDER-MEC-MTM2007-61193.

References

[1] B. De Baets and J. Fodor. Generator triplets of additive fuzzy preference structures. In *Proc. RelMiCS6*.

[2] B. De Baets and B. Van de Walle. Weak and strong fuzzy interval orders. *Fuzzy Sets and Systems*, 79:213–225, 1996.

[3] S. Díaz, B. De Baets, and S. Montes. Additive decomposition of fuzzy pre-orders. *Fuzzy Sets and Systems*, 158:830–842, 2007.

[4] S. Díaz, S. Montes, and B. De Baets. Transitivity bounds in additive fuzzy preference structures. *IEEE Transactions on fuzzy systems*, 15(2):275–286, 2007.

[5] J. Fodor. Traces of binary fuzzy relations. *Fuzzy Sets and Systems*, 50:331–341, 1992.

[6] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers, 1994.

[7] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, Dordrecht, 2000.

[8] B. Monjardet. Axiomatiques et propriétés des quasi-ordres. *Math. Sci. Hum.*, 63:51–82, 1978.

[9] M. Roubens and Ph. Vincke. Preference modelling. In *Lecture Notes in Economics and Mathematical Systems*, volume 250. Springer, Berlin, 1985.

[10] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

ISBN: 978-989-95079-6-8

Decomposition of the transitivity for additive fuzzy preference structures

Susana Díaz, Susana Montes[1] Bernard De Baets[2]

1.Dept. of Statistics and O.R., University of Oviedo
Oviedo, Spain
2.Dept. of Applied Mathematics, Biometrics and Process Control, Ghent University
Gent, Belgium
Email: {diazsusana,montes}@uniovi.es, Bernard.DeBaets@UGent.be

Abstract— *Transitivity is a very important property in order to provide coherence to a preference relation. Usually, t-norms are considered to define the transitivity of fuzzy relations. In this paper we deal with conjunctors, a wider family than t-norm, to define the transitivity. This more general definition allows to impove the results found in the literature. We characterize the behaviour with respect to transitivity of the strict preference and indifference relations of any fuzzy preference structure associated to any large preference relation. Those two characterizations provide very general expressions. We also obtain easier expressions in some particular cases.*

Keywords— Fuzzy preference relation, transitivity, conjunctor, indifference generator.

1 Introduction

Decision making is present in many situations in life. Preference models are an essential part of the design phase of a decision making process. The departure point in preference modelling is the comparison by pairs of possible alternatives. If this point of the process lacks of coherence, the whole process makes no sense. One of the most important property introduced to ensure coherence in the two-by-two comparison is transitivity. In this paper we focus on this property.

Given its relevance in preference modelling, transitivity has been widely studied (see [3, 13, 14], among others). In classical or crisp preference modelling there exists a basic relation R, called *large preference relation* that allows to compare a pair of possibilities: given the alternatives a and b, $(a, b) \in R$ expresses that alternative a is considered to be at least as good as alternative b. From R, three relations can be defined: the strict preference relation P, the indifference relation I and the incomparability relation J. It is well known that the transitivity of R is completely characterized by the transitivity of I and P and two additional relational inequalities involving I and P; in case of completeness, the transitivity of R is only characterized by the transitivity of I and P [2].

All those relations are crisp and therefore, they are not always appropriate to model human decisions. The lack of flexibility in crisp set theory lead to introduce fuzzy sets in preference modeling and to study the concept of fuzzy preference structure [8, 18, 24]; see [6] for a historical account of its development. Transitivity is traditionally defined for fuzzy relations by means of t-norms. In this context, the transitivity of the (fuzzy) large preference relation R has been characterized in a similar way as for crisp relations, in a particular case: when R is strongly complete [12]. But strong completeness is a quite restrictive condition. Subsequent works have treated more general preference relations [3, 4]. Those authors center their study on the translation of a fixed transitivity from a more general (fuzzy) large preference relation R to different associated indifference and strict preference relations (I and P respectively). We have continued along this line but we have treated the problem from a totally different viewpoint. On the one hand, we work with conjunctors, a wider family of operators than t-norms, so we handle a very general notion of transitivity. On the other hand, we do not study the preservation of the transitivity when decomposing R, but we explore the strongest transitivity we can assure I and P satisfy. In a previous work [15], we have fenced in the transitivity of I and P according to the transitivity R satisfies. In this work we go further: we provide an explicit expression for the transitivity that any I (respectively any P) satisfies. We prove that no stronger transitivity is fulfilled by all I (respectively for all P). Transitivity is a fundamental assumption in some decision-making models, like rational model (founded in quantitative disciplines like economics, mathematics and statistics) or political model (primarily based on the disciplines of political science, philosophy, psychology and sociology). Without this property, it is not possible to obtain a coherent order of the alternatives and therefore the choice of the best alternative is more complicate or even impossible. Thus, the study made in this work is a first step to develop a consistent theory to order the alternatives and therefore to provide a decision, when the comparison between alternatives can be graded.

The work is structured in six sections. In Section 2 we recall the most relevant concepts concerning additive fuzzy preference structures. In Section 3 we introduce conjunctors and we discuss some properties they satisfy and that will be useful in next sections. In Section 4 we include the general result we have obtained for the transitivity of the indifference relation I and we discuss the appearance of the general expression for the three most important particular cases. In Section 5 we characterize the transitivity of the strict preference relation P in a very general context. We also prove that the general expression gets much simplified for the most important particular cases. In Section 6 we briefly address some conclusions and future work.

2 Preference structures

2.1 Crisp preference structures

Let us consider a decision maker who is given a set of alternatives A. Let us suppose that this person compares the alternatives two by two. Given two alternatives, the decision maker

ISBN: 978-989-95079-6-8

can act in one of the following three ways: (i) he/she clearly prefers one to the other; (ii) the two alternatives are indifferent to him/her; (iii) he/she is unable to compare the two alternatives. According to these cases, three binary relations can be defined on A: the strict preference relation P, the indifference relation I and the incomparability relation J. Thus, for any $(a,b) \in A^2$, we classify:

$$\begin{array}{rcl}(a,b) \in P & \Leftrightarrow & \text{he/she prefers } a \text{ to } b;\\ (a,b) \in I & \Leftrightarrow & a \text{ and } b \text{ are indifferent to him/her;}\\ (a,b) \in J & \Leftrightarrow & \text{he/she is unable to compare } a \text{ and } b.\end{array}$$

We recall that for a binary relation Q on A, its converse is defined as $Q^t = \{(b,a) \mid (a,b) \in Q\}$, its complement as $Q^c = \{(a,b) \mid (a,b) \notin Q\}$ and its dual as $Q^d = (Q^t)^c$. If we consider the set A^2 ordered, *i.e.* assuming (a,b) and (b,a) as different pairs, one easily verifies that the triplet (P,I,J) and P^t establish a particular partition of A^2 [21].

Definition 1 *A preference structure on A is a triplet (P,I,J) of binary relations on A that satisfy:*

(i) *P is irreflexive, I is reflexive and J is irreflexive;*

(ii) *P is asymmetrical, I and J are symmetrical;*

(iii) *$P \cap I = \emptyset$, $P \cap J = \emptyset$ and $I \cap J = \emptyset$;*

(iv) *$P \cup P^t \cup I \cup J = A^2$.*

Every preference structure has associated a reflexive relation that completely characterizes this structure. A preference structure (P,I,J) on A is characterized by the reflexive binary relation $R = P \cup I$, its large preference relation, in the following way:

$$(P,I,J) = (R \cap R^d, R \cap R^t, R^c \cap R^d). \qquad (1)$$

Conversely, for any reflexive binary relation R on A, the triplet (P,I,J) constructed in this way from R is a preference structure on A such that $R = P \cup I$. As R is the union of the strict preference and the indifference, $(a,b) \in R$ means that a is at least as good as b.

Given a binary relation Q on A, we say that Q is transitive if $(aQb \wedge bQc) \Rightarrow aQc$, for any $(a,b,c) \in A^3$. Given two binary relations Q_1 and Q_2 on A, the composition is a binary relation denoted $Q_1 \circ Q_2$ such that for any $(a,b) \in A^2$ $a\,(Q_1 \circ Q_2)\,b \Leftrightarrow \exists c / aQ_1c \wedge cQ_2b$. Then, it is clear that Q is transitive if and only if $Q \circ Q \subseteq Q$. The transitivity of the large preference relation R can be characterized as follows [2].

Theorem 1 *For any reflexive relation R with corresponding preference structure (P,I,J) it holds that*

$$R \circ R \subseteq R \Leftrightarrow (P \circ P \subseteq P \wedge I \circ I \subseteq I \wedge P \circ I \subseteq P \wedge I \circ P \subseteq P).$$

In case R is complete, *i.e.* $R \cup R^t = A^2$, this characterization can be simplified as follows. Note that the completeness of R is equivalent to establish that any two elements are comparable, that is, $J = \emptyset$.

Theorem 2 *For any complete reflexive relation R with corresponding preference structure $(P,I,\emptyset)$ it holds that*

$$R \circ R \subseteq R \Leftrightarrow (P \circ P \subseteq P \ \wedge \ I \circ I \subseteq I).$$

Next we recall an important characterization of a preference structure. Let us consider for every relation its characteristic mapping *i.e.* $Q(a,b) = 1 \Leftrightarrow aQb$. Definition 1 can be written in the following minimal way [11]: I is reflexive and symmetrical, and for any $(a,b) \in A^2$:

$$P(a,b) + P^t(a,b) + I(a,b) + J(a,b) = 1.$$

Classical preference structures can therefore also be considered as Boolean preference structures, employing 1 and 0 for describing presence or absence of strict preferences, indifferences and incomparabilities.

2.2 Additive fuzzy preference structures

In the classical model, relations only express presence or absence of relationship, while fuzzy relations capture the nuances of human choices. In fuzzy preference modelling, strict preference, indifference and incomparability are a matter of degree. These degrees can take any value between 0 and 1 and fuzzy relations are used for capturing them (for a complete review about fuzzy relations see [18]).

The intersection of fuzzy relations is usually defined pointwisely based on some t-norm, *i.e.* an increasing, commutative and associative binary operation on $[0,1]$ with neutral element 1. The three most important t-norms are the minimum operator $T_{\mathbf{M}}(x,y) = \min\{x,y\}$, the algebraic product $T_{\mathbf{P}}(x,y) = xy$ and the Łukasiewicz t-norm $T_{\mathbf{L}}(x,y) = \max\{x+y-1,0\}$. The minimum operator is the greatest t-norm; the smallest t-norm is the drastic product defined by

$$T_{\mathbf{D}}(x,y) = \begin{cases} \min\{x,y\} & \text{, if } \max\{x,y\} = 1\,,\\ 0 & \text{, otherwise.}\end{cases}$$

The above t-norms can be ordered (usual ordering of functions) as follows: $T_{\mathbf{D}} \leq T_{\mathbf{L}} \leq T_{\mathbf{P}} \leq T_{\mathbf{M}}$. Similarly, the union of fuzzy relations is based on a t-conorm, *i.e.* a non-decreasing, commutative and associative binary operation on $[0,1]$ with neutral element 0. T-norms and t-conorms come in dual pairs: to any t-norm T there corresponds a t-conorm S through the relationship $S(x,y) = 1 - T(1-x, 1-y)$. For the above three t-norms this yields the maximum operator $S_{\mathbf{M}}(x,y) = \max\{x,y\}$, the probabilistic sum $S_{\mathbf{P}}(x,y) = x + y - xy$ and the Łukasiewicz t-conorm (bounded sum) $S_{\mathbf{L}}(x,y) = \min\{x+y,1\}$. For more information on t-norms and t-conorms, we refer to [19]. Along this paper we use the notations for t-norms presented there.

The definition of a fuzzy preference structure has been a topic of debate during several years (see e.g. [18, 23, 24]). Accepting the *assignment principle* — for any pair of alternatives (a,b) the decision maker is allowed to assign at least one of the degrees $P(a,b)$, $P(b,a)$, $I(a,b)$ and $J(a,b)$ freely in the unit interval — has finally led to a graduation of Definition 1 with the intersection based on the Łukasiewicz t-norm and the union based on the Łukasiewicz t-conorm. Interestingly, a minimal definition is identical to the classical one if we replace ordinary by fuzzy binary relations: a triplet (P,I,J) of fuzzy binary relations on A is a fuzzy preference structure on A if and only if I is reflexive and symmetrical, and for any $(a,b) \in A^2$:

$$P(a,b) + P^t(a,b) + I(a,b) + J(a,b) = 1,$$

ISBN: 978-989-95079-6-8

where $P^t(a,b) = P(b,a)$. This identity explains the name *additive fuzzy preference structures*.

Another topic of controversy has been how to construct such a structure from a reflexive fuzzy relation. Alsina [1] proved a kind of impossibility theorem showing that a construction based on a single t-norm is unfeasible. As a reaction, Fodor and Roubens adopted an axiomatic approach [18]. The most recent and most successful approach is that of De Baets and Fodor based on (indifference) generators [8].

Definition 2 *A generator i is a commutative $[0,1]^2 \to [0,1]$ mapping bounded by the Łukasiewicz t-norm $T_{\mathbf{L}}$, and the minimum operator $T_{\mathbf{M}}$: $T_{\mathbf{L}} \leq i \leq T_{\mathbf{M}}$.*

Note that the definition of a generator does not speak of monotonicity and therefore they are not necessarily t-norms, albeit having neutral element 1. For any reflexive fuzzy relation R on A it holds that the triplet (P,I,J) of fuzzy binary relations on A defined by:

$$\begin{aligned} P(a,b) &= R(a,b) - i(R(a,b),R(b,a)), \\ I(a,b) &= i(R(a,b),R(b,a)), \\ J(a,b) &= i(R(a,b),R(b,a)) - (R(a,b)+R(b,a)-1), \end{aligned}$$

is an additive fuzzy preference structure on A such that $R = P \cup_{S_{\mathbf{L}}} I$ *i.e.* $R(a,b) = P(a,b) + I(a,b)$.

Popular generators (see e.g. [18]) are the Frank t-norms. For the sake of completeness, we recall that the Frank t-norms are given by

$$T_\lambda^{\mathbf{F}}(x,y) = \begin{cases} T_{\mathbf{M}}(x,y) & \text{, if } \lambda = 0, \\ T_{\mathbf{P}}(x,y) & \text{, if } \lambda = 1, \\ T_{\mathbf{L}}(x,y) & \text{, if } \lambda = \infty, \\ \log_\lambda(1 + \frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}) & \text{, otherwise.} \end{cases}$$

We also recall that for any $\lambda \in [0,\infty]$ and for any $(x,y) \in [0,1]^2$,

$$T_{1/\lambda}^{\mathbf{F}}(x,y) = x - T_\lambda^{\mathbf{F}}(x,1-y).$$

3 Conjunctors

3.1 Generalizing T-transitivity

The usual way of defining the transitivity of a fuzzy relation is by means of a t-norm. Recall that a fuzzy relation Q on A is T-transitive if $T(Q(a,b),Q(b,c)) \leq Q(a,c)$ for any $(a,b,c) \in A^3$. However, such a framework is too restrictive in the setting of fuzzy preference modelling. On the one hand, even when the large preference relation R is T-transitive w.r.t. a t-norm T, the transitivity of the generated P and I cannot always be expressed w.r.t. a t-norm [14, 16, 15]. On the other hand, the results we will present in the following sections also hold when R itself is transitive w.r.t. a more general operator. From a fuzzy preference modelling point of view, it is not that surprising that the class of t-norms is too restrictive, as a similar conclusion was drawn when identifying suitable generators, as briefly explained in the previous section.

As it was discussed in [15, 16], suitable operators for defining transitivity for fuzzy relations are conjunctors:

Definition 3 *A binary operation $f : [0,1]^2 \to [0,1]$ is called a conjunctor if*

1. *it coincides with the Boolean conjunction on $\{0,1\}^2$;*
2. *it is increasing in each variable.*

Given a conjunctor f, we say that a fuzzy relation Q defined on A is f-transitive if for any $(a,b,c) \in A^3$,

$$f(Q(a,b),Q(b,c)) \leq Q(a,c).$$

It is clear that the definition of conjunctor largely extends the notion of t-norm. However, conjunctors and generators are not connected. The smallest conjunctor c_S and greatest conjunctor c_G are given by

$$c_S(x,y) = \begin{cases} 0 & \text{, if } \min\{x,y\} < 1, \\ 1 & \text{, otherwise,} \end{cases}$$

and

$$c_G(x,y) = \begin{cases} 0 & \text{, if } \min\{x,y\} = 0, \\ 1 & \text{, otherwise.} \end{cases}$$

As it is logical, $c_S \leq T_{\mathbf{D}} \leq T_{\mathbf{M}} \leq c_G$. Moreover, for two conjunctors $f \leq g$, it obviously holds that g-transitivity implies f-transitivity.

Defining the composition $Q_1 \circ_f Q_2$ of two fuzzy relations Q_1 and Q_2 w.r.t. a conjunctor f by

$$Q_1 \circ_f Q_2(a,c) = \sup_b f(Q_1(a,b),Q_2(b,c)),$$

still allows us to use the shorthand $Q \circ_f Q \subseteq Q$ to denote f-transitivity.

If we restrict our study for reflexive fuzzy relations, such as a large preference relations, the set of all the conjunctors that allow us to define transitivity is a proper subset of the set of all the conjunctors. Thus, the upper bound of this subset is $T_{\mathbf{M}}$ instead of c_G.

3.2 Dominance and bisymmetry

Dominance is a well-known relation for t-norms (see e.g. [19]) and its usefulness has been demonstrated several times (see e.g. [9, 22]). It can be generalized to conjunctors without any problem.

Definition 4 *A conjunctor f_1 is said to dominate a conjunctor f_2, denoted $f_1 \gg f_2$, if for any $(x,y,z,t) \in [0,1]^4$ it holds that*

$$f_1(f_2(x,y),f_2(z,t)) \geq f_2(f_1(x,z),f_1(y,t)).$$

But conjunctors do not verify the same properties as t-norms do with respect to this property. For instance, not every conjunctor dominates itself (see [15]). Dominance and the classical order of binary operators are not related for conjunctors neither. The minimum for example (which is not the greatest conjunctor), dominates any other conjunctor (see [15]).

The notion of self-dominance of conjunctors is obviously equivalent to another well-known property: bisymmetry (see e.g. [19]).

Definition 5 *A conjunctor f is said to be bisymmetric if for any $(x,y,z,t) \in [0,1]^4$ it holds that*

$$f(f(x,y),f(z,t)) = f(f(x,z),f(y,t)).$$

ISBN: 978-989-95079-6-8

3.3 Implications

Given a t-norm T, an implication (also called R-implication or T-residuum) based on T is defined (see e.g. [18, 19]) as follows:

$$I_T(x,y) = \sup\{z \mid T(x,z) \leq y\}.$$

This concept aims at generalizing the notion of Boolean implication. The definition usually concerns only left continuous t-norms. When T is left-continuous it holds that $T(x,z) \leq y \Leftrightarrow z \leq I_T(x,y)$.

Given a t-conorm S, a co-implication based on S is defined (see again e.g. [18, 19]) as follows:

$$J_S(x,y) = \inf\{z \mid S(x,z) \geq y\}.$$

When S is right-continuous it holds that $S(x,z) \geq y \Leftrightarrow z \geq J_S(x,y)$.

Following these ideas we can introduce two operators associated not only for a t-norm or a t-conorm, but to any commutative conjunctor.

Definition 6 *Given a commutative conjunctor $f : [0,1]^2 \to [0,1]$, we define the following operator from $[0,1]^2$ into $[0,1]$:*

$$I_f(x,y) = \sup\{z \mid f(x,z) \leq y\},$$

and

$$J_f(x,y) = \inf\{z \mid f(x,z) \geq y\}.$$

Let us notice that I_f is an implication, but J_f is not a co-implication.

Lemma 1 *Given a commutative conjunctor f and its associated operators I_f and J_f, we have that*

1. *I_f and J_f are decreasing in their first arguments and increasing in their second arguments.*
2. *If f is left continuous then*

$$f(x,z) \leq y \quad \Leftrightarrow \quad z \leq I_f(x,y).$$

3. *If f is upper bounded by the minimum t-norm, then*

$$x \leq y \quad \Rightarrow \quad I_f(x,y) = 1.$$

Moreover, if f is left-continuous and has 1 as neutral element, then we have the equivalence:

$$x \leq y \quad \Leftrightarrow \quad I_f(x,y) = 1.$$

4 Transitivity of I

In this section we consider indifference relations. As it was shown in Subsection 2.2, for a generator i, the symmetric component of a large preference relation is obtained as $I = i(R, R^t)$. Here we study the transitivity we can derive for this relation I when we fix the transitivity of R defined by a conjunctor h.

We begin recalling some upper and lower bounds for the transitivity of I. We know from [15] that I is at least c_S-transitive when R is h-transitive for any conjunctor h. We also showed in that paper that the transitivity we can assure for I is defined by a conjunctor upper bounded both by the conjunctor h and by the generator i that builds I from R. These are upper and lower bounds. We next provide a characterization for the transitivity of I.

Theorem 3 *Let i be an increasing generator and h a conjunctor. For any reflexive fuzzy relation R with corresponding indifference relation I generated by means of i, it holds that*

$$R \text{ is } h\text{-transitive} \quad \Rightarrow \quad I \text{ is } f_h^i\text{-transitive,}$$

where the conjunctor f_h^i is:

$$f_h^i(x,y) = \inf_{\substack{1 \geq u \geq x \\ 1 \geq v \geq y}} \left(i\left(h(u,v), h\left(J_i(v,y), J_i(u,x)\right)\right)\right).$$

Moreover, if i is continuous, the previous one is the strongest possible implication.

An interesting problem is to know when the transitivity of R is inherited by I, that is, to know when, departing from an h-transitive R we can assure that I is also h-transitive. We have already answered that question in [15]:

Theorem 4 *Let i be an increasing generator and let h be a commutative conjunctor upper bounded by the minimum t-norm. The associated conjunctor f_h^i is equal to h if and only if i dominates h.*

In particular this result could be applied to $i = T_{\mathbf{M}}$. Thus, when the indifference relation is obtained from the reflexive relation R by the minimum t-norm, $I = \min\{R, R^t\}$, it satisfies the same transitivity as R does.

Not only dominance allows us to obtain some general result, also the usual order among conjunctor, as it is showed in the following corollary.

Corollary 1 *For any increasing and bisymmetric generator i and any conjunctor h upper bounded by the minimum t-norm, if $i \leq h$, then $f_h^i = i$.*

Since every t-norm is increasing and bisymmetric, this result can be applied in particular to any generator that is a t-norm. In that case, transitivity of R w.r.t. that t-norm is preserved.

We will now consider Theorem 4 and Corollary 1 to study the particular case of a t-norm of the Frank family as generator.

Corollary 2 *Let $T_\lambda^{\mathbf{F}}$ be a t-norm of the Frank family, for any $\lambda \in [0,\infty]$. It holds that $f_{T_{\mathbf{P}}}^{T_{\mathbf{M}}} = T_{\mathbf{P}}$, $f_{T_{\mathbf{L}}}^{T_\lambda^{\mathbf{F}}} = T_{\mathbf{L}}$ and $f_h^{T_\lambda^{\mathbf{F}}} = T_\lambda^{\mathbf{F}}$ for any $h \geq T_\lambda^{\mathbf{F}}$.*

Remark 1 *Combining the results in Corollary 2 leads to the following table, where the entries are the conjunctors f_h^i.*

$h \setminus i$	$T_{\mathbf{L}}$	$T_{\mathbf{P}}$	$T_{\mathbf{M}}$
$T_{\mathbf{L}}$	$T_{\mathbf{L}}$	$T_{\mathbf{L}}$	$T_{\mathbf{L}}$
$T_{\mathbf{P}}$	$T_{\mathbf{L}}$	$T_{\mathbf{P}}$	$T_{\mathbf{P}}$
$T_{\mathbf{M}}$	$T_{\mathbf{L}}$	$T_{\mathbf{P}}$	$T_{\mathbf{M}}$

5 Transitivity of P

In this section we focus on the strict preference relation P and the transitivity it can satisfy once the transitivity of R is fixed and defined by a conjunctor h.

Let us recall that the strict preference relation P is obtained from R by means of the generator i as follows: $P = R - i(R, R^t)$.

ISBN: 978-989-95079-6-8

We showed in [15] that the transitivity that can be assured for the strict preference relation associated to any h-transitive large preference relation R is upper bounded by the conjunctor h, i.e., we cannot assure for P a stronger transitivity than h-transitivity. Next we provide not a bound but the explicit expression for the transitivity of P.

Theorem 5 *Let i be an increasing 1-Lipschitz generator and h a commutative conjunctor with neutral element 1. For any reflexive fuzzy relation R with corresponding strict preference relation P generated by means of i, it holds that*

$$R \text{ is } h\text{-transitive} \quad \Rightarrow \quad P \text{ is } g_h^i\text{-transitive}\,,$$

where the conjunctor g_h^i is:

$$g_h^i(x,y) = \inf_{\substack{1 \geq u \geq x \\ 1 \geq v \geq y}} (h(u,v) - i(h(u,v), \min\{I_h(v, I_i(u, u-x)), I_h(u, I_i(v, v-y))\}))\,.$$

Moreover, the previous one is the strongest possible implication.

According to our result from [15], g_h^i should always be not greater than h. Again, we will focus our attention on the cases the transitivity of R is totally inherited by P.

Proposition 1 *Let h be a rotation-invariant conjunctor, i.e., for all $(x,y,z) \in [0,1]^3$ it holds that*

$$h(x,y) \leq z \quad \Leftrightarrow \quad h(y, 1-z) \leq 1-x.$$

If $i = T_{\mathbf{L}}$, then $g_h^{T_{\mathbf{L}}} = h$.

Two well-known rotation-invariant t-norms are $T_{\mathbf{L}}$ and the minimum nilpotent $T_{\mathbf{nM}}$ defined by $T_{\mathbf{nM}}(x,y) = \min\{x,y\}\chi_{\{(x,y)|x+y>1\}}$, where χ_B denotes the characteristic function of any set B. Therefore, from Proposition 1, $g_{T_{\mathbf{L}}}^{T_{\mathbf{L}}} = T_{\mathbf{L}}$ and $g_{T_{\mathbf{nM}}}^{T_{\mathbf{L}}} = T_{\mathbf{nM}}$. This last conjunctor was directly obtained in [16].

If $i = T_{\mathbf{M}}$, only some special conjuntors satisfy the equality $g_h^{T_{\mathbf{M}}} = h$. The minimum t-norm is one of those t-norms. The following theorem characterizes all the conjunctors that verify the previous equality.

Proposition 2 *Let h be a commutative conjunctor with neutral element 1. Then h satisfies the equality $g_h^{T_{\mathbf{M}}} = h$ if and only if h is of the form*

$$h_k^d(x,y) = \begin{cases} 0 & \text{if } max(x,y) < d \\ k \cdot min(x,y) & \text{if } max(x,y) = d \\ min(x,y) & \text{otherwise} \end{cases}$$

where $k \in \{0,1\}$ and $d \in [0,1) \cup \{k\}$.

For the particular case when the conjunctor defining the transitivity of R and the indifference generator i are the same continuous t-norm the thorny general expression obtained in Theorem 5 gets much simpler.

Theorem 6 *Let T be a continuous t-norm. For any reflexive fuzzy relation R with corresponding strict preference relation P generated by means of T, it holds that*

$$R \text{ is } T\text{-transitive} \quad \Rightarrow \quad P \text{ is } g_T^T\text{-transitive}\,,$$

where

$$g_T^T(x,y) = \inf_{0 \leq \alpha \leq \min\{1-x, 1-y\}} \max\{T(x+\alpha, y+\alpha) - \alpha, 0\}$$

and this is the strongest possible implication.

In [5] the operator g_T^T was studied in depth not only for a t-norm T but in general for any binary aggregation operator A. In that general case, it was denoted as $\mathcal{D}[A]$. We can find there that for the Frank family of t-norms the expression can be simplified:

Proposition 3 *In case $i = T_\lambda^{\mathbf{F}} = h$ for $\lambda \in [0,\infty]$, it holds that*

$$g_{T_\lambda^{\mathbf{F}}}^{T_\lambda^{\mathbf{F}}}(x,y) = \begin{cases} T_\lambda^{\mathbf{F}}(x,y), & \text{if } x+y > 1\,, \\ S_M(T_\lambda^{\mathbf{F}}\left(\frac{1+x-y}{2}, \frac{1+y-x}{2}\right) - \frac{1-x-y}{2}\,,\, 0), & \text{otherwise}. \end{cases}$$

In particular, it holds that $g_{T_{\mathbf{L}}}^{T_{\mathbf{L}}} = T_{\mathbf{L}}$ (also obtained as a consequence of Proposition 1) , $g_{T_{\mathbf{M}}}^{T_{\mathbf{M}}} = T_{\mathbf{M}}$ (directly obtained in [13]) and

$$g_{T_{\mathbf{P}}}^{T_{\mathbf{P}}}(x,y) = \begin{cases} T_{\mathbf{P}}(x,y) - \left(\frac{T_{\mathbf{L}}(1-x,1-y)}{2}\right)^2, \text{ if } \sqrt{x} + \sqrt{y} > 1 \\ 0, \text{ otherwise.} \end{cases}$$

Other particular cases involving the three most important t-norms, i.e. concerning $h \in \{T_{\mathbf{L}}, T_{\mathbf{P}}, T_{\mathbf{M}}\}$ and $i \in \{T_{\mathbf{L}}, T_{\mathbf{P}}, T_{\mathbf{M}}\}$ are presented in the following propositions.

Proposition 4 *If $i = T_\lambda^{\mathbf{F}}$ for $\lambda \in [0,\infty]$ and $h = T_{\mathbf{L}}$, it holds that*

$$g_{T_{\mathbf{L}}}^{T_\lambda^{\mathbf{F}}}(x,y) = T_{1/\lambda}^{\mathbf{F}}(T_{\mathbf{L}}(x,y), S_{\mathbf{M}}(x,y)).$$

Proposition 5 *If $i = T_\lambda^{\mathbf{F}}$ and $h = T_{\mathbf{M}}$, it holds, for any $\lambda \in [0,\infty]$, that*

$$g_{T_{\mathbf{M}}}^{T_\lambda^{\mathbf{F}}}(x,y) = T_{\mathbf{nM}}^{\varphi_{1/\lambda}}(x,y)$$

where φ_λ is the automorphism of the interval $[0,\infty]$ defined as follows

$$\varphi_\lambda(x) = \begin{cases} x & \text{if } \lambda = 0, \\ \sqrt{x} & \text{if } \lambda = 1, \\ \frac{x+1}{2}\chi_{(0,1]}(x) & \text{if } \lambda = \infty, \\ \log_\lambda\left(\sqrt{\frac{\lambda^x - 1}{\lambda - 1}}\,(\lambda - 1) + 1\right) & \text{otherwise}, \end{cases}$$

and $T_{\mathbf{nM}}^{\varphi_{1/\lambda}}$ is the transformation by the automorphism $\varphi_{1/\lambda}$ of the t-norm $T_{\mathbf{nM}}$. Moreover, it holds that $g_{T_{\mathbf{M}}}^{T_\lambda^{\mathbf{F}}}$ is a t-norm.

As a consequence of this proposition, we obtain that $g_{T_{\mathbf{M}}}^{T_{\mathbf{L}}} = T_{\mathbf{nM}}$ and

$$g_{T_{\mathbf{M}}}^{T_{\mathbf{P}}}(x,y) = T_{\mathbf{nM}}^{\varphi_1}(x,y) = \begin{cases} \min\{x,y\} & \text{if } \sqrt{x} + \sqrt{y} > 1, \\ 0 & \text{otherwise}\,. \end{cases}$$

ISBN: 978-989-95079-6-8

In order to complete the study of all the combinations of the three most important t-norms, two cases are missing, namely $g_{T_{\mathbf{P}}}^{T_{\mathbf{L}}}$ and $g_{T_{\mathbf{P}}}^{T_{\mathbf{M}}}$, which will be revisited in Propositions 6 and 7.

Proposition 6 *It holds that*

$$g_{T_{\mathbf{P}}}^{T_{\mathbf{L}}}(x,y) = T_{\mathbf{M}}\left(T_{\mathbf{P}}(x,y), \frac{T_{\mathbf{L}}(x,y)}{T_{\mathbf{M}}(x,y)}\right) \cdot \chi_{(0,1]}(T_{\mathbf{M}}(x,y)).$$

When $T_{\mathbf{L}}$ is replaced by $T_{\mathbf{M}}$ as generator, the expression of $g_{T_{\mathbf{P}}}^{i}$ gets much more complicated.

Proposition 7 *It holds that*

$$g_{T_{\mathbf{P}}}^{T_{\mathbf{M}}}(x,y) = \max\{\min_{\substack{1 \geq u \geq x \\ 1 \geq v \geq y}} \left[uv - \min\left\{\tfrac{u-x}{v}, \tfrac{v-y}{u}\right\}\right], 0\} \cdot \chi_{(0,1]^2}(x,y).$$

Remark 2 *Combining the results in Propositions 1, 3, 4, 5, 6 and 7 leads to the following table, where the entries are the conjunctors g_h^i.*

$h \setminus i$	$T_{\mathbf{L}}$	$T_{\mathbf{P}}$	$T_{\mathbf{M}}$
$T_{\mathbf{L}}$	$T_{\mathbf{L}}$	$T_{\mathbf{P}}(T_{\mathbf{L}}, S_{\mathbf{M}})$	$T_{\mathbf{L}}(T_{\mathbf{L}}, S_{\mathbf{M}})$
$T_{\mathbf{P}}$	$g_{T_{\mathbf{P}}}^{T_{\mathbf{L}}}$	$g_{T_{\mathbf{P}}}^{T_{\mathbf{P}}}$	$g_{T_{\mathbf{P}}}^{T_{\mathbf{M}}}$
$T_{\mathbf{M}}$	$T_{\mathbf{nM}}$	$T_{\mathbf{nM}}^{\varphi_1}$	$T_{\mathbf{M}}$

6 Conclusions

This paper combines very general results with propositions concerning the most relevant particular cases. The general results can be applied to any conjunctor employed to define the transitivity of a large preference relation and any generator used to decompose that relation. The specific results concern the most important particular cases: those for which the conjunctors and the generators are the most important t-norms. The most general theorems we have introduced close the study of the transitivity that can satisfy the symmetric and asymmetric components of a (large) preference relation since those theorems involve (almost) any conjunctor and any generator we can use. The results concerning more particular operators provide more easy-to-use expressions for the most usual conjunctors and generators, i.e. for some t-norms. Only the last example leads to an unwieldy expression.

Despite the ugly general formulae obtained in Theorems 3 and 5, we have already proven that those conjunctors satisfy interesting properties for some particular cases. In future works we would like to study in depth the general expressions and the properties they satisfy.

Acknowledgment

The research reported on in this paper has been partially supported by Project FEDER-MEC-MTM2007-61193.

References

[1] C. Alsina. On a family of connectives for fuzzy sets. In *Fuzzy Sets and Systems*, 16:231–235, 1985.

[2] K. J. Arrow Social Choice and individual values. Wiley, 1951.

[3] M. Dasgupta and R. Deb. Transitivity and fuzzy preferences. In *Social Choice and Welfare*, 13:305–318, 1996.

[4] M. Dasgupta and R. Deb. Factoring fuzzy transitivity. In *Fuzzy Sets and Systems*, 118:489–502, 2001.

[5] B. De Baets, H. De Meyer, and S. Díaz. On an idempotent transformation of aggregation functions and its application on absolutely continuous archimedean copulas. In *Fuzzy Sets and Systems*, 160:733–751, 2009.

[6] B. De Baets and J. Fodor. Twenty years of fuzzy preference structures (1978-1997). *Belg. J. Oper. Res. Statist. Comput. Sci.*, 37:61–82, 1997.

[7] B. De Baets and J. Fodor. Generator triplets of additive fuzzy preference structures. In *Proc. RelMiCS6*.

[8] B. De Baets and J. Fodor. Additive fuzzy preference structures: the next generation. In *Principles of Fuzzy Preference Modelling and Decision Making*, 15–25, Academia Press, 2003.

[9] B. De Baets and R. Mesiar. t-partitions. In *Fuzzy Sets and Systems*, 97:211–223, 1998.

[10] B. De Baets and B. Van de Walle. Weak and strong fuzzy interval orders. *Fuzzy Sets and Systems*, 79:213–225, 1996.

[11] B. De Baets and B. Van de Walle. Minimal definitions of classical and fuzzy preference structures. In *Proceedings of the Annual Meeting of the Horth Amercan Fuzzy Information Precessing Society (Syracuse, New York, USA*, 299–304, 1997.

[12] B. De Baets, B. Van de Walle, and E. Kerre Fuzzy preference structures without incomparability. *Fuzzy Sets and Systems*, 76:333-348, 1995.

[13] S. Díaz, B. De Baets, and S. Montes. Additive decomposition of fuzzy pre-orders. *Fuzzy Sets and Systems*, 158:830–842, 2007.

[14] S. Díaz, B. De Baets, and S. Montes. On the compositional characterization of complete fuzzy pre-orders. *Fuzzy Sets and Systems*, 159:2221–2239, 2008.

[15] S. Díaz, S. Montes, and B. De Baets. Transitivity bounds in additive fuzzy preference structures. *IEEE Transactions on fuzzy systems*, 15(2):275–286, 2007.

[16] S. Díaz, S. Montes, and B. De Baets. Transitive decomposition of fuzzy preference relations: the case of nilpotent minimum. *Kybernetika*, 40:71–88, 2004.

[17] J. Fodor. Traces of binary fuzzy relations. *Fuzzy Sets and Systems*, 50:331–341, 1992.

[18] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers, 1994.

[19] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, Dordrecht, 2000.

[20] B. Monjardet. Axiomatiques et propriétés des quasi-ordres. *Math. Sci. Hum.*, 63:51–82, 1978.

[21] M. Roubens and Ph. Vincke. Preference modelling. In *Lecture Notes in Economics and Mathematical Systems*, volume 250. Springer, Berlin, 1985.

[22] S. Saminger, R. Mesiar, and U. Bodenhofer. Domination of aggregation operators and preservation of transitivity. *Int. J. Uncertainty Fuz. and Knowledge-Based Syst.*, 10:11–35, 2002.

[23] B. Van de Walle, B. De Baets, and E. Kerre A plea for the use of Łukasiewicz triplets in the definition of fuzzy preference structures. Part 1: General argumentation. *Fuzzy Sets and Systems*, 97:349–359, 1998.

[24] B. Van de Walle, B. De Baets, and E. Kerre Characterizable fuzzy preference structures. *Ann. Oper. Res.*, 80:105–136, 1998.

[25] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

ISBN: 978-989-95079-6-8

Fuzzy Tendency based Time Series Model for Forecasting Server Traffic

Nadezhda G. Yarushkina[1], Tagir R.Unusov[2], Tatyana V.Afanasyeva[3]

1 Ulyanovsk State Technical University
32, Severny Venets street, Ulyanovsk, 432027, Russia
2 Ulyanovsk State Technical University
32, Severny Venets street, Ulyanovsk, 432027, Russia
3 Ulyanovsk State Technical University
32, Severny Venets street, Ulyanovsk, 432027, Russia
Email: jng@ulstu.ru, tagir@ulstu.ru, tv.afanaseva@mail.ru

Abstract *- For modeling of change of terminal server load, the approach including representation of time series of server parameters in the form of fuzzy time series is used. Further in the article the analysis of fuzzy time series is considered. The model of fuzzy tendencies is offered for terminal server traffic modeling. This model reflects change of the volume of terminal server traffic expressed linguistically, and it is used for its forecasting. The results of the forecast permit to determine linguistically expressed fuzzy tendencies on the basis of the offered model.*

Keywords— Fuzzy Time Series, Fuzzy Tendency, Fuzzy Neural Network, Forecasting, Terminal Server.

1 Introduction

During the last decade development of computing networks (CN) of different levels (from local area networks to global area networks) has resulted in the fact that the quality of production processes depends on the quality of network services more and more. Correspondence of scales of trade processes to the computer network can be achieved due to constant analysis of quality of CN functioning (including the application layer) and forecasting its development. Such analysis of existent network often rests upon results of time series (TS) analysis of server parameters.

Time series of server parameters have a number of peculiarities: non-stationarity, non-homogeneity, the complex form of dynamics, which limit the applicability of classical statistic methods of TS analysis such as regression and spectral analysis, exponential smoothing. In this connection, at present promising technologies of time series analysis are intelligent methods which permit to use expert knowledge and have no rigid requirements to initial data.

Modeling of experts' knowledge in the form of linguistic terms (evaluations) formalized as fuzzy sets became a prerequisite to TS modeling in the form of fuzzy time series.

Since 1993 a number of works have been dedicated to forecasting of fuzzy TS [1 - 4] constructed on the basis of numerical TS. In these investigations let us note the following problems which can be solved: increase of accuracy of numerical forecast owing to specification of the quantity and form of membership functions, decrease of computing owing to decrease of the number of fuzzy rules, increase of the level of automation when solving forecast problems. The indicated problems can be solved owing to application of methods of intelligent data analysis such as a fuzzy cluster analysis, genetic algorithms and neural networks.

The study [5] proposes the fuzzy neural network with initial weights generated by genetic algorithm (GFNN) for the sake of learning fuzzy IF–THEN rules. The result from GFNN is further integrated with an ANN artificial neural networks (ANN) forecast using the time series data from another ANN. Model evaluation results for a convenience store (CVS) company indicate that the proposed system can perform more accurately than the conventional statistical method and a single ANN.

In the work [6] the method of forecasting of TS inflation on the basis of fuzzy systems and a neural network is offered. To determine the fuzzy relation of the first-order fuzzy time series model, the neural network with SCL clustering technique was used to derive the rules directly from the database. The method may be of real usefulness in practical applications, where usually the expert can not explain linguistically, what control actions the process takes or there is no knowledge of the process.

To forecast fuzzy time series, in study [7] a backpropagation neural network is applied because of its nonlinear structures. Authors propose a hybrid model consisting of a neural network approach to forecast the known patterns as well as a simple method to forecast the unknown patterns.

Then the authors published the work [8] in which they proposed bivariate models in order to improve forecasting. The stock index and its corresponding index futures were taken as the inputs to forecast the stock index for the next day.

In this work there is offered a new hybrid forecasting model of TS in the form of a time series of fuzzy tendencies representing context-dependent interval linguistic evaluations of the traffic of the terminal server of a complex network. For the time series of FT, the model of forecasting of the first order is defined in terms of fuzzy tendencies to make more reasonable decisions on planning of network structure change.

In the part 2 the model and base operations of FT processing are defined.

In the part 3 a fuzzy neural network for generation of fuzzy rules of forecasting fuzzy tendencies is offered. The developed mathematical simulation model of the terminal server as an element of a computing network on the basis of time series of fuzzy tendencies presented in part 4 permits to forecast the traffic of the terminal server both in terms of fuzzy tendencies and in terms of the initial TS.

ISBN: 978-989-95079-6-8

2 Models and base operations of fuzzy tendencies processing

In the work [8] the evaluation of TS is analyzed which is expressed in linguistic terms, the concepts fuzzy time series (FTS) and fuzzy tendency (FT) are introduces for that. The fuzzy tendency of a fuzzy time series is the fuzzy label which expresses the dynamics (systematic motion) of a fuzzy time series. The concept "tendency", being modeled by the fuzzy set, can be described by the set of constructed empiric rules. To find the fuzzy tendencies, it is offered to use Fuzzy Neurons Network (FNN). The analysis of the FT permits to forecast the behavior of the time series on the basis of process dynamics by means of prediction of tendencies of development of the object being studied. With it forecasting of numerical values of TS is not foreseen, and it can be considered as a shortcoming of the offered methods.

The base operations of fuzzy tendencies processing are algorithmic operations of fuzzy tendencies processing, namely the operation of forming the time series of fuzzy tendencies by the initial time series and the inverse operation of generation of the series – a representative of the fuzzy tendency.

Definition 1. A fuzzy time series $\widetilde{Y}$ is the ordered sequence of observations of some phenomenon the character of which changes in the course of time if values which some quantity takes on at the instant of time are expressed by the fuzzy label:

$$\widetilde{y}_t^i = \{(\widetilde{y}^i, U_y, \mu_i(y_t)), t\},$$

where $\widetilde{y}^i$ is the *i*-th element of the term-set of the linguistic variable $(\widetilde{Y}, T_y, U_y, G_y, M_y)$.

Thus, a FTS is the vector time series of values of all fuzzy variables:

$$\widetilde{Y} = \{\widetilde{y}_t\},$$

where $\widetilde{y}_t = \{\widetilde{y}_t^1, ..., \widetilde{y}_t^n\}$, *n* is the number of terms.

Definition 2. Fuzzy tendency. Let $\widetilde{y}_\Delta = \{\widetilde{y}_1, ..., \widetilde{y}_m\}$ be the fuzzy time series of the linguistic variable $(\widetilde{Y}, T_y, U_y, G_y, M_y)$, $\widetilde{Y}_\Delta = \{\widetilde{y}_\Delta\}$ is the set of fuzzy time series of the same length. Then the fuzzy tendency τ, determined on the $\widetilde{Y}_\Delta$, is the aggregate of ordered pairs:

$$\tau = \{\widetilde{y}_\Delta, \mu_\tau(\widetilde{y}_\Delta)\},$$

where $\mu_\tau(\widetilde{y}_\Delta)$ is the degree of membership of $\widetilde{y}_\Delta$ to FT.

Determining the fuzzy tendency on all intervals [*t-m+1, t*] of the time series and positioning the beginning and the end of the interval to the time scale, we receive the time series of the fuzzy tendency.

Definition 3. The time series of a fuzzy tendency (TSFT). Let $\{\widetilde{y}_{\Delta t}\}$ be the set of fuzzy time series which length is equal to *m*, where $\widetilde{y}_{\Delta t} = \{\widetilde{y}_{t-m+1}, ..., \widetilde{y}_t\}$, $\widetilde{y}_t \in \widetilde{Y}$ t. Then the time series of a fuzzy tendency is the ordered in time fuzzy set:

$$\tau_t = \{t, \mu_\tau(\widetilde{y}_{\Delta t})\}.$$

For the linguistic variable FT, the time series of fuzzy tendencies is determined as the vector TS of values of all kinds of fuzzy tendencies:

$$\tau_t = \{t, \tau_t^1, ..., \tau_t^p\}.$$

Let us assume that there is dependence between fuzzy tendencies observed at different instants of time. Using the scheme of the differed equation (1) let us represent the model of the time on the basis of FT for one variable:

$$\tau_t = f(\tau_{t-1}, ...\tau_{t-l}), \quad (1)$$

where *l* is the time interval. For construction of the model of TS on the basis of FT, it is necessary to solve the following problems:

-to determine the procedure of fuzzyfication and defuzzyfication of the TS;

-to determine the procedure of description and identification of the FT;

-to determine the procedure of receiving the crisp time series from values of FT;

-to identify the functional dependence.

Receiving a FTS from a crisp TS is based on the operation of fuzzyfication, i.e. transition from quantitative values to linguistic evaluations. The procedure of fuzzyfication is widely enough described in literature. Let us denote the process of finding the value of membership function of elements *Uy* to the terminal set T_y by the functional Fuzzy:

$$\widetilde{Y} = Fuzzy[Y].$$

The functional supposes solving the problems of linguistic terms description and rules of use, choosing the corresponding membership functions. When determining the linguistic terms the following methods can be used: subjective description of the universe U_y on the basis of expert knowledge about the system and descriptive terminology accepted in the field being investigated; uniform partition of the universe (into 3-7 intervals, as a rule); clusterisation of values of the time series for determination of the most significant intervals on the universe.

Defuzzyfication is the problem inverse to fuzzyfication, i.e. receiving the crisp TS from a fuzzy one, is given by the functional deFuzzy:

$$Y' = deFuzzy[\widetilde{Y}].$$

The problem of determination of the FT is fundamental in the construction of the system of data analysis, decision of which will permit to start disclosing the connections between parameters of the object being investigated. This problem consists of the following sub-problems:

-determination of linguistic variables of the FT;

-construction of the time series of the FT.

Let us denote solving the problem by the functional Tend:

$$\tau = Tend[\widetilde{Y}_\Delta]$$

For determination of the linguistic variable expressing the fuzzy tendency, it is necessary to single out the typical behavior of the fuzzy time series, that is to make up the terminal set. The analysis of the initial series for evaluation of the dynamics can be done by the following methods:

-using the evaluations of the dynamics (increase, decrease, stability);

ISBN: 978-989-95079-6-8

-subjective description on the basis of expert knowledge about the systematic behavior of system parameters;

-clusterisation of values of the fuzzy time series for determination of the most significant tendencies on the universe.

For construction of TSFT, it is necessary not only to single them out (to denote linguistic terms), but to give a description and be able to find the tendency on the FTS. Thus, the functional Tend includes the tool of FT description, the algorithm of finding the correspondence of fuzzy time series to the chosen description, that is the algorithm of fuzzy evaluation.

For the analysis and construction of crisp TS by the fuzzy model, the operation of receiving the FTS from tendency evaluation is necessary. Let us denote solving of this problem by the functional *deTend*, which is inverse to the functional Tend:

$$\widetilde{Y}`_{\Delta} = deTend[\tau],\ \widetilde{Y}`_{\Delta} \in \widetilde{Y}_{\Delta},$$

where $\widetilde{Y}`_{\Delta} = \{\widetilde{y}`_{\Delta}\}$ is the set of typical (characteristic) fuzzy time series for FT which have the maximum value of membership function $\mu_{\tau}(\widetilde{y}`_{\Delta}) = \max(\mu_{\tau}[\widetilde{Y}_{\Delta}])$.

Using the functional of defuzzyfication *deFuzzy*, receiving the fuzzy time series from tendencies *deTend* solves the problem of crisp TS forecasting.

Let us denote the aggregate of components and equations:

$$\widetilde{y}_t^i = Fuzzy[y_t],\ y_t = deFuzzy[\widetilde{y}_t^1,...\widetilde{y}_t^n], i = 1..n,$$

$$\tau_t^j = Tend[\widetilde{y}_{t-m_j+1},...,\widetilde{y}_t],$$

$$\widetilde{y}_t^i = deTend[\tau_t,...,\tau_{t+m-1}], m = \max(m_j),\ j = 1..p,$$

$$\tau_t = f(\tau_{t-1},...\tau_{t-l}),$$

where n is the number of terms of FTS, p is the number of terms of FT, m_j is the interval of definition of FT, l is the time log, by the *model of fuzzy tendencies* (MFT) with characterizing parameters (n, p, m, l). In more detailed form these parameters can be denoted as (n, $\{p_k\}$, $\{m_k\}$, l), where p_k is the number of kinds of tendencies which have the interval of definition m_k.

For the analysis and construction of the crisp time series by the fuzzy model (the functional *deTend)* let us make the fuzzy time series which has the maximum value of membership function in the form of the rule $\widetilde{y}_{\Delta t} = (\tau_t) \circ R_2$ correspond to each kind of fuzzy tendency. For a example, the procedure *deTend* can be represented in the form of the system of rules:

if τ_t is jump, then y_{t-2} is low,

if τ_t is jump, then y_{t-1} is high,

if τ_t is jump, then y_t is low.

Let us represent the equation (1) also by the fuzzy relation:

$$\tau_t = (\tau_{t-1},...\tau_{t-l}) \circ R_3.$$

The MFT is realized by multilevel system of logical relations:

$$R_1 \Rightarrow R_3 \Rightarrow R_2,$$

where outputs in the form of fuzzy variables of one set of rules are applied to inputs of the next set of rules without defuzzyfication and fuzzyfication. Transformation into fuzzy and crisp values happens only in relations R_1 and R_2 correspondingly.

3 Fuzzy neural network

Let us use classical fuzzy neurons in which operations of addition and multiplication are replaced by triangular norms:

AND-neuron $\beta = T(S(\tau_1, w_1), S(\tau_2, w_2))$,

OR-neuron $\tau^0 = S(T(\beta_1, z_1), T(\beta_2, z_2))$.

It can be interpreted in the linguistic form correspondingly as:

if ($\tau 1$ or $w1$) and (τ_2 or w_2) then β,

if (β_1 *and* z_1) or (β_2 *and* z_2) then τ_0.

On the basis of such neurons it is possible to construct the network of logical inference by Mamdani for finding the fuzzy tendency τ by adding weight coefficients. In a formalized way such network can be expressed in the following form:

$$\tau^0 = \underset{i=1}{\overset{k}{S}}[T(\beta_i, z_i)],\ \beta_i = \underset{j=1}{\overset{m}{T}}[S(\tau_j, w_{j,i})],$$

where T is the operator of conjunction, S is the operator of disjunction, k is the number of rules, m is the number of inputs. Weight coefficients are interpreted in the following way: z_i is the degree of influence of i-th rule on the general result (0 – has no influence, 1 – has influence); $w_{j,i}$ are degrees of no influence of j-th input on i-th rule (0 – has influence, 1 – has no influence). The basic idea of FNN learning consists in the iterative procedure of weights optimization (z_i, $w_{j,i}$) and removal of insignificant connections (network reduction), as a result of which the necessary composition and number of rules are formed.

The order of search fuzzy dependences on the basis of FNN is the following:

-Initialization of the fuzzy neural network.

-Learning: optimization of network weights.

-Analysis of the network: network reduction.

At the initialization stage it is necessary to determine input variables: kinds of tendencies and the time log. It is necessary to generate an excessive number of "complete" rules, which include all inputs with arbitrary weights in the interval (0,1).

The stage of learning is the process of FNN weights change on the basis of the learning sample. Network learning is possible by the method of back error propagation. For this purpose let us determine the function of error $E = \frac{1}{2}(\tau^0 - \tau)^2$ which is to be minimized by the method of gradient lowering. For decrease of the squared error E, it is necessary to change weights (z_i, $w_{j,i}$) in the direction of antigradient of the function E:

$$w_{j,i}^{+1} = w_{i,j} - \eta \frac{\partial E}{\partial w_{j,i}},\ z_i^{+1} = z_i - \eta \frac{\partial E}{\partial z_i},$$

where η is the speed of learning with the limitation $w_{j,i}^{+1} \in [0,1]$, $z_i^{+1} \in [0,1]$. For calculation of the derivative error, it is necessary to choose the corresponding functions of t-norms and t-conorms. The indication of completion of learning is reaching the error level.

ISBN: 978-989-95079-6-8

At the stage of analysis processing of FNN is carried out with the purpose of its simplification. The principle of removal of insignificant connections and neurons from the network forms the basis of reduction algorithms. One of simplest reduction methods is the method of projections which is realized in the following way. The synoptic weight is nulled if its value has fallen into the given range

$$w_{j,i} = \begin{cases} 1, w_{j,i} \geq (1-\varepsilon) \\ w_{j,i}, w_{j,i} < \varepsilon \end{cases}, \; z_i = \begin{cases} 0, z_i \leq \varepsilon \\ z_i, z_i > \varepsilon \end{cases},$$

where ε is some constant. On the basis of the chosen level of weights ε connections "input – AND-neuron", when $w_{j,i}=1$, and "AND-neuron – OR-neuron", when $z_i = 0$, are removed. Not used inputs and rules are also removed.

As a result the system of logical inference is got, which is the explaining function in the model of time series. Fuzzy rules are easily interpreted for an expert since they are expressed in terms inherent to the investigated field.

4 Server traffic modeling on the basis of the time series of fuzzy tendencies

For time series analysis on the basis of the offered model of a fuzzy time series, the software *FuzzyTendNet* is created.

Thus, for each time series a user assigns the set of linguistic terms with corresponding membership functions (trapezoid function form is used) and the set of fuzzy tendencies. The possibility of using several rules of identification and inverse rules for each tendency permits to describe various interval expert evaluations flexibly enough.

The equation of the model of a time series is the fuzzy neural network which has subnetworks for each output tendency. Each subnetwork is the system of logical deduction which consists of the set of rules. These rules can be formed manually and in the automatic mode with optimization on the basis of the algorithm of back error propagation. The method of projections is used for cancellation of the network. The following logical connectives (functions of triangular norms) are used: minimum and maximum — for rules of identification and inverse rules, product and probabilistic sum — for FNN.

For modeling of the traffic of the terminal server on the basis of the offered approach, seven parameters of the work of the terminal server were chosen, and time series were formed from them.

So, the server parameters of work were chosen (Table 1) and the statistics with 15 seconds interval during one day was collected.

Table 1. Description of variables

Parameter
X1 - Memory\ Pages exchange per second
X2 - Record accesses to disc per second
X3 - Read accesses from disc per second
X4 - % of processor load
X5 - Reading operations
X6 - Recording operations
Y - Traffic

For all parameters fuzzy variables of FTS describing the values "high" and "low" are determined. Fuzzy tendencies "load", "idle" are determined. Additional fuzzy tendencies "increase" and "decrease" are also determined for traffic. In Table 2 description of functionals of the MFT is presented.

Let us note the following:

-the functional deTend is described only for the dependent variable (traffic);

-FT determined on the unit interval, which have the semantic load analogous to variables of FTS, are used in the model.

Table 2. Description of tendencies of terminal-server model parameters change

Parameter	FTS	TSFT	
	$\tilde{y}$	τ	Tend
X1	low	idle	low(1) &low(2)
	high	load	high(1) & high(2)
X2	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
X3	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
X4	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
X5	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
X6	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
Y	low	idle	low(1) & low(2)
	high	load	high(1) & high(2)
		increase	low(1) & high(2)
		decrease	high(1) & low(2)

For comparison of models on the basis of fuzzy time series and fuzzy tendencies, a number of experiments were conducted.

The conclusion can be made that the model based on fuzzy tendencies describes the modeled process more successfully. On Fig. 1 the forecast for one step forward is presented for two models.

ISBN: 978-989-95079-6-8

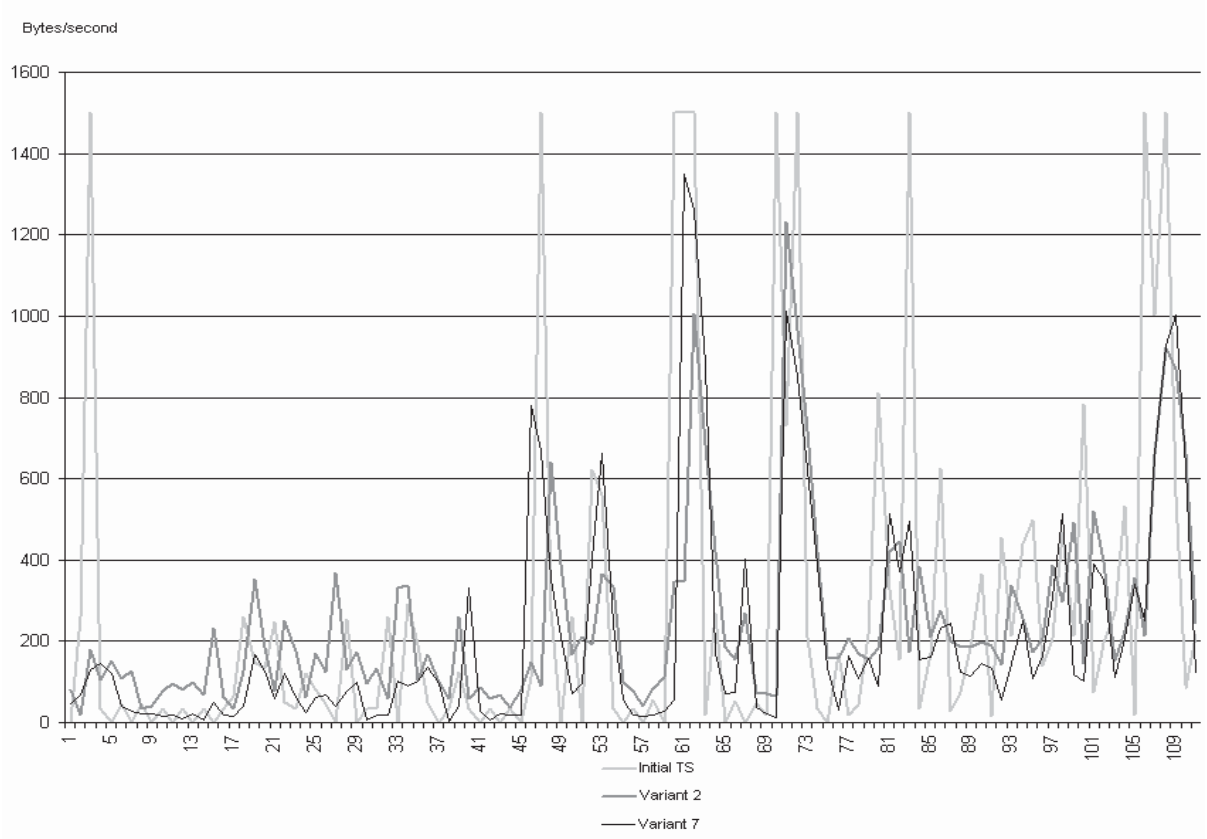


Figure 1:The graph of tested and forecasted TS

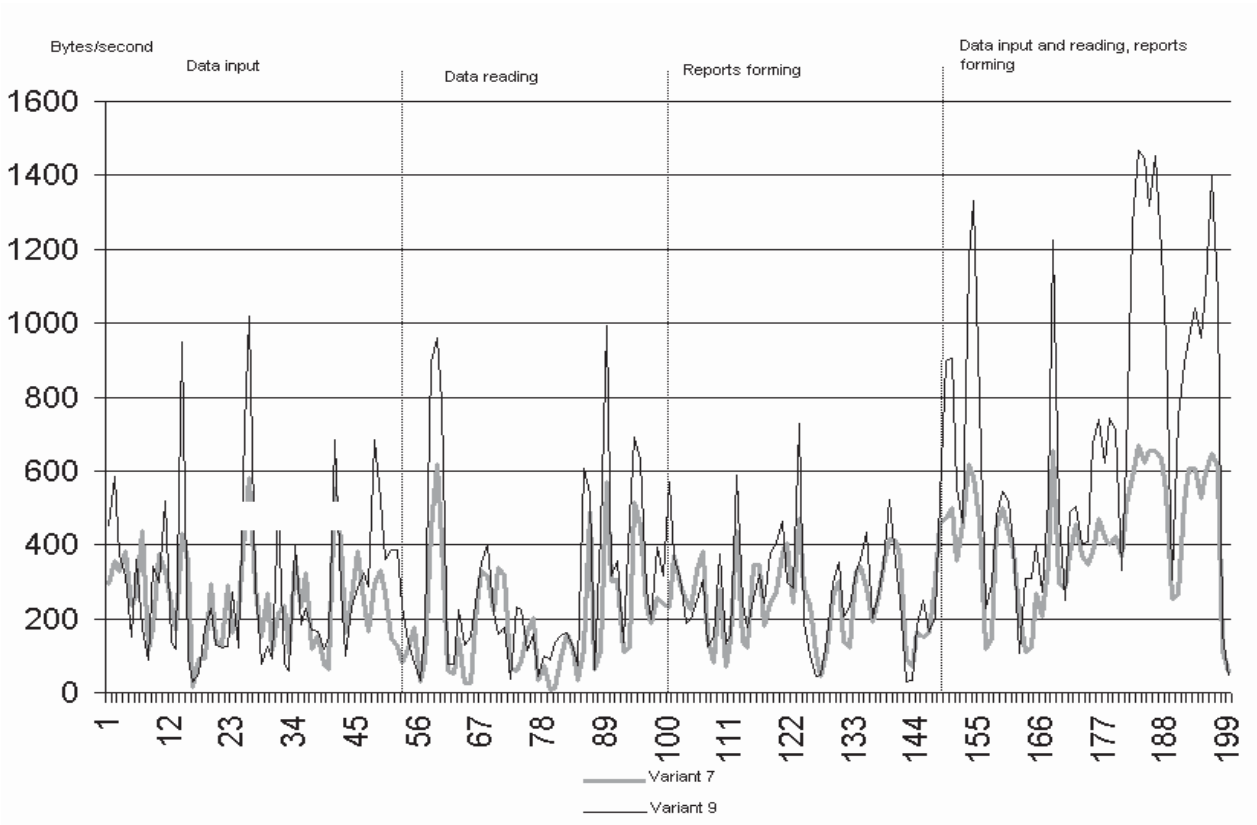


Figure 2:Load modeling

For server traffic volume modeling, the following modes of server work were described by the expert: data input, data reading, reports forming — with high values of corresponding groups of parameters. On Fig 2 results of modeling are presented.

By fuzzy tendencies low and high values of the traffic and its crisp values have been forecasted in different modes of work. The analysis of linguistic series of fuzzy tendencies showed increase of the traffic of the terminal server in the mode of data input and reports forming.

ISBN: 978-989-95079-6-8

Thus the conclusion can be made that increase of users performing the mentioned operations will be accompanied by increase of the traffic. When performing all operations on the server simultaneously, we can simultaneously observe both high and low traffic of the network, which can be explained by instability in this mode of work.

5 Conclusions

Thus, the time series of fuzzy tendencies is an effective indicator of development of a modeled process in a complex technical system described qualitatively by linguistic terms. Basic operations of fuzzy tendencies processing are algorithmic operations of fuzzy tendencies processing, namely the operation of forming the time series of fuzzy tendencies by initial time series and the inverse operation of generation of the series — a representative of the fuzzy tendency. The fuzzy neural network of the offered architecture is an effective generator of rules of fuzzy tendencies identification. The developed mathematical simulation model of the terminal-server as the element of the computing network on the basis of time series of fuzzy tendencies permits to forecast processor load, the outgoing and incoming traffic of the server.

References

[1] Q. Song, B. Chissom, Fuzzy time series and its models, Fuzzy Sets and Systems 54 (1993) 269–277.

[2] K. Huarng. Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets and Systems, 2001, 123: 387–394.

[3] K. Huarng, Heuristic models of fuzzy time series for forecasting, Fuzzy Sets and Systems 123 (2001) 369–386.

[4] Şah, M., Degtiarev, K.Y. : Forecasting Enrollment Model Based on First-Order Fuzzy Time Series. Proc. Int. Conf. Computational Intelligence (ICCI) (2004) 375–378

[5] R. J. Kuo . A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm// European Journal of Operational Research Volume 129, Issue 3, 16 March 2001, Pages 496-517

[6] Dušan Marček, Modelling of inflation fuzzy time series using the competitive neural network, Proc. International conference on softcomputing applied in computer and economic environments (ISCS) (2004) 107-111.

[7] Kun-Huang Huarng, Tiffany Hui-Kuang Yu. The application of neural networks to forecast fuzzy time series \\ Physica A: Statistical Mechanics and its Applications, Volume 363, Issue 2, 1 May 2006, Pages 481-491

[8] Tiffany Hui-Kuang Yu, Kun-Huang Huarng. A bivariate fuzzy time series model to forecast the TAIEX \\ Expert Systems with Applications Volume 34, Issue 4, May 2008, Pages 2945-2952

[9] N.G. Yarushkina Principles of the theory of fuzzy and hybrid systems: Training aid. - Moscow: Finances and statistics, 2004.

ISBN: 978-989-95079-6-8

Continuity and additivity of the trapezoidal approximation preserving the expected interval operator

Adrian I. Ban Lucian C. Coroianu
Department of Mathematics and Informatics, University of Oradea,
Universităţii 1, 410087 Oradea, Romania
Email: aiban@uoradea.ro, luccoro@yahoo.com

Abstract— *The nearest trapezoidal fuzzy number to a fuzzy number, with respect to a well-known metric and preserving the expected interval, was determined in recent articles. In the present paper the properties of additivity and continuity of the trapezoidal approximation operator are studied.*

Keywords— Additivity, Approximation, Continuity, Fuzzy number, Trapezoidal fuzzy number.

1 Introduction

In the papers [1] and [10] the problem to find the nearest (with respect to a well-known metric) trapezoidal fuzzy number $T(A)$ to a fuzzy number A such that $EI(T(A)) = EI(A)$, where $EI(B)$ denotes the expected interval of a fuzzy number B, was completely solved. Algorithms for computing the proper approximations are proposed in [6]. The properties of translation and scale invariance, identity, nearness criterion, expected interval invariance, order invariance, correlation invariance, uncertainty invariance, value and ambiguity invariance were considered in [1]. The properties of additivity and continuity are studied in the present paper.

In Section 3 we prove, by a simple example, that the additivity is not satisfied. The main result in this section contains a property of partial additivity. In the paper [5] the property of continuity of the approximations of fuzzy numbers is considered of extreme importance in applications, especially in fuzzy control systems, where it is sometimes indicated as robustness. The authors of the paper [2] pointed out that the method to study the continuity of the trapezoidal approximation without condition in [11] is not applicable in the case of trapezoidal approximation preserving the expected interval. In Section 4 we prove that the discussed approximation operator has the property to be Lipschitz, such that it is continuous.

2 Preliminaries

A fuzzy number A is a fuzzy subset of the real line $\mathbb{R}$ with the membership function μ_A which is normal, fuzzy convex, upper semicontinuous and supp A is bounded, where supp $A = cl\{x \in \mathbb{R} : \mu_A(x) > 0\}$ and cl is the closure operator.

Every λ-cut, $\lambda \in]0,1]$ of a fuzzy number A is a closed interval $A_\lambda = [A^-(\lambda), A^+(\lambda)]$, where

$$A^-(\lambda) = \inf\{x \in \mathbb{R} : \mu_A(x) \geq \lambda\},$$
$$A^+(\lambda) = \sup\{x \in \mathbb{R} : \mu_A(x) \geq \lambda\}.$$

We denote

$$A_0 = \left[A^-(0), A^+(0)\right] = \text{supp } A.$$

For two arbitrary fuzzy numbers A, B

$$A_\lambda = \left[A^-(\lambda), A^+(\lambda)\right]$$

and

$$B_\lambda = \left[B^-(\lambda), B^+(\lambda)\right]$$

the addition $A + B$ is introduced by

$$(A+B)_\lambda = \left[A^-(\lambda) + B^-(\lambda), A^+(\lambda) + B^+(\lambda)\right]$$

and the quantity

$$D^2(A,B) = \int_0^1 \left(A^-(\lambda) - B^-(\lambda)\right)^2 d\lambda + \int_0^1 \left(A^+(\lambda) - B^+(\lambda)\right)^2 d\lambda \quad (1)$$

gives a distance between A and B (see, e.g., [4]). The expected interval $EI(A)$ of a fuzzy number $A, A_\lambda = [A^-(\lambda), A^+(\lambda)]$, is defined by (see [3], [7])

$$EI(A) = \left[\int_0^1 A^-(\lambda)\, d\lambda, \int_0^1 A^+(\lambda)\, d\lambda\right].$$

An often used fuzzy number is the trapezoidal fuzzy number, completely characterized by four real numbers $t_1 \leq t_2 \leq t_3 \leq t_4$, denoted by $T = (t_1, t_2, t_3, t_4)$, with

$$T^-(\lambda) = t_1 + (t_2 - t_1)\lambda,$$
$$T^+(\lambda) = t_4 - (t_4 - t_3)\lambda, \lambda \in [0,1].$$

Sometimes (see [10]) a trapezoidal fuzzy numbers is denoted by $T = (l, u; x, y)$, with $l, u, x, y \in \mathbb{R}$ such that $x, y \geq 0, x + y \leq 2(u - l)$,

$$T^-(\lambda) = l + x\left(\lambda - \frac{1}{2}\right),$$
$$T^+(\lambda) = u - y\left(\lambda - \frac{1}{2}\right), \lambda \in [0,1].$$

It is immediate that

$$l = \frac{t_1 + t_2}{2}, \quad (2)$$
$$u = \frac{t_3 + t_4}{2}, \quad (3)$$
$$x = t_2 - t_1, \quad (4)$$
$$y = t_4 - t_3, \quad (5)$$

ISBN: 978-989-95079-6-8

the expected interval of a trapezoidal fuzzy number represented in this way is the real interval $[l,u]$ and the distance between $T=(l,u;x,y)$ and $T'=(l',u';x',y')$ becomes ([10])

$$D^2(T,T') = (l-l')^2 + (u-u')^2 + \frac{1}{12}(x-x')^2 + \frac{1}{12}(y-y')^2. \quad (6)$$

We denote by $F(\mathbb{R})$ the set of all fuzzy numbers, by $F^T(\mathbb{R})$ the set of all trapezoidal fuzzy numbers and by $T(A)$ the nearest (with respect to the metric D) trapezoidal fuzzy number to fuzzy number A preserving the expected interval of A. To express in a simplified form the approximation operator $T: F(\mathbb{R}) \to F^T(\mathbb{R})$ we denote

$$I(A) = I = \int_0^1 A^-(\lambda)d\lambda, \quad (7)$$

$$S(A) = S = \int_0^1 A^+(\lambda)d\lambda, \quad (8)$$

$$L(A) = L = \int_0^1 \lambda A^-(\lambda)d\lambda \quad (9)$$

and

$$U(A) = U = \int_0^1 \lambda A^+(\lambda)d\lambda, \quad (10)$$

for every fuzzy number $A, A_\lambda = [A^-(\lambda), A^+(\lambda)], \lambda \in [0,1]$ and we consider the following sets of fuzzy numbers

$$\Omega_1 = \{A \in F(\mathbb{R}) : 2I+S-3L-3U>0\},$$
$$\Omega_2 = \{A \in F(\mathbb{R}) : -I-2S+3L+3U>0\},$$
$$\Omega_3 = \{A \in F(\mathbb{R}) : -I+S+3L-3U\le 0\},$$
$$\Omega_4 = \Omega_1^c \cap \Omega_2^c \cap \Omega_3^c.$$

The set $\{\Omega_1,\Omega_2,\Omega_3,\Omega_4\}$ is a partition of $F(\mathbb{R})$ (see the proof of Theorem 7 in [1]).

The following result was proved in [1] and [10].

Theorem 1 (i) *If* $A \in \Omega_1$ *then*

$$T(A) = T_1(A) = (I, I, I, 2S-I);$$

(ii) *If* $A \in \Omega_2$ *then*

$$T(A) = T_2(A) = (2I-S, S, S, S);$$

(iii) *If* $A \in \Omega_3$ *then*

$$T(A) = T_3(A) = (4I-6L, -2I+6L, -2S+6U, 4S-6U);$$

(iv) *If* $A \in \Omega_4$ *then*

$$T(A) = T_4(A) = (3I+S-3L-3U, -I-S+3L+3U, -I-S+3L+3U, I+3S-3L-3U).$$

Four different operators $T_i, i \in \{1,2,3,4\}$ give us the nearest trapezoidal approximation preserving the expected interval: T_1 and T_2 lead to triangular fuzzy numbers with right side only or left side only, respectively, T_3 produces proper trapezoidal fuzzy numbers and T_4 produces triangular fuzzy numbers.

3 Additivity

The approximation operator $T: F(\mathbb{R}) \to F^T(\mathbb{R})$ given in Theorem 1 is scale invariant (see [1], Theorem 12, (ii)) and invariant to translations (see [1], Theorem 12, (i)), that is

$$T(\alpha A) = \alpha T(A)$$

and

$$T(A+z) = T(A) + z,$$

for every $A \in F(\mathbb{R}), z \in \mathbb{R}$ and $\alpha \in \mathbb{R} \setminus \{0\}$. In this section we point out a partial result of additivity and prove, be a simple example, that the operator T is not additive.

Example 2 *Let us consider the fuzzy numbers* A *and* B *given by*

$$A^-(\lambda) = 1+\sqrt{\lambda},$$
$$A^+(\lambda) = 30-27\sqrt{\lambda},$$
$$B^-(\lambda) = -1+\sqrt{\lambda},$$
$$B^+(\lambda) = 1-\sqrt{\lambda}, \lambda \in [0,1].$$

Then (see [1], Example 10 and Example 11) $A \in \Omega_1$,

$$T(A) = \left(\frac{5}{3}, \frac{5}{3}, \frac{5}{3}, \frac{67}{3}\right)$$

$B \in \Omega_4$,

$$T(B) = \left(-\frac{2}{3}, 0, 0, \frac{2}{3}\right),$$

therefore

$$T(A)+T(B) = \left(1, \frac{5}{3}, \frac{5}{3}, 23\right).$$

Because

$$(A+B)^-(\lambda) = 2\sqrt{\lambda},$$
$$(A+B)^+(\lambda) = 31-28\sqrt{\lambda}, \lambda \in [0,1],$$

we obtain

$$I(A+B) = \frac{4}{3},$$
$$S(A+B) = \frac{37}{3},$$
$$L(A+B) = \frac{4}{5}$$

and

$$U(A+B) = \frac{43}{10}.$$

We get $A+B \in \Omega_4$ *and*

$$T(A+B) = \left(\frac{31}{30}, \frac{49}{30}, \frac{49}{30}, \frac{691}{30}\right)$$

which implies

$$T(A+B) \neq T(A)+T(B).$$

Theorem 3 *If* $A, B \in \Omega_i$ *(*$i \in \{1,2,3,4\}$*) then*

$$T(A+B) = T(A)+T(B).$$

ISBN: 978-989-95079-6-8

Proof. The equalities

$$I(A+B) = I(A) + I(B),$$
$$S(A+B) = S(A) + S(B),$$
$$L(A+B) = L(A) + L(B),$$
$$U(A+B) = U(A) + U(B), A, B \in F(\mathbb{R})$$

are immediate. Then $A, B \in \Omega_i$ ($i \in \{1,2,3,4\}$) implies $A+B \in \Omega_i$ and $T(A)+T(B) = T(A+B)$ in every case $(i)-(iv)$ in Theorem 1. ■

4 Continuity

An extended trapezoidal fuzzy number ([10]) is an order pair of polynomial functions of degree less than or equal to 1. An extended trapezoidal fuzzy number may be not a fuzzy number, but the distance between two extended trapezoidal fuzzy numbers is similarly defined as in (1) or (6). The extended trapezoidal approximation $T_e(A) = (l_e, u_e; x_e, y_e)$ of a fuzzy number A is defined ([10]) as the extended trapezoidal fuzzy number which minimizes the distance $D(A,B)$, where B is an extended trapezoidal fuzzy number, and preserves the expected interval of A. The real numbers x_e and y_e are non-negative (see [11]). In [5], the authors proved that $T_e(A) = T_3(A)$.

Let us denote by $T(A) = (l_0, u_0; x_0, y_0)$ the nearest (with respect to the metric D) trapezoidal fuzzy number to fuzzy number A, preserving the expected interval, given in Theorem 1, in the form proposed in [10] (see also Section 2). We have ([10])

$$l_0 = l_e = \int_0^1 A^-(\lambda)d\lambda \tag{11}$$

$$u_0 = u_e = \int_0^1 A^+(\lambda)d\lambda. \tag{12}$$

Let us consider in the Euclidean space $\mathbb{R}^2$ (it is a finite dimensional Hilbert space) the points $A_e(x_e, y_e)$, $A_0(x_0, y_0)$ and the set

$$M = \{(x,y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + y \leq 2u_e - 2l_e\}$$

which is a closed convex subset of $\mathbb{R}^2$. Then $P_M(A_e)$ is the unique element in M (see e.g. [8], Theorem 4.10, p. 79) which minimizes the Euclidean distance in $\mathbb{R}^2$, $D_E(A_e, P)$, where $P \in M$. Taking into account the remarks in [10] we get

$$A_0 = P_M(A_e). \tag{13}$$

We present the main result of this section.

Theorem 4 *The nearest trapezoidal approximation operator preserving the expected interval $T : F(\mathbb{R}) \to F^T(\mathbb{R})$ is continuous.*

Proof. Let us consider two fuzzy numbers A and B,

$$A_\lambda = \left[A^-(\lambda), A^+(\lambda)\right],$$
$$B_\lambda = \left[B^-(\lambda), B^+(\lambda)\right], \lambda \in [0,1],$$

$T_e(A) = (l_e, u_e; x_e, y_e)$, $T_e(B) = (l'_e, u'_e; x'_e, y'_e)$ the extended trapezoidal approximations of A and B and $T(A) = (l_0, u_0; x_0, y_0)$, $T(B) = (l'_0, u'_0; x'_0, y'_0)$ the trapezoidal approximations preserving the expected interval of A and B. The relations (11) and (12) imply

$$D^2(T(A), T(B)) = (l_0 - l'_0)^2 + (u_0 - u'_0)^2 + \frac{1}{12}(x_0 - x'_0)^2 + \frac{1}{12}(y_0 - y'_0)^2$$
$$= (l_e - l'_e)^2 + (u_e - u'_e)^2 + \frac{1}{12}(x_0 - x'_0)^2 + \frac{1}{12}(y_0 - y'_0)^2$$

Because (see [9], Proposition 4.4)

$$D(T_e(A), T_e(B)) \leq D(A,B) \tag{14}$$

we obtain

$$D^2(T(A), T(B)) \leq D^2(A,B) + \frac{1}{12}(x_0 - x'_0)^2 + \frac{1}{12}(y_0 - y'_0)^2$$

or

$$D^2(T(A), T(B)) \leq D^2(A,B) + \frac{1}{12}D_E^2(A_0, B_0), \tag{15}$$

where $D_E(A_0, B_0)$ denotes the Euclidean distance between $A_0(x_0, y_0)$ and $B_0(x'_0, y'_0)$. Let us assume (contrariwise the proof is similar)

$$2u'_e - 2l'_e \geq 2u_e - 2l_e.$$

We consider

$$M_A = \{(x,y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + y \leq 2u_e - 2l_e\},$$
$$M_B = \{(x,y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + y \leq 2u'_e - 2l'_e\}$$

and

$$C(2u_e - 2l_e, 0),$$
$$C'(0, 2u_e - 2l_e),$$
$$G(2u'_e - 2l'_e, 0),$$
$$G'(0, 2u'_e - 2l'_e),$$

the points which define the closed convex sets M_A and M_B in the Euclidean space $\mathbb{R}^2$. The pairs (l_e, u_e), (l'_e, u'_e) being already fixed ((11), (12)), we denote $A_e(x_e, y_e)$ and $B_e(x'_e, y'_e)$ the points in $\mathbb{R}^2$ which represent the others components of the extended trapezoidal approximations of A and B, respectively. It is known that $x_e, y_e, x'_e, y'_e \geq 0$.

According to (13) we get

$$A_0 = P_{M_A}(A_e)$$
$$B_0 = P_{M_B}(B_e)$$

We denote by B_1 the projection of B_0 on the convex set M_A, that is the unique element in M_A which minimizes $D_E(B_0, Q)$, where $Q \in M_A$. We prove that B_1 is the

ISBN: 978-989-95079-6-8

projection of B_e on the set M_A, that is $B_1 \in M_A$ and $\min_{R \in M_A} D_E(B_e, R) = D_E(B_e, B_1)$, and

$$D_E(B_1, B_0) \leq D_E(C, G) = D_E(C', G').$$

Indeed, if $x'_e + y'_e \leq 2(u_e - l_e)$ then $B_e \in M_A$. Because $M_A \subseteq M_B$ we get

$$B_e \in M_B,$$
$$B_0 = B_1 = B_e$$

and

$$D_E(B_1, B_0) = 0 \leq D_E(C, G).$$

Contrariwise, if $x'_e + y'_e > 2(u_e - l_e)$ then the following situations are possible:

(i) $x'_e - y'_e > 2(u'_e - l'_e)$. Then

$$B_0 = G,$$
$$P_{M_A}(B_e) = C$$

and

$$B_1 = P_{M_A}(B_0) = P_{M_A}(G) = C.$$

In addition,

$$D_E(B_1, B_0) = D_E(C, G).$$

(ii) $2(u_e - l_e) \leq x'_e - y'_e \leq 2(u'_e - l'_e)$. Then

$$B_1 = P_{M_A}(B_e) = P_{M_A}(B_0) = C$$

and

$$D_E(B_1, B_0) = D_E(C, B_0) \leq D_E(C, G).$$

(iii) $-2(u_e - l_e) < x'_e - y'_e < 2(u_e - l_e)$. Because $B_e B_0$ is perpendicular on GG', $B_e B_1$ is perpendicular on CC' and GG', CC' are parallel we get

$$B_1 = P_{M_A}(B_e).$$

In addition,

$$D_E(B_1, B_0) < D_E(C, G).$$

(iv) $-2(u'_e - l'_e) \leq x'_e - y'_e \leq -2(u_e - l_e)$. Then

$$B_1 = P_{M_A}(B_e) = P_{M_A}(B_0) = C'$$

and

$$D_E(B_1, B_0) = D_E(C', B_0) \leq D_E(C', G').$$

(v) $x'_e - y'_e < -2(u'_e - l'_e)$. Then

$$B_0 = G',$$
$$P_{M_A}(B_e) = C'$$

and

$$B_1 = P_{M_A}(B_0) = P_{M_A}(G') = C'.$$

In addition,

$$D_E(B_1, B_0) = D_E(C', G').$$

We have

$$\begin{aligned} & D_E^2(C, G) \\ &= (2u'_e - 2l'_e - 2u_e + 2l_e)^2 \\ &= 4\left[(u'_e - u_e) - (l'_e - l_e)\right]^2 \\ &= 4\left[\int_0^1 \left(B^+(\lambda) - A^+(\lambda)\right) d\lambda \right. \\ &\quad \left. - \int_0^1 \left(B^-(\lambda) - A^-(\lambda)\right) d\lambda\right]^2 \\ &\leq 8\left[\int_0^1 \left(B^+(\lambda) - A^+(\lambda)\right) d\lambda\right]^2 \qquad (16) \\ &\quad + 8\left[\int_0^1 \left(B^-(\lambda) - A^-(\lambda)\right) d\lambda\right]^2 \\ &\leq 8\int_0^1 \left(B^+(\lambda) - A^+(\lambda)\right)^2 d\lambda \qquad (17) \\ &\quad + 8\int_0^1 \left(B^-(\lambda) - A^-(\lambda)\right)^2 d\lambda \end{aligned}$$

therefore

$$D_E^2(B_1, B_0) \leq 8D^2(A, B).$$

Because M_A is a closed convex subset of R^2 we obtain (see [11], Appendix C)

$$D_E(P_{M_A}(A_e), P_{M_A}(B_e)) \leq D_E(A_e, B_e)$$

that is

$$D_E(A_0, B_1) \leq D_E(A_e, B_e).$$

We get

$$\begin{aligned} D_E(A_0, B_0) &\leq D_E(A_0, B_1) + D_E(B_1, B_0) \\ &\leq D_E(A_e, B_e) + 2\sqrt{2}D(A, B). \end{aligned}$$

From (14) we obtain

$$D_E^2(A_e, B_e) \leq 12D^2(A, B), \qquad (18)$$

therefore

$$D_E(A_0, B_0) \leq 2(\sqrt{2} + \sqrt{3})D(A, B). \qquad (19)$$

The inequalities (15) and (19) imply

$$D(T(A), T(B)) \leq \sqrt{1 + \frac{(\sqrt{2} + \sqrt{3})^2}{3}} D(A, B) \qquad (20)$$

and the proof is complete. ■

If $D(A, B) = 0$ then (20) becomes equality.

If

$$D(T(A), T(B)) = \sqrt{1 + \frac{(\sqrt{2} + \sqrt{3})^2}{3}} D(A, B)$$

ISBN: 978-989-95079-6-8

then (16)-(18) must be equalities (contrariwise, inequality (20) becomes strict). Equality in (16) implies

$$\int_0^1 \left(B^+(\lambda) - A^+(\lambda)\right) d\lambda + \int_0^1 \left(B^-(\lambda) - A^-(\lambda)\right) d\lambda = 0 \quad (21)$$

and equality in (17) implies

$$B^+(\lambda) - A^+(\lambda) = k^+, \text{ a.e. } \lambda \in [0,1]$$

and

$$B^-(\lambda) - A^-(\lambda) = k^-, \text{ a.e. } \lambda \in [0,1],$$

where $k^+, k^- \in \mathbb{R}$. Substituting in (21) we get $k^+ + k^- = 0$, therefore

$$B^+(\lambda) = A^+(\lambda) + k, \text{ a.e. } \lambda \in [0,1]$$

and

$$B^-(\lambda) = A^-(\lambda) - k, \text{ a.e. } \lambda \in [0,1],$$

where $k \in \mathbb{R}$. According to (7)-(10) we obtain

$$I(B) = I(A) - k,$$
$$S(B) = S(A) + k,$$
$$L(B) = L(A) - \frac{k}{2},$$

and

$$U(B) = U(A) + \frac{k}{2}.$$

Taking into account Theorem 1 (iii) and (4), (5),

$$x_e^{'} = -6I(B) + 12L(B) = -6I(A) + 12L(A) = x_e,$$

$$y_e^{'} = 6S(B) - 12U(B) = 6S(A) - 12U(A) = y_e.$$

We get $A_e = B_e$ and equality in (18) implies $D(A,B) = 0$.

We conclude that equality in (20) holds if and only if $D(A,B) = 0$.

5 Conclusion

Many approximation methods for fuzzy numbers were proposed in the last years. Because the quality of approximation is important, a list of criteria that the approximation operator should or just can possess has been given in [5]. Continuity and additivity of nearest trapezoidal approximation operator which preserves the expected interval are studied in the present paper. Because the property of scale invariance was already proved in [1], the property of partial additivity, given in Theorem 3, assures the linearity of the trapezoidal approximation operator, when the fuzzy numbers under study are all with the same kind of asymmetry of membership functions (see [6]). The property of continuity, given in Theorem 4, means that if two fuzzy numbers are close, with respect to metric D, then their approximations are also close, with respect to metric D. Continuity is of extreme importance in applications, especially in fuzzy control systems, where it is sometimes indicated as robustness and discontinuous approximation operators seem unnatural.

References

[1] A. Ban. Approximation of fuzzy numbers preserving the expected interval. *Fuzzy Sets and Systems*, 159:1327-1344, 2008.

[2] A. Ban and L. Coroianu. Continuity of trapezoidal approximation operators, submitted.

[3] D. Dubois and H. Prade. The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24:279-300, 1987.

[4] P. Grzegorzewski. Metrics and orders of fuzzy numbers. *Fuzzy Sets and Systems*, 97:83-94, 1998.

[5] P. Grzegorzewski and E. Mrówka. Trapezoidal approximations of fuzzy numbers. *Fuzzy Sets and Systems*, 153:115-135, 2005.

[6] P. Grzegorzewski. Trapezoidal approximations of fuzzy numbers preserving the expected interval-Algorithms and properties. *Fuzzy Sets and Systems*, 159:1354-1364, 2008.

[7] S. Heilpern. The expected value of a fuzzy number. *Fuzzy Sets and Systems*, 47:81-86, 1992.

[8] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 1986.

[9] C.-T. Yeh. A note on trapezoidal approximations of fuzzy numbers. *Fuzzy Sets and Systems*, 158:747-754, 2007.

[10] C.-T. Yeh. Trapezoidal and triangular approximations preserving the expected interval. *Fuzzy Sets and Systems*, 159:1345-1353, 2008.

[11] C.-T. Yeh. On improving trapezoidal and triangular approximations of fuzzy numbers. *International Journal of Approximate Reasoning*, 48:297-313, 2008

ISBN: 978-989-95079-6-8

A baseline learning genetic fuzzy classifier based on low quality data

Ana M. Palacios[1] Luciano Sánchez[1] Inés Couso[2]

1. Departamento de Informática, Universidad de Oviedo, Gijón, Asturias, Spain
2. Departamento de Estadística e I.O. y D.M, Universidad de Oviedo, Gijón, Asturias Spain
Email: apalaciosjimenez@hotmail.com, luciano@uniovi.es, couso@uniovi.es

***Abstract*— Obtaining fuzzy rules from low quality data is a topic that has been recently formalized. This paper contains the first application of these principles to classification problems. We intend that the classifier proposed here serves as a baseline for future developments in the field. For that reason, we have extended a simple crisp genetic fuzzy classifier to imprecise data, paying special attention to the computational details. In particular, we will discuss some issues about the fuzzy-valued fitness function that is used in our formalism. A synthetic problem, plus two real-world datasets of low and medium complexities are also proposed, and used to evaluate the algorithm.**

1 Introduction

The term "low quality data" refers to datasets where some or all of the features are imprecisely perceived. It embodies different concepts: censored data, binning, missing values, significant digits, confidence intervals, contradictory information from different sources, fuzzy numbers, linguistic information, and others.

We are interested in those cases where the imprecision in the perception of the data is defined by a family of confidence intervals. This includes most of the preceding situations as particular cases. For instance, we can model few significant digits, censored or missing data by means of a single interval that spans the range of the unknown measure: "the weight of an object is in $[1.1, 1.2]$" or "the pressure is lower than 3". We can also reconcile different measurements of the same feature by means of a set of confidence intervals [8]. Certain (but not all) kinds of fuzzy data can also be regarded as imprecise perceptions of crisp values, known through a family of α-cuts [1].

To make clear the extent of this concept, let us recall the statistical framework of a standard Artificial Intelligence-based classifier. The purpose of a classifier is to predict the class of an object, given the values of other properties of the same object. To that end, a probability distribution is defined on the set of objects, and the mentioned properties are assumed to be random variables. The classifier, in turn, is a decision rule that depends on the posterior probability of each class, given a vector of properties.

The algorithms in this paper differ from the standard case because we do not suppose that we can accurately observe all the properties of the object. Otherwise, the same assumptions hold. In the most simple case (interval-valued data) we will perceive sets that contain these values. In the general case, we will be given a nested family of sets, each one of them containing the true value with certain probability. From a theoretical point of view, we hence understand that low quality data is a kind of data for which we can not achieve a precise knowledge about the posterior probability distributions induced by the mentioned random variables, but we can obtain families of probability distributions that are compatible with them. Accordingly, the mentioned decision rule can be inconclusive if the input data is not specific enough.

Recent works in fuzzy statistics suggest using a fuzzy representation when the data is known through a family of confidence intervals [1]. This representation assumes that a fuzzy set can be interpreted as a possibility distribution (which, in turn, is a family of probability distributions) and each α-cut of a fuzzy feature is a random set that contains the unknown crisp value of the feature with probability $1 - \alpha$. [7, 8].

1.1 GFS and Fuzzy fitness functions

Our particular fuzzy representation of low quality data is tied to the use of a fuzzy or interval-valued measurement of the accuracy of a classifier or model. Let us use an example: we have a classification system, defined by these rules:

$$\begin{array}{l}\text{if } x < 1 \text{ then class is } A\\ \text{if } x \in [1,2] \text{ then class is } B\\ \text{if } x > 2 \text{ then class is } C\end{array} \quad (1)$$

and the input that follows:

$$x < 1.8 \quad (2)$$

The output of the classifier is the set of classes $\{A, B\}$. If the object being classified is of class C, we know that the classifier has failed. Otherwise, we cannot know. Nonetheless, we can use a set-valued variable "number of errors", and state that the error of the classifier in that example is the set $\{0, 1\}$. The number of errors of the whole classifier can be obtained by adding these individual errors with interval arithmetic operators. If the output of the classifier is a fuzzy set, the number of errors is a fuzzy number too, and we must use fuzzy arithmetic [6].

Roughly speaking, estimating a classifier from data requires a numerical technique that finds the minimum of the classification error with respect to the free parameters of the classifying system. In our case, this function is interval-valued or fuzzy. But there are not many techniques for optimizing interval-valued or fuzzy valued functions. In the genetic algorithms field, the solutions are related to precedence operators between imprecise values [3, 4, 11]. We have previous works where we have jointly optimized a mix of crisp and fuzzy objectives with genetic algorithms [7]. We have also proposed a number of different algorithms for learning regression models from low quality data and the fuzzy representation mentioned before [5, 10, 8]. However, to the best of our knowledge there

ISBN: 978-989-95079-6-8

```
function GFS
1    Initialize population
2    for iter in {1, ..., Iterations}
3      for sub in {1, ..., subPop}
4        Select parents
5        Crossover and mutation
6        assignConsequent(offspring)
7      end for sub
8      Replace the worst subPop individuals
9      assignFitness(population,dataset)
10   end for iter
11   Purge unused rules
return population
```

Figure 1: Outline of the GFS that will be generalized [2]. Each chromosome codifies one rule. The fitness of the classifier is distributed among the rules at each generation.

```
function assignConsequent(rule)
1    for example in {1, ..., N}
2      m = membership(Antecedent,example)
3      weight[class[example]] = weight[class[example]] + m
4    end for example
5    mostFrequent = 0
6    for c in {1, ..., N_c}
7      if (weight[c]>weight[mostFrequent]) then
8        mostFrequent = c
9      end if
10   end for c
11   Consequent = mostFrequent
return rule
```

Figure 2: The consequent of a rule is not codified in the GA, but it is assigned the most frequent class label, between those compatible with the antecedent of the rule [2].

have not been previous GFSs where those principles have been applied to learn classification problems.

1.2 Summary

The structure of this paper is as follows: in the next section we generalize the crisp GFS defined in [2] to low quality data. In Section 3 we evaluate the generalized algorithm in both crisp and imprecise datasets, and propose two real-world datasets of classification with imprecise data. The paper finishes with the concluding remarks, in Section 4.

2 Generalizing a Genetic Fuzzy Classifier to imprecise data

Generalizing a GFS to imprecise data involves, at the very least, changes in the inference mechanism and the fitness function, as we have discussed in [6]. In this section we will generalize the GFS outlined in Figure 1, which was introduced in [2]. This is a very compact algorithm that allows us to focus in the subject of this paper (extending Genetic Fuzzy Classifiers to imprecise data) without getting lost in the details.

Observe that this algorithm depends on two functions: "assignConsequent" (line 6) and "assignFitness" (line 9). These functions are also listed in Figures 2 and 3. This algorithm does not codify the consequent of the fuzzy rules in the genetic individual. Instead, the function "assignConsequent"

```
function assignFitness(population,dataset)
1    for example in {1, ..., N}
2      winnerRule = 0
3      bestMatch = 0
4      for rule in {1, ..., M}
5        m = membership(Antecedent[rule],example)
6        if (m>bestMatch) then
7          winnerRule = rule
8          bestMatch = m
9        end if
10     end for rule
11     if (consequent(winnerRule)==class(example)) then
12       fitness[winnerRule] = fitness[winnerRule] + 1
13     end if
14   end for example
return fitness
```

Figure 3: The fitness of an individual is the number of examples that it classifies correctly. Single-winner inference is used, thus at most one rule changes its fitness when the rule base is evaluated in an example [2].

determines the class label that matches an antecedent with a maximum confidence. The function "assignFitness," in turn, determines the winner rule for each object in the training set and increments the fitness of the corresponding individual if its consequent matches the class of the object. In the remainder of this section, we will study the impact of the imprecise knowledge about the independent and dependent variables in the structure of the rules, and how to extend these two functions. That is to say: we analyze the reasoning method, the assignment of the consequents, the computation of a set-valued fitness and the genetic selection and replacement of the worst individuals.

2.1 Analysis of the reasoning method

The objective of the extended GFS is to obtain a fuzzy rule base from objects, when there is imprecise knowledge about some or all of the attributes of these objects. The set of classes that is produced when the input value is a fuzzy set can be computed in some different ways, but not all of them are consistent with our representation of an imprecise value.

Having an imprecise knowledge about the input variables differs from the standard case because:

1. The output of the FRBS will not be completely determined.

2. The number of errors of the FRBS in the training data will be partially known. The same happens if any other quality function is used instead of the number of errors, i.e. likelihood, logistic loss functions, etc.

Both issues have been introduced with the crisp classifier commented in the introduction: if the input is a set X that contains a range of inputs $x \in X$, the output is not a class but a set of classes:

$$\text{class(X)} = \{\text{class}(x) \mid x \in X\}. \tag{3}$$

The first difference is apparently a trivial issue: after all, one of the advantages of fuzzy rules is dealing with imprecision. Nevertheless, the standard reasoning method does not produce

ISBN: 978-989-95079-6-8

the set of classes that we need. To make our point clearer, let us study a fuzzy classifier comprising M rules:

$$\textbf{if } (x \textbf{ is } \widetilde{A}_i \,) \textbf{ then class is } C_i, \tag{4}$$

and let us use the single-winner inference mechanism:

$$\text{class(x)} = C_{\arg\max_i \{\widetilde{A}_i(x)\}}. \tag{5}$$

Observe that, classifying the set X by means of the standard mechanism produces the class label that follows:

$$\text{class'(X)} = C_{\arg\max_i \{\min\{\widetilde{A}_i(x) | x \in X\}\}} \tag{6}$$

while we need this set of labels:

$$\text{class(X)} = \{C_{\arg\max_i \{\widetilde{A}_i(x)\}} \mid x \in X\} \tag{7}$$

which is different than 6. Furthermore, the fuzzy set of classes obtained when aggregating the rules, before the defuzzification stage, is not what we need either, because our set class(X) does not coincide with the fuzzy set

$$\sum C / \max\{\min\{\widetilde{A}_i(x) \mid x \in X\} | C_i = C\}. \tag{8}$$

The code we propose to use is included in lines 2–23 in Figure 5, that we will explain later.

It is remarked that, in this paper, each rule will contain a single consequent. This is a result of our interpretation of "low quality" as a family of confidence intervals. In words: in this paper, if a point is labeled as "class {A,C}" we are not stating that it belongs to both categories at the same time (which is not an imprecise assert). We are expressing that we are not sure about the class of the object, i.e. we only know that it is not in class "B". Therefore, it makes not sense in this context to produce a rule with a double consequent, since this rule necessarily will have non-zero error at any example.

2.2 Assignment of consequents

```
function assignImpreciseConsequent(rule)
1    for example in {1,...,N}
2      m̃ = fuzMembership(Antecedent,example)
3      weight[class[example]] = weight[class[example]] ⊕ m̃
4    end for example
5    mostFrequent = {1,...,Nc}
6    for c in {1,...,Nc}
7      for c1 in {c+1,...,Nc}
8        if (weight[c] dominates weight[c1]) then
9          mostFrequent = mostFrequent - { c1}
10       end if
11     end for c1
12   end for c
13   Consequent = select(mostFrequent)
return rule
```

Figure 4: If the examples are imprecise, we might not know the most frequent class label –lines 5 to 12–. In this paper we have used the dominance proposed in [4] to reduce this set to one element.

The assignment of consequents seen in Figure 2 is extended in Figure 4. The original assigment consists in computing the confidences of the rules "if (x is $\widetilde{A}$) then class is C" for all the values of "C", then selecting the alternative with maximum confidence. In this case, the confidence of a rule is a set of values. The operation "dominates" used in line 8 can have different meanings, ranging from the strict dominance (A dominates B iff $a < b$ for all $a \in A, b \in B$) [11] to other definitions that induce a total order in the set of confidences. Generally speaking, we have to select one of the values in the set of nondominated confidences and use its corresponding consequent. In this paper, we have used the uniform dominance defined in [4], that induces a total order and thus the set of nondominated consequents has size 1.

2.3 Computation of fitness

```
function assignImpreciseFitnessApprox(population,dataset)
1    for example in {1,...,N}
2      setWinnerRule = ∅
3      for r in {1,...,M}
4        dominated = FALSE
5        rule.m̃ = fuzMembership(Antecedent[r],example)
6        for sRule in setWinnerRule
7          if (sRule dominates rule) then
8            dominated = TRUE
9          end if
10       end for sRule
11       if (not dominated) then
12         for sRule in setWinnerRule
13           if (m̃ dominates sRule) then
14             setWinnerRule = setWinnerRule −{ sRule }
15           end if
16         end for sRule
17         setWinnerRule = setWinnerRule ∪{ rule }
18       end if
19     end for r
20     setOfCons= ∅
21     for sRule in setWinnerRule
22       setOfCons= setOfCons ∪{ consequent(rule) }
23     end for sRule
24     deltaFit= 0
25     if ({class(example)} == setOfCons and
           size(setOfCons)==1) then
26       deltaFit = {1}
27     else
28       if ({class(example)}∩ setOfCons ≠ ∅) then
29         deltaFit = {0, 1}
30       end if
31     end if
32     Select winnerRule ∈ setWinnerRule
33     fitness[winnerRule] = fitness[winnerRule] ⊕ deltaFit
34   end for example
return fitness
```

Figure 5: Generalization of the function "assignFitness" to imprecise data. If the example is imprecisely perceived, there are three ambiguities that must be resolved: (a) some different crisp values compatible the same example might correspond to different winner rules –lines 3 to 19—, (b) these rules might have different consequent, thus we do not know if the rule base fails in the example –lines 20 to 31– and (c) we must assign credit to just one of these rules –lines 32 and 33–.

The output of the FRBS at the i-th object of the training set

ISBN: 978-989-95079-6-8

is a set of classes:

$$C_{\text{FRBS}}(X_i) = \{C_{\arg\max_j\{\widetilde{A}_j(x)\}} \mid x \in X_i\}. \quad (9)$$

The theoretical expression of the fitness function of the FRBS is:

$$\text{fitness} = \bigoplus e_i \quad (10)$$

where

$$e_i = \begin{cases} 1 & C_{\text{FRBS}}(X_i) = C_i \text{ and } \#(C_i) = 1 \\ 0 & C_{\text{FRBS}}(X_i) \cap C_i = \emptyset \\ \{0,1\} & \text{else} \end{cases} \quad (11)$$

In words, if the output of the FRBS is a single class label that matches the class label of the example, this point scores 1. If the set of classes emitted by the FRBS does not intersect with that of the object, this point scores 0. Otherwise, it scores the set $\{0,1\}$.

The evaluation of this function is computationally very expensive, and we will use an approximation, described in Figure 5. This algorithm computes an interval of values of matching between each rule and the input, then discards all rules that can not be the winner rule, and approximates the output of the FRBS by the set of the consequents of the non-discarded rules. This set includes the theoretical output, but sometimes it also includes extra class labels. In Figure 6 we have also included a more accurate approximation which is based on a sample of values of the support of the input. This second approximation will be used in the next section to better determine the quality of a classifier, but our learning will be guided by the function in Figure 5, because of its lower cost.

2.4 Genetic selection and replacement

There are two other parts in the original algorithm that must be altered in order to use an imprecise fitness function: (a) the selection of the individuals in [2] is based on a tournament, that depends on a total order on the set of fitness values. And (b) the same happens with the removal of the worst individuals. In both cases, we have used the uniform dominance defined in [4] to impose such a total order. We leave for future works the application of a multicriteria genetic algorithm similar to those used in our previous works in regression modeling [8, 10].

3 Numerical results

This section contains a brief numerical analysis of the generalized algorithm. We have performed the experiments that follow:

1. Synthetic datasets: Gaussian distribution, known Bayesian error, and different amounts of observation error.

2. Crisp datasets: three standard benchmarks, for testing that the extended algorithm has the same performance as the original version in crisp problems.

3. Imprecise, real world datasets: we propose two real world datasets, of small and medium size. One of them has been specifically designed for the purpose of this research, and the other is part of a practical problem of medical diagnosis.

```
function assignImpreciseFitnessExhaustive(population,dataset)
1     for example in {1,...,N}
2       S = sample(example)
3       maxScore = 0
4       for s in S
5         winnerRule = 0
6         bestMatch = 0
7         for rule in {1,...,M}
8           m = membership(Antecedent[rule],s)
9           if (m > bestMatch) then
10            winnerRule = rule
11            bestMatch = m
12          end if
13        end for rule
14        if (consequent(winnerRule) == class(example)) then
15          score[winnerRule] = score[winnerRule] ⊕ 1
16        elif (consequent(winnerRule) ⊂ class(example)) then
17          score[winnerRule] = score[winnerRule] ⊕ {0,1}
18        end if
19        if (max(score[winnerRule]) > max(score[maxScore]))
20          then maxScore = winnerRule
21        end if
22      end for s
23      if (score[maxScore] > 0) then
24        if (score[maxScore] == size(S)) then
25          fitness[maxScore] = fitness[maxScore] ⊕ 1
26        else
27          fitness[maxScore] = fitness[maxScore] ⊕ {0,1}
28        end if
29      end if
30    end for example
return fitness
```

Figure 6: Other generalization of the function "assignFitness" to interval-valued data. This function is computationally too expensive for being used as a fitness function; it will be used instead for obtaining better estimations of the train and test errors of the final rule bases. Lines 14–18 deal with the case where an object has imprecise output, i.e. "the class is A or C"; otherwise, the value of the variable "score" is crisp.

All the datasets used in this paper are available in the website of the KEEL project: `http://www.keel.es`. All the experiments have been run with a population size of 100, probabilities of crossover and mutation of 0.9 and 0.1, respectively, and limited to 200 generations. The fuzzy partitions of the labels are uniform and their size is 3, except when mentioned otherwise.

3.1 Synthetic datasets

The set "Gaussian" comprises 699 points of two classes. The distribution of both classes is bidimensional Gaussian, with unity covariance matrix, and centered in $(0,0)$ and $(3,0)$ respectively. To this data we have added interval-valued imprecision of sizes $\beta = 0.03, 0.05, 0.1, 0.2, 0.5$. A 10-cv experimental design was applied, and the mean values of the test errors are shown in Table 1. The training error has been also included, to show the differences between the approximation of the fitness function seen before and the exhaustive computation that has been used to compute the test error. Observe that the approximate error computed by the fitness function is less specific than the actual error, and the difference is relevant

ISBN: 978-989-95079-6-8

	Crisp			Low Quality		
β	Theoretical	Train	Test	Exh. Train	Exh. Test	Approx. Train
0	0.084	0.083	0.086	[0.086,0.086]	[0.082,0.082]	[0.086,0.086]
0.03				[0.047,0.086]	[0.083,0.094]	[0.076,0.091]
0.05				[0.075,0.089]	[0.081,0.098]	[0.071,0.094]
0.1				[0.070,0.103]	[0.068,0.104]	[0.076,0,093]
0.2				[0.052,0.116]	[0.055,0.128]	[0.075,0.089]
0.5				[0.022,0.183]	[0.022,0.179]	[0.014,0.225]

Table 1: Results of the extended GFS in the synthetic dataset "Gauss" for crisp data (first column) and different degrees of observation error (second column). The approximate error computed by the fitness function is less specific than the actual error, and the difference is relevant when the observation error is high ($\beta = 0.2$ and $\beta = 0.5$, nevertheless it still guides the evolution correctly.

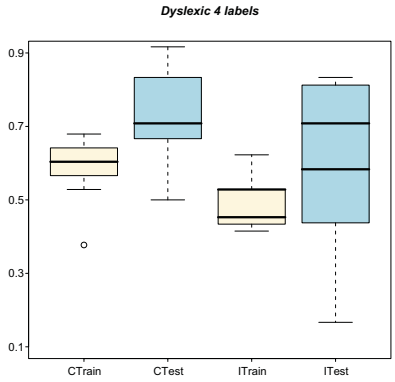


Figure 7: Boxplots illustrating the dispersion of the 10 repetitions of crisp and extended GFS in the problem "dyslexia-12", with 4 labels/partition. The boxplot of the imprecise experiments is not standard: we show respectively the 75% of the maximum and 25% percentile of the minimum fitness, thus the box displays at least the 50% of data; there are two marks inside the box, because the median of the data is an interval.

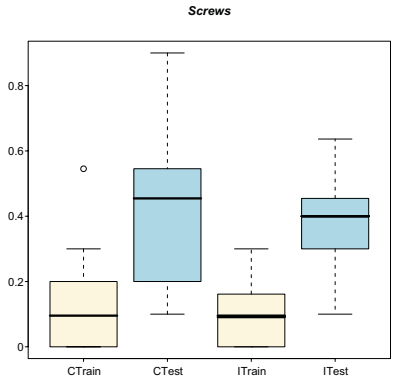


Figure 8: Boxplots illustrating the dispersion of the 10 repetitions of crisp and extended GFS in the problem "screws-50".

when the observation error is high ($\beta = 0.2$ and $\beta = 0.5$), nevertheless it still guides the evolution correctly.

3.2 Crisp datasets

Crisp datasets are included for assessing the performance of the generalized algorithm in standard problems. Since, in this case, the particularization of the algorithm to crisp data recovers the original algorithm in [2], the results are expected to be adequate. Nonetheless, we have included some estimations in Table 2. As expected, the results of both algorithms are similar; the differences are originated in the different random seeds.

3.3 Real world datasets

We propose two datasets for testing this and future learning algorithms with low quality data:

1. Dataset "Screws-50": 21 objects, 3 classes, 2 features (weight and length). We have weighed and measured 21 screws of three different types, taking into account the accuracy of the physical measurement: each feature is an interval. The class labels are precise. There are not outliers, i.e. 0% of error is attainable in absence of imprecision.

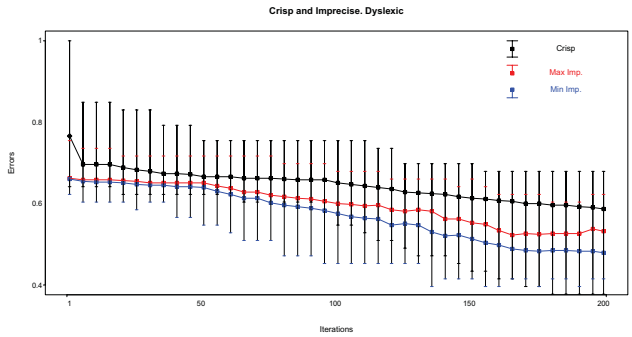


Figure 9: Compared evolution of crisp and imprecise GFS in the dataset "dyslexia-12". The ranges and means of 10 repetitions of the learning are shown, for both the crisp and the imprecise versions of the algorithm. The upper bound of the mean imprecise fitness is consistently lower than the mean of the crisp fitness.

2. Dataset "Dyslexia-12": 65 objects, 4 classes, 12 features. This is a selection of the original dataset described in [9], where the 12 most relevant variables have been hand-picked by a psychologist. There are imprecision in both the input and the output. The theoretical error is

ISBN: 978-989-95079-6-8

	Crisp		Low Quality		
Dataset	Train	Test	Exh. Train	Exh. Test	Approx. Train
Pima	0.254	0.287	[0.258,0.258]	[0.288,0.288]	[0.258,0.258]
Glass	0.323	0.365	[0.321,0.321]	[0.352,0.352]	[0.321,0.321]
Haberman	0.239	0.255	[0.238,0.238]	[0.248,0.248]	[0.238,0.238]

Table 2: Results in some crisp benchmarks, where the imprecise fitness function reduces to the crisp fitness function. The results of "crisp" and "low quality" columns are similar.

Crisp			Low Quality		
Dataset	Train	Test	Exh. Train	Exh. Test	Approx. Train
Screws-50	0.133	0.427	[0.096,0.096]	[0.377,0.377]	[0.068,0.106]
Dyslexia-12 (4 labels)	0.584	0.724	[0.481, 0.616]	[0.541, 0.675]	[0.477,0.516]
Dyslexia-12 (5 labels)	0.745	0.658	[0.581,0.667]	[0.608,0.641]	[0.537,0.549]

Table 3: Means of 10 repetitions of the generalized GFS for the imprecise datasets "Screws-50" and "Dyslexia-12" with 4 and 5 labels/variable

unknown.

We have compared the performance of the generalized algorithm to that of the original crisp algorithm. To that end, we have built a crisp dataset by removing the uncertainty in the imprecise dataset: each imprecise measurement was replaced by the mid-point of the corresponding interval, and those examples with imprecision in the independent variable were replicated for the different options. For instance, a point $(X = [1,3], C = \{A,B\})$ is converted into two points $(x = 2, c = A)$, $(x = 2, c = B)$.

We have used a 5x2cv design for the first problem, because of its small size, and 10cv for the second. The boxplots of the compared results, in both train and test sets, are depicted in Figures 7 and 8. Observe that the boxplots of the imprecise experiments are not standard. We propose using a box showing the 75% of the maximum and 25% percentile of the minimum fitness (thus the box displays at least the 50% of data) and also drawing two marks inside the box, because the median of the data is an interval. In Figure 9 the ranges and means of 10 repetitions of the learning are shown, for both the crisp and the imprecise versions of the algorithm. The upper bound of the mean imprecise fitness is consistently lower than the mean of the crisp fitness.

4 Concluding remarks

Extending a GFS to imprecise data in classification problems is based on the use of an interval or fuzzy valued fitness function. Most GFSs can be extended to low quality data if some changes are made in their reasoning method, and the genetic algorithm can deal with an imprecisely known fitness function. We have shown in detail how to apply this changes to a simple GCCL-type algorithm, and evaluated it with some synthetic and real-world benchmarks. The numerical results are as expected for an elementary algorithm like this; there is room for improvement and future works will address more complex GFSs that are based on a multicriteria fitness function.

Acknowledgements

This work was supported by the Spanish Ministry of Education and Science, under grants TIN2008-06681-C06-04, TIN2007-67418-C03-03, and by Principado de Asturias, PCTI 2006-2009.

References

[1] Couso, I., Sánchez, L. Higher order models for fuzzy random variables. Fuzzy Sets and Systems 159: 237–258 (2008)

[2] Ishibuchi, H., Nakashima, T., Murata, T, A fuzzy classifier system that generates fuzzy if-then rules for pattern classification problems. In Proc. of 2nd IEEE International Conference on Evolutionary Computation, 759-764 (1995)

[3] Koeppen, M., Franke, K., and Nickolay, B., Fuzzy-Pareto-Dominance driven multi-objective genetic algorithm. in Proc. 10th International Fuzzy Systems Assotiation World Congress (IFSA), Istanbul, Turkey, 2003: 450–453. (2003)

[4] Limbourg, P., Multi-objective optimization of problems with epistemic uncertainty. in EMO 2005: 413–427. (2005)

[5] Sánchez, L., Otero, J., Villar, J. R., Boosting of fuzzy models for high-dimensional imprecise datasets. Proc. IPMU 2006, Paris, France: 1965–1973. (2006)

[6] Sánchez L., Couso I. Advocating the use of imprecisely observed data in genetic fuzzy systems IEEE Transactions on Fuzzy Systems 15 (4): 551–562. (2007)

[7] Sánchez, L., Couso, I., Casillas, J. Modelling vague data with genetic fuzzy systems under a combination of crisp and imprecise criteria Proc. 2007 IEEE Symp. on Comp. Int. in Multicriteria Decision Making, Honolulu, USA: 30–37. (2007)

[8] Sánchez, L., Couso, I., Casillas, J. Genetic Learning of Fuzzy Rules based on Low Quality Data. Fuzzy Sets and Systems. Submitted.

[9] Sánchez, L., Palacios, A., Couso, I., A Minimum Risk Wrapper Algorithm for Genetically Selecting Imprecisely Observed Features, applied to the Early Diagnosis of Dyslexia. Lecture Notes in Computer Science 5271, 608–615 (2008)

[10] Sánchez, L., Otero, J., Couso, I., Obtaining linguistic fuzzy rule-based regression models from imprecise data with multiobjective genetic algorithms. Soft Computing. 13(5): 467–479 (2008)

[11] Teich, J., Pareto-front exploration with uncertain objectives. in EMO, 2001: 314–328. (2001)

ISBN: 978-989-95079-6-8

Pattern Recognition in Blur Motion Noisy Images using Fuzzy Methods for Response Integration in Ensemble Neural Networks

M. Lopez[1, 2] P. Melin[2] O. Castillo[2]

[1] PhD Student of Computer Science in the Universidad Autonoma de Baja California, Tijuana, B. C., México.
[2] Computer Science in the Graduate Division Tijuana, Institute of Technology Tijuana, B. C., Mexico,
Email: danym23@aol.com, pmelin@tectijuana.mx, ocastillo@tectijuana.mx

Abstract—*Linear Blur Motion is one of the most common degradation functions that corrupt images. Since 1976 many researchers have tried to estimate blur motion parameters and this problem can be solved for noise free images but in the case of noisy images this can be done when the image SNR is low. In this paper, we consider pattern recognition with ensemble neural networks for the case of fingerprints; we propose the use of fuzzy methods for Response Integration in Ensemble Neural Networks for blur motion noisy images. An ensemble neural network of three modules is used; each module is a local expert on person recognition based on a biometric measure (the fingerprints). The Response Integration method of the ensemble neural networks has the goal of combining the responses of the modules to improve the recognition rate of the individual modules when the SNR rate blur motion signal increases to a high level.*

Keywords— Pattern Recognition, Ensemble Neural Networks, Fuzzy Logic, Ratio SNR.

1 Introduction

Blur Motion occurs when there is relative motion between the camera and the object being captured [1]. When a changing scene is observed by a camera, all the classical algorithms assume that is possible to take pictures every δt instant, which means that every picture is taken with a dt ≈ 0 exposure time. If that is not the case, then the exposure time (dt = T) is large enough that different points in the scene are moving far enough and consequently their corresponding projections on the image plane travel several pixels. Therefore, during the capture process of an image, at any single image point, a certain number of scene points is projected during the exposure time, each one contributing to the final brightness of the image point; this effect is shown in figure 1. More formally, during the exposure time T in front of the pixel $P_{i,j}$ we could assume that they pass k scene points with brightness (C_1….C_k) respectively, then the resulting brightness value for pixel $P_{i,j}$ is given in equation (1), in the case it continues movement the summation is replaced by integration. This holds in general for every pixel that is moving points in the scene. It is clear that the blurring of the image exists only across the direction of the motion; this one dimensional blur is called Blur Motion.

$$P_{i,j} = \frac{1}{K}\sum_{l=1}^{k} C_l \qquad (1)$$

The Result of blur motion is shown in figure 1b where an image consistent of different value pixels is shown in figure 1a and then the blurred image is shown in figure 1b.

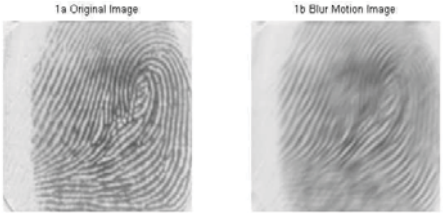


Figure 1: a) Original Image without noise, b) Blur Motion with 20 pixels distance d and 50 degrees angle α.

The blur motion can be described mathematically as the result of a linear filter b(x, y) = i(x, y)*h(x, y) where i is the theoreticalcal image taken with an exposure time T_e=0, b **is** the real blurred image and h the point spread function (PSF**).** Given an angle = α and the length d = V_o x T_e, which is the number of scene points that affect a specific pixel, the point spread function of blur motion is given in equation (2).

$$h(x,y) = \begin{cases} \frac{1}{d}, 0 \le |x| \le d * \cos(\alpha)\ \ y = \sin(\alpha) * d \\ 0, \text{otherwhise} \end{cases} \qquad (2)$$

In practical terms, this mean computing accurate estimates for the two parameters of the blur motion PSF, namely the length, d, and the angle, α. From these quantities, the relative velocity at this point can be easily recovered knowing the exposure time. In this paper for the experimental results we use fingerprint images and adding different levels of blur motion noise from 10 to 90 pixels in distance d and the angle α is zero. The measurement for noise in an image is the signal-to-noise ratio, SNR. The SNR measures the relative strength of the signal in a blurred and noisy image to the strength of the signal in a blurred image without noise. A SNR of 16 db is a low noise level, while a SNR of 4 db is a high noise level [1].

2 Noise and Registration Error

2.1 Signal-to-noise ratio

Noise is present in all digital imaging systems due to a number of sources such as photon shot noise, readout noise, dark current noise and quantization noise. While some sources can be effectively suppressed, such as the dark

ISBN: 978-989-95079-6-8

current noise by cooling and the quantization noise by using more bits, others not. Photon shot noise, for example, is unavoidable due to the particle nature of light. Readout noise increases with a higher readout rate, which is desirable in high-speed cameras. In equation 3 the calculus of the Variance and Standard deviation of the image are shown [2].

$$\sigma_I{}^2 = \frac{1}{2}\, var(I_1 - I_2), where\ I_1 = I + \eta_1, I_2 = I + \eta_2 \quad (3)$$

σ_I = the Standard deviation of the Fingerprint Image Original.
var = the variance between the Fingerprint Image Original and The Fingerprint Image with Blur Motion Noise.
$\eta_1 = 0$, is the noise to the Fingerprint Image Original.
η_2 = blur motion noise 10 to 90 distance pixels.
I_2 , is the Fingerprint Image with Blur Motion Noise.

The Signal-to-Noise Ratio (SNR) of an image I is then computed as:

$$SNR = 10\log\left(\frac{varI}{\sigma^2 I}\right) \quad (4)$$

The combined effect of these noise sources is often modeled from two images of the same scene captured under the same degradation is shown in figure 2.

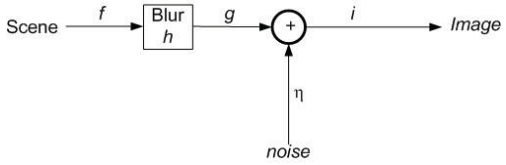


Figure 2: Image degradation by blur and additive noise.

2.2 Calculus of the SNR for Fingerprint Images

In this paper we use fingerprint images from the FCV2000 database of the Biometric System Laboratory at the University of Bologna [3, 4]. The image size is 300 pixels wide and 300 pixels high with a resolution of 500 ppi, and of a gray scale representation. The fingerprints were acquired by using a low-cost optical sensor; up to four fingers were collected from each volunteer (forefinger and middle finger of both the hands). The database is 10 fingers wide (w) and 8 impressions per finger deep (d) (80 fingerprints in all); the acquired fingerprints were manually analyzed to assure that the maximum rotation is approximately in the range [-15°, 15°] and that each pair of impressions of the same finger have a non-null overlapping area.
We added Blur Motion Level Noise from 10 to 90 distance pixels to all 80 fingerprint images, based on equations (3) and (4) the calculus of the Average SNR in decibels for the Fingerprint Image of each of the person obtained is shown in table 1.

Table I: The Average SNR in db of the Fingerprint Images.

Average SNR in db Person N° Fingerprint Image	Blur Motion Distance Pixels								
	10	20	30	40	50	60	70	80	90
P1	10.9	7.57	8.08	7.65	7.43	7.28	7.11	6.99	6.89
P2	8.36	8.02	7.52	7.38	7.16	7.01	6.89	6.78	6.68
P3	8.01	7.02	6,76	6.56	6.44	6.34	6.21	6.10	5.98
P4	7.67	5.75	5.57	5.46	5.26	5.17	5.07	5.01	4.93
P5	11.6	10.62	10.07	9.77	9.47	9.28	9.08	8.88	8.71
P6	7.29	7.28	6.88	6.75	6.58	6.46	6.33	6.20	6.09
P7	8.21	6.11	6.26	6.18	5.93	5.85	5.79	5.67	5.57
P8	10.6	10.07	9.4	9.08	8.8	8.56	8.47	8.15	7.97
P9	12.7	12	11.65	11.35	11.7	10.9	10.81	10.6	10.4
P10	8.94	6.43	6	5.79	5.52	5.33	5.24	5.14	5.03

In figure 3 we show the plot of the Average SNR of the Person Fingerprint with the blur motion level noise from 10 to 90 distance pixels.

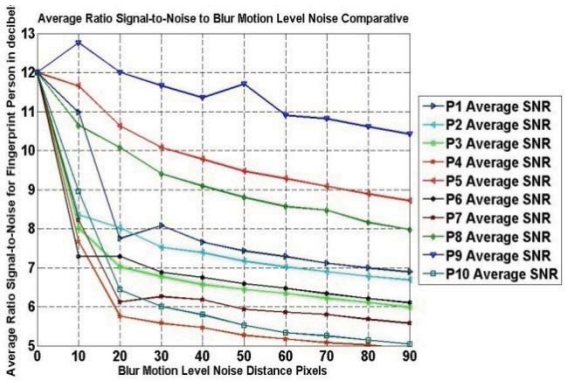


Figure 3: Average SNR with blur motion noise.

3 Ensemble Neural Networks Architecture

3.1 ANN´s Architecture

The Ensemble ANN´s architecture consists of three main modules [5, 6], in which each of them in turn consists of a set of neural networks trained with the same data (fingerprints image), If we now using different parameters for each module of the Ensemble Neural Networks, for module 1 the used parameters are learning rate =.001 , Goal Error=.001, for the module 2 learning rate =.0001 , Goal Error=.0001, for the module 3 the learning rate =.0001, and Goal Error=.00001, using 2 hidden layers, 36 neurons in the first layer and 17 neurons in the second layer for each module of the Ensemble Neural Network and is shown in figure 4.

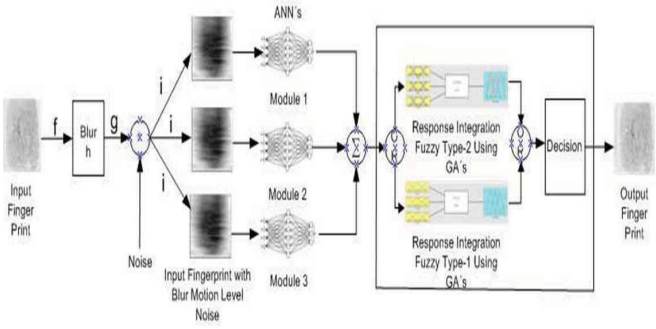


Figure 4: Ensemble Neural Networks with blur Motion Level Noise Added.

ISBN: 978-989-95079-6-8

The procedure to perform the tests was first to train the modules of the ensemble neural networks with the databases of the fingerprints without noise, in total using 8 fingerprints of 10 people, that is 80 fingerprints in total, until being able to find which architecture of the ensemble neural network responded better to arrive to the desired error.

We Added Blur Motion Level Noise from 10 to 90 distance pixels to the Fingerprint Input, to obtain a blur motion image that went into each of three modules ensemble neural networks. The output of each of the ensemble neural networks was obtained response integration with type-1 or type-2 fuzzy logic, and finally with the decision module obtained the output fingerprint.

3.2 Response Integration with Type-1 Fuzzy Logic

We also used a Type-1 Fuzzy Inference System as method of response integration of the ensemble neural network output, considering as input three linguistic variables, Activation Low, Activation Medium, and Activation High, and one output linguistic variable, Winning Activation of the three modules.

We show below one of the rules for the fuzzy inference system.

If (Module1 is ActMod1Low) and (Module2 is ActMod2Medium) and (Module3 is ActMod3Medium) then (Winner Module is Module3)

Three membership functions were used for each linguistic variable (input and output) of the triangular type, to be managed in a range from 0 to 1.

We show in figure 5 the membership functions designed using the editor of the fuzzy logic toolbox of MATLAB [14].

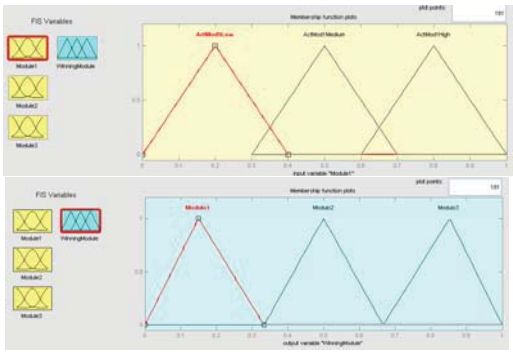


Figure 5: The Input Membership function and The Output Membership function of Fuzzy Logic Type-1.

Once the Ensemble Neural Network is trained, the fuzzy inference system integrates the outputs of the modules. We used the same 80 people's images to which we had applied noise with blur motion, and the type-1 fuzzy inference system gives an answer for the stage of the final decision.

3.3 Response Integration with Type-2 Fuzzy Logic

We used a Type-2 Fuzzy Inference System as method of response integration of the ensemble neural network output, considering as input three linguistic variables, Activation Low, Activation Medium, and Activation High, and one output linguistic variable, Winning Activation of the three modules, used the same rules of the type-1 fuzzy logic.

Three membership functions were used for each linguistic variable of the input and output of the Gaussian type managed in a range from 0 to 1.

We show in figure 6 the membership functions designed using the editor of IT2FUZZY fuzzy logic toolbox [12, 13].

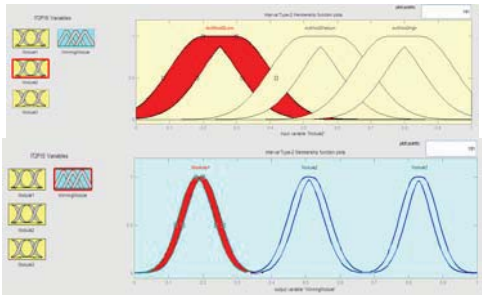


Figure 6: The Input Membership function and The Output Membership function of Type-2 Fuzzy Logic.

4 Simulation Results

4.1 Response Integration with Type-1 Fuzzy Logic

Once the Ensemble Neural Network is trained, the fuzzy inference system integrates the outputs of the modules. We used the same 80 persons images to which we had applied different levels of noise with blur motion, the type-1 inference systems give an output for the stage of the final decision, and we show the result if the fingerprint input was recognized. We show in figure 7 the simulation results.

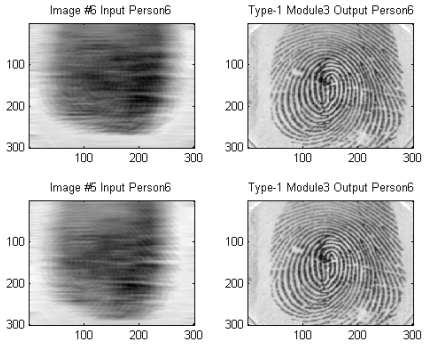


Figure 7: Simulation results for the fingerprints using the type-1 Fuzzy Inference System (blur motion 50 distance pixels).

4.2 Response Integration with Type-2 Fuzzy Logic

In the same way when the Ensemble Neural Network is trained, the fuzzy inference system integrates the outputs of the modules. We used the same 80 persons images to which we had applied different levels of noise with blur motion, the type-2 inference systems give an output for the stage of the final decision, and show the result if the fingerprint input was recognized. We show in figure 8 the simulation results.

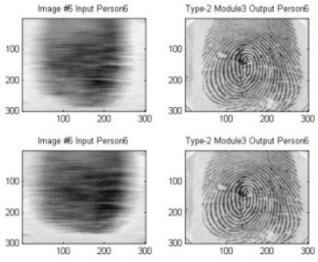


Figure 8: Simulation results for the fingerprints using the type-2 Fuzzy Inference System (blur motion 50 distance pixels).

ISBN: 978-989-95079-6-8

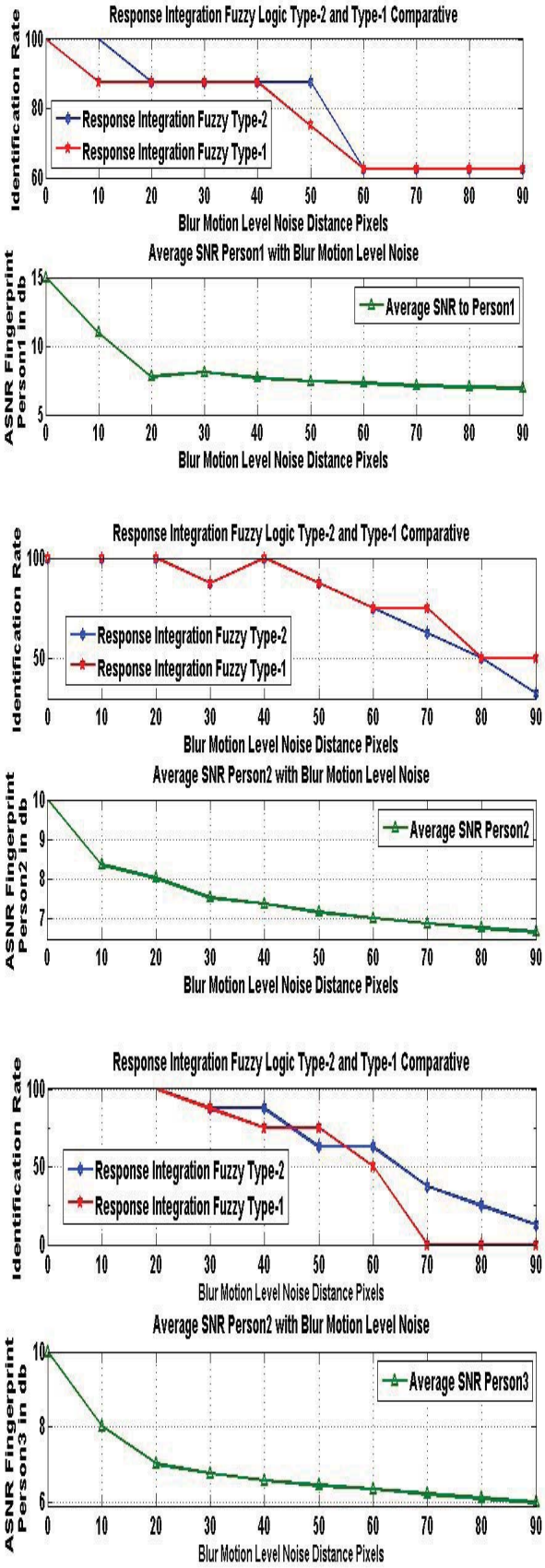
4.3 Comparison Pattern Recognition Blur Motion Noisy Images using Fuzzy Methods for Response Integration in Ensemble Neural Networks

The Response Integration method of the ensemble neural networks has the goal of combining the responses of the modules to improve the recognition rate of the individual modules when the blur motion signal increases to high level 10 to 90 distance pixels. In Table 2 we show a Comparison between Response Integration with Type-2 and Type-1 Fuzzy logic when the blur motion level noise increases and the Average SNR in db decrease.

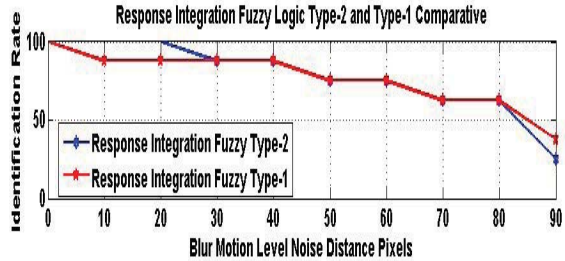
Table II: Comparison Response Integration Type-2 and Type-1 Fuzzy Logic with Blur Motion Level Noise and the Average SNR in db of the Fingerprint Images.

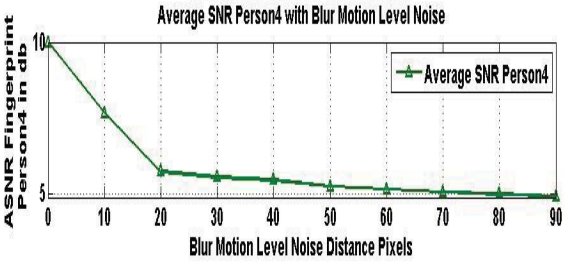
# Person Fingerprint	Blur Motion Level Noise								
	10	20	30	40	50	60	70	80	90
Average SNR P1 db	10.97	7.75	8.08	7.65	7.43	7.28	7.11	6.99	6.89
Fuzzy Type-2 %	100	87.5	87.5	87.5	87.5	62,5	62.5	62.5	62.5
Fuzzy Type-1 %	87.5	87.5	87.5	87.5	75	62.5	62.5	62.5	62.5
Average SNR P2 db	8.33	8.02	7.52	7.38	7.16	7.01	6.89	6.78	6.68
Fuzzy Type-2 %	100	100	87.5	100	87.5	75	62.5	50	37.5
Fuzzy Type-1 %	100	100	87.5	100	87.5	75	75	50	50
Average SNR P3 db	8.01	7.02	6.76	6.56	6.44	6.34	6.21	6.10	5,98
Fuzzy Type-2 %	100	100	87.5	87.5	62.5	62.5	37.5	25	12.5
Fuzzy Type-1 %	100	100	87.5	75	75	50	0	0	0
Average SNR P4 db	7.67	5.75	5.57	5.46	5.26	5.17	5.075	5.01	4.94
Fuzzy Type-2 %	100	100	87.5	87.5	75	75	62.5	62.5	25
Fuzzy Type-1 %	87.5	87.5	87.5	87.5	75	75	62.5	62.5	37.5
Average SNR P5 db	11.65	10.62	10.08	9.77	9.47	9.28	9.08	8.89	8.71
Fuzzy Type-2 %	100	100	100	100	100	87.5	87.5	75	50
Fuzzy Type-1 %	100	100	100	100	100	87.5	75	75	50
Average SNR P6 db	7.29	7.28	6.88	6.75	6.58	6.46	6.33	6.21	6.1
Fuzzy Type-2 %	100	100	100	100	100	100	100	87.5	75
Fuzzy Type-1 %	100	100	100	100	100	100	100	87.5	75
Average SNR P7 db	8.21	6.11	6.26	6.18	5.93	5.85	5.79	5.67	5.57
Fuzzy Type-2 %	100	100	100	87.5	87.5	87.5	62.5	50	37.5
Fuzzy Type-1 %	100	100	100	75	75	87.5	62.5	50	37.5
Average SNR P8 db	10.64	10.07	9.40	9.09	8.8	8.56	8.47	8.15	7.97
Fuzzy Type-2 %	100	100	100	87.5	87.5	87.5	62.5	50	37.5
Fuzzy Type-1 %	100	100	100	75	75	87.5	62.5	50	37.5
Average SNR P9 db	12.76	12	11.65	11.35	11.7	10.9	10.81	10.6	10.42
Fuzzy Type-2 %	100	100	100	100	100	100	100	100	100
Fuzzy Type-1 %	100	100	100	100	100	100	100	100	100
Average SNR P10 db	8.21	6.11	6.26	6.18	5.93	5.85	5.79	5.67	5.57
Fuzzy Type-2 %	100	100	100	100	100	100	100	100	100
Fuzzy Type-1 %	100	100	100	100	100	100	100	100	100

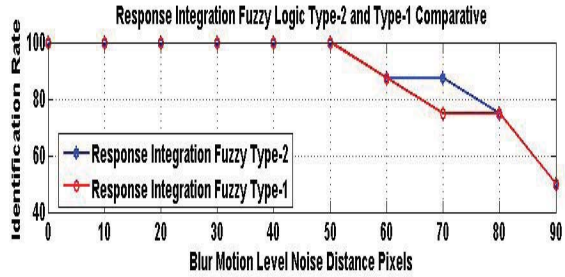
In Figure 9 we show in the up side the response integration of the type-2 and type-1 fuzzy system when the blur motion noise increases and the down side the Average SNR in db decrease for each of the Person Fingerprint Images.

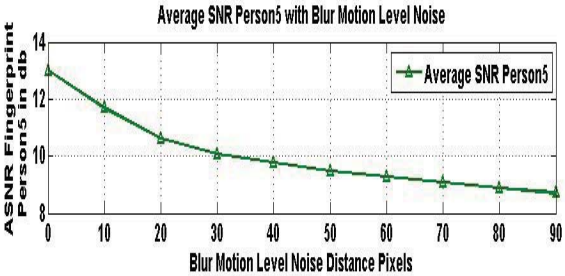


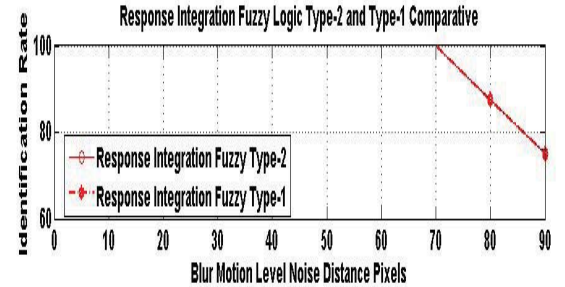
ISBN: 978-989-95079-6-8

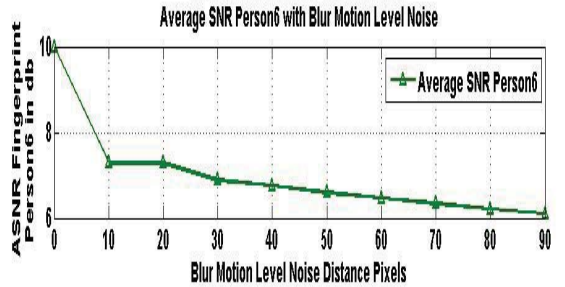


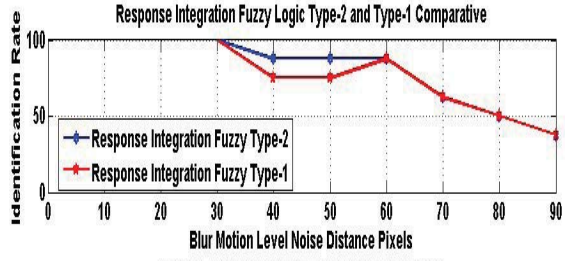


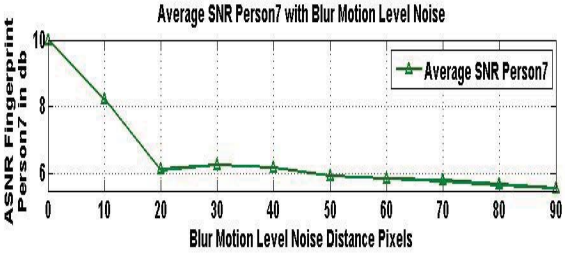


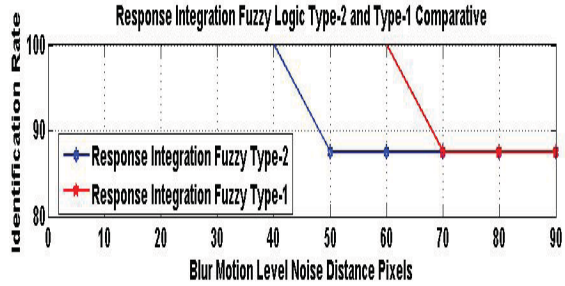


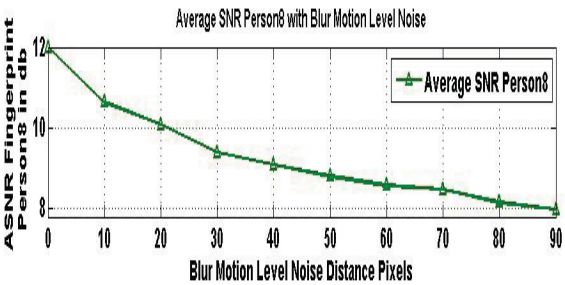


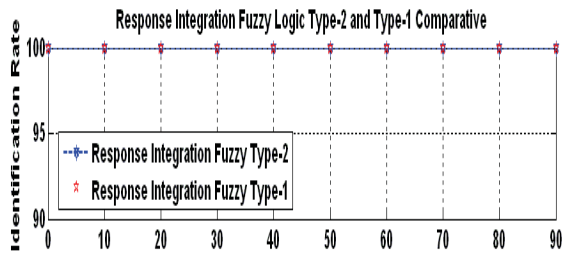


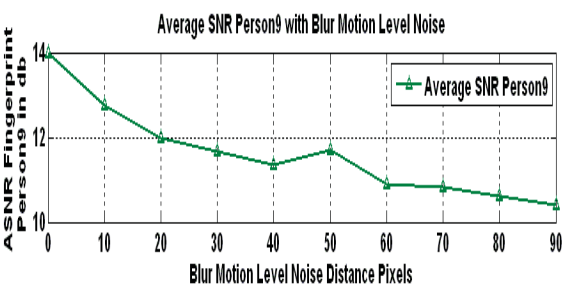












ISBN: 978-989-95079-6-8

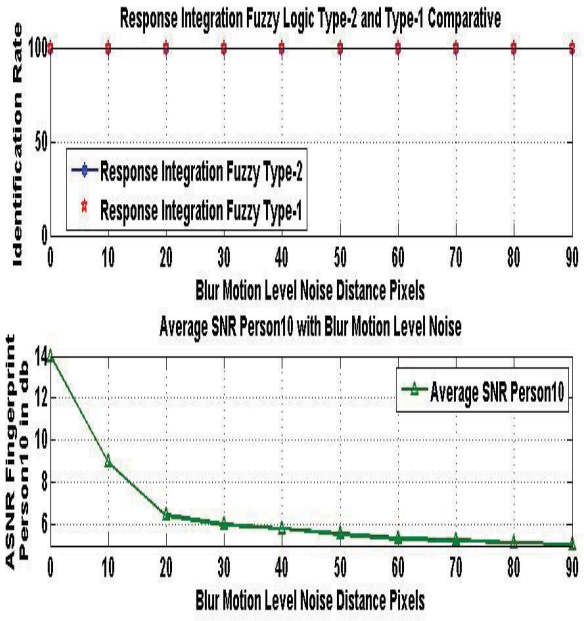


Figure 9: Response Integration Type-2 and Type-1 Fuzzy logic with the blur Motion Level Noise 10 to 90 Distance Pixels and Average SNR in db of each 10 Person Fingerprint Image.

5 Conclusions

Based on the experimental results, we can conclude that the behavior of the type-2 and type-1 fuzzy systems, as methods of response integration of ensemble neural networks for the fingerprints when added blur motion level noise increase and the Average SNR of the Fingerprint decrease, we obtain excellent results.

The identification rates obtained are between 37.5% to 100%for response integration type-2 fuzzy logic and 0% to 100% for response integration with type-1 fuzzy logic. The difference between the type-1 fuzzy and type-2 fuzzy inference systems for response integration of ensemble neural networks is appreciated when we uses the blur motion level noise. We think that there is an advantage in using a type-2 fuzzy inference system to manage the uncertainty of the knowledge base in pattern recognition problems. Future work will include, testing with more kinds of noise, other methods of response integration, using feature extraction, and image compression, with the goal of improving the identification rate.

Acknowledgment

We would like to express our gratitude to CONACYT, Universidad Autonoma de Baja California and Tijuana Institute of Technology for the facilities and resources granted for the development of this research project.

References

[1] Felix Krahmer_, Youzuo Lin, Bonnie McAdoo, Katharine Ott, Jiakou Wang, David Widemannk, Brendt Wohlberg, *Blind Image Deconvolution: Motion Blur Estimation, August 2006.*

[2] Tuan Q. Pham, Lucas J. van Vliet, Klamer Schutte, *Influence of signal-to-noise ratio and point spread function on limits of super resolution*, *Image Processing, Algorithms and Systems IV, Vol. 5672, pp. 169-180, 2005.*

[3] D. Maltoni, D. Maio, A.K. Jain, S. Prabhakar, *The full FVC2000 and FVC2002 databases are available in the DVD included in: Handbook of Fingerprint Recognition, Springer, New York, 2003.*

[4] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman and A. K. Jain, *"FVC2004: Third Fingerprint Verification Competition", Proc. International Conference on Biometric Authentication (ICBA), pp. 1-7, Hong Kong, July 2004.*

[5] M. Lopez, P. Melin: *Response integration in Ensemble Neural Networks using interval type-2 Fuzzy logic. IJCNN 2008: 1503-1508*

[6] O. Mendoza, P. Melin, G. Licea, *Modular Neural Networks and Type-2 Fuzzy Logic for Face Recognition, In Marek Reformat, editor, Proceedings of NAFIPS 2007, Number 1, pages CD Rom, San Diego, June 2007.*

[7] J. Urias, D. Solano, .M. Soto, M. Lopez, P. Melin, *¨Type-2 Fuzzy Logic as a Method of Response Integration in Modular Neural Networks¨. IC-AI 2006: 584-590.*

[8] P. Melin, F. González, G. Martínez: *Pattern Recognition Using Modular Neural Networks and Genetic Algorithms. IC-AI 2004: 77-83.*

[9] P. Melin, J. Urias, D. Solano, M. Soto, M. Lopez, O. Castillo: *Voice Recognition with Neural Networks, Type-2 Fuzzy Logic and Genetic Algorithms. Engineering Letters 13(2): 108-116 (2006).*

[10] P. Melin, A. Mancilla, M. Lopez, D. Solano, M. Soto, O. Castillo, *¨Pattern Recognition for Industrial Security using the Fuzzy Sugeno Integral and Modular Neural Networks¨, WSC11 11th Online World Conference on Soft Computing in Industrial Applications September 18-October 6, 2006.*

[11] P. Melin and O. Castillo, 2005*, ¨Hybrid Intelligent Systems for Pattern Recognition using Soft Computing¨, ISBN 3-540-24121-3 Springer Berlin Heidelberg New York.*

[12] J. R. Castro, O. Castillo, P. Melin, L.G. Martinez, S. Escobar, I. Camacho*, Building Fuzzy Inference Systems with Interval Type-2 Fuzzy Logic Toolbox Number 1 in Studies in Fuzziness and Soft Computing, 6, pages 53-62, Springer-Verlag, Germany, 1 edition, June 2007.*

[13] J. R. Castro, O. Castillo, P. Melin, *An Interval Type-2 Fuzzy Logic Toolbox for Control Applications; In Proc. FUZZ-IEEE 2007.*

[14] MATLAB Trade Marks, ©1994-2007 by The Math Works, Inc.

[15] S. Farsiu, D. Robinson, M. Elad and P. Milanfar. *Fast and robust multi-frame super-resolution. IEEE Trans. on Image Processing, 13(10):1327-1344, 2004.*

[16] R.C. Hardie, K.J. Barnard and E.E. Armstrong. *Joint map registration and high-resolution image estimation using a sequence of undersampled images. IEEE Trans.on Image Processing, 6(12):1621–1633, 1997.*

[17] R. Neelamani, H. Choi and R. Baraniuk. Forward: *Fourier-wavelet regularized deconvolution for Ill-Conditioned Systems. In IEEE Trans. Signal Processing, 52(2):418–433, 2004.*

[18] J. Biemond. *Iterative methods for image deblurring. Proceedings of the IEEE, 78(5):856–883*, 1990.

[19] M.Ebrahimi Moghadam, M.Jamzad, *Linear motion blur parameter estimation in noisy images using fuzzy sets and power spectrum, EUROSIP Journal on applied signal processing, Article number 68985, Volume 2007, pages 862–866..*

[20] Y. Yitzhaky, I. Mor, A. Lantzman, and N. S. Kopeika. *Direct method for restoration of motion blurred images. Journal of the Optical Society of America. A, Optics and image science, 15(6):1512–1519, 1998.*

Fuzzy Logic Adaptation of a Hybrid Evolutionary Method for Pattern Recognition

Fevrier Valdez[1] Patricia Melin[2] Oscar Castillo[2]

[1]PhD Student of Computer Science in the Universidad Autónoma de Baja California
Tijuana, B.C.
[2]Computer Science in the Graduate Division, Tijuana Institute of Technology
Tijuana, B.C.
Email: fvaldez0125@yahoo.com.mx, pmelin@tectijuana.mx, ocastillo@tectijuana.mx

Abstract— *We describe in this paper a new hybrid approach for optimization combining Particle Swarm Optimization (PSO) and Genetic Algorithms (GAs) using Fuzzy Logic to integrate the results. The new evolutionary method combines the advantages of PSO and GA to give us an improved FPSO+FGA hybrid method. Fuzzy Logic is used to combine the results of the PSO and GA in the best way possible. The new hybrid FPSO+FGA approach is compared with the PSO and GA methods with a set of benchmark mathematical functions. The proposed hybrid method is also tested with the problem of modular neural network optimization. The new hybrid FPSO+FGA method is shown to be superior with respect to both the individual evolutionary methods.*

Keywords— Fuzzy Logic, Evolutionary Computing, Genetic Algorithms**.**

1 Introduction

We describe in this paper a new evolutionary method combining PSO and GA, to give us an improved FPSO+FGA hybrid method. We apply the hybrid method to mathematical function optimization to validate the new approach. Also in this paper, the application of a Genetic Algorithm (GA) [1] and Particle Swarm Optimization (PSO) [2] for the optimization of mathematical functions is considered. In this case, we are using the Rastrigin's function, Rosenbrock's function, Ackley's function, Sphere's function Griewank's function, Michalewics's function and Zakharov's function [4][13] to compare the optimization results between a GA, PSO and FPSO+FGA. We also consider the problem of neural network architecture optimization, which is very important in the applications.

The paper is organized as follows: in section 2 a description about the Genetic Algorithms for optimization problems is given, in section 3 the Particle Swarm Optimization is presented, the neural networks is presented in section 4, in section 5 we can appreciate the proposed method FPSO+FGA and the fuzzy system, in section 6 we can appreciate the full model of FPSO+FGA that was used for this research, in section 7 the simulations results are described, finally we can see in section 8 the conclusions reached after the study of the proposed evolutionary computing methods.

2 Genetic Algorithms for Optimization

John Holland, from the University of Michigan initiated his work on genetic algorithms at the beginning of the 1960s. His first achievement was the publication of *Adaptation in Natural and Artificial System* [7] in 1975.

He had two goals in mind: to improve the understanding of natural adaptation process, and to design artificial systems having properties similar to natural systems [8].

The basic idea is as follows: the genetic pool of a given population potentially contains the solution, or a better solution, to a given adaptive problem. This solution is not "active" because the genetic combination on which it relies is split between several subjects. Only the association of different genomes can lead to the solution.

Holland's method is especially effective because it not only considers the role of mutation, but it also uses genetic recombination, (crossover) [9]. The crossover of partial solutions greatly improves the capability of the algorithm to approach, and eventually find, the optimal solution.

The essence of the GA in both theoretical and practical domains has been well demonstrated [1]. The concept of applying a GA to solve engineering problems is feasible and sound. However, despite the distinct advantages of a GA for solving complicated, constrained and multi-objective functions where other techniques may have failed, the full power of the GA in application is yet to be exploited [12] [14].

In Fig. 1 we show the reproduction cycle of the Genetic Algorithm.

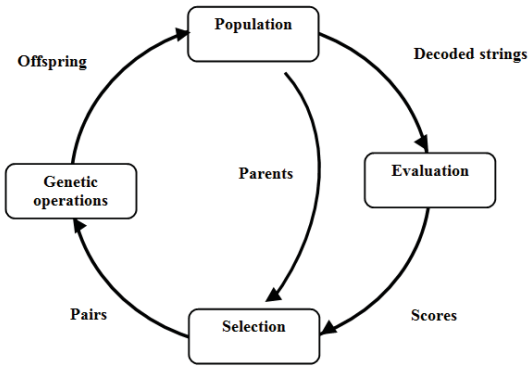


Figure 1: The Reproduction cycle

ISBN: 978-989-95079-6-8

3 Particle Swarm Optimization

Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Eberhart and Kennedy in 1995, inspired by social behavior of bird flocking or fish schooling [3].

PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA) [6]. The system is initialized with a population of random solutions and searches for optima by updating generations.

However, unlike GA, the PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles [10].

Each particle keeps track of its coordinates in the problem space, which are associated with the best solution (fitness) it has achieved so far (The fitness value is also stored). This value is called *pbest*. Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the neighbors of the particle. This location is called *lbest*. When a particle takes all the population as its topological neighbors, the best value is a global best and is called *gbest* [19].

The particle swarm optimization concept consists of, at each time step, changing the velocity of (accelerating) each particle toward its p*best* and *lbest* locations (local version of PSO). Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward *pbest* and *lbest* locations [17].

In the past several years, PSO has been successfully applied in many research and application areas. It is demonstrated that PSO gets better results in a faster, cheaper way compared with other methods [11] [18].

Another reason that PSO is attractive is that there are few parameters to adjust. One version, with slight variations, works well in a wide variety of applications. Particle swarm optimization has been used for approaches that can be used across a wide range of applications, as well as for specific applications focused on a specific requirement [20].

4 Neural Networks

The discipline of neural networks originates from an understanding of the human brain. The average human brain consists of 3 × 1010 neurons of various types, with each neuron connecting to up to 104 synapses [21]. Many neural-network models, also called connectionist models, have been proposed [6].

Neural networks are attractive since they consist of many neurons, with each one processing information separately and simultaneously. All the neurons are connected by synapses with variable weights. Thus, neural networks are actually parallel distributed processing systems.

Research on neural networks dates back to the 1940s when McCulloch and Pitts found that the neuron can be modeled as a simple threshold device to perform logic function [22].

In this paper, we used a supervised learning for training the neural network. Supervised learning is based on a direct comparison between the actual network output and the desired output. Some recent advances in supervised learning have been reviewed in [5].

5 FPSO+FGA Method

This method combines the characteristics of PSO and GA using several fuzzy systems for integration of results and parameter adaptation. In this section, the proposed FPSO+FGA method is presented.

The general idea of the proposed FPSO+FGA method can be seen in Fig. 2. The method can be described as follows:

1. A problem optimization to be considered is received in this case; we are testing with a modular neural network and mathematical functions.
2. The method is randomly initialized to start with FPSO or FGA.
3. The system will solve the problem by FPSO or FGA depending of Error (E) and Derivate of Error (DE) generated after applying FPSO or FGA.
4. The function evaluated is called FEVAL, then the system records the Errors (E and DE) obtained to compare with the value desired.
5. After that the function is evaluated, E and DE are taken as inputs to one fuzzy system called 'FSDM', which the main function is to decide if is suitable to continue solving the problem with FPSO or FGA depending of the information generated by the fuzzy rules to calculate when can be necessary do a change
6. Another two fuzzy systems receive the values of Error and DError as inputs to evaluate if it is necessary to change the values of the parameters in FPSO or FGA depending of the fuzzy rules. The fuzzy systems responsible to do change the value of parameters in FPSO and FGA are called 'fuzzyga' and 'fuzzypso'. On Fig. 2 the two fuzzy systems are represented by el block FSPA (Fuzzy System Parameter Adaptation) but while the method is running in the FPSO the fuzzy system changes the values of the cognitive acceleration 'c_1', and social acceleration 'c_2'; and while the method is running in the FGA the fuzzy system change the value of crossover and mutation.
7. Repeat the above steps until the termination criterion of the algorithm is met. Finally BVALUE receive the best value obtained and when the method is finished we are obtaining the best result.

The fundamental characteristic of this method is to adapt the parameters to reach the best result, in this case; we are doing changes on the parameters of crossover (cr), mutation (mu), cognitive acceleration (c_1) and social acceleration (c_2). We are taking this parameters because are very important to the convergence with good results of the method, for example 'c_1' and 'c_2' are working as a memory able of record the knowledge of the particles more important in a population in the FPSO; also 'cr' and 'mu' are important to reach the best solution in the FGA. The values of these four parameters are fuzzy because depending of the output of a fuzzy system able of adaptation parameters.

ISBN: 978-989-95079-6-8

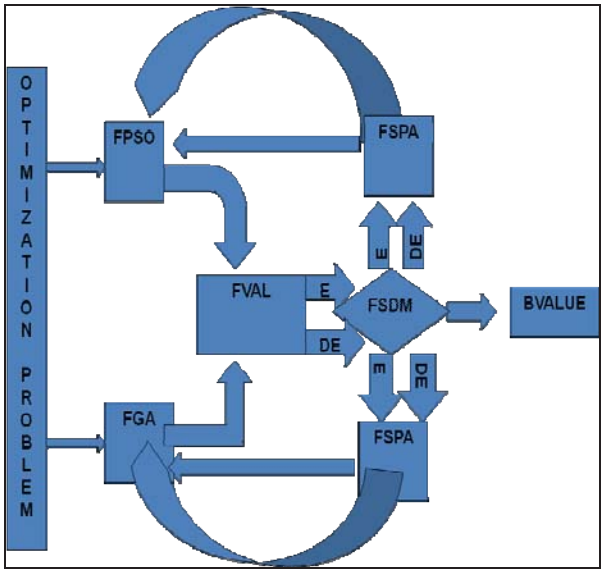


Figure 2: The FPSO+FGA scheme

6 Full Model of FPSO+FGA

The basic idea of the FPSO+FGA scheme is to combine the advantages of the individual methods using a fuzzy system for decision making and the others two fuzzy systems to improve the parameters of the FGA and FPSO when is necessary.

As can be seen in the proposed hybrid FPSO+FGA method, it is the internal fuzzy system structure, which has the primary function of receiving as inputs (Error and DError) the results of the outputs FGA and FPSO. The fuzzy system is responsible for integrating and deciding which are the best results being generated at run time of the FPSO+FGA. It is also responsible for selecting and sending the problem to the "fuzzypso" fuzzy system when the FPSO is activated or to the "fuzzyga" fuzzy system when FGA is activated. Also activating or temporarily stopping depending on the results being generated. The fuzzy system is of Mamdani type because it is more common in this type of fuzzy control and the defuzzification method is the centroid. In this case, we are using this type of defuzzification because in other papers we have achieved good results [4]. Also, the membership functions in the fuzzy main system were chosen of triangular form based on past experiences in this type of fuzzy control. However, in the simulation results we can see different architectures to the 'fuzzyga' and the 'fuzzypso'; where, we are changing the membership functions to observe which is more reliable to this application. In this research, we are testing with triangular and gaussian membership functions.

6.1 FPSO (Fuzzy Particle Swarm Optimization)

This section presents a detailed description of the FPSO model. The classical representation scheme for GAs is binary vectors of fixed length. In the case of an n_x – dimensional search space, each individual consists of n_x variables with each variable encoded as a binary string. The swarm is typically modeled by particles in multidimensional space that have a position and a velocity. These particles fly through hyperspace (i.e., R^n) and have two essential reasoning capabilities: their memory of their own best position and knowledge of the global or their neighborhood's best. In Fig. 3 we can see a simulation of Ackley's function. In a minimization optimization problem, "best" simply means the position with the smallest objective value. Members of a swarm communicate good positions to each other and adjust their own position and velocity based on these good positions. So a particle has the following information to make a suitable change in its position and velocity:

- A global best that is known to all and immediately updated when a new best position is found by any particle in the swarm.
- The neighborhood best that the particle obtains by communicating with a subset of the swarm.
- The local best, which is the best solution that the particle has seen.

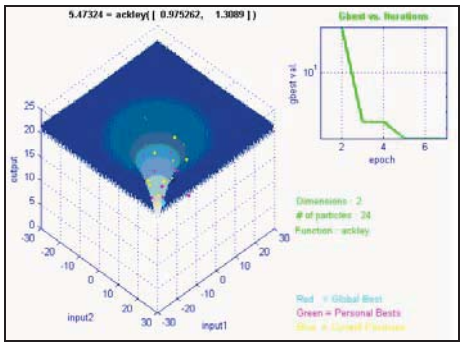


Figure 3: Simulation of ackley's function with FPSO

In this case, the social information is the best position found by the swarm, referred as $\hat{y}(t)$. For gbest FPSO, the velocity of particle i is calculated as

$$v_{ij}(t+1) = vi_j(t) + c_1 r_{1j}(t)[y_{ij}(t) - x_{ij}(t)] + c_2 r_{2j}(t)[\hat{y}_j(t) - x_{ij}(t)] \quad (1)$$

Where $vi_j(t)$ is the velocity of particle i in dimension j = 1,…,n_x at time step t, $xi_j(t)$ is the position of particle i in dimension j at time step t, 'c_1' and 'c_2' represents the cognitive and social acceleration. In this case, these values are fuzzy because they are changing dynamically when the FPSO is running, and $r_{1j}(t)$, r_{2j}~U(0,1) are random values in the range [0,1]. We are adding a fuzzy system called 'fuzzypso' that is able of change the 'c_1' and 'c_2' rate.

6.2 FGA (Fuzzy Genetic Algorithm)

This section presents a brief description of the FGA model. Several crossover operators have been developed for GAs, depending on the format in which individuals are represented. For binary representations, uniform crossover, one point crossover and two point crossover are the most popular. In this case we are using two points crossover with fuzzy crossover rate because we are adding a fuzzy system called 'fuzzyga' that is able of change the crossover and mutation rate.

6.3 Definition of the Fuzzy Systems used in FPSO+FGA

'fuzzypso': In this case we are using a fuzzy system called 'fuzzypso', and the structure of this fuzzy system is as follows:

Number of Inputs: 2
Number of Outputs: 2
Number of membership functions: 3

ISBN: 978-989-95079-6-8

Type of the membership functions: Triangular and Gaussian
Number of rules: 9
Defuzzification: Centroid
The main function of the fuzzy system called 'fuzzypso' is to adjust the parameters of the PSO. In this case, we are adjusting the following parameters: 'c_1' and 'c_2'; where:
'c_1' = Cognitive Acceleration
'c_2' = Social Acceleration
We are changing these parameters to test the proposed method. In this case, with 'fuzzypso' is possible to adjust in real time the 2 parameters that belong to the PSO.
'fuzzyga': In this case we are using a fuzzy system called **'fuzzyga'**, the structure of this fuzzy system is as follows:
Number of Inputs: 2
Number of Outputs: 2
Number of membership functions: 3
Type of membership functions: Triangular and Gaussian
Number of rules: 9
Defuzzification: Centroid
The main function of the fuzzy system called 'fuzzypso' is to adjust the parameters of the GA. In this case, we are adjusting the following parameters: 'mu', 'cr'; where:
'mu' = mutation
'cr' = crossover
'fuzzymain': In this case, we are using a fuzzy system called **'fuzzymain'**. The structure of this fuzzy system is as follows:
Number of Inputs: 2
Number of Outputs: 1
Number of membership functions: 3
Type of membership functions: Triangular
Number of rules: 9
Defuzzification: Centroid
The main function of the fuzzy system, called 'fuzzymain' is to decide on the best way for solving the problem, in other words if it is more reliable to use the FPSO or FGA. This fuzzy system is able to receive two inputs, called error and derror, it is to evaluate the results that are generated by FPSO and FGA in the last step of the algorithm. Fig. 4 shows the membership functions of the main fuzzy system that is implemented in this method. All fuzzy system consists of 9 rules. For example, one rule is if error is P and DError is P then best value is P (view Fig. 5). Fig. 6 shows the fuzzy system rule viewer. Fig. 7 shows the surface corresponding to the fuzzy main system. The other two fuzzy systems are similar to the main fuzzy system.

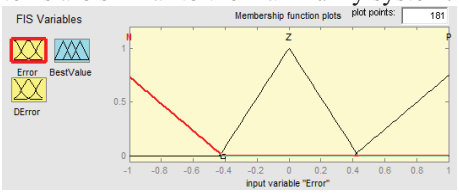


Figure 4: Fuzzy system membership functions

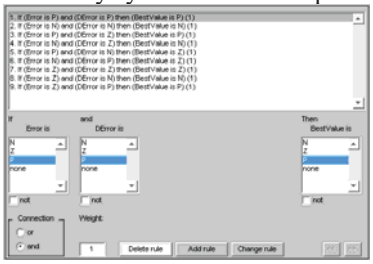


Figure 5: Fuzzy system rules

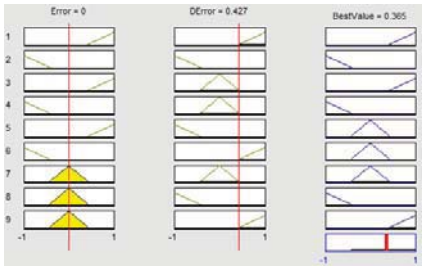


Figure 6: Fuzzy system rules viewer

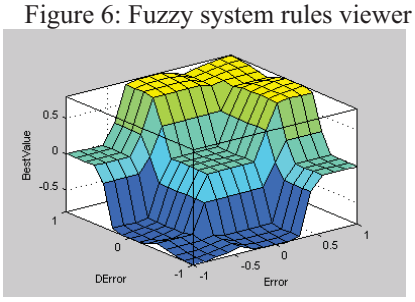


Figure 7: Surface of fuzzy system

7. Simulation Results for Modular Neural Network for Optimization

Several tests of the FPSO+FGA method for MNN optimization were made in the Matlab programming language.

All the implementations were developed using a computer with processor Intel core 2 quad of 64 bits that works to a frequency of clock of 2.5 GHz, 4 GB of RAM Memory and Windows Vista operating system.

We describe below simulation results of our approach for face recognition with modular neural networks (MNNs). We used two-layer feed-forward MNNs with the Conjugated-Gradient training algorithm [4]. The challenge is to find the optimal architecture of this type of neural network, which means finding out the optimal number of layers and nodes of the neural network [6]. We are using the Yale face database [20] that contains 165 grayscale images in GIF format of 15 individuals, for this paper only 10 subjects were used for training the MNN. There are 5 images per subject, one per different facial expression: center-light, happy, left-light, normal and right-light. In total 50 images were used (view Fig. 8). Three images per subject were used for training the MNN and the other two for the recognition. Regarding the genetic algorithm for NN evolution, we used a hierarchical chromosome for representing the relevant information of the network. First, we have the bits for representing the number of layers of the MNN; in this case, the initial topology was of 3 modules with 2 layers per module with 500 neurons in the first layer, 300 neurons in the second layer in each module. Therefore we used a representation the 2415 bits in total (view Fig. 9). The PSO is organized in a similar fashion, but there is less number of parameters. In Fig. 10 we can see the architecture of a MNN that we are using with the evolutionary proposed method FPSO+FGA.

The fitness function used in this case for the MNN combines the information of the error objective and also the information about the number of nodes as a second objective. This is shown in the following equation.

$$f(z) = \left(\frac{1}{\alpha * Ranking(ObjV1) + \beta * ObjV2} \right) * 10 \qquad (1)$$

ISBN: 978-989-95079-6-8

The first objective is basically the average sum of squared of errors as calculated by the predicted outputs of the MNN compared with real values of the function. This is given by the following equation.

$$f_1 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - y_i)^2 \qquad (2)$$

The second objective is the complexity of the neural network, which is measured by the total number of nodes in the architecture.

The final topology of the neural network for the problem of face recognition is obtained by the hybrid evolutionary method FPSO+FGA. The comparison of the final objective values (errors) will be shown in the following section. In the final architecture, the result of the MNN evolution is a particular architecture with different number of nodes by layers. Several tests were made; we obtained optimized different architectures for this Modular Neural Network; the best architecture obtained was the following:

Layers = 2 x module

NNL1M1 = 90, NNL2M1 = 50. NNL1M2 = 100, NNL2M2 = 150. NNL1M3 = 70, NNL2M3 = 90. Total bits = 565

Where: NNL1M1= Number of neurons of the layer 1 in module 1. NNL1M1= Number of neurons of the layer 1 in module 1. NNL2M1= Number of neurons of the layer 2 in module 1. NNL1M2= Number of neurons of the layer 1 in module 2. NNL2M2= Number of neurons of the layer 2 in module 2. NNL1M3= Number of neurons of the layer 1 in module 3. NNL2M3= Number of neurons of the layer 2 in module 3. We can see in the Fig. 11 the binary representation for this optimized architecture. With this final topology the Neural Network was trained and the ten images were recognized. It can be seen in Table 1 the different architectures obtained with this method using a 'fuzzyga' and a 'fuzzypso' with triangular membership functions and a 'fuzzymain' with triangular membership functions, the other parameters to the 3 fuzzy systems are above described. Table 2, shows the simulation results with all fuzzy systems with gaussian membership functions. The proposed method optimizes the initial architecture proposed for the problem of face recognition.

Figure 8: Images of the Yale face database

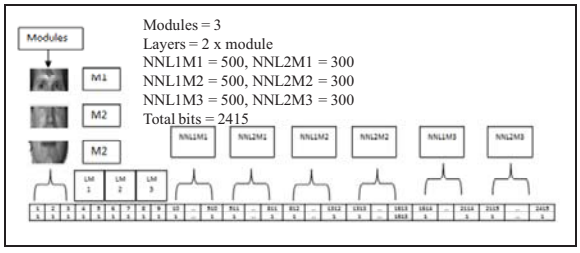


Figure 9: Binary representation for FPSO+FGA (No optimized)

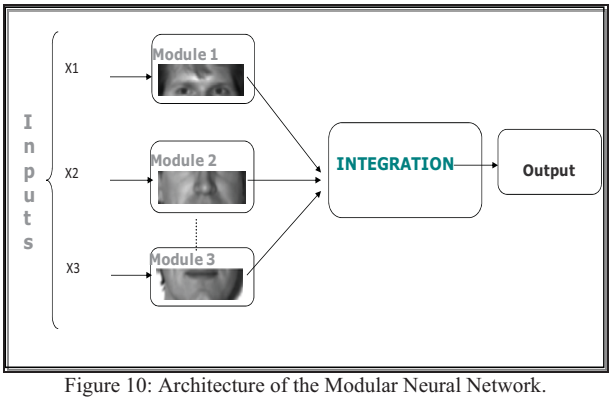


Figure 10: Architecture of the Modular Neural Network.

The Parameters of Table 1 and 2, are as follows:

LMod = Layers per module

NNL = Number of neurons Layer 1, 2 or 3.

GE = Goal Error.

RE = Reached Error.

IDENT = Number of images recognized.

VAR = Number of variables for the mathematical function.

It can be seen in Table 1 and Table 2 that this method is good alternative to solve this type of problems, in our case to optimize Modular Neural Networks. Also this FPSO+FGA have been applied, for optimization of complex mathematical functions to validate our approach. The parameters in FPSO+FGA; as crossover, mutation, cognitive acceleration and social are fuzzy, the population size was of 100 individuals, Table 3 shows the simulation results with 7 mathematical functions to testing this hybrid evolutionary method. The mathematical functions are evaluated with 2, 4, 8 and 16 variables.

Table 1. Simulation Results for the MNN with triangular membership functions

LMod	NN L1 M1	NN L2 M1	NN L1 M2	NN L2 M2	NN L1 M3	NN L2 M3	GE	RE	IDENT
2	20	60	80	50	60	120	0.01	0.03	8
2	90	50	100	150	70	90	0.01	0.005	10
2	70	40	80	40	90	30	0.01	0.02	8
2	150	135	200	90	84	40	0.01	0.003	9
2	100	120	100	145	100	70	0.01	0.001	10

Table 2. Simulation Results for the MNN with gaussian membership functions

LMod	NN L1 M1	NN L2 M1	NN L1 M2	NN L2 M2	NN L1 M3	NN L2 M3	GE	RE	IDENT
2	30	50	90	70	50	115	0.01	0.05	8
2	85	60	103	140	80	130	0.01	0.02	9
2	90	50	95	50	85	40	0.01	0.08	8
2	130	120	180	90	70	50	0.01	0.02	9
2	50	90	75	70	40	40	0.01	0.01	9

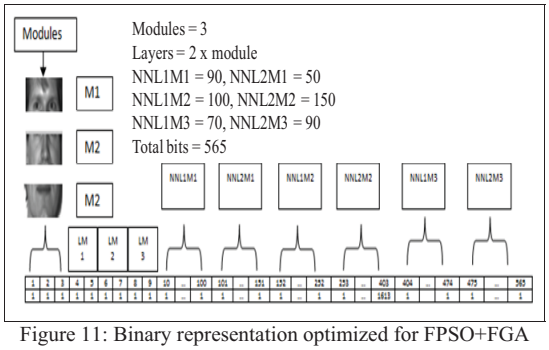


Figure 11: Binary representation optimized for FPSO+FGA

ISBN: 978-989-95079-6-8

Table 3. Simulation results for Mathematical Functions

Math F	NVar = 2		NVar = 4		NVar = 8		NVAR = 16	
Math	BEST	MEAN	BEST	MEAN	BEST	MEAN	BEST	MEAN
Ras	1.45E-06	3.05E-04	0.0003	0.0755	0.01318	0.9311	0.5483	5.0946
Ros	1.17E-02	1.17E-02	0.0285	0.5991	0.15800	3.8925	0.2555	4.33334
Ack	8.42E-04	4.98E-03	8.42e-01	4.98E-02	0.7	1.56	2.35	2.63
Sph	5.75E-11	1.05E-10	1.946e-05	4.5109e-004	0.00059	0.0057	0.0024 8	0.0211
Gri	7.88E-11	1.07E-07	7.18e-06	1.1182e-004	0.00016	9.299e-004	0.00040	0.0043
Mich	-1.8010	-1.8201	-1.80129	-1.8002	-1.80130	-1.8005	-1.801301	-1.8004
Zak	6.00E-07	0.00168	3.237e-07	8.4129e-005	1.3308e-07	6.4901e-005	8.63410e-07	7.0065e-005

In Table 4, It can be seen a comparison results between the 3 methods analyzed in this research, we can observe FPSO+FGA as a good alternative to resolve optimization problems. Also, the Neural Network (NN) was included to test this approach. Table 4 shows the simulation results for two variables.

Table 4. Comparison results between GA, PSO and FPSO+FGA

Math Funct	GA	PSO	FPSO+FGA	Objective Value
Ras	2.15E-03	5.47E-05	3.05E-04	0
Ros	1.02E-05	1.97E-03	1.17E-02	0
Ack	2.98	2.98	4.98E-03	0
Sph	1.62E-04	8.26E-11	1.05E-10	0
Grie	2.552E-05	2.56E-02	1.07E-07	0
Mich	-1.7829	-7.44E-01	-1.8201	-1.8013
Zakh	0.00146674	8.10	0.00168	0
NN	3.33E-01	2.3E-01	1.01E-03	0

8 Conclusions

The analysis of the simulation results of the evolutionary method considered in this paper, FPSO+FGA lead us to the conclusion that for the optimization of Modular Neural Networks with this method is a good alternative because it is easier to optimize the architecture of Modular Neural Network than to try it with PSO or GA separately. This is, because the combination PSO and GA with fuzzy rules gives a new hybrid method FPSO+FGA. It can be seen in Table 1 that the second and five architectures obtained after applying FPSO+FGA recognizes the ten images; we are demonstrating that it is reliable uses for this type of applications. Recently we are working with more images for test the effectiveness of this approach. In Table 4 it can be seen a comparison with PSO and GA used separately, we can observed as FPSO+FGA was better.

Acknowledgment

We would like to express our gratitude to CONACYT, Universidad Autónoma de Baja California and Tijuana Institute of Technology for the facilities and resources granted for the development of this research.

References

[1] Man K.F., Tang K.S., and Kwong S., "Genetic Algorithms: Concepts and Designs", Springer Verlag. 1999.

[2] Eberhart, R. C. and Kennedy, J. A new optimizer using particle swarm theory. Proceedings of the Sixth International Symposium on Micromachine and Human Science, Nagoya, Japan. pp. 39-43, 1995.J.-G. Lu, "Title of paper with only the first word capitalized," *J. Name Stand. Abbrev.*, in press.

[3] Kennedy, J. and Eberhart, R. C. Particle swarm optimization. Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ. pp. 1942-1948, 1995.

[4] Holland J.H., Adaptation in natural and artificial system, Ann Arbor, The University of Michigan Press, 1975.

[5] Valdez, F. and Melin P. "Parallel Evolutionary Computing using a cluster for Mathematical Function Optimization", Nafips. San Diego CA USA, 598-602. June 2007.

[6] Castillo, O.; Melin, P. "Hybrid intelligent systems for time series prediction using neural networks, fuzzy logic, and fractal theory" Neural Networks, IEEE Transactions on Volume 13, Issue 6, Page(s): 1395 – 1408. Nov. 2002.

[7] D.B. Fogel, "An introduction to simulated evolutionary optimization", IEEE transactions on neural networks, vol 5, n 1, Page(s):3 – 14, jan 1994.

[8] Goldberg D., Genetic Algorithms, Addison Wesley, 1988.

[9] Emmeche C., Garden in the Machine. The Emerging Science of Artificial Life, Princeton University Press, pp. 114. 1994.

[10] Angeline P. J., Using Selection to Improve Particle Swarm Optimization , In Proceedings 1998 IEEE World Congress on Computational Intelligence, Anchorage, Alaska, IEEE, 84-89. 1998.

[11] Angeline P.J., Evolutionary Optimization Versus Particle Swarm Optimization: Philosophy and Performance Differences, Evolutionary Programming VII, Lecture Notes in Computer Science 1447, Springer, :601-610. 1998.

[12] Back T., Fogel D.B., and Michalewicz Z., (Eds), Handbook of Evolutionary Computation. Oxford University Press, 1997.

[13] Montiel O, Castillo O, Melin P, Rodriguez A and Sepulveda R: "Human evolutionary model: A new approach to optimization." Inf. Sci. 177 (10): 2075-2098, 2007.

[14] Castillo O., Valdez F. and Melin P., "Hierarchical Genetic Algorithms for topology optimization in fuzzy control systems". International Journal of General Systems, Volume 36, Issue 5., pag 575-591, October 2007.

[15] Valdez F., Melin P. and Castillo O. "Evolutionary Computing for the Optimization of Mathematical functions". Analysis and Design of Intelligent Systems Using Soft Computing Techniques. Advances in Soft Computing 41. June 2007.

[16] Castillo O, Huesca G, and Valdez F. "Proceedings of the International Conference on Artificial" Intelligence, IC-AI '04, Las Vegas, Nevada,USA, Volume 1, pag. 98-104,June 21-24, 2004.

[17] Veeramachaneni K, Osadciw L and Yan W "Improving Classifier Fusion Using Particle Swarm Optimization", IEEE Fusion Conference, Italy, July, 2006.

[18] M and Kennedy J. "The particle swarm-explosion, stability, and convergence in a multidimensional complex space". IEEE Transactions on Evolutionary Computation, 6(1):58-73. 2002.

[19] Kennedy J and Mendes R. "Population structure and particle swarm performance".Proceeding of IEEE conference on Evolutionary Computation, pag. 1671-1676. 2002.

[20] Wei W, Jiatao S, Zhongzxiu Y and Zheru . "Wavelet-based Illumination Compensation for Face Recognition using Eifenface Method Intelligent Control and Automation". WCICA 2006. Volume 2, Issue , 21-23 June 2006 Page(s): 10356 – 10360, 2006.

[21] Muller B, Reinhardt J, StricklandM. Neural networks: An introduction. 2nd edn, Springer-Verlag, Berlin. 1995.

[22] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull of Math Biophysics 5:115–133. 1943.

Fuzzy Subgroups and the Teichmüller Space

Paavo Kukkurainen

Department of Mathematics and Physics, Lappeenranta University of Technology
Lappeenranta, Finland
Email: paavo.kukkurainen@lut.fi

Abstract— *There exists a generalization of the Teichmüller space of a covering group. In this paper we combine this generalized Teichmüller space $T(G)$ and any fuzzy subgroup $\mathcal{A}: G \longrightarrow \mathcal{F}$ where G is a subgroup of the group consisting of such orientation preserving and orientation reversing Möbius transformations which act in the upper half-plane of the extended complex plane. A partially ordered set $\mathcal{F} = (\mathcal{F}, \leq)$ consists of stabilizers of G and all of their intersections. After preliminaries we present two new results concerning this special case of fuzzy subgroups. These conclusions are then applied to the known theory of the parametrization of the generalized Teichmüller space . As consequence, the equivalence classes of fuzzy subgroups (with an equivalence relation) become the elements of $T(G)$ where G is generated by a finite set of hyperbolic Möbius transformations. Let the number of the generators be n. Then there is an embedding $\psi: T(G) \longrightarrow \mathbb{R}^{3n-3}$ and therefore a homeomorphism $T(G) \longrightarrow \psi(T(G))$. Through parametrization of $T(G)$ $\psi(\mathcal{A}(G))$ has $3n-3$ real coordinates which are also the coordinates of $\mathcal{A}(G)$ up to identification.*

Keywords— Fuzzy subgroups, Möbius groups, Möbius transformations, compact Riemann surfaces, Teichmüller spaces.

1 Introduction

Orientation preserving and orientation reversing Möbius transformations form a group denoted by $\widehat{M}$. Following [2], we say that every subgroup of $\widehat{M}$ is a Möbius group. In Section 2 we first take some basic knowledge of Möbius transformations and then list all the types of Möbius transformations acting in the upper half-plane H which means that they are self-mappings of H. Let G and G' be Möbius groups fixing H. Then we define geometric isomorphisms $j: G \longrightarrow G'$ which play an important role in our representation.

In Section 3 we present two new Propositons 1 and 2 which the author of this paper has proved in [1]. It is not relevant to give the proofs in this connection because they are too long to show as well as many preliminaries are required. The propositions constitute the relationship between subgroups of the group of Möbius transformations and certain kind of fuzzy subgroups.

Let G, G', G_1 and G_2 be Möbius groups with a fixed G. Defining a delation $\delta(j)$ of an isomorphism $j: G \longrightarrow G'$ by means of multipliers of Möbius transformations $g \in G$ we say that any two isomorphisms $j_1: G \longrightarrow G_1$ and $j_2: G \longrightarrow G_2$ with some properties are equivalent precisely when $\delta(j_2 j_1^{-1}) = 1$. The set $E(G)$ of equivalence classes $[j]$ forms a metric space $(E(G), d)$ with a metric d. In the theory of compact Riemann surfaces this space can be regarded as a generalization of the Teichmüller space of a covering group G. So, $(E(G), d)$ is denoted by $T(G)$. All the presented premiliminaries and the parametrization of the generalized Teichmüller space can be found from [2], [3] and [4]. We will consider them in Section 4 so much as needed for Section 5.

Suppose that there exists a set F in G such that the parametrization condition in Proposition 3 is satisfied. Then we obtain an embedding $\psi: T(G) \longrightarrow \mathbb{R}^{3n-3}$ where G is finitely generated by n hyperbolic elements. More precisely, $T(G) \longrightarrow \psi(T(G))$ is a homeomorphism. Then, up to homeomorphism, $[j]$ is an element of $\mathbb{R}^{3n-3}$ with coordinates $k(j(g))$ which are the multipliers of Möbius transformations $j(g)$ for every $g \in F$.

In Section 5 we apply fuzzy subgroups to the theory of the parametrization of the generalized Teichmüller space. According to Corollary 1 the type-preserving isomorphisms $j_1: G \longrightarrow G_1$ and $j_2: G \longrightarrow G_2$ as well as the fuzzy subgroups $\mathcal{A}(j_1(g))$ and $\mathcal{A}(j_2(g))$ are simultaneously equivalent. The definition of the fuzzy subgroups is given by (3) in Proposition 1 and the equivalence between two fuzzy subgroups is defined by conjugation with a Möbius transformation. Roughly speaking, it is now possible to replace any element $[j]$ by the corresponding $[\mathcal{A}(j(g))]$ in Proposition 4. This leads to the representation of $[\mathcal{A}(j(g))]$ with real coordinates (up to identification) in the Euclidean space $\mathbb{R}^{3n-3}$.

In the following special case, we refer to [3]: If G is a covering group of the upper half-plane H over some compact Riemann surface S, then $[j]$ has coordinates in $\mathbb{R}^{6s-6+3n}$ where S is of genus s and n is the number of conformal disks removed from S. The coordinates of $[\mathcal{A}(j(g))]$ are given in subsection 5.2.

2 Preliminaries for Möbius transformations

In this section we refer to [2].

2.1 Möbius transformations

Directly conformal automorphisms of the extended complex plane $\widehat{\mathbb{C}}$ are orientation preserving Möbius transformations

$$g(z) = \frac{az+b}{cz+d} \quad a,\, b,\, c,\, d \in \mathbb{C}, \quad ad-bc = 1, \qquad (1)$$

and $g(-\frac{d}{c}) = \infty$, $g(\infty) = \frac{a}{c}$.

Indirectly conformal automorphisms of $\widehat{\mathbb{C}}$ are of the form

$$g(z) = \frac{a\bar{z}+b}{c\bar{z}+d} \quad a,\, b,\, c,\, d \in \mathbb{C}, \quad ad-bc = -1 \qquad (2)$$

which are orientation reversing Möbius transformations. The mappings (1) and (2) form the group $\widehat{M}$. Two transformations g_1 and g_2 in $\widehat{M}$ are conjugate if $g_1 = h g_2 h^{-1}$ for some Möbius transformation h.

ISBN: 978-989-95079-6-8

A point z is said to be a fixed point of g if $g(z) = z$. Every non-identity orientation preserving Möbius transformation g has one fixed point or two fixed points. Then

- g with one fixed point is parabolic and it is conjugate to $z \mapsto z+1$.
- If g has two fixed points and it is conjugate to $z \mapsto kz$ for some $k \in \mathbb{C} \setminus \{0\}$, then g is loxodromic if $|k| \neq 1$ and elliptic if $|k| = 1, k \neq 1$.
- Loxodromic transformations fixing the upper half-plane H are called hyperbolic. Otherwise they are strictly loxodromic.

In the loxodromic case we set

$$P(g) = \lim_{n\to\infty} g^n \text{ and } N(g) = \lim_{n\to\infty} g^{-n},$$

which are the attracting and repelling fixed points of g. Because the fixed points of any loxodromic g are real iff $g(H) = H$, hyperbolic transformations have fixed points on the real axis $\mathbb{R}$. Therefore $P(g)$ and $N(g)$ are real if g is hyperbolic. A circle or line in H perpendicular to to the real axis is a non-Euclidean line. Now the non-Euclidean line through $P(g)$ and $N(g)$ is called the axis $ax(g)$ of g. In the elliptic case $g(H) = H$ iff the fixed points of g are complex conjugates.

The multiplier of a Möbius transformation g is defined by means of the cross-ratio

$$k(g) = (g(z), z, x, y) = \frac{g(z)-x}{g(z)-y}\,\frac{z-y}{z-x},$$

where x and y are two different fixed points of g. In the parabolic case we set $k(g) = 1$. For the above complex number k we have $k = k(g)$. The multiplier is invariant in conjugation.

We are only interested in such orientation reversing Möbius transformations which act in H (fix H). This leads to the next consideration: the axis $ax(\sigma) = \{z \mid \sigma(z) = z\}$ of a reflection σ fixing H is also a circle or line orthogonal to $\mathbb{R}$. Denote by x and y the real fixed points of σ. Then

- the reflection $\sigma(z) = \eta(\bar{z})$ where η is the elliptic transformation defined by $k(\eta) = -1$, $\eta(x) = x$ and $\eta(y) = y$,
- a glide-reflection s fixing H is of the form $s = \tau\sigma$ where τ is a hyperbolic transformation fixing H, σ is a reflection fixing H and $ax(\tau) = ax(\sigma)$.

The reflection σ has an infinite number of fixed points but two real fixed points (if $ax(\sigma)$ is not a line). For the multiplier we set $k(\sigma) = -1$. The glide-reflection s and the hyperbolic τ have the same two fixed points. Moreover, the multiplier $k(s) = -k(\tau)$.

It is known that transformations (1) and (2) fix the upper-half plane H iff the coefficients a, b, c, d are real. In fact, there are the following types of Möbius transformations fixing H: hyperbolic, parabolic and elliptic transformations, the identity transformation, reflections and glide-reflections.

Suppose that G and G' are groups of Möbius transformations (or Möbius groups) acting in H. An isomorphism $j : G \longrightarrow G'$ is induced by a Möbius transformation h if $j(g) = hgh^{-1}$ for all $g \in G$. We say that $j : G \longrightarrow G'$ is type-preserving if g and $j(g)$ are of the same type for all $g \in G$. A type-preserving isomorphism $j : G \longrightarrow G'$ is geometric on $\mathbb{R}$ if there exists a homeomorphism $\varphi : \widehat{\mathbb{R}} \longrightarrow \widehat{\mathbb{R}}$ inducing j on $\widehat{\mathbb{R}}$, which is the boundary of the upper half-plane. Especially $\varphi(P(g)) = P(j(g))$ and $\varphi(N(g)) = N(j(g))$ for any hyperbolic g. Since every Möbius transformation is a homeomorphism $\widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}$, then for every φ which induces a geometric isomorphism j we have $h \mid \widehat{\mathbb{R}} = \varphi$ ($h : \widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}$ is a Möbius transformation).

3 Connection between fuzzy subgroups and Möbius transformations

Let $(P, \leq)$ be a partially ordered set and A a nonempty set.

(i) A mapping $\mathcal{A} : A \longrightarrow P$ is a P-(fuzzy)set on A.

(ii) For every $p \in P$, $A_p = \{x \in A \mid \mathcal{A}(x) \geq p\}$ is a p-level set subset (in short a level subset) of $\mathcal{A}$.

(iii) Let $\mathcal{G} = (G, \circ)$ be a group such that $(A_p, \circ)$ is a subgroup of $\mathcal{G}$ for every $p \in P$. Then a P-set on G, $\mathcal{A} : G \to P$, is said to be a P-(fuzzy) subgroup of $\mathcal{G}$.

Proposition 1. *[1] Let G be a subgroup of the group of Möbius transformations and let $G_z = \{g \in G \mid g(z) = z\}$ be a stabilizer of G at $z \in \widehat{\mathbb{C}}$. If $\mathcal{F}$ is a set of stabilizers of G and all of their intersections, and $\mathcal{F} = (\mathcal{F}, \leq)$ is a partially ordered set under $p \leq q$ iff $p \supseteq q$ ($p, q \in \mathcal{F}$), then $\mathcal{A} : G \to \mathcal{F}$*

$$\mathcal{A}(g) = \bigcap (p \in \mathcal{F} \mid g \in p) = \bigcap (G_z \mid g \in G_z) \qquad (3)$$

is a $\mathcal{F}$-subgroup of G.

We set an equivalence relation between fuzzy subgroups (3) by conjugation: $\mathcal{A}(g_1) \sim \mathcal{A}(g_2)$ iff there exists a Möbius transformation h such that $\mathcal{A}(g_1) = h\mathcal{A}(g_2)h^{-1}$. Observe that stabilizers form a group and the conjugation of the intersection of the groups G_z occurs by elements in formula (3). Recall that a Möbius group is a subgroup of the group $\widehat{M}$.

Proposition 2. *[1] Let G and G' be Möbius groups acting in the upper half-plane, $\mathcal{F}$ a set of stabilizers of G and all of their intersections. Let $\mathcal{A} : G \longrightarrow \mathcal{F}$ be defined by (3) and suppose that there is a type-preserving isomorphism $j : G \longrightarrow G'$. Then the following conditions are equivalent:*

(i) $\mathcal{A}(g) \sim \mathcal{A}(j(g))$ for every $g \in G$,

(ii) $k(g) = k(j(g))$ for every $g \in G$,

(iii) j is induced by a Möbius transformation,

(iv) $j : G \longrightarrow G'$ is a geometric isomorphism.

Further, let $j_1 : G \longrightarrow G_1$ and $j_2 : G \longrightarrow G_2$ be type-preserving isomorphisms. Then

$$\mathcal{A}(j_1(g)) \sim \mathcal{A}(j_2(g)) \quad \text{for every } g \in G$$

iff

$$j_2 j_1^{-1} : G_1 \longrightarrow G_2 \quad \text{is induced by a Möbius transformation.}$$

4 Coordinates of the generalized Teichmüller space

In this section we refer to [3].

ISBN: 978-989-95079-6-8

4.1 The set E

Let G and G' be Möbius groups acting in the upper half-plane H and $j : G \longrightarrow G'$ an isomorphism. We suppose that G has a set $E = \{g_1, g_2, \cdots\}$ of hyperbolic generators satisfying the following conditions:

(i) $ax(g_1) \cap ax(g_2)$ is a set of two points,

(ii) $(ax(g_1) \cup ax(g_2)) \cap ax(g_i) = \emptyset$, $i = 3,4,\cdots$,

(iii) $(N(g_i), P(g_i), N(g_1), P(g_1)) < 1$, $i = 3,4,\cdots$.

4.2 The dilation

Let $j : G \longrightarrow G'$ be an isomorphism between Möbius groups. The dilation $\delta(j)$ of j is the smallest number of the numbers $a \geq 1$ satisfying

$$|k(g)|^{\frac{1}{a}} \leq |k(j(g))| \leq |k(g)|^a \tag{4}$$

for all $g \in G$. If there exists no number $a \geq 1$ we define $\delta(j) = \infty$.

For a fixed G, let $J(G)$ be a set of all isomorphisms $j : G \longrightarrow G'$ with the following properties:

(i) G and G' are Möbius groups acting in H,

(ii) the dilation $\delta(j)$ is finite,

(iii) there exists a homeomorphism $\varphi : \widehat{\mathbb{R}} \longrightarrow \widehat{\mathbb{R}}$ such that $\varphi(P(g_i)) = P(j(g_i))$ and $\varphi(N(g_i)) = N(j(g_i))$ for all $g_i \in E$.

Define in $J(G)$ an equivalence relation $\sim$ by setting $j_1 \sim j_2$ iff $\delta(j_2 j_1^{-1}) = 1$. Then the set $E(G)$ of equivalence classes $[j]$ becomes a metric space $(E(G), d)$ with the dilation metric d defined by $d([j_1],[j_2]) = log\ \delta(j_2 j_1^{-1})$ [4].

Definition 1. *[3] The space $T(G) = (E(G), d)$ is called the generalized Teichmüller space of a Möbius group G.*

4.3 Parametrization of the Teichmüller space $T(G)$

Let us construct a set F containing the next elements:

(1) g_i, $i = 1,2,\cdots$,

(2) $g_i g_1$, $i = 2,3,\cdots$,

(3) $g_2 g_i g_2^{-1} g_i^{-1}$, $i = 3,4,\cdots$,

where every $g_i \in E$. The following proposition shows the significance of the set F if we want to find the condition for the parametrization of $T(G)$. We say that F parametrizes the space $T(G)$ or the set $J(G)$.

Proposition 3. *[3] Suppose that there exists a homeomorphism $\varphi : \widehat{\mathbb{R}} \longrightarrow \widehat{\mathbb{R}}$ for which $\varphi(P(g_i)) = P(j(g_i))$ and $\varphi(N(g_i)) = N(j(g_i))$ for all $g_i \in E$. If*

$$k(j(g)) = k(g)$$

for all $g \in F$, then j is induced by a Möbius transformation $h : \widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}$.

Since multipliers are invariant in conjugation the converse holds trivially.

Proposition 4. *[3] Suppose that $E = \{g_1, \cdots, g_n\}$ is finite. Then there is an embedding*

$$\psi : T(G) \longrightarrow \mathbb{R}^{3n-3},\ \psi([j]) = (x_1, \cdots, x_{3n-3}),$$

where

$$\begin{aligned} x_i &= k(j(g_i)),\ i = 1,\cdots,n, \\ x_{n-1+i} &= k(j(g_i)j(g_1)),\ i = 2,\cdots n, \\ x_{2n-3+i} &= k(j(g_2)j(g_i)j(g_2)^{-1}j(g_i)^{-1}),\ i = 3,\cdots,n. \end{aligned}$$

Proof. The mapping $\psi : T(G) \longrightarrow \mathbb{R}^{3n-3}$ is continuous with the dilation metric d [3], [4]. By Proposition 3, ψ is injective [3]. Furthermore, the inverse $\psi^{-1} : \psi(T(G)) \longrightarrow T(G)$ is continuous. Hence $\psi : T(G) \longrightarrow \psi(T(G))$ is a homeomorphism. □

5 Fuzzy subgroups in the Teichmüller space

5.1 The main results

Let G, G_1 and G_2 be Möbius groups. According to Proposition 2 and inequalities (4) we conclude

Corollary 1. *Let $j_1 : G \longrightarrow G_1$ and $j_2 : G \longrightarrow G_2$ be type-preserving isomorphisms. Then the following statements are equivalent:*

(i) $j_1 \sim j_2$,

(ii) $\delta(j_2 j_1^{-1}) = 1$,

(iii) $k(j_2 j_1^{-1}(g)) = k(g)$ *for every* $g \in G$,

(iv) $j_2 j_1^{-1} : G_1 \longrightarrow G_2$ *is induced by a Möbius transformation,*

(v) $\mathcal{A}(j_1(g)) \sim \mathcal{A}(j_2(g))$ *for every* $g \in G$.

Let $j \in J(G)$ and suppose that $j : G \longrightarrow G'$ is geometric. Then j is type-preserving and satisfies the condition *(iii)* in subsection 4.2. Following the definition of $E(G)$, let us denote the set of equivalence classes $[\mathcal{A}(j(g))]$ by $E'(G)$. By Corollary 1,*(i)*,*(v)*, $(E'(G), d)$ becomes also a metric space having the same metric d. This leads to the new form of the generalized Teichmüller space, $T'(G) = (E'(G), d)$. Reformulation of Proposition 4 for the fuzzy subgroups $\mathcal{A}(j(g))$ yields

Proposition 5. *Suppose that $E = \{g_1, \cdots, g_n\}$ is finite. Then there is an embedding*

$$\psi' : T'(G) \longrightarrow \mathbb{R}^{3n-3},\ \psi'([\mathcal{A}(j(g)]) = (x_1, \cdots, x_{3n-3}),$$

for every $g \in G$, and where

$$\begin{aligned} x_i &= k(j(x_i)),\ i = 1,\cdots,n, \\ x_{n-1+i} &= k(j(g_i)j(g_1)),\ i = 2,\cdots n, \\ x_{2n-3+i} &= k(j(g_2)j(g_i)j(g_2)^{-1}j(g_i)^{-1}),\ i = 3,\cdots,n. \end{aligned}$$

ISBN: 978-989-95079-6-8

5.2 The covering group of a compact Riemann surface

Let G be a Möbius group acting in the upper half-plane H and suppose that G is a covering group of H over the compact Riemann surface S. Since any non-identity cover transformation of G has no fixed points in H, we know that all elements of G are hyperbolic or parabolic Möbius transformations, or it is the identity mapping. We say that G is of signature (s,n) if $S = H/G$ is of genus s from which n conformal disks are removed.

Assume that G is a fixed covering group of signature (s,n) with $s > 0$ and $n > 0$. Then G has $2s+n$ hyperbolic generators satisfying the defining relation:

$$c_n \cdots c_1\, b_s^{-1}\, a_s^{-1}\, b_s\, a_s \cdots b_1^{-1}\, a_1^{-1}\, b_1\, a_1 \;=\; id$$

from which c_n can be solved. As conclusion the set

$$E \;=\; \{a_1, b_1, a_2^{-1}, b_2, \cdots, a_s^{-1}, b_s, c_1^{-1}, \cdots, c_{n-1}^{-1}\}$$

generates G freely. Moreover, E satisfies the conditions $(i)-(iii)$ in subsection 4.1. Then the mapping $\psi' : T'(G) \longrightarrow \mathbb{R}^{3n-3}$ in Proposition 5 takes the form

$$\psi' : T'(G) \longrightarrow \mathbb{R}^{6s-6+3n},\;\; \psi'([\mathcal{A}(j(g)]) = (x_1, \cdots, x_{6s-6+3n})$$

with coordinates

$$\begin{aligned}
x_t \;=\;& k(j(a_i)),\;\; i = 1, \cdots, s,\\
& k(j(b_i)),\;\; i = 1, \cdots, s\\
& k(j(c_i)),\;\; i = 1, \cdots, n-1,\\
& \text{for } t = 1, \cdots, 2s+n-1,\\
x_{2s+n-2+t} \;=\;& k(j(b_i)j(a_1)),\;\; i = 1, \cdots, s,\\
& k(j(a_i)^{-1})j(a_1)),\;\; i = 2, \cdots, s,\\
& k(j(c_i)^{-1})j(a_1))\;\; i = 1, \cdots, n-1,\\
& \text{for } t = 2, \cdots 2s+n-1,\\
x_{4s+2n-5+t} \;=\;& k(j(b_1)j((a_i)^{-1})j((b_1)^{-1})j(a_i)),\\
& i = 2, \cdots, s\\
& k(j(b_1)(j(b_i)j((b_1)^{-1})j((b_i)^{-1})),\\
& i = 2, \cdots, s\\
& k(j(b_1)j((c_i)^{-1})j((b_1)^{-1})j(c_i)),\\
& i = 1, \cdots, n-1,\\
& \text{for } t = 3, \cdots, 2s+n-1.
\end{aligned}$$

Example 1. Let us construct a compact Riemann surface $S = H/G$ with a covering group G of signature $(1,1)$. Then S consists of one handle from which one conformal disk is removed. Therefore G has a set $E = \{a, b, c^{-1}\}$ of hyperbolic generators with the defining relation

$$c^{-1}\, a\, b \;=\; id,$$

where c is a boundary mapping of the disk satisfying $c = ab$. Then a and b generates G freely, $G =< a, b >$. In fact, the hyperbolic a and b have no comman fixed points and G is a purely hyperbolic group. It does not contain parabolic or elliptic elements. The first one is concluded from [3] and the second one holds because G is a covering group.

An embedding

$$\begin{aligned}
&\psi' : E'(G) \longrightarrow \mathbb{R}^3,\\
&\psi'([\mathcal{A}(j(g)]) = (k(j(a_1)), k(j(b_1)), k(j(b_1)j(a_1)))
\end{aligned}$$

gives the coordinates $(k(j(a_1)), k(j(b_1)), k(j(b_1)j(a_1)))$ of $[\mathcal{A}(j(g)]$ up to homeomorphism for any $g \in G$ and some geometric isomorphism j of G and $j \in J(G)$.

References

[1] P. Kukkurainen, Fuzzy subgroups, algebraic and topological points of view and complex analysis. Thesis for the degree of Doctor of Philosophy, Acta Universitatis Lappeenrantaensis, 2009. In printed.

[2] M. Seppälä and T. Sorvali, Geometry of Riemann Surfaces and Teichmüller Spaces, Elsevier Science Publishers B.V., North-Holland Amsterdam, 1992.

[3] T. Sorvali, Parametrization of free Möbius groups, Annales Academiae Scientiarum Fennicae. Series A. I.Mathematica 579, 1974, 1-12.

[4] T. Sorvali, The boundary mapping induced by an isomorphism of covering groups, Annales Academiae Scientiarum Fennicae. Series A. I.Mathematica 526, 1972, 4-31.

ISBN: 978-989-95079-6-8

Reasoning about Actions in Fuzzy Environment

Emad Saad[1] Shaimaa A. Elmorsy[2] Mahmoud M. H. Gabr[2] Yasser F. M. Hassan[2]

1. Department of Computer Science, Gulf University for Science and Technology
West Mishref, Kuwait
2. Department of Mathematics and Computer Science, Faculty of Science, Alexandria University
Alexandria, Egypt
Email: saad.e@gust.edu.kw, shaimaa_elmorsy@hotmail.com, mahgabr@yahoo.com, y.fouad@sci.alex.edu.eg

Abstract— *Reasoning in the presence of imprecision and vagueness is inevitable in many real-world applications including those in robotics and intelligent agents. Although, reasoning about actions is a major component in these real-world applications, current actions languages for reasoning about actions lack the ability to represent and reason about actions in the presence of imprecision and vagueness that stem from effects of actions in these real-world applications. In this paper we present a new action language called fuzzy action language, $\mathcal{A}_F$, that allows the representation and reasoning about actions with vague (fuzzy) effects. In addition we define the notions of fuzzy planning and fuzzy plan in the fuzzy action language $\mathcal{A}_F$. Furthermore, we describe a fuzzy planner based on the fuzzy action language A_F that is developed by translating a fuzzy action theory,* $\mathbf{FT}$*, in A_F into a normal logic program with answer set semantics,* Π*, where trajectories in* $\mathbf{FT}$ *are equivalent to the answer sets of* Π*. In addition, we formally prove the correctness of our translation. Furthermore, we show that fuzzy planning problems can be encoded as a SAT problem.*

Keywords— Answer set programming, fuzzy planning, reasoning about actions with fuzzy effects, reasoning under uncertainty.

1 Introduction

Reasoning about the properties of actions is fundamental in many real-world applications. Therefore, many action languages that allow representing and reasoning about actions have been developed which include [2, 5, 9, 10, 11, 12, 18, 21, 22, 24]. Furthermore, uncertainty is a main issue in representing and reasoning about actions in these real-world applications. Uncertainty in these real-world applications stems from different sources including incompleteness, erroneous, or imprecision (vagueness). For reasoning under uncertainty, probability theory is used to reason in the presence of incompleteness or erroneous, however, fuzzy set theory is used to reason in the presence of imprecision or vagueness. To deal with probabilistic uncertainty in reasoning about actions, different proposals have been presented. These proposals include [2, 5, 12, 21, 22]. Although, these proposals deal with uncertainty probabilistically, they are inappropriate in situations where actions have imprecise or vague effects, which need a different formalism from the one required in reasoning about actions with probabilistic effects as in [2, 5, 12, 21, 22]. This is because the underlying formalism in reasoning about actions with probabilistic effects is the probability theory, however, the formalism required for reasoning about actions with imprecise effects is the fuzzy set theory. Actions with imprecise effects appear in many domains including robotics and intelligent agents. Consider the following example adapted from [23]. Consider a planner that controls an autonomous car that is moving in a highway with the goal to keep the car close to the middle of a lane. The autonomous car planner's uses the action navigate whose effect is to make the car close to the middle of the lane. The effect of the action navigate is neither probabilistically nor precisely defined. Therefore, action languages, for example [9, 10, 11, 18, 24], that reason about actions whose effects are precisely defined (either entirely true or entirely false), and action languages, for example [2, 5, 12, 21, 22], that reason about actions whose effects are probabilistically defined (either true or false to a probabilistic degree), can not be used to correctly represent and reason about these kind of actions whose effects are imprecisely or vaguely defined.

In this paper we present a new action language called fuzzy action language, $\mathcal{A}_F$, that allows the representation and reasoning about actions with vague (fuzzy) effects. In addition, we define the notions of fuzzy planning and fuzzy plan in the fuzzy action language $\mathcal{A}_F$. Furthermore, we describe a fuzzy planner based on the fuzzy action language $\mathcal{A}_F$ that is developed by translating a fuzzy action theory, $\mathbf{FT}$, in $\mathcal{A}_F$ into a normal logic program with answer set semantics, Π, where trajectories in $\mathbf{FT}$ are equivalent to the answer sets of Π. In addition, we formally prove the correctness of our translation. Furthermore, we show that fuzzy planning problems can be encoded as SAT problems.

This paper is organized as follows. Section 2 reviews the answer set semantics of normal logic programs and the fundamental notions of the fuzzy set theory. The syntax and semantics of the fuzzy action language, $\mathcal{A}_F$, is introduced in section 3. In section 4, a translation from a fuzzy planning problem in the fuzzy action language $\mathcal{A}_F$ into a normal logic program with answer set semantics is presented. The correctness of the translation is introduced in section 5. Finally, conclusion and related work is presented in section 6.

2 Preliminaries

In this section we review the answer set semantics of normal logic programs [7] and the basic notions of fuzzy set theory [27].

2.1 Normal Logic Programs with Answer Set Semantics

Let $\mathcal{L}$ be a first-order language with finitely many predicate symbols, function symbols, constants, and infinitely many variables. The Herbrand base of $\mathcal{L}$ is denoted by $\mathcal{B}$. A Her-

ISBN: 978-989-95079-6-8

brand interpretation is a subset of the Herbrand base $\mathcal{B}$. A normal logic program is a finite set of rules of the form

$$a \leftarrow a_1, \ldots, a_n, not\ b_1, \ldots, not\ b_m$$

where a, $a_1, \ldots, a_n, b_1, \ldots, b_m$ are atoms and not is the negation-as-failure. Intuitively, the meaning of the above rule is that if it is believable (provable) that a_i is true and it is not believable that b_i is true then it is believable that a is true. A normal logic program is ground if no variables appear in any of its rules.

Let Π be a ground normal logic program, S be a Herbrand interpretation, and r be a rule as above. Then, we say that

- S satisfies a_i (denoted by $S \models a_i$) iff $a_i \in S$.
- S satisfies $not\ b_j$ (denoted by $S \models not\ b_j$) iff $b_j \notin S$.
- S satisfies $(a_1, \ldots, a_n, not\ \ b_1, \ldots, not\ \ b_m)$ (denoted by $S \models (a_1, \ldots, a_n, not\ \ b_1, \ldots, not\ \ b_m)$) iff $\{a_1, \ldots, a_n\} \subseteq S$ and $\{b_1, \ldots, b_m\} \nsubseteq S$.
- S satisfies $a \leftarrow a_1, \ldots, a_n, not\ b_1, \ldots, not\ b_m$ iff $S \models a$ or $S \not\models (a_1, \ldots, a_n, not\ b_1, \ldots, not\ b_m)$.

A Herbrand interpretation is said to satisfy a normal logic program Π iff it satisfies every rule in Π. A Herbrand model of a normal logic program Π is a Herbrand interpretation of Π that satisfies Π. A Herbrand interpretation S of a normal logic program Π is said to be an answer set of Π if S is the minimal Herbrand model (with respect to the set inclusion) of the reduct, denoted by Π^S, of Π w.r.t. S, where Π^S is a set of rules of the form

$$a \leftarrow a_1, \ldots, a_n$$

such that

$$a \leftarrow a_1, \ldots, a_n, not\ b_1, \ldots, not\ b_m \in \Pi$$

and $\{b_1, \ldots, b_m\} \nsubseteq S$

Intuitively, for any $not\ \ b_j$ in the body of a rule in Π with $b_j \notin S$ is simply satisfied by S, and $not\ b_j$ is safely removed from the body of the rule. If $b_j \in S$ then the body of the rule is not satisfied and the rule is trivially ignored.

2.2 Fuzzy Sets

In this section we review the basic notions of fuzzy sets as presented in [27]. Let U be a set of objects. A fuzzy set, F, in U is defined by the grade membership function $\mu_F : U \rightarrow [0, 1]$, where for each element $x \in U$, μ_F assigns to x a value $\mu_F(x)$ in $[0, 1]$. The support for F denotes the set of all objects x in U for which the grade membership of x in F is a non-zero value. Formally, $support(F) = \{x \in U \mid \mu_F(x) > 0\}$. The intersection (conjunction) of two fuzzy sets F and F' in U, denoted by $F \wedge_f F'$ is a fuzzy set G in U where the grade membership function of G is $\mu_G(x) = \min(\mu_F(x), \mu_{F'}(x))$ for all $x \in U$. However, the union (disjunction) of two fuzzy sets F and F' in U, denoted by $F \vee_f F'$ is a fuzzy set G in U where the grade membership function of G is $\mu_G(x) = \max(\mu_F(x), \mu_{F'}(x))$ for all $x \in U$. The complement (negation) of a fuzzy set F in U is a fuzzy set in U denoted by $\overline{F}$ where the grade membership function of $\overline{F}$ is $\mu_{\overline{F}}(x) = 1 - \mu_F(x)$ for all $x \in U$. A fuzzy set F in U is said to be contained in another fuzzy set G in U if and only if $\mu_F(x) \leq \mu_G(x)$ for all $x \in U$. Notice that we use the notations $\wedge_f$ and $\vee_f$ to denote fuzzy conjunction and fuzzy disjunction respectively to distinguish them from $\wedge$ and $\vee$ for propositional conjunction and disjunction respectively. Furthermore, other function characterizations for the fuzzy conjunction and fuzzy disjunction operators can be used. However, we will stick with the min and max function characterizations for the fuzzy conjunction and fuzzy disjunction as originally proposed in [27].

3 Fuzzy Action Language $\mathcal{A}_F$

In this section we develop a novel action language, called *fuzzy action language*, $\mathcal{A}_F$, that allows the representation of actions with vague (fuzzy) effects. An action theory in $\mathcal{A}_F$ is capable of representing the grade membership distribution of the possible initial states, the executability conditions of actions, and the grade membership distribution of the fuzzy effects of actions. The semantics of $\mathcal{A}_F$ is based on a transition function that maps an action and a set of states to a set of states. The fuzzy action language $\mathcal{A}_F$ is inspired by the action languages in [21, 22].

3.1 Language syntax

A predicate that describes a property of the environment, which may contain variables, is called a fluent. Let $\mathcal{A}$ be a set of action names that can contain variables and $\mathcal{F}$ be a set of fluents. A fluent $f \in \mathcal{F}$ or $\neg f$, the negation of f is called a fluent literal. A conjunction of fluent literals of the form $l_1 \wedge \ldots \wedge l_n$ is called a conjunctive fluent formula, where $l_1, \ldots, l_n$ are fluent literals. Sometimes we abuse the notation and refer to a conjunctive fluent formula as a set of fluent literals ($\emptyset$ denotes $true$).

A fuzzy action theory, $\mathbf{FT}$, in $\mathcal{A}_F$ is a tuple of the form $\mathbf{FT} = \langle S_0, \mathcal{AD} \rangle$, where S_0 is a proposition of the form (1), $\mathcal{AD}$ is a set of propositions from (2-3) as follows:

$$\textbf{initially} \begin{cases} \psi_1 \ : \ v_1 \\ \psi_2 \ : \ v_2 \\ \quad \ldots \\ \psi_n \ : \ v_n \end{cases} \tag{1}$$

$$\textbf{executable}\ a\ \textbf{if}\ \psi \tag{2}$$

$$a\ \textbf{causes} \begin{cases} \phi_1 \ : \ v_1\ \textbf{if}\ \psi_1 \\ \phi_2 \ : \ v_2\ \textbf{if}\ \psi_2 \\ \quad \ldots \\ \phi_n \ : \ v_n\ \textbf{if}\ \psi_n \end{cases} \tag{3}$$

where $\psi, \psi_1, \ldots, \psi_n, \phi_1, \ldots, \phi_n$ are conjunctive fluent formulae, $a \in \mathcal{A}$ is an action, and for all $1 \leq i \leq n$, we have $v_i \in [0, 1]$. The set of all ψ_i must be mutually exclusive. Let $\mathcal{S}$ be the set of all states formed from the fluents in $\mathcal{F}$.

Proposition (1) is a fuzzy set in the set of all states $\mathcal{S}$. It represents the grade membership of the possible initial states. We consider that only ψ_i with grade membership that is non-zero is listed in (1). Proposition (1) says that the grade membership of the possible initial state ψ_i is v_i for all $1 \leq i \leq n$. Proposition (2) represents the *executability condition* of actions, where each variable that appears in a also appears in ψ. It states that an action a is executable in any state in which ψ holds. A proposition of the form (3) represents the fuzzy (vague) effects resulting from executing an action a in the

ISBN: 978-989-95079-6-8

states in which a is executable. For each $1 \leq i \leq n$, all variables that appear in ϕ_i also appear in a and ψ_i. It describes that a causes ϕ_i to hold with grade membership v_i in a successor state to a state in which a is executed and ψ_i holds for all $1 \leq i \leq n$. For each $1 \leq i \leq n$, ψ_i is called a precondition of an action a that corresponds to an effect ϕ_i, ϕ_i is called an effect of a, v_i is the grade membership that ϕ_i holds given that ψ_i holds, where v_i is a non-zero value in $[0, 1]$. For any proposition of the form (3), the set of ground preconditions ψ_i are mutually exclusive and exhaustive.

It will be more convenient for the subsequent results to represent an action a as a set of the form $a = \{a_1, \ldots, a_n\}$, where each a_i corresponds to ϕ_i, v_i, and ψ_i. Therefore, alternatively, for each $1 \leq i \leq n$, proposition (3) can be represented as

$$a_i \textbf{ causes } \phi_i \ : \ v_i \textbf{ if } \psi_i$$

A fuzzy action theory is ground if it does not contain any variables.

3.2 Semantics

A consistent set of ground literals ϕ is a set of literals that does not contain a pair of complementary literals, i.e., l and $\neg l \notin \phi$. If a literal $l \in \phi$, then we say l is true (holds) in ϕ (denoted by $\phi \models l$), and l is false (does not hold) in ϕ if $\neg l$ is in ϕ (denoted by $\phi \models \neg\, l$). If a set of literals σ is contained in ϕ then we say σ is true (holds) in ϕ (denoted by $\phi \models \sigma$), otherwise, σ is false (does not hold) in ϕ (denoted by $\phi \not\models \sigma$). A state s is a complete and consistent set of literals that describes the world at a certain time point.

Definition 1 *Let $\mathbf{FT} = \langle S_0, \mathcal{AD} \rangle$ be a ground fuzzy action theory, $\mathcal{S}$ is the set of all states in $\mathbf{FT}$, s be a state whose grade membership $\mu_S(s)$ is v, a_i **causes** ϕ_i : v_i : **if** ψ_i ($1 \leq i \leq n$) be a proposition in $\mathcal{AD}$, and $a = \{a_1, \ldots, a_n\}$ be an action, where each a_i corresponds to ϕ_i, v_i, and ψ_i for $1 \leq i \leq n$. Then, $s' = \Phi(a_i, s)$ is the state resulting from executing a in s, given that a is executable in s, where:*

- *$l \in \Phi(a_i, s)$ and $\neg\ l \notin \Phi(a_i, s)$ if $l \in \phi_i$ and the precondition ψ_i holds in s.*
- *$\neg\ l \in \Phi(a_i, s)$ and $l \notin \Phi(a_i, s)$ if $\neg\ l \in \phi_i$ and the precondition ψ_i holds in s.*
- *Otherwise, $l \in \Phi(a_i, s)$ iff $l \in s$ and $\neg\ l \in \Phi(a_i, s)$ iff $\neg\ l \in s$.*

where the grade membership of the resulting state $s' = \Phi(a_i, s)$ is $\mu_{\mathcal{S}}(s') = \min(\mu_{\mathcal{S}}(s), v_i) = \min(v, v_i)$. We call Φ a fuzzy transition function.

Example 1 *Consider the following fuzzy planning task adapted from [14]. An arm of a robot is grasping a block from a table, where the* pickup *action the robot performs has effects that are imprecisely defined. The arm is able to tightly hold a block (hb) with a grade membership 0.9 after executing the* pickup *action in the state in which the gripper is dry (gd), and the arm tightly cannot hold the block ($\neg hb$), after executing the* pickup *action in the same state, with grade membership value 0.4. On the other hand, when executing the* pickup *action in the state while the gripper is not dry ($\neg\ gd$) causes the block to be tightly held (hb) with grade of membership equal to 0.3 and tightly not held ($\neg\ hb$) with grade membership 0.6. We assume the initial grade membership distribution of the initial states of the world is given by the grade membership function $\mu_{\mathcal{S}}$ such that $\mu_{\mathcal{S}}(s_1) = 0.9$ and $\mu_{\mathcal{S}}(s_2) = 0.5$, where $s_1 = \{gd, \neg hb\}$ and $s_2 = \{\neg gd, \neg hb\}$, with the understanding that the grade membership of the other states of the world is 0. This fuzzy planning domain can be represented in the fuzzy action language $\mathcal{A}_F$ as the fuzzy action theory $\mathbf{FT} = \langle S_0, \mathcal{AD} \rangle$ where*

$$S_0 = \textbf{initially} \left\{ \begin{array}{lcl} \{gd, \neg hb\} & : & 0.9 \\ \{\neg gd, \neg hb\} & : & 0.5 \end{array} \right.$$

and $\mathcal{AD}$ consists of:

$$\textbf{executable } pickup \textbf{ if } \emptyset$$

$$pickup \textbf{ causes} \left\{ \begin{array}{lcllll} \{hb\} & : & 0.9 & \textbf{if} & \{gd\} \\ \{\neg hb\} & : & 0.4 & \textbf{if} & \{gd\} \\ \{hb\} & : & 0.3 & \textbf{if} & \{\neg\, gd\} \\ \{\neg\, hb\} & : & 0.6 & \textbf{if} & \{\neg\, gd\} \end{array} \right.$$

The $pickup$ action can be represented as the set $pickup = \{pickup_1, pickup_2, pickup_3, pickup_4\}$, where

$$\begin{array}{llllll} pickup_1 & \textbf{causes} & \{hb\} & : & 0.9 & \textbf{if } \{gd\} \\ pickup_2 & \textbf{causes} & \{\neg hb\} & : & 0.4 & \textbf{if } \{gd\} \\ pickup_3 & \textbf{causes} & \{hb\} & : & 0.3 & \textbf{if } \{\neg\, gd\} \\ pickup_4 & \textbf{causes} & \{\neg\, hb\} & : & 0.6 & \textbf{if } \{\neg\, gd\} \end{array}$$

The grade membership distribution resulting from executing the $pickup$ action in the initial states s_1 and s_2 is given by:

- $s_1^{'} = \{gd, hb\} = \Phi(pickup_1, \{gd, \neg hb\})$, where $\mu_S(s_1^{'}) = \min(\mu_S(s_1), 0.9) = \min(0.9, 0.9) = 0.9$.
- $s_2^{'} = \{gd, \neg hb\} = \Phi(pickup_2, \{gd, \neg hb\})$, where $\mu_S(s_2^{'}) = \min(\mu_S(s_1), 0.4) = \min(0.9, 0.4) = 0.4$.
- $s_3^{'} = \{\neg gd, hb\} = \Phi(pickup_3, \{\neg gd, \neg hb\})$, where $\mu_S(s_3^{'}) = \min(\mu_S(s_2), 0.3) = \min(0.5, 0.3) = 0.3$.
- $s_4^{'} = \{\neg gd, \neg hb\} = \Phi(pickup_4, \{\neg gd, \neg hb\})$, where $\mu_S(s_4^{'}) = \min(\mu_S(s_2), 0.6) = \min(0.5, 0.6) = 0.5$.

Definition 2 (Fuzzy Plan) *A sequence of actions $\langle a_0, a_1, \ldots, a_{n-1} \rangle$ is called a fuzzy plan, where each $(0 \leq i \leq n-1)$ a_i is an action with fuzzy effects.*

The grade membership that a fuzzy plan P satisfies a conjunctive fluent formula $\mathcal{G}$ after executing P in a given state s is given by the following definition.

Definition 3 *Let $\mathbf{FT}$ be a ground fuzzy action theory, s, s' be states, s_I be a variable ranging over the possible initial states, $\mathcal{G}$ be a conjunctive fluent formula, and $\langle a_0, a_1, \ldots, a_{n-1} \rangle$ be a fuzzy plan. Then the grade membership that $\mathcal{G}$ is true in a state s' that results after executing $\langle a_0, a_1, \ldots, a_{n-1} \rangle$ in the possible initial states s_I is given by*

$$\mu'_{\mathcal{P}}(s' | s_I, \langle a_0, a_1, \ldots, a_{n-1} \rangle) =$$

$$\max_{s''}(\min_{s_I = s}(\mu'_{\mathcal{P}}(s'' | s, a_0),\ \mu'_{\mathcal{P}}(s' | s'', \langle a_1, \ldots, a_{n-1} \rangle)))$$

where $\mathcal{P}$ is the set of all plans in $\mathbf{FT}$, and $\mu'_{\mathcal{P}}(s'' | s, a_0) = \mu_{\mathcal{S}}(s'') = \min(\mu_S(s), v)$ such that $s'' = \Phi(a_0, s)$. In general, for any action $a \in \mathcal{A}$, $\mu'_{\mathcal{P}}(s'' | s, a) = \mu_{\mathcal{S}}(s'') = \min(\mu_S(s), v)$ such that $s'' = \Phi(a, s)$.

ISBN: 978-989-95079-6-8

Definition 4 *A fuzzy planning problem is a 4-tuple* $\mathbf{FP} = \langle S_0, \mathcal{AD}, \mathcal{G}, \mathcal{T} \rangle$*, where* S_0 *is a fuzzy set in the set of all states* S *that represents the initial agent knowledge about the world states at the time of execution (the initial grade membership distribution over states),* $\mathcal{AD}$ *is a fuzzy action description,* $\mathcal{G}$ *is conjunctive fluent formula represents the goal to be satisfied, and* $0 \leq \mathcal{T} \leq 1$ *is the fuzzy threshold for the goal* $\mathcal{G}$ *to be achieved. We say* $\langle a_0, \ldots, a_{n-1} \rangle$ *is a fuzzy plan for* $\mathbf{FP}$ *iff each* a_i *appears in* $\mathcal{AD}$ *and* $\mu'_{\mathcal{P}}(s'|s_I, \langle a_0, a_1, \ldots, a_{n-1} \rangle) \geq \mathcal{T}$*, where* s_I *is a variable ranging over the possible initial states, and* $\mathcal{G}$ *is true in* s'*.*

4 Fuzzy Planning Using Answer Set Programming

This section presents a translation from a fuzzy planning problem $\mathbf{FP} = \langle S_0, \mathcal{AD}, \mathcal{G}, \mathcal{T} \rangle$ into a normal logic program with answer set semantics, $\Pi_{\mathbf{FP}}$, where the rules in $\Pi_{\mathbf{FP}}$ encode (1) the initial grade membership distribution S_0, (2) the fuzzy transition function Φ, (3) the fuzzy action description $\mathcal{AD}$, (4) and the goal $\mathcal{G}$. The answer sets of $\Pi_{\mathbf{FP}}$ correspond to valid trajectories in $\mathbf{FP}$. The normal logic program translation of a fuzzy planning problem $\mathbf{FP}$ is mainly adapted from [25]. We assume that the length of the fuzzy plan that we are looking for is known. We use the predicates $holds(L, T)$ to represent the fact that a literal L holds at time moment T and $occ(AC, T)$ to describe that an action AC executes at time moment T. We use lower case letters to represent constants and upper case letters to represent variables.

Let $\Pi_{\mathbf{FP}}$ be the normal logic program translation of a fuzzy planning problem $\mathbf{FP} = \langle S_0, \mathcal{AD}, \mathcal{G}, \mathcal{T} \rangle$, where $\Pi_{\mathbf{FP}}$ is the set of rules described as follows. In addition, given p is a predicate and $\psi = \{l_1, \ldots, l_n\}$, we use $p(\psi)$ to denote $p(l_1), \ldots, p(l_n)$.

- For each action $a = \{a_1, \ldots, a_n\} \in \mathcal{A}$, we add to $\Pi_{\mathbf{FP}}$ the set of facts

$$action(a_i) \leftarrow \tag{4}$$

for each $1 \leq i \leq n$. States of the world are described by literals that are encoded in $\Pi_{\mathbf{FP}}$ by the rules

$$literal(A) \leftarrow atom(A) \tag{5}$$
$$literal(\neg A) \leftarrow atom(A) \tag{6}$$

where $atom(A)$ is a set of facts that describe the properties of the world. To present that A and $\neg A$ are contrary literals, the following rules are added to $\Pi_{\mathbf{FP}}$.

$$contrary(A, \neg A) \leftarrow atom(A) \tag{7}$$
$$contrary(\neg A, A) \leftarrow atom(A) \tag{8}$$

- The initial grade membership distribution **initially** ψ_i : v_i for $1 \leq i \leq n$ is represented in $\Pi_{\mathbf{FP}}$ as follows. Consider that $s_1, s_2, \ldots, s_n$ form the set of possible initial states, where for each $1 \leq i \leq n$, $s_i = \{l_1^i, \ldots, l_m^i\}$, and the grade membership of s_i is $\mu_{\mathcal{S}}(s_i) = v_i$. Moreover, let $s = s_1 \cup s_2 \cup \ldots \cup s_n$, $s' = s_1 \cap s_2 \cap \ldots \cap s_n$, $\widehat{s} = s - s'$, and $s'' = \{\, l \,|\, l \in \widehat{s} \vee \neg l \in \widehat{s}\}$. To generate the set of all possible initial states, the following set of rules are added to $\Pi_{\mathbf{FP}}$. For each literal $l \in s'$, the fact

$$holds(l, 0) \leftarrow \tag{9}$$

is in $\Pi_{\mathbf{FP}}$. This fact presents that the literal l holds at time moment 0. This set of facts specifies the set of literals that hold in every possible initial state. Moreover, for each literal $l \in s''$, we add to $\Pi_{\mathbf{FP}}$ the rules

$$holds(l, 0) \leftarrow not\ holds(\neg l, 0) \tag{10}$$
$$holds(\neg l, 0) \leftarrow not\ holds(l, 0) \tag{11}$$

The above rules say that the literal l (similarly $\neg l$) holds at time moment 0, if $\neg l$ (similarly l) does not hold at the time moment 0.

- Each executability condition proposition of an action $a = \{a_1, \ldots, a_n\}$ of the form (2) is encoded in $\Pi_{\mathbf{FP}}$ for each $1 \leq i \leq n$ as

$$exec(a_i, T) \leftarrow holds(\psi, T) \tag{12}$$

- For each proposition of the form a_i **causes** ϕ_i : v_i **if** ψ_i ($1 \leq i \leq n$), in $\mathcal{AD}$, we proceed as follows. Let $\phi_i = \{l_i^1, \ldots, l_i^m\}$. Then, $\forall$ $(1 \leq i \leq n)$, we have for each $1 \leq j \leq m$,

$$holds(l_i^j, T+1) \leftarrow occ(a_i, T), exec(a_i, T), holds(\psi_i, T) \tag{13}$$

belongs to $\Pi_{\mathbf{FP}}$. This rule states that if the action a occurs at time moment T and the precondition ψ_i holds at the same time moment, then the literal l_i^j holds at the time moment $T+1$.

- The frame axioms are presented in $\Pi_{\mathbf{FP}}$ as below. For any literal L we have the rule

$$holds(L, T+1) \leftarrow holds(L, T), not\ holds(L', T+1), contrary(L, L') \tag{14}$$

in $\Pi_{\mathbf{FP}}$. The above rule states that L holds at the time moment $T+1$ if it holds at the time moment T and its contrary does not hold at the time moment $T+1$.

- To encode the fact that a literal A and its negation $\neg A$ cannot hold at the same time, we add the following rule to $\Pi_{\mathbf{FP}}$

$$\leftarrow holds(A, T), holds(\neg A, T) \tag{15}$$

- Action generation rules are described by

$$occ(AC^i, T) \leftarrow action(AC^i), not\ abocc(AC^i, T) \tag{16}$$
$$abocc(AC^i, T) \leftarrow action(AC^i), action(AC^j), occ(AC^j, T), AC^i \neq AC^j \tag{17}$$

The above rules generate action occurrences once at a time, where AC^i and AC^j are variables representing actions.

- Let $\mathcal{G} = g_1 \wedge \ldots \wedge g_m$ be a goal expression, then $\mathcal{G}$ is encoded in $\Pi_{\mathbf{FP}}$ as

$$goal \leftarrow holds(g_1, T), \ldots, holds(g_m, T) \tag{18}$$

ISBN: 978-989-95079-6-8

Example 2 *The normal logic program translation, $\Pi_{\mathbf{FP}}$, of the fuzzy planning problem $\mathbf{FP} = \langle S_0, \mathcal{AD}, \mathcal{G}, \mathcal{T} \rangle$, described in Example 1 proceeds as follows, where S_0 and $\mathcal{AD}$ are as presented in Example 1, $\mathcal{G} = \{hb\}$, and $\mathcal{T} \in [0,1]$. In addition to the rules (5), (6), (7), (8), (14), (15), (16), and (17), $\Pi_{\mathbf{FP}}$ contains the rules*

$$\begin{aligned} action(pickup_1) &\leftarrow \\ action(pickup_2) &\leftarrow \\ action(pickup_3) &\leftarrow \\ action(pickup_4) &\leftarrow \end{aligned}$$

where pickup is in $\mathcal{A}$. Properties of the world are described by the atoms gd (gripper dry) and hb (holding block), which are encoded in $\Pi_{\mathbf{FP}}$ by the rules

$$\begin{aligned} atom(gd) &\leftarrow \\ atom(hb) &\leftarrow \end{aligned}$$

Executability conditions of action pickup are encoded in $\Pi_{\mathbf{FP}}$ by the rules

$$\begin{aligned} exec(pickup_1, t) &\leftarrow \\ exec(pickup_2, t) &\leftarrow \\ exec(pickup_3, t) &\leftarrow \\ exec(pickup_4, t) &\leftarrow \end{aligned}$$

where $0 \leq t \leq n$. The set of possible initial states are encoded in $\Pi_{\mathbf{FP}}$ by the rules:

$$\begin{aligned} holds(\neg hb, 0) &\leftarrow \\ holds(gd, 0) &\leftarrow not\ holds(\neg gd, 0) \\ holds(\neg gd, 0) &\leftarrow not\ holds(gd, 0) \end{aligned}$$

Effects of the pickup action are encoded in $\Pi_{\mathbf{FP}}$ by the rules

$$\begin{aligned} holds(hb, T+1) &\leftarrow occ(pickup_1, T), \\ &\quad exec(pickup_1, T), holds(gd, T) \\ holds(\neg hb, T+1) &\leftarrow occ(pickup_2, T), \\ &\quad exec(pickup_2, T), holds(gd, T) \\ holds(hb, T+1) &\leftarrow occ(pickup_3, T), \\ &\quad exec(pickup_3, T), holds(\neg gd, T) \\ holds(\neg hb, T+1) &\leftarrow occ(pickup_4, T), \\ &\quad exec(pickup_4, T), holds(\neg gd, T) \end{aligned}$$

The goal is encoded in $\Pi_{\mathbf{FP}}$ by the rule

$$goal \leftarrow holds(hb, T)$$

5 Correctness

In this section we prove the correctness of our translation. We show that the answer sets of the normal logic program translation, $\Pi_{\mathbf{FP}}$, of a fuzzy planning problem, $\mathbf{FP}$, correspond to trajectories in $\mathbf{FP}$. Consider that the domain of T is $\{0, \ldots, n\}$. Assume that Φ is the fuzzy transition function associated with $\mathbf{FP}$, s_0 is a possible initial state, and $a_0, \ldots, a_{n-1}$ be a set of actions in $\mathcal{A}$. Any action a_i can be represented as a set where $a_i = \{a_{1_i}, \ldots, a_{m_i}\}$. Therefore, a trajectory in $\mathbf{FP}$ is $s_0\ a_{j_0}\ s_1 \ldots a_{j_{n-1}}\ s_n$ for $(1 \leq j \leq m)$ and $(0 \leq i \leq n)$, such that $\forall (0 \leq i \leq n)$, s_i is a state, a_i is an action, $a_{j_i} \in a_i = \{a_{1_i}, \ldots, a_{m_i}\}$, and $s_i = \Phi(a_{j_{i-1}}, s_{i-1})$.

Theorem 1 *Let $\mathbf{FP} = \langle S_0, \mathcal{AD}, \mathcal{G}, \mathcal{T} \rangle$ be a fuzzy planning problem, P be a fuzzy plan in $\mathbf{FP}$, and T_P be the set of all trajectories in P. Then, $s_0\ a_{j_0}\ s_1 \ldots a_{j_{n-1}}\ s_n$ is a trajectory in T_P iff $occ(a_{j_0}, 0), \ldots, occ(a_{j_{n-1}}, n-1)$ is true in an answer set of $\Pi_{\mathbf{FP}}$.*

Theorem 1 presents that any fuzzy planning problem, $\mathbf{FP}$, can be translated into a normal logic program with answer set semantics, $\Pi_{\mathbf{FP}}$, such that a trajectory in $\mathbf{FP}$ is equivalent to an answer set of $\Pi_{\mathbf{FP}}$. Theorem 1 shows that normal logic programs with answer set semantics can be used to find fuzzy plans for fuzzy planning problems in two steps. The first step is to translate a fuzzy planning problem, $\mathbf{FP}$, into a normal logic program whose answer sets correspond to valid trajectories in $\mathbf{FP}$. From the answer sets of the normal logic program translation of $\mathbf{FP}$, the set of trajectories T_P that correspond to a fuzzy plan P in $\mathbf{FP}$ is determined. The second step is to calculate the grade membership of the fuzzy plan P using the formula

$$\max_{s_0\ a_{j_0}\ s_1 \ldots a_{j_{n-1}}\ s_n \in T_P} \left(\min_{0 \leq i \leq n-1} (\mu_{\mathcal{S}}(s_i), \mu_{\mathcal{S}}(s_{i+1})) \right)$$

Furthermore, we show that any fuzzy planning problem can be encoded as a SAT formula. Hence, state-of-the-art SAT solvers can be used to find fuzzy plans for fuzzy planning problems. Any normal logic program, Π, can be translated into a SAT formula, $\mathcal{S}$, where the models of $\mathcal{S}$ are equivalent to the answer sets of Π [19]. Therefore, the normal logic program translation of a fuzzy planning problem $\mathbf{FP}$ can be encoded into an equivalent SAT formula, where the models of $\mathcal{S}$ correspond to valid trajectories in $\mathbf{FP}$.

Theorem 2 *Let $\mathbf{FP}$ be a fuzzy planning problem and $\Pi_{\mathbf{FP}}$ be the normal logic program encoding of $\mathbf{FP}$. Then, the models of the SAT encoding of $\Pi_{\mathbf{FP}}$ are equivalent to valid trajectories in $\mathbf{FP}$.*

6 Conclusions and Related Work

We described a novel action language called fuzzy action language, $\mathcal{A}_F$, that allows the representation and reasoning about actions with fuzzy effects. In addition we introduced the notions of fuzzy planning and fuzzy plan in the fuzzy action language $\mathcal{A}_F$. Furthermore, we described a fuzzy planner based on the fuzzy action language A_F that is developed by translating a fuzzy planning problem, $\mathbf{FP}$, in $\mathcal{A}_F$ into a normal logic program with answer set semantics, $\Pi_{\mathbf{FP}}$, where trajectories in $\mathbf{FP}$ are equivalent to the answer sets of $\Pi_{\mathbf{FP}}$. In addition, we formally proved the correctness of our planner. Furthermore, we showed that a fuzzy planning problem can be encoded as a SAT problem.

The literature is rich with action languages that are capable of representing and reasoning about actions in the presence of probabilistic uncertainty, which include [2, 3, 5, 12, 14, 21, 22]. In [22], a probabilistic action language $\mathcal{P}$ is described that allows the representation of imperfect sensing actions (with probabilistic outcomes), non-sensing actions with probabilistic effects, the initial probability distribution over the possible initial states, and the indirect effects of actions. The action language $\mathcal{E}^+$ [12] allows sensing actions under the assumption that the agent's sensors are perfect, actions with probabilistic effects, and actions with non-deterministic effects. In

ISBN: 978-989-95079-6-8

addition, the semantics of $\mathcal{E}^+$ is based on description logic. Other high level probabilistic action description languages are described in [2, 5]. These languages are similar to $\mathcal{E}^+$ in the sense that they represent and reason about actions with probabilistic effects, except that they do not allow actions with non-deterministic effects or sensing actions. In [21], a high level action language called $\mathcal{A}_{MD}$ is presented that allows the factored representation and reasoning about Markov Decision Processes for reinforcement learning. In addition to the fact that $\mathcal{A}_F$ is a high level language, the major difference between $\mathcal{A}_F$ and these languages is that $\mathcal{A}_F$ represent and reason about actions under the presence of fuzzy uncertainty or imprecision. This is achieved in $\mathcal{A}_F$ by allowing description of actions with fuzzy effects, the executability conditions of actions, and the initial grade membership distribution of the initial states.

References

[1] C. Baral. *Knowledge representation, reasoning, and declarative problem solving*. Cambridge University Press, 2003.

[2] C. Baral, N. Tran, L. C. Tuan. *Reasoning about actions in a probabilistic setting*. In *AAAI*, 2002.

[3] C. Boutilier, T. Dean, and S. Hanks. Decision-theoretic planning: structural assumptions and computational leverage. *Journal of AI Research*, 11:1–94, 1999.

[4] C. da Costa Pereira and A. Tettamanzi. Reasoning about actions with imprecise and incomplete state descriptions. *Fuzzy Sets and Systems*, 160(10):13831401, 2009.

[5] T. Eiter and T. Lukasiewicz. Probabilistic reasoning about actions in nonmonotonic causal theories. In *19th Conference on Uncertainty in Artificial Intelligence*, 2003.

[6] T. Eiter et al. Declarative problem solving in dlv. In *Logic Based Artificial Intelligence*, 2000.

[7] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. *ICSLP*, 1988, MIT Press.

[8] M. Gelfond and V. Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9(3-4):363-385, 1991.

[9] M. Gelfond and V. Lifschitz Representing action and change by logic programs. *Journal of Logic Programming*, 17:301–321, 1993.

[10] M. Gelfond and V. Lifschitz. Action languages. In *Electronic Transactions on AI*, 3(16), 1998.

[11] E. Giunchiglia and V. Lifschitz. An action language based on causal explanation: preliminary report. In *AAAI*, 1998.

[12] L. Iocchi, T. Lukasiewicz, D. Nardi, and R. Rosati. Reasoning about actions with sensing under qualitative and probabilistic uncertainty. In *16th European Conference on Artificial Intelligence*, 2004.

[13] H. Kautz and B. Selman. Pushing the envelope: planning, propositional logic, and stochastic search. In *13th National Conference on Artificial Intelligence*, 1996.

[14] N. Kushmerick, S. Hanks, and D. Weld. An algorithm for probabilistic planning. *Artificial Intelligence*, 76(1-2):239-286, 1995.

[15] J. Lee and V. Lifschitz. Describing additive fluents in action language C+. In *International Joint Conference in AI*, 2003.

[16] V. Lifschitz. Answer set planning. In *ICLP*, 1999.

[17] V. Lifschitz. Two components of an action language. *Annals of Mathematics and Artificial Intelligence*, 21:305–320, 1997.

[18] V. Lifschitz and W. Ren. A modular action description language. In *AAAI*, 2006.

[19] F. Lin and Y. Zhao. ASSAT: Computing answer sets of a logic program by SAT solvers. *Artificial Intelligence*, 157(1-2):115-137, 2004.

[20] I. Niemela and P. Simons. Efficient implementation of the well-founded and stable model semantics. *In Joint International Conference and Symposium on Logic Programming*, 289-303, 1996.

[21] E. Saad. A logical framework to reinforcement learning using hybrid probabilistic logic programs. In *Second International Conference on Scalable Uncertainty Management*, 2008.

[22] E. Saad. Probabilistic planning with imperfect sensing actions using hybrid probabilistic logic programs. In *Ninth International Workshop on Computational Logic in Multi-Agent Systems (CLIMA'08)*, 2008.

[23] A. Saffiotti, K. Konolige, and E. Ruspini. A Multivalued logic approach to integrating planning and control. *Artificial Intelligence*, 76(1-2): 481–526, 1995.

[24] T. Son and C. Baral. Formalizing sensing actions - a transition function based approach. *Artificial Intelligence*, 125:(1-2): 19 - 91, 2001.

[25] T. Son, C. Baral, T. Nam, and S. McIlraith. Domain-dependent knowledge in answer set planning. *ACM Transactions on Computational Logic*, 7(4):613–657, 2006.

[26] V. S. Subrahmanian and C. Zaniolo. Relating stable models and AI planning domains. *In International Conference of Logic Programming*, 233-247, 1995.

[27] L. Zadeh. Fuzzy Sets. *Information and Control*, 8(3): 338 –353, 1965.

[28] L. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Systems, Man, and Cybernetics*, SMC-3:28–44, 1973.

ISBN: 978-989-95079-6-8

Interval-valued restricted equivalence functions applied on Clustering Techniques

Aranzazu Jurio[1] Miguel Pagola[1] Daniel Paternain[1] Carlos Lopez-Molina[1] Pedro Melo-Pinto[2]

1.Departamento de Automática y Computación, Universidad Pública de Navarra
Pamplona, Spain
2.Department of Engineering, CITAB, Universidade de Trás-os-Montes e Alto Douro
Vila Real, Portugal
Email: aranzazu.jurio@unavarra.es, miguel.pagola@unavarra.es, pmelo@utad.pt

Abstract— *In this work we use interval-valued fuzzy sets in the Fuzzy C-Means algorithm for image segmentation. We introduce interval-valued restricted equivalence functions as a way of measuring the equivalence between the intervals associated to different pixels. We propose two construction methods for those new functions. We also prove experimentally that with these new concepts, the algorithm provides better results in ultrasound image segmentation that than those not using intervals.*

Keywords— Decomposability, Fuzzy C-Means, Interval-Valued Fuzzy Sets, Restricted Equivalence Functions, Ultrasound image.

1 Introduction

Ultrasound images are used to evaluate breast anomalies during the diagnosis in breast exams. When any breast anomaly is studied to determine whether it is benign or not, the radiologist studies its size, area and other morphological characteristics [12]. In this sense, the development of algorithms that automatically separates the lesion and the rest of the image is very important for a quick and correct diagnosis.

Ultrasound images segmentation is very limited by its bad quality. Speckle noise, shades and attenuation make the segmentation problem a very difficult topic [10]. There exist some methods used for this purpose; a comprehensive survey can be found in [9].

A very common method for segmenting images is clustering. The purpose of this work is to adapt the well known algorithm Fuzzy C-Means (FCM) [1, 2] to ultrasound image segmentation. This algorithm tries to classify a set of objects, in this case image pixels, in a determined number of clusters. That number must be chosen a priori. The main idea is to find as many centers as clusters must be, trying to minimize the distance between each pixel and the cluster centroids.

Due to the complexity of ultrasound images, there exists a lot of uncertainty within them. This uncertainty is also present in the images obtained after processing the ultrasound images. We are going to use interval-valued fuzzy sets [13] in the FCM algorithm to deal with this uncertainty.

As already mentioned, our proposed algorithm is based on the FCM. The new contributions we propose are: the use of interval-valued fuzzy sets and the search of the greatest interval equivalence between the image pixels and the centroids of the clusters. To find the greatest equivalence between intervals we introduce the concept of interval-valued restricted equivalence function.

This paper is organized in the following way: first of all we present some preliminary concepts. In Section 3 we define the concept of interval-valued restricted equivalence functions. In Section 4, we explain the way to apply these functions to ultrasound images segmentation and we show some experimental results. Finally some conclusions and future researchs are shown.

2 Preliminaries

We know that in fuzzy set theory a function $n : [0,1] \to [0,1]$ with $n(0) = 1$, $n(1) = 0$ that is strictly decreasing and continuous is called strict negation. If, in addition, n is involutive, then it is said that it is a strong negation.

We denote by $L([0,1])$ the set of all closed subintervals of the closed interval $[0,1]$; that is,

$$L([0,1]) = \{\mathbf{x} = [\underline{x}, \overline{x}] \mid (\underline{x}, \overline{x}) \in [0,1]^2 \text{ and } \underline{x} \leq \overline{x}\}$$

$L([0,1])$ is a partially ordered set with respect to the relation $\leq_L$ defined in the following way: given $\mathbf{x}, \mathbf{y} \in L([0,1])$,

$$\mathbf{x} \leq_L \mathbf{y} \text{ if and only if } \underline{x} \leq \underline{y} \text{ and } \overline{x} \leq \overline{y};$$

$$\mathbf{x} =_L \mathbf{y} \text{ if and only if } \underline{x} = \underline{y} \text{ and } \overline{x} = \overline{y};$$

The relation above is transitive, antisymmetric and it expresses the fact that $\mathbf{x}$ strongly links to $\mathbf{y}$, so that $(L([0,1]), \leq_L)$ is a complete lattice, where the smallest element is $0_L = [0,0]$, and the largest is $1_L = [1,1]$.

Definition 1 *An interval-valued fuzzy set A on the universe $U \neq \emptyset$ is a mapping $A : U \to L([0,1])$.*

Definition 2 *An IV negation is a function $N : L([0,1]) \to L([0,1])$ that is decreasing (with respect to $\leq_L$) and such that $N(1_L) = 0_L$ and $N(0_L) = 1_L$. If for all $\boldsymbol{x} \in L([0,1])$, $N(N(\boldsymbol{x})) = \boldsymbol{x}$, it is said that N is involutive.*

Theorem 1 *[7, 3] A function $N : L([0,1]) \to L([0,1])$ is an involutive IV negation if and only if there exists an involutive negation n such that*

$$N(\boldsymbol{x}) = [n(\overline{x}), n(\underline{x})].$$

Definition 3 *[4] A function $REF : [0,1]^2 \to [0,1]$ is called a restricted equivalence function, if it satisfies the following conditions:*

(1) $REF(x,y) = REF(y,x)$ for all $x, y \in [0,1]$;

(2) $REF(x,y) = 1$ if and only if $x = y$;

ISBN: 978-989-95079-6-8

(3) $REF(x,y) = 0$ *if and only if* $x = 1$ *and* $y = 0$ *or* $x = 0$ *and* $y = 1$*;*

(4) $REF(x,y) = REF(n(x), n(y))$ *for all* $x, y \in [0,1]$*,* n *being a strong negation;*

(5) For all $x, y, z \in [0,1]$*, if* $x \leq y \leq z$*, then* $REF(x,y) \geq REF(x,z)$ *and* $REF(y,z) \geq REF(x,z)$*.*

Definition 4 *[8] A function* $F : L([0,1])^2 \to L([0,1])$ *is said decomposable if there exist two functions* $F_1, F_2 : [0,1]^2 \to [0,1]$ *such that for all* $[\underline{x}, \overline{x}], [\underline{y}, \overline{y}] \in L([0,1])$ *they satisfy that*

$$F([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = [F_1(\underline{x}, \underline{y}), F_2(\overline{x}, \overline{y})]$$

Definition 5 *[6] A binary aggregation function is defined as a function* $Ag : [0,1]^2 \to [0,1]$ *such that:*

(1) $Ag(x_1, x_2) \leq Ag(y_1, y_2)$ *whenever* $x_1 \leq y_1$ *and* $x_2 \leq y_2$*;*

(2) $Ag(0,0) = 0$ *and* $Ag(1,1) = 1$*.*

This is an idempotent aggregation function if

(3) $Ag(x,x) = x$ *for all* $x \in [0,1]$

3 Interval Valued Restricted Equivalence Functions

Definition 6 *An Interval Valued Restricted Equivalence Function (*REF_{IV}*) is a function*

$$REF_{IV} : L([0,1]) \times L([0,1]) \to L([0,1])$$

such that:

(1) $REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) = REF_{IV}(\boldsymbol{y}, \boldsymbol{x})$ *for all* $\boldsymbol{x}, \boldsymbol{y} \in L([0,1])$*;*

(2) $REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) = 1_L$ *if and only if* $\boldsymbol{x} = \boldsymbol{y}$*;*

(3) $REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) = 0_L$ *if and only if* $\boldsymbol{x} = 1_L$ *and* $\boldsymbol{y} = 0_L$ *or* $\boldsymbol{x} = 0_L$ *and* $\boldsymbol{y} = 1_L$*;*

(4) $REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) = REF_{IV}(N(\boldsymbol{x}), N(\boldsymbol{y}))$ *being N an involutive IV negation;*

(5) For all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in L([0,1])$*, if* $\boldsymbol{x} \leq_L \boldsymbol{y} \leq_L \boldsymbol{z}$*, then* $REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) \geq_L REF_{IV}(\boldsymbol{x}, \boldsymbol{z})$ *and* $REF_{IV}(\boldsymbol{y}, \boldsymbol{z}) \geq_L REF_{IV}(\boldsymbol{x}, \boldsymbol{z})$*.*

Theorem 2 *There is not any decomposable* REF_{IV}*.*

Proof. We suppose there exists a decomposable REF_{IV}. This means that there exist two functions $F_1, F_2 : [0,1] \times [0,1] \to [0,1]$ such that:

$$REF_{IV}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}] = [F_1(\underline{x}, \underline{y}), F_2(\overline{x}, \overline{y})]$$

in such form that REF_{IV} fulfills the five properties demanded in Definition 6. It means:

(1) As $REF_{IV}(\mathbf{x}, \mathbf{y}) = REF_{IV}(\mathbf{y}, \mathbf{x})$, then F_1 and F_2 must satisfy that $F_1(a,b) = F_1(b,a)$ and $F_2(a,b) = F_2(b,a)$ for all $a, b \in [0,1]$.

(2) As $REF_{IV}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = 1_L$ if and only if $\underline{x} = \underline{y}$ and $\overline{x} = \overline{y}$, then $F_1(\underline{x}, \underline{x}) = 1$ and $F_2(\overline{x}, \overline{x}) = 1$, so for all $x \in [0,1]$ $F_1(x,x) = F_2(x,x) = 1$.

(3) As $REF_{IV}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = 0_L$ if and only if $\mathbf{x} = 0$ and $\mathbf{y} = 1$ or vice versa, then $F_1(0,1) = 0$ and $F_2(0,1) = 0$.

(4) As $REF_{IV}([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = REF_{IV}([n(\overline{x}), n(\underline{x})], [n(\overline{y}), n(\underline{y})]$ then $[F_1(\underline{x}, \underline{y}), F_2(\overline{x}, \overline{y})] = [F_1(n(\overline{x}), n(\overline{y})), F_2(n(\underline{x}), n(\underline{y}))]$. Separating each term, $F_1(\underline{x}, \underline{y}) = F_1(n(\overline{x}), n(\overline{y}))$ and $F_2(\overline{x}, \overline{y}) = F_2(n(\underline{x}), n(\underline{y}))$. In particular, if $\underline{x} = 1$ and $\underline{y} = 0$, and $\overline{x} = \overline{y} = 1$ then $F_2(\overline{x}, \overline{y}) = F_2(1,1) = 1$ and $F_2(n(\underline{x}), n(\underline{y})) = F_2(0,1) = 0$. $1 \neq 0$, so there is not any decomposable REF_{IV}.

Theorem 3 *Let* REF *be a restricted equivalence function, and* Ag_1 *and* Ag_2 *be two symmetric, idempotent aggregation functions such that:*

(i) $Ag_1 \leq Ag_2$*;*

(ii) $Ag_1(a,b) = 1$ *if and only if* $a = b = 1$*;*

(iii) $Ag_2(a,b) = 0$ *if and only if* $a = b = 0$*.*

Under these conditions, the function

$$REF_{IV} : L([0,1]) \times L([0,1]) \to L([0,1])$$

given by

$$REF_{IV}(\boldsymbol{x}, \boldsymbol{y}) = \begin{array}{l}[Ag_1(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})), \\ Ag_2(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y}))]\end{array}$$

is an interval-valued restricted equivalence function in the sense of Definition 6.

Proof. By hypothesis, $Ag_1 \leq Ag_2$, so the REF_{IV} is a well defined interval.

(1) REF, Ag_1 and Ag_2 are commutative functions, so REF_{IV} is a commutative function too.

(2) (Sufficiency) If $REF_{IV}(\mathbf{x}, \mathbf{y}) = 1_L$, then $Ag_1(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})) = 1$. So, by (ii), $REF(\underline{x}, \underline{y}) = REF(\overline{x}, \overline{y}) = 1$. Any REF is 1 if and only if both arguments are equal, so $\underline{x} = \underline{y}$ and $\overline{x} = \overline{y}$. This means $\mathbf{x} = \mathbf{y}$.

(Necessity) If $\mathbf{x} = \mathbf{y}$, it means $\underline{x} = \underline{y}$ and $\overline{x} = \overline{y}$, then $REF_{IV}(\mathbf{x}, \mathbf{x}) = \begin{array}{l}[Ag_1(REF(\underline{x}, \underline{x}), REF(\overline{x}, \overline{x})), \\ Ag_2(REF(\underline{x}, \underline{x}), REF(\overline{x}, \overline{x}))]\end{array}$ $[Ag_1(1,1), Ag_2(1,1)]$ and by the property (2) in Definition 5 this is $[1,1]$.

(3) (Sufficiency) If $REF_{IV}(\mathbf{x}, \mathbf{y}) = 0_L$, then $Ag_2(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})) = 0$, therefore, by (iii), $REF(\underline{x}, \underline{y}) = REF(\overline{x}, \overline{y}) = 0$. Any REF is 0 if and only if one of the arguments is 0 and the other is 1, so $\underline{x} = 1$ and $\underline{y} = 0$ or vice versa, and $\overline{x} = 1$ and $\overline{y} = 0$ or vice versa. This means $\underline{x} = \overline{x} = 1$ and $\underline{y} = \overline{y} = 0$ or vice versa.

(Necessity) If $\mathbf{x} = 1_L$ and $\mathbf{y} = 0_L$, then $REF_{IV}(\mathbf{x}, \mathbf{y}) = [Ag_1(REF(1,0), REF(1,0)), Ag_2(REF(1,0), REF(1,0))] = [Ag_1(0,0), Ag_2(0,0)]$, which is equal to $[0,0]$ by property (2) in Definition 5.

ISBN: 978-989-95079-6-8

(4) $REF_{IV}(N(\mathbf{x}), N(\mathbf{y})) = [Ag_1(REF(n(\overline{x}), n(\overline{y})), REF(n(\underline{x}), n(\underline{y}))), Ag_2(REF(n(\overline{x}), n(\overline{y})), REF(n(\underline{x}), n(\underline{y})))]$. As $REF(\underline{x}, \underline{y}) = REF(n(\underline{x}), n(\underline{y}))$ and $REF(\overline{x}, \overline{y}) = REF(n(\overline{x}), n(\overline{y}))$, then $REF_{IV}(N(\mathbf{x}), N(\mathbf{y})) = [Ag_1(REF(\overline{x}, \overline{y}), REF(\underline{x}, \underline{y})), Ag_2(REF(\overline{x}, \overline{y}), REF(\underline{x}, \underline{y}))]$. As Ag_1 and Ag_2 are commutative functions, this expression is equal to $[Ag_1(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})), Ag_2(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y}))] = REF_{IV}(\mathbf{x}, \mathbf{y})$.

(5) If $\mathbf{x} \leq_L \mathbf{y} \leq_L \mathbf{z}$, then $\underline{x} \leq \underline{y} \leq \underline{z}$ and $\overline{x} \leq \overline{y} \leq \overline{z}$. By (5) in Definition 3, $REF(\underline{x}, \underline{y}) \geq REF(\underline{x}, \underline{z})$ and $REF(\overline{x}, \overline{y}) \geq REF(\overline{x}, \overline{z})$. By (1) in Definition 5 $Ag_1(REF(\underline{x}, \underline{z}), REF(\overline{x}, \overline{z})) \leq Ag_1(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y}))$ and $Ag_2(REF(\underline{x}, \underline{z}), REF(\overline{x}, \overline{z})) \leq Ag_2(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y}))$, so $REF_{IV}(\mathbf{x}, \mathbf{z}) \leq REF_{IV}(\mathbf{x}, \mathbf{y})$. The reasoning for $REF_{IV}(\mathbf{x}, \mathbf{z}) \leq REF_{IV}(\mathbf{y}, \mathbf{z})$ is analogous.

Corollary 1 *Given a restricted equivalence function REF, and given T and S any t-norm and any t-conorm,*

$$REF_{IV}(\boldsymbol{x},\boldsymbol{y}) = [T(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y})), S(REF(\underline{x}, \underline{y}), REF(\overline{x}, \overline{y}))]$$

is an interval-valued restricted equivalence function.

Proof. We need to see that any t-norm fulfills the properties demanded to Ag_1 and any t-conorm fulfills the properties demanded to Ag_2.

For the t-norm T it is necessary that $T(a, b) = 1$ if and only if $a = b = 1$. (Sufficiency) If the t-norm is the minimum, if $min(a, b) = 1$, then $a = b = 1$. As the minimum is the greatest of all t-norms, to be 1 the arguments of any t-norm must be greater than or equal to 1. As they are defined in the unit interval, both arguments must be 1.

(Necessity) We know that $T(1, x) = x$ for any $x \in [0, 1]$. If $x = y = 1$, then $T(1, 1) = 1$.

For the t-conorm S it is necessary that $S(a, b) = 0$ if and only if $a = b = 0$. (Sufficiency) If the t-conorm is the maximum, if $max(a, b) = 0$, then $a = b = 0$. As the maximum is the smallest of all t-conorms, to be 0 the arguments of any t-conorm must be smaller than or equal to 0. As they are defined in the unit interval, both arguments must be 0.

(Necessity) We know that $S(x, 0) = x$ for any $x \in [0, 1]$. If $x = y = 0$, then $S(0, 0) = 0$.

Example 1 *If we choose $T = min$, $S = max$ and $REF(x, y) = 1 - |x - y|$, we have that:*

$$REF_{IV}(\boldsymbol{x},\boldsymbol{y}) = [min(1 - |\underline{x} - \underline{y}|, 1 - |\overline{x} - \overline{y}|), max(1 - |\underline{x} - \underline{y}|, 1 - |\overline{x} - \overline{y}|)] =$$

$$[1 - max(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|), 1 - min(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|)] =$$

$$N([min(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|), max(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|)].$$

4 Application of $\mathbf{REF_{IV}}$ to clustering algorithm

The algorithm proposed in this paper is a modification of the Fuzzy C-Means (FCM) algorithm [1, 2], which aims to find the most characteristic point of each cluster, considered its centroid. Such a way, the membership degree of every object to each cluster is achieved via minimizing the target function:

$$J_m(U, V) = \sum_{j=1}^{n} \sum_{i=1}^{K} w_{ij}^m \parallel x_j - C_i \parallel^2$$

In the new algorithm we propose, the function to minimize is quite similar. The only difference is the way of calculating the distance between each pixel and the cluster centroids. Instead of using the euclidean distance, we use a negation of an interval restricted equivalence function. Therefore, our objective function to minimize is:

$$J_m(U, V) = \sum_{j=1}^{n} \sum_{i=1}^{K} w_{ij}^m (1 - Ag(REF_{IV}(\mathbf{x_j}, \mathbf{C_i})))$$

The main purpose of the new algorithm is to segment the image into K areas represented by K interval centroids, maximizing the equivalence in each area, using the interval valued restricted equivalence functions, that we have previously defined. As the algorithm is partially supervised, the number of areas must be selected by the user before the execution. Notice that the REF_{IV} is an interval. As we want to maximize the equivalence, we are going to use an aggregation of the bounds of the equivalence interval.

The algorithm starts with the initialization of all the required parameters. First of all it is necessary to build the interval-valued fuzzy set associated with the image. For that, each pixel is assigned an interval depending on the values of its neighborhood. The rest of the parameters to be initialized are the number of areas to be found, the fuzzification degree, the weight matrix, the maximum number of iterations and the finishing threshold.

After the initialization, the algorithm runs in a loop. In this loop the centroid of each cluster and the weight matrix are calculated until the improvement is not significant. To update that weight matrix, we use the interval-valued restricted equivalence functions.

When the loop is finished, the algorithm labels each pixel to show the results. For this purpose, each pixel is associated to the cluster with the biggest weight.

4.1 Algorithm structure

1. Given an image X with n pixels, build its interval-valued fuzzy set $A = \{[\underline{x_i}, \overline{x_i}] | x_i \in X\}$ with $i = 1 \ldots n$.
2. Initialize
   - 2.1 K = number of areas to be segmented.
   - 2.2 T = maximum number of iterations.
   - 2.3 m = fuzzification degree, $1 \leq m \leq \infty$.
   - 2.4 ε = finishing threshold, $\varepsilon > 0$.
   - 2.5 W = weight matrix, $0 \leq w_{ij} \leq 1$ with $i = 1 \ldots n$ and $j = 1 \ldots K$.

ISBN: 978-989-95079-6-8

2.6 t = current iteration = 0.

3. REPEAT

3.1 Increase current iteration

3.2 Calculate the centroids

$$C_j = \frac{\sum_{i=1}^{n} w_{ij}^m [\underline{x_i}, \overline{x_i}]}{\sum_{i=1}^{n} w_{ij}^m}$$

3.3 Update the weight matrix

$$w_{ij} = \left(\sum_{k=1}^{K} \left(\frac{1 - (Ag(REF_{IV}(\mathbf{x_i}, \mathbf{C_j})))}{1 - (Ag(REF_{IV}(\mathbf{x_i}, \mathbf{C_k})))} \right)^{\frac{1}{m-1}} \right)^{-1}$$

where $\mathbf{x_i} = [\underline{x_i}, \overline{x_i}]$, $\mathbf{C_j} = [\underline{C_j}, \overline{C_j}]$, $\mathbf{C_k} = [\underline{C_k}, \overline{C_k}]$ and $Ag(REF_{IV}(\mathbf{x}, \mathbf{y})) = \frac{REF_{IV}(\mathbf{x},\mathbf{y}) + \overline{REF_{IV}(\mathbf{x},\mathbf{y})}}{2}$.

3.4 Calculate the error.

UNTIL $t = T$ or $error \leq \varepsilon$.

4. Label the image.

4.2 Preprocessing

To prove the performance of the new algorithm, we are going to segment some ultrasound images. In this case we work with breast ultrasound images, where the goal is to classify every pixel of the image. It means, decide whether they belong to the lesion or not.

Due to the characteristics of these images, which may contain several similar regions, it is necessary to select the central pixel of the desired zone. Since the user has to contribute with some information, we say that the proposed system is partially supervised.

To improve the results the images are preprocessed. In order to do so, we create two new images for every ultrasound image. The first one is a false edges image while the second one is a enhanced image. An example of these new images can be seen in Fig. 1.

Figure 1: False edges and enhanced images.

4.2.1 Enhanced image

To create the enhanced image we use the algorithm proposed by Sahba et al. [11] for ultrasound images. The main idea is to remove the image noise and to enhance the gray level in the selected area. First of all the noise is removed using the median filter (7x7 or 9x9). After that, every pixel is fuzzified depending on its gray level using a membership function considering the average gray level in the pixel neighborhood and the position of the central pixel selected.

4.2.2 False Edges image

The false edges image represents to what degree each pixel is part of an edge. In [5] a method is presented to extract false edges from t-norms and t-conorms. For each pixel, the first step is to build a submatrix (3x3, 5x5, etc.). Applying any t-norm to the elements of that submatrix, the lower bound of an interval is obtained. Analogously, the upper bound is got applying any t-conorm to the same submatrix. The width of that interval, it means, the upper bound minus the lower bound, is called false edge. If the t-norm and the t-conorm chosen are the minimum and maximum respectively, the false edge is basically the substraction of the biggest element in the submatrix minus the lowest one.

4.3 Experiment

The experiment we show in this section is based on 8 ultrasound images, shown in Fig. 2. They are breast images in which there exist some zones that can be tumors. To create the interval-valued fuzzy set associated to the image, we use a net of 5x5 neighbors for every pixel. In this sense, each pixel is corresponded with an interval value that depends not only on its value but also on its neighbors' values. To construct the lower bound we take the minimum of the values of the mesh, and to calculate the upper bound we take the greatest one.

The REF_{IV} we use is the one shown in example 1:

$$\begin{aligned} REF_{IV}(\mathbf{x}, \mathbf{y}) = \quad & [min(1 - |\underline{x} - \underline{y}|, 1 - |\overline{x} - \overline{y}|), \\ & max(1 - |\underline{x} - \underline{y}|, 1 - |\overline{x} - \overline{y}|)] = \\ N([min(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|), & max(|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|)]. \end{aligned}$$

To calculate the performance of the segmentation method we propose, we use the areas overlapping measure (S_A) between the image we have obtained and the ideal one. The latter is created by an expert radiologist.

$$S_A = \frac{|Ideal \ \& \ Obtained|}{|Ideal \ + \ Obtained|} \tag{1}$$

where $Ideal$ is the binary image segmented by the expert radiologist and $Obtained$ is the binary image got by the studied method.

We are going to compare the results from our algorithm with the ones obtained with the FCM algorithm by comparing their accuracy with respect to the ideal images. In Fig. 2 some results are illustrated. The first column matches the original image, while the second and the third columns match the segmentation using the FCM algorithm and the proposed one respectively. Finally, the fourth column shows the ideal segmentation.

As it can be seen, the results from our algorithm are more similar to the ideal ones than the ones got by the FCM. This improvement can be numerically observed by the areas overlapping percentage between each image and its corresponding ideal one (see table 1, where A, B, ...H refer to the images shown in Fig. 2).

With these results, our proposed algorithm improves the results obtained by the FCM in 8%. This percentage is the difference between the average overlap of the ideal image with the one obtained by the simple FCM and with the one obtained by our method in the images of Fig. 2. We can say this is a very good result, as it means a significant progress towards the correct segmentation of this kind of images.

ISBN: 978-989-95079-6-8

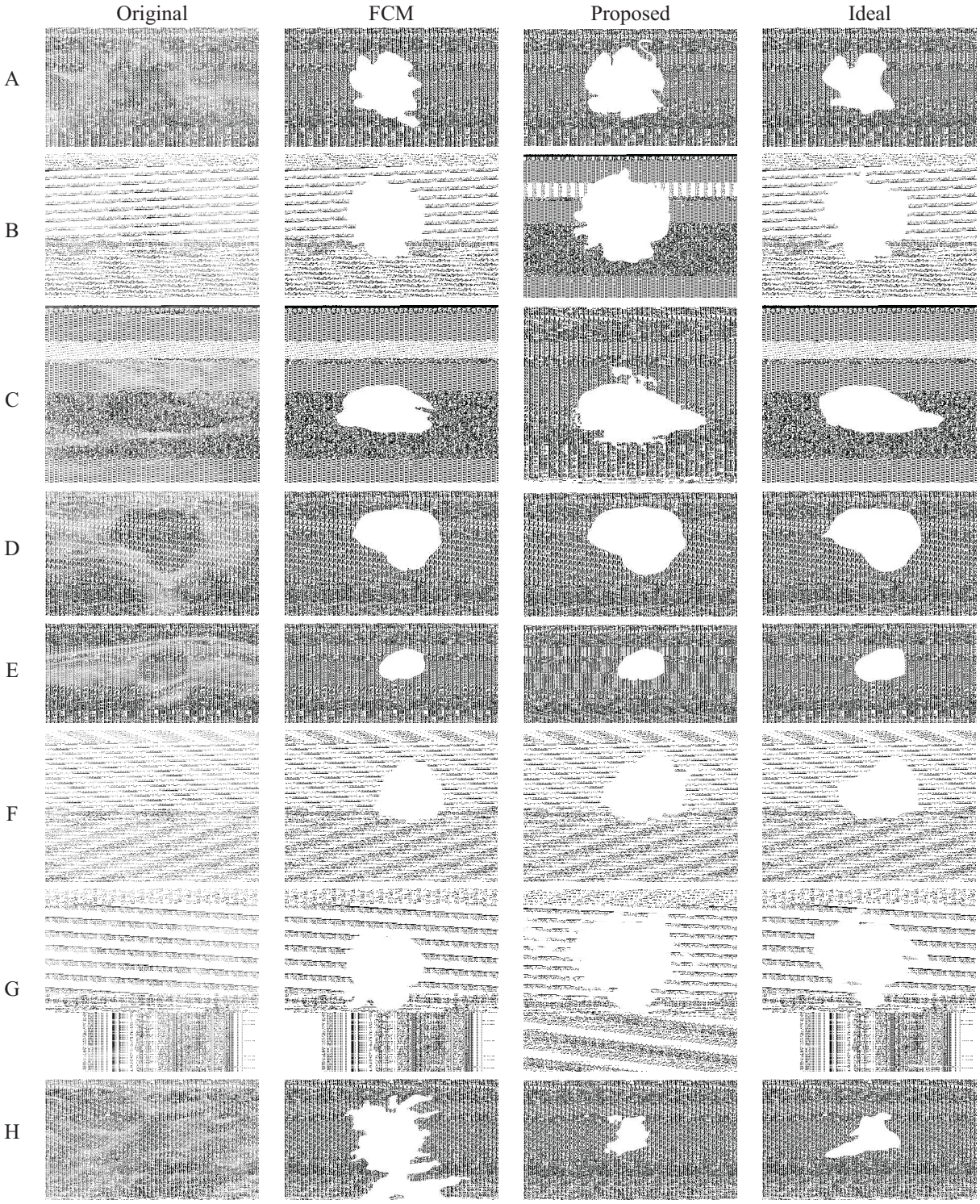


Figure 2: Obtained results.

Table 1: Analytic solutions.

	A	B	C	D
FCM	75.17	76.73	73.28	84.48
Proposed	70.81	89.43	76.37	91.47
	E	F	G	H
FCM	74.11	72.65	57.87	27.55
Proposed	76.77	79.63	71.68	51.33

The use of interval-valued fuzzy sets in this sort of images increases the accuracy of the segmentation. In some cases, it allows us to discover regions that are not easily identified at first sight, as it can be seen in the images we have segmented.

5 Conclusions and future research

In this work we show the advantage of using interval-valued fuzzy sets in the FCM algorithm. For this reason, we have introduced the concept of interval-valued restricted equivalence function and we have presented two different construction

ISBN: 978-989-95079-6-8

methods. The effectiveness of this method has been proved with the experimental results.

As future research we want to work with intervals within the FCM instead of aggregating the bounds of the equivalence interval. Also, we are going to study other processings of ultrasound images to improve the results.

Acknowledgment - This paper has been partially supported by the National Science Foundation of Spain, Reference TIN2007-65981.

References

[1] J.C. Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, NY (1981).

[2] J.C. Bezdek, J. Keller, R. Krisnapuram, N.R. Pal. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. (Kluwer Academic Publishers) (1999).

[3] H. Bustince, E.Barrenechea, M. Pagola. Generation of Interval-Valued Fuzzy and Atanassov's Intuitionistic Fuzzy Connectives from Fuzzy Connectives and from K_α Operators: Laws for Conjunctions and Disjunctions, Amplitude. International Journal of Intelligent Systems, vol. 23 (2008) 680-714.

[4] H. Bustince, E. Barrenechea, M. Pagola. Restricted equivalence functions. Fuzzy Sets and Systems 157 (2006) 2333-2346.

[5] H. Bustince, E. Barrenechea, M. Pagola, R. Orduna. Construction of interval type 2 fuzzy images to represent images in grayscale. False edges. FUZZY IEEE 2007. London, U.K.

[6] T. Calvo, A. Kolesarova, M.Komornikova, R. Masiar. Aggregation operators: properties, classes and construction methods. In T. Calvo, G. Mayor and R. Mesiar (eds): Aggregation Operators New Trends and Applications (Physica-Verlag, Heidelberg, 2002) 3-104.

[7] G. Deschrijver, C. Cornelis, EE.Kerre. On the representation of intuitionistic fuzzy T-norms and T-conorms. IEEE Trans Fuzzy Syst. 12(1) (2004) 45-61.

[8] P. Drygas. The problem of distributivity between binary operations in bi-fuzzy set theory. Proceedings of IPMU'08. L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds). (2008) 1648-1653.

[9] J.A. Noble, D. BouKerroui. Ultrasound image segmentation: A survey. IEEE Transactions on Medical Imaging 25 (2006) 987-1010.

[10] K. Ogawa, M. Fukushima, K. Kubota, N. Hisa. Computer-aided diagnosis system for diffuse liver-diseases with ultrasonography by neural networks. IEEE Transactions on Nuclear Science. 45(6) (1998) 3069-3074.

[11] F. Sahba, H.R. Tizhoosh, M.M. Salama. A coarseto-fine approach to prostate boundary segmentation in ultrasound images. BioMedical Engineering OnLine. (2005) 4-58.

[12] M.H. Yap, E.A. Edirisinghe, H.E. Bez. A novel algorithm for initial lesion detection in ultrasound breast images.Journal of applied clinical medical physics (Fall 2008). Vol. 9, number 4.

[13] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-I. Information Science. 8 (1975) 199-249.

ISBN: 978-989-95079-6-8

Transitive Closure of Interval-valued Relations

Ramón González del Campo[1] Luis Garmendia[2] Jordi Recasens[3]

1.DSIC Universidad Complutense de Madrid
2.DISIA Universidad Complutense de Madrid
3.Universitat Politécnica de Catalunya
Email: rgonzale@estad.ucm.es, lgarmend@fdi.ucm.es, j.recasens@upc.edu

Abstract— *This paper introduces some concepts of interval-valued fuzzy relations and some of their properties: reflexivity, symmetry, T-transitivity, composition and locally reflexivity. It is also introduced the concept of T-transitive closure for an interval-valued fuzzy relation. An algorithm to compute the T-transitive closure of finite interval-valued relations, some properties and some examples are given.*

Keywords— Generalized t-norms, Interval-valued Fuzzy Relations, Interval-valued Fuzzy Sets, t-norms, t-representable t-norms, t-transitive closure.

1 Introduction

Fuzzy sets, $\mathcal{FS}$, were introduced by Zadeh in 1.965 [1]. Since then many generalizations of fuzzy sets have been proposed to model the uncertainty and the vagueness in linguistic variables replacing the unit interval by another structure such as posets or lattices [2, 3, 4]. One of these generalizations are type-2 fuzzy sets, $\mathcal{FS}2$, [5, 6, 7] were introduced by Zadeh. A Type-2 fuzzy set on a universe of discourse X, $\mathcal{FS}2$, is a fuzzy set whose membership function is another fuzzy set on [0,1]:

$$A = \{((x,u), \mu_A(x,u)) \mid \forall x \in X, \forall u \in [0,1]\}$$

Type-2 fuzzy sets have been widely studied and applied since in many cases the uncertainty can be better expressed by a fuzzy set than by a single numeric value. The problem with type2 fuzzy sets though is their computational complexity and the difficulty for an expert to select the adequate fuzzy subset as membership degree of an object to a linguistic label. This is why some simplifications have been proposed, such as the use of only some families of fuzzy sets such as triangular and trapezoidal ones. One of the most popular and useful simplification is the use of interval-valued fuzzy sets.

Interval-valued fuzzy sets ($\mathcal{IVFS}$) were introduced in the 60s by Grattan-Guinness [8], Jahn [9], Sambuc [10] and Zadeh [5]. They are extensions of classical fuzzy sets where the membership value between 0 and 1 is replaced an interval in [0,1]. They easily allow to model uncertainty and vagueness because sometimes it is easier for experts to give a "membership interval" than a membership degree to objects on a universe. $\mathcal{IVFS}$ are a special case of type-2 fuzzy sets that simplifies the calculations while preserving their richness as well. Intuitionistic fuzzy sets on X ($\mathcal{IFS}$) introduced by Atanassov [11]. In intuitionistic fuzzy sets each element has a membership degree, μ, and a non-membership degree, ν satisfying: $\mu + \nu \leq 1$.

$A = \{(x, \mu(x), \nu(x)) \mid x \in X, \mu(x), \nu(x) \in [0,1]\}$

The value $\pi = 1 - \mu - \nu$ is a measure of the uncertainty. This paper is organized as follows: In section three we introduce some important properties of interval-valued fuzzy relations such as reflexivity, symmetry, T-transitivity, composition and local reflexivity. Traditionally, the study of conjunctions between interval-valued fuzzy sets has been reduced to be modeled with t-representable t-norms. However, not all generalized t-norms are t-representable. Moreover, some of the non t-representable t-norms sometimes satisfy even more properties than t-representable t-norms [12]. Probably the most important property a fuzzy relation can fulfil is transitivity with respect a given t-norm. Since many times the data are given by a proximity relation P (i.e.: a reflexive and symmetric but not necessarily transitive fuzzy relation), there are some methods to obtain a transitive relation close to P to replace it when transitivity is required. The most popular way to do this is calculating its transitive closure. In section four we introduce the concept of T-transitive closure for an interval-valued fuzzy relation and its expression in a finite universe for any t-norm. A few methods to compute it and some examples are given.

2 Preliminaries

2.1 Interval-valued fuzzy sets

Definition 2.1 *[13] Let $\mathcal{L}^I = (L, \leq_L)$ be a lattice that satisfies:*

1. $L = \{[x_1, x_2] \in [0,1]^2$ *with* $x_1 \leq x_2\}$.
2. $[x_1, x_2] \leq_L [y_1, y_2]$ *if and only if* $x_1 \leq y_1$ *and* $x_2 \leq y_2$

Trivially:

$[x_1, x_2] <_L [y_1, y_2] \Leftrightarrow x_1 < y_1, x_2 \leq y_2$ or $x_1 \leq y_1, x_2 < y_2$
$[x_1, x_2] =_L [y_1, y_2] \Leftrightarrow x_1 = y_1, x_2 = y_2$.

$0_L =_L [0,0]$ and $1_L =_L [1,1]$ are the smallest and the greatest elements in L respectively.

$\mathcal{L}^I$ is a complete lattice an the supremum and infimum are defined as follows.

Definition 2.2 *[12] Let $\{[v_i, w_i]\}$ be a set of intervals on L. Then*

1. *Infimum:*
   $Inf_L\{[v_i, w_i]\} \equiv [infimun\{v_i\}, infimun\{w_i\}]$
2. *Supremum:*
   $Sup_L\{[v_i, w_i]\} \equiv [supremun\{v_i\}, supremun\{w_i\}]$

Definition 2.3 *[13] An interval-valued fuzzy set A on a universe X is a mapping:*

ISBN: 978-989-95079-6-8

$$A = \{(a, [x_1, x_2]) \mid a \in X, [x_1, x_2] \in L\}$$

Definition 2.4 *[13] Let X be a universe and A and B two interval-valued fuzzy sets. The equality between A and B is defined as: $A =_L B$ if and only if $A(a) =_L B(a)\ \forall a \in X$.*

Definition 2.5 *[13] Let X be a universe and A and B two interval-valued fuzzy sets. The inclusion of A in to B is defined as: $A \subseteq_L B$ if and only if $A(a) \leq_L B(a)\ \forall a \in X$.*

Definition 2.6 *[13] A negation function for interval-valued fuzzy sets $\mathcal{N}$ is a decreasing function, $\mathcal{N} : L \to L$, that satisfies:*

1. $\mathcal{N}(0_L) =_L 1_L$
2. $\mathcal{N}(1_L) =_L 0_L$

If $\mathcal{N}(\mathcal{N}([x_1, x_2])) =_L [x_1, x_2]$ for all $[x_1, x_2]$ in L then $\mathcal{N}$ is called an involutive negation.

Definition 2.7 *A strong negation function for interval-valued fuzzy sets, $\mathcal{N}$, is a involutive function, $\mathcal{N} : L \to L$, that satisfies:*

1. $\mathcal{N}(0_L) =_L 1_L$
2. $\mathcal{N}(1_L) =_L 0_L$

T-norms are generalized to the lattice $\mathcal{L}^I$ in a straightforward way.

Definition 2.8 *A generalized t-norm function [13], $\mathcal{T}$, is a monotone increasing, symmetric and associative operator, $\mathcal{T} : L^2 \to L$, that satisfies: $\mathcal{T}(1_L, [x_1, x_2]) =_L [x_1, x_2]$ for all $[x_1, x_2]$ in L.*

Due to monotony, it is easy to show:

$$\mathcal{T}(Sup_L\{[v_i, w_i]\}, [y_1, y_2]) \geq_L Sup_L\{\mathcal{T}([v_i, w_i], [y_1, y_2])\}$$
$$\mathcal{T}(Inf_L\{[v_i, w_i]\}, [y_1, y_2]) \leq_L Inf_L\{\mathcal{T}([v_i, w_i], [y_1, y_2])\}$$

Due to the associativity of $\mathcal{T}$ the conjunction of three or more intervals can be defined inductively as:

$$\mathcal{T}(a, \mathcal{T}(b, c)) =_L \mathcal{T}(\mathcal{T}(a, b), c) =_L a \triangle b \triangle c \text{ where } \triangle =_L \mathcal{T}.$$

where $a =_L [a_1, a_2]$, $b =_L [b_1, b_2]$ and $c =_L [c_1, c_2]$.

$\underline{T([x_1, x_2], [y_1, y_2])}$ and $\overline{T([x_1, x_2], [y_1, y_2])}$ will denote the lower and the higher values of $T([x_1, x_2], [y_1, y_2])$.

Definition 2.9 *[14] Let $\{x_i\}$ in $[0, 1]$. A t-norm T in $([0, 1], \leq)$ is left-continuous if it satisfies:*

$$T(Sup\ x_i, y) = Sup\ T(x_i, y)$$

Right-continuity can be defined in a similar way. This property is also called sup-preserving.

Definition 2.10 *[13] A generalized t-norm operator $\mathcal{T}$ is t-representable in $\mathcal{L}^I$ if there are two t-norms: T_1 and T_2 (T_1, T_2, in ([0,1],$\leq$)) that satisfy:*

$$\mathcal{T}([x_1, x_2], [y_1, y_2]) =_L [T_1(x_1, y_1), T_2(x_2, y_2)]$$

where $T_1(v, w) \leq T_2(v, w)\ \forall v, w \in [0, 1]$.

Let $x =_L [x_1, x_2]$ and $y =_L [y_1, y_2]$ be two intervals on L:

Example 2.1
$Inf_L(\{[x_1, x_2], [y_1, y_2]\}) =_L [min(x_1, y_1), min(x_2, y_2)]$ is t-representable by means of $T = min$ in ([0,1],$\leq$):

Example 2.2 *The following product generalized t-norm $*_L$ is t-representable:*

$$*_L([x_1, x_2], [y_1, y_2]) =_L [x_1 * y_1, x_2 * y_2]$$

Example 2.3 *There are two generalizations of the Lukasiewicz t-norm [12]:*

- $T_w([x_1, x_2], [y_1, y_2]) =_L$ $[max(0, x_1 + y_1 - 1), max(0, x_2 + y_2 - 1)]$
- $T_W([x_1, x_2], [y_1, y_2]) =_L$ $[max(0, x_1 + y_1 - 1), max(0, x_1 + y_2 - 1, x_2 + y_1 - 1)]$

Note that T_w is t-representable but T_W is not.

Definition 2.11 *[13] A generalized t-conorm $\mathcal{S}$ is an increasing, commutative, and associative operator $\mathcal{S} : L^2 \to L$, that satisfies: $\mathcal{S}(0_L, [x_1, x_2]) =_L [x_1, x_2]$ and $\mathcal{S}(1_L, [x_1, x_2]) =_L 1_L$.*

Due to the associativity of $\mathcal{S}$ we can write:

$\mathcal{S}(a, \mathcal{S}(b, c)) =_L \mathcal{S}(\mathcal{S}(a, b), c) =_L a \bigtriangledown b \bigtriangledown c$ where $\bigtriangledown =_L \mathcal{S}$.

For example, $\mathcal{S} = Sup_L$ is a generalized t-conorm.
In this paper the next definition is proposed.

Definition 2.12 *Let $\{[v_i, w_i]\}$ be in L. A generalized t-norm operator $\mathcal{T}$ is left-continuous if and only if:*

$$\mathcal{T}(Sup_L\{[v_i, w_i]\}, [y_1, y_2]) =_L Sup_L\{\mathcal{T}([v_i, w_i], [y_1, y_2])\}$$

Right-continuity can be defined in a similar way. A t-norm, a negation and the dual t-conorm of $\mathcal{T}$ with respect to $\mathcal{N}$ is called a De Morgan triplet.

3 Interval-valued fuzzy relations

This section contains some definitions on interval-valued fuzzy relations that generalize the ones of fuzzy relations valued on [0,1]. In the next section, $\mathcal{T}$-transitive relations will be generated using the $Sup - \mathcal{T}$ product.

Definition 3.1 *[15] Let X_1 and X_2 be two universes of discourse. An interval-valued fuzzy relation $R : X_1 \times X_2 \to L$ (where $\mathcal{L}^I = (L, \leq_L)$ is a lattice) is a mapping:*

$$R = \{((a, b), [x, y]) \mid a \in X_1, b \in X_2, [x, y] \in L\}$$

where $x = \underline{R(a, b)}$ and $y = \overline{R(a, b)}$.
In the rest of the paper, it is assumed that $X_1 = X_2 = X$.

Definition 3.2 *An interval-valued relation $R : X^2 \to L$ is reflexive if:*

$$R(a, a) =_L 1_L\ \forall a \in X$$

ISBN: 978-989-95079-6-8

Definition 3.3 *An interval-valued relation $R : X^2 \to L$ is $[\alpha_1 - \alpha_2]-$reflexive if:*

$$R(a,a) \geq_L [\alpha_1, \alpha_2] \; \forall a \in X$$

where $\alpha_1, \alpha_2 \in [0,1]$ and $\alpha_1 \leq \alpha_2$.

Definition 3.4 *An interval-valued relation $R : X^2 \to L$ is symmetric if:*

$$R(a,b) =_L R(b,a) \; \forall a, b \in X$$

Definition 3.5 *Let $\mathcal{T}$ be a generalized t-norm operator and let R interval-valued fuzzy relation on X. R is $\mathcal{T}$-transitive if:*

$$\mathcal{T}(R(a,b), R(b,c)) \leq_L R(a,c) \; \forall a, b, c \in X$$

Proposition 3.1 *If $\mathcal{T}$ is t-representable with T_1 and T_2 ($\mathcal{T} = [T_1, T_2]$) then an interval-valued relation $R : X^2 \to L$ is $\mathcal{T}$-transitive if and only if $\underline{R}$ is T_1-transitive and $\overline{R}$ is T_2-transitive:*

Definition 3.6 *Let $\mathcal{T}$ a generalized t-norm. Let $\mathcal{S}$ a generalized t-conorm from $\mathcal{S}$ and let be $\bigtriangledown$ the n-ary generalized t-conorm defined by associativity. Let R_1 and R_2 be two interval-valued fuzzy relations on a finite set $X = \{c_1, ...c_m\}$. The $\mathcal{S} - \mathcal{T}$-composition of R_1 and R_2 is defined as follows:*

$$(R_1 \diamond_{(\mathcal{ST})} R_2)(c_q, c_r) =_L \bigtriangledown_{c_k \in X} \mathcal{T}(R_1(c_q, c_k), R_2(c_k, c_r))$$

where $\bigtriangledown_{c_k \in X} \mathcal{T}(R_1(c_q, c_k), R_2(c_k, c_r)) =_L \mathcal{T}(R_1(c_q, c_1), R_2(c_1, c_r)) \bigtriangledown ... \bigtriangledown \mathcal{T}(R_1(c_q, c_m), R_2(c_m, c_r))$.

Definition 3.7 *The local equality relation of a fuzzy interval-valued fuzzy relation R on X is the fuzzy relation E_R is:*

$$E_R(a,b) = \begin{cases} Sup_{L\ \forall c \in X} max(R(a,c), R(c,a)) & a = b; \\ [0,0], & a \neq b. \end{cases}$$

Definition 3.8 *An interval-valued relation R is locally reflexive if $E_R \subseteq_L R$.*

It is easy to prove that if R is an reflexive interval-valued relation then R is a locally reflexive interval-valued relation.

4 $\mathcal{T}$-transitive closure for IVFSs

In this section it is proved the existence of the $\mathcal{T}$-transitive closure of an interval valued fuzzy relation and its calculation via $sup - \mathcal{T}$. The $sup - \mathcal{T}$ product is associative if $\mathcal{T}$ is left continuous of the universe is finite. Moreover, it is proved that the $sup - \mathcal{T}$ product is continuous and so is the calculation of the $\mathcal{T}$-transitive closure for finite universes. An algorithm to calculate the $\mathcal{T}$-transitive closure is provided.

Definition 4.1 *Let $\mathcal{T}$ be a generalized t-norm and let $R : X^2 \to L$ be an interval-valued relation on a finite universe X. The $\mathcal{T}$-transitive closure of R is the relation $R^{\mathcal{T}} : X \times X \to L$ that satisfies:*

1. $R \subseteq_L R^{\mathcal{T}}$.
2. *If $R \subseteq_L R'$ and R' is $\mathcal{T}$-transitive then $R^{\mathcal{T}} \subseteq_L R'$.*

Lemma 4.1 *R is $\mathcal{T}$-transitive if and only if $R \diamond_{Sup_L \mathcal{T}} R \subseteq_L R$.*

Proof.

- R is $\mathcal{T}$-transitive $\Rightarrow R \diamond_{Sup_L \mathcal{T}} R \subseteq_L R$:
  $(R \diamond_{Sup_L \mathcal{T}} R)(a,b) =_L$
  $Sup_{L\ c \in X}\{\mathcal{T}(R(a,c), R(c,b))\} \leq_L R(a,b)$
  due to $\mathcal{T}(R(a,c), R(c,b)) \leq_L R(a,b) \; \forall a, b, c \in X$
- $R \diamond_{Sup_L \mathcal{T}} R \subseteq_L R \Rightarrow R$ is $\mathcal{T}$-transitive:
  $Sup_{L\ d \in X}\{\mathcal{T}(R(a,d), R(d,b))\} \leq_L R(a,b) \; \forall a, b \in X$
  Then $\forall c \in X \; \mathcal{T}(R(a,c), R(c,b)) \leq_L R(a,b) \square$

Lemma 4.2 *Let R be an interval-valued relation. If $R^{\mathcal{T}}$ exists then it is unique.*

Proof. Let S_1 and S_2 be two relations. If S_1 and S_2 are transitive closures of R then according to definition 4.1: $S_1 \subseteq_L S_2$ and $S_2 \subseteq_L S_1$, consequently $S_1 =_L S_2 \square$.

Lemma 4.3 *Let R be an interval-valued relation in a universe X and an let $\mathcal{T}$ be an arbitrary generalized t-norm. Then the transitive closure of R always exists.*

Proof. Consider the set Ω_R of $\mathcal{T}$-transitive fuzzy relations containing R. Let us define the fuzzy relation $S_R^*(a,b) =_L Inf_{L;S \in \Omega_R}\{S(a,b)\}$. We will prove that S_R^* is $\mathcal{T}$-transitive. Due to monotony of generalized $\mathcal{T}$-norms it is obtained:
$\mathcal{T}(S_R^*(a,b), S_R^*(b,c)) =_L$
$\mathcal{T}(Inf_{L\ S_1 \in \Omega_R}\{S_1(a,b)\}, Inf_{L\ S_2 \in \Omega_R}\{S_2(b,c)\})$
$\leq_L Inf_{L\ S_1 \in \Omega_R} Inf_{L\ S_2 \in \Omega_R}\{\mathcal{T}(S_1(a,b), S_2(b,c))\}$
$\leq_L Inf_{L\ S \in \Omega_R}\{\mathcal{T}(S(a,b), S(b,c))\} =_L S_R^*(a,c) \square$

Proposition 4.1 *Let X be an arbitrary universe. Let $\{S_1, ..., S_n\}$ a set of interval-valued relations. If $\mathcal{T}$ is a left-continuous generalized $\mathcal{T}$-norm, then it satisfies:*

$$R \diamond_{Sup_L \mathcal{T}} (Sup_{L\ \forall i=1..n}\{S_i\}) =_L Sup_{L\ \forall i=1..n}\{R \diamond_{Sup_L \mathcal{T}} S_i\}$$

Proof.

$$\begin{aligned} &R \diamond_{Sup_L \mathcal{T}} (Sup_{L\ \forall i=1..n}\{S_i\})(a,c) =_L \\ &Sup_{L\ b \in X}\{\mathcal{T}(R(a,b), Sup_{L\ \forall i=1..n}\{S_i(b,c)\})\} \\ &=_L Sup_{L\ \forall i=1..n}\ Sup_{L\ b \in X}\{\mathcal{T}(R(a,b), S_i(b,c))\} \\ &=_L (Sup_{L\ \forall i=1..n}\{R \diamond_{Sup_L \mathcal{T}} S_i\})(a,c) \square \end{aligned}$$

Definition 4.2 *Given a generalized t-norm $\mathcal{T}$, the $\mathcal{T}$-power $R^{(n)_{\mathcal{T}}}$ of a fuzzy relation R on X is recursively defined as follows:*

1. $R^{(1)_{\mathcal{T}}} \equiv R$
2. $R^{(n)_{\mathcal{T}}} \equiv R^{(n-1)_{\mathcal{T}}} \diamond_{Sup_L \mathcal{T}} R$

Lemma 4.4 *If $A \subseteq_L B$ then $A^{(k)} \subseteq_L B^{(k)}$*

Proof.
Trivial due to monotony of $\mathcal{T} \square$.

Theorem 4.1 *Let X be an arbitrary universe and $\mathcal{T}$ a left-continuous t-norm. The transitive closure of R is:*

$$R^{\mathcal{T}} =_L Sup_{L\ \forall k \in \mathbb{N}}\{R^{(k)_{\mathcal{T}}}\}$$

Proof.
Let $S = Sup_{L\ \forall k \in \mathbb{N}}\{R^{(k)_{\mathcal{T}}}\}$ be. By proposition 4.1:

ISBN: 978-989-95079-6-8

$$S^{(2)_\mathcal{T}} =_L$$
$$(Sup_{L\ \forall k\in\mathbb{N}}\{R^{(k)_\mathcal{T}}\}) \diamond_{Sup_L\mathcal{T}} (Sup_{L\ \forall l\in\mathbb{N}}\{R^{(l)_\mathcal{T}}\}) =_L$$
$$Sup_{L\ \forall k,l\in\mathbb{N}}\{R^{(k+l)_\mathcal{T}}\} =_L Sup_{L\ \forall m\in\mathbb{N}\setminus\{0,1\}}\{R^{(m)_\mathcal{T}}\} \subseteq S$$

By lemma 4.1, it follows that S is T-transitive.

Moreover, if $R \subseteq S'$ and S' is T-transitive, then again by lemma 4.1 and 4.4 it holds that:

$$R^{(2)_\mathcal{T}} \subseteq_L S'^{(2)_\mathcal{T}} \subseteq_L S', ..., R^{(k)_\mathcal{T}} \subseteq_L S'^{(k)_\mathcal{T}} \subseteq_L S'$$

hence $S \subseteq_L S'$ and $R^\mathcal{T} =_L S\Box$.

Lemma 4.5 *Let $\mathcal{T}$ be any generalized t-norm, then:*

$$\mathcal{T}([x_1,x_2],[y_1,y_2]) \leq_L Inf_L([x_1,x_2],[y_1,y_2])$$
$$\forall [x_1,x_2],[y_1,y_2] \in L$$

Proof.
$\mathcal{T}([x_1,x_2],[y_1,y_2]) \leq_L \underline{\mathcal{T}([x_1,x_2],1_L) =_L [x_1,x_2]}$
$\Rightarrow [\underline{\mathcal{T}([x_1,x_2],[y_1,y_2])}, \overline{\mathcal{T}([x_1,x_2],[y_1,y_2])}] \leq_L [x_1,x_2]$
$\underline{\mathcal{T}([x_1,x_2],[y_1,y_2])} =_L \overline{\mathcal{T}([y_1,y_2],[x_1,x_2])} \leq_L$
$\underline{\mathcal{T}([y_1,y_2],1_L) =_L [y_1,y_2]}$
$\Rightarrow [\underline{\mathcal{T}([x_1,x_2],[y_1,y_2])}, \overline{\mathcal{T}([x_1,x_2],[y_1,y_2])}] \leq_L [y_1,y_2]$
Therefore:
$\underline{\mathcal{T}([x_1,x_2],[y_1,y_2])} \leq minimum[x_1,y_1]$
$\overline{\mathcal{T}([x_1,x_2],[y_1,y_2])} \leq minimum[x_2,y_2] \Rightarrow$
$\mathcal{T}([x_1,x_2],[y_1,y_2]) \leq_L Inf_L([x_1,x_2],[y_1,y_2])\Box$

Lemma 4.6 *Let $\mathcal{S}$ be any generalized t-conorm, then:*

$$\mathcal{S}([x_1,x_2],[y_1,y_2]) \geq_L Sup_L([x_1,x_2],[y_1,y_2])$$
$$\forall [x_1,x_2],[y_1,y_2] \in L$$

Proof.
$\mathcal{S}([x_1,x_2],[y_1,y_2]) \geq_L \underline{\mathcal{S}([x_1,x_2],1_L) =_L [x_1,x_2]}$
$\Rightarrow [\underline{\mathcal{S}([x_1,x_2],[y_1,y_2])}, \overline{\mathcal{S}([x_1,x_2],[y_1,y_2])}] \geq_L [x_1,x_2]$
$\underline{\mathcal{S}([x_1,x_2],[y_1,y_2])} =_L \overline{\mathcal{S}([y_1,y_2],[x_1,x_2])} \geq_L$
$\underline{\mathcal{S}([y_1,y_2],1_L) =_L [y_1,y_2]}$
$\Rightarrow [\underline{\mathcal{S}([x_1,x_2],[y_1,y_2])}, \overline{\mathcal{S}([x_1,x_2],[y_1,y_2])}] \geq_L [y_1,y_2]$
Therefore:
$\underline{\mathcal{S}([x_1,x_2],[y_1,y_2])} \geq maximun[x_1,y_1]$
$\overline{\mathcal{S}([x_1,x_2],[y_1,y_2])} \geq maximun[x_2,y_2] \Rightarrow$
$\mathcal{S}([x_1,x_2],[y_1,y_2]) \geq_L Sup_L([x_1,x_2],[y_1,y_2])\Box$

Due associativity of $\mathcal{T}$:

$$\mathcal{T}([x_1,x_2],\mathcal{T}([y_1,y_2],[z_1,z_2])) =_L$$
$$\mathcal{T}(\mathcal{T}([x_1,x_2],[y_1,y_2]),[z_1,z_2]) =_L [x_1,x_2] \triangle [y_1,y_2] \triangle [z_1,z_2]$$
$$Inf_L([x_1,x_2],Inf_L([y_1,y_2],[z_1,z_2])) =_L$$
$$Inf_L(Inf_L([x_1,x_2],[y_1,y_2]),[z_1,z_2])) =_L [x_1,x_2] \triangle_{L-inf} [y_1,y_2] \triangle_{L-inf} [z_1,z_2]$$

Lemma 4.7 *Let $\mathcal{T}$ be an arbitrary generalized t-norm. Let P be a path with a cycle:*

$$P \equiv R(a,a_1) \triangle R(a_1,a_2) \triangle ... \triangle R(a_{t-1},a_t) \triangle R(a_t,c_1) \triangle R(c_1,c_2) \triangle ... \triangle R(c_q,a_t) \triangle R(a_t,a_{t+1}) \triangle ... \triangle R(a_k,b)$$

then

$$P \equiv R(a,a_1) \triangle R(a_1,a_2) \triangle ... \triangle R(a_{t-1},a_t) \triangle R(a_t,c_1) \triangle R(c_1,c_2) \triangle ... \triangle R(c_q,a_t) \triangle R(a_t,a_{t+1}) \triangle ... \triangle R(a_k,b) \leq_L R(a,a_1) \triangle R(a_1,a_2) \triangle ... \triangle R(a_{t-1},a_t) \triangle R(a_t,a_{t+1}) \triangle ... \triangle R(a_k,b)$$

Proof.
Trivial due to $\mathcal{T}([x_1,x_2],[y_1,y_2]) \leq_L [y_1,y_2]$ forall $[x_1,x_2],[y_1,y_2]$ in L and associativity of $\mathcal{T}\Box$.

Theorem 4.2 *Let X be a finite universe with cardinality n. The transitive closure of R, $R^\mathcal{T}$, is:*

$$R^\mathcal{T} =_L Sup_{L\ k=1..n}\{R^{(k)_\mathcal{T}}\}$$

Proof.

$R^{(k)_\mathcal{T}}(a,b) =_L$

$=_L Sup_{L\ a_1,a_2,...,a_{k-1}} R(a,a_1) \triangle R(a_1,a_2) \triangle ... \triangle R(a_{k-1},b) \leq_L$
$=_L Sup_{L\ a_1,a_2,...,a_{k-1}} R(a,a_1) \triangle_{Inf_L} R(a_1,a_2) \triangle_{Inf_L} ... \triangle_{Inf_L} R(a_{k-1},b)$ according to lemma 4.5.

Now suppose $k > n$. Then $R^{(k)} \subseteq_L R^{(n)}$ because it exists a cycle (lemma 4.7). Therefore, the cases $k > n$ need not be considered $\Box$.

Corollary 4.1 *Let X be a finite universe with cardinality n. If R is a locally reflexive relation, the transitive closure of R is:*

$$R^\mathcal{T} =_L Sup_{L\ k=1..n-1}\{R^{(k)_\mathcal{T}}\}$$

Proof
If R is a locally reflexive relation then $R(a,b) \leq_L R(a,a)$. Consider a cycle of length n, then its strength is:
$R^{(n)_\mathcal{T}}(a,a) =_L$

$=_L Sup_{L\ a_1,a_2,...,a_{k-1}} R(a,a_1) \triangle R(a_1,a_2) \triangle ..., \triangle R(a_{k-1},a) \leq_L$
$Sup_{L\ a_1,a_2,...,a_{k-1}} R(a,a_1) \triangle_{Inf_L} R(a_1,a_2) \triangle_{Inf_L} ... \triangle_{Inf_L} R(a_{k-1},a) \leq_L R(a,a)\Box$

Theorem 4.3 *Let X be a finite universe with cardinality n. If R is a reflexive fuzzy relation on X then the transitive closure of R is:*

$$R^\mathcal{T} =_L R^{(n-1)_\mathcal{T}}$$

Proof.
$R(a,b)^{(k)_\mathcal{T}} =_L$

$=_L \mathcal{T}(R(a,a), R^{(k)_\mathcal{T}}(a,b))$
$\leq_L Sup_{L\ c\in X}\{T(R(a,c), R^{(k)_\mathcal{T}}(c,b))\}$
$=_L R^{(k+1)_\mathcal{T}}(a,b)$

Therefore $R(a,b)^{(k)_\mathcal{T}} \leq_L R^{(k+1)_\mathcal{T}}(a,b) \Leftrightarrow R^{(k)_\mathcal{T}} \subseteq R^{(k+1)_\mathcal{T}}\Box$

Theorem 4.4 *Let $\mathcal{T}$ be a t-representable t-norm ($\mathcal{T} = [T_1,T_2]$) and let $R = [\underline{R},\overline{R}]$ be a interval-valued relation. Then $R^\mathcal{T} = [\underline{R}^{T_1}, \overline{R}^{T_2}]$.*

Proof.

$R^{(k)_\mathcal{T}}(a,b) =_L$

$=_L \ Sup_{L\ a_1,a_2,...,a_{k-1}} R(a,a_1) \ \triangle \ R(a_1,a_2) \ \triangle$
$..., \triangle \ (a_{k-1},b)$
$=_L \ Sup_{L\ a_1,a_2,...,a_{k-1}} [\underline{R(a,a_1)} \ \triangle_1 \ \underline{R(a_1,a_2)} \ \triangle_1$
$..., \triangle_1 \ \underline{R(a_{k-1},b)}, \overline{R(a,a_1)} \ \triangle_2 \ \overline{R(a_1,a_2)} \ \triangle_2$
$..., \triangle_2 \ \overline{R(a_{k-1},b)}]$
$=_L \ [max_{a_1,a_2,...,a_{k-1}} \underline{R(a,a_1)} \ \triangle_1 \ \underline{R(a_1,a_2)} \ \triangle_1$
$..., \triangle_1 \ \underline{R(a_{k-1},b)}$
$, max_{a_1,a_2,...,a_{k-1}} \overline{R(a,a_1)} \ \triangle_2 \ \overline{R(a_1,a_2)} \ \triangle_2 ..., \triangle_2$
$\overline{R(a_{k-1},b)}]$
$=_[\ \underline{R^{(k)_{T_1}}(a,b)}, \overline{R^{(k)_{T_2}}(a,b)}]$

where $\triangle = [\triangle_1, \triangle_2]$. Then $\mathcal{T} = [T_1, T_2] \Rightarrow R^{\mathcal{T}} = [\underline{R}^{T_1}, \overline{R}^{T_2}]\Box$.

Definition 4.3 *The distance d of the supremum on L is defined for all $x_1, x_2, y_1, y_2 \in [0,1]$ by $d([x_1,x_2],[y_1,y_2]) = max(|x_1 - y_1|, |x_2 - y_2|)$.*

Definition 4.4 *Let R,S be two interval-valued fuzzy relatons on a set X. The distance d between R and S us defined by:*

$$d(R,S) = sup_{x,y \in X} d(R(x,y), S(x,y))$$

Lemma 4.8 *d is a distance on the set R_X of interval-valued fuzzy relations on X $\Box$.*

Proof.
The supremum of distances is a distance.

Theorem 4.5 *Let T be a generalized continuous t-norm and R_X the set of interval-valued fuzzy relations on X. R_X with the sup-$\mathcal{T}$ product is an ordered topological semigroup.*

Proof.

- Associativity is a straightforward exercise.
- The interval-valued relation $E(x,y) = \begin{cases} [1,1], & x=y; \\ [0,0], & \text{otherwise.} \end{cases}$ is the identity element of R_X.
- Continuity: Since T is defined on a compact set, it is uniformly continuous. Therefore:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that}$$
$$\forall m, n, m', n', a, b, a', b' \in [0,1]$$
$$max(|m-a|, |m'-a'|, |n-b|, |n'-b'|) \Rightarrow$$
$$|\mathcal{T}([m,m'],[n,n']) - \mathcal{T}([a,a'],[b,b'])| < \epsilon (*)$$

We want to prove that given two interval-valued fuzzy relations $A, B \in R_X$:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } \forall M, N \in R_X$$
$$max(d(M,A), d(N,B)) < \delta \Rightarrow d(M \circ N, A \circ B) < \epsilon$$

Given ϵ, we take $\delta > 0$ satisfying (*). Then:

$$d(M \circ N, A \circ B) =$$
$$sup_{x,y \in X} |sup_{z \in X} \mathcal{T}(M(x,z), N(z,y)) -$$
$$sup_{z \in X} \mathcal{T}(A(x,z), B(z,y))| \leq$$
$$sup_{x,y \in X} sup_{z \in X} |\mathcal{T}(M(x,z), N(z,y)) -$$
$$\mathcal{T}(A(x,z), B(z,y))| \leq \epsilon$$

- Monotonicity is an immediate consequence of the monotonicity of $\mathcal{T}\Box$.

Corollary 4.2 *For any positive integer n the map assigning to an interval-values fuzzy relation M on a finite set its n^{th} power is non-decreasing and continuous if the corresponding generalized t-norm $\mathcal{T}$ is continuous.*

Corollary 4.3 *If a generalized t-norm $\mathcal{T}$ is continuous, the map that assigns the $\mathcal{T}$-transitive closure to the interval-valued fuzzy relations on a finite set is non-decreasing and continuous.*

Theorem 4.6 *Given a method to calculate the transitive closure of a fuzzy relation. Nevertheless, there are more effective methods and algorithms. For example it is possible to use an extension of the Floyd-Warshall algorithm [16]*

4.1 Algorithm to compute the transitive closure of an interval-valued fuzzy relation

Let R be an interval-valued relation on a finite universe X with cardinality n and let $\mathcal{T}$ be a generalized t-norm. It is possible to compute $R^{\mathcal{T}}$ using the next algorithm:

for k $\leftarrow$ 1 **until** n **do**
  for i $\leftarrow$ 1 **until** n **do**
    for j $\leftarrow$ 1 **until** n **do**
      R(i,j)$\leftarrow Sup_L(R(i,j), \mathcal{T}(R(i,k), R(k,j)))$
    end for
  end for
end for

Example 4.1 *Let $\mathcal{T}$ be a generalized t-norm $T_w([x_1,x_2],[y_1,y_2]) =_L [max(0, x_1 + y_1 - 1), max(0, x_2 + y_2 - 1)]$ and let $R : X \times X \rightarrow L$ be the following interval-valued relation:*

$$R = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0,0] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

The computation of the $\mathcal{T}$-transitive closure of R is the following:

$$R^{\mathcal{T}} = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0.2,0.8] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

Note that by theorem 4.4, $R^{\mathcal{T}} = [\underline{R^W}, \overline{R^W}]$

Example 4.2 *Let $\mathcal{T}$ be a generalized t-norm $T([x_1,x_2],[y_1,y_2]) =_L [W(x_1,y_1), Min(x_2,y_2)]$ and let $R : X \times X \rightarrow L$ be the following interval-valued relation:*

$$R = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0,0] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

The computation of the $\mathcal{T}$-transitive closure of R is the following:

ISBN: 978-989-95079-6-8

$$R^{\mathcal{T}} = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0.2,0.9] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

Note that by theorem 4.4, $R^{\mathcal{T}} = [\underline{R^W}, \overline{R^{Min}}]$

Example 4.3 *Let* $\mathcal{T}$ *be a generalized t-norm* $T_W([x_1,x_2],[y_1,y_2]) =_L [max(0, x_1 + y_1 - 1), max(0, x_1 + y_2 - 1, x_2 + y_1 - 1)]$ *and let* $R : X \times X \to L$ *be the following interval-valued relation:*

$$R = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0,0] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

The computation of the $\mathcal{T}$*-transitive closure of* R *is the following:*

$$R^{\mathcal{T}} = \begin{pmatrix} [1,1] & [0.6,0.8] & [0.6,0.9] & [0.2,0.5] \\ & [1,1] & [0.4,0.9] & [0.6,0.9] \\ & & [1,1] & [0.6,0.9] \\ & & & [1,1] \end{pmatrix}$$

Note that the generalized $\mathcal{T}$*-norm in this example is not t-representable.*

5 An Application

It is possible to model social relations using Interval-valued Fuzzy Sets. For example, it is possible to model "friendship". Let R be a relation which describes the degree of friendship between pairs of people. If A and B are friends and B and C are friends, then it will probably exist some degree of friendship between A and C. In this context, the $\mathcal{T}$-transitive closure can be used to learn friendship of individuals from a chain of friends, for example, from common friends.

Example 5.1 *Let* R *be an interval-valued relation which models the degree of friendship between four people* $\{a, b, c, d\}$*. It is possible to model the transitivity of friendship using the generalized t-norm* $*_L([x_1,x_2],[y_1,y_2]) =_L [x_1 * y_1, x_2 * y_2]$*:*

$$R = \begin{pmatrix} [1,1] & [0.6,0.9] & [0.6,0.7] & [0,0] \\ & [1,1] & [0,0] & [0.2,0.8] \\ & & [1,1] & [0,0] \\ & & & [1,1] \end{pmatrix}$$

Initially, d *has only friendship with* b*. In addition, the uncertainty of degree of friendship between* b *and* d *is wide. It is interesting to see how friendship is propagated.*

$$R^{\mathcal{T}} = \begin{pmatrix} [1,1] & [0.6,0.9] & [0.6,0.7] & [0.16,0.72] \\ & [1,1] & [0.48,0.63] & [0.2,0.8] \\ & & [1,1] & [0.096,0.5] \\ & & & [1,1] \end{pmatrix}$$

In $R^{\mathcal{T}}$ d *has some degree of friendship with* a, c *learned by a common friend* b*. The uncertainty of degree of friendship of* d *with the rest of them has been propagated as well.*

6 Conclusions

Traditionally, the study of interval-valued fuzzy sets has been reduced to t-representable t-norms. We have generalized the main properties of interval-valued fuzzy relations: reflexivity, symmetry, composition, local reflexivity and t-transitivity. We also have generalized the T-transitive closure for any t-norm and an algorithm to compute it is provided. It is analyzed the case of t-representable t-norms and proved that in this case the T-transitive closure of R can be obtained from the two classical T-transitive closures of the lower and higher relations (theorem 4.4). Some examples are provided.

References

[1] L.A. Zadeh. Fuzzy sets. *Information And Control*, 8:338–353, 1965.

[2] H. Bustince, F. Herrera, and J. Montero. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, volume 220 of *Studies in Fuzziness and Soft Computing*. Hardcover, 2008.

[3] D. Gomez J. Montero and Bustince H. On the relevance of some families of fuzzy sets. *Fuzzy Sets And Systems*, 158:2429–2442, 2007.

[4] J. Goguen. L-fuzzy sets. *Journal Of Mathematical Analysis And Applications*, 18:145–174, 1967.

[5] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning i. *Information Sciences*, 8:199–249, 1975.

[6] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning ii. *Information Sciences*, 8:301–357, 1975.

[7] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning iii. *Information Sciences*, 9:43–80, 1975.

[8] I. Grattan-Guiness. Fuzzy membership mapped onto interval and many-valued quantities. *Math. Logik. Grundladen Math*, 22:149–160, 1975.

[9] K.U. Jahn. Intervall-wertige mengen. *Math. Nach.*, 68:115–132, 1975.

[10] E. Sanchez and R. Sambuc. Fuzzy relationships. phi -fuzzy functions. application to diagnostic aid in thyroid pathology. *Proceedings of an International Symposium on Medical Data Processing*, pages 513–524, 1976.

[11] K.T. Atanassov. *Intuitionistic Fuzzy Sets*. Physica-Verlag, Heidelberg, New York,, 1999.

[12] C. Cornelis, G. Deschrijver, and E. Kerre. Advances and challenges in interval-valued fuzzy logic. *Fuzzy Sets and Systems*, 157(5):622–627, 2006.

[13] C. Cornelis, G.Deschrijver, and E. Kerre. Implication in intuitionistic fuzzy and interval-valued fuzzy set theory: construction, classification, application. *Int. J. Approx. Reasoning*, 35(1):55–95, 2004.

[14] B. De Baets and H. De Meyer. On the existence and construction of t-transitive closures. *Inf. Sci.*, 152(1):167–179, 2003.

[15] H. Bustince and P. Burillo. Mathematical analysis of interval-valued fuzzy relations: application to approximate reasoning. *Fuzzy Sets Syst.*, 113(2):205–219, 2000.

[16] H. Naessens, H. De Meyer, and De B. Baets. Algorithms for the computation of t-transitive closures. *IEEE Transactions on Fuzzy Systems*, 10(1):541–551, 2002.

ISBN: 978-989-95079-6-8

Choquet-integral-based Evaluations by Fuzzy Rules

Eiichiro Takahagi

School of Commerce, Senshu University,
Tamaku, Kawasaki 214-8580, Japan
Email: takahagi@isc.senshu-u.ac.jp

Abstract— *Choquet-integral-based evaluation models are proposed. The evaluation parameters – fuzzy measures – are assigned from a fuzzy rule table. There are three variations in this model: TF-, BP-, and AV-type models. The TF-type model is a natural extension of ordinal Choquet integrals. The BP-type model involves an evaluation using a reference point. The AV-type model involves a neutral evaluation method. These methods have a number of advantages: Output values are continuous and piecewise linear; If the fuzzy rules have monotonicity, output values are also monotone. We compare these models with some fuzzy reasoning models.*

Keywords— Choquet Integral, Fuzzy rule, Simplified fuzzy reasoning, Cumulative prospect theory, Choquet integral with respect to bi-capacities

1 Introduction

Fuzzy reasoning models are very useful tools for developing fuzzy control models. A part of a fuzzy control model involves sensing status, evaluation using fuzzy reasoning models, and producing output. By repeatedly applying the feedback process, fuzzy control models help approach the control goal (Approach Principal [1]). However, to be applied in the social sciences, evaluation models must be able to determine one global evaluation value by a single calculation process. It is important that the global evaluation values adhere obsequiously to fuzzy rules. For example, if a fuzzy rule is monotone with respect to the inputs, the output of the evaluation model must satisfy the monotonicity property with respect to the inputs. Almost all the fuzzy reasoning models do not satisfy the monotonicity property [2].

In this paper, we propose Choquet-integral-based evaluation methods by using fuzzy rules. The fuzzy rule table is of the same form as ordinal simplified fuzzy reasoning models, but the calculations use (extended) Choquet integrals instead of min-max calculations or product-sum calculations. The global evaluation values satisfy continuous and piecewise linear outputs. Moreover, if fuzzy rules are monotone respect to the inputs, the output values also satisfy the monotonicity property.

2 Definitions

2.1 Notations

X**:** Set of evaluation items (n: number of evaluation items)

x_i**:** Input value of the i^{th} item

y**:** Global evaluation value

2.2 Fuzzy Space Division Constraint

Each input i is divided into m_i fuzzy sets in which the membership functions are p_i^j for $i = 1, \ldots, n, j = 1, \ldots, m_i$. For each input item, all the membership functions satisfy the following conditions:

1. All fuzzy sets are normal and convex. The vertex of a membership function is unique $\forall i, j$, that is, there is a unique point v_i^j, where $p_i^j(v_i^j) = 1\ \forall i$ and j.

2. The sum of the membership values is 1:

$$\sum_j p_i^j(x_i) = 1, \forall x_i, i. \tag{1}$$

3. There are one or two active membership functions – $p_i^j(x_i) > 0 - \forall x_i, \forall i$.

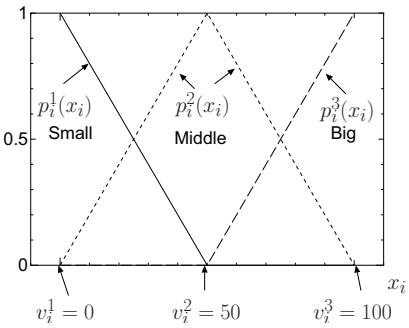


Figure 1: Fuzzy space division ($i = 1, 2$)

An example of the fuzzy space division is shown in figure 1. In this example, there are two inputs, $X = \{1, 2\}$, the membership functions are of the triangle type, and the membership functions are the same for all inputs, $v_i^1 = 0, v_i^2 = 50, v_i^3 = 100, i = 1, 2$.

2.3 Fuzzy Rule Table

Representative points $(k_1, \ldots, k_n)$ are defined as pairs of vertex numbers. For example, the representative point $(2, 3)$ $(n = 2)$ is the point at which input 1 is "Middle" and input 2 is "Big." The fuzzy rule table c is defined as a function from the representative points to the output values:

$$c : \{1, \ldots, m_1\} \times \ldots \times \{1, \ldots, m_n\} \to \mathbb{R}. \tag{2}$$

Table 1 is an example of the fuzzy rule tables,

- If input 1 is "Big" and input 2 is "Big," the output is 100.

ISBN: 978-989-95079-6-8

- If input 1 is "Big" and input 2 is "Middle," the output is 80.
- ...

Table 1: Fuzzy rule table

Input 2 \ 1	Small (1)	Middle (2)	Big (3)
Small (1)	0 (= $c(1,1)$)	50 (= $c(2,1)$)	80 (= $c(3,1)$)
Middle (2)	30 (= $c(1,2)$)	60 (= $c(2,2)$)	90 (= $c(3,2)$)
Big (3)	70 (= $c(1,3)$)	80 (= $c(2,3)$)	100 (= $c(3,3)$)

2.4 Choquet Integral and Extended Choquet Integral

The Choquet integral [3] is a useful tool for use in global evaluation methods. The integral can be used to represent complementary and substitute relations among evaluation items [4]. A fuzzy measure $\mu^\dagger$ is defined as

$$\mu^\dagger : 2^X \to \mathbb{R} \tag{3}$$

$$\mu^\dagger(\emptyset) = 0. \tag{4}$$

In ordinal fuzzy measure definitions, this definition is known as a non-monotonicity fuzzy measure, but in this paper, the monotonicity of the fuzzy measure is not assumed.

The Choquet integral is defined as

$$y = (\mathrm{C}) \int h d\mu \equiv \int_0^\infty \mu(\{x \mid h(x) > r\}) dr. \tag{5}$$

The extended fuzzy measure and extended Choquet integral [5, 6] are proposed for handling the cases in which $\mu(\emptyset) \neq 0$. An extended fuzzy measure μ is defined as

$$\mu : 2^X \to \mathbb{R}. \tag{6}$$

Since it is not assumed by definition that $\mu(\emptyset) = 0$, the integration interval is limited to $[0, 1]$. The extended Choquet integral is defined as

$$y = (\mathrm{EC}) \int h d\mu \equiv \int_0^1 \mu(\{x \mid h(x) > r\}) dr. \tag{7}$$

The extended Choquet integral can be calculated by using an ordinal Choquet integral:

$$(\mathrm{EC}) \int h d\mu = (\mathrm{C}) \int h d\mu^\dagger + \mu(\emptyset) \tag{8}$$

$$\text{where } \mu^\dagger(A) = \mu(A) - \mu(\emptyset), \forall A \in 2^X.$$

3 Segment Division Calculation Method

3.1 Segmentation and Segment Selection

Segmentation is performed at vertex points v_i^j for all inputs. Figure 2 shows an example of the segmentation. Segment $S_{(k_1,\ldots,k_n)}(k_i < m_i, \forall i)$ is the n-th rectangle whose vertices are the representative points $(k_1 + l_1, \ldots, k_2 + l_n), \forall l_i \in \{0, 1\}, i = 1, \ldots, n$. In the example, there are four segments, $S_{(1,1)}$, $S_{(2,1)}$, $S_{(1,2)}$ and $S_{(2,2)}$.

First, we select the segment that includes the input values $((x_1, \ldots, x_n))$. If $x_1 = 40$ and $x_2 = 20$, $S_{(1,1)}$ is selected.

In this model, each segment has a different extended fuzzy measure $\mu^{(k_1,\ldots,k_n)}$ and integrand $h^{(k_1,\ldots,k_n)}$, but the Choquet integral is caclculated only for the selected segment.

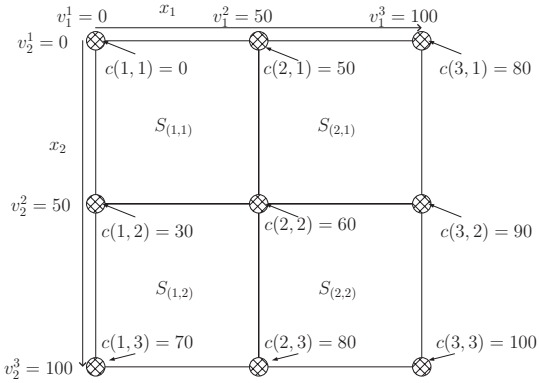


Figure 2: Segmentation

3.2 Segment Calculation

Figure 3 shows segment $S_{(1,1)}$. The output is calculated by interpolating the output values of four representative points; those values are $c(1,1), c(2,1), c(1,2)$, and $c(2,2)$.

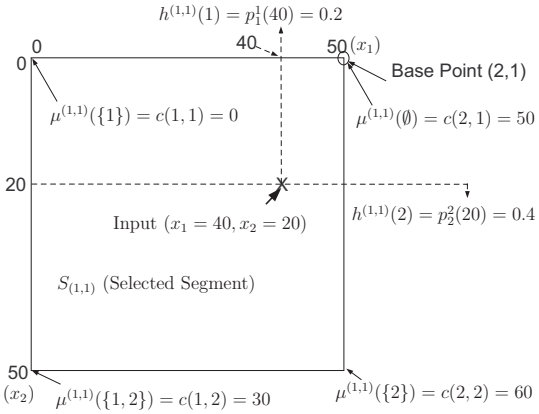


Figure 3: Calculation process (Base Point: $(2,1)$)

Base point selection The base point is the origin of the fuzzy measure and integrand for the segment. First, the base point is selected from among the representative points in the segment. Alternative base points of segment $S_{(k_1,\ldots,k_n)}$ are $(k_1 + l_1, \ldots, k_n + l_n), \forall l_i \in \{0, 1\}, i = 1, \ldots, n$.

Alternative base points of segment $S_{(1,1)}$ in figure 3 are $(1,1)$,$(2,1)$,$(1,2)$, and $(2,2)$.

Extended fuzzy measure $\mu^{(k_1,\ldots,k_n)}$ assignment When the segment $S_{(k_1,\ldots,k_n)}$ and the base point $(q_1, \ldots, q_n)$ are selected, extended fuzzy measures $\mu^{(k_1,\ldots,k_n)}$ are assigned as follows:

$$\mu^{(k_1,\ldots,k_n)}(A) = c(l_1, \ldots, l_n), \forall A \in 2^X \tag{9}$$

$$\text{where } l_i = \begin{cases} q_i & \text{if } i \notin A \\ k_i & \text{if } i \in A \text{ and } k_i \neq q_i \\ k_i + 1 & \text{if } i \in A \text{ and } k_i = q_i. \end{cases}$$

If $(2,1)$ is selected as the base point in figure 3,

$$\mu^{(1,1)}(\emptyset) = c(2,1),\ \mu^{(1,1)}(\{1\}) = c(1,1)$$
$$\mu^{(1,1)}(\{2\}) = c(2,2),\ \mu^{(1,1)}(\{1,2\}) = c(1,2)$$

ISBN: 978-989-95079-6-8

Integrand $h^{(k_1,\ldots,k_n)}$ assignment When the segment $S_{(k_1,\ldots,k_n)}$ and the base point $(q_1,\ldots,q_n)$ are selected, integrands $h^{(k_1,\ldots,k_n)}$ are assigned as follows:

$$h^{(k_1,\ldots,k_n)}(i) = p_i^j(x_i), i = 1,\ldots,n \quad (10)$$
$$\text{where } j = \begin{cases} k_i & \text{if } k_i \neq q_i \\ k_i + 1 & \text{otherwise.} \end{cases}$$

If $(2,1)$ is selected as the base point of segment $S_{(1,1)}$,

$$h^{(1,1)}(1) = p_1^1(x_1), \;\; h^{(1,1)}(2) = p_2^2(x_2) \quad (11)$$

Extended Choquet integral The output value y is calculated from the extended Choquet integral by using the fuzzy measure $\mu^{(k_1,\ldots,k_n)}$ and integrand $h^{(k_1,\ldots,k_n)}$ of the selected segment.

$$y = (\mathrm{EC}) \int h^{(k_1,\ldots,k_n)} d\mu^{(k_1,\ldots,k_n)} \quad (12)$$

The output values change with the methods for selection the base points. We propose the following type of Choquet-integral-based evaluation methods: TF-, BP- and BP-type.

3.3 TF-type

In TF-type models, the smallest (upper left) representative points are selected as base points for all segments, that is, the base point of a segment $S_{(k_1,\ldots,k_n)}$ is $(k_1,\ldots,k_n)$.

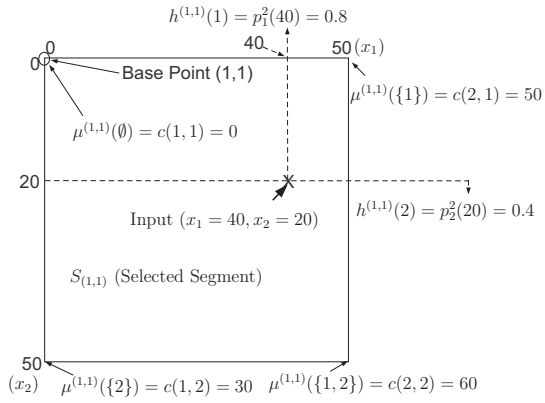


Figure 4: Extended fuzzy measure assignment in TF-type models

When $x_1 = 40$ and $x_2 = 20$, from figure 2, segment $S_{(1,1)}$ is selected. In the TF-type model, $(1,1)$ is selected as the base point. Figure 4 shows an example of the calculation process. Since the origin is $(1,1)$, the membership functions "Middle" ($p_1^2(x_1)$ and $p_2^2(x_2)$) are used. Therefore, the integrand $h^{(1,1)}(1) = p_1^2(x_1) = 0.8$ and $h^{(1,1)}(2) = p_2^2(x_2) = 0.4$. Since the origin is $(1,1)$, $\mu^{(1,1)}(\emptyset) = c(1,1) = 0$, $\mu^{(1,1)}(\{1\}) = c(2,1) = 50$, $\mu^{(1,1)}(\{2\}) = c(1,2) = 30$, and $\mu^{(1,1)}(\{1,2\}) = c(2,2) = 60$ (figure 4). The output is calculated as

$$y = (\mathrm{EC}) \int h^{(1,1)} d\mu^{(1,1)}. \quad (13)$$

Since $\mu^{11}(\emptyset) = 0$, this equation is calculated by using the ordinal Choquet integral.

$$y = [h^{(1,1)}(1) - h^{(1,1)}(2)]\mu^{(1,1)}(\{1\}) + h^{(1,1)}(2)\mu^{(1,1)}(\{1,2\}) = 44. \quad (14)$$

Table 2 lists integrands and extended fuzzy measures for all segments. Figure 5 shows a graph corresponding to Table1.

Table 2: Integrand and extended fuzzy measure for TF ($n = 2$)

		Integrand		μ			
Segment	Base Point	$h(1)$	$h(2)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$S_{(1,1)}$	$(1,1)$	$p_1^2(x_1)$	$p_2^2(x_2)$	$c(1,1)$	$c(2,1)$	$c(1,2)$	$c(2,2)$
$S_{(2,1)}$	$(2,1)$	$p_1^3(x_1)$	$p_2^2(x_2)$	$c(2,1)$	$c(3,1)$	$c(1,2)$	$c(3,2)$
$S_{(1,2)}$	$(1,2)$	$p_1^2(x_1)$	$p_2^3(x_2)$	$c(1,2)$	$c(2,2)$	$c(1,3)$	$c(2,3)$
$S_{(2,2)}$	$(2,2)$	$p_1^3(x_1)$	$p_2^3(x_2)$	$c(2,2)$	$c(3,2)$	$c(2,3)$	$c(3,3)$

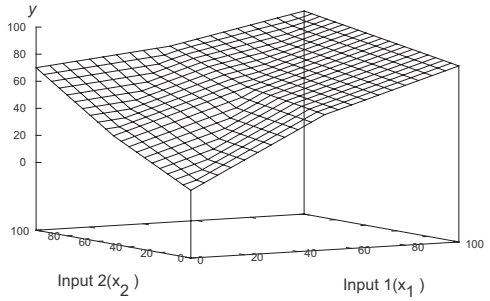


Figure 5: 3D graph corresponding to Table1 (TF-Type)

3.4 Base Point Change

The output values y are dependent on the selection of the base point. If the base point $(2,1)$ is selected for segment $S_{(1,1)}$, the extended fuzzy measure $\mu^{(1,1)}$ is

$$\mu^{(1,1)}(\emptyset) = c(2,1) = 50, \;\; \mu^{(1,1)}(\{1\}) = c(1,1) = 0,$$
$$\mu^{(1,1)}(\{2\}) = c(2,2) = 60, \;\; \text{and } \mu^{(1,1)}(\{1,2\}) = c(1,2) = 30.$$

If the input values are $x_1 = 40, x_2 = 20$, $h^{(1,1)}(1) = p_1^1(x_1) = 0.2$ and $h^{(1,1)}(2) = p_2^2(x_2) = 0.4$. The output value is

$$y = (\mathrm{EC}) \int h^{(1,1)} d\mu^{(1,1)} = 48,$$

which is not the equal to the output value 44 given by the TF-type models (14).

3.5 BP-type

In cumulative prospect theory [7], the evaluation methods differ for values above and below the reference point. The model has been developed for analyzing such situations. BP-type calculations can represent the generalized Choquet integral with respect to bi-capacity [8, 9], and the generalized Choquet integral is an extension of cumulative prospect theory.

In the BP-type model, each input has three membership functions: negative, reference, and positive; $m_i = 3, \forall i$. In prospect theory, it is important to evaluate the distance from the reference point, and therefore, the base point for the all segments is the reference point (figure 6 and table 4), namely $(2,\ldots,2)$.

ISBN: 978-989-95079-6-8

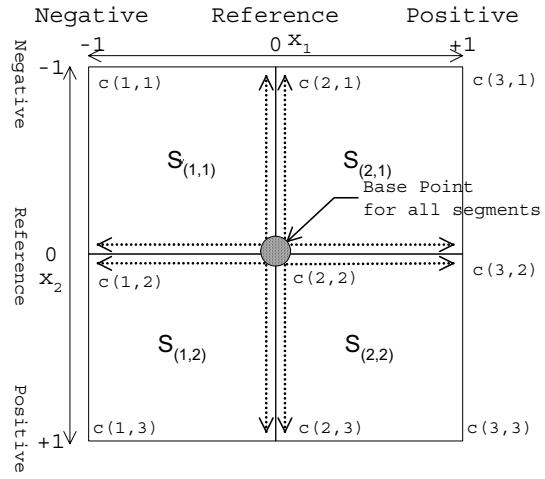


Figure 6: BP-type

BP type model Table 3 is an example of a BP-type evaluation rule table. In thoose BP-type models, membership functions are of the triangle type, such as figure 1; the base points for all segments is the reference point $(2,\ldots,2)$, and the output of the reference point is 0, that is, $c(2,\ldots,2) = 0$ and $x_i \in [-1,1]$. The integrands $h^{(k_1,\ldots,k_n)}$ are

$$h^{(k_1,\ldots,k_n)}(i) = \mid x_i \mid, \forall(k_1,\ldots,k_n), i = 1,\ldots,n. \quad (15)$$

Table 4 lists the integrands and extended fuzzy measures for all segments when $n = 2$. If $x_1 = 0.2$ and $x_2 = -0.6$, $S_{(2,1)}$ is selected and $h^{(2,1)}(1) = 0.2$ and $h^{(2,1)}(2) = 0.6$;

$$y = (\mathrm{EC}) \int h^{(2,1)} d\mu^{(2,1)} = -0.44.$$

Figure 7 shows a graph corresponding to table 3.

Table 3: BP-type fuzzy rule table

Input 2 \ 1	Negative	Reference	Positive
Negative	-1.0	-0.8	-0.6
Reference	-0.8	0	$+0.6$
Positive	-0.6	$+0.8$	$+1.0$

Table 4: Integrands and extended fuzzy measures for BP

		Integrands		μ			
Segments	Base Points	$h(1)$	$h(2)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$S_{(1,1)}$	$(2,2)$	$-x_1$	$-x_2$	$c(2,2)$	$c(1,2)$	$c(2,1)$	$c(1,1)$
$S_{(2,1)}$	$(2,2)$	x_1	$-x_2$	$c(2,2)$	$c(3,2)$	$c(2,1)$	$c(3,1)$
$S_{(1,2)}$	$(2,2)$	$-x_1$	x_2	$c(2,2)$	$c(1,2)$	$c(2,3)$	$c(1,3)$
$S_{(2,2)}$	$(2,2)$	x_1	x_2	$c(2,2)$	$c(3,2)$	$c(2,3)$	$c(3,3)$

Bi-capacity Model Let $Q(X) = \{(A,B) \in 2^X \times 2^X \mid A \cap B = \emptyset\}$. Bi-capacity [8] is defined as a function $v : Q(X) \to [-1,1]$ that satisfies the following conditions:

- If $A \subset A'$, $v(A,B) \leq v(A',B)$ and if $B \subset B'$, $v(A,B) \geq v(A,B')$,
- $v(\emptyset,\emptyset) = 0$ and

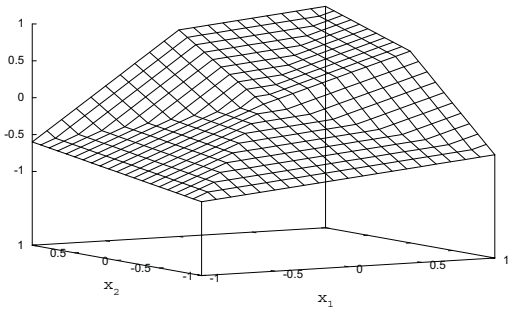


Figure 7: Outputs corresponding to Table 3 (BP-type))

- $v(X,\emptyset) = 1, v(\emptyset,X) = -1$.

The Choquet integral with respect to bi-capacity[9] is defined as

$$C_v(x) = \sum_{i=1}^{n} \mid x_{\sigma(i)} \mid [v(A_{\sigma(i)} \cap X^+, A_{\sigma(i)} \cap X^-) - v(A_{\sigma(i+1)} \cap X^+, A_{\sigma(i+1)} \cap X^-)] \quad (16)$$

where $X^+ \equiv \{i \in X \mid x_i \geq 0\}$, $X^- = X \setminus X^+$ and σ is a permutation on X satisfying $\mid x_{\sigma(1)} \mid \leq \ldots \leq \mid x_{\sigma(n)} \mid$ and $A_{\sigma(i)} \equiv \{\sigma(i),\ldots,\sigma(n)\}$.

BP type and bi-capacity model If a fuzzy rule table c is assigned from a bi-capacity v in the following manner:

$$c(k_1,\ldots,k_n) = v(A,B), \quad (17)$$
$$\text{where } A = \{i \mid k_i = 3\}, B = \{i \mid k_i = 1\},$$

the output y is equal to $C_v(x)$. Table 5 shows the correspondence when $n = 2$.From (17), the extended fuzzy measures are as follows:

$$\mu^{(k_1,\ldots,k_n)}(A) = v(E,F), \forall(k_1,\ldots,k_n), \forall A \in 2^X \quad (18)$$

where $E = \{i \mid k_i = 2\} \cap A$ and $F = \{i \mid k_i = 1\} \cap A$

For the input $x_1,\ldots,x_n$, the selected segment is $(k_1,\ldots,k_n)$, where

$$k_i = \begin{cases} 1 & \text{if } x_k < 0 \\ 2 & \text{otherwise} \end{cases} \quad (19)$$

Since $h^{(k_1,\ldots,k_n)}(i) = \mid x_i \mid$, $\mu^{(k_1,\ldots,k_n)}(\emptyset) = 0$, and from (18) and (19), $\mu^{(k_1,\ldots,k_n)}(A_{\sigma(i)}) = v(A_{\sigma(i)} \cap X^+, A_{\sigma(i)} \cap X^-)$,

$$\begin{aligned} y &= (\mathrm{C}) \int h^{(k_1,\ldots,k_n)} d\mu^{(k_1,\ldots,k_n)} \\ &= \sum_{i=1}^{n} \mid x_{\sigma(i)} \mid [\mu^{(k_1,\ldots,k_n)}(A_{\sigma(i)}) - \mu^{(k_1,\ldots,k_n)}(A_{\sigma(i+1)})] \\ &= C_v(x). \end{aligned}$$

3.6 AV-type

TF-type and BP-type calculations depend on the selection of the base points. Since the output value in the AV-type model is the average value for all base point selections, this output

ISBN: 978-989-95079-6-8

Table 5: Fuzzy rule table by a bi-capacity ($n = 2$)

Input 2 \ 1	Negative	Reference	Positive
Negative	$v(\emptyset, \{1,2\})$	$v(\emptyset, \{2\})$	$v(\{1\}, \{2\})$
Reference	$v(\emptyset, \{1\})$	$v(\emptyset, \emptyset)$	$v(\{1\}, \emptyset)$
Positive	$v(\{2\}, \{1\}))$	$v(\{2\}, \emptyset)$	$v(\{1,2\}, \emptyset)$

does not depend on the base point selection. Therefore, the AV-type method is a neutral evaluation method in terms of the selection of base points.

If $n = 2$, there are four options for selections the base point for each segment. For $S_{(1,1)}$, $(1,1)$, $(2,1)$, $(1,2)$, and $(2,2)$ are the options for selecting the base points. Table 6 lists the integrands and extended fuzzy measures for each base point selection.

Table 6: Integrand and extended fuzzy measure for AV ($S_{(1,1)}$)

	Integrand		μ			
Base Point	$h(1)$	$h(2)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$(1,1)$	$p_1^2(x_1)$	$p_2^2(x_2)$	$c(1,1)$	$c(2,1)$	$c(1,2)$	$c(2,2)$
$(2,1)$	$p_1^1(x_1)$	$p_2^2(x_2)$	$c(2,1)$	$c(1,1)$	$c(2,2)$	$c(1,2)$
$(1,2)$	$p_1^2(x_1)$	$p_2^1(x_2)$	$c(1,2)$	$c(2,2)$	$c(1,1)$	$c(2,1)$
$(2,2)$	$p_1^1(x_1)$	$p_2^1(x_2)$	$c(2,2)$	$c(1,2)$	$c(2,1)$	$c(1,1)$

The calculation process when $x_1 = 30$, $x_2 = 40$, and the rule in table 1 is followed is given in Table 7. The output value of AV-type model, 44, is the average value for four selected base points. Figure 8 shows the graph of the AV-type outputs.

Table 7: Calculation process for $x_1 = 30$ and $x_2 = 40$

	Integrand		μ				output
Base Point	$h(1)$	$h(2)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$	y
$(1,1)$	0.6	0.8	0	50	30	60	42
$(2,1)$	0.4	0.8	50	0	60	30	46
$(1,2)$	0.6	0.2	30	60	0	50	46
$(2,2)$	0.4	0.2	60	30	50	0	42
			Average (y)				44

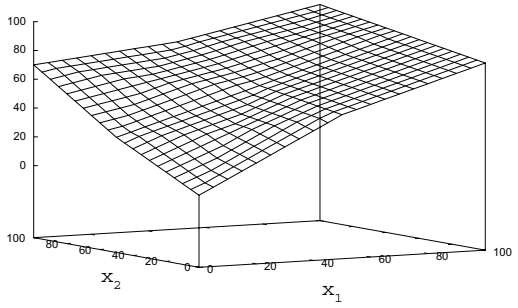


Figure 8: AV-type corresponding to Table 1

4 Properties

Continuity Choquet integral models satisfy the continuity property. Therefore, the proposed models satisfy the continuity property in each segment. For calculating the output values of a point on a bordering hyperplane, there are two or more options for selecting the calculation segment. However, in TF-, BP- and AV-type models, the output values remains the same regardless of the selected segment, that is, the output value on a bordering hyperplane does not depend on the selection of segments (The proof is omitted.) Therefore, Choquet integral based evaluations satisfy the continuity property.

Monotonicity It is known that almost all fuzzy reasoning models do not satisfy the monotonicity property when the fuzzy rule table is monotone. Choquet-integral-based evaluations satisfy the monotonicity property when the fuzzy rule table is monotone. A monotone fuzzy rule table is defined as $c(k_1, \ldots, k_n) \geq c(k'_1, \ldots, k'_n), \forall k_i \geq k'_i$. Examples of monotone rule tables are Table 1 and 3. When the fuzzy rule table is monotone, it is easy to confirm that the values of the extended Choquet integral are monotone and continuous for each segment. From the continuity of the proposed model, it follows that the Choquet-integral-based evaluation models have monotonicity if the fuzzy rule table is monotone.

Piecewise linear If the all membership functions are linear in that case in figure 1, the outputs are piecewise linear. This property is derived from the piecewise linearity of the Choquet integral.

5 Comparisons with Fuzzy Reasoning models

Fuzzy-rule-based global evaluation methods have been developed by using fuzzy reasoning models [10] such as "Min-Max Gravity Method," "Product-Sum Gravity Method," and "Simplified Fuzzy Reasoning." The outputs of the fuzzy rules for the min-max gravity method and product-sum gravity method are fuzzy sets. Therefore, to obtain the output (non fuzzy) value, a defuzzfication (gravity) calculation is performed. However, these calculations are complex. To analyze properties, the simplified fuzzy reasoning is utilized.

Simplified fuzzy reasoning is based on the product-sum method [11]. In the simplified method, it is possible to use min-max calculations instead of product-sum calculations. To demonstrate the properties of min-max and product-sum calculations, we utilize the two types of simplified fuzzy reasoning models.

5.1 Product-sum-type Simplified Fuzzy Reasoning

Simplified fuzzy reasoning is the most popular type of fuzzy reasoning. If the membership function shown in figure 1 and monotone fuzzy rules are assumed, product-sum-type simplified fuzzy reasoning satisfies the monotonicity property.

Product-sum-type simplified fuzzy reasoning can also be perfoermed using the segment division method. If $S_{(1,1)}$ in figure 2 is selected, the output given by the method is

$$\begin{aligned} y = & [1 - p_1^2(x_1)][1 - p_2^2(x_2)]c(1,1) \\ + & p_1^2(x_1) \times [1 - p_2^2(x_2)]c(2,1) \\ + & [1 - p_1^2(x_1)]p_2^2(x_2)c(1,2) + p_1^2(x_1)p_2^2(x_2)c(2,2) \end{aligned} \quad (20)$$

This equation involves an expectation calculation, that is, $p_1^2(x_1)$ and $p_2^2(x_2)$ are random variables, and the expected value of four events ($c(1,1), c(2,1), c(1,2)$, and $c(2,2)$) is calculated by this eqation.

In the segment, if the fuzzy rules are monotone, the outputs satisfy the monotonicity and continuities. Therefore, product-sum methods with monotone fuzzy rules satisfy the monotonicity property.

ISBN: 978-989-95079-6-8

Since product-sum-type calculations involve random variables calculation, output values vary unpredictably. Table 8 lists the AND-type fuzzy rule "If Input 1 and 2 are big, the output is 1." Figure 9 shows the output values for the rule table. Product-sum-type calculations are nonlinear because of product calculations. The input variation from $x_1 = x_2 = 0$ to $x_1 = x_2 = 0.01$ increases the output value by only 0.001, but the variation from $x_1 = x_2 = 0.99$ to $x_1 = x_2 = 1$ increases the output value by 0.0199. These results are inconsistent with the rule.

Table 8: AND-type rule table

Input 2 \ Input 1	Small	Big
Small	0.0	0.0
Big	0.0	1.0

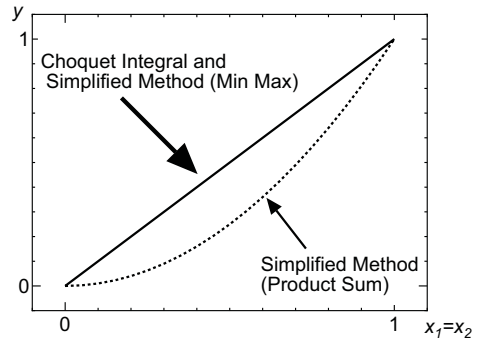


Figure 9: Comparison with product-sum-type method

Since $x_1 = x_2 = 0.5$ only satisfies half of the rule condition, it is expected that the output value is 0.5, but while the product-sum-type is 0.25. Moreover, if the number of inputs for the condition is increased, the output given by the product-sum method decreases. For example, if the number of inputs is 3 or 4, the product-sum output is $0.5^3 = 0.125$ or $0.5^4 = 0.0625$, respectively.

5.2 Min-max type Simplified Fuzzy Reasoning

In ordinal simplified fuzzy reasoning, product-sum calculation is performed, but it is possible to use min and max operators instead of product and sum operators.

Table 9: Monotone fuzzy rule table

Input 2 \ Input 1	Small	Big
Small	0.6 ($c(1,1)$)	0.7 ($c(2,1)$)
Big	0.8 ($c(1,2)$)	1.0 ($c(2,2)$)

Table 9 is represented by the fuzzy switching function

$$y = (c(1,1) \wedge x_1^1 \wedge x_2^1) \vee (c(2,1) \wedge x_1^2 \wedge x_2^1) \vee (c(1,2) \wedge x_1^1 \wedge x_2^2) \vee (c(2,2) \wedge x_1^2 \wedge x_2^2),$$

where $x_1^1 + x_1^2 = 1$ and $x_2^1 + x_2^2 = 1$. The equation does not satisfy the monotonicity property because fuzzy switching functions do not satisfy the complementary laws. For example, if $x_1^2 = 0.5$ and $x_2^2 = 1.0$, $y = 0.5$. However, if $x_1^2 = 0.4$ and $x_2^2 = 1.0$, $y = 0.6$.

Therefore, if fuzzy-value calculations are necessary, it is preferable to use Choquet-integral-based calculations.

6 Conclusions

This model is very useful for applications in social science, such as performance evaluation and educational evaluations. Since social science applications require strict validity of the output values, the global evaluation functions should not exhibit non-monotonicity and unexpected nonlinearity, such as that seen in figure 9. The Choquet integral models can be used to ensure the validity of the outputs. However, the attitude of evaluators might change according to the situation or the input values, and this might change the weights and the relationships among the evaluation parameters. The Choquet integral models keep those validities. Fuzzy rule tables can be used to represent these attitudes. Therefore, the proposed model inherits the merits of both the Choquet integral model and the fuzzy rule table.

References

[1] N. Nakajima, Inference, Reasoning, and Fuzzy Reasoning, *Journal of Japan Society for Fuzzy Theory and Systems*, Vol.6, No.6, 1105-1114, 1994.

[2] Y. Okuda et al, Monotonicity Analysis of Fuzzy Reasoning, *Waseda journal of mathematics education*, Vol.21, No.1, 28-33, 2003.

[3] G. Choquet, Theory of Capacities, Annales de l'Institut Fourier 5, 131-295, 1954.

[4] M. Grabisch et al, Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference, Kluwer, 1995.

[5] E.Takahagi, Fuzzy Integral Based Fuzzy Switching Functions, in *Transactions on Rough Sets II. Lecture Notes in Computer Science 3135 edited by James F. Peters et al*, pp.129-150, Springer, 2004.

[6] E. Takahagi, Fuzzy three-valued switching functions using Choquet integral, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.7 No.1, 47-52, 2003.

[7] A. Tversky and D. Kahneman, Advances in Prospect Theory: Cumulative Representation of Uncertainty, *J. of Risk and Uncertainty*, 5, 297-323, 1992.

[8] M. Grabisch and Ch. Labreuche, Bi-capacities — Part I: definition, Möbius transform and interaction, *Fuzzy Sets and Systems*, 151, 211-236, 2005.

[9] M. Grabisch and Ch. Labreuche, Bi-capacities — Part II: the Choquet integral, *Fuzzy Sets and Systems*, 151, 237-259, 2005.

[10] E.H. Mamdani, Applications of fuzzy algorithm for control of a simple dynamic plant, *Proc. of IEEE* ,121,1585-1588, 1974.

[11] M. Mizumoto, Product-Sum-Gravity Method = Fuzzy Singleton-type Reasoning Method = Simplified Fuzzy Reasoning Method, *Proc. of Fifth IEEE International Conference on Fuzzy Systems*, 2098-2102, 1996.

ISBN: 978-989-95079-6-8

Fuzzy modelling of sensor data for the estimation of an origin-destination matrix

Krystyna Biletska[124] Sophie Midenet[4] Marie-Hélène Masson[23] Thierry Denœux[12]

1. Université de Technologie de Compiègne
2. UMR CNRS 6599 Heudiasyc,
BP 20529, 60205 Compiègne cedex, France
3. Université de Picardie Jules Verne, IUT de l'Oise
4. Université Paris-Est, Institut National de Recherche sur les Transports et leur Sécurité (INRETS), Laboratoire GRETIA
2 rue de la Butte Verte, 93166 Noisy le Grand cedex, France
Email: krystyna.biletska@inrets.fr (corresponding author)

Abstract— *This paper examines a short-time estimation problem of an origin-destination (OD) matrix, where each element is a volume of vehicle flow between one of the OD pair of zones of a signalised junction. The estimation is based on the use of traffic measurements provided by video sensors and on the knowledge of the traffic lights. This data is subject to redundancy, imprecision and uncertainty. The main purpose of this paper is to obtain the best estimates of the OD matrix by modelling the data imperfection, using a two-step method. First, relationships between the observed data are built in real-time using High-Level Petri Nets. Due to the imperfection of data the system obtained is underdetermined and inconsistent. Second, the fuzzy sets theory is used to model this imperfection and to overcome the inconsistency of the system.*

Keywords— Fuzzy least squares, Fuzzy linear programming, Fuzzy modelling, Origin-destination matrix

1 Introduction

Knowledge about origins and destinations (OD) of vehicles in a road network is important in most transport systems. In the case of a junction, its mathematical representation is OD matrix $\mathbf{B}$, each element b_{ij} of which is a proportion of the flow of vehicles that come from entrance (origin) i and go to exit (destination) j. Such a proportion is called the *OD flow rate*. Since the OD matrix changes in time following the changes in traffic demand, the estimation period of the OD matrix has to be as short as possible. In particular, the period should be equal to a traffic light cycle (duration of green-amber-red sequence) when we deal with a signalised junction. At INRETS/GRETIA we use such a short-time estimation of the OD matrix as part of a diagnostic system for signalised junctions [1]. This system compares the impacts of different traffic control strategies expressed in terms of CO_2 and pollutant emissions.

The OD matrix is generally deduced from vehicle counts made on each entrance and each exit of the junction during a given time interval. These counts are usually provided by magnetic loops embedded in the road surfaces and sensitive to metallic masses. The estimation can be obtained from a conservation law of vehicles which is a set of relationships between exit and entrance flow counts. In general, when loops are installed on every entrance and exit of the junction, the estimation problem is underdetermined. Thus a solution is not unique and additional information such as a *prior* OD matrix is used to choose the OD matrix which corresponds best to the actual matrix. Some of the existing methods are based on the information minimisation principle [2], on maximisation of likelihood [3] or on Bayesian inference [4]. Other methods propose a recursive estimation of the OD matrix [5, 6]. Furthermore, the estimation problem can be represented as a constraint optimisation problem [7, 8]. Methods which use traffic lights are described in [9, 10].

This paper considers the problem of reconstituting the origins and destinations of vehicle flows crossing a signalised junction, at each traffic light cycle. OD flow volumes are estimated using traffic lights and traffic measurements from video cameras installed at the junction. These measurements, provided every second, are the vehicle counts made on each entrance and exit of the junction and the number of vehicles stopped at each inner section of the junction. The data is subject to redundancy, uncertainty and inaccuracy. Such an imperfection is linked to the reliability of video sensors, measurement conditions, traffic characteristics and drivers' behaviour.

In order to take into account all available information it is necessary to consider the nature and possible interdependence of the data. None of the cited methods takes into account the lake of imprecision of vehicle counts, the possible physical complexity of the junction and the traffic lights at the same time. Moreover they cannot be applied to the problem because the period of estimation is quite short.

The main purpose of this paper is to obtain the best estimates of the OD matrix by modelling the data imperfection, using a two-stage method. First, a conservation law of vehicles, which is represented by an inconsistent and underdetermined system of equations, is built by High-Level Petri Nets at each traffic light cycle. Second, the fuzzy sets theory is used to overcome the inconsistency of the system and to model the data imperfection. Three different approaches have been analysed to solve this system of equations: ordinary least squares, fuzzy least squares and fuzzy linear programming. A numerical study of the proposed methods has been done using the data collected in a real experimental junction fitted out with video cameras and a traffic light controller.

The rest of the paper is organised as follows. Section 2

ISBN: 978-989-95079-6-8

presents the problem and describes the real data. Sections 3 and 4 introduce the fundamental principles of High-Level Petri Nets and show how they are applied to model the OD flows through the experimental site. Section 5 proposes three methods for OD matrix estimation and the paper concludes by proposing further lines of research.

2 The problem

The experimental site is an isolated signalised junction of two double-lane roads situated in the south suburb of Paris (Fig. 1). The main road B-D which connects the suburbs to Paris has a high traffic volume, whereas the road A-C has lower traffic volume. Traffic lights control four incoming links and four inner zones. Note that right-turning vehicles use special lanes and are not taken into account in this study. Only eight OD flows are statistically significant: $AC, AD, BD, BB, CB, DB, DC, DD$.

Eight video cameras are installed at the junction in order to capture all the entrance and exit links and the inner zones. The location, height and angle of each camera depend on the geometry of the junction and are chosen to favour the measurement of space traffic parameters such as queue length on incoming links. The camera views are analysed in real time using image processing techniques developed at INRETS [11]. They provide several measurements every second:

- $\mathcal{X}_i(\tau)$ vehicle counts measured at the end of an entrance i at second τ ($i = 1, \ldots, n$),
- $\mathcal{Y}_j(\tau)$ number of vehicles that have passed through the beginning of the exit j at second τ ($j = 1, \ldots, m$),
- $\mathcal{Z}_k(\tau)$ number of stopped vehicles at inner zone k at second τ ($k = 1, \ldots, p$),

where n, m and p are the numbers of entrances, exits and inner zones respectively. Here, a *traffic light cycle* is a period of time between two sequential onsets of the red light on the main road B-D .

Many phenomena influence the quality measurement of traffic parameters. Traffic conditions (peak or off-peak periods) are the reason for many traffic count errors. If the traffic flow is heavy, the gaps between vehicles are small and it is difficult to distinguish these gaps on the video images. Thus the number of vehicles measured is lower than the actual number.

The characteristics of vehicles are also a source of measurement errors. High vehicles passing in front of camera will hide the smaller vehicles or the whole camera field, i.e. they will produce a *masking effect*. Two-wheeled vehicles are only seldom counted because they are small. The heterogeneous colours of vehicle roofs also add to the problem of detection.

Meteorological conditions inevitably have an influence on all types of traffic measurements and the video are blurred: the wind shakes the posts the cameras are fixed to, the sun's rays cause the reflections on the vehicle surfaces and camera lenses, rain, snow and fog obscure a camera field. Changes in brightness caused by the position of the sun, clouds and headlights at night also determine the reliability of the measurements.

Let $x_i(c)$ be the flow volume at entrance i during a traffic light cycle c and $y_j(c)$ be the flow volume which entered the junction during cycle c and leaves it by exit j. *OD flow rate*

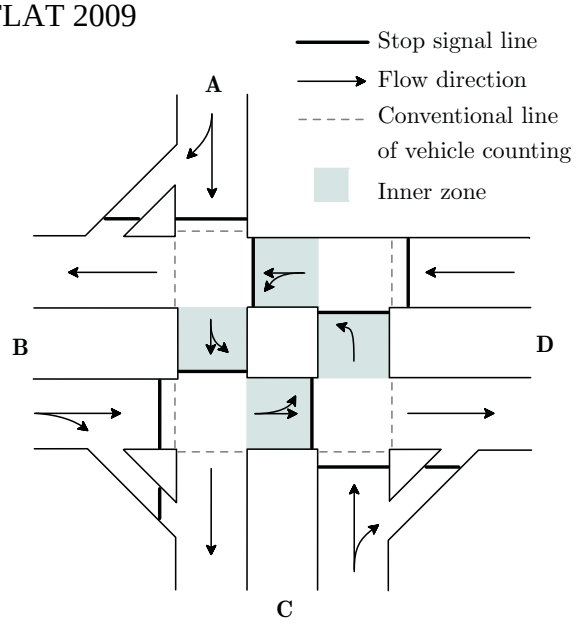


Figure 1: The experimental junction

b_{ij} is the proportion of the flow of vehicles that come from entrance i and go to exit j. The problem is to estimate OD flow rates b_{ij} ($\forall i \in [\![1, n]\!], \forall j \in [\![1, m]\!]$) at the end of each traffic light cycle c, such that

$$y_j(c) = \sum_{i=1}^{n} b_{ij}(c) x_i(c), \tag{1a}$$

$$x_i(c) b_{ij}(c) \geq z_{kij}(c) \quad \forall k \text{ s.t. } \delta_{kij} = 1, \tag{1b}$$

$$\sum_{j=1}^{m} b_{ij}(c) = 1, \tag{1c}$$

$$b_{ij}(c) \geq 0, \tag{1d}$$

where $z_{kij}(c)$ is the number of vehicles which cross the junction from i to j and stop at inner zone k during cycle c, $\delta_{kij} = 1$ if OD flow from i to j can pass through inner zone k and is 0 otherwise. For a given cycle c the value of variable $x_i(c)$ can be obtained from instantaneous vehicle counts

$$x_i(c) = \sum_{\tau=1}^{\mathcal{G}_i(c)} \mathcal{X}_i(\tau),$$

where $\mathcal{G}_i(c)$ is a duration of the green light of cycle c in entrance i. The values of $y_j(c)$ and $z_{kij}(c)$ cannot be obtained directly from traffic measurements, because it is impossible to know the period of time when the vehicle flow $x_i(c)$ leaves the junction or stops at inner zones.

This paper proposes a method to obtain the values of $y_j(c)$ and $z_{kij}(c)$ from $\mathcal{Y}_j(\tau)$ and $\mathcal{Z}_k(\tau)$ respectively and then to estimate the OD flow rates. First, a tool for vehicle flow segmentation is built via two High-Level Petri Nets (HLPN). The first net indicates the set of vehicle flows which may be present in each zone at any given second. It makes possible to associate these flows to the measurements taken in the corresponding zones and provides the beginnings of the flows. The second net provides the ends of the flows. For a given cycle we thus know the duration of the presence of the flows in each zone

ISBN: 978-989-95079-6-8

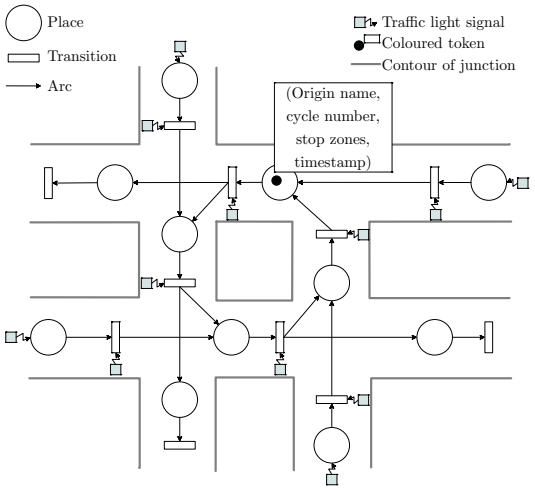


Figure 2: High Level Petri Net 1

and can collect the corresponding measurements. A consistent underdetermined system of equations, whose unknowns are the OD flow rates, can thus be built dynamically and solved at each traffic light cycle.

3 Fundamentals of High-Level Petri Nets

A High-Level Petri Net (HLPN) is a graphical and mathematical tool used to model discrete-event systems [12, 13]. In this paper, the HLPN is a classical Petri Net [14] complemented by the notions of time and colour. In graphic representation, an HLPN is a directed graph composed of two kinds of nodes: *places*, drawn by circles, and *transitions* represented as bars. The arcs connect either a place to a transition or, inversely, a transition to a place. A place may be *marked* by coloured *tokens* endowed with timestamps and characterises the system state. A transition represent an event that can change the state of the system modelled. The set of input and output places of a transition is also interpreted as a set of pre- and post-conditions of an event. When the event occurs the transition is *enabled* and the marking of input and output places changes. Thereby, the dynamic behaviour of the system is expressed by means of time-varying marking.

A 7-tuple $N = (\Sigma, \mathcal{P}, \mathcal{T}, \mathcal{A}, \mathcal{C}, \mathcal{W}, \mathcal{F})$ is a formal representation of HLPN, where:

- Σ is a finite non-empty colour set,
- $\mathcal{P}$ and $\mathcal{T}$ are the finite sets of places and transitions respectively, such that $\mathcal{P} \cap \mathcal{T} = \emptyset$ and $\mathcal{P} \cup \mathcal{T} \neq \emptyset$,
- $\mathcal{A} \subset (\mathcal{P} \times \mathcal{T}) \cup (\mathcal{T} \times \mathcal{P})$ is the finite set of arcs,
- $\mathcal{C} : \mathcal{P} \rightarrow 2^{\Sigma}$ is a *colour function* which associates a set of colours to a place,
- $\mathcal{W} : \mathcal{A} \rightarrow 2^{\Sigma}$ is an *arc function*.
- $\mathcal{F}$ is a delay associated to an arc.

A token is represented by a triple (p, s, θ), where $p \in \mathcal{P}$ is the place marked by the token, $s \subseteq \mathcal{C}(p)$ is a colour set which describes the system state in p, called *colour*, and θ is the time at which the token becomes available and the system state changes, called *timestamp*. Marking $\mathcal{M}(p, \tau)$ of place $p \in \mathcal{P}$ at time τ is a multiset represented by the pair $(S, \mathbf{n})$ where S is the set of colours of tokens arrived in p at $\tau = \theta$ and $\mathbf{n}$ is the vector every element of which is the number of occurrences of the token's colour s in set S.

Function $\mathcal{W}(p, t)$, corresponding to the arc from output place p to t, is a multiset of colours that are elements of the set $\mathcal{C}(p)$. The function $\mathcal{W}(t, p)$ is defined similarly.

Transition $t \in \mathcal{T}$ is enabled if $\mathcal{M}(p, \tau) \supseteq \mathcal{W}(p, t)$. It removes $\mathcal{W}(p, t)$ tokens from each of its input places p of t and adds $\mathcal{W}(t, p)$ tokens to its output places p with delay $\mathcal{F}(t, p)$ associated to an arc from t to p.

4 Modelling the traffic flows crossing a signalised junction

In this paper two HLPNs are used to model and segment the OD flows of vehicles crossing the junction in order to built the vehicle conservation law in a dynamic way. In the first net, a place stands for a zone of the experimental junction (Fig. 2). A token in a place is represented by a colour set $s = (s_1, s_2, s_3)$, where $s1 \in \{A, B, C, D\}$ is the name of flow origin, $s2 \in \mathbb{N}$ is a cycle number and $s3$ is a set of stop zones. In addition, it has a timestamp θ. A token in place p means the possible presence of a flow of origin $s1$ in the zone corresponding to p. Some transitions are related to the change of traffic lights in entrances and inner zones, other transitions represent the flow departures. An arc function $\mathcal{W}(t, p)$ is defined over a set of colours $s1$ representing the origins of the flows that pass through this arc. Delay $\mathcal{F}(t, p)$ is an average vehicle travel time between two successive zones of the junction corresponding to the input and output places of transition t.

The rules for changing colours of tokens, enabling transitions and marking places are defined as follows. During the green light at entrance i of the junction, the corresponding place is marked by one token with colour $s = (s1, s2, s3)$, $s_3 = \emptyset$, every second. Transition t associated to entrance i becomes enabled and the token is transmitted to output places of t. The transition associated with an inner traffic light is enabled if its input place contains at least one token and the corresponding traffic light is green. The enabled transition transmits the token downstream according to $\mathcal{W}(t, p)$ and $\mathcal{F}(t, p)$. If the light in inner zone k is red, tokens are stacked in the corresponding place. Their colours are changed by adding the name of inner zone k to the colour set s_3. At the beginning of the green light these tokens are shifted to the exit place p with predefined delay $\mathcal{F}(t, p)$. The transition related to the exit of junction is enabled when each of its input places contains at least one token.

Marking $\mathcal{M}(p, \tau)$ of place p characterizes a set of entrance flows that can be present at time τ in a junction zone corresponding to p. When the marking of p does not change during certain period of time $\Delta\tau$, all measurements taken at the corresponding zone during $\Delta\tau$ are associated to the set of flows. Thus, for a given cycle c, each exit flow $y_j(c)$ can be seg-

ISBN: 978-989-95079-6-8

mented into L_j vehicle platoons, such that

$$y_j^l(c) = \sum_{\tau \in \Delta\tau_l(c)} \mathcal{Y}_j(\tau) \qquad \forall l \in [\![1, L_j(c)]\!],$$

$$z_{kij}^l(c) = \max_{\tau \in \mathcal{R}_k(c)} \mathcal{Z}_k(\tau)\delta_{kij} \quad \begin{array}{l} \forall l \in [\![1, L_j(c)]\!], \\ \forall i \in \mathcal{M}(p, \Delta\tau_l), \\ \forall k \text{ s.t. } \delta_{kij} = 1, \end{array}$$

where $\Delta\tau_l(c)$ is a period of time during which platoon l has left the junction, $z_{kij}^l(c)$ is a maximum number of vehicles that belong to platoon l and have stopped at inner zone k, $\mathcal{R}_k(c)$ is a duration of red light in inner zone k, δ_{kij} indicates if zone k is situated between OD pair of zones (i, j), $\mathcal{M}(p, \Delta\tau_l)$ is a marking of place $p \in \mathcal{P}$ corresponding to exit j. The number of platoons $L_j(c)$ depends on the number of entrances of the junction, and on traffic volume and traffic light command.

Note that the first occurrence of token $(p, \{s1, s2, s3\}, \theta)$ denotes that the beginning of the flow with origin $s1$ crosses the zone corresponding to place p. The ends of the flows are provided by the second HLPN, which has the identical topology and the same representation of places and transitions as the first HLPN. However, the meaning of the tokens is different. A token stands for the end of a flow and is represented by a set of two colours without a timestamp: the name of flow origin $s1 \in \{A, B, C, D\}$ and a cycle number $s2 \in \mathbb{N}$.

The onsets of the beginning and the end of a flow in place p allow us to determine the duration of the flow presence in zone related to p. Thus we can collect the measurements made in this zone and, at each cycle c, can built the following set of equations

$$y_j^l = \sum_{i \in \mathcal{M}(p_j, \Delta\tau_l)} b_{ij}^l x_i \qquad \forall j \in [\![1, m]\!], \forall l \in [\![1, L_j]\!], \quad (2a)$$

$$x_i b_{ij}^l \geq z_{kij}^l \quad \begin{array}{l} \forall j \in [\![1, m]\!], \forall l \in [\![1, L_j]\!], \\ \forall i \in \mathcal{M}(p_j, \Delta\tau_l), \\ \forall k \text{ s.t. } \delta_{kij} = 1, \end{array}, \quad (2b)$$

$$\sum_{j,l} b_{ij}^l = 1 \qquad \forall i \in [\![1, n]\!], \quad (2c)$$

$$b_{ij}^l \geq 0 \quad \begin{array}{l} \forall j \in [\![1, m]\!], \forall l \in [\![1, L_j]\!], \\ i \in \mathcal{M}(p_j, \Delta\tau_l), \end{array} \quad (2d)$$

where the cycle c has been omitted to simplify the notations.

Let J and K be the numbers of constraints (2a) and (2b) respectively, and I be the number of unknowns b_{ij}^l, such that

$$\begin{aligned} J &= \sum_{j=1}^{m} jL_j, \\ K &= \sum_{j=1}^{m} \sum_{l=1}^{L_j} \sum_{i \in \mathcal{M}(p_j, \tau_l)} \sum_{k=1}^{p} \delta_{kij}, \\ I &= \sum_{j=1}^{m} \sum_{l=1}^{L_j} \sum_{i \in \mathcal{M}(p_j, \tau_l)} i. \end{aligned}$$

Since in general $I \geq J$ (in this paper $I \approx 10$), the obtained system of equations (2a) is underdetermined.

5 OD matrix estimation

The estimation problem of OD flow rates b_{ij}^l from the relationship (2a) can be viewed as an estimation of coefficients of the regression model

$$\mathbf{y} = \mathbf{X_1 b}, \quad (3)$$

where $\mathbf{y}$ is a vector composed of J output variables y_j^l, $\mathbf{b}$ is a vector containing the I unknowns b_{ij}^l, $\mathbf{X_1}$ is an $J \times I$ matrix rearranged in such a way that (2a) corresponds to (3).

The elements of model (3) must satisfy the constraints:

$$\mathbf{X_2 b} \geq \mathbf{z}, \quad (4a)$$

$$\mathbf{I b} = \mathbb{1}, \quad (4b)$$

$$\mathbf{b} \geq 0, \quad (4c)$$

where $\mathbf{z}$ is a vector composed of K variables z_{ijk}^l, $\mathbf{X_2}$ is an $K \times I$ matrix built so that (2b) is equal to (4a), $\mathbf{I}$ is an indicator $n \times I$ matrix organised in such a manner that (4b) is equivalent to (2c) and $\mathbb{1}$ is an identity n-vector.

Considering $\mathbf{b}$ to be an m-vector of crisp elements, three situations are studied in this section of the paper:

A) values of $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{y}$ and $\mathbf{z}$ are crisp

B) values of $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{y}$ and $\mathbf{z}$ are fuzzy

C) values of $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{y}$ and $\mathbf{z}$ are crisp, relationships (3) and (4a) are fuzzy, where the strict mathematical relations are replaced by fuzzy equivalents $\lessapprox$ and $\gtrapprox$ that are read "essentially smaller or equal to" and "essentially greater than or equal to".

In the first case, we use the least squares method to estimate the crisp regression coefficients $\mathbf{b}$. The fuzzy least squares (FLS) model based on commonly used Diamond ρ_2-metric [15] has been applied to the second situation. For the last case, we formulate a fuzzy linear programming problem (FLP) following the approach proposed by Zimmernann [16]. We present in the sequel the three methods and their results on our application.

5.1 Least squares method

Let the input and output variables $\mathbf{X}_1$ and $\mathbf{y}$ of the regression model (3) be non-negative crisp numbers. The unknown coefficients $\mathbf{b}$ of the model and the elements of $\mathbf{X}_2$ and $\mathbf{z}$ are also considered to be crisp. A first approach to estimate $\mathbf{b}$ is to use a least squares method. To insure the existence of feasible solutions, we introduce slack variables, e_i ($\forall i = 1, \ldots, K$), and we propose to solve the following problem:

$$\min_{\mathbf{b},\mathbf{e}} \| \mathbf{y} - \mathbf{X_1 b} \|^2 + \sum_{i=1}^{K} e_i \quad (5)$$

subject to

$$\begin{cases} \mathbf{X_2 b} + \mathbf{e} \geq \mathbf{z}, \\ \mathbf{I b} = \mathbb{1}, \\ \mathbf{b} \geq 0, \\ \mathbf{e} \geq 0. \end{cases} \quad (6)$$

5.2 Fuzzy least squares method

To take into account model errors and the inherent imprecision of data measurements, we choose to represent the input and output variables of the regression model (3) by triangular fuzzy numbers $\tilde{\mathbf{X}}_1 = (\underline{\mathbf{X}}_1, \mathbf{X}_1^m, \overline{\mathbf{X}}_1)$, $\tilde{\mathbf{y}} = (\underline{\mathbf{y}}, \mathbf{y}^m, \overline{\mathbf{y}})$ and

ISBN: 978-989-95079-6-8

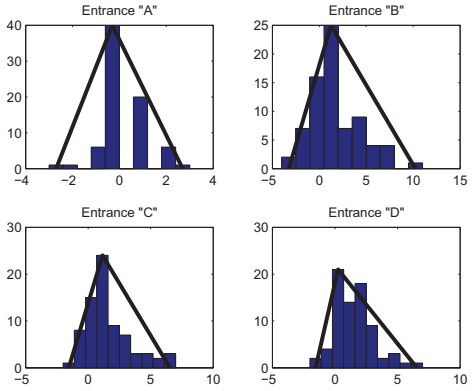


Figure 3: Empirical distributions of error counts for the entrances of the junction

$\tilde{\mathbf{z}} = (\underline{\mathbf{z}}, \mathbf{z}^m, \overline{\mathbf{z}})$, where $[\underline{a}, \overline{a}]$ is the support and a^m is the mode of fuzzy number $\tilde{a}$. The form of fuzzy numbers $\tilde{\mathbf{X}}_1$ was derived from empirical distributions of the error counts as shown in Fig. 3. In most cases, counts of vehicles at the entrances are smaller than the true value of $\mathbf{X}_1$. The fuzzy number $\tilde{\mathbf{X}}_2$ is determined similarly to $\tilde{\mathbf{X}}_1$. Since there is no available histograms of error counts for $\mathbf{y}$ and $\mathbf{z}$, the fuzzy numbers $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{z}}$ are supposed to be symmetrical with the spreads chosen experimentally. Note that $\underline{\mathbf{X}}_1, \underline{\mathbf{X}}_2, \underline{\mathbf{y}}, \underline{\mathbf{z}} \in \mathbb{R}^+$ because of the nature of the data.

According to the fuzzy least squares (FLS) method proposed by Diamond [15, 17] and supposing the unknown coefficients $\mathbf{b}$ to be crisp, the following minimisation problem is written:

$$\min_{\mathbf{b}} \left(\| \underline{\mathbf{y}} - \underline{\mathbf{X}}_\mathbf{1}\mathbf{b} \|^2 + \| \mathbf{y}^m - \mathbf{X}_1^m\mathbf{b} \|^2 + \right. \\ \left. + \| \overline{\mathbf{y}} - \overline{\mathbf{X}}_\mathbf{1}\mathbf{b} \|^2 \right) \quad (7)$$

subject to

$$\begin{cases} \overline{\mathbf{X}}_\mathbf{2}\mathbf{b} \geq \underline{\mathbf{z}}, \\ \mathbf{I}\mathbf{b} = \mathbb{1}, \\ \mathbf{b} \geq 0. \end{cases} \quad (8)$$

Note that the first constraint in (8) is the least conservative translation of the strong relation (4a).

5.3 Fuzzy linear programming approach

Another approach to estimate the OD flow rates is to model the constraints (3) and (4a) by fuzzy sets. Assuming that the values of $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{y}$ and $\mathbf{z}$ are strict numbers, we consider the estimation problem of $\mathbf{b}$ that satisfies:

$$\mathbf{X_1 b} \gtrsim \mathbf{y} \quad -\mathbf{X_1 b} \gtrsim -\mathbf{y}, \quad (9a)$$
$$\mathbf{X_2 b} \gtrsim \mathbf{z}, \quad (9b)$$
$$\mathbf{I b} = \mathbb{1}, \quad (9c)$$
$$\mathbf{b} \geq 0, \quad (9d)$$

where the sign $\gtrsim$ means that we accept a small violation of the strict relation $\geq$ in a sense that is described below..

Denoting

$$\mathbf{H} = (-\mathbf{X}_1, \mathbf{X}_1, -\mathbf{X}_2)^\top,$$
$$\mathbf{h} = (-\mathbf{y}, \mathbf{y}, -\mathbf{z})^\top,$$

we write (9a)-(9b) as

$$\mathbf{H b} \lesssim \mathbf{h} \quad (10)$$

Each row i of (10) is assumed to be a fuzzy set with a membership function $\mu_i(\mathbf{b})$:

$$\mu_i(\mathbf{b}) = \begin{cases} 1 & H_i\mathbf{b} \leq h_i \\ 1 - \frac{H_i\mathbf{b} - h_i}{\xi_i} & h_i < H_i\mathbf{b} \leq h_i + \xi_i \\ 0 & H_i\mathbf{b} > h_i + \xi_i \end{cases} \quad (11)$$

where each $\xi_i > 0$ is a given constant reflecting an acceptable degree of constraint violation, $i = 1, \ldots, 2J + K$. Here $\mu_i(\mathbf{b})$ is defined as the degree to which $\mathbf{b}$ satisfies the relationship i.

With respect to Bellman-Zadeh decision-making rule for fuzzy sets [18], the membership function of the fuzzy decision set $\tilde{B}$ is defined as follows

$$\mu_{\tilde{B}}(\mathbf{b}) = \min_i \left(\mu_i(\mathbf{b})\right) = \min_i \left(1 - \frac{H_i\mathbf{b} - h_i}{\xi_i}\right), \\ \forall i = 1, \ldots, 2J + K. \quad (12)$$

According to the symmetric FLP method proposed by Zimmermann [16, 19], the crisp optimal solution can be given by

$$\max_{\mathbf{b} \geq 0, \mathbf{I}\mathbf{b} = \mathbb{1}} \mu_{\tilde{B}}(\mathbf{b}) \quad (13)$$

which can be obtained by solving the following problem of linear programming:

$$\max_{\lambda, \mathbf{b}} \lambda \quad (14)$$

subject to

$$\begin{cases} \lambda < 1 - \frac{H_i\mathbf{b} - h_i}{\xi_i} & \forall i = 1, \ldots, 2J + K, \\ \mathbf{I}\mathbf{b} = \mathbb{1}, \\ \mathbf{b} \geq 0. \end{cases} \quad (15)$$

5.4 Results

These methods have been tested using the *raw data* collected at the experimental junction. The estimation is made on 25 consecutive traffic light cycles, equivalent to 30 minutes. In addition, the actual values of OD flow rates b_i^* ($\forall i = 1, \ldots, 8$), derived manually from video images, are available for all cycles.

Since the measure of the number of stopped vehicles is more accurate than the measure of exit flow volume, the following parameters of the FLS method where chosen: $\mathbf{y}^m - \underline{\mathbf{y}} = \overline{\mathbf{y}} - \mathbf{y}^m = 3$, $\mathbf{z}^m - \underline{\mathbf{z}} = \overline{\mathbf{z}} - \mathbf{z}^m = 2$. The acceptable degree of constraint violation ξ_i ($\forall i = 1, \ldots, 2J + K$) in FLP method has been fixed to 5 for constraints (9a) and to 3 for constraints (9b).

The estimation error has been calculated for the OD flow rates (Fig. 4): $E = \mathbf{b}\mathbf{X}^* - \mathbf{b}^*\mathbf{X}^*$, where $\mathbf{X}^*$ is a vector of actual vehicle counts at the entrances of the experimental junction. The best results are obtained with the FLS method for which the median error is almost zero for all OD flows. The results of the least squares method are less accurate than those of the FLS method. Surprisingly, the FLP method, although it seems to be a more natural fuzzy translation of the crisp initial problem, does not provide very good results. Note that the estimation error E of all methods is higher, if the flow volume is lower, like for the OD flows "AD" and "BB".

ISBN: 978-989-95079-6-8

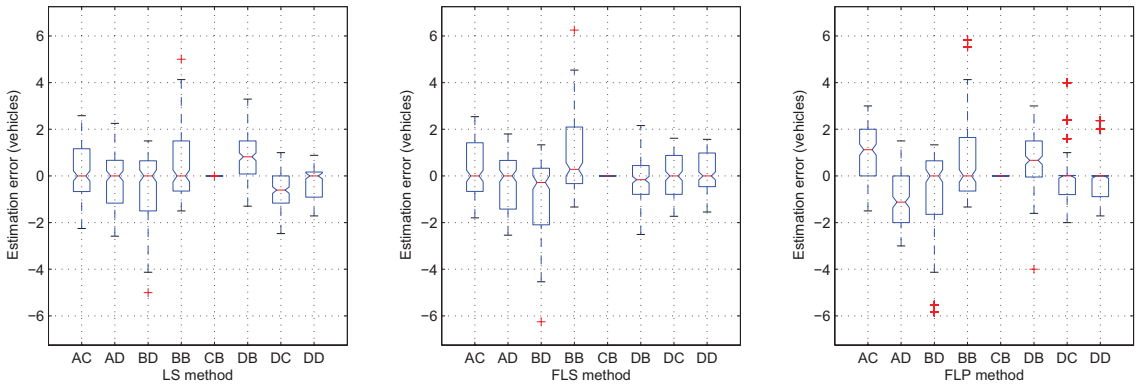


Figure 4: Estimation errors for 8 OD flow rates

6 Conclusion and future lines of research

In this paper we have proposed a new short-time estimation method of the OD matrix for a signalised junction. We have built a model of the segmentation of traffic flows crossing the junction in order to draw up a vehicle conservation law, represented by a underdetermined system of equations, for each traffic light cycle. The model, made using High-Level Petri Nets, is based on the use of traffic lights.

Real traffic measurements collected at the real experimental junction are used to estimate the OD flow rates. The experimental comparison shows that a small gain is obtained when the data imprecision and uncertainty is taken into account using a fuzzy modelling. Among the two proposed fuzzy approaches, the fuzzy least squares yields the most accurate results. In spite of the important imprecision of the real data, the first results are very encouraging to continue the research on improvement of the estimation accuracy.

The main future line of research is to fuzzify the token timestamps of the first HLPN in order to model the temporal imprecision of data measurements. The application of our method should be also extended to a sequence of junctions.

References

[1] S. Midenet *et al.* Signalized intersection with real-time adaptive control: on-field assessment of CO_2 and pollutant emission reduction. *Transportation Research Part D*, 9:29–47, 2004.

[2] H. J. Van Zuylen. The estimation of turning flows on a junction. *Traffic Engineering Control*, 20(11):539–541, 1979.

[3] E. Hauer, E. Pagitsas, and B. T. Shin. Estimation of turning flows from automatic counts. *Transportation Research Record*, (795):1–7, 1981.

[4] M. J. Maher. Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. *Transportation Research Part B*, 17(6):435–447, 1983.

[5] M. Cremer and H. Keller. A new class of dynamic methods for the identification of origin-destination flows. *Transportation Research Part B*, 21(2):117–132, 1987.

[6] N. L. Nihan and G. A. Davis. Recursive estimation of origin-destination matrices from input/output counts. *Transportation Research Part B*, 21(2):149–163, 1987.

[7] B. Li and B. De Moor. Recursive estimation based on the equality-constrained optimization for intersection origin-destination matrices. *Transportation Research Part B*, 33(3):203–214, 1999.

[8] H. D. Sherali *et al.* Parameter optimization methods for estimating dynamic origin-destination trip-tables. *Transportation Research Part B*, 31(2):141–157, 1997.

[9] H. Keller and G. Ploss. Real-time identification of O-D network flows from counts for urban traffic control. In N.H.Gartner and N.H.M.Wilson, editors, *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, pages 267–289. Elsevier Science Publishing Co., 1987.

[10] I. Matschke and B. Friedrich. Dynamic OD estimation using additional information from traffic signal lights timing. *TRISTAN IV Conference*, 2001.

[11] D. Aubert *et al.* Automatic vehicle queue measurement at intersections using image-processing. *8th IEE International Conference on Road Traffic Monitoring and Control, London*, 422:100–104, 1996.

[12] M. Dotoli and M. P. Fanti. An urban traffic network model via coloured timed Petri nets. *Control Engineering Practice*, 14:1213 – 1229, 2006.

[13] K. Jensen *et al.* Coloured Petri nets and CPN tools for modelling and validation of concurrent systems. In *International Journal on Software Tools for Technology Transfer*, pages 98–132. Springer Verlag, 1998.

[14] T. Murata. Petri nets: properties, analysis and applications. In *Proceedings of the IEEE*, volume 77, pages 541–580, April 1989.

[15] P. Diamond and R. Korner. Extended fuzzy linear models and least squares estimates. *Computers and Mathematics with Applications*, 33(9):15–32, May 1997.

[16] H.-J. Zimmermann. Description and optimization of fuzzy systems. *International Journal of General Systems*, 2:209–215, 1976.

[17] R. Slowinski, editor. *Fuzzy sets in decision analysis, operation research and statistics*. The handbooks of fuzzy sets. Kluwer Academic Publishers, 1998.

[18] R. Bellman and L. A. Zadeh. Decision-making in a fuzzy environment. *Management Science*, 17:141–164, 1970.

[19] H.-J. Zimmermann, editor. *Fuzzy set theory and its applications*. Kluwer Academic Publishers, second edition, 1991.

ISBN: 978-989-95079-6-8

Classification with reject options in a logical framework: a fuzzy residual implication approach

Hoel Le Capitaine Carl Frélicot

MIA Laboratory, University of La Rochelle
La Rochelle, France
Emails: {hoel.le_capitaine, carl.frelicot}@univ-lr.fr

Abstract— *In many classification problems, overlapping classes and outliers can significantly decrease a classifier performance. In this paper, we introduce the possibility of a given classifier to reject patterns either for ambiguity or for distance. From a set of typicality degrees for a pattern to be classified, we propose to use fuzzy implications to quantify the similarity of the degrees. A class-selective scheme based on this new family is presented, and experimental results showing the efficiency of the proposed algorithm are given.*

Keywords— Ambiguity measures, Dombi and Hamacher implications, fuzzy r-implications, pattern classification, reject options

1 Introduction

Usual classification of objects, or patterns, consists in assigning incoming samples to one class belonging to a set of known ones. Since classes may overlap in the feature space and all of them do not appear during the learning step of most real classification problems, some samples may have to be not classified but rejected. In practice, two situations can lead to rejection, either the incoming sample could belong to unknown class(es) (*distance rejection*) or it should be assigned to several classes (*ambiguity rejection*) [1]. In this paper, we focus our attention on this latter situation. Initially introduced by Chow [2], it consists in rejecting the most unreliable patterns that have a low posterior probability, or more generally a low membership degree value, to its closest class. Such a reject classification rule is optimal if posterior probabilities are estimated without error, unfortunately this is not the case in practice.

We investigate some new measures that aim at quantifying to what extent such a pattern is ambiguous from its membership degrees, or more generally from its multi-class classifier soft outputs, using fuzzy implications. These measures assess the relationships between membership values with various combination of triangular norm operators (t-norms for short, see [3] for an overview).

The remainder of this paper is organized as follows. In section 2, we start by presenting distance-based class models that we will use in the sequel, and then discuss about standard approaches to pattern rejection. Next, we give a brief recall on some basic definitions of triangular norms and fuzzy residual implications, as well as some new parametrical implications (section 3). Section 4 is devoted to the proposal of using implications as ambiguity measures. In section 5, an experimental evaluation and comparisons to well known class-selective pattern recognition schemes are provided. We finally conclude and give some perspectives in section 6.

2 Classification with reject options

2.1 Classification

Let $\mathbf{x}$ be a pattern described by a p (real) features and let $\Omega = \{\omega_1, \cdots, \omega_c\}$ be a set of classes of cardinality c. The objective of classification is to assign any pattern $\mathbf{x} \in \mathbb{R}^p$ to one of the c classes of Ω. It generally consists of two steps L (*labeling*) and H (*hardening*):

- $L : \mathbf{x} \mapsto \mathbf{u}(\mathbf{x}) = {}^t(u_1(\mathbf{x}), \ldots, u_c(\mathbf{x})) \in \mathcal{L}_{\bullet c}$
- $H : \mathbf{u}(\mathbf{x}) \mapsto \mathbf{h}(\mathbf{x}) = {}^t(h_1(\mathbf{x}), \ldots, h_c(\mathbf{x})) \in \mathcal{L}_{hc}$.

where $\mathcal{L}_{hc} = \{\mathbf{h}(\mathbf{x}) \in \mathcal{L}_{fc} : h_i(\mathbf{x}) \in \{0,1\}\}$ and $\mathcal{L}_{\bullet c}$ depends on the mathematical framework the classifier relies on, e.g. $\mathcal{L}_{pc} = [0,1]^c$ for degrees of typicality or $\mathcal{L}_{fc} = \{\mathbf{u}(\mathbf{x}) \in \mathcal{L}_{pc} : \sum_{i=1}^{c} u_i(\mathbf{x}) = 1\}$ for posterior probabilities and membership degrees. Posterior probabilities can be obtained either from (known) class-conditional densities whose parameters are estimated using a learning set X of patterns, i.e. patterns for which the assignment is known, or from class-density estimates using the classes of their neighbors in X. However, for high dimensional feature spaces, many patterns are required to perform good class-density estimates (curse of dimensionality). When the learning set is small, distances to prototypes of each class are used, e.g.

$$d^2(\mathbf{x}, \mathbf{v}_i) = (\mathbf{x} - \mathbf{v}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{v}_i) \tag{1}$$

where $\mathbf{v}_i$ and Σ_i are the mean vector and covariance matrix of the class ω_i estimated from X. In order to have a common scaling, label values are often obtained in the unit interval, e.g.

$$u_i(\mathbf{x}) = \frac{\alpha_i}{\alpha_i + d^2(\mathbf{x}, \mathbf{v}_i)} \tag{2}$$

where α_i are user-defined parameters.

We focus on the $H-$step because it is in charge of the classification. This step often reduces to the class of maximum label value selection, resulting in an exclusive classification rule which is not efficient in practice because it supposes that Ω is exhaustively defined (closed-world assumption) and classes do not overlap (separability assumption). Such untrue assumptions can lead to very undesired decisions. In many real applications, it is more convenient to with-hold making a decision than making a wrong assignment, e.g. in medical diagnosis where a false negative outcome can be much more costly than a false positive. Reject options have been proposed to overcome these difficulties and to reduce misclassification risk. The first one, called *distance rejection* [1], is dedicated

ISBN: 978-989-95079-6-8

to outlying patterns. If $\mathbf{x}$ is far from all the class prototypes, this option allows to assign it to no class. The second one, called *ambiguity rejection*, allows to assign inlying patterns to several or all classes [2, 4]. If $\mathbf{x}$ is close to two or more class prototypes, it is associated with the corresponding classes. Including reject options leads to partition the feature space into as many regions as subsets of classes, i.e. at most 2^c ones, to which a pattern can be assigned. Formally, it consists in modifying H such that $\mathbf{h}(\mathbf{x})$ can take values in the set of vertices of the unit hypercube $\mathcal{L}_{hc}^c = \{0,1\}^c$ instead of the exclusive subset $\mathcal{L}_{hc} \subset \mathcal{L}_{hc}^c$. Different strategies can be adopted to handle these options at hand, but they all lead to a three types decision system: distance rejection when $\mathbf{h}(\mathbf{x}) = {}^t(0,\ldots,0) = \underline{0}$, exclusive classification when $\mathbf{h}(\mathbf{x}) \in \mathcal{L}_{hc}$, ambiguity rejection when $\mathbf{h}(\mathbf{x}) \in \mathcal{L}_{hc}^c \backslash \{\mathcal{L}_{hc} \cup \underline{0}\}$. The resulting classification rule is then a matter of selecting the appropriate number of classes varying from zero (distance rejection) to c (total ambiguity rejection).

2.2 Usual reject schemes

Since the work by Chow [2], many rejection schemes have been proposed. In its general form, a class-selective procedure is defined by

$$n^\star(\mathbf{x},t) = \min_{k\in[1,c]} \{k : \mathcal{A}_k(\mathbf{x}) \leq t\} \tag{3}$$

where $\mathcal{A}_k$ is a given ambiguity measure on membership degrees, $n^\star(\mathbf{x},t)$ is the number of selected classes for the pattern $\mathbf{x}$ to be classified, and t is a user-defined threshold which can be class dependent (e.g. t_k). This threshold can be set conditionally to cost functions relative to error, reject and correct classification rates. Propositions from the literature mainly consist in new definitions of ambiguity measures $\mathcal{A}_k$. For Chow, $\mathcal{A}_k$ was one minus the maximum value of the membership degrees. In the paper by Ha [4], the second highest value was tested to decide wether one or several classes are selected. Since this scheme is leading to unnatural decisions, Horiuchi [5] proposed a new measure defined by the difference of the membership degrees which is actually a disambiguity measure. In [6], Frélicot proposed to use the ratio of membership degrees. As usual, we use the convention that if $\mathcal{A}_k(x) > t$ for all k, then we set $n^\star(\mathbf{x},t) = c$. Table 1 summarizes the existing ambiguity measures used for pattern rejection, where the membership degrees are assumed to be sorted in decreasing order for writing convenience, i.e. $u_1(\mathbf{x}) \geq \cdots \geq u_c(\mathbf{x})$.

Table 1: Existing ambiguity measures related to Eq. (3).

Scheme	**Ambiguity Measure** $\mathcal{A}_k(\mathbf{x})$
Chow [2]	$1 - u_1(\mathbf{x})$
Ha [4]	$u_{k+1}(\mathbf{x})$
Horiuchi [5]	$1 - (u_k(\mathbf{x}) - u_{k+1}(\mathbf{x}))$
Frélicot [6]	$u_{k+1}(\mathbf{x})/u_k(\mathbf{x})$

3 Fuzzy residual implications

3.1 Basic definitions

Let us recall basic definitions of fuzzy operators that will be used to combine the values of interest, i.e. the pattern class-degrees of typicality. Depending on properties, aggregation functions can be classified into several categories: conjunctive, disjunctive, compensatory, and so on. We restrict on conjunctive and disjunctive functions. By definition, the output of a conjunctive operator is lower or equal than the minimum value, whereas the output of a disjunctive operator is greater or equal than the maximum value. Beyond these operators, we choose to use the triangular norms because of their ability to generalize the logical AND and OR crisp operators to fuzzy sets, see [3] for a survey. Briefly, a triangular norm (or t-norm) is a binary operation on the unit interval $\top : [0,1]^2 \to [0,1]$ which is commutative, associative, non decreasing and has 1 for neutral element. Thus, a t-norm $\top$ is conjunctive and the minimum operator $\wedge$ is the greatest t-norm. Alternatively, a triangular conorm (or t-conorm) is the dual binary operation $\perp : [0,1]^2 \to [0,1]$ having the same properties except that its neutral element is 0. Thus, a t-conorm $\perp$ is disjunctive and the maximum operator $\vee$ is the lowest t-conorm. Typical examples of dual couples (t-norm, t-conorm) that will be used in the sequel are given in Table 2.

Table 2: Examples of dual couples, including parametrical ones.

Standard	$a\top_S b = \min(a,b)$ $a\perp_S b = \max(a,b)$
Algebraic	$a\top_A b = a\,b$ $a\perp_A b = a + b - a\,b$
Łukasiewicz	$a\top_L b = \max(a+b-1,0)$ $a\perp_L b = \min(a+b,1)$
Hamacher	$a\top_{H_\gamma} b = \frac{a\,b}{\gamma+(1-\gamma)\,(a+b-a\,b)}$ $a\perp_{H_\gamma} b = \frac{a+b+(\gamma-2)\,a\,b}{1+(\gamma-1)\,a\,b}$
Dombi	$a\top_{D_\gamma} b = \left(1+\left(\left(\frac{1-a}{a}\right)^\gamma + \left(\frac{1-b}{b}\right)^\gamma\right)^{1/\gamma}\right)^{-1}$ $a\perp_{D_\gamma} b = 1-\left(1+\left(\left(\frac{a}{1-a}\right)^\gamma + \left(\frac{b}{1-b}\right)^\gamma\right)^{1/\gamma}\right)^{-1}$

A fuzzy residual implication, denoted R-implication (or $\rightarrow$) is defined by:

$$I(a,b) = \sup_t \{t \in [0,1] : \top(a,t) \leq b\} \tag{4}$$

Note that if $\top$ is a left-continuous t-norm, the supremum operation can be substituted by maximum. If we use additive generating functions, i.e. a strictly decreasing function $f : [0,1] \to [0,+\infty]$ with $f(1) = 0$, and admitting an inverse (or pseudo-inverse) function f^{-1}, Eq. (4) can be written as

$$I(a,b) = \max(f^{-1}(f(b) - f(a)), 0) \tag{5}$$

because f is strictly monotonic. We generally speak about an implication function if I is non-increasing in the first variable, non-decreasing in the second variable and $I(0,0) = I(1,1) = 1$, and $I(1,0) = 0$, see [7] for a large survey on fuzzy implication functions. Within theses implications, the well-known *Gödel* and *Goguen* ones are respectively given by

$$I(a,b) = \begin{cases} 1 & \text{if } b \geq a \\ b & \text{if } b < a \end{cases} \tag{6}$$

and

$$I(a,b) = \begin{cases} 1 & \text{if } b \geq a \\ \frac{b}{a} & \text{if } b \leq a \end{cases} \tag{7}$$

ISBN: 978-989-95079-6-8

which are obtained with the minimum and algebraic (or product) triangular norms, respectively. In the sequel, we will assume for writing convenience that the fuzzy values are sorted in decreasing order, e.g. $a \geq b$.

3.2 Parametrical implications

Proposition 1. *Let $(\top_{H_\gamma})$, $\gamma \in [0, +\infty[$, be the family of Hamacher t-norms. The residual Hamacher implication is given by*

$$I_{H_\gamma}(a,b) = \begin{cases} 1 & \text{if } b \geq a \\ \frac{b\,(\gamma+a-\gamma a)}{b\,(\gamma+a-\gamma a)+a-b} & \text{if } b \leq a \end{cases} \tag{8}$$

Proof. By definition of R-implications (4), we can write $I_{H_\gamma}(a,b) = \sup_t\{t \in [0,1] : \top_{H_\gamma}(a,t) \leq b\}$. We also have $I_{H_\gamma}(a,b) = \max_t\{t \in [0,1] : \top_{H_\gamma}(a,t) \leq b\}$ because $\top_{H_\gamma}$ is a left-continuous t-norm. Then, solving

$$\frac{a\,t}{\gamma + (1-\gamma)\,(a+t-a\,t)} \leq b \tag{9}$$

gives

$$t \leq \frac{b\,(\gamma+a-\gamma a)}{b\,(\gamma+a-\gamma a)+a-b}. \tag{10}$$

Since $a \geq b$, it is easy to show that the right part of Eq. (10) is in $[0,1]$, hence we obtain Eq.(8). □

Proposition 2. *Let $(\top_{D_\gamma})$, $\gamma \in]0, +\infty[$, be the family of Dombi t-norms. The residual Dombi implication is given by*

$$I_{D_\gamma}(a,b) = \begin{cases} 1 & \text{if } b \geq a \\ \left(1 + \left(\left(\frac{1-b}{b}\right)^\gamma - \left(\frac{1-a}{a}\right)^\gamma\right)^{1/\gamma}\right)^{-1} & \text{if } b < a \end{cases} \tag{11}$$

Proof. $I_{D_\gamma}(a,b) = \sup_t\{t \in [0,1] : \top_{H_\gamma}(a,t) \leq b\}$ by (4). Since $\top_{D_\gamma}$ is a left-continuous t-norm, we can write $I_{D_\gamma}(a,b) = \max_t\{t \in [0,1] : \top_{D_\gamma}(a,t) \leq b\}$. Then, solving

$$\left(1 + \left(\left(\frac{1-a}{a}\right)^\gamma + \left(\frac{1-t}{t}\right)^\gamma\right)^{1/\gamma}\right)^{-1} \leq b \tag{12}$$

gives

$$t \leq \left(1 + \left(\left(\frac{1-b}{b}\right)^\gamma - \left(\frac{1-a}{a}\right)^\gamma\right)^{1/\gamma}\right)^{-1}. \tag{13}$$

Since, $a \geq b$, it is easy to show that $\left(\left(\frac{1-b}{b}\right)^\gamma - \left(\frac{1-a}{a}\right)^\gamma\right)^{1/\gamma} \geq 0$, hence the right part of Eq. (13) is in $[0,1]$ and (11) is obtained. □

Note that R-implications are mostly used in fuzzy inference systems, see [8] for a large overview on the use of parametrical R-implications in fuzzy rule based systems.

4 Some parametrical measures of ambiguity

In this section, the concept of similarity measure and its relationship with ambiguity measures are described. Then we propose to use fuzzy parametrical implications as new families of ambiguity measures to be used for pattern classification with reject options. The resulting class-selection algorithm (H-step) is presented and we finally discuss the choice of the parameter for parametrical implications with the help of numerical examples.

4.1 Proposition and properties

A similarity measure $\mathcal{S}$ generally satisfies the following properties:

(P1) $\mathcal{S}(a,b) = \mathcal{S}(b,a)$, *(symmetry)*

(P2) $\mathcal{S}(a,a) \geq \mathcal{S}(a,b)$, *(minimality)*

(P3) $\mathcal{S}(a,b) = 1 \Leftrightarrow a = b$, *(identity)*

However, the symmetry property (P1) is still subject to experimental investigation: if $\mathcal{S}(a,b)$ is the answer to the question *how is a similar to b?*, then, when making comparisons, subjects focus more on the feature a than on b. This corresponds to the notion of *saliency of a and b* [9]: if b is more salient than a, then a is more similar to b than vice versa, which is experimentally confirmed. Property (P2) is also open because it can be violated experimentally, see [9] for details.

There are several ways to deal with the comparison of fuzzy values, or fuzzy quantities. The first one is based on a broad class of measures of equality based on a distance measure which is specified for membership functions of fuzzy sets. This approach takes its roots from studies on how to measure the distance between two real functions and do not refer to any specific interpretation. The general form of a Minkowski r-metric is usually taken and leads to well known distance functions (Hamming, Euclidean, Chebyshev). A second way to compare fuzzy values comes from some basic set-theoretic considerations where union, intersection and complement are defined for fuzzy sets. Cardinal and possibility based measures belong to this category. In this paper, we focus on the third way to deal with fuzzy values comparison: the logical framework. Accordingly to [10], for a certain universe of discourse $\mathcal{D}$, the degree of equality of two fuzzy elements a and b can be defined by implications as follows:

$$(a \equiv b) = \frac{1}{2}\big((a \to b) \wedge (b \to a) + (\bar{a} \to \bar{b}) \wedge (\bar{b} \to \bar{a})\big) \tag{14}$$

where $\wedge$ stands for minimum, $\to$ is an implication and $\bar{a}$ is the strict negation $\bar{a} = 1 - a$.

Since 1 is the neutral element of t-norms, applying Eq. (4) with $a \geq b$ gives

$$(a \equiv b) = \frac{1}{2}\big((a \to b) + (\bar{b} \to \bar{a})\big). \tag{15}$$

A convenient way to define an ambiguity measure is to quantify to which extent two fuzzy membership degrees are similar, so that it is closely related to the problem of matching fuzzy quantities, or fuzzy sets similarity. So we propose to use fuzzy R-implications as generalized ambiguity measures.

ISBN: 978-989-95079-6-8

Given a set of c truth values assumed to be sorted in decreasing order, i.e. $u_1(\mathbf{x}) \geq \cdots \geq u_c(\mathbf{x})$, with no loss of generality, let us have two predicates ($\mathbf{x}$ *is* ω_i), with the truth value $u_i(\mathbf{x})$, and ($\mathbf{x}$ *is* ω_k), with the truth value $u_k(\mathbf{x})$. The truth value of the implication *if the pattern* $\mathbf{x}$ *is* ω_i*, then the pattern* $\mathbf{x}$ *is also* ω_j*,* $\forall j$ *varying from* $i+1$ *to* k is an ambiguity measure given by $I(u_i(\mathbf{x}), u_k(\mathbf{x}))$. By convention, we assume that the assignment of $\mathbf{x}$ to ω_i is more probable than the assignment of $\mathbf{x}$ to ω_k when using this implication because $u_i(\mathbf{x}) \geq u_k(\mathbf{x})$ and obviously we have $I(u_i(\mathbf{x}), u_{i+1}(\mathbf{x})) \geq I(u_i(\mathbf{x}), u_{k>i+1}(\mathbf{x}))$.

Proposition 3. *Given* $t \in [0,1]$*, the optimum cardinality of the generalized class-selective rejection rule is given by*

$$n^\star(\mathbf{x}, t) = \min_{k \in [1,c]} \left\{ k : I\Big(u_k(\mathbf{x}), u_{k+1}(\mathbf{x})\Big) \leq t \right\} \qquad (16)$$

with $u_1(\mathbf{x}) \geq \cdots \geq u_c(\mathbf{x})$*, and the convention* $u_{c+1}(\mathbf{x}) = 0$*.*

Since $I(a,0) = 0$ if $a \neq 0$, c classes will be selected if none were previously selected.

Property 1. *If we use the Standard triangular norm* min*, we obtain the Ha class-selective rejection scheme [4].*

Property 2. *If we use the Łukasiewicz triangular norm, we obtain the Horiuchi class-selective rejection scheme [5].*

Property 3. *If we use the Algebraic triangular norm, we obtain the Frélicot class-selective rejection scheme [6].*

Proofs are straightforward and will be given in a longer forthcoming paper. Note that modifying Eq.(16) such as

$$n^\star(\mathbf{x}, t) = \min_{k \in [0,c]} \left\{ k : I\Big(u_k(\mathbf{x}), u_{k+1}(\mathbf{x})\Big) \leq t \right\} \qquad (17)$$

with the convention $u_0(\mathbf{x}) = 1$ allows to select none of the classes, i.e. to proceed to distance rejection, since $I(1,a) = a$ whatever the triangular norm.
The resulting generalized pattern classification rule with reject options ($H-$step) is presented in Algorithm 1. It can be used to compare various schemes, depending on the choice of the triangular norm.

Algorithm 1: $H-$step classification algorithm.

Data: a vector of soft class-labels $\mathbf{u}(\mathbf{x}) \in \mathcal{L}_{pc}$ and a reject threshold t
Result: a vector of class-selective assignments $\mathbf{h}(\mathbf{x}) \in \mathcal{L}_{hc}^c$
begin
  set $\mathbf{h}(\mathbf{x})$ to $\underline{0}$
  find $n^\star(\mathbf{x}, t)$ − Eq.(16) or Eq.(17)
  foreach $j = 1 : n^\star(\mathbf{x}, t)$
  in decreasing order of $u_j(\mathbf{x})'s$ **do**
    set $h_j(\mathbf{x}) = 1$
  end
end

4.2 Discussion and examples

One of the main difficulties when using t-norms is to choose the dual couple and if needed to set the parameter value. Let us study how the choice of γ for Hamacher and Dombi implications modifies the resulting implication strength.
• *Hamacher*: increasing the value of γ will make two fuzzy values more similar because $I_{H_\gamma}(a,b) = ab/(a-b+ab)$ if $\gamma = 0$ while $\lim_{\gamma \to +\infty} I_{H_\gamma}(a,b) = 1$ whatever $(a,b) \in [0,1]^2$. Indeed, I_{H_γ} is non-decreasing with γ since

$$\frac{\partial I_{H_\gamma}}{\partial \gamma} = \frac{(b-ab)(a-b)}{(b(\gamma + a - \gamma a) + a - b)^2} \geq 0.$$

The influence of γ is much more significant for low values of a and b than for high ones because $b(\gamma + a - \gamma a)$ appears to be of order $b\,a$ (respectively $b\,(\gamma + a)$) for high (respectively low) values of a and b. It follows that if $\gamma_1 >> \gamma_2$, we have $\frac{\gamma_1}{\gamma_1+\varepsilon} >> \frac{\gamma_2}{\gamma_2+\varepsilon}$, so that large values of γ associated to low values of (a,b) will result in a high value of I_{H_γ}, see Table 3 for examples.

Table 3: Hamacher implications examples for $a = 0.9$, $b = 0.8$, $c = 0.1$ and $d = 0.05$.

γ	0	2	10
$I_{H_\gamma}(a,b)$	0.87	0.89	0.93
$I_{H_\gamma}(a,c)$	0.10	0.12	0.19
$I_{H_\gamma}(c,d)$	0.09	0.65	0.90

• *Dombi*: decreasing the value of γ will make two fuzzy values more similar because $\lim_{\gamma \to +\infty} I_{D_\gamma}(a,b) = b$ while $\lim_{\gamma \to 0} I_{D_\gamma}(a,b) = 1$ whatever $(a,b) \in [0,1]^2$. Analogously to Hamacher family, I_{D_γ} is non-increasing with γ since

$$\frac{\partial I_{D_\gamma}}{\partial \gamma} \leq 0.$$

Thus, decreasing γ for the Dombi family has the same impact as increasing γ for the Hamacher family, so that we expect the opposite tendency, see Table 4 for examples. The appropriate tuning of γ can be achieved using a gradient procedure, as in [11].

Table 4: Dombi implications examples for $a = 0.9$, $b = 0.8$, $c = 0.1$ and $d = 0.05$.

γ	0.5	2	10
$I_{D_\gamma}(a,b)$	0.97	0.81	0.80
$I_{D_\gamma}(a,c)$	0.12	0.10	0.10
$I_{D_\gamma}(c,d)$	0.35	0.06	0.05

The simple graphical example shown in Fig.1 illustrates how the different implications behave. The top-plot shows a one-dimensional dataset composed of two classes (□ and ×) described by a distance-based model (Eqs.(1-2)) with $\alpha_i = 1$ ($\forall i = 1, c$). In the middle, implication truth values for the standard, algebraic, and Łukasiewicz triangular norms, corresponding to Ha, Frélicot and Horiuchi class-selective schemes respectively, as well as Chow's scheme are shown. Truth values when using parametrical implications are shown in the bottom-plot. Observing the data points in the top-plot allows to obtain the bounds of intervals of x values where the classes do not overlap, therefore for which no misclassification should

ISBN: 978-989-95079-6-8

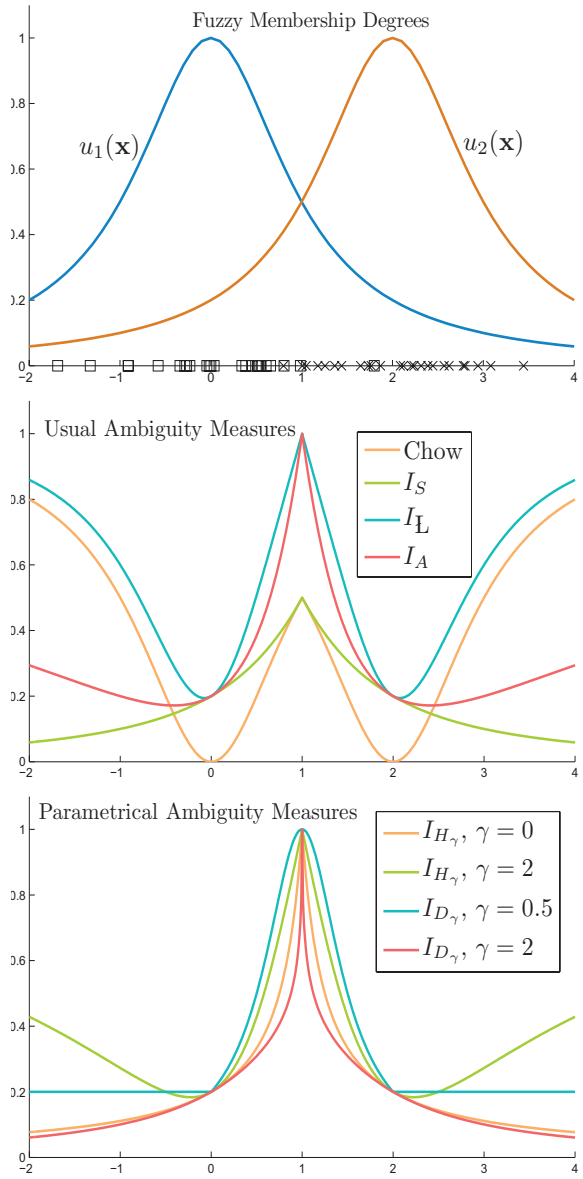


Figure 1: Original outputs using Eqs.(1-2) for two classes with normal distributions (*top*), implication truth values with usual triangular norms (*middle*) and implication truth values with Hamacher and Dombi families (*bottom*).

happen, here $\mathbf{x}_{low} \approx 0.8$ and $\mathbf{x}_{up} \approx 1.9$. Then, the corresponding $u_1(\mathbf{x}_{low})$ and $u_2(\mathbf{x}_{up})$ ordinate values can be used to set the classification threshold t. Then, referring to measures of ambiguity in the middle - and bottom - plots, the behaviour of the different schemes can be analyzed. In such a situation, Chow's and Horiuchi's (Łukasiewicz implication) schemes would lead to reject too many points in low density areas, while other schemes do not show these drawbacks. On the other hand, Ha's scheme (standard implication) would not allow to reject simultaneously highly ambiguous and outlying patterns, whereas schemes based on algebraic or parametrical implications do (provided an appropriate value for the parameter), so a better classification performance can be expected. One can also see in the bottom-plot that tuning the parameter allows to reject patterns for ambiguity as well as for distance (a higher value of γ for the Hamacher family and a lower value for the Dombi family) as pointed out in the discussion.

5 Experimental results

In this section, we report experiments carried out on both artificial and real benchmark datasets for which it is beneficial to introduce the ambiguity reject option because the classes overlap in the feature space. The proposed class-selective rejection scheme is compared to the usual rejection schemes presented in section 2. For additional comparison, results obtained using the two very recent classification rules proposed by Tax and Duin [12] called *Outlier-norm* and *Target-norm* are given, the former being especially designed for distance rejection. Note we did not detailed these (not usual) rules here because they do not rely on an ambiguity measure, therefore they cannot be derived using fuzzy R-implications, by contrast to the previous ones. In all cases, the $L-$step was performed using Eqs.(1-2) with $\alpha_i = 1$ $(\forall i = 1, c)$.

5.1 The datasets

Two synthetic datasets are used: D which contains 2000 points drawn from two normal seven-dimensional distributions of 1000 points each with means $\mathbf{v}_1 = {}^t(1 \;\; 0 \;\cdots\; 0)$ and $\mathbf{v}_2 = {}^t(-1 \;\; 0 \;\cdots\; 0)$, and equal covariance matrices $\Sigma_1 = \Sigma_2 = I$, and DH which consists of two overlapping gaussian classes with different covariance matrices according to the Highleyman distribution, each composed of 800 observations in $\mathbb{R}^2$, see [13]. Real datasets are taken from the UCI Machine Learning Repository [14]. The characteristics (number n of patterns, number p of features, number c of classes, degree of overlap) are reported in Table 5.

Table 5: The datasets and their characteristics.

Dataset	n	p	c	**Overlap**
D	2000	7	2	slight
DH	1600	2	2	very slight
Ionosphere	351	34	2	very strong
Forest	495411	10	2	moderate
Vowel	528	10	11	very slight
Digits	10992	16	10	very slight
Thyroid	215	5	3	very slight
Pima	768	9	2	strong
Statlog	6435	36	6	slight
Glass	214	9	6	moderate

5.2 Results

Table 6 shows the classification performance of the different rejection schemes obtained by a 10-fold cross-validation procedure on the different datasets. In all cases, the threshold t was set to reject 10% of the data, so that 90% is the best achievable correct classification rate performance, then the error rate is $(90 - \text{correct})\%$. The best results are indicated in bold. Note that there are no outliers in the datasets, so that a part of the rejected points are considered as outliers whereas they belong to classes.

It appears from these results that parametrical implications generally outperform usual rejection schemes as well as non usual ones (Outlier-norm, Target-norm), e.g. $I_{H_{\gamma=2}}$ and $I_{D_{\gamma=0.5}}$. As expected in the previous section, these schemes enable to reject both ambiguous and outlying patterns. Furthermore, the tradeoff to be found between a scheme which

ISBN: 978-989-95079-6-8

Table 6: Classification performance (%) on synthetic and real datasets.

Scheme	D	DH	Ionosphere	Forest	Vowel	Digits	Thyroid	Pima	Statlog	Glass
Chow	75.15	83.75	54.42	67.44	87.88	88.03	84.19	59.51	75.66	66.82
Ha	77.35	86.63	54.99	68.35	88.18	87.38	86.98	60.55	75.56	69.63
Horiuchi	77.55	84.31	56.13	69.27	89.09	88.76	86.98	62.89	77.68	68.22
Frélicot	79.65	87.12	**58.12**	69.73	89.09	89.13	**87.91**	63.28	77.48	71.03
Hamacher_0	**80.00**	87.12	55.56	69.76	88.99	88.26	**87.91**	63.15	77.26	70.09
Hamacher_2	79.75	**87.30**	**58.12**	**70.01**	**89.49**	**89.17**	**87.91**	63.28	77.78	**71.33**
$\text{Dombi}_{0.5}$	79.20	86.68	56.41	69.76	89.09	88.96	**87.91**	**63.40**	**78.12**	71.03
Dombi_2	78.85	86.38	55.56	69.65	88.89	87.91	**87.91**	62.50	76.94	70.09
Outlier-norm	76.85	83.31	57.54	67.44	87.88	86.19	84.65	61.33	75.66	69.16
Target-norm	78.80	87.19	56.83	69.69	87.27	86.39	84.19	63.15	75.86	66.82

does not reject outliers (Ha) and others which reject too much patterns (Chow, Horuichi) appears to favour the choice of implications based on a t-norm which is lower than $\top_S$ (the highest t-norm) and greater than $\top_L$, e.g. the algebraic or the parametrical implications (provided an appropriate value for γ). More generally, rejection schemes that take into account relationship between fuzzy membership degrees (Horiuchi, Frélicot, Hamacher, Dombi) perform better than all the others schemes we tested. Looking at the degrees of overlap, $I_{D_{\gamma=0.5}}$ (respectively $I_{H_{\gamma=2}}$) is more efficient for datasets presenting a slight/very slight (respectively strong/very strong) overlap.

6 Conclusion and perspectives

In this paper, a generalized class-selective rejection scheme based on a logical approach to pattern assignation is presented allowing to either reject only ambiguous patterns or ambiguous and outlying patterns. For this purpose, we propose to design a family of ambiguity measures based on fuzzy residual implication functions, which indicate to which extent a pattern should be assigned to $n^\star$ classes depending on its membership degrees from the truth value of the implication. These measures assess the relationships between membership values with various combination of triangular norm operators, including parametrical families. It is shown that the proposed scheme generalizes well-known ones of the literature on pattern classification with reject options. It appears from experiments on both synthetic and real datasets that using, as the basis for residual implication computation, triangular norms greater than the Łukasiewicz triangular norm and lower than the standard one, gives better classification performance. Furthermore, measures taking in consideration several membership degrees, so that interactions between classes are taken into account, also give better results.

A future work will consist in defining new class-selective rejection schemes based on other parametrical triangular families (Sklar, Frank, Yager and so on) and compare their classification performance. We also think that, depending on the context of the pattern recognition problem, other implications functions than the residual implication ones would be suitable. We plan to make an extensive study on the behavior of *S-implications* which are an immediate generalization of the usual boolean implication, *QL-implications* coming from quantum mechanic logic, and *D-implications* which are the contraposition with respect to a negation of QL-implications, see [7] for definitions.

References

[1] B. Dubuisson and M. Masson. A statistical decision rule with incomplete knowledge about classes. *Pattern Recognition*, 26:155–165, 1993.

[2] C.K. Chow. On optimum error and reject tradeoff. *IEEE Transactions on Information Theory*, 16:41–46, 1970.

[3] E.P Klement and R. Mesiar. *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Elsevier, 2005.

[4] T. Ha. The optimum class-selective rejection rule. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):608–615, 1997.

[5] T. Horiuchi. Class-selective rejection rule to minimize the maximum distance between selected classes. *Pattern Recognition*, 31:1579–1588, 1998.

[6] C. Frélicot and B. Dubuisson. A multi-step predictor of membership function as an ambiguity reject solver in pattern recognition. In *4th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 709–715. IPMU'92, July 1992.

[7] M. Mas, M. Monserrat, J. Torrens, and E. Trillas. A survey on fuzzy implication functions. *IEEE Transactions on Fuzzy Systems*, 15(6):1107–1121, 2007.

[8] Th. Whalen. Parameterized r-implications. *Fuzzy Sets and Systems*, 134(2):231–281, 2003.

[9] A. Tversky. Features of similarity. *Psychological review*, 84(4):327–352, 1977.

[10] K. Hirota and W. Pedrycz. Matching fuzzy quantities. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(6):1580–1586, 1991.

[11] L. Rutkowski and K. Cpalka. Flexible neuro-fuzzy systems. *IEEE Transactions on Neural Networks*, 14(3):554–574, 2003.

[12] D.M.J. Tax and R.P.W. Duin. Growing a multi-class classifier with a reject option. *Pattern Recognition Letters*, 29(10):1565–1570, 2008.

[13] W. Highleyman. Linear decision functions, with application to pattern recognition. *Proc. IRE 50*, 50(6):1501–1514, 1962.

[14] C. Blake and C. Merz. Uci repository of machine learning databases, 1998.

ISBN: 978-989-95079-6-8

Fuzzy c-Lines for Data with Tolerance

Yuchi Kanzawa[1] Yasunori Endo[2] Sadaaki Miyamoto[2]

1.Shibaura Institute of Technology, Japan
2.University of Tsukuba, Japan
Email: kanzawa@sic.shibaura-it.ac.jp

Abstract— *This paper presents a new clustering algorithm ,which is based on fuzzy c-lines, can treat data with some errors. First, the tolerance is formulated and introduce into optimization problem of clustering. Next, the problem is solved using Karush-Kuhn-Tucker conditions. Last, the algorithm is constructed based on the results of solving the problem. Some numerical examples for the proposed method are shown.*

Keywords— Data with Tolerance, Fuzzy c-Lines, Fuzzy Clustering

1 Introduction

Fuzzy c-means (FCM) [1] is one of the well-known fuzzy clustering and many FCM variants have been proposed after FCM. In these variants, an algorithm allowing cluster prototypes to be linear varieties instead of points has been proposed [1]. This algorithm is called fuzzy c-varieties (FCV) and regarded as the combination of clustering and principal component analysis. If linear varieties of the prototypes are limited to one dimensional, FCV is called fuzzy c-lines(FCL).

In general, any data that is represented by numeric have some errors. In order to classify such data, an algorithm has been proposed by some of the authors [2]. This algorithm is called the FCM for data with tolerance (FCM-T). In FCM-T, the tolerance is regarded as another decision variable composing the optimization problem like membership grades and the cluster centers, and is determined as minimizing the objective function.

In this paper, on the analogy of the story from FCM to FCM-T, we propose fuzzy c-lines for data with tolerance (FCL-T).

The contents of this paper are the followings. In the second section, we define some notations and introduce FCM-T and FCL. In the third section, we propose FCL-T. In the fourth section, some numerical examples are shown. In the last section, we conclude this paper.

2 Preliminaries

In this section, we define some notations, introduce fuzzy c-lines (FCL), and also introduce fuzzy c-means with tolerance (FCM-T) derived from conventional fuzzy c-means (FCM). In the first subsection, we define some notations which are the data for clustering, the membership by which the each data belongs to the each cluster, the cluster prototypes of one dimensional linear manifold. In the second subsection, we introduce FCL, whose prototypes are one dimensional linear manifolds. In the third subsection, we define the tolerance for the data and the maximum tolerance for the data, and introduce FCM-T, which is the story that clustering algorithms were oriented for data with some errors.

2.1 Notations

In this subsection, we define some notations which are the data for clustering, the membership by which the each data belongs to the each cluster, the cluster prototypes of one dimensional linear manifold.

The data set $x = \{x_i \mid x_i \in \mathbf{R}^p, i \in \{1, \ldots, N\}\}$ is given. The membership by which x_i belongs to the j-th cluster is denoted by $u_{i,j}$ $(i \in \{1, \cdots, N\}, j \in \{1, \cdots, C\})$ and the set of $u_{i,j}$ is denoted by $u \in \mathbf{R}^{N \times C}$ called the partition matrix. The constraint for u is

$$\sum_{j=1}^{C} u_{i,j} = 1 \quad (0 \leq u_{i,j} \leq 1).$$

The cluster prototype set is denoted by $V = \{V_1, \cdots, V_C\}$, where

$$V_j = \{y_j \in \mathbf{R}^p \mid y_j = v_j + t_j s_j, t_j \in \mathbf{R}\}$$

with $v = \{v_j \mid v_j \in \mathbf{R}^p, j \in \{1, \ldots, C\}\}$ and $s = \{s_j \in \mathbf{R}^p \mid \|s_j\| = 1, j \in \{1, \cdots, C\}\}$.

2.2 Fuzzy c-Lines

In this subsection, we introduce FCL (FCV), whose prototypes are one dimensional linear manifolds.

FCL is the algorithm obtained by solving the following optimization problem:

$$\underset{u,v,s}{\text{minimize}}\, J_\ell(u, v, s) \tag{1}$$

$$\text{subject to} \sum_{j=1}^{C} u_{i,j} = 1, \quad \|s_j\| = 1, \tag{2}$$

where

$$J_\ell(u, v, s) = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{i,j}^m \left(\|x_i - v_j\|^2 - ((x_i - v_j) s_{j,\ell})^2 \right). \tag{3}$$

The optimal solutions u, v and s are obtained by the following algorithm.

Algorithm 1 (FCL)

Step 1 *Give the number of clusters C and the fuzzifier parameter m. Set the initial values of v and s.*

Step 2 *Calculate u such that*

$$u_{i,j} = 1 / \sum_{k=1}^{C} \left(\frac{d_{i,j}}{d_{i,k}} \right)^{1/(m-1)}, \tag{4}$$

where

$$d_{i,j} = \|x_i - v_j\|^2 - ((x_i - v_j) s_{j,\ell})^2. \tag{5}$$

ISBN: 978-989-95079-6-8

Step 3 *Calculate v such that*

$$v_j = U_j^{-1} \sum_{i=1}^{N} u_{i,j}^m x_i, \tag{6}$$

where

$$U_j = \sum_{i=1}^{N} u_{i,j}^m. \tag{7}$$

Step 4 *Calculate s as the eigenvector corresponding to the maximal eigenvalue of the following matrix:*

$$\sum_{i=1}^{N} u_{i,j}^m (x_i - v_j)(x_i - v_j)^{\mathrm{T}}. \tag{8}$$

Step 5 *Check the stopping criterion for (u, v, s). If the criterion is not satisfied, go back to* Step 2.

2.3 Fuzzy c-Means for Data with Tolerance

In this subsection, we define the tolerance for the data and the maximum tolerance for the data, and introduce fuzzy c-means for data with tolerance (FCM-T) [2]. This algorithm is oriented toward data with error. The tolerance for the data x is denoted by $\varepsilon = \{\varepsilon_i \mid \varepsilon_i \in \mathbf{R}^p, i \in \{1, \ldots, N\}\}$. The maximum tolerance is denoted by $\kappa = \{\kappa_i \mid \kappa_i \in \mathbf{R}_+, i \in \{1, \ldots, N\}\}$.

FCM-T is the algorithm obtained by solving the following optimization problem:

$$\underset{u,\varepsilon,v}{\text{minimize}}\, J_{\mathrm{t}}(u, \varepsilon, v)$$

$$\text{subject to} \quad \begin{cases} \displaystyle\sum_{j=1}^{C} u_{i,j} = 1, \\ \|\varepsilon_i\|^2 \le \kappa_i^2 \quad (\kappa_i > 0), \end{cases}$$

where

$$J_{\mathrm{t}}(u, \varepsilon, v) = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{i,j}^m \|x_i + \varepsilon_i - v_j\|^2. \tag{9}$$

The optimal solutions u, ε and v are obtained by the following algorithm.

Algorithm 2 (FCM-T)

Step 1 *Give the number of clusters C, the fuzzifier parameter m and the maximum tolerance set κ for data x. Set the initial values of ε and v.*

Step 2 *Calculate u such that*

$$u_{i,j} = 1 / \sum_{k=1}^{C} \left(\frac{d_{i,j}}{d_{i,k}} \right)^{1/(m-1)}, \tag{10}$$

where

$$d_{i,j} = \|x_i + \varepsilon_i - v_j\|^2. \tag{11}$$

Step 3 *Calculate ε such that*

$$\varepsilon_i = -\alpha_i \left(x_i - \sum_{j=1}^{C} u_{i,j}^m v_j \right), \tag{12}$$

where

$$\alpha_i = \min \left\{ \kappa_i \left\| x_i - \sum_{j=1}^{C} u_{i,j}^m v_j \right\|^{-1}, \left(\sum_{j=1}^{C} u_{i,j}^m \right)^{-1} \right\}. \tag{13}$$

Step 4 *Calculate v such that*

$$v_j = U_j^{-1} \sum_{i=1}^{N} u_{i,j}^m (x_i + \varepsilon_i), \tag{14}$$

where

$$U_j = \sum_{i=1}^{N} u_{i,j}^m. \tag{15}$$

Step 5 *Check the stopping criterion for (u, ε, v). If the criterion is not satisfied, go back to* Step 2.

If the maximal tolerance κ_i is set to zero, this algorithm is coincides with FCM without tolerance. This algorithm is executing FCM without tolerance for data moving within a region as shown in Fig. 1.

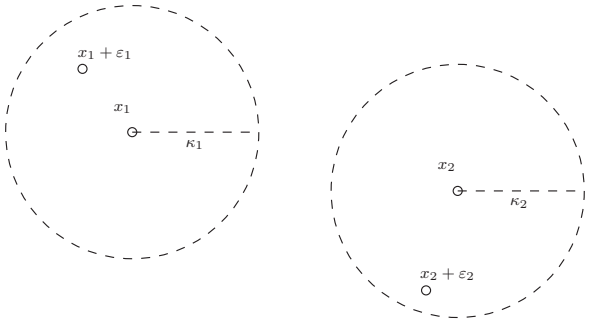


Figure 1: Data with Tolerance

3 Fuzzy c-Lines for Data with Tolerance

In this section, fuzzy c-lines for data with tolerance (FCL-T) is proposed. First, the tolerance is formulated and introduce into optimization problem of FCL. Next, the problem is solved using Karush-Kuhn-Tucker conditions. Last, the algorithm is constructed based on the results of solving the problem.

3.1 Optimization Problem and Its KKT conditions

In this subsection, a new optimization problem is proposed such that the tolerance is formulated and introduce into optimization problem of FCL, and its Karush-Kuhn-Tucker conditions are considered.

Let us consider the following optimization problem:

$$\underset{u,\varepsilon,v,s}{\text{minimize}}\, J_{\ell,\mathrm{t}}(u, \varepsilon, v, s)$$

$$\text{subject to} \quad \begin{cases} \displaystyle\sum_{j=1}^{C} u_{i,j} = 1, \\ \|s_j\| = 1, \\ \|\varepsilon_i\|^2 \le \kappa_i^2, \end{cases} \tag{16}$$

ISBN: 978-989-95079-6-8

where

$$J_{\ell,\mathrm{t}}(u,\varepsilon,v,s) = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{i,j}^{m}\left(\|x_i+\varepsilon_i-v_j\|^2-((x_i+\varepsilon_i-v_j)s_j)^2\right). \tag{17}$$

Lagrange function of this problem $L_{\ell,\mathrm{t}}(u,\varepsilon,v,s)$ is described as below:

$$\begin{aligned} L_{\ell,\mathrm{t}}(u,\varepsilon,v,s) =& J_{\mathrm{s},\ell,\mathrm{t}}(u,\varepsilon,v,s) + \sum_{i=1}^{N}\gamma_i\left(\sum_{j=1}^{} u_{i,j}-1\right) \\ &+\sum_{j=1}^{C}\xi_j\left(s_j^{\mathrm{T}}s_j-1\right) \\ &+\sum_{i=1}^{N}\delta_i\left(\|\varepsilon_i\|^2-\kappa_i^2\right). \end{aligned} \tag{18}$$

where $\gamma = (\gamma_1,\cdots,\gamma_N)$, $\xi = (\xi_1,\cdots,\xi_C)$ and $\delta = (\delta_1,\cdots,\delta_N)$ are Karush-Kuhn-Tucker vectors. Karush-Kuhn-Tucker (KKT) conditions are described as below:

$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial u_{i,j}} =0, \tag{19}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial v_j} =0, \tag{20}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial s_j} =0, \tag{21}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial \varepsilon_i} =0, \tag{22}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial \gamma_i} =0, \tag{23}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial \xi_j} =0, \tag{24}$$
$$\frac{\partial L_{\ell,\mathrm{t}}}{\partial \delta_i} =0, \tag{25}$$
$$\delta_i\frac{\partial L_{\ell,\mathrm{t}}}{\partial \delta_i} =0, \tag{26}$$
$$\delta_i \geq 0. \tag{27}$$

3.2 Optimal Solution of $u_{i,j}$, v_j and s_j

In this subsection, we derive the optimal solution of $u_{i,j}$, v_j and s_j from the above KKT conditions, and we show that these solutions are just substituting $x_i+\varepsilon_i$ to x_i in FCL (Algorithm 1).

From KKT conditions (19) and (23), we have the optimal solution of $u_{i,j}$ as

$$u_{i,j} = 1/\sum_{k=1}\left(\frac{d_{i,j}}{d_{i,k}}\right)^{1/(m-1)}, \tag{28}$$

where

$$d_{i,j} = \|x_i+\varepsilon-v_j\|^2-(s_j^{\mathrm{T}}(x_i+\varepsilon_i-v_j))^2. \tag{29}$$

From KKT conditions (20), we have

$$(E-s_js_j^{\mathrm{T}})\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i-v_j)=0, \tag{30}$$

where E is the p-dimensional unit matrix. The null space of the matrix $E-s_js_j^{\mathrm{T}}$ include s_j since

$$(E-s_js_j^{\mathrm{T}})s_j = s_j-(s_j^{\mathrm{T}}s_j)s_j=0. \tag{31}$$

For the p-dimensional vector w such as $w^{\mathrm{T}}s_j=0$ and as $w^{\mathrm{T}}w=1$ and for a scalar value α, we have

$$\begin{aligned}(E-s_js_j^{\mathrm{T}})(\alpha w) =& \alpha(w-(s^{\mathrm{T}}w)s_j)\\ =&\alpha w,\end{aligned} \tag{32}$$

which is equal to 0 if and only if $\alpha=0$, hence, the null space of $E-s_js_j^{\mathrm{T}}$ is just s_j. Therefore, Eq.(30) indicates that

$$\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i-v_j)=as_j \tag{33}$$

for an arbitrary value of a. From the above, we have the optimal solution of v_j as

$$v_j=\frac{\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i)-as_j}{\sum_{i=1}^{N}u_{i,j}^{m}}, \tag{34}$$

in which we may set $a=0$ and have

$$v_j=\frac{\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i)}{\sum_{i=1}^{N}u_{i,j}^{m}}. \tag{35}$$

From KKT conditions (21) and (24), we have

$$\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i-v_j)(x_i+\varepsilon_i-v_j)^{\mathrm{T}}s_j=-\xi_js_j, \tag{36}$$

from which we can obtain the optimal solution of s_j as the eigenvector for the maximal eigenvalue of the matrix

$$\sum_{i=1}^{N}u_{i,j}^{m}(x_i+\varepsilon_i-v_j)(x_i+\varepsilon_i-v_j)^{\mathrm{T}}. \tag{37}$$

Comparing between Eq. (28) and (4), Eq. (29) and (5), Eq. (35) and (6), and Eq. (37) and (8), we can see that optimal solutions in FCL-T is obtained just by substituting $x_i+\varepsilon_i$ to x_i in FCL (Algorithm 1).

3.3 Optimal Solution of ε_i

In this subsection, we obtain the optimal solution of ε_i from the KKT conditions.

From KKT conditions (22), (25), (26) and (27), we have

$$\left(\delta_iE+\sum_{j=1}^{C}u_{i,j}^{m}(E-s_js_j^{\mathrm{T}})\right)\varepsilon_i+ \sum_{j=1}^{C}u_{i,j}^{m}(E-s_js_j^{\mathrm{T}})(x_i-v_j)=0, \tag{38}$$

$$\delta_i(\|\varepsilon_i\|^2-\kappa_i^2)=0. \tag{39}$$

ISBN: 978-989-95079-6-8

In order to solve this equation, we discuss in each case that the matrix $\left(\sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)$ is invertible or not as the followings.

The matrix $\left(\sum_{j=1}^{C} u_{i,j}^m \left(E - s_j s_j^{\mathrm{T}}\right)\right)$ is not invertible if $s_j = s_{\tilde{j}}$ for any $j \neq \tilde{j}$ or if there exists j^* such as $u_{i,j^*} = 1$, though the proof is omitted because of the sake of the pages. Here, we write $s = s_j$ in the case of $s_j = s_{\tilde{j}}$ for any $j \neq \tilde{j}$ and also write $s = s_{j^*}$ in the case that there exists j^* such as $u_{i,j^*} = 1$. In these cases, Eq. (38) is rewritten by setting $\varepsilon_i = \alpha s + \beta w$ with certain values α and β, and a p-dimensional vector w such as $s^{\mathrm{T}} w = 0$ and as $w^{\mathrm{T}} w = 1$, as

$$\beta(\sum_{j=1}^{C} u_{i,j}^m) w + \delta_i(\alpha s + \beta w) + (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j) = 0. \quad (40)$$

Multiplying s^{T} to Eq.(40), we have $\delta_i \alpha = 0$, which indicates $\delta_i = 0$ or $\alpha = 0$. $\alpha = 0$ means that ε_i is orthogonal to s. In the case of $\delta_i = 0$, we have

$$(E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i + \varepsilon_i - v_j) = 0. \quad (41)$$

which implies

$$\sum_{j=1}^{C} u_{i,j}^m (x_i + \varepsilon_i - v_j) = as \quad (42)$$

with an arbitrary value a. Hence, we have the optimal solution of ε_i in the case that $\left(\sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)$ is not invertible with $\delta_i = 0$, as

$$\varepsilon_i = \frac{as - \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j)}{\sum_{j=1}^{C} u_{i,j}}. \quad (43)$$

If we set a such that ε_i is orthogonal to s, we can discuss both the case of $\delta_i \alpha = 0$ and the base of $\alpha = 0$ as the followings. Since $s^{\mathrm{T}} \varepsilon_i = 0$, Eq.(38) is rewritten as

$$\left(\sum_{j=1}^{C} u_{i,j}^m\right) \varepsilon_i + \delta_i \varepsilon_i + (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j) = 0, \quad (44)$$

from which we have

$$\varepsilon_i = -\frac{(E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j)}{\sum_{j=1}^{C} u_{i,j}^m + \delta_i}. \quad (45)$$

If $\delta_i = 0$, we have

$$\varepsilon_i = -\frac{(E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j)}{\sum_{j=1}^{C} u_{i,j}^m}. \quad (46)$$

If $\delta_i \neq 0$, then $\|\varepsilon_i\| = \kappa_i^2$ from Eq. (39), by which we have

$$\|\varepsilon_i\|^2 = \left\| \frac{(E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j)}{\sum_{j=1}^{C} u_{i,j}^m + \delta_i} \right\|^2 = \kappa_i^2. \quad (47)$$

By solving Eq.(47) as

$$\sum_{j=1}^{C} u_{i,j}^m + \delta_i = \frac{\left\| (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j) \right\|}{\kappa_i}, \quad (48)$$

we have

$$\varepsilon_i = -\kappa_i \left\| (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j) \right\|^{-1} \cdot (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j). \quad (49)$$

Getting Eq.(46) and (49) together, we have the optimal solution of ε_i in the case that the matrix $\left(\sum_{j=1}^{C} u_{i,j}^m \left(E - s_j s_j^{\mathrm{T}}\right)\right)$ is not invertible

$$\varepsilon_i = -\min \left\{ \kappa_i \left\| (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j) \right\|^{-1}, \left(\sum_{j=1}^{C} u_{i,j}^m\right)^{-1} \right\} (E - ss^{\mathrm{T}}) \sum_{j=1}^{C} u_{i,j}^m (x_i - v_j). \quad (50)$$

In the case that $\left(\sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)$ is invertible, we cannot obtain the explicit form of the optimal solution ε_i, so, we apply some numerical method for Eq. (38) and (39) as the followings. First, we set $\delta_i = 0$ and solve Eq. (38) as

$$\varepsilon_i = -\left(\sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)^{-1} \cdot \sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})(x_i - v_j). \quad (51)$$

If the obtained value of ε_i satisfies $\|\varepsilon_i\|^2 \leq \kappa_i^2$, we adopt this value as the optimal solution of ε_i. Otherwise, we solve Eq. (38) with $\|\varepsilon_i\|^2 = \kappa_i^2$ numerically.

ISBN: 978-989-95079-6-8

3.4 Algorithm

In this subsection, we propose an iterative algorithm of fuzzy c-lines for data with tolerance from the above discussion.

Algorithm 3 (FCL-T)

Step 1 *Give the values of the fuzzifier parameter m and the maximal tolerance set κ. Set the initial cluster center v, s and ε.*

Step 2 *Calculate $u_{i,j}$ such that*

$$u_{i,j} = 1/\sum_{k=1}^{C}\left(\frac{d_{i,j}}{d_{i,k}}\right)^{1/(m-1)}, \tag{52}$$

where

$$d_{i,j} = \|x_i + \varepsilon - v_j\|^2 - (s_j^{\mathrm{T}}(x_i + \varepsilon_i - v_j))^2. \tag{53}$$

Step 3 *Calculate v_j such that*

$$v_j = \frac{\sum_{i=1}^{N} u_{i,j}^m (x_i + \varepsilon_i)}{\sum_{i=1}^{N} u_{i,j}^m}. \tag{54}$$

Step 4 *Obtain s_j as the eigenvector with the maximal eigenvalue of the matrix*

$$\sum_{i=1}^{N} u_{i,j}^m (x_i + \varepsilon_i - v_j)(x_i + \varepsilon_i - v_j)^{\mathrm{T}}. \tag{55}$$

Step 5 *If $s_j = s_{\tilde{j}}$ for any $j \neq \tilde{j}$ or if there exists j such that $u_{i,j} = 1$, obtain ε_i as*

$$\varepsilon_i = -\min\left\{\kappa_i \left\|(E - ss^{\mathrm{T}})\sum_{j=1}^{C} u_{i,j}^m (x_i - v_j)\right\|^{-1}, \left(\sum_{j=1}^{C} u_{i,j}^m\right)^{-1}\right\}(E - ss^{\mathrm{T}})\sum_{j=1}^{C} u_{i,j}^m (x_i - v_j), \tag{56}$$

where

$$s = \begin{cases} s_j & \text{for} s_j = s_{\tilde{j}} \quad (j \neq \tilde{j}), \\ s_{j^*} & \text{for } u_{i,j^*} = 1, \end{cases} \tag{57}$$

and go to Step 7. *Otherwise, go to* Step 6.

Step 6 *Calculate*

$$\varepsilon_i = -\left(\sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)^{-1} \cdot \sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})(x_i - v_j). \tag{58}$$

If $\|\varepsilon_i\|^2 \leq \kappa_i^2$ holds, adopt this ε_i and go to Step 7. *Otherwise, solve the equation*

$$\left(\delta_i E + \sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})\right)\varepsilon_i + \sum_{j=1}^{C} u_{i,j}^m (E - s_j s_j^{\mathrm{T}})(x_i - v_j) = 0, \tag{59}$$

$$\|\varepsilon_i\|^2 = \kappa_i^2 \tag{60}$$

Step 7 *Check the stopping criterion. If the criterion is not satisfied, go back to* Step 2.

If the maximal tolerance κ_i is set to zero, this algorithm is coincides with FCL without tolerance (Algorithm 1). Remark that Eq. (56) in Algorithm 3 and Eq. (13) for obtaining the tolerance in Algorithm 2 are essentially the same with each other, because Eq. (56) means that the tolerance is obtaining in the orthogonal complementary space of s in the pattern space. Remark also that there is the quite similarity between Eq. (58) in Algorithm 3 and Eq. (9.29) in [3] for dealing with the missing data in linear fuzzy clustering followed by the optimal completion strategy [4], because Eq. (58) means that the optimal solution of the tolerance in a certain condition is not affected by the constraint of the tolerance $\|\varepsilon_i\| \leq \kappa_i^2$, while Eq. (9.29) in [3] does not consider the constraint for missing component of data from the beginning.

4 Numerical Example

In this section, we show some examples of classification by fuzzy c-lines for data with tolerance (Algorithm 3). In each example, 100 trials for Algorithm 3 with different initial cluster centers are tested and the solution with the minimal objective function value is selected as the final result. The classified dataset is Gustaffson's cross [5] shown in Fig. 2. This dataset consists of 20 points ($N = 20$) in $\mathbf{R}^2$ and forms two visually apparent linear clusters with the crossing shape.

We fix $m = 2$ and test four different values of $\kappa_i \in \{0, 0.016, 0.032, 0.064\}$. All cases of κ_i produce the correctly classified results shown in Fig. 3, 4 and 5, respectively. Especially, the cases of $\kappa_i \in \{0.032, 0.064\}$ produce the same results shown in Fig. 5. From Fig. 4, we can see that each data moves to the direction of its own prototype. From Fig. 5, we can see that each data moves just on its own prototype.

From these results, we consider the following hypothesis: if Gustaffson's cross is generated by adding error followed by uniform distribution with a width after sampling from the two crossing exactly linear shaped population, FCL-T (Algorithm 3) produces the shape of the original population by $v_j + ts_j$ and also produces the width of the uniform distribution by κ_i, though we have not proved it yet.

5 Conclusion

In this paper, we proposed fuzzy c-lines for data with tolerance. The proposed method is obtained by introducing the idea of tolerance, which means that each datum moves within a region, into fuzzy c-lines. This is the analogy with that fuzzy c-means for data with tolerance is obtained by introducing the idea of tolerance into fuzzy c-means. In the numerical examples, we can see that each data moves to the direction of its own prototype in setting the positive maximal tolerance, and that each data moves on its own prototype in setting larger maximal tolerance.

ISBN: 978-989-95079-6-8

From this feature of the proposed method, we consider the following hypothesis: if dataset is generated by adding error followed by uniform distribution with a width after sampling from the exactly linear shaped population, FCL-T (Algorithm 3) produces the shape of the original population and also produces the width of the uniform distribution. But we must investigate the proposed method theoretically and through many numerical experiments, which is the first future work.

As another future works, (1) entropy regularlized fuzzy c-means [6] for data with tolerance will be proposed by the similar discussion, (2) the cases of error followed by other probabilistic distribution will be considered, e.g., the case of error followed by normal distribution can be considered by adding the penalty term of squared tolerance, and Laplace one by the penalty term of absolute value of tolerance.

References

[1] Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms, Plenun, New York, (1981).

[2] Endo, Y., Miyamoto, S.: "Fuzzy c-Means for Data with Tolerance", Proc. 2005 International Symposium on Nonlinear Theory and Its Applications, pp.123–127, 2005.

[3] Miyamoto, S., Ichihashi, H., Honda, K.: "Algorithms for Fuzzy Clustering", Springer, Berlin (2008).

[4] Hathaway, R.J., Bezdek, J.C.: "Fuzzy c-means Clustering of Incomplete Data", IEEE Trans. on Systems, Man and Cybernetics Part B 31(5), pp.735–744 (2001).

[5] Gustaffson, D.E., Kessel, W.: "Fuzzy Clustering with a Fuzzy Covariance Matrix", Proc. IEEE–CDC, Vol.2, pp.761–766 (1979).

[6] Miyamoto, S. and Mukaidono, M.: "Fuzzy c-Means as a Regularization and Maximum Entropy Approach", Proc. of the 7th International Fuzzy Systems Association World Congress (IFES'97), Vol. II, pp. 86–92, Prague, 1997.

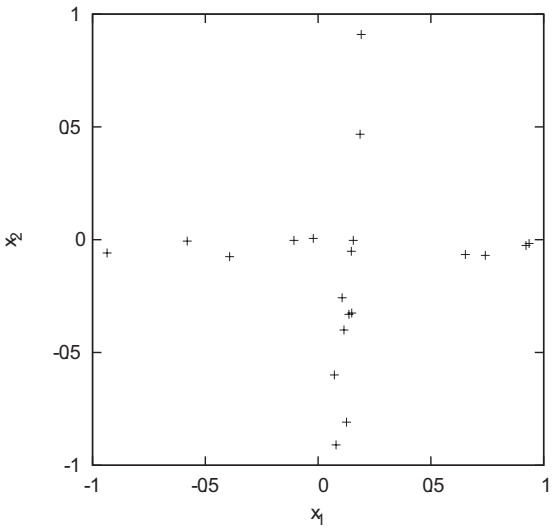


Figure 2: Gustaffson's Cross

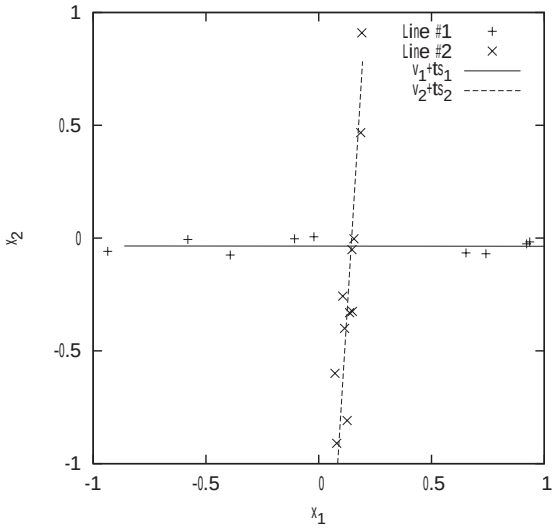


Figure 3: Successful Classification Result of Fig. 2 by FCL-T 3 with $m = 2$ and $\kappa_i = 0$

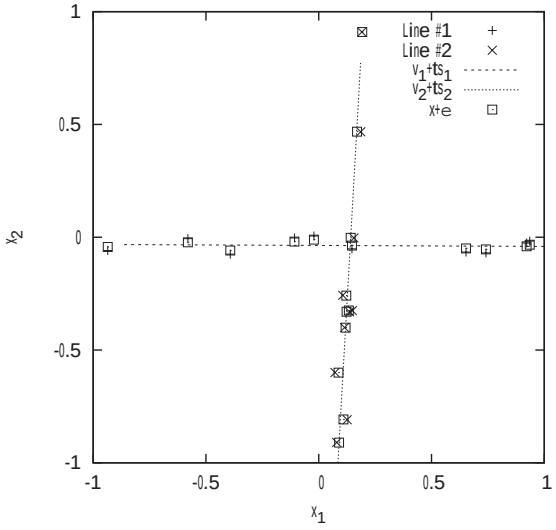


Figure 4: Successful Classification Result of Fig. 2 by FCL-T 3 with $m = 2$ and $\kappa_i = 0.016$

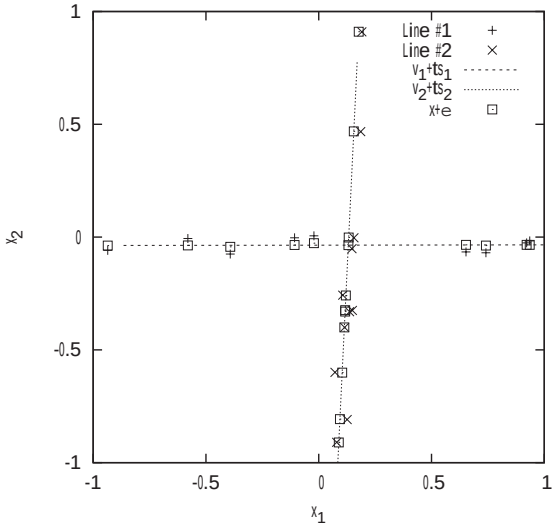


Figure 5: Successful Classification Result of Fig. 2 by FCL-T 3 with $m = 2$ and $\kappa_i = 0.032$

ISBN: 978-989-95079-6-8

Geographic Information Retrieval based on two orthogonal criteria

Gloria Bordogna [1] Giorgio Ghisalberti [2] Marco Pagani [1] Giuseppe Psaila [3]

1 CNR - IDPA, via Pasubio 5, c/o POINT, Dalmine 24044 (BG), Italy
2 Politecnico di Milano, Dip. Di Elettronica, via Bassini 16, 20133 Milano, Italy
3 Università di Bergamo, Facoltà di Ingegneria, via G. Marconi 5, 24044 Dalmine (BG) Italy
Email: gloria.bordogna@idpa.cnr.it, ghisalberti@elet.polimi.it, marco.pagani@idpa.cnr.it, psaila@unibg.it

Abstract— *A Geographic Information Retrieval (GIR) model is defined and the architecture of a GIR system, named Geo-Finder, is described. The GIR system is constituted by two main modules. The geo-indexing module applies bipolar criteria to automatically identify the geo-reference focus (footprint) of textual documents. The geo-retrieval module applies a context dependent matching function to evaluate queries consisting of two orthogonal constraints, a content constraint and a spatial constraint; the spatial constraint is defined by modeling the user's perception of geographic "closeness" between the documents' footprint and the query footprint. For each retrieved document, two relevance scores are computed with respect to the two query conditions, that can be combined to generate an overall ranked list of documents in a flexible way, by allowing users to specify a trade off between them.*

Keywords— bipolar criteria evaluation, geo-footprint, geo-indexing, geographic information retrieval, perception distance, spatial query evaluation.

1 Introduction

The representation and management of Geographic Information is becoming a hot topic in the research area of information retrieval [9]. Given that about 15% of the queries submitted to search engines contain geographic names [20], the effective indexing and retrieval of geographical information poses new challenges to the design of *location-based search services* [18].

Location based search services allow users to search into specialized repositories, and more generally on the Internet, documents describing both resources and specific contents in the *neighbourhood of the user location*, or in the *neighbourhood of a geographic location* that is explicitly specified in the request. "*In the neighbourhood of a location*" specifies a spatial constraint on the geographic content of the retrieved documents, typically a constraint on the geographic distance from the desired location, like in the request "*find Indian restaurants near Bergamo university*".

Nevertheless, actual search engines do not enhance the influence of the geographical information content in evaluating requests containing geographic names. For example, the previous request submitted to *Google* retrieves as first ranked web page "*hotels near Indiana University East*" that does not satisfy the spatial constraint to be close to Bergamo. Generally, search engines are able to find specific resources such as hotels, restaurants, hospitals, in the neighbourhood of a locality, e.g. "*find hospital near Milan*", but this works only for specific resources, and moreover the ranking does not depend on the distance from the specified locality.

Furthermore, the indexing process does not extract most of these places automatically; consequently many web pages related to the query are not displayed as a result. In fact, many of the results obtained by a Google search are relative to resources that have directly informed Google of their geographical position. More specialized services has been developed for automatically mapping documents. For example, *MetaCarta* [11] is a system that geo-tags news, and allows searches in which one can distinctly specify the content constraint and the location of the news. However, this system is unable to retrieve news in the neighbourhood of a desired locality.

The identification of the geographic names, and successively the selection of the *geographic reference focus* (footprint) of web pages have been dealt with in several papers [1][2] [8][15]. Other papers considered the problem of spatial query evaluation [3][13][23]. In [15] the problem of implicit location identification is considered.

In this paper, after introducing the main problems involved in the design of a GIR system, we describe our GIR model and the system, named *Geo-Finder*, implementing it.

The main characteristics of the proposed geo-indexing model is the integration of multiple bipolar criteria satisfaction degrees [4][6][12][17][25], computed based on context dependent rules. Some criteria have a positive influence on the selection of the geographic names as footprints of the document; others have a negative influence. The positive and the negative constraints are heterogeneous and express pieces of information of a different nature [4]. The footprint of a document is represented as a fuzzy set of geographic coordinates (i.e., latitude and longitude, identifying the location and extent on the Earth surface of the geographic names occurring in the document) with membership degrees expressing their strength in defining the document footprint.

Second, we propose a distance measure to model the evaluation of the spatial query constraint, named "user's perception" distance. This distance measure depends on several aspects modelling the user context, such as the *spatial scope* of the query.

Finally, the system user interface maps the bi-dimensional relevance of retrieved documents in a Cartesian plane. The axes correspond to two orthogonal criteria

ISBN: 978-989-95079-6-8

(content constraint, geographic constraint), and the distance from the origin is inversely proportional to the global relevance of the documents. The user can also obtain a unique ranked list by choosing a trade-off of the two criteria satisfaction degrees based on their linear combination.

2 Why Geographic Information Retrieval is difficult

Geographic Information Retrieval (GIR) can be considered a specialized area of Information Retrieval, with an emphasis on the geographic indexing and geographic retrieval. GIR deals with any kind of information, i.e., not just maps or images but also texts, that have some relation to one or more locations on the Earth's surface, i.e., geo-referenced information [22]. Most of the information available on the Internet and in digital libraries is implicitly geo-referenced.

Often the link to the place (geographic footprint) is encoded by a geographic name. Geographic indexing implies the identification of the geographic names in a text and their translation into footprints, which are two operations that imply the management of imprecision, ambiguity, and incompleteness.

Often geographic names are ambiguous [14][21], some of them are homonymous of general terms (e.g. "*Los Angeles"*), some others identify distinct places on the Earth (e.g."*Rome*") or are temporal (e.g."*Leningrad*", "*Petersburg*", "*St Petersburg*"), or even are local names whose recognition relies on the knowledge of the local language.

Further, some pseudo-names are imprecise (e.g. "*around Milan*") or implicitly mentioned (e.g. "*the capital of Italy*"), or depending on the context (e.g. "*highest peak*" implicitly identifies "*Mont Blanc*" in a text describing the Alps). In [19] it has been argued that, given all these characteristics, geographic indexing can be feasibly faced by considering large corpora of geographic knowledge and heuristic rules.

On the other side, geographic retrieval implies being able to retrieve documents whose geographic focus satisfies a spatial constraint specified in a query. Generally, the spatial constraint demands the document footprint to be "*in the neighbourhood of a place*", i.e., it is a constraint on the distance. However, in this context the geographic distance is not merely Euclidean, but is related to the user's spatial context of interest, that can be related to the human perception of the time needed to cover it [5].

When a user searches resources or documents on the internet that are close to his/her current location, the judgment on the distance is related to his/her perception, depending on the fact that he/she is walking, driving or flying. Then, the constraint on the distance depends on the request and users' context.

Besides distance, other topological constraints could be defined, e.g., inclusion, overlapping, at the south/north of a specific region. Also these constraints must be interpreted in a tolerant, approximate way.

Then, geographic retrieval can draw benefits by defining the spatial constraints as soft, context dependent constraints, admitting degrees of satisfaction.

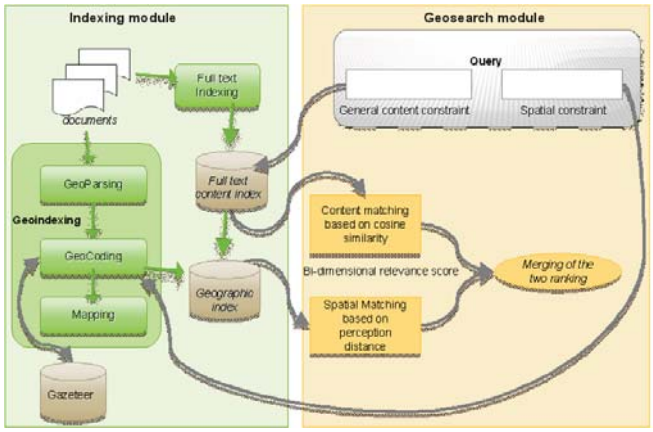


Figure1. Architecture of the Geographic Information Retrieval system *Geo-Finder*

3 The Architecture of the Geographic Information Retrieval System *Geo-Finder*

In this section we present the general architecture of the GIR system named *Geo-Finder*, that we designed and implemented. The architecture is depicted in Figure 1.

The system has the typical structure of an IRS, consisting of two main components: the *Indexing Module* and the *Retrieval Module* (named *GeoSearch*).

The Indexing Module has two main sub-modules: the *Full-Text Indexing sub-module* performs the full text indexing of the documents to represent their generic content, and generates the textual inverted index.

The *GeoIndexing sub-module* is the novel component, specialized in the identification of the footprints of documents, representing their geographic focus. This sub-module makes use of a *gazetteer* stored into a PostgreSQL database, containing the names of geographic entities (both administrative and physical entities) of all over the world, in English language and local languages (*GeoNames*) [10].

Once the *GeoIndexing* module has identified the footprints, it also stores them into the geographic index, that is a posting file containing the documents' identifiers and their footprints.

On the other side, the *GeoSearch* module interprets queries composed of two conditions. On the left hand side of the *GeoSearch* box in Figure 1, the generic content condition can be specified as a set of keywords, which define the *generic content constraint* on the documents' full text content representation. On the right hand side, the spatial condition can be specified as a geographic name, which is used to define the *spatial constraint* on the documents' footprints.

The *GeoIndexing* module parses the terms in the spatial condition to identify their footprint. These two conditions are evaluated by the sub-modules named *Content Matching module* (based on the *Lucene* library) and *Spatial Matching module*, based on perception distance (that is based on the original geo-retrieval model introduced in the next section), respectively.

In order to be retrieved, a document must satisfy, at least a little, the first content based condition (necessary condition), while the spatial condition is used for conditioning the ranking of the documents (optional

ISBN: 978-989-95079-6-8

condition). Then, the two conditions are *merged* by an "*and possibly*" aggregator, applied by the *Merging* sub-module that combines the satisfaction degrees of the two constraints, in order to obtain a unique value that performs the global ranking of documents. In the combination, one can give more preference to one of the two conditions, in order to emphasize the influence on the global ranking of either the document content or the document footprint.

4 The Geo-Retrieval Model

In this section, we describe the geo-retrieval model at the basis of the system implementation. In the first subsection the *GeoIndexing* model is described. In the second subsection, we introduce the geo-retrieval model at the basis of the *GeoSearch* component.

4.1 The GeoIndexing model

The GeoIndexing model is defined to identify for each document its footprint. A footprint of a document *d*, *Foot(d)*, is as a fuzzy set of geographic coordinates gc=(*lat,lon*), *lat*=latitude *lon*=longitude, expressed in degrees, with a membership degree $\mu_{Foot(d)}(gc)=GeoRef(gc) \in [0,1]$ representing the strength by which the geographic location gc, named gw, belongs to the footprint of the document d.

$$Foot(d)=\{GeoRef(gc_1)/gc_1,...,GeoRef(gc_n)/gc_n\}$$

A document is represented as a stream of tokens <*t*>. Some terms *t* have been selected as content indexes and thus are in the dictionary. For each of them, the frequency in the collection is known, while in the posting list we enter the documents in which they appear with their significance degrees *F(d,t)*. The significance degree is usually defined based on statistic analysis of the document text [16].

Besides this information, in the posting list we can also find the positions of the occurrences of the index terms in the document text, $occ_k(t_i,d)$, the k-th occurrence of t_i in *d*.

The identification of the document footprint is achieved in three steps, where distinct sets of heuristic rules are evaluated. Each set of rules acts as a filter on the input terms, so that only those terms whose global satisfaction degree of the set of rules is above a threshold are selected as input to the second step.

The *GeoIndexing* module implements this geoindexing model. It operates in two subsequent phases: first, it performs a G*eoParsing* that applies the first two sets of rules to detect the candidate geographic names (gw). Then, the *GeoCoding* sub-module identifies the document footprint *Foot(d)* and stores it in a file with the document unique identifier *d*.

The first set of rules consists of Name Entity Recognition (NER) rules, aimed at reducing the set of terms among which to successively select the candidate geographic names e.g.:

```
if Language(d)="English"∧
   FirstChar(t)=Capital then return(t).
```

The second set of rules receives, in input, a stream of previously selected terms, hereafter indicated by gw, and filters the candidate geographic names, a subset of the input terms. It applies bipolar [4] context dependent rules (r_i) exploiting a *gazetteer* [10] and computing independently a satisfaction degree *s(gw)* and a dissatisfaction degree *d(gw)* that denote to what extent gw is a geo-name and is not a geo-name, respectively.

The aggregation of the positive (negative) rules is done based on a Generalized Conjunction Disjunction function (GCD) [12], that, for distinct values of the parameter *p*, can model aggregations from completely compensative (or), where each rule can replace any other, to completely not compensative (and), where all rule must be satisfied simultaneously [6]:

$$s(gw)=\left(\sum_{i=1}^{m}\lambda_i*\left(r_i(gw)\right)^{ps}\right)^{1/ps}$$
$$d(gw)=\left(\sum_{j=1}^{n}\overline{\lambda}_j*\left(\overline{r}_j(gw)\right)^{pd}\right)^{1/pd} \quad (1)$$

The rules r_i and $\overline{r}_j$ assume a value in [0,1]. *ps* and *pd* are set so as to define (partially) compensative aggregations. In our experiment we used *ps=pd*=20, i.e., towards or like aggregation [6].

$\lambda_i,\overline{\lambda}_j \in [0,1]$ with $\sum_{i=1}^{m}\lambda_i=1, \sum_{i=1}^{n}\overline{\lambda}_j=1$ are the weights of the rules, i.e. their importance degrees in the aggregation, and are determined based on statistical analysis on a sample set of documents of a collection, and are set in a configuration file that is read by the *GeoParsing* sub-module during index generation. This way, the geoindexing can be suited to the characteristics of a collection.

The satisfaction of a rule r_i with *i=1,m* is interpreted as a hint of evidence that gw is a geographic name; thus the first *m* rules have a positive influence on the recognition of gw as a candidate geoname, like, e.g., the following two rules:

$$\text{If } gw \in gazetteer \wedge \exists gw_k \in d$$
$$\wedge\ gw_k = administrative\_distr(gw) \Rightarrow r_2(gw)=0.5$$
$$\wedge\ |occ_i(gw,d)-occ_j(gw_k,d)|<\Delta \Rightarrow r_2(gw)=r_2(gw)+0.5$$

E.g. if gw="*San Francisco*" and gw_k="*California*", its administrative district, occurs in the same document *d* at a maximum distance $\Delta=3$ words, then $r_2(gw)=1$, otherwise if the occurrences are at a greater distance than Δ the rule is only partially satisfied, i.e., $r_2(gw)=0.5$.

Another rule is the following:

$$\text{If } gw \in gazetteer \wedge gw_{-1} \in prefix \Rightarrow r_3(gw)=1$$

e.g. if gw="*Blanc*" is preceded by gw_{-1}="*Mount*" that belongs to the set prefix of prefixes of geographic names such as *Mount*, *lake*, *city*, *river*, then $r_3(gw)=1$.

Conversely, the satisfaction of a rule $\overline{r}_j$ with *j=1,n* is interpreted as a hint of evidence that gw is not a geographic name. Thus, these rules have a negative influence, like the following one:

$$\text{If } gw \in gazetteer \wedge gw \in Stopwords \Rightarrow \overline{r}_4(gw)=1$$

ISBN: 978-989-95079-6-8

e.g. gw ="Nice" is a stop-word too.

For each input gw, a geo-score *GeoScore*(gw)∈[0,1] is computed based on the values of *s*(gw) and *d*(gw) as follows:

$$GeoScore(\mathtt{gw}) = \begin{cases} s(\mathtt{gw}) - d(\mathtt{gw}) & if\ s(\mathtt{gw}) > d(\mathtt{gw}) \\ 0 & if\ otherwise \end{cases}$$

We select the gw with *GeoScore*(gw)$>\tau>0$ as reliable geographic names. This threshold allows restricting the footprint of a document to reduce the possibility of identifying false positives. τ must be set based on experimentations, and is also specified in the configuration file. The greater it is, the smaller is the possibility of selecting false geo-names. Generally, it is better to lose some true geo-names than to select false ones.

The third set of rules, used in the third step by the *GeoCoding* sub-module, is aimed at identifying, from the reliable geographic names, the geographic locations that belong to the document footprint (i.e., the fuzzy set *Foot*(*d*) of pairs of geographic coordinates).

For each of the selected geo-names, gw, a (set) of pair(s) of geographic coordinates $GC_{gw}=\{gc_1,\ldots,gc_n\}$, with gc=(lat,lon), is retrieved from the gazetteer. Each pair gc is a geocode of gw, uniquely identifying a geographic place on the geographic domain. Notice that homonymous geo-names have the same name but distinct gc pairs.

For each geocode associated with a selected gw, a *geo-reference score (GeoRef*(gc)) ∈[0,1] is computed, that expresses the strength by which gw, located in gc, belongs to the document footprint (i.e. it is marginal or central in defining the geographic focus of the document).

A true geographic name can be correctly identified in a document, but it can be meaningless in defining the geographic focus of the document itself. Let us consider, for example, the geographic names that are often present at the very end of web pages or in their footnotes: they generally have nothing to do with the document content, but are related with the affiliation of the web master, and thus must not define the document footprint.

Also these rules have a satisfaction degree $r'_1(gc)\in$ [0,1] that is dependent on some variable. For example, the frequency F(*d*,gw) of a geoname gw in a document *d*, increases the strength of its geocodes in the footprint proportionally to the degree by which the geoname was recognized as a candidate geo-name *s*(gw), and inversely proportional to the degree by which it was recognized as not being a geo-name *d*(gw). This is represented by the following rule:

$$r'_1(gc) = \mu_{significant}(F(d,gw) * GeoScore(gw))$$

with $gc\in GC_{gw}$ and $\mu_{significant}$ is a monotonic non-decreasing membership function.

Another rule evaluates the presence of another geoname gw_k in *d*, whose geo-code is geographically near to that of gw (at a maximum geographic distance *dist* equal top *Δ*). When this occurs, it increases *GeoRef*(gc). That is:

$$r'_2(gc) = max(0, \Delta - dist(gc, gc_k))$$

with $gc\in GC_{gw}$, $gc_k\in GC_{gwk}$ and $gw_k \in d$.

The population attribute of an administrative name gw, that has multiple geocodes, indicated by *population*(gc), is used to favour the influence of big cities w.r.t. small ones having the same geoname in the footprint:

$$r'_3(gc) = (population(gc))/max_k(population(gc_k))$$

$gw \in gazetteer,\ gc, gc_k \in GC_{gw}$

This rule allows us to resolve ambiguities, generated by homonymous geographic names, that is, to favour the geographic coordinates (geo-codes) that most likely belong to the document footprint.

These rules are aggregated based on a GCD aggregation, in which each rule has a weight $\lambda'_i\in[0,1]$, with $\sum_{i=1}^{m}\lambda'_i = 1$, determined based on statistical analysis and the parameter *p*=1, neutral aggregation (all configurable):

$$GeoRef(gc) = \left(\sum_{i=1}^{m}\lambda'_i\left(r'_i(gc)\right)^p\right)^{1/p} \quad (2)$$

A minimum threshold *Φ*>0 on *GeoRef*(gc) (configurable) restricts the footprint of a document to reduce its extent. Therefore, at the end of this model, for each geoname we have two different scores: *GeoScore*(gw)$>\tau>0$ and *GeoRef*(gc)$>\Phi>0$. Figure 2 shows the geonames extracted from a sample paper. The map was generated by the *Mapping module*: this module creates distinct graphic metafiles for each document to map the footprint in distinct graphic environments, such as Gmaps for Google maps (see figure 2), KML for Google Earth, and GPX for GPS data format environments .

Score	GeoLocation	Name	Latitude	Longitude	Country code	Population
0,54	Milano	Milan	45.461216	9.187832	IT	1306661
1,00	Italy	Italian Republic	42.833332	12.833333	IT	58103033
0,87	Abano	Abano Terme	45.360992	11.791407	IT	18206
0,76	Ferrara	Ferrara	44.82678	11.620713	IT	130992
0,78	Vicenza	Vicenza	45.557293	11.540898	IT	107223
0,92	Tuscany	Tuscany	43.416668	11.0	IT	3573454
0,63	Apulia	Andria	41.23	16.29	IT	4021957

Figure 2. Google map of a document footprint, generated by *Geo-Finder* mapping module. The hue of the red pins represents the degree of the *geo-Reference* scores of the localities in the document footprint. A table with the geographic coordinates in the footprint and their country code and population attribute is also reported.

ISBN: 978-989-95079-6-8

4.2 The GeoRetrieval model

The GeoRetrieval model takes the footprint (*Foot(q)*) of the spatial query condition *q*, that consists of a single or a set of geographic names. The footprint *Foot(q),* if not directly found in the geographic index, is identified by applying the *GeoParsing* and *GeoCoding* rules described in the previous subsection.

For each document d that has been retrieved by the *Content-based Matching* module, we match its footprint, *Foot(d)* w.r.t. the query Footprint *Foot(q)* by applying formula (3). This matching function defines the semantics of the soft spatial constraint "*close*" that computes a degree of satisfaction $GRS(d) \in [0,1]$ (*GeoRelevanceScore*) as follows:

$$GRS(d) = \mu_{close}\left(Foot(d), Foot(q)\right) \tag{3}$$

$$\mu_{close}(Foot(d), Foot(q)) =$$

$$= \max_{i,j}\left[\mu_{Foot(d)}[i] * \mu_{Foot(q)}[j] * \left(1 - \frac{qscope(dist(i,j))}{Maxdist(Foot(d) \cup Food(q))}\right)\right]$$

with

$$\mu_{Foot(e)}[i] = GeoRef(gc_i) > 0$$

and

$$qscope(x) = \begin{cases} x & if \quad x \le \delta + k * Maxdist(Foot(q)) \\ & with \qquad \delta \ge 0 \quad k > 0 \\ 0 & otherwise \end{cases}$$

and *Maxdist(X)* is the maximum geographic distance (*dist*) of any two elements in the footprint *X*.

i and *j* represent the i-th and j-th pairs of geographic coordinates *latitude* and *longitude* in the footprints of the document *d,* with membership degrees *GeoRef*(*i*), and of the query *q*, with membership degrees *GeoRef*(*j*) that is assumed equal to 1.

The *qscope* function models the "user perception" distance measure, where *Maxdist(Foot(q))* is the query footprint maximum dispersion. δ and k are constant values that define the query scope and are set without user explicit input.

We consider four main query scopes, a *full* scope considering the whole globe, a *large* scope, considering an area covered by a continent or a big country like Russia, a *meso* scope considering an area of a nation or a big region, and a *small* scope considering a city and its surroundings. Each scope has specific values for the parameters (e.g. *k=3, δ=10km* is associated with a small scope, *k=30*, *δ=100km* with a *meso* scope).

δ is the query *range*, and is useful in the case of a query footprint consisting of a single geographic coordinate pair in order to retrieve also documents with footprint in the surrounding places. Distinct δ can adapt the evaluation of the spatial constraint "*close*" to the user perception; thus, modelling strict or relaxed interpretations of the *close* surrounding of a point. *k* allows to model a tolerance on the geographic distance between a document footprint and the query footprint that is equal to *k* times the query maximum dispersion, i.e., *Maxdist(Foot(q))*. This allows enlarging or reducing the query scope. This parameter can be related to the scale of the map needed to represent the minimum bounding box of the Minkowski sum of *Foot(q)* and a circle of radius *k* [7].

For example, if one specifies the two geonames *Bergamo*, *Como* (*Como* being at about 40km from Bergamo*)* as spatial condition, and the query scope is *small* (i.e. *k=3* and *δ=10km*) documents with footprints at a maximum distance of 130 km from the query footprint are retrieved (e.g. both documents in *Milano* and *Lugano* are retrieved while a document with a footprint in *Rome* is not).

On the other side, a query with footprint in *Bergamo, Dalmine* (10 *km* from *Bergamo* in *Milano* direction) will retrieve documents at a maximum distance from the query footprint of 40 km, (e.g. it will retrieve just the document in "*Milano*" and not the one in "*Lugano*").

Figure 3 depicts the *GeoSearch* user interface of the GIR system. At the top, there are two text forms for submitting the content (left) and spatial (right) query condition.

Bottom, on the right panel the bi-dimensional relevance domain is depicted, in with each point corresponds to a retrieved document. The origin identifies the query. The document's X coordinate (Y coordinate) is its relevance degree w.r.t. spatial (general content) query constraint. The closest the document is to the origin, the most relevant it is with respect to at least one query constraint.

On the left panel, the ordered list of retrieved documents is reported corresponding with a merging of the two relevance degrees giving equal importance to the two conditions. By moving the sliding bar at the top of this panel, it is possible to modify the preference between the two conditions and thus to re-rank the documents accordingly. This is achieved by a linear combination of the relevance rankings.

5 Conclusions

The system has undergone a first evaluation based on a collection of 1100 documents in *Italian* and *English* with an average length of 800 words. The collection comprehends research papers on geological studies carried out at IDPA CNR, Reuters news of the RCV1 collection, and web pages of the Open Directory Project.

This first evaluation was aimed at estimating the ability of the *GeoParser* to identify the correct geonames in documents text. For each document, we classified its geo-names and we compared the classification with respect to the footprints of the documents. We achieved a recall of 91% and a precision of 93%. Nevertheless, these results are preliminary, and further tests are needed to complete the evaluation.

The contributions of this proposal with respect to current practice are several: first of all, the computation of a fuzzy footprint to represent the geo-reference focus of a textual document, based on a bipolar criteria decision process; second, the use of a user's perception based distance measure at the basis of the computation of the degree of satisfaction of the spatial query constraint; third, the evaluation of a bi-dimensional relevance score, and the possibility to flexibly merge the two rankings into a single one by specifying a relative importance weight of the two query constraints.

ISBN: 978-989-95079-6-8

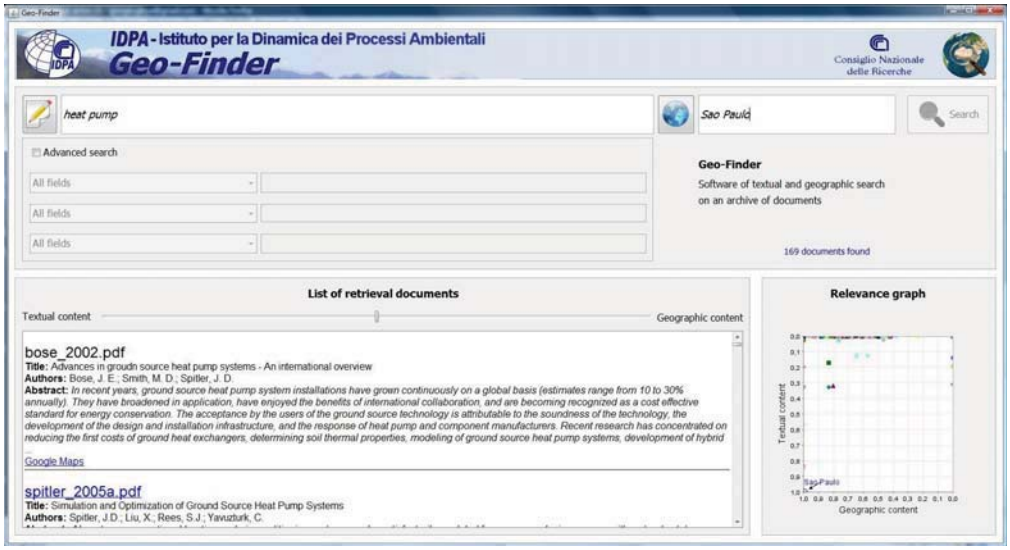


Figure 3. *GeoSearch* user interface of *Geo-Finder*: at the top the two query fields for specifying the content based condition="*heat pump*" and spatial condition="*Sao Paulo*"); below the two results panels. On the right side the bi-dimensional relevance graph, on the left the ranked list (merging with equal priority the two relevance scores).

References

[1] Amitay E., Har N., et al Web-a-where: Geotagging Web Content. SIGIR 2004.

[2] C. Wang, X. Xie et al., Detecting Geographical Locations from Web Resources. GIR'05, Bremen, Germany, 2005.

[3] Cai, G., "GeoVSM: An integrated Retrieval model for geographic information" M.J. Egenhofer and D.M. Marks Eds: GIScience 2002, LNCS 2478, Springer Verlag, 65-79, 2002.

[4] D. Dubois, H. Prade, An Introduction to Bipolar Representations of Information and Preference, Int. J. of Intelligent Systems, 23, 866–877, 2008.

[5] Egenhofer M., J., Mark D., M., Naïve Geography, eds Frank A.U., Kuhn W., Spatial Information Theory: a theoretical foundation for GIS . Springer Verlag LNCS 998, 1-15. 1995,

[6] G.Bordogna, G. Psaila, Modeling Soft Conditions with Unequal Importance in Fuzzy Databases based on the Vector p-norm, in Proc. of the IPMU Malaga, 2008.

[7] G.Bordogna, M. Pagani, G.Pasi, G. Psaila, Modeling uncertainty in Location-based spatial queries evaluation, Fuzzy Sets and Systems, accepted, september 2008.

[8] Hill L., "Introduction to Georeferencing in Digital Libraries Tutorial Document", University of California, Santa Barbara Aug. 2003.

[9] Himmelstein M.. Local search: The internet is the yellow pages. Computer, 38(2):26–34, 2005.

[10] http://www.library.uu.nl/kaartenzaal/Softwaresite/Beschrijvingen/geoname.html.

[11] http://www.metacarta.com/

[12] J.J. Dujmovic´, H. Larsen, Properties and modelingof partial conjunction/disjunction, in: B. De Baets (Ed.), Proceedings of Eurofuse Workshop on Data and Knowledge Engineering. Published as Current Issues in Data and Knowledge Engineering, Exit Publishers, Warsaw, 215–224, 2004.

[13] Jones et al, . "Spatial Information Retrieval and Geographical Ontologies: An Overview of the SPIRIT project" SIGIR 2002: Proc. of the 25th ACM SIGIR Conf.,Tampere, Finland', ACM Press, 387 – 388, 2002.

[14] Li H., Srihari R.K., Niu C., Li W., Infoxtract location normalization: a hybrid approach to geographic references in information extraction. In Proc. of the HLT-NAACL 2003 Workshop on Analysis of geographic references, 39-44.2003.

[15] Li Z., Wang C., Xie X., Wang X., Ma W.Y., Indexing implicit locations for geographical information retrieval, in Proc. of GIR'06, August 10–11, 2006, Seattle, USA, 2006.

[16] Lucene library http://lucene.apache.org/java/docs/

[17] Matthé T., De Tré G., (2009) Bipolar Query Satisfaction Using Satisfaction and Dissatisfaction Degrees, Proc. of the ACM SAC 2009, IAR track, 1699-1703.

[18] Mountain D. and MacFarlane A.. Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals. Journal of Information Science, 33(5):515–530, 2007.

[19] Reid J., A., "geoXwalk – A Gazetteer Server and Service for UK Academia" Koch/Sølvberg (Eds) Research and Advanced Technology for Digital Libraries: in Proc. of the 7th European Conf. ECDL 2003 Trondheim, Norway, August 17-22, 2003.

[20] Sanderson M. and Kohler J.. Analyzing geographic queries. In Workshop on Geographic Information Retrieval SIGIR, 2004.

[21] Smith D.A., Crane G., Disambiguating geographic names in historical digital library. In P. Costantoupolos and I. Solvberg eds, ECDL, LNCS 2163, Springer Verlag, 127-136, 2001.

[22] Vestavik Ø., Geographic Information Retrieval: an Overview, 2008.www.idi.ntnu.no/~oyvindve/article.pdf,

[23] Woodruff A.G., Plaunt C., GIPSY: Georeferenced Information Processing System, Journal of the American Society for Information Science, 1-21, 1994.

[24] Yasuda N., Toda H., Hirano T., Geographic Information Retrieval for just your surroundings, in Proc of the WWW08, April 21-25, Beijing China, 2008

[25] Zadrozny S., Kacprzyk J., (2007) Bipolar Queries Using Various Interpretations of Logical Connectives. IFSA (1), 181-190.

ISBN: 978-989-95079-6-8

A Semi-infinite Programming Approach to Possibilistic Optimization under Necessity Measure Constraints

Masahiro Inuiguchi[1]

1.Graduate School of Engineering Science, Osaka University
Toyonaka, Osaka, Japan
Email: inuiguti@sys.es.osaka-u.ac.jp

Abstract— *In this paper, possibilistic linear programming problems are investigated. After reviewing relations among conjunction and implication functions, necessity fractile optimization models with various implication functions are applied to the possibilistic linear problems. We show that the necessity fractile optimization models are reduced to semi-infinite linear programming problems. A simple numerical example is given to demonstrate the correctness of the result. The paper is concluded with some remarks for further developments.*

Keywords— Possibilistic linear programming, semi-infinite linear programming, necessity measure, implication function, conjunction function

1 Introduction

Fuzzy and possibilistic programming approaches are proposed to mathematical programming problems with ambiguity and vagueness [1, 2, 3]. By those approaches, we obtain reasonable solutions under conflicting soft constraints and goals, robust solutions under hard and soft constraints, hopeful solutions of attaining high-level goals, and so on. In possibilistic programming approaches, possibility and necessity measures are used to reduce the problems to the conventional programming problems. Many results demonstrate that possibilistic linear programming problems preserve the linearity in the reduced problems when possibility and necessity measures are defined by minimum operation and Dienes implication function. However, cases with the other conjunction and implication functions have not yet considerably investigated while several alternative approaches [4, 5] have been proposed in calculation of linear functions with fuzzy coefficients.

In this paper, we treat necessity fractile optimization model of possibilistic linear programming problems with soft constraints. We note that soft constraints include hard constraints as a special case. Then the problems are more general than problems with hard constraints. To construct necessity measures, we allow arbitrary implication satisfying weak requirements. Moreover, implication functions can be different among constraints. By the properties of implication functions, we show that necessity fractile optimization problems are reduced to a semi-infinite linear programming problems. Therefore the problems can be approximated by linear programming problems or the solutions can be obtained by a relaxation procedure together with the simplex method.

In next section, we briefly review the closure of generation procedures among conjunction and implication functions. Moreover, we introduce possibility and necessity measures. Possibilistic linear programming problems with soft constraints are given and reduced to the conventional programming problems through necessity fractile optimization models in Section 3. The main results are shown together with a conceivable solution procedure. in Section 4. In Section 5, a simple numerical example is given to demonstrate the correctness of the main result.

2 Conjunction and Implication Functions

2.1 Definitions

In this paper, a conjunction function is defined as a two-place function $T : [0,1] \times [0,1] \to [0,1]$ satisfying

(T0) T is lower semi-continuous, (semi-continuity)
(T1) $T(0,0) = T(0,1) = T(1,0) = 0$ and $T(1,1) = 1$, (boundary condition)
(T2) $T(a,b) \le T(c,d)$ if $0 \le a \le c \le 1$ and $0 \le b \le d \le 1$. (monotonicity)

A conjunction function T satisfies the following properties (t1) $T(a,1) = T(1,a) = a$ for any $a \in [0,1]$, (T3) $T(a,b) = T(b,a)$ for any $a, b \in [0,1]$ (commutativity) and (T4) $T(a,T(b,c)) = T(T(a,b),c)$ for any $a, b, c \in [0,1]$ (associativity) is known as a triangular norm.

In this paper, an implication function is a two-place function $I : [0,1] \times [0,1] \to [0,1]$ satisfying

(I0) I is upper semi-continuous, (semi-continuity)
(I1) $I(0,0) = I(0,1) = I(1,1) = 1$ and $I(1,0) = 0$, (boundary condition)
(I2) $I(a,b) \le I(c,d)$ if $0 \le c \le a \le 1$ and $0 \le b \le d \le 1$. (monotonicity)

Given a conjunction function and/or a strong negation n, an implication function can be generated through a transformation. The following three transformations are frequently adopted in the literature for implication generators:

$$I^{\mathrm{R}}[T](a,b) = \sup\{s \in [0,1] \mid T(a,s) \le b\}, \quad (1)$$

$$I^{\mathrm{S}}[T](a,b) = n(T(a,n(b))), \quad (2)$$

$$I^{\mathrm{r-R}}[T](a,b) = \sup\{s \in [0,1] \mid T(n(b),s) \le n(a)\}. \quad (3)$$

The first one, I^{R}, is encountered in the maximum solution of a fuzzy relation equation [6] and adopted in view of modus ponens. The second one, I^{S}, is introduced in analogy to Boolean logic. The last one, $I^{\mathrm{r-R}}$ is reciprocal to the first one which is obtained by taking a contraposition of the first one. When T is a t-norm, $I^{\mathrm{R}}[T]$, $I^{\mathrm{S}}[T]$ and $I^{\mathrm{r-R}}[T]$ are called R-implication (residual implication), S-implication and reciprocal R-implication, respectively. Whereas I^{S} produces an implication function from an arbitrary conjunction function T,

ISBN: 978-989-95079-6-8

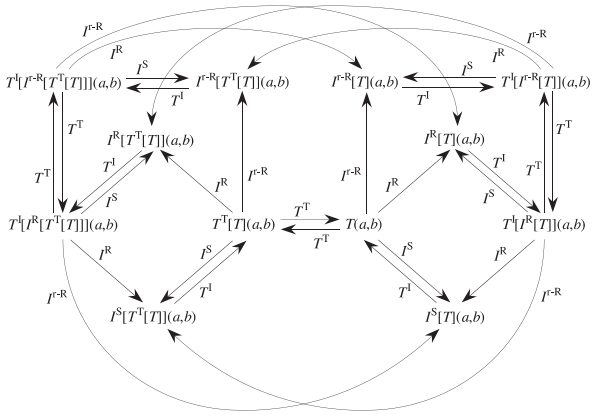


Figure 1: Conjunction and implication generation

I^{R} and $I^{\mathrm{r-R}}$ produce an implication function from a conjunction function which satisfies

$$T(1,a) > 0 \text{ for any } a > 0. \tag{4}$$

On the other hand, a conjunction function can be generated from an implication function through a transformation,

$$T^{\mathrm{I}}[I](a,b) = n(I(a,n(b))). \tag{5}$$

This transformation is symmetrical to I^{S}. From an implication function, a conjunction function is produced through T^{I}.

A conjunction function in this paper is not commutative. Thus, a new conjunction function may be generated from a conjunction function through

$$T^{\mathrm{T}}[T](a,b) = T(b,a). \tag{6}$$

2.2 Closure of generation processes

For the transformations (1)–(3), (5) and (6), the following equalities can be easily shown:

$$T^{\mathrm{I}} \circ I^{\mathrm{S}} = \mathrm{id.},\ I^{\mathrm{S}} \circ T^{\mathrm{I}} = \mathrm{id.},\ T^{\mathrm{T}} \circ T^{\mathrm{T}} = \mathrm{id.}, \tag{7}$$

$$I^{\mathrm{S}} \circ T^{\mathrm{T}} \circ T^{\mathrm{I}} \circ I^{\mathrm{R}} = I^{\mathrm{r-R}}, \tag{8}$$

where '$\circ$' denotes a composition, for example, $T^{\mathrm{I}} \circ I^{\mathrm{R}}$ is a composite transformation of I^{R} and T^{I}, i.e.,

$$T^{\mathrm{I}} \circ I^{\mathrm{R}}[T](a,b) = T^{\mathrm{I}}[I^{\mathrm{R}}[T]](a,b).$$

The notation 'id.' stands for the identical transformation.

From (T0), we have the following equalities [7]:

$$I^{\mathrm{R}} \circ T^{\mathrm{I}} \circ I^{\mathrm{R}}[T] = I^{\mathrm{S}}[T], \tag{9}$$

$$I^{\mathrm{r-R}} \circ T^{\mathrm{I}} \circ I^{\mathrm{R}} \circ T^{\mathrm{T}}[T] = I^{\mathrm{S}}[T], \tag{10}$$

$$I^{\mathrm{R}} \circ T^{\mathrm{I}} \circ I^{\mathrm{r-R}} \circ T^{\mathrm{T}}[T] = I^{\mathrm{r-R}}[T], \tag{11}$$

$$I^{\mathrm{r-R}} \circ T^{\mathrm{I}} \circ I^{\mathrm{r-R}} \circ T^{\mathrm{T}}[T] = I^{R}[T], \tag{12}$$

Equations (7)–(12) are summed up by Figure 1. As shown in Figure 1, the generation processes from a lower semi-continuous conjunction function as well as from an upper semi-continuous implication function are closed. This result is given by Inuiguchi and Sakawa [7] as the generalization of the result by Dubois and Prade [8]. Note that the semi-continuity is preserved through the generation processes [7].

As shown in Figure 1, we have six conjunction functions and six implication functions. The mappings among those twelve functions are given in Table 1. The inverse mappings are also indicated in Table 1. Some of those functions often appear in literatures on fuzzy relation equations, fuzzy logic, approximate reasoning and so on.

2.3 Possibility and necessity measures

Given a possible range V, the possibility and necessity measures $\Pi_V(S)$ and $N_V(S)$ of a fuzzy event S are defined by

$$\Pi_V(S) = \sup_{u \in U} T(\mu_V(u), \mu_S(u)), \tag{13}$$

$$N_V(S) = \inf_{u \in U} I(\mu_V(u), \mu_S(u)), \tag{14}$$

where μ_V and μ_S are membership functions of fuzzy sets V and S, respectively. U is the universal set. $\Pi_V(S)$ indicates the degree to what extent S is possible under possible range V, while $N_V(S)$ indicates the degree to what extent S is certain under possible range V.

Neither possibility measure nor necessity measure is unique. Therefore, we need to select possibility and necessity measures suitable for the given problem. By their definitions, selections of possibility and necessity measures are equivalent to selections of conjunction and implication functions. In the literature, minimum operation, i.e., $T(a,b) = \min(a,b)$ and Dienes implication, i.e., $I(a,b) = \max(1-a,b)$ are often used because they are simple and used in the original definitions of possibility and necessity measures [9, 10]. However, there is no guarantee that those possibility and necessity measures represent the decision maker's preferences well. Considering the variety of decision maker's preferences, the original possibility and necessity measures are not sufficiently flexible. Then the selections of possibility and necessity measures are important issues.

A conceivable approach to their selections are proposed by Inuiguchi and Tanino [11] and by Inuiguchi et al. [12]. The approach is based on the images of relations between two fuzzy sets represented by two modifier functions. Because of space limitations, we do not introduce this approach but we treat various kinds of necessity measures in the framework of possibilistic linear programming problems.

3 Linear Program with Necessity Measures

We consider a possibilistic linear programming problem,

$$\begin{array}{ll} \text{maximize} & \mathbf{c}^{\mathrm{T}}\mathbf{x}, \\ \text{subject to} & \mathbf{a}_i^{\mathrm{T}}\mathbf{x} \lesssim_i b_i,\ i = 1,2,\ldots,m, \\ & \mathbf{x} \geq \mathbf{0}, \end{array} \tag{15}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ is a decision vector. b_i, $i = 1,2,\ldots,m$ are constants. Components c_j and a_{ij} of $\mathbf{c}$ and $\mathbf{a}_i$ are not known exactly but the possible ranges of those values are known as fuzzy numbers C_j and A_{ij}, respectively. A fuzzy number is a normal, convex and bounded fuzzy set on the real line whose membership function is upper semi-continuous. The notation $\lesssim_i$ is a fuzzified inequality so that $\lesssim_i\ b_i$ corresponds to a fuzzy set B_i with verbal expression 'a set of real numbers which are roughly smaller than b_i'. We assume that the membership function μ_{B_i} of B_i is non-increasing and upper semi-continuous and satisfies $\mu_{B_i}(b_i) = 1$.

ISBN: 978-989-95079-6-8

Table 1: Mappings among twelve conjunction and implication functions

conjunction → implication		
Φ	$\Phi(a,b)$	inverse
$I^{\mathrm{R}}[T]$	$\sup\{s\in[0,1] \mid T(a,s)\le b\}$	$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{I}}[I]]]$
$I^{\mathrm{S}}[T]$	$n(T(a,n(b)))$	$T^{\mathrm{I}}[I]$
$I^{\mathrm{r-R}}[T]$	$\sup\{s\in[0,1] \mid T(n(b),s)\le n(a)\}$	$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]]$
$I^{\mathrm{R}}[T^{\mathrm{T}}[T]]$	$\sup\{s\in[0,1] \mid T(s,a)\le b\}$	$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{I}}[I]]]$
$I^{\mathrm{S}}[T^{\mathrm{T}}[T]]$	$n(T(n(b),a))$	$T^{\mathrm{T}}[T^{\mathrm{I}}[I]]$
$I^{\mathrm{r-R}}[T^{\mathrm{T}}[T]]$	$\sup\{s\in[0,1] \mid T(s,n(b))\le n(a)\}$	$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]]$
implication → conjunction		
Φ	$\Phi(a,b)$	inverse
$T^{\mathrm{I}}[I]$	$n(I(a,n(b)))$	$I^{\mathrm{S}}[T]$
$T^{\mathrm{T}}[T^{\mathrm{I}}[I]]$	$n(I(b,n(a)))$	$I^{\mathrm{S}}[T^{\mathrm{T}}[T]]$
$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{I}}[I]]]$	$\inf\{s\in[0,1] \mid I(a,s)\ge b\}$	$I^{\mathrm{R}}[T]$
$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{I}}[I]]]$	$\inf\{s\in[0,1] \mid I(b,s)\ge a\}$	$I^{\mathrm{R}}[T^{\mathrm{T}}[T]]$
$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]]$	$\inf\{s\in[0,1] \mid I(n(s),n(a))\ge b\}$	$I^{\mathrm{r-R}}[T]$
$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]]$	$\inf\{s\in[0,1] \mid I(n(s),n(b))\ge a\}$	$I^{\mathrm{r-R}}[T^{\mathrm{T}}[T]]$
conjunction → conjunction		
Φ	$\Phi(a,b)$	inverse
$T^{\mathrm{T}}[T]$	$T(b,a)$	$T^{\mathrm{T}}[T]$
$T^{\mathrm{I}}[I^{\mathrm{R}}[T]]$	$n(\sup\{s\in[0,1] \mid T(a,s)\le n(b)\})$	$T^{\mathrm{I}}[I^{\mathrm{R}}[T]]$
$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T]]$	$n(\sup\{s\in[0,1] \mid T(b,s)\le n(a)\})$	$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{T}}[T]]]$
$T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{T}}[T]]]$	$n(\sup\{s\in[0,1] \mid T(s,a)\le n(b)\})$	$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T]]$
$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{T}}[T]]]$	$n(\sup\{s\in[0,1] \mid T(s,b)\le n(a)\})$	$T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{T}}[T]]]$
implication → implication		
Φ	$\Phi(a,b)$	inverse
$I^{\mathrm{R}}[T^{\mathrm{I}}[I]]$	$\sup\{s\in[0,1] \mid I(a,n(s))\ge n(b)\}$	$I^{\mathrm{R}}[T^{\mathrm{I}}[I]]$
$I^{\mathrm{r-R}}[T^{\mathrm{I}}[I]]$	$\sup\{s\in[0,1] \mid I(n(b),n(s))\ge a\}$	$I^{\mathrm{R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$
$I^{\mathrm{R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$	$\sup\{s\in[0,1] \mid I(s,n(a))\ge n(b)\}$	$I^{\mathrm{r-R}}[T^{\mathrm{I}}[I]]$
$I^{\mathrm{S}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$	$I(n(b),n(a))$	$I^{\mathrm{S}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$
$I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$	$\sup\{s\in[0,1] \mid I(s,b)\ge a\}$	$I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I]]]$

By the extension principle, the possible ranges of $\mathbf{c}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{a}_i^{\mathrm{T}}\mathbf{x}$ are obtained as fuzzy sets $\mathbf{C}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{A}_i^{\mathrm{T}}\mathbf{x}$, respectively, defined by the following membership functions:

$$\mu_{\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}}(y)=\sup_{\substack{r_1,\ldots,r_n\\ \boldsymbol{r}^{\mathrm{T}}\boldsymbol{x}=y}}\min\left(\mu_{C_1}(r_1),\ldots,\mu_{C_n}(r_n)\right), \quad (16)$$

$$\mu_{\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{x}}(y)=\sup_{\substack{r_1,\ldots,r_n\\ \boldsymbol{r}^{\mathrm{T}}\boldsymbol{x}=y}}\min\left(\mu_{A_{i1}}(r_1),\ldots,\mu_{A_{in}}(r_n)\right), \quad (17)$$

where $\mathbf{r}=(r_1,\ldots,r_n)^{\mathrm{T}}$.

Using a necessity measure N^i defined by an upper semi-continuous implication function I^i, in this paper, we formulate Problem (15) as a necessity fractile optimization model (see Inuiguchi and Ramík [3]):

$$\begin{array}{ll}\text{maximize} & q,\\ \text{subject to} & N^0_{\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}}([q,+\infty))\ge h^0,\\ & N^i_{\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{x}}(B_i)\ge h^i,\ i=1,2,\ldots,m,\\ & \mathbf{x}\ge\mathbf{0},\end{array} \quad (18)$$

where q is an auxiliary variable. $h^0\in(0,1]$ and $h^i\in(0,1]$, $i=1,2,\ldots,m$ are certainty levels of goal achievement and constraint satisfactions specified by the decision maker. Note that we obtain

$$N^0_{\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}}([q,+\infty))=\inf_{r<q}I^0(\mu_{\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}}(r),0), \quad (19)$$

$$N^i_{\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{x}}(B_i)=\inf_{r}I^i(\mu_{\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{x}}(r),\mu_{B_i}(r)). \quad (20)$$

The selections of necessity measures and certainty levels depend on the required robustness of goal achievement/constraint satisfactions, the meanings of total goal achievement/constraint satisfactions, the estimation of fuzzy coefficients and so on. The method proposed by Inuiguchi and Tanino [11] and by Inuiguchi et al. [12] would be useful for selecting suitable necessity measures.

Let $[S]_h$ be a h-level set of a fuzzy set S, i.e., $[S]_h=\{u\in U \mid \mu_S(u)\ge h\}$. Then, because fuzzy numbers C_i and A_{ij} are bounded and have upper semi-continuous mem-

ISBN: 978-989-95079-6-8

bership functions, we have (see Dubois and Prade [13])

$$[\mathbf{C}^{\mathrm{T}}\mathbf{x}]_h = \sum_{j=1}^{n} [C_j]_h x_j, \tag{21}$$

$$[\mathbf{A}_i^{\mathrm{T}}\mathbf{x}]_h = \sum_{j=1}^{n} [A_{ij}]_h x_j. \tag{22}$$

Let $c_j^{\mathrm{L}}(h) = \inf[C_j]_h$, $c_j^{\mathrm{R}}(h) = \sup[C_j]_h$, $a_{ij}^{\mathrm{L}}(h) = \inf[A_{ij}]_h$ and $a_{ij}^{\mathrm{R}}(h) = \sup[A_{ij}]_h$. Then, considering the non-negativity of $\mathbf{x}$, we have

$$[\mathbf{C}^{\mathrm{T}}\mathbf{x}]_h = \left[\sum_{j=1}^{n} c_j^{\mathrm{L}}(h)x_j, \sum_{j=1}^{n} c_j^{\mathrm{R}}(h)x_j \right], \tag{23}$$

$$[\mathbf{A}_i^{\mathrm{T}}\mathbf{x}]_h = \left[\sum_{j=1}^{n} a_{ij}^{\mathrm{L}}(h)x_j, \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(h)x_j \right]. \tag{24}$$

To reduce Problem (18), the following theorem is useful.

Theorem 1. *Let N^i be a necessity measure defined by an implication function I^i. Then for any fuzzy sets V and S of a universal set U, we have*

$$\begin{aligned} & N_V^i(S) \geq h \\ & \Leftrightarrow \forall u \in U,\ \forall k \in [0,1]; \\ & \quad \mu_V(u) \geq k \text{ implies } \mu_S(u) \geq T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{I}}[I^i]]](k,h), \\ & \Leftrightarrow \forall k \in [0,1];\ [V]_k \subseteq [S]_{f^i(k,h)}, \end{aligned} \tag{25}$$

where $f^i(k,h) = T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{I}}[I^i]]](k,h)$.

(**Proof**) Because of the upper semi-continuity of I^i, Figure 1 and Table 1, we have

$$\begin{aligned} & N_V^i(S) = \inf_{u\in U} I^i(\mu_V(u), \mu_S(u)) \geq h \\ & \Leftrightarrow \forall u \in U;\ I^i(\mu_V(u), \mu_S(u)) \geq h \\ & \Leftrightarrow \forall u \in U;\ \mu_V(u) \leq \sup\{s \in [0,1] \mid I^i(s, \mu_S(u)) \geq h\} \\ & \Leftrightarrow \forall u \in U;\ \mu_V(u) \leq I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I^i]]](h, \mu_S(u)) \\ & \Leftrightarrow \forall u \in U,\ \forall k \in [0,1]; \\ & \quad \mu_V(u) \geq k \text{ implies } I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I^i]]](h, \mu_S(u)) \geq k \\ & \Leftrightarrow \forall u \in U,\ \forall k \in [0,1];\ \mu_V(u) \geq k \text{ implies} \\ & \quad \mu_S(u) \geq \inf\{s \in [0,1] \mid I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I^i]]](h,s) \geq k\} \\ & \Leftrightarrow \forall u \in U,\ \forall k \in [0,1];\ \mu_V(u) \geq k \text{ implies} \\ & \quad \mu_S(u) \geq T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{I}}[I^{\mathrm{r-R}}[T^{\mathrm{T}}[T^{\mathrm{I}}[I^i]]]]]](k,h) \\ & \Leftrightarrow \forall u \in U,\ \forall k \in [0,1];\ \mu_V(u) \geq k \text{ implies} \\ & \quad \mu_S(u) \geq T^{\mathrm{I}}[I^{\mathrm{R}}[T^{\mathrm{I}}[I^i]]](k,h). \end{aligned}$$

(Q.E.D.)

From the assumptions of B_i, we have $[B_i]_k = (-\infty, \bar{b}_i(k)]$, where $\bar{b}_i : [0,1] \to [0,+\infty)$ is defined by $\bar{b}_i(k) = \sup\{r \mid \mu_{B_i}(r) \geq k\}$. From (23), (24) and Theorem 1, Problem (18) is reduced to the following linear semi-infinite programming problem:

$$\begin{aligned} \text{maximize} \quad & q, \\ \text{subject to} \quad & \sum_{j=1}^{n} c_j^{\mathrm{L}}(k)x_j \geq \bar{q}(f^0(k,h^0)),\ \forall k \in [0,1], \\ & \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(k)x_j \leq b_i(f^i(k,h^i)),\ \forall k \in [0,1], \\ & \qquad\qquad i = 1,2,\ldots,m, \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{26}$$

where $\bar{q} : [0,1] \to \{-\infty, q\}$ is defined by

$$\bar{q}(k) = \begin{cases} q, & \text{if } k > 0, \\ -\infty, & \text{if } k = 0. \end{cases} \tag{27}$$

We showed that fractile optimization models based on necessity measures defined by any implication functions of possibilistic linear programming problems are reduced to linear semi-infinite programming problems. Thus, we can solve the problems by using linear semi-infinite programming techniques [14]. For example, let $\varepsilon > 0$ be a sufficiently small number and let $k_v = (v-1)(1-\varepsilon)/(l-1)+\varepsilon$, $v = 1,2,\ldots,l$. Then Problem (26) can be approximated by the following linear programming problem:

$$\begin{aligned} \text{maximize} \quad & q, \\ \text{subject to} \quad & \sum_{j=1}^{n} c_j^{\mathrm{L}}(k_v)x_j \geq \bar{q}(f^0(k_v,h^0)),\ v = 1,\ldots,l, \\ & \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(k_v)x_j \leq b_i(f^i(k_v,h^i)), \\ & \qquad\qquad v = 1,\ldots,l,\ i = 1,\ldots,m, \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{28}$$

Let $M = \{1,2,\ldots,m\}$. As another approach to computation of an approximate solution to Problem (26), the following algorithm based on a relaxation procedure is conceivable:

Algorithm

S1. Let $z = 0$ and $\varepsilon > 0$ be a sufficiently small positive number. Solve the following linear programming problem:

$$\begin{aligned} & \text{maximize } q, \\ & \text{subject to} \\ & \sum_{j=1}^{n} c_j^{\mathrm{L}}(k_0)x_j \geq \bar{q}(f^0(k_0,h^0)), k_0 = \varepsilon, 1, \\ & \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(k_i)x_j \leq b_i(f^i(k_i,h^i)),\ k_i = \varepsilon, 1,\ i \in M, \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Let $\mathbf{x}^z = (x_1^z,\ldots,x_n^z)^{\mathrm{T}}$ be the obtained optimal solution and q^z be the optimal value.

S2. Calculate $d_0 = \inf_{r\in[0,1]} \sum_{j=1}^{n} c_j^{\mathrm{L}}(r)x_j^z - \bar{q}(f^0(r,h^0))$. If $d_0 < 0$, we define $k_0^z = \arg\inf_{r\in[0,1]} \sum_{j=1}^{n} c_j^{\mathrm{L}}(r)x_j^z - \bar{q}(f^0(r,h^0))$ and otherwise, $k_0^z = 0$.

S3. For $i = 1,2,\ldots,m$, the following processes (a) and (b) are applied. (a) calculate $d_i = \sup_{r\in[0,1]} \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(r)x_j^z - b_i(f^i(r,h^i))$. (b) If $d_0 > 0$, we define $k_0^z = \arg\sup_{r\in[0,1]} \sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(r)x_j^z - b_i(f^i(r,h^i))$ and otherwise, $k_i^z = 0$.

ISBN: 978-989-95079-6-8

Table 2: Fuzzy numbers and constraints

$C_1 = \langle 5, 1\rangle$	$C_2 = \langle 7, 0.7\rangle$	
$A_{11} = \langle 2, 0.7\rangle$	$A_{12} = \langle 3, 0.5\rangle$	$B_1 = (230, 60\rangle$
$A_{21} = \langle 4, 1.5\rangle$	$A_{22} = \langle 2, 0.3\rangle$	$B_2 = (370, 100\rangle$
$A_{31} = \langle 1, 0.5\rangle$	$A_{32} = \langle 3, 0.3\rangle$	$B_3 = (140, 90\rangle$

S4. If $\sum_{i=0}^{m} k_i^z = 0$, terminate the algorithm. An approximate solution is obtained as $\mathbf{x}^z$ and q^z.

S5. Let $z = z + 1$. Solve the following linear programming problem:

$$
\begin{aligned}
&\text{maximize } q,\\
&\text{subject to}\\
&\sum_{j=1}^{n} c_j^{\mathrm{L}}(k_0)x_j \geq \bar{q}(f^0(k_0, h^0)), k_0 = \varepsilon, 1,\\
&\sum_{j=1}^{n} c_j^{\mathrm{L}}(k_0^w)x_j \geq \bar{q}(f^0(k_0^w, h^0)), w = 1, \ldots, z\\
&\sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(k_i)x_j \leq b_i(f^i(k_i, h^i)),\ k_i = \varepsilon, 1,\ i \in M,\\
&\sum_{j=1}^{n} a_{ij}^{\mathrm{R}}(k_i^w)x_j \leq b_i(f^i(k_i^w, h^i)),\ w = 1, \ldots, z,\ i \in M,\\
&\mathbf{x} \geq \mathbf{0}.
\end{aligned}
$$

Let $\mathbf{x}^z = (x_1^z, \ldots, x_n^z)^{\mathrm{T}}$ be the obtained optimal solution and q^z be the optimal value. Go to S2.

4 A Numerical Example

In order to demonstrate that Problem (26) is equivalent to Problem (18), in other words, an optimal solution to Problem (28) is an approximate solution to Problem (18), we consider the possibilistic linear programming problem (15) with $m = 3$. Fuzzy numbers C_j and A_{ij} are given by symmetric triangular fuzzy numbers whose center values and spreads are shown in Table 2, where a symmetric triangular fuzzy number Q=$\langle q^{\mathrm{C}}, q^{\mathrm{S}}\rangle$ ($q^{\mathrm{S}} > 0$) is characterized by a membership function,

$$
\mu_Q(r) = \begin{cases} 1 - \dfrac{|r - q^{\mathrm{C}}|}{q^{\mathrm{S}}} & \text{if } r \in [q^{\mathrm{C}} - q^{\mathrm{S}}, q^{\mathrm{C}} + q^{\mathrm{S}}],\\ 0 & \text{otherwise.}\end{cases} \tag{29}
$$

As shown in Table 2, fuzzy constraints B_i are defined by fuzzy sets $B_i = (b_i^{\mathrm{C}}, b_i^{\mathrm{S}}\rangle$ with a linear membership function,

$$
\mu_{B_i}(r) = \begin{cases} 1 & \text{if } r < b_i^{\mathrm{C}},\\ 1 - \dfrac{r - b_i^{\mathrm{C}}|}{b_i^{\mathrm{S}}} & \text{if } r \in [b_i^{\mathrm{C}}, b_i^{\mathrm{C}} + b_i^{\mathrm{S}}],\\ 0 & \text{otherwise.}\end{cases} \tag{30}
$$

We apply Problem (18) with reciprocal Goguen implication $I^0(a, b) = \min(1, (1-a)/(1-b))$ with definition $(1-a)/0 = +\infty$ for any $a \leq 1$, Lukasiewicz implication $I^1(a, b) = \min(1, 1-a+b)$, Reichenbach implication $I^2(a, b) = 1-a+ab$ and Goguen implication $I^3(a, b) = \min(1, b/a)$ with definition $b/0 = +\infty$ for any $b \geq 0$. Then, for $h > 0$ and $k > 0$, we obtain $f_0(k, h) = \max(0, 1 - (1 - k)/h)$, $f_1(k, h) = \max(0, k + h - 1)$, $f_2(k, h) = \max(0, (k + h - 1)/k)$ and $f_3(k, h) = kh$. Moreover, we set $h^0 = 0.5$, $h^1 = 0.5$, $h^2 = 0.62$ and $h^3 = 0.62$.

Let $\varepsilon = 0.0000001$ and $l = 20$, we solve Problem (28), i.e.,

$$
\begin{aligned}
&\text{maximize } q,\\
&\text{subject to}\\
&4.500x_1 + 6.650x_2 \geq q, && 4.550x_1 + 6.685x_2 \geq q,\\
&4.600x_1 + 6.720x_2 \geq q, && 4.650x_1 + 6.755x_2 \geq q,\\
&4.700x_1 + 6.790x_2 \geq q, && 4.750x_1 + 6.825x_2 \geq q,\\
&4.800x_1 + 6.860x_2 \geq q, && 4.850x_1 + 6.895x_2 \geq q,\\
&4.900x_1 + 6.930x_2 \geq q, && 4.950x_1 + 6.965x_2 \geq q,\\
&5x_1 + 7x_2 \geq q, && 2.350x_1 + 3.250x_2 \leq 290.00,\\
&2.315x_1 + 3.225x_2 \leq 287.00, && 2.280x_1 + 3.200x_2 \leq 284.00,\\
&2.245x_1 + 3.175x_2 \leq 281.00, && 2.210x_1 + 3.150x_2 \leq 278.00,\\
&2.175x_1 + 3.125x_2 \leq 275.00, && 2.140x_1 + 3.100x_2 \leq 272.00,\\
&2.105x_1 + 3.075x_2 \leq 269.00, && 2.070x_1 + 3.050x_2 \leq 266.00,\\
&2.035x_1 + 3.025x_2 \leq 263.00, && 2x_1 + 3x_2 \leq 260,\\
&4.900x_1 + 2.180x_2 \leq 465.00, && 4.825x_1 + 2.165x_2 \leq 454.44,\\
&4.750x_1 + 2.150x_2 \leq 446.00, && 4.675x_1 + 2.135x_2 \leq 439.09,\\
&4.600x_1 + 2.120x_2 \leq 433.33, && 4.525x_1 + 2.105x_2 \leq 428.46,\\
&4.450x_1 + 2.090x_2 \leq 424.29, && 4.375x_1 + 2.075x_2 \leq 420.67,\\
&4.300x_1 + 2.060x_2 \leq 417.50, && 4.225x_1 + 2.045x_2 \leq 414.71,\\
&4.150x_1 + 2.030x_2 \leq 412.22, && 4.075x_1 + 2.015x_2 \leq 410.00,\\
&4x_1 + 2x_2 \leq 408, && 1.500x_1 + 3.300x_2 \leq 230.00,\\
&1.475x_1 + 3.285x_2 \leq 227.21, && 1.450x_1 + 3.270x_2 \leq 224.42,\\
&1.425x_1 + 3.255x_2 \leq 221.63, && 1.400x_1 + 3.240x_2 \leq 218.84,\\
&1.375x_1 + 3.225x_2 \leq 216.05, && 1.350x_1 + 3.210x_2 \leq 213.26,\\
&1.325x_1 + 3.195x_2 \leq 210.47, && 1.300x_1 + 3.180x_2 \leq 207.68,\\
&1.275x_1 + 3.165x_2 \leq 204.89, && 1.250x_1 + 3.150x_2 \leq 202.10,\\
&1.225x_1 + 3.135x_2 \leq 199.31, && 1.200x_1 + 3.120x_2 \leq 196.52,\\
&1.175x_1 + 3.105x_2 \leq 193.73, && 1.150x_1 + 3.090x_2 \leq 190.94,\\
&1.125x_1 + 3.075x_2 \leq 188.15, && 1.100x_1 + 3.060x_2 \leq 185.36,\\
&1.075x_1 + 3.045x_2 \leq 182.57, && 1.050x_1 + 3.030x_2 \leq 179.78,\\
&1.025x_1 + 3.015x_2 \leq 176.99, && 1x_1 + 3x_2 \leq 174.2,\\
&x_1 \geq 0,\ x_2 \geq 0.
\end{aligned}
\tag{31}
$$

Solving this problem, we obtain an optimal solution $\mathbf{x}^* = (x_1^*, x_2^*) = (79.508, 31.564)$ with the optimal value $q^* = 567.69$. As shown in Figures 2∼5, we can observe that the obtained solution approximately satisfies constraints, $N^0_{\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}}([q^*, +\infty)) \geq 0,5$ and $N^i_{\boldsymbol{A}_i^{\mathrm{T}}\mathbf{x}}(B_i) \geq h^i$, $i = 1, 2, 3$.

5 Concluding Remarks

Utilizing the closure of generation processes of conjunction and implication functions, we have shown that necessity fractile optimization models of possibilistic linear programming problems are reduced to semi-infinite linear programming problems. Therefore, we can solve the problems approximately by linear programming techniques. As a solution procedure for the reduced semi-infinite linear programming problem, we have described a relaxation procedure. In order to demonstrate the main result, we have given a simple numerical example. In the example, the reduced semi-infinite linear programming problem is approximated by a linear programming problem having coarsely-sampled constraints. It has been confirmed that the obtained solution approximately satisfies the required constraints on necessity measures by figures.

The investigation on possibilistic linear programming problems with necessity measures is important to have flexible

ISBN: 978-989-95079-6-8

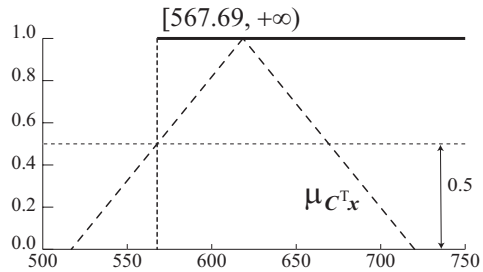


Figure 2: The relation between $\mathbf{C}^{\mathrm{T}}\mathbf{x}^*$ and the optimal value

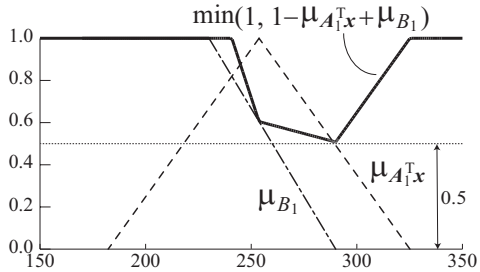


Figure 3: The satisfaction of $N^1_{\boldsymbol{A}_1^{\mathrm{T}}\mathbf{x}}(B_1) \geq 0.5$

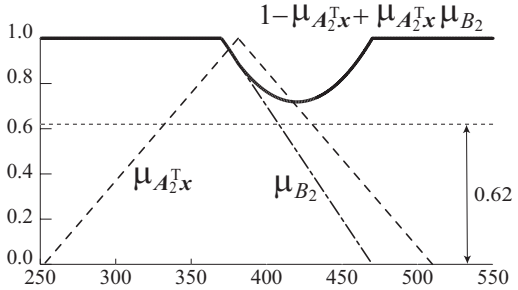


Figure 4: The satisfaction of $N^2_{\boldsymbol{A}_2^{\mathrm{T}}\mathbf{x}}(B_2) \geq 0.62$

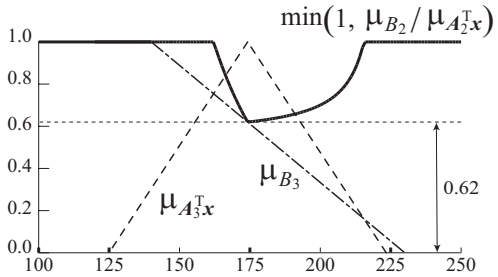


Figure 5: The satisfaction of $N^3_{\boldsymbol{A}_3^{\mathrm{T}}\mathbf{x}}(B_3) \geq 0.62$

models of robust optimization reflecting decision makers' various attitudes toward the uncertainty. Fortunately, models with necessity measures are often easier than models with possibility measures. Analyzing some special implication functions defining necessity measures and special shapes of fuzzy coefficients, the reduced semi-infinite constraints can further reduced to finite constraints. Moreover, we may solve necessity measure optimization models of possibilistic linear programming problems by the simplex method together with the bisection method. These would be future topics in our studies.

Acknowledgment

The author acknowledges that this work has been partially supported by the Grant-in-Aid for Scientific Research (B) No. 17310098, and thanks Yuji Matsumoto (Osaka University) for his help in solving a linear programming problem.

References

[1] M. Inuiguchi, H. Ichihashi and H.Tanaka. Fuzzy programming: A survey of recent developments. in: R. Słowiński and J. Teghem (eds.). *Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty*. Kluwer Academic Publishers, Dordrecht, 45–68, 1990.

[2] H. Rommelfanger and R. Słowiński. Fuzzy linear programming with single or multiple objective functions. in: R. Słowiński (ed.). *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*. Kluwer Academic Publishers, Boston, 179–213, 1998.

[3] M. Inuiguchi and J. Ramík. Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem. *Fuzzy Sets and Systems*, 111: 29–45, 2000.

[4] H. Rommelfanger and T. Keresztfalvi. Multicriteria fuzzy optimization based on Yager's parameterized t-norm. *Foundations of Computing and Decision Sciences*, 16: 99–110, 1991.

[5] M. Inuiguchi. Linear programming with interactive fuzzy numbers. *Proceedings of 40th Annual Convention of ORSI*, 27–37, 2007.

[6] E. Sanchez. Resolution of composite fuzzy relation equations. *Information and Control*, 30: 38–48, 1976.

[7] M. Inuiguchi and M. Sakawa. On the closure of generation processes of implication functions from a conjunction function. in: T. Yamakawa and G. Matsumoto (eds.). *Methodologies for the Conception, Design, and Application of Intelligent Systems*. World Scientific, Singapore, 1: 327–330, 1996.

[8] D. Dubois and H. Prade. A theorem on implication functions defined from triangular norms. *Stochastica*, 8: 267–279, 1984.

[9] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1: 3–28, 1978.

[10] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, 1980.

[11] M. Inuiguchi and T. Tanino. Necessity measures and parametric inclusion relations of fuzzy sets. *Fundamenta Informaticae*. 42: 279–302, 2000.

[12] M. Inuiguchi, S. Greco, R. Słowiński and T. Tanino. Possibility and necessity measure specification using modifiers for decision making under fuzziness. *Fuzzy Sets and Systems*. 137: 151–175, 2003.

[13] D. Dubois and H. Prade. Fuzzy numbers: An overview. in: J.C. Bezdek (ed.). *Analysis of Fuzzy Information, Vol.I: Mathematics and Logic*. CRC Press, Boca Raton, FL, pp.3–39, 1987.

[14] M.A. Goberna and M.A. López. *Linear Semi-infinite Optimization*. John Wiley & Sons, Chichester, 1998.

ISBN: 978-989-95079-6-8

On Aristotle's NC and EM Principles in Three-valued Logics*

E. Trillas, C. Moraga, E. Renedo†

European Centre for Soft Computing
Mieres (Asturias), Spain
Email: {enric.trillas, claudio.moraga, eloy.renedo }@softcomputing.es.

Abstract— *By interpreting* 'p *is* impossible' *by* 'p *is* self-contradictory'*, and* 'p *is* always' *by* 'not p *is* self-contradictory'*, this paper studies which of the three-valued systems of Łucasiewicz, Gödel, Kleene, Bochvar, and Post, do verify the Aristotle's principles of Non-Contradiction (NC), and Excluded-Middle (EM).*

Keywords— Systems of three-valued logic, Non-Contradiction, Excluded-Middle.

1 Introduction

As it is well known, the systems of multiple-valued logic do not verify the *principles of non-contradiction* and *excluded-middle*, once presented in the forms (see [5])

$$a \cdot a' = 0,\ a + a' = 1.$$

Nevertheless, in [7] a different way of looking at these two principles, and based on the concept of self-contradiction, respectively $a \cdot a' \leq (a \cdot a')'$, and $(a + a')' \leq ((a + a')')'$, for all a, b, was introduced. These expressions, that are even in more agreement with the original formulation given by Aristotle (see [1]), allow the verification of the two principles for a wide class of structures that include all algebras of fuzzy sets, as well as De Morgan algebras (see [5]). All that meant some progress for what concerns the verification of the two principles.

This paper tries to study the verification of the two Aristotle's principles once interpreted as in [7] , by some of the most well known systems of three-valued logic, namely, by those of the Łukasiewicz, Gödel, Kleene, Bochvar and Post. It is first shown that in Łukasiewicz, Gödel and Kleene cases the two principles are verified, that in the case of Bochvar only one of them is verified, and that in the case of Post no one is verified. Everything is done in agreement with the implication's semantics, as given at each case by its truth-table. At the end, although far from the semantical interpretation at each case, it is shown how the five systems could verify the two principles in the sense introduced in [7].

2 Systems of MV logic

2.1

Let Ω be a set consisting of propositions $p, q, r, \ldots$ and such that,

- If $p \in \Omega$, then '*not p*'$\in \Omega$
- If $p, q \in \Omega$, then '*p and q*', '*p or q*', and '*If p, then q*', are in Ω

2.2

Let L be a set with at least three elements, and endowed with a unary operation $' : L \to L$, and three binary operations $\cdot, +, \to: L \times L \to L$. Suppose there exists 0, 1 in L such that $0 \neq 1$, and $0' = 1$.

2.3

(Ω, L, t) is a *system of multiple-valued logic*, if $t : \Omega \to L$ (truth-function) verifies:

- $t(\text{not } p) = t(p)'$
- $t(p \text{ and } q) = t(p) \cdot t(q)$
- $t(p \text{ or } q) = t(p) + t(q)$
- $t(\text{if } p, \text{ then } q) = t(p) \to t(q)$

for all p, q in Ω. If L is with n elements, the system is an *n-valued* one.

2.4

Two systems of multiple-valued logic, (Ω_1, L_1, t_1) and (Ω_2, L_2, t_2), are *isomorphic* if there exist a pair of bijections (Φ, φ) such that

- $\Phi : \Omega_1 \to \Omega_2, \quad \varphi : L_1 \to L_2$
- $\varphi(a') = \varphi(a)', \varphi(a \cdot b) = \varphi(a) \cdot \varphi(b), \varphi(a + b) = \varphi(a) + \varphi(b), \varphi(a \to b) = \varphi(a) \to \varphi(b)$ with the corresponding operations in L_1, L_2
- $t_1 = \varphi^{-1} \circ t_2 \circ \Phi$, or $t_2(\Phi(p)) = \varphi(t_1(p))$ for all $p \in \Omega_1$.

Example.
The three-valued system of Łukasiewicz, Ł$_3$, (see [5]) is obtained with $L = \{T, F, I\}$ and the four operations given by the following tables:

	$'$
T	F
I	I
F	T

$\bullet$	T	I	F
T	T	I	F
I	I	I	F
F	F	F	F

$+$	T	I	F
T	T	T	T
I	T	I	I
F	T	I	F

$\to$	T	I	F
T	T	I	F
I	T	T	I
F	T	T	T

*This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias, Spain), and CICYT (Spain) under project TIN2008-06890-C02-01

†Corresponding author.

ISBN: 978-989-95079-6-8

where $T, I, F \in L, T \neq F$, and $F' = T$.

With $L_0 = \{1, \frac{1}{2}, 0\}$, and $\varphi : L \to L_0$ given by $\varphi(T) = 1, \varphi(I) = \frac{1}{2}, \varphi(F) = 0$, the tables are translated into,

$'$	
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

$\bullet$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
0	0	0	0

$+$	1	$\frac{1}{2}$	0
1	1	1	1
1	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\to$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	1	$\frac{1}{2}$
0	1	1	1

Then, *any system* $Ł_3(\Omega, L, t)$ *is isomorphic to* (Ω, L_0, t^*), with $t^*(p) = \varphi(t(p))$, for all $p \in \Omega$.

Notice that it is,
$a' = 1 - a, a \cdot b = \min(a, b), a + b = \max(a, b)$,
$a \to b = \min(1, 1 - a + b)$, for all $a, b \in L_0$.
Hence,
$t(\text{not } p) = 1 - t(p), t(p \text{ and } q) = \min(t(p), t(q))$,
$t(p \text{ or } q) = \max(t(p), t(q))$,
$t(\text{If } p, \text{ then } q) = \min(1, 1 - t(p) + t(q))$,
for all $p, q \in \Omega$.
Notice that $(a')' = a'' = a$.

3 The other L_0 three-valued systems

Analogously, we can obtain the corresponding systems of Gödel, Bochvar, Kleene and Post (see [5], [2]), by their isomorphic images with $\varphi : L \to L_0$ given by $\varphi(T) = 1, \varphi(I) = \frac{1}{2}, \varphi(F) = 0$.

3.1

Gödel (G_3), is translated into L_0 by

$'$	
1	0
$\frac{1}{2}$	0
0	1

$\bullet$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
0	0	0	0

$+$	1	$\frac{1}{2}$	0
1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\to$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	1	0
0	1	1	1

or, for all $a \in L_0$,

$$a' = \begin{cases} 1, & \text{if } a = 0 \\ 0, & \text{if } a \neq 0 \end{cases}$$

$a \cdot b = \min(a, b)$
$a + b = \max(a, b)$

$$a \to b = \begin{cases} 1, & \text{if } a \leq b \\ b, & \text{if } a > b \end{cases}$$

Notice that, the negation $'$ is not strong, since $(\frac{1}{2}')' = 0' = 1$.

3.2

Kleene (K_3), is translated into L_0 by

$'$	
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

$\bullet$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
0	0	0	0

$+$	1	$\frac{1}{2}$	0
1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\to$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	1	1

or, for all $a \in L_0$,
$a' = 1 - a$,
$a \cdot b = \min(a, b)$,
$a + b = \max(a, b)$,
$a \to b = \max(1 - a, b)$.

Notice that $(a')' = a'' = a$, for all $a \in L_0$ (the negation is strong).

3.3

Bochvar (B_3) is translated into L_0 by

$'$	
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

$\bullet$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	0	$\frac{1}{2}$	0

$+$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	1
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\to$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	1

or,
$a' = 1 - a$,

$$a \cdot b = \begin{cases} \frac{1}{2}, & \text{if } (a, b) \in \{(\frac{1}{2}, 0), (0, \frac{1}{2})\} \\ \min(a, b), & \text{otherwise} \end{cases},$$

$$a + b = \begin{cases} \frac{1}{2}, & \text{if } (a, b) \in \{(1, \frac{1}{2}), (\frac{1}{2}, 1)\} \\ \max(a, b), & \text{otherwise} \end{cases},$$

$a \to b = a' + b$.

Notice that $(a')' = a'' = a$, for all $a \in L_0$ (the negation is strong).

3.4

Post (P_3), is translated into L_0 by

$'$	
1	$\frac{1}{2}$
$\frac{1}{2}$	0
0	1

$\bullet$	1	$\frac{1}{2}$	0
1	0	0	$\frac{1}{2}$
$\frac{1}{2}$	0	1	$\frac{1}{2}$
0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

$+$	1	$\frac{1}{2}$	0
1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\to$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{2}$	1	$\frac{1}{2}$	0
0	1	1	1

or,

$$a' = \begin{cases} 1, & \text{if } a = 0 \\ |\frac{1}{2} - a|, & \text{if } a \neq 0 \end{cases},$$

$a \cdot b = (a' + b')'$,
$a + b = \max(a, b)$,
$a \to b = a' + b$.

Notice that, for all $a \in L_0$, is $(a')' = a'' \neq a$ (the negation is not strong).

Remark. 3.1. In the systems $Ł_3, K_3, G_3, B_3$ but not in P_3, holds the following:

- $0 \cdot 1 = 1 \cdot 0 = 0 \cdot 0 = 0, 1 \cdot 1 = 1$
- $0 + 1 = 1 + 0 = 1 + 1 = 1, 0 + 0 = 0$
- $0' = 1, 1' = 0$

ISBN: 978-989-95079-6-8

- $1 \to 1 = 0 \to 1 = 0 \to 0 = 1, 1 \to 0 = 0$

Then, these four three-valued systems do contain the classical system of two-valued logic based on the values $\{0, 1\}$ that, notwithstanding, is not contained in P_3, a system in which only the operation + preserves such classical case.

4 The two principles

4.1

Triplets $(L, \vDash, ')$ consisting of a non-empty set L, a transitive relation $\vDash \subset L \times L$, and a mapping $' : L \to L$ reversing $\vDash$: if $a \vDash b$, then $b' \vDash a'$, are called *transitive reversing systems* (TRS for short) (see [7])

4.2

An operation $\cdot : L \times L \to L$, is a *type-1 operation* in $(L, \vDash, ')$, provided $a \cdot a' \vDash a$ and $a \cdot a' \vDash a'$, for all $a \in L$.

An operation $+ : L \times L \to L$, is a *type-2 operation* in $(L, \vDash, ')$, provided $a \vDash a + a'$ and $a' \vDash a + a'$, for all $a \in L$.

Theorem 4.1. *(ANC, for Aristotle's Non-contradiction). If $(L, \vDash, ')$ is a TRS, and $\cdot$ is a type-1 operation, it holds*

$$a \cdot a' \vDash (a \cdot a')' \tag{1}$$

for all $a \in L$ (see [7])

Theorem 4.2. *(AEM, for Aristotle's Excluded-middle). If $(L, \vDash, ')$ is a TRS, and $+$ is a type-2 operation, it holds*

$$(a + a')' \vDash ((a + a')')' \tag{2}$$

for all $a \in L$ (see [7])

Remark. 4.3. By defining 'a is selfcontradictory' as $a \vDash a'$, and by taking $'$ representing 'not', $\cdot$ representing 'and', it is clear that (1) can be read

'a and not a, is self-contradictory'.

Hence, interpreting 'impossible' by 'self-contradictory' (1) represents 'a and not a, is impossible'. That is, (1) can be read as an algebraic translation of the Aristotle's Non-contradiction principle.

Remark. 4.4. Like in the previous remark, and taking + representing 'or', (2) can be read as 'not (a or not a) is self - contradictory', or 'not (a or not a) is impossible', an algebraic translation of the Aristotle's Excluded-middle principle usually stated as 'a or not a is always'.

4.3

Theorem 4.5. *Given a triplet $(L, \cdot, ')$, with $L \neq \emptyset$, $\cdot : L \times L \to L$, and $' : L \to L$, the relation $\vDash_{NC} \subset L \times L$, given by the set of pairs*

$$\vDash_{NC} = \{(a \cdot a', (a \cdot a')'); a \in L\},$$

assures the verification of the principle ANC.

Proof. Obviously, $a \cdot a' \vDash_{NC} (a \cdot a')'$, for all $a \in L$. □

Theorem 4.6. *Given a triplet $(L, +, ')$, with $L \neq \emptyset$, $+ : L \times L \to L$, and $' : L \to L$, the relation $\vDash_{EM} \subset L \times L$, given by the set of pairs*

$$\vDash_{EM} = \{((a + a')', ((a + a')')'); a \in L\},$$

assures the verification of the principle AEM

Proof. Obviously, $(a + a')' \vDash_{NC} (a + a')')'$, for all $a \in L$. □

Theorem 4.7. *Given a quadruplet $(L, \cdot, +, ')$, with $L \neq \emptyset$,$\cdot : L \times L \to L$, $+ : L \times L \to L$, and $' : L \to L$, the relation*

$$\vDash = \vDash_{NC} \cup \vDash_{EM} \subset L \times L$$

assures the verification of the two principles ANC and AEM.

Proof. Obviously, for all $a \in L$, is

- $a \cdot a' \vDash_{NC} (a \cdot a')'$ and $a \cdot a' \vDash (a \cdot a')'$
- $(a + a')' \vDash_{EM} ((a + a')')'$, and $(a + a')' \vDash ((a + a')')'$. □

It is to be noticed that *when $'$ verifies $(a')' = a'' = a$, ($'$ is strong), the triplets $(L, \vDash_{NC}, ')$* and $(L, \vDash_{EM}, ')$ are TRS, since:

- $a \cdot a' \vDash_{NC} (a \cdot a')'$, and $(a \cdot a')'' = a \cdot a'$
- $(a+a')' \vDash_{EM} ((a+a')')'$ is equivalent to $(a+a')' \vDash_{EM} a + a'$

Of course, *this does not mean that those triplets are TRS only when $'$ is strong*.

Remark. 4.8. It is obvious that the set of relations $\vDash$ for which $(L, \cdot, +, ')$ verifies, respectively, the principle ANC or the principle AEM, is not empty. If $\vDash$ allows to verify ANC, it is $\vDash_{NC} \subset \vDash$, if $\vDash$ allows to verify AEM is $\vDash_{EM} \subset \vDash$, and if $\vDash$ allows both principles is $\vDash_{NC} \cup \vDash_{EM} \subset \vDash$

Remark. 4.9. It should be pointed out that the last theorems ignore what follows. *The unary operation $'$ should represent 'not', and the relation $a \vDash b$ should represent* '*If a, then b*'. Otherwise, the interpretation of 'If a, then not a' by means of $a \vDash a'$ (a is self-contradictory) has no sense at all. Hence, *and provided $a \cdot b$ does represent* 'a and b', *those representations are crucial for asserting that $a \cdot a' \vDash (a \cdot a')'$ is an interpretation of* "'a and not a' is a self-contradictory statement". And analogously with $(a + a')' \vDash ((a + a')')'$.

4.4

Let us call *Modern Non-contradiction Principle* (MNC, for short), the statement '$a \cdot a' = 0$ for all $a \in L$', and *Modern Excluded-middle Principle* (MEM, for short), '$a + a' = 1$ for all $a \in L$'

Theorem 4.10. *Given a triplet $(L, \vDash, ')$, an operation $\cdot : L \times L \to L$, and $0 \in L$, such that*

- $0 \vDash 0'$
- $a \cdot a' = 0$ *(MNC)*

then, it holds ANC.

ISBN: 978-989-95079-6-8

Proof. $a \cdot a' = 0 \vDash 0' = (a \cdot a')'$. $\square$

Theorem 4.11. *Given a triplet* $(L, \vDash, ')$*, an operation* $+ : L \times L \to L$*, and* $1 \in L$*, such that*

- $1' \vDash (1')'$
- $a + a' = 1$ *(MEM)*

then, it holds AEM.

Proof. $(a + a')' = 1' \vDash (1')' = ((a + a')')'$. $\square$

Remark. 4.12. If $1' = 0$, then $1' \vDash (1')'$ is equivalent to $0 \vDash 0'$

Remark. 4.13. In the sense of theorems 4.10 and 4.11, MNC and MEM are particular cases of ANC and AEM, respectively.

Remark. 4.14. Apart of the three operations $(\cdot, +, ')$, the principles ANC and AEM only involve the relation $\vDash$, but the principles MNC and MEM do involve the relation =, and the 'singular' elements 0,1.

4.5

In what follows this paper is devoted to study if the three-valued systems $Ł_3, G_3, K_3, B_3$, and P_3, do verify the principles ANC or AEM once expressed in the above algebraic terms.

5 Posing the problem for the five three-valued systems

Given a MVL system (Ω, L_0, t), consider its 'values' part $(L_0, ', \cdot, +, \to)$, and the binary relations $\leq$ and $\vDash$ given, respectively, by

$a \vDash b \Leftrightarrow a \to b = 1$, and $\leq$ is the linear order in $L_0 = \{1, \frac{1}{2}, 0\}$, inherited from $\mathbb{R}$.

- *Case* $Ł_3$.

  It is $a \to b = 1 \Leftrightarrow \min(1, 1 - a + b) = 1 \Leftrightarrow a \leq b$. That is, $\vDash_{Ł} = \leq$.

  Since $a' = 1 - a$ implies "$a \leq b \Leftrightarrow b' \leq a'$", $(L_0, \leq, ')$ is a TRS in which, obviously, the operation $\cdot = \min$ is one of type-1, and $+ = \max$ is one of type-2. Hence, it holds

  $$a \cdot a' \leq (a \cdot a')', \ (a + a')' \leq ((a + a')')',$$

  for all $a \in L_0$.

- *Case* G_3.

  It is $a \to b = 1 \Leftrightarrow a \leq b$. That is, $\vDash_G = \leq$, like in the case $Ł_3$. Since it is also $\cdot = \min$, $+ = \max$, it also holds

  $$a \cdot a' \leq (a \cdot a')', \ (a + a')' \leq ((a + a')')',$$

  for all $a \in L_0$.

- *Case* K_3.

  Since, $\vDash_K = \{(1,1), (\frac{1}{2}, 1), (0,1), (0, \frac{1}{2}), (0,0)\} \subset \leq$, and $\vDash'_K = \{(0,0), (0, \frac{1}{2}), (0,1), (\frac{1}{2}, 1), (1,1)\} = \vDash_K$, it means that $(L_0, \vDash_K, ')$ is a TRS. Since, $a' = 1 - a$, it is also $(L_0, \leq, ')$ a TRS.

  From $\frac{1}{2} \cdot \frac{1}{2}' = \frac{1}{2}$, $(\frac{1}{2} \cdot \frac{1}{2}')' = \frac{1}{2}$, and $(\frac{1}{2}, \frac{1}{2}) \notin \vDash_K$, follows that it is not $a \cdot a' \vDash (a \cdot a')'$, for all $a \in L_0$.

  Analogously, $(\frac{1}{2} + \frac{1}{2}')' = \frac{1}{2}$, $(\frac{1}{2} + \frac{1}{2})'' = \frac{1}{2}$, shows that it is not $(a + a')' \vDash ((a \cdot a')')$, for all $a \in L_0$.

  Nevertheless, since $\cdot$ denoting $\min$ is a type-1 operation in $(L_0, \leq, ')$, and $+$ denoting $\max$ is a type-2 operation in $(L_0, \leq, ')$, it follows

  $$a \cdot a' \leq (a \cdot a')', \ (a + a')' \leq ((a + a')')',$$

  for all $a \in L_0$.

- *Case* B_3.

  It is $\vDash_B = \{(1,1), (0,1), (0,0)\} \subset \leq$, and $\vDash'_B = \{(b', a'); (a, b) \in \vDash_B\} = \{(0,0), (0,1), (1,1)\} = \vDash_B$. That is, $(L_0, \vDash_B, ')$ and $(L_0, \leq, ')$ are TRSs.

  Since, $\frac{1}{2} \cdot \frac{1}{2}' = \frac{1}{2}$ and $(\frac{1}{2} \cdot \frac{1}{2}')' = \frac{1}{2}$, but $(\frac{1}{2}, \frac{1}{2}) \notin \vDash_B$, it is not $a \cdot a' \vDash_B (a \cdot a')'$, for all $a \in L_0$.

  Since, $(\frac{1}{2} + \frac{1}{2}')' = \frac{1}{2}' = \frac{1}{2}$ and $((\frac{1}{2} + \frac{1}{2}')')' = \frac{1}{2}$, it is not $(a + a')' \vDash_B ((a + a')')'$, for all $a \in L_0$.

  Notwithstanding,

  $$\left.\begin{array}{lcl} 1 \cdot 1' = 0, (1 \cdot 1')' = 0' = 1 & \Rightarrow & 0 \leq 1 \\ \frac{1}{2} \cdot \frac{1}{2}' = \frac{1}{2}, (\frac{1}{2} \cdot \frac{1}{2}')' = \frac{1}{2} & \Rightarrow & \frac{1}{2} \leq \frac{1}{2} \\ 0 \cdot 0' = 0, (0 \cdot 0')' = 0' = 1 & \Rightarrow & 0 \leq 1 \end{array}\right\} \Rightarrow$$

  $a \cdot a' \leq (a \cdot a')'$, for all $a \in L_0$.

  Analogously,

  $$\left.\begin{array}{lcl} (1 + 1')' = 1, ((1 + 1')')' = 1' + 1 = 1 & \Rightarrow & 1 \leq 1 \\ (\frac{1}{2} + \frac{1}{2}')' = \frac{1}{2}, ((\frac{1}{2} + \frac{1}{2}')')' = \frac{1}{2} & \Rightarrow & \frac{1}{2} \leq \frac{1}{2} \\ (0 + 0')' = 0, ((0 + 0')')' = 0 + 0' = 1 & \Rightarrow & 0 \leq 1 \end{array}\right\} \Rightarrow$$

  $(a + a')' \leq ((a + a')')'$, for all $a \in L_0$.

- *Case* P_3.

  It is $\vDash_P = \{(1,1), (\frac{1}{2}, 1), (0,1), (0, \frac{1}{2}), (0,0)\} \subset \leq$, but since $(\frac{1}{2}, \frac{1}{2}) \in \vDash'_P$ it is not $\vDash'_P \subset \vDash_P$, and $(L_0, \vDash_P, ')$ is not a TRS. In addition, also $(L_0, \leq_P, ')$ it is not a TRS, since $\frac{1}{2} \leq 1$, but $1' = \frac{1}{2} \not\leq 0 = \frac{1}{2}'$.

  From, $0 \cdot 0' = 0 \cdot 1 = \frac{1}{2}, (0 \cdot 0')' = \frac{1}{2}' = 0$, and neither $(\frac{1}{2}, 0) \vDash 0$, nor $\frac{1}{2} \leq 0$, it follows that it is not $a \cdot a' \vDash_P (a \cdot a')'$, nor $a \cdot a' \leq (a \cdot a')'$, for all $a \in L_0$.

  From, $(0 + 0')' = \frac{1}{2}, ((0 + 0')')' = 0$, it follows that it is neither $(a + a')' \vDash_P ((a + a')')'$, nor $(a + a')' \leq ((a + a')')'$, for all $a \in L_0$.

6 The solutions to the problem

6.1

Given a multiple-valued system (Ω, L_0, t), consider the two relations:

$$p \vDash q \Leftrightarrow t(p) \vDash t(q), \text{ and } p \leq q \Leftrightarrow t(p) \leq t(q),$$

and define:

ISBN: 978-989-95079-6-8

- The MV-system (Ω, L_0, t) verifies the principle ANC respect to $\vDash, \leq$, respectively, whenever

  p *and not* $p \vDash$ *not* (p *and not* p),
  or, p *and not* $p \leq$ *not* (p *and not* p)

- The MV-system (Ω, L_0, t) verifies the principle AEM respect to $\vDash, \leq$, respectively, whenever

  not (p *or not* p) $\vDash$ *not*(*not* (p *or not* p)),
  or, *not*(p *or not* p) $\leq$ *not*(*not* (p *or not* p)),

  for all $p \in \Omega$.

6.2

Hence, from the results in section 5, follows:

1. Three-valued *Łukasiewicz* systems *do verify* ANC *and* AEM, *with respect to* $\vDash_{\mathbf{Ł}} = \leq$
2. Three-valued *Gödel* systems *do verify* ANC *and* AEM, *with respect to* $\vDash_G = \leq$
3. Three-valued *Kleene* systems *verify neither* ANC *,nor* AEM, *with respect to* $\vDash_K \subset \leq$
4. Three-valued *Kleene* systems *do verify* ANC *and* AEM, *with respect to* $\leq$
5. Three-valued *Bochvar* systems *verify neither* ANC, *nor* AEM, *with respect to* $\vDash_B \subset \vDash_K \subset \leq$
6. Three-valued *Bochvar* systems *do verify* ANC *and* AEM, *with respect to* $\leq$
7. Three-valued *Post* systems *verify neither* ANC, *nor* AEM, *with respect to* $\vDash_P = \vDash_K \subset \leq$
8. Three-valued *Post* systems *verify neither* ANC, *nor* AEM, *with respect to* $\leq$.

Anyway, and in the line of theorems in section 4.2, it is possible to go further for what concerns points 3, 5, 7 and 8.

6.3

Obviously, *the most surprising case is* P_3. By just adding $(\frac{1}{2}, 0)$ to $\vDash_P$, that is, with the relation

$$\vDash^*_P = \vDash \cup \{(\tfrac{1}{2}, 0)\},$$

it is easy to prove that *both* ANC *and* AEM *hold with respect to* $\vDash^*_P$. But the triplet $(L_=, \vDash^*_P, ')$ is not a TRS, since $(1,1) \in \vDash^*_P$ but $(1', 1') = (\frac{1}{2}, \frac{1}{2}) \notin \vDash^*_P$.

Taking $\vDash^{**}_P = \vDash^*_P \cup \{(\frac{1}{2}, \frac{1}{2})\} = \{(\frac{1}{2}, 0), (0, \frac{1}{2}), (0,0), (1,1), (\frac{1}{2}, 1), (0,1), (\frac{1}{2}, \frac{1}{2})\}$, it is obtained the TRS $(L_0, \vDash^{**}_P, ')$, and *it also holds* ANC *and* AEN *with respect to* $\vDash^{**}_P$. In addition, and as it is easy to check, *the operation* $\cdot$ *in* P_3 *is a type-1 operation in* $(L_0, \vDash^{**}_P, ')$, and *the operation* + in P_3 *is a type-2 operation in* $(L_0, \vDash^{**}_P, ')$.

It should be pointed out that neither $\cdot$ is a type-1 operation in $(L_0, \vDash^*_P, ')$, nor + is a type-2 operation in $(L_0, \vDash^*_P, ')$. The reasons are, respectively, that $\frac{1}{2} \cdot \frac{1}{2}' = \frac{1}{2}$, and $\frac{1}{2} + \frac{1}{2}' = \frac{1}{2}$, but $(\frac{1}{2} \cdot \frac{1}{2}) \notin \vDash^*_P$.

6.4

In the case K_3, taking $\vDash^*_K = \vDash_K \cup \{(\frac{1}{2}, \frac{1}{2})\} = \leq$, the triplet $(L_0, \vDash^*_K, ')$ is a TRS in which the operation $\cdot$ denoting min is a type-1 operation, and the operation + is a type-2 operation. Hence, *it holds ANC with respect to* $\vDash^*_K$ *, and also AEM with respect to the same* $\vDash^*_K$.

6.5

Concerning the case B_3, with $\vDash^*_B = \vDash_B \cup \{(\frac{1}{2}, \frac{1}{2})\}$, the TRS $(L_0, \vDash^*_B, ')$ is reached in which *it holds neither ANC, nor AEM, since*

$$(1+1')' = 1, ((1+1')')' = 1' = 0, \text{ but } (1,0) \notin \vDash^*_B .$$

Anyway, with $\vDash^{**}_B = \vDash^*_B \cup \{(1,0)\} = \{(1,1), (0,1), (0,0), (\frac{1}{2}, \frac{1}{2}), (1,0)\}$, it is obtained the TRS $(L_0, \vDash^{**}_B, ')$, *in which the two principles ANC and AEM obviously do hold.*

6.6

At the end, in the five cases there exist a TRS $(L_0, \vDash, ')$ for which the two principles ANC and AEM do hold. In the cases Ł$_3$ and G_3, it is $\vDash = \leq$ with $\leq$ the lineal order of the real line. In the case K_3, there is a solution with $\leq$, but also with $\vDash^*_K$ that, nevertheless, is again a part of the order $\leq$.

What is different are the cases of Bochvar and Post, where $\vDash^{**}_B$ is not a part of $\leq$, and $\vDash^{**}_P$ is not a part of $\leq$, but on the contrary, it is $\leq \subset \vDash^{**}_P$.

This fact, of not being a part of $\leq$, seems to be the red line between the three system Ł$_3, G_3, K_3$ and the two systems B_3, P_3. The trick to reach the principles ANC and AEM seems to be based on the identification of some pairs $(a, b) \in L_0 \times L_0$ that, added to $\vDash$ in the line of section 4.2, give a new relation for which (without altering the original negation) the two principles do hold.

6.7

Form Remark 4.8, it follows that the relation considered from 6.2 to 6.5 are larger than $\vDash_{NC}$, and $\vDash_{EM}$. As it is easy to check,

- In Ł$_3$, $\vDash_{NC} = \vDash_{EM} = \{(0,1), (\frac{1}{2}, \frac{1}{2})\} \subset \vDash_{\mathbf{Ł}} = \leq$
- In G_3, $\vDash_{NC} = \vDash_{EM} = \{(0,1)\} \subset \vDash_G = \leq$
- In K_3, $\vDash_{NC} = \vDash_{EM} = \{(0,1), (\frac{1}{2}, \frac{1}{2})\} \subset \leq$, but $\vDash_{NC} \nsubseteq \vDash_K$, and $\vDash_{NC} \nsubseteq \vDash^*_K$
- In B_3, $\vDash_{NC} = \vDash_{EM} = \{(0,1), (\frac{1}{2}, \frac{1}{2})\} \subset \leq$, but $\vDash_{NC} \nsubseteq \vDash_B$, and $\vDash_{NC} \nsubseteq \vDash^*_B$
- In P_3, $\vDash_{NC} = \vDash_{EM} = \{(0,1), (\frac{1}{2}, 0)\} \subset \vDash^*_P$, but $\vDash_{NC} \nsubseteq \leq$, and $\vDash_{NC} \nsubseteq \vDash_P$.

Remark. 6.1. An element $a \in L$ is $\vDash$-self-contradictory, if $a \vDash a'$.

In Ł$_3$, since $\vDash_{\mathbf{Ł}_3} = \leq$ and $a' = 1 - a$, it is $a \leq a'$ whenever $a \leq \frac{1}{2}$, that is, the self-contradictory elements are 0 and $\frac{1}{2}$. In G_3, with $\vDash_G = \leq$, the only element that is self-contradictory is 0.

In K_3, with $\vDash_K \subset \leq$, the only self-contradictory is 0, but with $\vDash^*_K$ also $\frac{1}{2}$ is self-contradictory.

ISBN: 978-989-95079-6-8

In B_3, with $\vDash_B \subset \leq$, the only element that is self-contradictory is 0, but with $\vDash_B^*$ and $\vDash_B^{**}$ also $\frac{1}{2}$ is self-contradictory.

In P_3, with $\vDash_P$, the only element that is self-contradictory is 0, but with $\vDash_P^*$ and $\vDash_P^{**}$ also $\frac{1}{2}$ is self-contradictory.

7 Conclusion

7.1

When the principles NC and EM are interpreted in the usual 'modern' forms,

$$a \cdot a' = 0,\ a + a' = 1$$

for all $a \in L_0$, neither Ł$_3$ nor G_3, K_3, B_3, P_3 do verify one of them. For example

- in Ł$_3, G_3, K_3, B_3$: $\frac{1}{2} \cdot \frac{1}{2}' = \frac{1}{2} \neq 0, \frac{1}{2} + \frac{1}{2}' = \frac{1}{2} \neq 1$
- in $P_3 : 0 \cdot 0' = \frac{1}{2} \neq 0, \frac{1}{2} + \frac{1}{2}' = \frac{1}{2} \neq 1$.

Hence, the results presented in this paper mean some progress in what concerns, jointly with what is in [7] and [6], the general verification of the non-contradiction and excluded-middle principles. If they can fail at the just mentioned 'modern' level, it seems to exist at least another level (that of self-contradiction) at which they do not fail.

7.2

Nevertheless, right now the authors are not yet able to completely answer the question of *why the systems Ł$_3, G_3, K_3$ do have solution in $\leq$, but the systems B_3, P_3 do have to exit from $\leq$ to reach a solution, and the system K_3 also admits a second solution in $\leq$*. In all these cases, the pairs that are to be added to the initial relation given by $a \to b = 1$, verify $a \to b \neq 1$. For example,

- In K_3, the added pair $(\frac{1}{2}, \frac{1}{2})$, verifies $\frac{1}{2} \to \frac{1}{2} = \frac{1}{2} \neq 1$
- In B_3, the added pairs $(\frac{1}{2}, \frac{1}{2})$ and $(1, 0)$, verify $\frac{1}{2} \to \frac{1}{2} = \frac{1}{2} \neq 1, 1 \to 0 = 0 \neq 1$
- In P_3, the added pairs $(\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$, verify $\frac{1}{2} \to 0 = 0 \neq 1, \frac{1}{2} \to \frac{1}{2} = \frac{1}{2} \neq 1$

Notice that in the cases Ł$_3$ and G_3 it is $a \to b = 1 \Leftrightarrow a \leq b$. Instead, in K_3 it is only $a \to b = \max(1-a, b) = 1 \Rightarrow a = 0$, or $b = 1 \Rightarrow a \leq b$, but not reciprocally. It seems that the relation 'If a, then b' given by $a \to b = 1$, needs to be enlarged to reach both principles but without an agreement with the truth table of $\to$. The case K_3, with the change of $\frac{1}{2} \to \frac{1}{2} = \frac{1}{2}$ by $\frac{1}{2} \to \frac{1}{2} = 1$, results to be coincidental with Ł$_3$. This is a subject that, related to the semantics of the system, deserves more thinking.

7.3

Another consideration about the operators $\to$ drives us to see into its conditional behaviour related to the $\leq$ or $\vDash$ relations. That is [4], whether or not the *Modus Ponens* rule of inference, $a \cdot (a \to b) \leq b$, or $a \cdot (a \to b) \vDash b$, do hold.

Only the Gödel system, G_3, verifies them, as is easy to prove. For the rest of the cases the following counter-examples answer the question.

1. Ł$_3$, with respect to $\vDash_{Ł} = \leq$. It is $\frac{1}{2} \cdot (\frac{1}{2} \to 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2} \not\leq 0$, that is $(\frac{1}{2}, 0) \notin \leq$.
2. K_3 , with respect to $\vDash_K$, and $\vDash_K^* = \leq$. It is $\frac{1}{2} \cdot (\frac{1}{2} \to 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$, and $(\frac{1}{2}, 0) \notin \leq = \vDash_K^*$ and $(\frac{1}{2}, 0) \notin \vDash_K$ too.
3. B_3 , with respect to $\vDash_B, \vDash_B^*, \vDash_B^{**}$, and $\leq$. It is $\frac{1}{2} \cdot (\frac{1}{2} \to 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$, and $(\frac{1}{2}, 0) \notin \leq$, $(\frac{1}{2}, 0) \notin \vDash_B$, $(\frac{1}{2}, 0) \notin \vDash_B^*$ and $(\frac{1}{2}, 0) \notin \vDash_B^{**}$ too.
4. P_3 , with respect to both $\vDash_P$ and $\leq$. It is $\frac{1}{2} \cdot (\frac{1}{2} \to 0) = \frac{1}{2} \cdot 0 = \frac{1}{2}$, and $(\frac{1}{2}, 0) \notin \leq$, $(\frac{1}{2}, 0) \notin \vDash_P$. With respect to both $\vDash_P^*$ and $\vDash_P^{**}$. It is $\frac{1}{2} \cdot (\frac{1}{2} \to \frac{1}{2}) = \frac{1}{2} \cdot \frac{1}{2} = 1$, and $(1, \frac{1}{2}) \notin \vDash_P^*$, $(1, \frac{1}{2}) \notin \vDash_P^{**}$.

Hence, it is difficult to see the interest of such systems from the point of view of inference.

7.4

Last, but not least significant, is to notice how the different operators $\to$ verify the five properties characterizing implication functions [3], namely, i) $0 \to y = 1$, ii) $1 \to y = y$, iii) $x \to (y \to z) = y \to (x \to z)$, iv) $\to$ is decreasing in the first variable, and v) $\to$ is increasing in the second variable.

A simple calculation shows that in Ł$_3$, G_3, and K_3 the operation $\to$ verifies the five properties, therefore in these systems $\to$ is an implication function. On the contrary, B_3 does verify only properties ii and iii, and P_3 only i, iii, and iv. Hence, neither in B_3, nor in P_3, $\to$ is an implication function.

In the same vein, since in P_3 it is $1' = \frac{1}{2}$, but not $1' = 0$, operation $'$ is not, properly speaking, what is usually designed as a negation.

Acknowledgment

The authors express their thanks to the three anonymous reviewers for their hints and comments.

References

[1] Aristotle. *Metaphysics*. Book IV,4. Translated by W. D. Ross, eBooks@Adelaide, 2007.

[2] L. Bolc and P. Borowik. *Many-valued Logics*. Springer-Verlag, Berlin, 1992.

[3] D. Dubois and H. Prade. *Fundamentals of Fuzzy Sets*. Kluwer, Upper Saddle River, NJ, 2000.

[4] G. M. Hardegree. Material implication in orthomodular (and boolean) lattices. *Notre Dame Journal of Formal Logic*, 22(2):163–182, 1981.

[5] N. Rescher. *Many-valued Logic*. McGraw-Hill, New York, 1969.

[6] E. Trillas. An algebraic view on aristotle's nc and em principles. Submitted to *Theoria*.

[7] E. Trillas, C. Alsina, and A. Pradera. Searching for the roots of non-contradiction and excluded-middle. *International Journal of General Systems*, 31(5):499–513, 2002.

ISBN: 978-989-95079-6-8

Parametric Fuzzy Modelling Framework for Complex Data–Inherent Structures

Arne–Jens Hempel Steffen F. Bocklisch
Department of Electrical Engineering, Chair for Systems Theory, Chemnitz University of Technology
Chemnitz, Germany
Email: arne-jens.hempel@etit.tu-chemnitz.de, steffen.bocklisch@etit.tu-chemnitz.de

Abstract— *The present article dedicates itself to fuzzy modelling of data–inherent structures. In particular two main points are dealt with: the introduction of a fuzzy modelling framework and the elaboration of an automated, data–driven design strategy to model complex data–inherent structures within this framework.*

The innovation concerning the modelling framework lies in the fact that it is consistently built around a single, generic type of parametrical and convex membership function. In the first part of the article this essential building block will be defined and its assets and shortcomings will be discussed.

The novelty regarding the automated, data–driven design strategy consist in the conservation of the modelling framework when modelling complex (nonconvex) data–inherent structures. Instead of applying current clustering methods the design strategy uses the inverse of the data structure in order to created a fuzzy model solely based on convex membership functions.

Throughout the article the whole model design process is illustrated, section by section, with the help of an academic example.

Keywords— pattern recognition, fuzzy classification, fuzzy modelling of data structures, data–driven fuzzy classifier design

1 Introduction

Nowadays the world suffocates in vast amounts of data. To give those data an interpretable and thus economical meaning it has to be analysed. The goal of such an analysis is the creation of a model or the classification of the considered phenomenon, e. g. modelling of the traffic flow in cities, medical or machine diagnosis [1, 2, 3].

Basically there are two main philosophies to deduce such a model, theoretical and experimental modelling. In experimental modelling it is assumed that measurement data (objects) reflect the complexity of the phenomenon under consideration through data–inherent structures. Unfortunately, the same data might also exhibit imprecision (e. g. measuring inaccuracies) or depict interesting phenomena characteristics just vaguely (because of missing information). With the help of fuzzy set theory these occurring inaccuracies can be taken into account as a supplementary model feature [4].

In this work the whole modelling problem is understood as a fuzzy classification task, where specific fuzzy sets form a model equivalent. As it is pointed out in [5], there are a lot of sophisticated solutions for such a task, which in general apply nonparametric fuzzy sets or a composition of different fuzzy sets. Contrary to those approaches, the main philosophy behind this work is the exclusive usage of one specific parametrical fuzzy set to model complex data–inherent structures as well as the data itself. Another aspect of the here pursued type of structure modelling is that it works in the original feature space without any transformation like fuzzy support vector classifiers or any assumption of fuzzy functions [6, 7].

2 Fuzzy Pattern Classes

In order to become acquainted with the modelling philosophy it is necessary to understand its core component, the so called *fuzzy pattern class* (FPC). The subsequent sections provide a basic survey about the definition, composition, capabilities and utilisation of *fuzzy pattern class models*.

From a fuzzy theoretical perspective fuzzy pattern classes correspond to a side–specific parametrical multivariate membership function. FPCs are referred to as classes since they emerge from an agglomeration of class supporting objects (see section 2.2), consequently they represent an superordinate entity.

2.1 Definition of a Fuzzy Pattern Class

Although the usual FPC–membership function is multidimensional it derives from one–dimensional basis functions. Hence, it is reasonable to study this basis functions being equivalent to one–dimensional fuzzy pattern classes first. Generally a one–dimensional fuzzy pattern class A is defined over its class space U based on a side–specific parametrical function concept, see (1).

$$\mu^A(a,\vec{p}) = \begin{cases} \dfrac{a}{1+\left(\dfrac{1}{b_l}-1\right)\left|\dfrac{u}{c_l}\right|^{d_l}}, & u<0 \\ \dfrac{a}{1+\left(\dfrac{1}{b_r}-1\right)\left|\dfrac{u}{c_r}\right|^{d_r}}, & u\geq 0 \end{cases} \tag{1}$$

The function concept comprises a set of seven parameters a and $\vec{p}=(b_l,b_r,c_l,c_r,d_l,d_r)$. The further specification of these parameters results from the fact that the parameter a characterises an entire fuzzy pattern class, whereas the parameters combined in $\vec{p}$ are related to a dimension of the class space [8]. Beyond their mere mathematical functionality all parameters possess the following semantical meaning:

- The parameter a represents the maximum membership value of the FPC μ^A. Regarding a structure of classes the parameter a expresses the weight of a specific class. Considering a dynamic classification process a embodies the topicality or authenticity of the information represented by that class [9, 10].
- In the normalised case $a=1$, the parameters b_l, b_r of $\vec{p}$ assign left– and right–sided membership values at the class borders $u=-c_l$ and $u=c_r$.
- c_l, c_r mark the support of a class in a crisp sense. Both parameters characterise the left– and right–sided expansions of a fuzzy pattern class.
- The continuous descent of the membership function is specified by the parameters d_l, d_r. From a graphical point of view d_l, d_r determine the shape of the membership function, or in

ISBN: 978-989-95079-6-8

other words, the fuzziness of a class. Fig. 1 illustrates the introduced concept of the membership function considering the general unidimensional case.

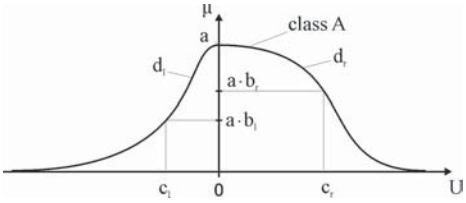


Figure 1: Membership function and parameters.

To obtain the common multidimensional fuzzy pattern class A the basis functions of each class dimension are accumulated using the N-fold compensatory Hamacher intersection operator (2), where n denotes the index of the basis functions and N the total number of dimensions [11].

$$k_{Ham\cap}\mu_N^A = \frac{1}{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{\mu_n^A}} \tag{2}$$

Regarding the main philosophy behind this paper, the most important feature of this intersection operation is the conservation of the parametrical function concept for the multidimensional case [11].

Considering the multidimensional FPC form the class describing set of parameters is supplemented by a class specific position $\vec{u}_0$ in the original feature space and a class specific orientation $\vec{\varphi}$. Fig. 2 depicts the influence of the additional parameters for a two dimensional three class structure. The different location of each class results from $\vec{u}_{c10} = (0.8, 0.2)^T$ $\vec{u}_{c20} = (0.5, 0.5)^T$ and $\vec{u}_{c30} = (0.2, 0.8)^T$, whereas an additional class orientation φ of 60° has been applied to the middle class.

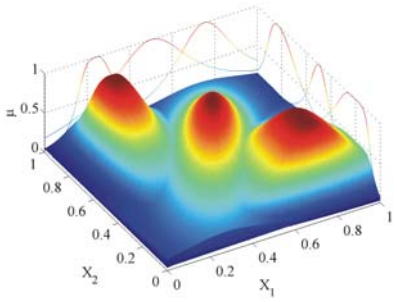


Figure 2: Two dimensional three class structure

2.2 *Data Driven Design of Fuzzy Pattern Classes*

Being familiar with the definition of fuzzy pattern classes a more intriguing question is how fuzzy pattern class models can be deduced. As a matter of principle such models can be obtained via two different approaches [8].

First they can be defined by expertise. That is an expert determines all class parameters based upon task and domain specific knowledge. This approach is not pursued here.

The second approach is a data–driven method, strongly advocating the here featured goal to model data–inherent structures. Based upon a class labelled set of learning data x_l (e.g. measurement objects) all class parameters are assigned automatically by a two step aggregation procedure [9, 10]. The class labels might result from a preliminary conducted cluster analysis.

For the sake of clarity only the basic aggregation principle will be outlined in the following. A full description can be found in [9, 12]. In order to perform the aggregation on sound mathematical foundations, the crisp learning dataset is extended to a set of fuzzy objects, using the introduced function concept (1). This extension is justified by the fact that every observation (measurement) inheres a so called "elementary fuzziness" (e.g impression of a sensor) [9]. In the first aggregation step the class position $\vec{u}_0$, alignment φ and extensions c_l, c_r are calculated in a dimension–wise manner. As exemplified in Fig. 3 the position $\vec{u}_0$ and alignment $\vec{\varphi}$ of the class space is obtained via a principal component analysis (PCA), where $\vec{u}_0$ is defined by the mean over all learning objects and $\vec{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_{N-1})^T$ by sequential rotation of the class space U into the principal axes.

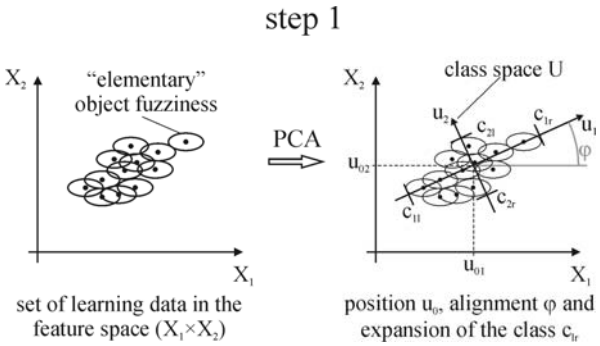


Figure 3: Step 1: aggregation procedure

The extensions c_l, c_r of the class are determined by the outermost objects in each class space dimension. In the subsequent second aggregation step the class shape d_l, d_r and the border memberships b_l, b_r are determined also dimension–wisely.

After the transformation of the objects x_l into their corresponding class space U the shape of a fuzzy patter class (d_l, d_r) is assigned based on their agglomeration properties. The more the data resembles an agglomeration according to a geometric series the smoother the class shape. The rate of resemblance is determined by the mean distance between two adjacent objects. The smoothest class shape is obtained for $d_{lr} = 2$ where the objects are cumulating in the centre of the class conform to a geometric series. The crisp case results for ($d_{lr} \to \infty$) where the objects are equally distributed over the class space, however for calculation purposes $d_{lr} = 20$ has proven to be a sufficient value to represent the crisp case.

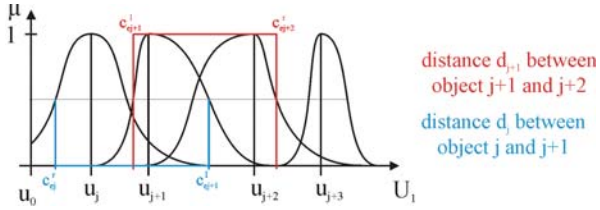


Figure 4: Step 2a: determination of the class form

The values for border memberships b_l, b_r are derived by the

ISBN: 978-989-95079-6-8

conservation of the object cardinality taking into account the results of c_{lr} and d_{lr}. Since the integral over the class membership function (1) cannot be solved analytically the border memberships b_l, b_r are estimated by a binary search over the unity interval $b_{lr} \in [0,1]$ [9].

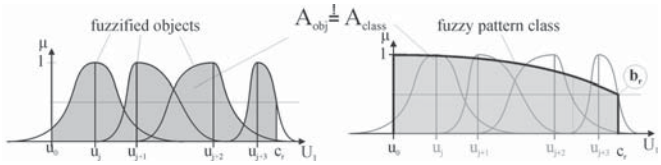


Figure 5: Step 2b: claim for cardinality

The class "weight" a is calculated by a logistic function given the number of the class supporting objects, see [9].

2.3 Application of FPC and Mode of Operation

For applicative purposes all task relevant fuzzy pattern classes are grouped together in a so called fuzzy pattern classifier. In operating mode the fuzzy pattern classifier assigns unknown objects to this class structure. The objects to be classified are each denoted by a vector $\vec{x}$ of their features:

$$\vec{x} = (x_1, x_2, ..., x_N)^T , \tag{3}$$

where N represents the number of feature dimensions. The results of the classification process are stored into a so called vector of sympathy $\vec{s}$. The components of $\vec{s}$ express the membership of a classified object to the corresponding class:

$$\vec{s} = (s_1, s_2, ..., s_K)^T , \tag{4}$$

where K is the total number of classes. The gradual membership of an object to a given class is calculated using (1).

$$s_k = \mu^k(\vec{x}) \quad \text{for } k = 1, 2, ..., K \tag{5}$$

Figure 6 illustrates the process of classification with the help of a one-dimensional three class structure. The object to be classified is situated in the centre of class two, the right outskirts of the first class and in the left centre of the third class. Alongside with the classification task the classification results are listed.

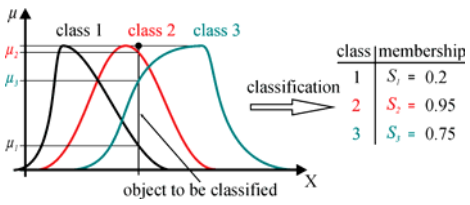


Figure 6: Object classification

According to Fig.6 the vector of sympathy describes a unique assignment of the object to the class structure with respect to its location in the feature space, since there are three classes it contains three values of membership.

2.4 Properties of Fuzzy Pattern Classes

In order to round off the comprehension about the afore introduced concept of fuzzy pattern class models its major properties (advantages and drawbacks) will be outlined subsequently. The main features of the fuzzy pattern class model lie in its versatility, its uniformity and its closed modelling framework. All these features can be attributed to the unimodal, side–specific and parametric class membership function. In the most general case the class membership function offers multivariate FPC models with various and asymmetric shapes, ranging from peak- over bell- to crisp shaped fuzzy sets. In connection with the introduced data–driven design procedure the class membership function allows to map class internal object distributions onto its shape and to model correlative relations without loosing its fuzzy logic basis. Besides this flexibility it has to be stressed that the parameters are semantically motivated or have at least a semantical meaning. It is therefore that the fuzzy pattern classes are considered to be well interpretable and transparent. Another feature resulting from the utilisation of the parametric membership concept is its good trade off between data compression, computational cost and generality. Due to the choice of the membership function (1) each fuzzy pattern class is defined on a set of eight parameters per dimension, providing a sufficient data compression, especially for high dimensional models. Due to the choice of the conjunction operator (2) the intersection operation is exclusively performed on *parameter level* saving computational cost. Both advantages are traded off for generality in so far as class membership functions are convex models, specifying a convex area of the feature space. Consequently FPC are best suited to model convex data–inherent structures. But their convex nature causes fuzzy pattern class models to be afflicted with significant errors when it comes to model nonconvex data–inherent structures.

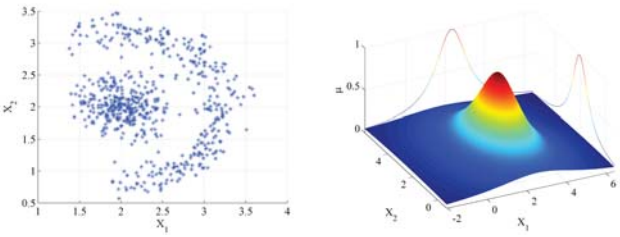


Figure 7: Nonconvex data structure and according FPC model

An example for such an error is depicted in Fig.7 where a central object accumulation enclosed by a half circle shaped data structure was aggregated to a fuzzy pattern class. Obviously the region between the object accumulations does not belong to the given data structure but the associated fuzzy pattern model μ^C will assign high grades of memberships for this region.

In order to circumvent this major drawback two possibilities can be thought of:
The first way, is to segment the data into convex subsets, for example with the help of cluster algorithms. Aside from the fact that this approach works on every data–inherent structure it might create considerably large structures of fuzzy pattern classes at the expenses of model clarity and computational costs.

3 Fuzzy Pattern Anti–classes

The hereafter elaborated access to dissolve the convexity drawback arises from the negation of a class assertion over its unsupported class space. The idea behind this approach

ISBN: 978-989-95079-6-8

can be stated as follows: *The difference of convex sets can be a non–convex set.*
Fig.8 sketches such a negation of the FPC model for a ring shaped problem.

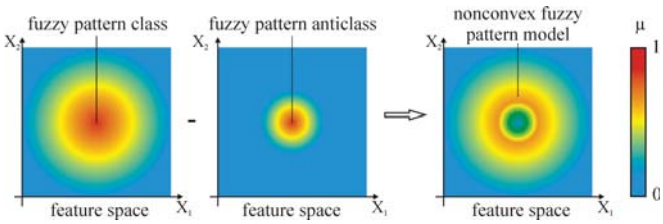


Figure 8: Nonconvex fuzzy pattern model via negation

As it is depicted the negation will be introduced in terms of fuzzy pattern anti–classes (FPAC). It works on semantical level and from this point of view FPACs can be seen as a further specification of a preceding FPC.
Mathematically fuzzy pattern anti–classes are defined upon the same membership function concept as fuzzy pattern classes, see (1). The main concern of of this definition is conservation of the fuzzy pattern modelling framework and with it the *automated model generation* and the *model properties* (such as flexibility, interpretability, computational efficiency, etc.). The enormous shape diversity of fuzzy pattern classes together with the mutual negation of such membership functions allows to model almost any form of data–inherent structures. The only task that needs to be solved can be formulated as follows:
Determine the fuzzy pattern anti–classes given a set of learning data, containing an arbitrary (nonconvex) data–inherent structure and the appendent FPC model.

3.1 Design of Fuzzy Pattern Anti–classes

Assuming that FPACs, like usual fuzzy pattern classes, can be supported by objects or, better so called "anti–objects", then it is in accordance with modelling framework that FPCAs can be designed in a data–driven manner. Aiming to elaborate an automated FPAC design method the already introduced automated databased algorithm can be exploited. However the introduced data–driven FPC design relies on class supporting objects. Correspondingly, in order to setup FPACs these "anti–objects" have to be found first. When considering the general case there are no "anti–objects" given. Consequently they have to be *generated*.
For the sake of clearness the anti–object generation procedure will be illustrated for the two–dimensional case with the help of an academic example. The general multidimensional case follows analogously.

3.1.1 Generation of Anti–objects

According to the task (section 3) the information given are the location of each class supporting object in the learning dataset and the modelling fuzzy pattern class, see Fig.7. Suppose that there is no prior information about the distribution of the learning data it is impossible to make an assumption about the shape of a data–inherent structure. That is why the central idea behind the anti–object generation is led by the assumption that the *anti–objects will agglomerate in the class space being unsupported by learning objects*, such that they will adopt a kind of inverse data–inherent structure and form FPACs.
To ascertain whether a partition of the class space is actually supported by learning objects requires a discretisation of the class space. The size of the class space to be discretised is determined based upon the class borders c_{lr} in each dimension, whereas the discretisation resolution amounts to 2% of the class space leading to a 50×50 matrix F with 2500 cells. A cell of an N–dimensional class space then posses the following the extend:

$$V_c = \frac{1}{50}\left((c_{l1}+c_{r1})\cdot(c_{l2}+c_{r2})\cdot\ldots\cdot(c_{lN}+c_{rN})\right) \qquad (6)$$

Fig.9 depicts the discretised class space for the example along with the class borders in green, the object supported cells highlighted in red and the unsupported class space cells coloured in dark blue.

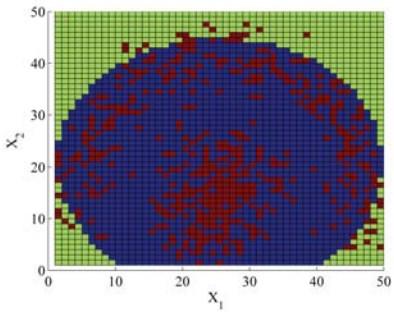


Figure 9: Discretised class space for the example

According to the central idea, the anti–objects have to agglomerate in these unsupported cells. In other words it is necessary to define an model for the agglomeration process. In particular two ways have been pioneered to model the accumulation of anti–objects, namely accumulation based on the elemental fuzziness and accumulation based on power series expansion. The most promising agglomeration model resulted from the assumption of a Fibonacci series.
Its leading thought is the expansion of Fibonacci numbers around the cell of interest until an object supported cell or a class border cell is reached. The result of the expansion around an arbitrary cell (i, j) is a 50×50 matrix $F_{d(i,j)}$ containing only the expanded Fibonacci numbers and zeros.

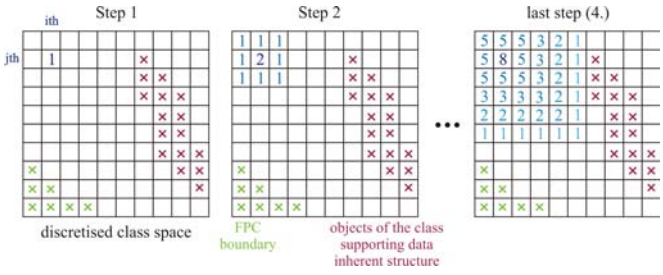


Figure 10: Fibonacci expansion for an arbitrary cell

In this context it has to be mentioned that the fuzzy pattern class borders are treated like objects. Purpose of this procedural manner is to provide an insight of the inner class anti–object distribution by preventing an anti–object accumulation close to the class borders.

ISBN: 978-989-95079-6-8

Fig.10 illustrates the implementation of the Fibonacci expansion for an arbitrary cell at position (i, j). The sum over all Fibonacci matrices $F_{d(i,j)}$ yields the result of the agglomeration process, as in (7).

$$F_d = \sum_{i=1}^{N} \sum_{j=1}^{N} F_{d(i,j)} \tag{7}$$

F_d can be interpreted as a kind of anti–object density matrix. It is broken down into an explicit number of anti–objects per cell based on the maximum object density of all cells d_{max}, see (8).

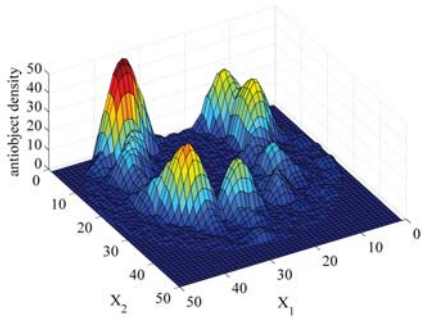


Figure 11: Anti–object density matrix F_d for the example

$$F_a = rd\left(\left(\frac{d_{max}}{max\,(F_d)}\right) F_d\right) \tag{8}$$

With the usual rounding operation rd, in (8), it is assured that each cell contains a natural number of anti–objects ($F_a \in \mathbb{N}^{50\times 50}$).

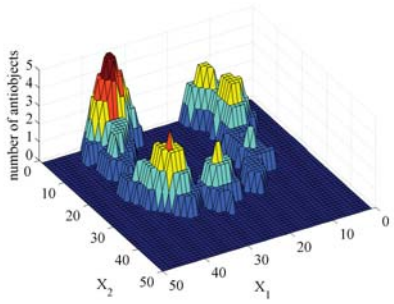


Figure 12: Anti–object matrix F_a for the example

Fig.12 presents the resulting number of anti–objects in their corresponding cells, after applying (8), for the example. The total number of anti–objects, being distributed around the centre region amounts to 887 with a maximum density of five anti–objects per cell.

The final step to complete the anti–object generation comprises the assignment of a position to each anti–object. For this purpose all anti–objects are uniformly distributed over their associated cell. In the here discussed two–dimensional case the anti–object position of an arbitrary cell with the coordinates (i, j) follows from (9), where *rand* draws an random number from the unity interval.

$$\begin{pmatrix} u_{a1} \\ u_{a2} \end{pmatrix} = \begin{pmatrix} i \cdot \frac{1}{50}\,(c_{l1} + c_{r1}) - rand \\ j \cdot \frac{1}{50}\,(c_{l2} + c_{r2}) - rand \end{pmatrix} \tag{9}$$

After their distribution the set of anti–objects forms an inverse of the data–inherent structure, see left–hand side of Fig.13.

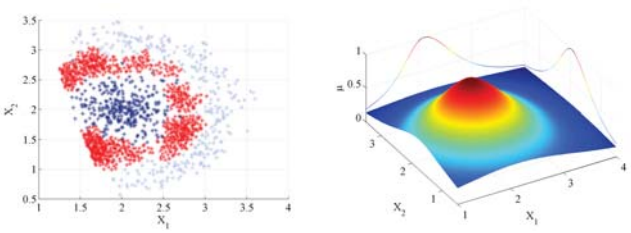


Figure 13: left: Objects, anti–objects, right: according FPAC

However, as it can be seen the anti–objects (in red) form a nonconvex data structure. Modelling this with the help of the automated FPC algorithm again creates a faulty FPAC μ^{AC}, see right–hand side of Fig.13. This might lead to the misconception that the problem was just shifted to the anti–objects. But that is by no means the case, since the whole procedure can be applied again to the created anti–class.

Like a block letter and its print anti–objects and objects complement each other to a complete convex model. Or with another emphasis anti–objects together with objects cover the *entire* class space. Because of that complementary relation, the original objects can be seen as the anti–objects of the anti–objects. Consequently, instead of running through the entire process again for the anti–class, the original objects can be recycled. The reutilisation of the original objects is justified by the facts that a part of the original objects is already complementing the set of anti–objects to a convex description and that their position is already known. Original objects being considered to negate the anti–class are required, firstly to posses a high membership to the anti–class and secondly to be located within the class borders of the anti–class. Generally this applies to original objects with a anti–class membership above $\mu = 0.5$.

For the considered example, the only objects meeting this requirement are the ones located in the centre region, highlighted in dark blue see Fig.13. After setting up the "anti–anti–class" μ^{AAC} by the selected original objects the negation procedure stops owing to the fact the average membership of possible anti–objects drops below a threshold of $\mu = 0.5$.

3.2 Combination of Fuzzy Pattern Anti–class Models

After their generation the anti–class models have to be combined together in order to from an overall model of the data–inherent structure. This combination process is characterised by the following keynote: *If an object does not belong to the anti–class it belongs to the preceding class.* A reasonable implementation of this key concept derives from the concatenation of the natural complement and minimum conjunction [13, 14].

$$\mu = min\left(\mu^{class}, \left(1 - \mu^{anti-class}\right)\right) \tag{10}$$

If the anti–class generation yields several levels of negation their sequence has to be respected for the concatenation. Or differently it has to be respected that *each level of negation creates an new level of hierarchy*. Equation (11) represents this aspect for the example.

ISBN: 978-989-95079-6-8

$$\mu = min\left(\mu^{C}, 1 - min\left(\mu^{AC}, \left(1 - \mu^{AAC}\right)\right)\right) \qquad (11)$$

Fig.14 depicts the nonconvex and fuzzy model, resulting from the class–anti–class combination, together with the supporting objects in black.

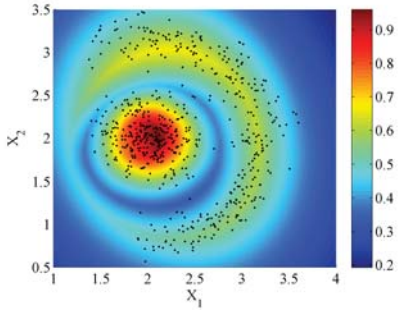


Figure 14: Resulting FPC Model

3.3 Properties of the Fuzzy Pattern Anti–class Design

The introduction of FPAC design is round off by having its assets and drawbacks summarised subsequently.
The main advantage of the introduced design approach lies in the exploitation of a single type of membership function. All the beneficial properties of fuzzy pattern classes, such as transparency, interpretability, computational efficiency etc. are imparted and conserved into the design of nonconvex FPC models. Equally important, by complementing the objects over the considered class space the FPAC design converts the major drawback of fuzzy pattern models (convexity) into an advantage. As a consequence the anti–objects have to be constructed only once. Furthermore the complete object anti–object covering of the class space guarantees convergence of the FPAC design method, since there is no further anti–object generation. Besides that, each anti–class is smaller or equal in size compared to its preceding class such that the area of interest is also converging to zero. Another aspect worth to mention is that the FPAC design works independent from clustering algorithms but features structure capturing.

The major drawback of the design strategy arises from its numerical character, the class space has to be discretised in order to distribute the anti–objects. As a consequence their creation is computationally costly.

4 Conclusions

This paper presents a data driven approach toward modelling of complex data–inherent structures. Its main philosophy is the exclusive usage of a general fuzzy modelling framework. Instead of applying cluster analysis techniques the design strategy aims to complete the convex fuzzy pattern model with the help of so called anti–objects. These anti–objects are not available prior to the design. They have been generated and distributed over the class space. For this purpose a Fibonacci expansion model was elaborated and demonstrated.
With the help of the introduced automated FPC design the anti–objects have been agglomerated to a negating fuzzy pattern classes (FPAC). By preserving the membership function concept the FPAC are on one hand afflicted again with the drawbacks of fuzzy pattern classes. But on the other hand the same design approach can be applied again, eventually leading to a pure convex fuzzy pattern model. A possible combination of the setup FPACs and the original fuzzy pattern model to a hierarchical nonconvex overall FPC model has been demonstrated with the help of an example.

Finally it has to be stressed that the introduced design strategy is a universal approach in so far as it can be applied to any multivariate unimodal convex parametrical membership function.

Further points of research are:
- reducing the numerical character of the method
- combination with cluster algorithms
- creation of fuzzy pattern class and anti–class networks

References

[1] Gernot Herbst and Steffen F. Bocklisch. Classification of keystroke dynamics - A case study of fuzzified discrete event handling. In Bengt Lennartson, Martin Fabian, Knut Akesson, Alessandro Giua, and Ratnesh Kumar, editors, *9th International Workshop on Discrete Event Systems*, 2008.

[2] Arne-Jens Hempel and Steffen F. Bocklisc. Hierarchical modelling of data inherent structures using networks of fuzzy classifiers. In *Tenth International Conference on Computer Modeling and Simulation, 2008. UKSIM 2008.*, pages 230–235. IEEE, 2008. ISBN: 0-7695-3114-8.

[3] C. Weihs & W. Gaul, editor. *Classification - the Ubiquitous Challenge: Proceedings of the 28th Annual Conference of the Gesellschaft Fur Klassifikation E.V., University of Dortmund*, volume 28. Springer, June 2005.

[4] Lotfi Asker Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[5] James C. Bezdek, James Keller, Raghu Krisnapuram, and Nikhil R. Pal. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing (The Handbooks of Fuzzy Sets)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[6] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2006.

[7] Xuehua Li and Lan Shu. Fuzzy theory based support vector machine classifier. *Fuzzy Systems and Knowledge Discovery, Fourth International Conference on*, 1:600–604, 2008.

[8] Steffen F. Bocklisch. *Prozeßanalyse mit unscharfen Verfahren*. VEB Verlag Technik, Berlin, 1987.

[9] Arne-Jens Hempel. Aggregation und Identifikation von Fuzzy−Objekten mit unterschiedlichen elementaren Unschärfen. Master's thesis, TU Chemnitz, 2005.

[10] Michael Päßler. Mehrdimensionale Zeitreihenmodellierung und Prognose mittels Fuzzy Pattern Modellen. Master's thesis, TU Chemnitz, 1998.

[11] Ulrich Scheunert. *Neue Verfahren der Fuzzy−Mengen−Verknüpfung und Fuzz−Arithmetik und ihrer Anwendung bei der Sensor−Daten−Fusion*. PhD thesis, TU Chemnitz, 2001.

[12] Steffen F. Bocklisch and Norman Bitterlich. *Fuzzy systems in computer science*, chapter Fuzzy pattern classification – methodology and application. Vieweg, Wiesbaden, 1994.

[13] Leszek Rutkowski. *Flexible Neuro-fuzzy Systems: Structures, Learning and Performance Evaluation (Kluwer International Series in Engineering and Computer Science)*. Kluwer Academic Publishers, Norwell, MA, USA, 2004.

[14] H. J. Zimmermann. *Fuzzy Set Theory and its Applications*. Springer, 4 edition, October 2001.

ISBN: 978-989-95079-6-8

On Triangular Norm Based Fuzzy Description Logics

Àngel García-Cerdaña Eva Armengol Francesc Esteva

IIIA, Artificial Intelligence Research Institute
CSIC, Spanish Council for Scientific research
Campus UAB, 08193 Bellaterra, Spain
Email: {angel, eva, esteva}@iiia.csic.es

Abstract— *Description Logics (DLs) are knowledge representation languages useful to represent concepts and roles. Fuzzy Description Logics (FDLs) incorporate both vague concepts and vague roles modeling them as fuzzy sets and fuzzy relations respectively. In the present paper, following ideas from Hájek, we propose the use of t-norm based (fuzzy) logics with truth constants in the language as logics underlying the fuzzy description language. We introduce the languages $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ as an adequate syntactical counterpart of some semantic calculi given in different works dealing with FDLs.*

Keywords— Description Logics, Fuzzy Description Logics, t-Norm Based Fuzzy Logics, Truth-constants, Involutive negation.

1 Introduction

Description Logics (DLs) are knowledge representation languages (particularly suited to specify formal ontologies), which have been studied extensively over the last two decades. A full reference manual of the field is [1]. The vocabulary of DLs consists of *concepts*, which denote sets of individuals, and *roles*, which denote binary relations among individuals. From atomic concepts and roles DL systems allow, by means of *constructors*, to build complex descriptions of both concepts and roles, which are used to describe a domain through a knowledge base (KB) containing the definitions of relevant domain concepts or some hierarchical relationships among them (TBox) and a specification of properties of the domain instances (ABox). One of the main issues of DLs is the fact that the semantics is given in a Tarski-style presentation and the statements in both TBox and ABox can be identified with formulae in first-order logic or a slight extension of it, and hence we can use reasoning to obtain implicit knowledge from the explicit knowledge in the KB.

A natural generalization to cope with vague concepts and relations consists in interpret DL concepts and roles as fuzzy sets and fuzzy relations, respectively. From this point of view, it is at the end of the last decade (from 1998) when several proposals of Fuzzy Description Logics (FDLs) were introduced (e.g., the first ones by Yen [19], Tresp and Molitor [18] and Straccia [14]). However, the logic framework behind these initial works is very limited. With the aim of enriching the expressive possibilities Hájek [9] proposes to take t-norm based fuzzy logics as logics underlying FDLs. This change of view gives a wide number of choices on which a DL can be based: for every particular problem we can consider the t-norm based (fuzzy) logic that seems to be more adequate. As an example, Hájek studies the FDL associated with the description language $\mathcal{ALC}$. After this work, several researchers on FDLs have developed approaches based on the spirit of Hájek's paper, even though their work is more related to expressiveness and algorithms than in its logical base (see for instance [16, 13, 17]).

The main motivation of the present work is based on the following consideration: since the axioms of the bases of knowledge in FDLs include truth degrees (see for instance [14]), a natural choice is to include symbols for these degrees in both, the description language and, as truth constants, in the t-norm based logic where that language is interpreted.

To this goal in the present paper we propose two new families of description languages, denoted by $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ that are extensions of the language $\mathcal{ALC}$ considered by Hájek in [9]. After some introductory notions we define their semantics and describe the corresponding knowledge base (TBox and ABox) from a syntactic and semantic perspective and, taking advantage of having truth constants in the logic, we define graded notions of validity, satisfiability and subsumption. We also give some representative example and some new results for the case of $\mathcal{ALC}$ language over Gödel logic with truth constants and an involutive negation.

2 Fuzzy Logic: basics

In the last decade a family of fuzzy logics as t-norm based fuzzy logics has defined and studied (see, for instance, the monograph [8] for the main notions used in this section). They are multi-valued systems with additive conjunction and disjunction, multiplicative conjunction, implication, negation and the constant $\bar{0}$ which are interpreted in $[0, 1]$ as $\min$, $\max$, a continuous t-norm $*$, its residuum $\rightarrow_*$, the negation function $n(x) = x \rightarrow_* 0$ and 0, respectively. This interpretation is mainly defined by a continuous t-norm and its residuum which justify the name t-norm based logics. Remember that the main continuous t-norms are the minimum, the product and the Lukasiewicz since all other continuous t-norm are ordinal sum of these three basic ones.

2.1 From BL to the logic of a continuous t-norm

The Basic fuzzy Logic (BL) (defined in [8]) has the following basic connectives: *multiplicative conjunction* ($\&$) *implication* ($\rightarrow$) (both binary) and *falsity* ($\bar{0}$) (nullary). It is defined by the inference rule of *Modus Ponens* and the following schemata (taking $\rightarrow$ as the least binding connective):

ISBN: 978-989-95079-6-8

(BL1) $(\varphi \to \psi) \to ((\psi \to \chi) \to (\varphi \to \chi))$
(BL2) $\varphi\&\psi \to \varphi$
(BL3) $\varphi\&\psi \to \psi\&\varphi$
(BL4) $\varphi\&(\varphi \to \psi) \to \psi\&(\psi \to \varphi)$
(BL5a) $(\varphi \to (\psi \to \chi)) \to (\varphi\&\psi \to \chi)$
(BL5b) $(\varphi\&\psi \to \chi) \to (\varphi \to (\psi \to \chi))$
(BL6) $((\varphi \to \psi) \to \chi) \to (((\psi \to \varphi) \to \chi) \to \chi)$
(BL7) $\bar{0} \to \varphi$

The usual defined connectives are introduced as follows:

$\varphi \vee \psi := ((\varphi \to \psi) \to \psi) \wedge ((\psi \to \varphi) \to \varphi)$,
$\varphi \wedge \psi := \varphi\&(\varphi \to \psi)$, $\varphi \leftrightarrow \psi := (\varphi \to \psi)\&(\psi \to \varphi)$,
$\neg\varphi := \varphi \to \bar{0}$, $\bar{1} := \neg\bar{0}$.

Łukasiewicz, Product and Gödel Logics can be obtained as axiomatic extensions of BL with the following axioms: $\neg\neg\varphi \to \varphi$ for Łukasiewicz; $(\varphi \wedge \neg\varphi) \to \bar{0}$ and $\neg\neg\chi \to (((\varphi\&\chi) \to (\psi\&\chi)) \to (\varphi \to \psi))$ for Product; and $\varphi \to \varphi\&\varphi$ for Gödel.

An *evaluation of propositional variables* is a mapping e assigning to each variable p a *truth value* $e(p) \in [0,1]$. Given a continuous t-norm $*$, the evaluation e is extended inductively to a mapping of all formulas into the so-called *standard algebra* $[0,1]_* = \langle [0,1], *, \to_*, \max, \min, 0, 1\rangle$ defined on [0,1] by the t-norm and its residuum in the following way: $e(\varphi\&\psi) = e(\varphi) * e(\psi)$; $e(\varphi \to \psi) = e(\varphi) \to_* e(\psi)$; $e(\bar{0}) = 0$. In [3] it is proved that the formal system BL is sound and complete w.r.t. the standard algebras defined in $[0,1]$ by continuous t-norms and their residua, i.e., a formula φ is provable in BL if and only if it is a common tautology of all standard algebras defined by a continuous t-norm and its residuum. Consequently we say that BL is the logic of continuous t-norms and their residua. It is also well known that Łukasiewicz (Product, Gödel) Logic is the logic of the t-norm of Łukasiewicz (Product, Gödel) and its residuum in the sense that a formula is provable in each logic if and only if it is a tautology over the standard algebra defined by the corresponding t-norm and its residuum.

In [6] the logic L^* of each continuous t-norm $*$ and its residuum is proved to be finitely axiomatizable as extension of BL. Moreover it is also given an algorithm to find a finite set of axioms characterizing each logic L^*. When the negation $\neg\varphi := \varphi \to \bar{0}$ defined in L^* is not involutive, a new logic $L^*_\sim$, expanding L^* by adding an involutive negation, can be considered. This negation, denoted by $\sim$, could be introduced, as is done in the context of intuitionistic logic (see [11]) or in the context of Gödel or Product logics (cf.[5]) by adding the axioms

(~ 1) $\sim\sim\varphi \to \varphi$
(~ 2) $\sim(\varphi \vee \psi) \leftrightarrow (\sim\varphi \wedge \sim\psi)$
(~ 3) $\neg\varphi \to \sim\varphi$

Notice that having an involutive negation in the logic enriches the representational power of the logical language in a non-trivial way because a multiplicative (or strong) disjunction $\varphi \veebar \psi$ is definable now (by duality) as $\sim(\sim\varphi\& \sim\psi)$, being the associated truth function $\oplus$ defined as $x \oplus y := n(n(x) * n(y))$, where n is the truth function of $\sim$ and, on the other hand, a contrapositive implication $\varphi \hookrightarrow \psi$ is definable as $\sim\varphi \veebar \psi$, with truth function $\hookrightarrow_\oplus$ defined as $x \hookrightarrow_\oplus y = n(x) \oplus y$.

2.2 The logics $BL\forall$ and $L^*\forall$

The language of the *basic fuzzy predicate logic* $BL\forall$ consists of a set of *predicate symbols* $\mathcal{P} = \{P, Q, \dots\}$, each together with its *arity* $n \geq 1$ and a set of *constant symbols* $\mathcal{C} = \{c, d, \dots\}$. The logical symbols are *variable symbols* $x, y, \dots$, *connectives* $\&, \to, \bar{0}$ and *quantifiers* $\forall, \exists$. Other connectives $(\vee, \wedge, \neg, \leftrightarrow, \bar{1})$ are defined as in BL. *Terms* are constant symbols and variable symbols. An *atomic formula* is an expression of the form $P(t_1, \dots, t_n)$, where P is a predicate letter of arity n and $t_1, \dots, t_n$ are terms. The set of predicate formulas is built from atomic formulas in the usual way.

A *fuzzy interpretation* for our language is a tuple $\mathbf{M} = \langle M, (r_P)_{P\in\mathcal{P}}, (m_c)_{c\in\mathcal{C}}\rangle$, where M is a set, for each n-ary predicate symbol P, r_P is a fuzzy n-ary relation $M^n \to [0,1]$; and for each constant symbol c, m_c is an element of M.

Given a continuous t-norm $*$, an $\mathbf{M}$-*evaluation* of the variables assigns to each variable x an element $v(x)$ of M. From M and v we define the *truth value of a term* t in the following way: $\|t\|_{\mathbf{M},v} = v(t)$ when t is a variable, and $\|t\|_{\mathbf{M},v} = m_c$ when t is a constant c. The *truth value of a formula* φ *for an evaluation* v, denoted by $\|\varphi\|^*_{\mathbf{M},v}$, is a value in $[0,1]$ defined inductively as follows:

$$\begin{array}{ll} \|\varphi\|_{\mathbf{M},v} & \text{if } \varphi \text{ is an atomic formula,} \\ 0 & \text{if } \varphi = \bar{0}, \\ \|\alpha\|^*_{\mathbf{M},v} * \|\beta\|^*_{\mathbf{M},v} & \text{if } \varphi = \alpha\&\beta, \\ \|\alpha\|^*_{\mathbf{M},v} \to_* \|\beta\|^*_{\mathbf{M},v} & \text{if } \varphi = \alpha \to \beta, \\ \inf\{\|\alpha\|^*_{\mathbf{M},v} : v \equiv_x v'\} & \text{if } \varphi = (\forall x)\alpha, \\ \sup\{\|\alpha\|^*_{\mathbf{M},v} : v \equiv_x v'\} & \text{if } \varphi = (\exists x)\alpha, \end{array}$$

where $v \equiv_x v'$ means that for all variables $y \neq x$, $v(y) = v'(y)$.

The *truth value of a formula* φ is defined by $\|\varphi\|^*_{\mathbf{M}} := \inf\{\|\varphi\|^*_{\mathbf{M},v} : v \text{ is an } \mathbf{M}\text{-evaluation}\}$. A formula φ is a $*$-*tautology* if $\|\varphi\|^*_{\mathbf{M}} = 1$ for every fuzzy interpretation $\mathbf{M}$; φ is a *standard tautology* if it is a $*$-*tautology* for each continuous t-norm $*$. The following standard tautologies are taken as axioms of the basic fuzzy predicate logic $BL\forall$ (see [8]): *a)* the axioms of BL; *b)* the following axioms on quantifiers:

$(\forall 1)$ $(\forall x)\varphi(x) \to \varphi(t)$ (t substitutable for x in $\varphi(x)$),
$(\exists 1)$ $\varphi(t) \to (\exists x)\varphi(x)$ (t substitutable for x in $\varphi(x)$),
$(\forall 2)$ $(\forall x)(\varphi \to \psi) \to (\varphi \to (\forall x)\psi)$ (x not free in φ),
$(\exists 2)$ $(\forall x)(\varphi \to \psi) \to ((\exists x)\varphi \to \psi)$ (x not free in ψ),
$(\forall 3)$ $(\forall x)(\varphi \vee \psi) \to (\forall x)\varphi \vee \psi$ (x not free in ψ).

Deduction rules are (as in classical logic) *Modus Ponens* and *Generalization*. Notions of *proof*, *provability*, *theory*, etc., are defined in the usual way. Let $\mathcal{C}$ be an axiomatic extension of BL. $\mathcal{C}\forall$ is obtained by taking the axioms and rules of $BL\forall$ plus the axioms characterizing $\mathcal{C}$. Thus, given a continuous t-norm $*$, the predicate logic $L^*\forall$ is the logic obtained from L^* by adding to its axiomatization the schemas for quantifiers and the rule of generalization.

2.3 Adding truth constants to the language

T-norm based logics are infinite-valued logics. However, the advantage of being a many-valued logic is not used in current approaches since the semantic deduction of formulas do not take into account the *intermediate or partial truth degrees*. That is to say, current approaches use a truth-preserving con-

ISBN: 978-989-95079-6-8

sequence relation in the same way as in the classical logic, i.e. deduce true formulas (having value 1) from sets of true formulas. An elegant way to take advantage from being many-valued is to introduce truth constants into the language, as it is done by Pavelka in [12] and more recently in [8, 6, 4]. The approach considered in this paper is based in these ideas.

Given a continuous t-norm $*$, its residuum $\rightarrow_*$ and its corresponding logic L^*, let $\mathbf{S} = \langle S, *, \rightarrow_*, \max, \min, 0, 1\rangle$ be a *countable* (i.e., finite or enumerable) *subalgebra* of the corresponding standard algebra $[0,1]_*$. The expansion of L^* adding into the language a truth constant $\bar{r}$ for each $r \in S$, denoted by $L^*(\mathbf{S})$, is defined as follows:

i) the language of $L^*(\mathbf{S})$ is the one of L^* plus a truth constant $\bar{r}$ for each $r \in S$,

ii) the axioms and rules of $L^*(\mathbf{S})$ are those of L^* plus the *book-keeping* axioms: for each $r, s \in S \setminus \{0,1\}$, $\bar{r}\&\bar{s} \leftrightarrow \overline{r * s}$ and $(\bar{r} \rightarrow \bar{s}) \leftrightarrow \overline{r \rightarrow_* s}$.

Completeness results for propositional logic $L^*(\mathbf{S})$ when $*$ is a continuous t-norm has been fully studied in [4].

2.4 *Defining the logics $L^*(\mathbf{S})\forall$ and $L^*_\sim(\mathbf{S})\forall$*

When the negation associated to the continuous t-norm $*$ is not involutive, the logic $L^*_\sim(\mathbf{S})$ can be defined in a similar way although in this case $\mathbf{S}$ has to be a countable subalgebra of the algebra obtained by adding the truth function of the involutive negation $n(x) := 1 - x$ to the operations of $[0,1]_*$. Moreover we need to add the book-keeping axioms for the involutive negation: $\sim \bar{r} \leftrightarrow \overline{n(r)}$. The corresponding predicate logics $L^*(\mathbf{S})\forall$ and $L^*_\sim(\mathbf{S})\forall$ are respectively obtained from $L^*\forall$ and $L^*_\sim\forall$ by expanding the language with a truth constant $\bar{r}$ for every $r \in S$ and by adding the book keeping axioms. The truth value of the formula $\bar{r}$ is given by $\|\bar{r}\|^*_{\mathbf{M}} = r$. The logics $L^*(\mathbf{S})\forall$ and $L^*_\sim(\mathbf{S})\forall$ will be the basis of our proposal for the description languages presented in the next section.

3 The description languages $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$

Similarly as first-order logic is the counterpart for interpreting the classic description language $\mathcal{ALC}$, the logics $L^*(\mathbf{S})\forall$ and $L^*_\sim(\mathbf{S})\forall$ will be the counterpart for interpreting the description languages $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ we will define in this section. In these languages, take an special role the so-called evaluated formulas. An *evaluated formula* is a formula of one of the types $\bar{r} \rightarrow \varphi$, $\varphi \rightarrow \bar{r}$ where φ is a formula without new truth constants (i.e., different from $\bar{0}$ and $\bar{1}$). In this setting $\bar{r} \leftrightarrow \varphi$ is definable as $(\bar{r} \rightarrow \varphi)\&(\varphi \rightarrow \bar{r})$. The name of evaluated formulas cames from the fact that $e(\bar{r} \rightarrow \varphi) = 1$ (resp. $e(\varphi \rightarrow \bar{r}) = 1$) if and only if $e(\varphi) \geq r$ (resp. $e(\varphi) \leq r$). Thus, evaluated formulas correspond to the type of formulas (under the notation $\langle \varphi, \succcurlyeq r\rangle$ and $\langle \varphi, \preccurlyeq r\rangle$) used for the knowledge bases in papers on FDLs (cf.[14, 15, 17]). Next we define the description languages $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ from a syntactic and semantic perspective. Then we introduce the notions of *TBox* and *ABox* for that languages.

Syntax. In the languages of description we start from *atomic concepts* and *atomic roles*. Complex descriptions are built inductively with constructors of concepts. We will use the letters A for atomic concepts, R for atomic roles and both C and D for descriptions of concepts. Using the connectives $\bar{0}, \&, \rightarrow$ (falsity, conjunction, implication), the quantifiers $\forall, \exists$ and the point . as an auxiliary symbol, the description of concepts in classic $\mathcal{ALC}$ can be built using the following syntactical rules

$$C, D \rightsquigarrow A \mid \bar{0} \mid C\&D \mid C \rightarrow D \mid \forall R.C \mid \exists R.C$$

Given a continuous t-norm $*$ and a countable subalgebra $\mathbf{S}$ of the corresponding standard algebra $[0,1]_*$, let us consider the logic $L^*(\mathbf{S})$. We define $\mathcal{ALC}_{L^*(\mathbf{S})}$ by adding to $\mathcal{ALC}$, for every $r \in S$, a nullary connective $\bar{r}$ and the rule $C \rightsquigarrow \bar{r}$.

The language $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ is defined by adding to $\mathcal{ALC}_{L^*(\mathbf{S})}$ the connective $\sim$ and the syntactic rule $C \rightsquigarrow \sim C$.

Following [8], the notions of *instance of a concept* and *instance of a role* allow us to read the formulas of both languages as formulas of the corresponding predicate fuzzy logic. For each term t (variable or constant). The *instance* $D(t)$ of a concept D is defined as follows:

$$\begin{array}{ll} A(t), & \text{if } D \text{ is an atomic concept } A,\\ \bar{0}, & \text{if } D = \bar{0},\\ \sim C(t), & \text{if } D = \sim C,\\ C_1(t) \circ C_2(t), & \text{if } D = C_1 \circ C_2, \text{ where } \circ \in \{\&, \rightarrow\},\end{array}$$

and, if y is a variable not occurring in $C(t)$,

$$\begin{array}{ll} (\forall y)(R(t,y) \rightarrow C(y)), & \text{if } D = \forall R.C,\\ (\exists y)(R(t,y)\&C(y)), & \text{if } D = \exists R.C,\end{array}$$

where, given two terms t_1 and t_2, $R(t_1, t_2)$ is an atomic formula corresponding to the atomic role R. We will refer to the expressions of the form $R(t_1, t_2)$ as *instances of the atomic role R*.

Semantics. According to semantics for $L^*(\mathbf{S})\forall$ and $L^*_\sim(\mathbf{S})\forall$ a fuzzy interpretation $\mathbf{M}$ associates a fuzzy set $A^{\mathbf{M}}$ to each atomic concept A and a fuzzy binary relation $R^{\mathbf{M}}$ to each atomic role R, and the truth value for complex descriptions is given as follows:

$$\begin{array}{rcl} \bar{0}^{\mathbf{M}}(a) & = & 0\\ \bar{r}^{\mathbf{M}}(a) & = & r, \text{ for every } r \in S\\ (C\&D)^{\mathbf{M}}(a) & = & C^{\mathbf{M}}(a) * D^{\mathbf{M}}(a)\\ (C \rightarrow D)^{\mathbf{M}}(a) & = & C^{\mathbf{M}}(a) \rightarrow_* D^{\mathbf{M}}(a)\\ (\forall R.C)^{\mathbf{M}}(a) & = & \inf\{R^{\mathbf{M}}(a,b) \rightarrow_* C^{\mathbf{M}}(b) : b \in M\}\\ (\exists R.C)^{\mathbf{M}}(a) & = & \sup\{R^{\mathbf{M}}(a,b) * C^{\mathbf{M}}(b) : b \in M\}\end{array}$$

In the case of complex descriptions in $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$ we must to add the following:

$$(\sim C)^{\mathbf{M}}(a) \quad = \quad 1 - C^{\mathbf{M}}(a)$$

Fuzzy TBox and Fuzzy ABox. Now we define the notions of $TBox$ and $ABox$ for $\mathcal{ALC}_{L^*(\mathbf{S})}$ and $\mathcal{ALC}_{L^*_\sim(\mathbf{S})}$. In these definitions we use the following graded notion of inclusion between fuzzy sets: $\textit{degree}(C \subseteq D) = \inf_x(C(x) \rightarrow_* D(x))$. Of course this degree is 1 if and only if $C(x) \leq D(x)$ for all x and 0 if the support[1] of the two fuzzy sets are disjoint. Having the truth constants in the language allows us to associate sentences like "*degree* $(C \subseteq D) \leq r$" with formulas such as $(\forall x)(C(x) \rightarrow D(x)) \rightarrow \bar{r}$.

A *fuzzy concept inclusion axiom* is a sentence of one of the following forms:

- $\bar{r} \rightarrow (\forall x)(C(x) \rightarrow D(x))$

[1] The *support* of a fuzzy set is the cardinal of the set of elements which membership degree is greater than 0.

ISBN: 978-989-95079-6-8

Table 1: The graded notation for fuzzy KB.

Evaluated Formula	FDL Graded Notation
$\bar{r} \rightarrow (\forall x)(C(x) \rightarrow D(x))$	$\langle C \sqsubseteq D, \succcurlyeq \bar{r} \rangle$
$(\forall x)(C(x) \rightarrow D(x)) \rightarrow \bar{r}$	$\langle C \sqsubseteq D, \preccurlyeq \bar{r} \rangle$
$\bar{r} \leftrightarrow (\forall x)(C(x) \rightarrow D(x))$	$\langle C \sqsubseteq D, \approx \bar{r} \rangle$
$\bar{r} \rightarrow C(a)$	$\langle a : C, \succcurlyeq \bar{r} \rangle$
$C(a) \rightarrow \bar{r}$	$\langle a : C, \preccurlyeq \bar{r} \rangle$
$\bar{r} \leftrightarrow C(a)$	$\langle a : C, \approx \bar{r} \rangle$
$\bar{r} \rightarrow R(a,b)$	$\langle (a,b) : R, \succcurlyeq \bar{r} \rangle$
$R(a,b) \rightarrow \bar{r}$	$\langle (a,b) : R, \preccurlyeq \bar{r} \rangle$
$\bar{r} \leftrightarrow R(a,b)$	$\langle (a,b) : R, \approx \bar{r} \rangle$

- $(\forall x)(C(x) \rightarrow D(x)) \rightarrow \bar{r}$
- $\bar{r} \leftrightarrow (\forall x)(C(x) \rightarrow D(x))$

A *fuzzy assertion axiom* is a sentence of one of the following forms:

- $\bar{r} \rightarrow C(a)$ or $\bar{r} \rightarrow R(a,b)$
- $C(a) \rightarrow \bar{r}$ or $R(a,b) \rightarrow \bar{r}$
- $C(a) \leftrightarrow \bar{r}$ or $R(a,b) \leftrightarrow \bar{r}$

Now a *fuzzy TBox* is defined as a finite set of *fuzzy concept inclusion axioms* while a *fuzzy ABox* is defined as a finite set of *fuzzy assertion axioms*. A *fuzzy KB* is a pair $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, where the first component is a *fuzzy TBox* and the second one is a *fuzzy ABox*.

Notice that all the axioms of the *fuzzy KB* are evaluated formulas. Thus, the syntactic notion of *fuzzy KB* according to our approach, both the $TBox$ and the $ABox$ can be seen as theories of the logic $L^*(\mathbf{S})\forall$ (or $L^*_{\sim}(\mathbf{S})\forall$).

In Tab. 3 we present an alternative graded notation for fuzzy inclusion and fuzzy assertions that we will use in the example. This notation is similar to the one used in some papers of FDLs (see for instance [17]). Moreover this notation is according to the semantical interpretation in the sense that, for instance,

$$\langle \sigma, \succcurlyeq \bar{r} \rangle^{\mathbf{M}} = r \rightarrow_* \sigma^{\mathbf{M}}$$

and so

$$\mathbf{M} \models \langle \sigma, \succcurlyeq \bar{r} \rangle \quad \text{iff} \quad r \rightarrow_* \sigma^{\mathbf{M}} = 1 \quad \text{iff} \quad \sigma^{\mathbf{M}} \geq r$$

It is interesting to remark that in $\mathcal{ALC}_{L^*_{\sim}(\mathbf{S})\forall}$ the involutive negation allows us to define graded expressions like, for instance, $\langle a : C, \succ \bar{r} \rangle$ as $\sim \langle a : C, \preccurlyeq \bar{r} \rangle$ which corresponds to the formula $\sim (C(a) \rightarrow \bar{r})$.

An example. We will use a data set composed of nine robots (Fig. 1), each one with either the same of different shape of head and body (i.e., they are homogeneous or not homogeneous respectively), they can or cannot wear a tie, they can or cannot smile, and they hold some object. Taking into account all these characteristics, robots can have different *friendliness* degree. The domain of interpretation of the robots is the set: $M_{\mathcal{R}} = \{r_i : 1 \leq i \leq 9\} \cup \{o_i : 1 \leq i \leq 9\}$, where the r_i are the robots and each o_i is the object that the robot r_i holds (e.g., the object o_4 is the flower that r_4 holds). Atomic concepts of the language are the following: Robot, Happy, Object, FriendlyObject, Homogeneous, Balloon, Flag, Flower, Sword, Ax and HasTie. There is only one atomic role: hasObject.

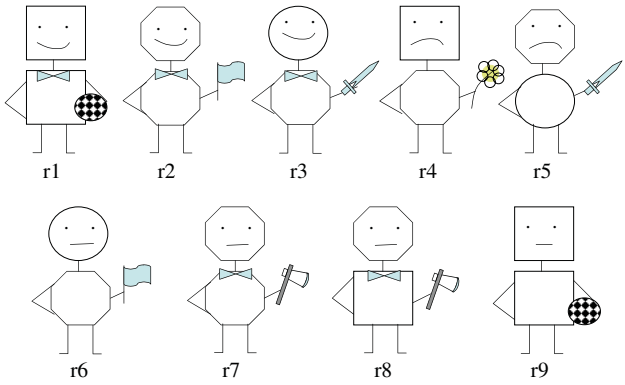


Figure 1: The 9 little robots.

The $TBox$ concerning the robots domain is the following:

Friendly $\equiv$ Robot&(∃hasObject.FriendlyObject)&(Happy $\veebar$ Homogeneous)
$\langle$Robot & Object $\sqsubseteq \bar{0}, \approx \bar{1}\rangle$
$\langle \bar{1} \sqsubseteq$ Robot $\veebar$ Object, $\approx \bar{1}\rangle$
$\langle$Flower $\sqsubseteq$ FriendlyObject, $\approx \bar{1}\rangle$
$\langle$Balloon $\sqsubseteq$ FriendlyObject, $\approx \overline{0.75}\rangle$
$\langle$Flag $\sqsubseteq$ FriendlyObject, $\approx \overline{0.50}\rangle$
$\langle$Sword $\sqsubseteq$ FriendlyObject, $\approx \overline{0.25}\rangle$
$\langle$Ax $\sqsubseteq$ FriendlyObject, $\approx \bar{0}\rangle$

where $C \veebar D$ is an abbreviation of $\sim (\sim C \& \sim D)$ and $C \equiv D$ is an abbreviation for the conjunction of the formulas $\langle C \sqsubseteq D, \approx \bar{1} \rangle$ and $\langle D \sqsubseteq C, \approx \bar{1} \rangle$.

Notice that objects have different friendliness degree. For instance, a sword is a friendly object with degree 0.25 and an ax is a friendly object with degree 0 (i.e., it should be considered unfriendly in the classical case). On the other hand, the $TBox$ also contains a definition of Friendly allowing to assess the friendliness degree of a robot.

The $ABox$ containing the descriptions of the robots is the following:

For each i, $1 \leq i \leq 9$, $\langle r_i$: Robot, $\approx \bar{1}\rangle$, $\langle (r_i, o_i)$: hasObject, $\approx \bar{1}\rangle$
$\langle r_1$: Homogeneous, $\approx \bar{1}\rangle$, $\langle o_1$: Balloon, $\approx \bar{1}\rangle$, $\langle r_1$: Happy, $\approx \bar{1}\rangle$
$\langle r_2$: Homogeneous, $\approx \bar{1}\rangle$, $\langle o_2$: Flag, $\approx \bar{1}\rangle$, $\langle r_2$: Happy, $\approx \bar{1}\rangle$
$\langle r_3$: Homogeneous, $\approx \overline{0.75}\rangle$, $\langle o_3$: Sword, $\approx \bar{1}\rangle$, $\langle r_3$: Happy, $\approx \bar{1}\rangle$
$\langle r_4$: Homogeneous, $\approx \overline{0.50}\rangle$, $\langle o_4$: Flower, $\approx \bar{1}\rangle$, $\langle r_4$: Happy, $\approx \bar{0}\rangle$
$\langle r_5$: Homogeneous, $\approx \overline{0.50}\rangle$, $\langle o_5$: Sword, $\approx \bar{1}\rangle$, $\langle r_5$: Happy, $\approx \bar{0}\rangle$
$\langle r_6$: Homogeneous, $\approx \overline{0.75}\rangle$, $\langle o_6$: Flag, $\approx \bar{1}\rangle$, $\langle r_6$: Happy, $\approx \overline{0.50}\rangle$
$\langle r_7$: Homogeneous, $\approx \bar{1}\rangle$, $\langle o_7$: Ax, $\approx \bar{1}\rangle$, $\langle r_7$: Happy, $\approx \overline{0.50}\rangle$
$\langle r_8$: Homogeneous, $\approx \overline{0.75}\rangle$, $\langle o_8$: Ax, $\approx \bar{1}\rangle$, $\langle r_8$: Happy, $\approx \overline{0.50}\rangle$
$\langle r_9$: Homogeneous, $\approx \bar{1}\rangle$, $\langle o_9$: Balloon, $\approx \bar{1}\rangle$, $\langle r_9$: Happy, $\approx \overline{0.50}\rangle$

Notice that, using truth constants, we can assess different degrees of homogeneity according to the shape of both head and body. In particular, we assess, in a subjective way, that a combination of round shapes of head and body (i.e., a circle and an octagon) give a more homogeneous aspect to the robot than combining round and square shapes. Thus, robots r_6 and r_8 are considered more homogeneous than robot r_4. Similarly, robots have different form of mouth that give them different degree of happiness (i.e., robot r_1 is assessed as more happy than robots r_8 and r_4).

Fuzzy reasoning. Reasoning in fuzzy description logics consists on the same kind of tasks than in the classical case but

ISBN: 978-989-95079-6-8

now they depend on the chosen continuous t-norm $*$. One of the advantages of introducing truth constants in the language is the possibility to define the graded versions of the notions of $*$-satisfiability, $*$-subsumption and $*$-validity defined in [9] without modifying the semantics. Thus, given a concept C, a t-norm $*$ and a truth value $r \in S$ we introduce the following graded notions with respect to a knowledge base $\mathcal{K}$:

• C is *$*$-satisfiable in a degree greater or equal than r* iff there is a model $\mathbf{M}$ of $\mathcal{K}$ such that $\|\bar{r} \to C(a)\|^*_{\mathbf{M}} = 1$ (being a a constant).

• C is *$*$-valid in a degree greater or equal than r* in a model $\mathbf{M}$ of $\mathcal{K}$ iff $\|\bar{r} \to (\forall x)C(x)\|^*_{\mathbf{M}} = 1$.

• C is *$*$-subsumed* by D *in a degree greater or equal than* r in a model $\mathbf{M}$ of $\mathcal{K}$ iff the concept $C \to D$ is *$*$-valid in a degree greater or equal than r* in the model $\mathbf{M}$, that is, iff $\|\bar{r} \to (\forall x)(C(x) \to D(x))\|^*_{\mathbf{M}} = 1$.

On the other hand, we can analogously define the notions of lower thresholds. For instance, a concept C is *$*$-satisfiable in a degree lower or equal than r* iff $\|C(a) \to \bar{r}\|^*_{\mathbf{M}} = 1$ for some model $\mathbf{M}$ of $\mathcal{K}$. Moreover, it is also possible to define an interval where a concept is either valid, satisfiable or subsumed. For instance a concept C is *$*$-satisfiable in an interval of degrees* $[r, s]$ iff $\|\bar{r} \to C(a)\|^*_{\mathbf{M}} = 1$ and $\|C(a) \to \bar{s}\|^*_{\mathbf{M}} = 1$ for some model $\mathbf{M}$. In particular, when $r = s$ the interval became a value called the *degree of satisfiability*.

Example 3.1 *The concept C below is $*$-satisfiable with degree 0.75 in the robots model using any continuous t-norm.*

$$C := \mathsf{Homogeneous} \,\&\exists \mathsf{hasObject}.\mathsf{FriendlyObject}$$

According to the semantics, C is $*$-satisfiable with degree 0.75 if there is at least one robot such that

$C(x) = \mathsf{Homogeneous}(x) * \sup_{y \in M_{\mathcal{R}}}(\mathit{hasObject}(x, y) * \mathit{FriendlyObject}(y)) = 0.75$

The equality is true since the robots r_1 and r_9 have $\mathsf{Homogeneous}(x) = 1$ and both hold a ballon that, as the $TBox$ states, it is a friendly object with degree 0.75. Thus, because 1 is the unity element of any t-norm, both robots have friendliness degree 0.75.

Example 3.2 *Let us to analyze the $*$-subsumption degree of the concept* Object & Sword & FriendlyObject *by* $\bar{0}$ *with respect to the $TBox$.*

According to the definition of $*$-subsumption, we have to analyze the following formula

$$\bar{r} \leftrightarrow (\forall x)((\mathit{Object} \,\&\, \mathit{Sword} \,\&\, \mathit{FriendlyObject})(x) \to \bar{0})$$

and thus, according to the interpretation, the $*$-subsumption has degree

$$r = \inf_{x \in M_{\mathcal{R}}} \{\mathit{Object}(x) * \mathit{Sword}(x) * \mathit{FriendlyObject}(x) \to_* 0\}$$

For all t-norms: a) when x is an element that is neither an object nor a sword, then r is $0 \to_* 0 = 1$, and b) when x is a sword, $r = \mathit{FriendlyObject}(x) \to_* 0$. In such case r depends on the t-norm. For instance, taking the Łukasiewicz's t-norm, $r = \min(1, 1 - \mathit{FriendlyObject}(x) + 0) = 0.75$. Taking either the minimum or the product t-norms $r = 0.25 \to_* 0 = 0$.

Concerning the entailment, we say that a fuzzy assertion α is *$*$-entailed* by the knowledge base $\mathcal{K}$ if every model $\mathbf{M}$ of $\mathcal{K}$ also satisfies α. For instance, the reader can easily prove that the fuzzy $ABox$ of the robots' example $*$-entails the assertion $\langle \mathsf{Friendly}(r_6), \succcurlyeq 0.50\rangle$ with respect to the fuzzy $TBox$ using either Łukasiewicz or minimum t-norms. However, using the product t-norm the above $*$-entailment is not satisfied.

4 A case study: Fuzzy description logics associated to $G_\sim\forall(S)$

The first FDL systems were related to the initial Zadeh proposal for fuzzy sets operations. The logic underlying this proposals is the logic associated to the calculus over $[0, 1]$ defined by the functions $\min$, $\max$, $n(x) = 1 - x$, the Kleene-Dienes implication $x \to y = \max(1 - x, y)$, and quantifiers interpreted as in section 2.2. This is referred as *minimalistic* apparatus of fuzzy logic by Hájek in [9]. But, as already noted by Hájek, the (non residuated) implication of this logic has some no nice behaviour. For example: (a) in this logic $\varphi \to \varphi$ is not a tautology and (b) an implication $\varphi \to \psi$ is evaluated as 1 only if either φ is evaluated as 0 or ψ is evaluated as 1. The languages $\mathcal{ALC}$ based on this logic was studied by Straccia in [14]. The notion of subsumption is defined using the notion of fuzzy subsets ($C \subseteq D$ if and only if $C(x) \leq D(x)$ for all x in the univers) with no relation to implication function due to the lack of a good relation between the implication function of this logic and the order relation in $[0, 1]$. Moreover it is not known any (Hilbert style) axiomatization of this logic.

For these reasons in this short case study we propose the interest of studying the $\mathcal{ALC}$ language associated to Gödel Logic with truth constants $G(\mathbf{S})\forall$ and specially the $\mathcal{ALC}$ language associated to its expansion with an involutive negation $G_\sim(\mathbf{S})\forall$. In [7] the authors prove the canonical completenes of the logic $G(\mathbf{S})\forall$. The canonical completeness can also be proved for $G_\sim(\mathbf{S})\forall$.

Theorem 4.1 *The logic $G_\sim(\mathbf{S})\forall$ has the Canonical Finite Strong completeness, i.e., for every finite set of formulas $\Gamma \cup \{\varphi\}$,*

$$\Gamma \vdash_{G_\sim(\mathbf{S})\forall} \varphi \quad \textit{iff} \quad \Gamma \models_{[0,1]_{G_\sim(\mathbf{S})}} \varphi$$

where
$[0, 1]_{G_\sim(\mathbf{S})} = \langle [0, 1], \min, \max, \to_G, n, \{r \mid r \in S\}, 0, 1\rangle$, *with $n(x) = 1 - x$ being the truth function associated to the involutive negation.*

The proof is a simplified version of [7, Theorem 11]. As a consequence of this theorem, we have an equivalence between $*$-entailment in $\mathcal{ALC}_{G_\sim(\mathbf{S})}$ and the semantics consequence relation on $[0, 1]_{G_\sim(\mathbf{S})}$.

Notice that in $G_\sim(\mathbf{S})\forall$ we have, in addition to the connectives of this logic, the ones being the counterpart of the minimalistic logic, since an implication having as truth function the Kleene-Dienes implication is definable as $\sim \varphi \vee \psi$. Moreover the truth constants allow us to have evaluated formulas and their negation by the involutive negation that makes possible to represent by formulas the semantical expressions saying that the interpretation of a formula is greater, greater or equal, less, and less or equal than a value r.

ISBN: 978-989-95079-6-8

Thus the description language based on $G_\sim\forall(S)$ seems to be a good choice because both it is actually very expressive and it maintains the canonical completeness. In this setting we have two implication connectives, that semantically correspond to the residuum of the minimum plus the Kleene-Dienes implication. Of course, as in the general case of $L^*_\sim(\mathbf{S})\forall$, the degree of subsumption between concepts is defined by means of the residuated implication, i.e, the Gödel implication. The Straccia's work [14] and the Bobillo's work [2] gives algorithms for proving satisfiability and subsumption in the $\mathcal{ALC}$ description language based on the minimalistic logic and in the $\mathcal{SROIQ}(D)$ description language based on Gödel logic respectively. As future work we want to find analogous algorithms for the $\mathcal{ALC}$ description languages based on $G_\sim(\mathbf{S})\forall$.

5 Conclusions

This paper is a first step on the direction proposed by Hájek concerning the relationships between some proposals of FDLs and the recent developments in mathematical fuzzy logics. The main contributions of our approach is the use of truth constants in the language of description and the introduction of an involutive negation in the required cases. This allows us to recover graded notions of satisfiability, validity and subsumption that have been used in the fuzzy logic setting. This choice is oriented to search for the syntactical counterpart of the semantic calculi proposed in some works dealing with FDLs.

Acknowledgments

This work has been supported by the Spanish projects MULOG2 TIN2007-68005-C04-01/04, "Agreement Technologies" (CONSOLIDER CSD 2007-0022, INGENIO 2010), and the MCYT-FEDER Project MID-CBR (TIN2006-15140-C03-01), the CSIC grant 200750I005, and the Generalitat de Catalunya grant 2005-SGr-00093. The authors thank Félix Bou for his assistance in the improvement of this paper.

References

[1] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi and P. F. Patel-Schneider, editors. The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, 2007.

[2] F. Bobillo. Managing vagueness in Ontologies *PhD disertation, Univ. Granada* , 2008.

[3] R. Cignoli, F. Esteva, L. Godo, and A. Torrens. Basic fuzzy logic is the logic of continuous t-norms and their residua. *Soft Computing*, 4(2):106–112, 2000.

[4] F. Esteva, J. Gispert, L. Godo, and C. Noguera. Adding truth-constants to logics of continuous t-norms: Axiomatization and completeness results. *Fuzzy Sets and Systems*, 158(6):597–618, 2007.

[5] F. Esteva, L. Godo, P. Hájek, and M. Navara. Residuated fuzzy logics with an involutive negation. *Archive for Mathematical Logic*, 39(2):103–124, 2000.

[6] F. Esteva, L. Godo, and F. Montagna. Equational characterization of the subvarieties of BL generated by t-norm algebras. *Studia Logica* 76(2): 161–200, 2004.

[7] F. Esteva, L. Godo, and C. Noguera. On completeness results for predicate Łukasiewicz, product, Gödel and nilpotent minimum logics expanded with truth-constants. *Mathware & Soft Computing*, 14(3):223–246, 2007.

[8] P. Hájek. *Metamathematics of fuzzy logic*, volume 4 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 1998.

[9] P. Hájek. Making fuzzy description logic more general. *Fuzzy Sets and Systems*, 154(1):1–15, 2005.

[10] P. Hájek. What does mathematical fuzzy logic offer to description logic? In Elie Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, chapter 5, pages 91–100. Elsevier, 2006.

[11] A. A. Monteiro. Sur les algèbres de Heyting symmétriques. *Port. Math.*, 39: 1–237, 1980.

[12] J. Pavelka. On fuzzy logic. I, II, III. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 25(1):45–52, 25(2):119–134, 25(5):447–464,1979.

[13] G. Stoilos, G. Stamou, J.Z. Pan, V. Tzouvaras, and I. Horrocks. Reasoning with very expressive fuzzy description logics. *Journal of Artificial Intelligence Research*, 30(8):273–320, 2007.

[14] U. Straccia. Reasoning within fuzzy description logics. *Journal of Artificial Intelligence Research*, 14:137–166, 2001.

[15] U. Straccia. Fuzzy alc with fuzzy concrete domains. In *Proceeedings of the International Workshop on Description Logics (DL-05)*, pages 96–103, Edinburgh, Scotland, 2005. CEUR.

[16] U. Straccia. A fuzzy description logic for the semantic web. In Elie Sanchez, editor, *Fuzzy Logic and the Semantic Web*, Capturing Intelligence, chapter 4, pages 73–90. Elsevier, 2006.

[17] U. Straccia and F. Bobillo. Mixed integer programming, general concept inclusions and fuzzy description logics. *Mathware & Soft Computing*, 14(3):247–259, 2007.

[18] C. B. Tresp and R. Molitor. A description logic for vague knowledge. Technical Report RWTH-LTCS Report 98-01, Aachen University of Technology, 1998.

[19] J. Yen. Generalizing term subsumption languages to fuzzy logic. In *Proc. of the 12th IJCAI*, pages 472–477, Sidney, Australia, 1991.

[20] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[21] L. A. Zadeh. Fuzzy logic and approximate reasoning. *Synthese*, 30:407–428, 1975.

ISBN: 978-989-95079-6-8

Evaluating premises, partial consequences and partial hypotheses *

I. García-Honrado, E. Trillas, S. Guadarrama and E. Renedo

European Centre for Soft Computing
Mieres (Asturias), Spain
Email: {itziar.garcia, enric.trillas, sergio.guadarrama, eloy.renedo}@softcomputing.es

Abstract— *To evaluate premises, consequences and hypotheses, on this paper relevance and support ratios are defined for each of them. This allows to distinguish consequences based on the number of premises that support them, and also to reduce the set of premises while maintaining the same consequences. Since the relation between premises and hypotheses is, in some sense, similar to the relation between consequences and premises, analogous ratios are defined for hypotheses and premises.*

Keywords— Conjectures, Consequences, Hypotheses, Relevance, Support.

1 Introduction

The aim of most problems is to make choice between possible solutions, a clear example is a medical diagnosis problem. In this paper we allocate degrees for the elements in the set of consequences, hypotheses or premises in order to choose the consequence, hypothesis or premise with the biggest degree. Papers [2] and [3] had dealt with that idea of graded consequences.

The following section will show Conjectures, Hypotheses, and Consequences (CHC) models introduced in [7], which was suggested, in part by Watanabe in [10], and takes into account the particular case of partial consequence's operator [8] [9]. Partial operators of consequences are that allow to get consequences of each premise, or subset of premises, and obtaining the final set of consequences as the union of all these partial consequences.

To define support for each consequence (section 3) we consider that consequences with bigger support, are those that are supported by more premises or subsets of premises. By the way, different degrees are allocated for consequences. So, for example, following with a medical diagnosis problem in which the premises are diseases and the consequences are symptoms, we can choose between consequences, and select as the stronger, the one with biggest support.

In section 4, we deal with a measure of relevance for premises that is useful for knowing which premises have more importance, in the sense of how many consequences can be deduced from them. Thanks to that measure, the set of premises can be reduced to a smaller set with the same relevance. This reduced set of premises gets rid of superfluous premises and yet allows to work with less premises, while getting the same set of consequences. Till now all premises seemed to have the same importance.

Finally, in section 5, we also consider partial hypotheses, that is, hypotheses of one premise, and not hypotheses of all premises. And analogous measure of support for premises, as well as of relevance for partial hypotheses, are defined. This allows to evaluate subset of partial hypotheses by counting how many premises they give as consequences.

2 Basic concepts

2.1 CHC models

Reasoning can be understood as a process allowing to get conjectures from a set of premises, P. There are three basic types of reasoning: deduction, abduction and induction. A process that allows to get consequences is a deductive reasoning, a process that allows to get hypotheses is an abductive reasoning, and finally, if the process allows to get speculations, it is an speculative reasoning.

In this paper, CHC models are defined on a preordered set $(L, \leq)$, endowed with a negation, $'$. And when it is necessary, the preordered set will be endowed with an infimum, $\cdot$, and a supremum, $+$, operations $(L, \leq, ', \cdot, +)$, a preorder with infimum and supremum operations is a partial ordered set (poset)with these operations. The infimum of L is called first element and it is denoted by 0, the supremum of L is called the last element and it is denoted by 1. The paper only deals with finite algebraic structures, that is, with a finite set L.

CHC models can be based on consequences operators [8, 9],

Definition 2.1 *If L is a set, and $\mathfrak{F} \subset \mathbb{P}(L)$, it is said that $(L, \mathfrak{F}, C)$ is a* structure of consequences*, provided that $C : \mathfrak{F} \to \mathfrak{F}$ verifies,*

1. *$P \subset C(P)$, for all $P \in \mathfrak{F}$ (C is* extensive*)*
2. *If $P \subset Q$, then $C(P) \subset C(Q)$, for all $P, Q \in \mathfrak{F}$ (C is* monotonic*)*
3. *$C(C(P)) = C(P)$, or $C^2 = C$, for all $P \in \mathfrak{F}$ (C is a* clausure*)*

i.e. C is an operator of consequences *(in the sense of Tarski) for $\mathfrak{F}$ in L.*

For each $\{q\} \in \mathfrak{F}$, let us write $C(q) = C(\{q\})$.

*This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias, Spain), and CICYT (Spain) under project TIN2008-06890-C02-01

ISBN: 978-989-95079-6-8

Definition 2.2 *A consequences' operator C is consistent in P, if for all $q \in C(P)$, $q' \notin C(P)$.*
A structure of consequences $(L, \mathfrak{F}, C)$ is consistent if C is consistent for all $P \in \mathfrak{F}$.

Let P be the set of premises, $P \neq \emptyset$, and $C(P)$ a set of consequences for P. Conjectures of P can be defined from a consistent consequences operator C, as those elements whose negation is not in $C(P)$, $Conj_C(P) = \{q \in L; q' \notin C(P)\}$. Hypothesis can be defined by $Hyp_C(P) = \{h \in Conj_C(P) - C(P); \{h\} \in \mathfrak{F},\ P \subset C(h)\}$.

Finally, speculations are those conjectures that are neither consequences, nor hypotheses. Hence, $Sp_C(P) = Conj_C(P) - [C(P) \cup Hyp_C(P)] = \{q \in L, q \notin C(P), q' \notin C(P), q \notin Hyp_C(P)\}$.

2.2

Definition 2.3 *A consequences' operation C is a partial consequences operator if $C(P) = \bigcup_{R \subset P, R \in \mathfrak{F}} C(R)$.*

Definition 2.4 *A decomposable consequences operator is a consequences' operator such that $C(P) = \bigcup_{p \in P} C(R)$.*

This paper considers *partial hypothesis*, elements that are hypotheses of a subset of the set of premises P. This idea comes from that of *partial consequences*.

Definition 2.5 *For each set P of premises the partial hypotheses set is,*

$$Hyp^*_C(P) = \{h \in \{L - 0\} - P; \{h\} \in \mathfrak{F}, \exists R \subset P,$$
$$R \subset C(h)\}.$$

Obviously, hypotheses are partial hypotheses, since $P \subset P$ and $P \subset C(h)$, provided h is a hypothesis.

Remark 2.6 *Although hypotheses are anti-monotonic ($P_1 \subset P_2$, implies $Hyp(P_2) \subset Hyp(P_1)$), partial hypotheses are (as it is easy to prove) monotonic ($P_1 \subset P_2$, implies $Hyp^*(P_1) \subset Hyp^*(P_2)$). That is why they can not be considered classical hypotheses.*

2.3

The paper deals with a general concept of measure [6], defined in a preordered set $(L, \leq)$. A measure is a mapping $m : L \to [0, 1]$, such that:

- There exists a minimal $x_0 \in L$, such that $m(x_0) = 0$
- There exists a maximal $x_1 \in L$, such that $m(x_1) = 1$
- If $x \leq y$, then $m(x) \leq m(y)$.

3 Consequences support

This section introduces a ratio in order to distinguish which consequences are the more supported by a given set of premises. And proof in which cases is a measure

Let's recall that in this paper L is assumed to be a finite set.

Definition 3.1 *The support of $q \in L$ is the ratio of subsets of premises that allow getting q as a consequence, to all possible subsets of premises.*

$$Supp_{C,P}(q) = \frac{|\{R \in \mathbb{P}(P); q \in C(R)\}|}{2^{|P|} - 1} = \quad (1)$$
$$\frac{|\{R \in \mathbb{P}(P); q \in C(R)\}|}{|\mathbb{P}(P) - \emptyset|}.$$

Since $P \neq \emptyset$, it is $|P| > 0$ and the quotient in the definition is possible.

The bigger support a consequence has, the more subsets of premises allow deducing it.
Notice that if $q \notin C(P), Supp_{C,P}(q) = 0$, since if there were $R \in \mathbb{P}(P)$ such that $q \in C(R)$, because of the monotonicity of the consequence operator, $C(R) \subset C(P)$ would imply $q \in C(P)$.

This ratio verifies the following properties,

- If $P \subset Q$, it is $Supp_{C,P}(q) \leq Supp_{C,Q}(q)$, for all $q \in L$.
- If $P = \{p\}, \forall q \in C(P)$, it is $Supp_{C,P}(q) = 1$.
- For all $q \in C(P)$, $Supp_{C,P}(q) > 0$.
- $Supp_{C,P}(q) = 1$ means that q is a consequence for all $R \in \mathbb{P}(P)$. Particularly, q is consequence of all $p \in P$.

The support defined by (1), is not a measure in general. For example, if $C(P) = P,\ \forall P \in \mathbb{P}(L)$, let P be a set with more than one element. If $q \in L$, it is either $q \in P$, or $Supp(q) = 0$. If $q \in P$, there exists $p \in P$ such that $p \neq q$, and $q \notin C(p)$ and $Supp(q) \neq 1$. Therefore, there is no $q \in L$ such that $Supp_{C,P}(q) = 1$.

Remark 3.2 *$Supp_{C,P}$ is monotonic with respect to the preorder given by C, $q_1 \leq_C q_2$ iff $q_2 \in C(q_1)$ [1],*

Proof. Since if $q_1 \leq_C q_2$, for each R such that $q_1 \in C(R)$, it is $C(q_1) \in C(C(R)) = C(R)$, and, as $q_2 \in C(q_1)$, it is also $q_2 \in C(R)$. So, $Supp_{C,P}(q_1) \leq Supp_{C,P}(q_2)$.
□

Since, given P, the relation defined between the pairs of elements in L with the same value of $Supp_{C,P}$, is an equivalence, the classes

$$[q] = \{v \in L;\ Supp_{C,P}(v) = Supp_{C,P}(q)\}$$

give a partition on L in a number of parts that is, at most, $2^{|P|}$.

3.1 The case of the operator $C_\bullet$

Let $(L, \leq)$ be now a preordered set in which is defined an infimum operation denoted by '$\cdot$'.

The partial consequences operator $C_\bullet$ gives consequences that are consequences for a subset of the set of premises P, it is $C_\bullet(P) = \{q \in L; \exists \{p_1, p_2, ..., p_n\} \subset P : p_1 \cdot p_2 \cdot ... \cdot p_n \leq q\}$. It is a partial consequences' operator and it obviously verifies $C_\bullet(P) = \bigcup_{R \subset P, R \in \mathfrak{F}} C_\bullet(R)$.

ISBN: 978-989-95079-6-8

Notice, that as L is finite, P is also finite and $C_\bullet(P) = \{q \in L; \inf P \leq q\}$, which equal to the infimum operator of consequences $C_\wedge$.

$$Supp_{C_\bullet,P}(q) = \frac{|\{R \in \mathbb{P}(P) \in P; \inf(R) \leq q\}|}{2^{|P|}-1} \quad (2)$$

Let's see what specific properties are verified by $Supp_{C_\bullet,P}$,

- If L has last element, it implies $1 \in L$, then $1 \in C_\bullet(P)$ and $Supp_{C_\bullet,P}(1) = 1$.
- If $q_1 \leq q_2$, then $Supp_{C_\bullet,P}(q_1) \leq Supp_{C_\bullet,P}(q_2)$. That is the function $Supp_{C_\bullet,P}$ is monotonic.
- $Supp_{C_\bullet,P}(\sup\{q_1,q_2\}) \geq \max\{Supp_{C_\bullet,P}(q_1), Supp_{C_\bullet,P}(q_2)\}$, provided $\sup\{q_1,q_2\}$ exists.
- $Supp_{C_\bullet,P}(\inf\{q_1,q_2\}) \leq \min\{Supp_{C_\bullet,P}(q_1), Supp_{C_\bullet,P}(q_2)\}$

Corollary 3.3 *Let $(L, \leq, ', \cdot, +)$ be a partial ordered set with infimum and supremum operations and first and last elements. If $P \neq \{0\}$, the function $Supp_{C_\bullet,P} : L \to [0,1]$ is a measure.*

Proof. It is monotonic, and it verifies the boundary conditions, since 0 is not a consequence $Supp_{C_\bullet,P}(0) = 0$ and $Supp_{C_\bullet,P}(1) = 1$. □

3.2 The case of the operator $C_\leq$

$C_\leq$ is the partial consequences operator that gives as consequences those elements that follow from at least one premise in P, formally, it is $C_\leq(P) = \{q \in L; \exists p \in P : p \leq q\}$, see [8]. Hence, it can be considered as a decomposable consequences' operator, since allows getting consequences that are not deduced from all premises. It is straightforward that $C_\leq(P) = \bigcup_{R \subset P, R \in \mathfrak{F}} C(R) = \bigcup_{p \in P} C_\leq(p)$.

In this case, a different definition of the support's ratio seems to be more convenient, since it deals only with consequences of each $p \in P$, nor with consequences of each subset of $\mathbb{P}(P)$.

Definition 3.4 *The support of $q \in C_\leq(P)$ is the ratio of premises that allow getting q as consequence to all premises.*

$$\widehat{Supp}_{C_\leq,P}(q) = \frac{|\{p \in P; p \leq q\}|}{|P|} \quad (3)$$

Since $P \neq \emptyset$, it is $|P| > 0$ and the quotient in the definition is possible.

If $q \notin C_\leq$, then $\widehat{Supp}_{C_\leq,P}(q) = 0$.

So, the bigger Support an element has, the more premises allow to reach it.

$\widehat{Supp}_{C_\leq,P}$ verifies,

- If $P = \{p\}, \forall q \in C_\leq(p)$, it is $\widehat{Supp}_{C_\leq,P}(q) = 1$.
- If L has last element, 1, then $1 \in C_\leq(P)$ and $\widehat{Supp}_{C_\leq,P}(1) = 1$.
- For no $q \in C_\leq(P)$ is $\widehat{Supp}_P(q) = 0$. That is, for all $q \in C_\leq(P)$, $\widehat{Supp}_{C_\leq,P}(q) > 0$.
- $\widehat{Supp}_{C_\leq,P}(q) = 1$ means that q is a consequence for all $p \in P$.
- If $q_1 \leq q_2$, then $\widehat{Supp}_{C_\leq,P}(q_1) \leq \widehat{Supp}_{C_\leq,P}(q_2)$. That is the function $\widehat{Supp}_{C_\leq,P}$ is monotonic with respect to $\leq$.

Remark 3.5 *In order to know what happens if we calculate the support for the infimum or supremum, of two consequences, provided it exists and it is a consequence, weak boundaries are found,*

- $\widehat{Supp}_{C_\leq,P}(\sup\{q_1,q_2\}) \geq \max\{\widehat{Supp}_{C_\leq,P}(q_1), \widehat{Supp}_{C_\leq,P}(q_2)\}$
- $\widehat{Supp}_{C_\leq,P}(\inf\{q_1,q_2\}) \leq \min\{\widehat{Supp}_{C_\leq,P}(q_1), \widehat{Supp}_{C_\leq,P}(q_2)\}$

Obviously, if the operator is consistent, that is, if $q \in C_\leq(P)$, then $q' \notin C_\leq(P)$, it follows $\widehat{Supp}_{C_\leq,P}(q') = 0$.

Theorem 3.6 *Let $(L, \leq, ', \cdot, +)$ be a partial ordered set with first and last elements and $P \neq \{0\}$. Function $\widehat{Supp}_{C_\leq,P} : L \to [0,1]$ is a measure.*

Proof. It is monotonic as it is stated above. Since 0 is not a consequence $\widehat{Supp}_{C_\leq,P}(0) = 0$. Finally, it is obvious that $1 \in C_\leq(P)$ and $\widehat{Supp}_{C_\leq,P}(1) = 1$. □

From $\widehat{Supp}_{C_\leq,P}$ we can calculate $Supp_{C_\leq,P}$. If $\widehat{Supp}_{C_\leq,P}(q) = k$ and $|P| = n$, it is $|\{R \in \mathbb{P}(P); q \in C(R)\}| = k \cdot n$. Hence,

$$Supp_{C_\leq,P} = \frac{2^n - \sum_{i \in \{0,\ldots,n-k\cdot n\}} \frac{(n-k\cdot n)!}{i!(n-k\cdot n-i)!}}{2^n - 1}. \quad (4)$$

The numerator in (4) is the number of subsets of premises that contain at least one of the premises supporting q.

Corollary 3.7 *Let $(L, \leq)$ be a preordered set with first and last elements and $P \neq \{0\}$. Function $Supp_{C_\leq,P} : L \to [0,1]$ is a measure.*

Proof. It is monotonic as it is proven at the beginning of this section, and it verifies boundary conditions, since 0 is not a consequence $Supp_{C_\leq,P}(0) = \frac{2^n-2^n}{2^n-1} = 0$ and $Supp_{C_\leq,P}(1) = \frac{2^n-1}{2^n-1} = 1$. □

ISBN: 978-989-95079-6-8

Example 3.8 *Figure 1 represents a preordered set of medical symptoms and deseases for patients. Let's calculate the support for the consequences for a patient with* $P = \{antibody, bacterium\}$. *To such an end, let us notice that,* $C_{\leq}(P) = \{antibody, bacterium, fever, eruption, 1\}$. *Then,*

- $\widehat{Supp}_{C_{\leq},P}(antibody) = \frac{1}{2}$, *and* $Supp_{C_{\leq},P}(antibody) = \frac{2^2-(1+1)}{2^2-1} = \frac{2}{3}$.
- $\widehat{Supp}_{C_{\leq},P}(bacterium) = \frac{1}{2}$, *and* $Supp_{C_{\leq},P}(bacterium) = \frac{2}{3}$.
- $\widehat{Supp}_{C_{\leq},P}(fever) = 1$, *and* $Supp_{C_{\leq},P}(fever) = \frac{3}{3} = 1$.
- $\widehat{Supp}_{C_{\leq},P}(eruption) = \frac{1}{2}$, *and* $Supp_{C_{\leq},P}(eruption) = \frac{2}{3}$.
- $\widehat{Supp}_{C_{\leq},P}(1) = Supp_{C_{\leq},P}(1) = 1$.

Hence, the consequence with greatest support is 'fever'.

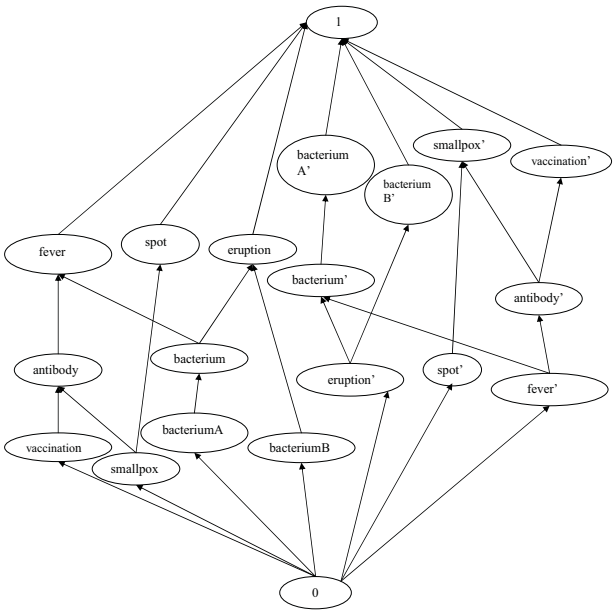


Figure 1: Preorder

4 Relevance for premises

This section introduces a measure to calculate the proportion of consequences that are gotten from a subset of premises, and what is more, it is shown how to reduce the set of premises using this ratio in order to give the same set of consequences.

Definition 4.1 *The relevance of a subset of premises* $R \in \mathbb{P}(P) - \{\emptyset\}$ *is the ratio of consequences deduced from a* R*, to all consequences.*

$$Rel_{C,P}(R) = \frac{|\{q \in L; q \in C(R)\}|}{|C(P)|} = \frac{|C(R)|}{|C(P)|}, \quad \text{if } R \in \mathbb{P}(P) - \{\emptyset\}, \tag{5}$$

and, $Rel_{C,P}(\emptyset) = 0$.

Since $|P| > 0$ *and* $P \subset C(P)$*, it is* $|C(P)| > 0$*, the quotient does exist.*

If a subset of premises allows to deduce all consequences, the set of premises can be reduced to it, since both have the same set of consequences.

There are many properties that $Rel_{C,P}$ verifies,

- If there exists $R \in \mathbb{P}(P)$, such that $Rel_{C,P}(R) = 1$, it means that all consequences for P are consequences of R. So, $C(P) = C(R)$.
- If $R_1 \subset R_2$, then $Rel_{C,P}(R_1) \leq Rel_{C,P}(R_2)$. That is function $Rel_{C,P}$ is monotonic.
- It is $Rel_{C,P}(P) = 1$, and $Rel_{C,P}(\emptyset) = 0$.

Theorem 4.2 *Function* $Rel_{C,P} : \mathbb{P}(P) \to [0,1]$ *is a measure.*

Proof. Straightforward, by the last properties. □

Remark 4.3 *In this case, the concept of fuzzy measure is defined in the preordered set* $(\mathbb{P}(L), \subset)$*, since relevance is defined for all subsets of premises and not only for single premises. Remember that the support is defined for each element.*

The ratio of relevance applying for each premise allows to define a partition into the set of premises, in classes whose elements have the same relevance, $[q] = \{p \in L;\ Supp_{C,P}(p) = Supp_{C,P}(q)\}$. Analogously, it can be built a partition in the set $\mathbb{P}(P)$, defining each class as $[S] = \{R \in \mathbb{P}(P);\ Supp_{C,P}(R) = Supp_{C,P}(S)\}$. The maximum number of classes that can exist is $|C(P)| + 1$.

4.1 Using the operator $C_{\leq}$

For the operator $C_{\leq}$ it is useful to calculate the relevance for each $p \in P$, since it is sufficient to get consequences for each premise and then join them. So, in this case is enough to deals with

$$Rel_{C_{\leq},P} : P \to [0,1]$$

Definition 4.4 *The relevance for a premise* $p \in P$ *is the proportion of consequences deduced from* p.

$$Rel_{C_{\leq},P}(p) = \frac{|\{q \in L; p \leq q\}|}{|C_{\leq}(P)|} = \frac{|C(p)|}{|C(P)|}. \tag{6}$$

If a premise allows to deduce all consequences, the set of premises can be reduced to that premise, since both give the same set of consequences.

There are many properties that $Rel_{C_{\leq},P}(p)$ verifies,

- For all $p \in P$, $Rel_{C_{\leq},P}(p) > 0$, since $p \in \{q \in L; p \leq q\}$ implies $|\{q \in L; p \leq q\}| > 0$.
- If there exists $p \in P$ such that $Rel_{C_{\leq},P}(p) = 1$, it means that all consequence for P is consequence of p. So, $C_{\leq}(P) = C_{\leq}(p)$.

ISBN: 978-989-95079-6-8

- If $p_1 \leq p_2$, then $Rel\,_{C_\leq,P}(p_2) \leq Rel\,_{C_\leq,P}(p_1)$. That is the function $Rel\,_{C_\leq,P}$ is anti-monotonic in this sense. Then, the function $1 - Rel\,_{C_\leq}$ is monotonic.
- Let L be endowed with an infimum operation. If $\inf P \in P$, as $Rel\,_{C_\leq,P}(\inf P) = 1$, because $\forall q \in C_\leq(P)$ there exists $p \in P$ such that $p \leq q$, and $\inf P \leq p \leq q$. Then, $C_\leq(P) = C_\leq(\inf P)$.
  A common consequence's operator is $C_\wedge$, defined by $C_\wedge(P) = \{q \in L; \inf P \leq q\}$, that can be defined as $C_\wedge(P) = C_\leq(\inf P)$, so in that case $C_\leq(P) = C_\wedge(P)$.

Example 4.5 *Using the same preset in figure 1. Let's calculate the relevance for premises. Here, we have an example that allows us to quantify the relevance of deseases of a patient.*

If the patient has $P = \{antibody, smallpox\}$*, then,* $C_\leq(P) = \{antibody, smallpox, spot, fever, 1\}$*. Hence,*

- $Rel\,_{C_\leq,P}(antibody) = \frac{|\{antibody,fever,1\}|}{|C_\leq(P)|} = \frac{3}{5}$.
- $Rel\,_{C_\leq,P}(smallpox) = \frac{|\{antibody,smallpox,spot,fever,1\}|}{|C_\leq(P)|} = \frac{5}{5} = 1$
- *Obviously,* $Rel\,_{C_\leq,P}(P) = 1$.

This example shows the case that a premise allows to deduce all consequences of P, since $C_\leq(P) = C_\leq(smallpox)$.

Theorem 4.6 *If* $(L, \leq, ', \cdot, +)$ *is a partial ordered set and* $P \subset L$ *such that* $\inf P \neq 0$.
There exists $p \in P$ *such that* $Rel\,_{C_\leq,P}(p) = 1$*, if and only if* $p = \inf P$.

Proof. If $Rel\,_{C_\leq}(p) = 1$, it is $p \leq q$ for all $q \in C_\leq(P)$. And since $C_\leq(P) \subset C_\wedge(P)$, it is $\inf P \leq q\ \forall q \in C_\leq(P)$. The infimum is the greatest lower bound of a set, then $p \leq \inf P$. It is also $\inf P \leq p$. Thus, $p = \inf P$, because L is a lattice, so verifies antisymmetric property and has an infimum for each subset.
On the other hand if $p = \inf P$, implies $\inf P \in P$, and as it is said, $Rel\,_{C_\leq,P}(\inf P) = 1$. □

In the theorem and in the above example, it is shown that the set of premises can be reduced to an only premise with relevance one, but if there is no one premise with relevance one, it could be found a subset of premises that allows to obtain the same consequences as the initial set of premises. When models deal with not a very big number of premises, a simple program can be used in order to find a minimal set of premises by calculating all combination of premises.

This algorithm is exponential in the number of premises. So, others algorithms can be designed in order to deal with a big number of premises.

,

The algorithm is as follows.

First of all, we look for premises with greatest relevance, we put one of these premises (p_1) into the set of reduced premises, then we calculate a relative relevance

$$Rel\,_{C_\leq,P-\{p_1\}}(p) = \frac{|\{q \in C_\leq(P) - C_\leq(p_1); p \leq q\}|}{|C_\leq(P) - C_\leq(p_1)|},$$

and we introduce a premise with the greatest relative relevance (p_2), and then we calculate other relative relevance,

$$Rel\,_{C_\leq,P-\{p_1,p_2\}}(p) =$$

$$\frac{|\{q \in (C_\leq(P) - C_\leq(p_1)) - C_\leq(p_2); p \leq q\}|}{|(C_\leq(P) - C_\leq(p_1)) - C_\leq(p_2)|},$$

and this process is repeated till the lowest r that verifies $C_\leq(P) = \bigcup_{i\in\{1,...,r\}} C_\leq(p_i)$. Then the reduced set searched in this way will be $\{p_1, ..., p_r\}$.

4.2 Using the operator $C_\bullet$

In this case we can particularize the definition of relevance for each subset of premises.

The relevance for a subsets of premises $R \subset \mathbb{P}(P) - \{\emptyset\}$ is the ratio of consequences deduced from R, to consequences deduced from P.

$$Rel\,_{C_\bullet,P}(R) = \frac{|\{q \in L; q \in C_\bullet(R)\}|}{|C_\bullet(P)|} =$$

$$\frac{|\{q \in L; \exists \tilde{R} \subset \mathbb{P}(R), \inf \tilde{R} \leq q\}|}{|C_\bullet(P)|}$$

Example 4.7 *Let* P *be* $\{c, d, a'\}$ *defined in the preorder in figure 2. So,* $C_\bullet(P) = \{c, d, f, g, b, e, d', a', 1\}$*. Relevance for all subset of premises are,*

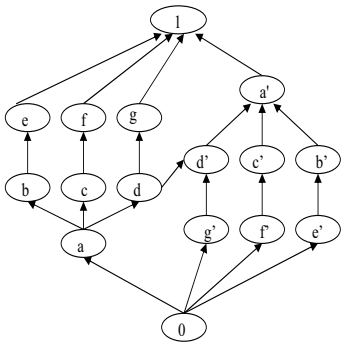


Figure 2: Preorder

- $Rel\,_{C_\bullet,P}(\{c\}) = \frac{3}{7}$, $Rel\,_{C_\bullet,P}(\{d\}) = \frac{5}{7}$, $Rel\,_{C_\bullet,P}(\{a'\}) = \frac{2}{7}$
- $Rel\,_{C_\bullet,P}(\{c, d\}) = 1$, $Rel\,_{C_\bullet,P}(\{c, a'\}) = 1$, $Rel\,_{C_\bullet,P}(\{d, a'\}) = \frac{5}{7}$

This example gives two reduced sets of premises, $\{c, d\}$ *and* $\{c, a'\}$*. Obviously* $C_\bullet(P) = C_\bullet(\{c, d\}) = C_\bullet(\{c, a'\})$.

5 Validating premises and partial hypotheses

Each premise is supported, or explained by hypotheses, so in this section, a support for each premise is defined. Then, we define the relevance of each partial hypothesis. This is analogously to the above section, because premises are consequences of hypotheses.

Let $Hyp^*{}_C(P) \neq \emptyset$.

ISBN: 978-989-95079-6-8

Definition 5.1 *The support of $p \in P$ is the ratio of subsets of hypotheses that allow getting p as consequence, to all subsets of partial hypotheses.*

$$Supp_{C,Hyp^*_C(P)}(p) = \frac{|\{H \subset Hyp^*_C(P); p \in C(H)\}|}{2^{|Hyp^*_C(P)|} - 1} \quad (7)$$

Since $|Hyp^*_C(P)| > 0$ the quotient in the definition is possible.

The bigger Support a premise has, the more hypotheses allow to deduce it.

The $Supp_{C,Hyp^*_C}$ verifies,

- If $P = \{p\}$, it is $Supp_{C,Hyp^*_C(P)}(p) = 1$.
- If $1 \in P$, then $Supp_{C,Hyp^*_C(P)}(1) = 1$.
- $Supp_{C,Hyp^*_C(P)}(p) = 1$ means that p is explained by all $h \in Hyp^*_C(P)$, in particular for all $h \in Hyp(P)$.
- If $p_1 \leq p_2$, then $Supp_{C,Hyp^*_C(P)}(p_1) \leq Supp_{C,Hyp^*_C(P)}(p_2)$. That is, function Sup is monotonic.
- $Supp_{C,Hyp^*_C(P)}(\sup\{p_1, p_2\}) \geq \max\{Supp_{C,Hyp^*_C(P)}(p_1), Supp_{C,Hyp^*_C(P)}(p_2)\}$.
- $Supp_{C,Hyp^*_C(P)}(\inf\{p_1, p_2\}) \leq \min\{Supp_{C,Hyp^*_C(P)}(p_1), Supp_{C,Hyp^*_C(P)}(p_2)\}$.

$Supp_{C,Hyp^*_C(P)}$, allows to compare premises in relation to hypotheses and to allocate different degrees to each premise.

Definition 5.2 *The relevance for a subset of partial hypotheses $H \subset Hyp^*_C(P)$ is the proportion of premises deduced from H.*

$$Rel_{C,Hyp^*_C(P)}(H) = \frac{|\{p \in P; p \in C(H)\}|}{|P|}, \quad (8)$$

if $H \in \mathbb{P}(Hyp^*_C(P)) - \{\emptyset\}$,

and, $Rel_{C,Hyp^*_C(P)}(\emptyset) = 0$.

Since $|P| > 0$, the quotient is possible.

There are many properties that $Rel_{C,Hyp^*_C(P)}$ verifies,

- If $P = \{p\}$, it is $Rel_{C,Hyp^*_C(P)}(H) = 1$, for all $H \subset Hyp^*_C(P)$.
- For all $H \subset Hyp^*_C(P)$, $Rel_{C,Hyp^*_C(P)}(H) > 0$.
- If $H_1 \subset H_2$, then $Rel_{C,Hyp^*_C(P)}(H_1) \leq Rel_{C,Hyp^*_C(P)}(H_2)$. That is, function $Rel_{C,Hyp^*_C(P)}$ is monotonic.
- If there exists $h \in Hyp^*_C(P)$ such that $Rel_{C,Hyp^*_C(P)}(\{h\}) = 1$, it means that h is a partial hypothesis that can explain all premises, so it is hypothesis.

Theorem 5.3 *If $Hyp_C(P) \neq \emptyset$, then $Rel_{C,Hyp^*_C(P)}$ is a measure.*

Proof. It is monotonic as it is stated above, $Rel_{C,Hyp^*_C(P)}(\emptyset) = 0$, and since there exists $h \in Hyp_C(P)$ $Rel_{C,Hyp^*_C(P)}(\{h\}) = 1$.
□

The measure, $Rel_{C,Hyp^*_C(P)}$, allows to compare partial hypotheses, and to distinguish which partial hypotheses are hypotheses in the classical sense, that are those one with relevance one.

6 Conclusions

In this paper, it is built a measure in the set of consequences, premises and partial hypotheses. That can be useful in decision problems in order to choose the consequence, premise or hypotheses with the biggest measure.

It should be also pointed out that by using a relevance measure, we can get rid of premises that do not add information, and still get the same set of consequences.

It is also introduced the concept of set of partial hypotheses, that contains the classical hypotheses one. The measure built allocate value one to that partial hypotheses that are hypotheses in the classical sense.

As future work it can be proposed to apply these measures to practical problems, for examples medical diagnosis problems much more bigger than the one that appears in this paper.

7 Acknowledgements

Authors thanks to the three anonymous reviewers for their hints and comments.

References

[1] J. L. Castro and E. Trillas. Sobre preordenes y operadores de consecuencias de Tarski. *Theoria*, 4(11):419–425, 1989. (In Spanish).

[2] M.K. Chakraborty. Use of fuzzy set theory in introducting graded consequence in multiple-valued logic. *Fuzzy Logic in Knowledge-Based Systems, Decision and Control*, (1):247–257, 1988.

[3] M.K. Chakraborty and S. Dutta. *Theory of graded consequence and fuzzy logics*. LATD, Italy, 2008.

[4] D. Qiu. A note on Trillas' CHC models. *Artificial Intelligence*, 171:239–254, 2007.

[5] A. R. de Soto, A. Álvarez, and E. Trillas. Short note: Counting conjectures. *Mathware and Soft Computing*, 14(2):165–170, 2007.

[6] E. Trillas and C. Alsina. A reflection on what is a membership function. *Mathware & Soft-Computing*, 6:201–215, 1999.

[7] E. Trillas, S. Cubillo, and E. Castiñeira. On conjectures in orthocomplemented lattices. *Artificial Intelligence*, 117:255275, 2000.

[8] E. Trillas, I. García-Honrado, and A. Pradera. Consequences and conjectures in peordered sets. *ECSC's Tecnical Report (available upon request)*.

[9] E. Trillas, M. Mas, and M. Monserrat. Conjecturing from consequences. *Forthcoming in Int. journal of General Systems*.

[10] Satosi Watanabe. *Knowing and guessing. A Quantitative Study of Inference and Information*. John Wiley and sons, New York, 1969.

ISBN: 978-989-95079-6-8

Flexible Retrieval of X-Ray Images Based on Shape Descriptors Using a Fuzzy Object-Relational Database

Juan Miguel Medina[1] Sergio Jaime-Castillo[1] Carlos D. Barranco[2] Jesús R. Campaña[1]

1.Department of Computer Science and Artificial Intelligence. University of Granada
C/ Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain.
2.Division of Computer Science, School of Engineering. Pablo de Olavide University
Utrera Rd. Km. 1, 41013 Seville, Spain.
Email: {medina,sjaime}@decsai.ugr.es, cbarranco@upo.es, jesuscg@decsai.ugr.es

Abstract— *This paper presents a novel approach for medical image storage using a Fuzzy Object-Relational Database Management System (FORDBMS). The system stores medical images along with a set of parameters describing their content. Flexible queries can be performed over these parameters to retrieve images matching visually. To illustrate the capabilities of the FORDBMS, parameter curves are obtained from X-Ray images of patients suffering from scoliosis, and queries are performed when looking for images with a determined curve pattern. Results show that retrieved images visually match the condition established in the query.*

Keywords— Image Retrieval, Fuzzy Databases

1 Introduction

Images are a fundamental tool in health care for diagnosis, clinical studies, research and learning. Currently, there are multiple techniques to capture images from patients to help diagnostic tasks such as X-Ray images, Computer Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Ultrasonography, etc. The diagnostic task generates a large amount of images which must be archived for future evaluations. Fortunately, most of these techniques produce digital images, which are more efficiently archived and handled through computer systems than physical ones. In medical imaging, these computer systems are called Picture Archiving and Communication Systems (PACS). PACS are computers or networks devoted to the storage, retrieval, distribution and presentation of images. These systems solve the problem of storaging digital images but do not provide mechanisms to retrieve them based on their content.

Content-Based image retrieval (CBIR) [1] is the application of computer vision techniques to the problem of digital image search in large databases. First image retrieval approaches were based on captions and textual descriptors collected by humans. Nowadays, image retrieval systems improve the textual-based ones using features such as color, texture or shape, which are automatically extracted from images [2]. In this regard, a very important point to take into account is the imprecision in the feature descriptions, as well as the storage and retrieval of that imprecise data. To deal with this vagueness, some interesting approaches introduce the use of fuzzy logic in the feature representation, as well as in the retrieval process [3, 4, 5]. These fuzzy approaches also allow to perform queries on the basis of linguistic terms, avoiding one of the drawbacks of the classical image retrieval systems, where the queries have to be defined on the basis of images or sketches similar to the one we are searching for.

Database Management Systems (DBMS) are crucial in CBIR for retrieval purposes and, as we have explained above, these systems must be able to represent fuzzy data as complex datatype structures to provide flexible content based retrieval. There is a wide variety of proposals for fuzzy data handling in databases [6, 7, 8, 9, 10, 11] but in general these models and/or implementations do not have enough modeling power and performance for image indexing applications. For these kind of applications we propose to use the Fuzzy Object-Relational Database System (FORDBMS) model introduced in [12, 13]. This model evolves classical fuzzy databases models to incorporate object-oriented features for a powerful representation and handling of data, fuzzy or not.

CBIR techniques get more semantical results when applied to a specific domain or application area, as knowledge on the domain and on the image characteristics helps the process to extract the relevant features for this specific area of application. Health care is an application area that may benefit from the CBIR techniques. If we focus the CBIR techniques on the analysis of a certain pathology we can get high level features processing certain types of images. For example, there are some proposals to extract vertebral and spinal shapes from X-rays [14, 15] (these algorithms need some user intervention). The paper [16] describes a technique to automatically measure Cobb angle [17] for scoliosis pathologies given the end vertebras of the curve.

This paper shows how our FORDBMS is suitable for easy image representation and retrieval, using the fuzzy descriptors obtained by means of computer vision algorithms [18] or provided by experts. To illustrate this, we show the process to represent a structure in the FORDBMS describing scoliosis measures (Cobb angles) obtained from anteroposterior X-rays. Aditionally, we will show that the use of fuzzy comparators implemented by the FORDBMS in queries let us retrieve images that match the expected visual characteristics searched.

The rest of the paper is organized as follows. Section 2 describes the scoliosis pathology and the use of X-rays for its diagnosis. Section 3 introduces the fuzzy object-relational database system used to store and retrieve the fuzzy data. Examples of queries are presented in Section 4 and, finally, the main conclusions and future works are summarized in Section 5.

ISBN: 978-989-95079-6-8

2 Idiopathic scoliosis and its X-ray based diagnosis

In order to show how the use of fuzzy databases can help in health care, we have focused on the study of representation and retrieval of images related with idiopathic scoliosis. In this section we will describe the most relevant characteristics of this pathology in relation with our purpose.

Scoliosis is a three-dimensional deformation of the spine that produces vertebral rotation and crushing, and lateral curvature. It is typically classified as congenital (caused by vertebral anomalies present at birth), idiopathic (sub-classified as infantile, juvenile, adolescent, or adult according to when onset occurred) or as having developed as a secondary symptom of another condition, such as cerebral palsy, spinal muscular atrophy or due to physical trauma. Depending on the severity and progression of the deformation may be necessary treatment, consisting of observation, orthotic (brace) treatment, or surgery. About 2-4% of the adolescent population has some degree of scoliosis. Approximately 2.2% of these adolescents will require treatment.

To diagnose and treat scoliosis it is necessary to perform measures of the spine deformity. There are physical examinations to initially detect the presence of the deformity, but a precise diagnosis and treatment needs the help of radiologic techniques. The most accurate technique to measure spinal deformity is Computed Tomography (CT), that provides a three-dimensional view of the spine. However this technique is expensive and exposes the patient to a high radiation. Taking into account that a patient having scoliosis may need observation and treatment for many years and many radiologic tests, the frequent use of this technique may be inappropriate. X-Rays expose the patient to a lower radiation and, because of this, full-length standing spine X-rays are the standard method for evaluating the severity and progression of the scoliosis. Anteroposterior X-rays (AP X-rays) project spinal deformities as curves. The standard method to quantitatively asses curvatures is the measurement of the Cobb angle on AP X-rays. The Cobb angle [17] can be manually measured by calculating the angle between the lines respectively drawn along the upper endplate of the superior end-vertebra and the lower endplate of the inferior end-vertebra, as shown in Fig. 1. Using this measure each curve present in the spine is characterized by means of four parameters: the side of the convexity of the curve (right or left), the superior end-vertebra, the inferior end-vertebra and the angle value. It is important also to identify the apical vertebra associated with the adjacent disc interspaces that have the greatest segmental angulation of all interspaces in the curve. This vertebra occurs at the horizon or apex of a curve (T9 in Fig. 1).

The manual measurement of Cobb angle depends on experience and personal judgment. Errors are due to selecting different end- vertebrae and estimating different slopes of the vertebrae. The standard measurement error is 3° to 5° for the same observer and 5° to 7° for different observers.

The use of AP X-rays is useful for diagnosis, clinical studies and health learning. To help these purposes, it will be interesting to perform storage and retrieval of X-rays from a database based on curves parameters present in the spine. The problem is that, as we have shown, measures are imprecise and the query parameters needs to be flexible. For this reason, classical databases are not suitable for this purpose. It is necessary a Database System that can store imprecise data and can perform flexible queries on them. Moreover, as spine parameter description is complex, the database system must handle complex structures to represent imprecise data and must provide flexible comparators on these structures. Our FORDBMS, as we will show in next sections, have those capabilities.

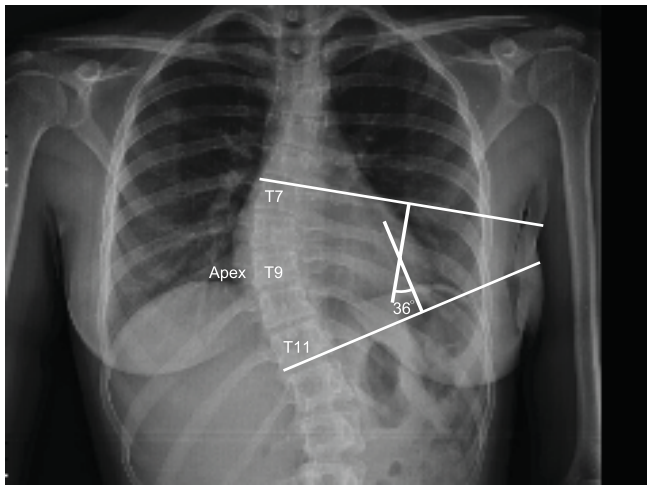


Figure 1: Cobb angle measurement.

3 The fuzzy object-relational database system

The recent proliferation of large image databases leads to the need for DBMS applied to multimedia libraries management to ensure high performance, scalability, availability with fault tolerance and distribution.

Nowadays, market leader DBMSs offer these required features transparently. However, the database models implemented by them, generally the relational model, are not suitable to manage fuzzy data, which is necessary for this kind of image description algorithms. In order to solve this drawback, some database models and DBMSs implementing them have been proposed. Nevertheless, the existing fuzzy DBMSs are in general research prototypes which do not match the high performance and other necessary requirements for this kind of applications.

In [12, 13] we introduce the strategy of implementation of our FORDMS model, that it is based on the extension of a market leader RDBMS (Oracle®) by using its advanced object-relational features. This strategy let us take full advantage of the host RDBMS features (high performance, scalability, etc.) and the ability for representing and handling fuzzy data provided by our extension, making this FORDBMS suitable to support systems for flexible content based retrieval of images.

3.1 Fuzzy datatype support

Our FORDBMS is able to handle and represent a wide variety of fuzzy datatypes, which allows to model any sort of fuzzy data easily. These types of fuzzy data, which are represented as classes with light gray background color in Fig. 2, are the following:

- Atomic fuzzy types (AFT), represented as possibility distributions over ordered (OAFT) or non ordered (NOAFT)

ISBN: 978-989-95079-6-8

domains.

- Fuzzy collections (FC), represented as fuzzy sets of objects with conjunctive (CFC) or disjunctive (DFC) semantics.
- Fuzzy objects (FO), whose attribute types could be crisp or fuzzy, and where each attribute is associated with a degree to weigh its importance in object comparison.

All fuzzy types define a Fuzzy Equal operator (FEQ) that computes the degree of fuzzy equality for each pair of instances. Each fuzzy datatype has its own implementation of this operator in accordance with its nature. Moreover, the FORDBMS provides parameters to adjust the fuzzy equality computation to the semantics of the data handled. For OAFTs the system implements other fuzzy comparators as FGT (fuzzy greater than), FLT (fuzzy less than), etc.

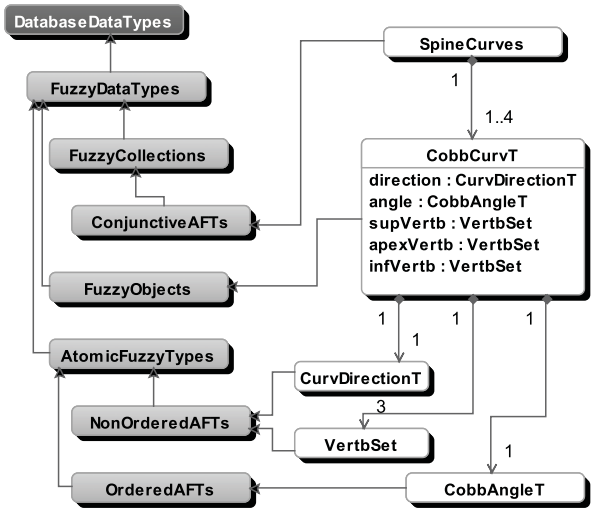


Figure 2: UML Class Diagram for Fuzzy types and SpineCurves Datatype.

3.2 Fuzzy datatype operators

Our FORDBMS can handle flexible comparisons over several types of fuzzy datatypes. Moreover, our system provides the possibility of adapting the behavior of some operators to adjust flexible comparisons to the semantics of the data modeled. The current section describes the most significant operators for the defined fuzzy datatypes, and their adaptation to handle flexible queries on complex objects like the spine curve description structure illustrated in this paper.

3.2.1 Fuzzy inclusion operator

The operator `FInclusion(A,B)` calculates the inclusion degree $A \subseteq B$, where A and B are instances of CFC. The calculus is done using a modification of the *Resemblance Driven Inclusion Degree* introduced in [19], which computes the inclusion degree of two fuzzy sets whose elements are imprecise.

Definition 1 (Resemblance Driven Inclusion Degree). *Let A and B be two fuzzy sets defined over a finite reference universe $\mathscr{U}$, μ_A and μ_B the membership functions of these fuzzy sets, S the resemblance relation defined over the elements of $\mathscr{U}$, $\otimes$ be a t-norm, and I an implication operator. The inclusion degree of A in B driven by the resemblance relation S is calculated as follows:*

$$\Theta_S(B|A) = \min_{x \in \mathscr{U}} \max_{y \in \mathscr{U}} \theta_{A,B,S}(x,y) \tag{1}$$

where

$$\theta_{A,B,S}(x,y) = \otimes(I(\mu_A(x), \mu_B(y)), \mu_S(x,y)) \tag{2}$$

We propose a modification that substitutes the minimum aggregation operator in equation 1 by a weighted mean aggregation operator, whose weight values are the membership degrees in A of the elements of $\mathscr{U}$, divided by the cardinal of A. This modification is made in order to obtain a less extreme resemblance inclusion degree, since it takes into account the importance of each included element. The *Modified Resemblance Inclusion Degree* is defined in equation 3.

$$\Theta_S(B|A) = \sum_{x \in \mathscr{U}} \frac{\mu_A(x)}{|A|} \cdot \max_{y \in \mathscr{U}} \theta_{A,B,S}(x,y) \tag{3}$$

with $|A| = \sum_{x \in \mathscr{U}} \mu_A(x)$. Our implementation of `FInclusion(A,B)` uses the minimum as t-norm, and the Gödel implication as the implication operator.

3.2.2 Fuzzy equality Operator

The operator `FEQ(A,B)` calculates the resemblance degree between two instances of a fuzzy datatype.

When `A` and `B` are two instances of CFC, this resemblance degree is calculated by means of the *Generalized Resemblance between Fuzzy Sets* proposed in [19], which is based on the concept of double inclusion.

Definition 2 *(Generalized resemblance between fuzzy sets). Let A and B be two fuzzy sets defined over a finite reference universe $\mathscr{U}$, over which a resemblance relation S is defined, and $\otimes$ be a t-norm. The generalized resemblance degree between A and B restricted by $\otimes$ is calculated by means of the following formulation:*

$$\beta_{S,\otimes}(A,B) = \otimes(\Theta_S(B|A), \Theta_S(A|B)) \tag{4}$$

Therefore, the implementation of the operator `FEQ(A,B)`, when A and B are instances of CFC, aggregates the results of `FInclusion(A,B)` and `FInclusion(B, A)` using a minimum t-norm.

If the operator `FEQ(A,B)` is applied when `A` and `B` are instances of the class *FO*, then, for the sake of flexibility, the resemblance degree between these objects is calculated as a weighted average of the resemblance degree of their attribute values. In the FORDBMS catalog we store some parameters for the attributes that belong to each fuzzy object. Specifically, the table *SDS_FO_ATTRIBUTES* stores into the column *relevance*, the relevance value of the considered attribute. This value attaches a weight to the attribute for the calculation of the `FEQ` operator. Moreover, if this value is set to -1 for an attribute, this attribute acts as determinant in the object comparison, in the way that, if the resemblance operator on this

ISBN: 978-989-95079-6-8

attribute returns 0, then the complete objects comparison returns 0. This parameter is useful, for example, in the case that we are searching for *"left curves"*, setting *Direction* attribute to -1, the query never retrieves a *"right curve"* even if resemblance for other object attributes is greater than 0. Another parameter that modifies the computation of the operator `FEQ` on fuzzy objects is the minimum percentage of attributes comparison greater than 0 (*min_pct*). This parameter, stored in the catalog table *SDS_FO_TYPES*, set the necessary number of attributes comparisons greater than 0 to get a value greater than 0 for the whole object comparison.

The following definition formalizes the ideas exposed above.

Definition 3 *(Parameterized Object Resemblance Degree). Let o_1 and o_2 be two objects of the class C, $obj.a_i$ the value of the i-th attribute of the object obj, $rel(a_i)$, relevance degree of the i-th attribute of the object obj, n the number of attributes defined in the class C, min_pct the necessary minimum percentage of attributes comparison greater than 0, and FEQ the resemblance operator.*

$$OR(o_1,o_2) = \begin{cases} 0 \text{ if} \\ (\exists i \in \mathcal{A} : FEQ(o1.a_i, o2.a_i) = 0 \wedge rel(a_i) = -1) \vee \\ (|\{a_i : i \in \mathcal{A} \wedge FEQ(o1.a_i, o2.a_i) > 0\}| < n.min\_pct) \\ \\ \sum_{i\in\mathcal{A}}^{n} FEQ(o1.a_i, o2.a_i).|rel(a_i)| / \sum_{i\in\mathcal{A}}^{n} |rel(a_i)| \\ otherwise \end{cases} \tag{5}$$

Our FORDBMS also let us relax the resemblance operator `FEQ` applied on *OAFT* values. To do this, the user must use the static method `relax(kernel_pct,support_pct)`, where `kernel_pct` sets the percentage that the kernel increases for each pair of values being compared, and `support_pct` sets the percentage increase for the support of these values.

4 Retrieving X-Ray images from database using its fuzzy description

Our FORDBMS can handle complex fuzzy data structures to represent rich semantic problems. It can store instances of these structures and provides a powerful set of operators to retrieve data based on flexible parameters. We have chosen an example based on the representation of the characteristics of the spine curves taken from AP X-rays to illustrate how flexible queries can retrieve interesting results for the specialists. Let us show how to model the data structure of this example on our FORDBMS, then we will perform some queries and we will analyze the retrieved results.

4.1 Database modeling for spine curve description

Our FORDBMS is capable of representing curves structures present in the spine measured from AP X-Rays, taking into account that the measures can be imprecise due to errors. It also provides parameters to adjust the fuzzy comparators used to get good query results.

The classes whose background color is white in Fig. 2 integrate the datatype structure of the database that represents spine parameters. As this figure illustrates, we model the curve description of the spine as a fuzzy conjunctive collection datatype (*SpineCurves*) which includes up to four fuzzy object datatypes (*CobbCurveT*). *CobbCurveT* datatype represents a spine curve measure. This datatype has five attributes: *Direction*, *Angle*, *SupVertb*, *ApexVertb* and *InfVertb* that store: the side of the convexity of the curve (right or left), the angle measure and, the superior, apical and inferior vertebra of the curve, respectively. Attribute *Direction* is of NOAFT type, this allows to perform queries with a not determined value for the side of convexity of the curve. The attributes *SupVertb*, *ApexVertb* and *InfVertb* store values of type NOAFT. The domain of this attributes (*vertbSetT*) includes 24 vertebrae that are enumerated as follows: $v_1 = L5, v_2 = L4, v_3 = L3, v_4 = L2, v_5 = L1, v_6 = T12, v_7 = T11, v_8 = T10, v_9 = T9, v_{10} = T8, v_{11} = T7, v_{12} = T6, v_{13} = T5, v_{14} = T4, v_{15} = T3, v_{16} = T2, v_{17} = T1, v_{18} = C7, v_{19} = C6, v_{20} = C5, v_{21} = C4, v_{22} = C3, v_{23} = C2, v_{24} = C1$. We will define a nearness relation on this underlined domain to take into account the adjacency between each pair of vertebrae as follows:

$$\forall i,j \in Cardinal(VertbSett): \quad nearness(v_i,v_j) = \begin{cases} 1 & if\ i = j \vee i = j+1 \vee \\ & i = j-1 \\ 0.66 & if\ i = j+2 \vee i = j-2 \\ 0.33 & if\ i = j+3 \vee i = j-3 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The attribute *Angle* is of type OAFT, so we can store on this attribute numerical values, crisp or fuzzy (using a trapezoidal membership function). Measures obtained with errors for the angle can be represented as a fuzzy value for the attribute. This attribute allows flexible comparisons also. We can use the method `relax` described in the previous section on this attribute.

Once we have defined the type structure for the spine description, we can create a table that stores X-rays images together with their fuzzy descriptions as follows:

```
create table APXRay (
 image# number, xray bfile, SpineDescription SpineCurves);
```

We have inserted twenty images with their curve description to illustrate query examples; the statement for one image insertion is like this:

```
Insert into apxray values (
 313701 ,BFILENAME('APXRays','313701.gif'),
 SpineCurves(
     1,cobbCurvT(
        CurvDirectionT('RIGHT'),
        CobbAngleT(trapezoid(10.1)),
        VertbSetT('T2'),
        VertbSetT('T5'),
        VertbSetT('T10')
             ),
     1,cobbCurvT(
        CurvDirectionT('LEFT'),
        CobbAngleT(trapezoid(16.5)),
        VertbSetT('T9'),
        VertbSetT('T12'),
        VertbSetT('L2')
             ),
     1,cobbCurvT(
        CurvDirectionT('RIGHT'),
        CobbAngleT(trapezoid(12.5)),
        VertbSetT('L2'),
        VertbSetT('L4'),
        VertbSetT('L5')
             )
  )
);
```

ISBN: 978-989-95079-6-8

4.2 *Examples of Queries*

In this section we will illustrate some capabilities by performing a query that uses the `FInclusion` operator and a second one that uses the `FEQ` operator on `CFC` datatypes. To get flexible results on queries that involve whole fuzzy object comparison, it is advisable to relax comparison on angle attribute. To perform it we invoke the static method `relax` described in Section 3.2.2. In this case we relax 40% the kernel and 70% the support. This is the SQL sentence: `Execute CobbAngleT.relax(0.4,0.7);`

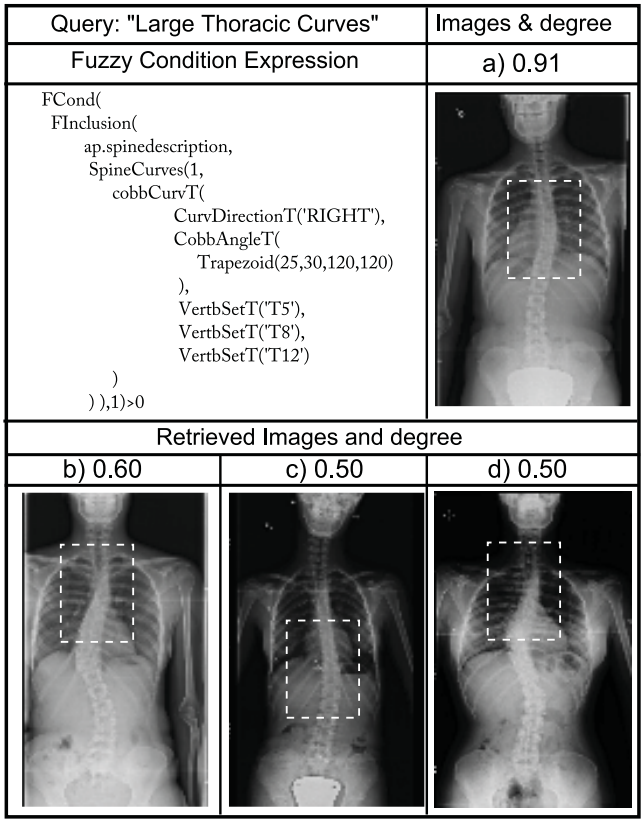


Figure 3: Searching images that include a thoracic pattern curve.

As shown in Fig. 3, we can perform queries that search images including a determined pattern of curves. In the example we illustrate the following query: *"Show X-Ray images and its compliance degree that include a large thoracic curve"*. The "thoracic curve" pattern proposed in [20] is described by the following approximate parameters: the convexity of the curve is on the right side, the superior end-vertebra is between T4 and T6, the apex vertebra is between T8 and T9 and, the inferior end-vertebra is between T11 and L2. The Fuzzy Condition Expression cell of the Fig. 3 shows the parameters used for the `WHERE` clause of the query statement, where the concept "large" is modeled as a trapezoidal distribution that matches Cobb angles that are approximately greater or equal than 25-30 degrees.

To perform the query we use the `FInclusion` operator, described in Section 3.2.1, which evaluates the similarity of the queried curve with respect to each curve included in the spine description. The query retrieves four images from the database. The image a), includes a curve, highlighted with a rectangular area, witch takes the values 'Right', 21.8, T5, T8, T10 for the parameters: orientation, angle size, superior, apex and inferior vertebrae, respectively. The compliance degree for this image with respect the query is 0.91. As we can see, this image includes a curve that appreciably matches the concept "large thoracic curve", this is the reason for the high degree computed. Visually, the marked curve in image b) weakly matches the concept "thoracic curve", this is the reason for a minor compliance degree. Image c) includes a curve that tends to a "thoracolumbar curve" concept; because of this, the retrieved compliance degree is low, 0.5. Finally, image d) includes a curve that tends to a "cervicothoracic curve" concept, the compliance degree (0.50) confirms this fact.

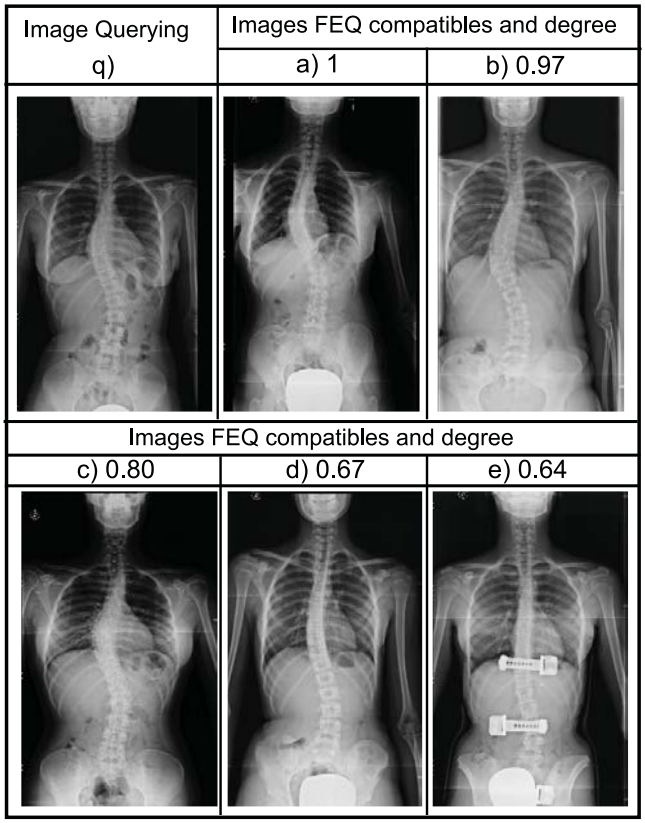


Figure 4: Searching images that present similar spine curvature that image q).

In some studies it may be interesting to find X-rays of patients that present a similar curve pattern to a given one. In our database, we model these curve patterns present in the spine as a *CFC*. To find similar curve patterns, our FORDBMS provides the operator `FEQ` which operates on *CFC* as has been described in Section 3.2.2. The query shown in Fig. 4 searches for X-rays of patients presenting similar curve patterns that the image query q). The SQL syntax is:

```
SELECT ap1.image#,ap1.xray,ap1.cdeg(1)
FROM apxray ap1, apxray ap2
WHERE ap1.image='q'
     AND FCOND(
          FEQ(ap1.spinedescription, ap2.spinedescription),1
         )>0 order by cdeg(1) desc;
```

As can be seen, higher compliance degree for an image denotes better visual matching with the queried image. Another interesting aspect of our results is the following: the images retrieved by the two kinds of queries evaluated hold that, the

ISBN: 978-989-95079-6-8

lower is the curve pattern matching, the lower the compliance degree computed for these one is.

5 Concluding remarks and future work

This paper shows that a FORDBMS is a powerful tool to represent flexible descriptors obtained from images by means of computer vision algorithms or provided by experts. The system also provides parameterized fuzzy comparators to retrieve images based on those descriptors in a flexible way. To the best of our knowledge, there is no system for flexible image retrieval that provides same capabilities on the subject considered in this paper. The prototype of FORDBMS used to show these results are implemented on the RDBMS Oracle® 10.2. Thereon, it is interesting to make the following remarks:

- The prototype has not implemented yet any index technique to accelerate the data retrieval based on fuzzy conditions. For this reason, the queries shown have been computed by means of a sequential search; moreover, the complex fuzzy datatype structure used for the spine curve representation more decreases the efficiency if none index technique is used. We are working on the implementation of the indexes techniques for fuzzy data proposed in [21, 22]; this will provide us mechanisms to optimize retrieval process.
- The scalability of our FORDBMS and CBIR system shown is guaranteed by the scalability of the host ORDBMS system, Oracle® 10.2.

Future work will involve too the enrichment of fuzzy features of the FORDBMS to enhance support for image content based retrieval. Finally we will study another application domains to perform content based retrieval using our FORDBMS.

Acknowledgment

This work has been partially supported by the Spanish "Ministerio de Ciencia y Tecnología" (MCYT) under grants TIN2006-07262 and TIN2007-68084-CO2-01, and the "Consejería de Innovación Ciencia y Empresa de Andalucía" (Spain) under research projects P06-TIC-01570 and P07-TIC-02611.

References

[1] Michael S. Lew, Nicu Sebe, Chabane Djeraba Lifl, and Ramesh Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1):1–19, 2006.

[2] J. Fuertes, M. Lucena, N. Perez de la Blanca, and J. Chamorro-Martínez. A scheme of colour image retrieval from databases. *Pattern Recognition Letters*, 22:323–337, 2001.

[3] Ju Han and Kai-Kuang Ma. Fuzzy color histogram and its use in color image retrieval. *IEEE Transactions on Image Processing*, 11(8):944–952, 2002.

[4] D. Sánchez, J. Chamorro-Martínez, and M.A. Vila. Modelling subjectivity in visual perception of orientation for image retrieval. *Information Processing and Management*, 39:251–266, 2003.

[5] B. Verma and S. Kulkarni. Fuzzy logic based interpretation and fusion of color queries. *Fuzzy sets and systems*, 147:99–118, 2004.

[6] S. Fukami, M. Umano, M. Muzimoto, and H. Tanaka. Fuzzy database retrieval and manipulation language. *IEICE Technical Reports, AL-78-85 (Automata and Language)*, 78(233):65–72, 1979.

[7] J.M. Medina, O. Pons, and M.A. Vila. Gefred. a generalized model of fuzzy relational databases. *Information Sciences*, 76(1):87–109, 1994.

[8] J.M. Medina, M.A. Vila, J.C. Cubero, and O. Pons. Towards the implementation of a generalized fuzzy relational database model. *Fuzzy Sets and Systems*, 75:273–289, 1995.

[9] H. Prade and C. Testemale. Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries. *Information Sciences*, 34:115–143, 1984.

[10] M. Umano. Freedom-o: A fuzzy database system. In Gupta-Sanchez, editor, *Fuzzy Information and Decision Processes*. North-Holand, 1982.

[11] M. Zemankova-Leech and A. Kandel. *Fuzzy Relational Databases - A Key to Expert Systems*. Verlag TUV Rheinland, Koln, 1984.

[12] J.C. Cubero, N. Marín, J.M. Medina, O. Pons, and M.A. Vila. Fuzzy object management in an object-relational framework. In *Proc. 10th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'2004*, pages 1767–1774, 2004.

[13] Carlos D. Barranco, Jesús R. Campaña, and Juan M. Medina. Towards a fuzzy object-relational database model. In José Galindo, editor, *Handbook of Research on Fuzzy Information Processing in Databases*, pages 435–461. IGI Global, 2008.

[14] B. Verdonck, R. Nijlunsing, F. A. Gerritsen, J. Cheung, D. J. Wever, A. Veldhuizen, S. Devillers, and S. Makram-Ebeid. Computer assisted quantitative analysis of deformities of the human spine. *Lecture Notes on Computer Sciences*, 1496:822–831, 1998.

[15] Mohammed Benjelloun and Sad Mahmoudi. Spine localization in x-ray images using interest point detection. *Journal of Digital Imaging. Published on-line, doi:10.1007/s10278-007-9099-3*, 2008.

[16] Junhua Zhang, Edmond Lou, Lawrence H. Le, Douglas L. Hill, James V. Raso, and Yuanyuan Wang. Automatic cobb measurement of scoliosis based on fuzzy hough transform with vertebral shape prior. *Journal of Digital Imaging. Published on-line, doi: 10.1007/s10278-008-9127-y*, 2008.

[17] JR Cobb. Outline for the study of scoliosis. *Am Acad Orthop Surg Inst Course Lect*, 5:261–275, 1948.

[18] J. Chamorro-Martínez, J.M. Medina, C.D. Barranco, E. Galán-Perales, and J.M. Soto-Hidalgo. Retrieving images in fuzzy object-relational databases using dominant color descriptors. *Fuzzy Sets and Systems*, 158:312324, 2007.

[19] N. Marín, J.M. Medina, O. Pons, D. Sánchez, and M.A. Vila. Complex object comparison in a fuzzy context. *Information and Software Technology*, 45(7):431–444, 2003.

[20] Ponseti and B. Friedman. Prognosis in idiopathic scoliosis. *J Bone Joint Surgery*, 32:381–395, 1950.

[21] Carlos D. Barranco, Jesús R. Campaña, and Juan M. Medina. A b+-tree based indexing technique for fuzzy numerical data. *Fuzzy Sets and Systems*, 159:14311449, 2008.

[22] Carlos D. Barranco, Jesús R. Campaña, and Juan M. Medina. B+-tree based indexing technique for necessity measured flexible conditions on fuzzy numerical data. In *Proc. 10th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'2008 Torremolinos (Spain)*, pages 1717 – 1724, 2008.

ISBN: 978-989-95079-6-8

Minimal Combination for Incremental Grammar Fragment Learning

Nurfadhlina Mohd Sharef Trevor Martin Yun Shen

Artificial Intelligence Group, University of Bristol
Bristol, BS8 1TR, United Kingdom
Email: ennms@bris.ac.uk, trevor.martin@bris.ac.uk, yun.shen@bris.ac.uk

Abstract-*The ability to identify text fragments can benefit many information extraction tasks. The main challenge in this matter is to recognise the multiple patterns that exist across possibly heterogeneous sources. Traditional machine learning methods can generate grammars for a set of training examples but the disadvantage is their potential heavy computational overhead. The proposed incremental grammar learning method offers immediate learning of new examples by slight alteration of the existing grammar collections by measuring the changes needed to equalize two grammars. This minimal combination approach avoids overgeneralization and maintains the coverage of newly learnt and past examples without a need to repeat the training process. Our experimental results suggest that the proposed algorithm outperforms a fuzzy grammar learning algorithm implemented in normal supervised learning setting.*

Keywords- evolving fuzzy grammar, grammar fragments, incremental learning, grammar similarity, tagging, text mining.

1 Introduction

The "digital obesity" phenomenon is a pressing problem at both the individual and corporate level [1]. This is partly because information is recorded using different formats for similar purposes and according to different hierarchies and categories. However, it is likely for the objects placed within similar categories to have similar properties and with some degree of correspondence.

These information sources often contain text segments with additional structure that can be utilized to assist in the identification of matching items. For example, within a text segment such as: 'Burger King, 12, Meredith Lane, Ipswich IP2 3AX Ipswich' lies an embedded structure: business name, number, street name, placename, post code and county name while a brief comment such as: 'the Voyager has stopped receiving signal' contains information on the product name and problem description.

Works on lexical semantics and word disambiguation have studied text structures by focusing on linguistic patterns using formal language models i.e in the subject-verb-object structure. However, these methods are normally well-suited for texts that are in complete grammatical form. On the other hand, text segments such as addresses, dates and times, names of products, people, as well as simple sentence forms such as questions, complaints, and news are frequently contained within "casual" text such as email, posts on a web forum, text messages, etc which often does not follow formal grammar rules.

Humans are able to use clues that identify structured segments without necessarily following a strict pattern. There can be many variations of the patterns and humans are normally better at seeing the answer in a specific case than being able to define a pattern that is universally applicable. As such, this is an ideal problem for machine learning; the key subproblem is converting small sections of texts into a more structured form by using grammar fragments to add tags. Note that we are not attempting to parse the entire piece of free text, merely to locate sections within it that can be tagged with an additional structure. For this reason, we refer to grammar fragments, as they are not intended to be a complete language model.

We define grammar fragment learning as the extraction of text fragments that comply with the predefined grammar. In contrast to in-depth whole sentence understanding, the grammar fragments learning focus only on portions of text that are relevant to specific domains. For example, UK address grammars might extract the following patterns:

a. Chicago Rock Cafe Northgate St Ipswich IP1 3BX Suffolk
b. Lattice Lodge Guest House 499 Woodbridge Rd Ipswich
c. 25 St Nicholas Street Ipswich Suffolk IP1 1TW

while grammar collections for posts on a telecom provider forum will recognize the following question entries:

a. Can I use an apple with BT Broadband Talk and Softphone?
b. When does BT plan to develop and support a Softphone client for NON windows platforms such as Mac OSX or Linux?
c. What is Softphone and how much does it cost?

Several authors have proposed the use of genetic algorithm [2,3,4,5], fuzzy [6,7,8], semantic net [9] and hybrid probabilistic [10] to learn grammars mostly from structured examples but these follow the traditional route of training a fixed system on a set of examples, followed by application of that system. Adaptation to new examples (when the training set is found to be inadequate) will be computationally expensive and hence inefficient. The training set needs to be sufficiently diverse to cover all unseen examples or the training phase will need to be repeated in order to improve the training set. Thus, it is highly desirable for a method that can support the incremental learning of a new example by slight modification of the learned rules without reconsidering past examples.

The purpose of this study is to investigate an evolutionary grammar fragment learning with *incremental* feature i.e. it can be updated each time a new pattern is encountered (or each time a string is read) without extensive re-computation as compared to other supervised learning methods. The evolved grammar ensures the ability to encompass the newly learnt, as well as the previously seen, examples. Main issues in evolutionary grammar learning are also highlighted which are defining the fuzzy grammar learning operators and its representation aspect. This algorithm will benefit

ISBN: 978-989-95079-6-8

works in intelligent text pattern recognition and its relevant applications such as intelligent agent and recommendation system.
The paper is organised as follows: the first part gives the introduction to incremental grammar learning while section two describes some of existing works relating to text pattern learning. The third section details the proposed evolutionary algorithm setting followed by observed results in section four. The final section summarises the investigated issues and some future works related to the study.

2 Related Works

Text annotation and text tagging are among the active text extraction research areas. Information tagging has been applied for sentence structure learning in question answering system [11] as well as specific domains including finding medical terms in legal text [12,13], detecting contact information and address [14] and gene and protein interaction information identification [15,16]. Among the popular techniques that have been employed are CFG, statistical, and evolutionary method.
Viola and Narasimhan [14] reported the advantage of marrying statistical natural language processing to the analysis of non-natural language text through the Discriminative Context Free Grammar (Discriminative CFG) approach which has extracted contact information more accurately than a similar conditional Markov model. The hierarchically structured CFG method [3,7,14] is more powerful than the formal English language model such as the subject-verb-object and noun-adverb-adjective model because these models have limitations when more complex sentence forms are involved.
The CFG method provides a rich collection of features that can measure and tag the properties of a sequence of tokens. Another advantage is that the CFG can propagate long range token dependencies efficiently. CFG method can also result in the detection of semantic information (encoded in the form of a parse tree) within a sentence - for example the word *fish and chips* in the restaurant address entry can be classified as broad as a type of business, or a food business genre, or even as specific as a food name.
Backus-Naur-Form (BNF) is a simple method for specifying CFG. It consists of a set of production rules with various operators such as "[ ]" which means "*this part is optional*", "|" for *or*, etc. A BNF grammar is usually more easily deciphered than a regular expression. A method for generating BNF grammars from data was described in [17]. As with regular expressions, a BNF grammar does not generally allow partial matching and there are no methods for finding the similarity between two BNF grammars.
An alternative approach to this is the fuzzy grammar similarity algorithm [2,3,9] which can measure the degree of similarity between grammars. The incorporated fuzzy approach supports the uncertainty and redundancy notion in deriving the grammar. The fuzzy grammar method was applied in evolutionary grammar learning using genetic algorithm.
However, in common with other supervised learning algorithms, the grammars generated using GA method [2,3,4,5,6,10], can only recognize data belonging to the trained pattern and needs to be rebuilt to include newly added patterns. Our preliminary study on GA [6] shows that GA is unfeasible because of the slow convergence towards creation of a good parsing grammar. The genetic operators and random concept in GA had also ignored the semantic properties (sequence of the tokens that bears meaning) in the dataset.
The minimal combination approach discussed in this paper is a revised version from the maximal combination approach [6] which avoids overgeneralization. On the contrary, [6] has generated a lot of unseen examples which has also led to low accuracy. The preliminary experiment on genetic algorithm in [6] has shown that genetic algorithm method is unsuitable for the grammar learning task because it cannot support grammar similarity and is also burden by the slow convergence. Work described in this paper concentrates on an alternative minimal combination approach and demonstrates that different permutations of the training set have no significant effect on incremental evolutionary grammar learning.

3 Brief Introduction to Grammar Fragment Learning

The fuzzy parsing concept allows for the learning of the pattern in the strings and can be seen as an extension of sequence matching, in which: i) we allow one sequence to be defined by a grammar, so that the matching process is potentially more complex and ii) we allow the standard edit operations (insert, delete, substitute) on the string in order to make it conform to the grammar.
It is adapted from the Cocke-Younger-Kasami (CYK) algorithm to allow for:

a. fuzzy grammar fragments that can contain fuzzy sets of terminal symbols;
b. calculation of the overlap between fuzzy grammar fragments;
c. partial parsing, measured by the proportion of a string that has to be altered by Levenshtein-style operations in order to make it conform to a fuzzy grammar fragment.

We identify a grammar fragment with the set of strings that are parsed by the grammar fragment. Because we allow partial parsing, this is a fuzzy set. The fuzzy membership, $\mu_G(Si)$ calculates the similarity between string and grammar, estimated by the minimum number of edit operations needed to convert a string into one which is parsed by the grammar. A string's membership decreases as it needs more changes to make it conform to the grammar.

$$\mu_G(Si) = 1 - \min\left(1, \frac{\min_{P \in Ext(G)} (totalCost(Si, P))}{length(Si)}\right)$$

We define a cost to be a five-tuple *(I D S Rs Rt)* where *I*, *D* and *S* are nonnegative real numbers representing, respectively, the approximate number of insertions, deletions, and substitutions needed to convert a string parsed by the source grammar into one that satisfies the target grammar. *Rs* and *Rt* represent sequences of grammar elements remaining (respectively) in the source and target after the match; at least one of *Rs* and *Rt* is null in every valid cost. Table 1 shows a simple example of grammar similarity comparison between a source and a target grammar in a UK postal address. In this case, *Cost(sg,tg)=(0 1 1 countyName Null)* to substitute *word* with *placename*, and leave *word* and countyName unmatched.

ISBN: 978-989-95079-6-8

Table 1: Example of Grammar Edit Distance Operation
(*I:Insert, D:Delete, S:Substitute, Rs: remainder of source grammar, Rt: remainder of target grammar)

Source grammar, sg	Number	Word	Word	Streetending	Placename	
Target grammar, tg	Number	Placename		Streetending	Placename	Countyname
Edit distance*	0	S=1	D=1	0	0	Rt=1

The *totalCost* function defines a total order on costs by $C1 \le C2$ iff $totalCost(C1) \le totalCost(C2)$.

$$\text{totalCost}((I, D, S, Rs, Rt)) = I + D + S + \text{minLength}(Rs) + \text{minLength}(Rt)$$

Addition of costs is order dependent and is defined by

$$(I1, D1, S1, Rs1, Rt1) + (I2, D2, S2, Rs2, Rt2) = (I3, D3, S1 + S2, Rs2, Rt2)$$

where

$$I3 = I1 + I2 + \text{minLength}(Rt1)$$

$$D3 = D1 + D2 + \text{minLength}(Rs1)$$

For two grammars, *GS* and *GT*, *Overlap(GS, GT)* is the approximate degree to which an arbitrary string parsed by the source grammar *GS* is also parsed by the target grammar *GT*. We define the degree of overlap of a source grammar GS with a target grammar GT as

$$\text{Overlap}(GS, GT) = 1 - \min(1, \text{totalCost}(C)/\text{minLength}(GS))$$

where $C = CostGG(GS,GT)$ is an estimate of the cost of changing a string parsed by the grammar *GS* into one parsed by the grammar *GT*.

The *CostGG* function is defined as follows, where S and T are arbitrary sequences of grammar elements, *s*, *t* are terminal symbols, and *TSi* and *TSj* are (fuzzy) sets of terminal symbols.

$$\text{CostGG}(\text{nu}\mathbf{l}, T) = (0, 0, 0, \text{null}, T)$$

$$\text{CostGG}(S, \text{null}) = (0, 0, 0, S, \text{null})$$

$$\text{CostGG}(s, t) = (0, 0, \delta(s, t), null, null)$$

where $\delta(s, t)$ measures the cost of the symbol *s* replacing symbol *t*. By default $\delta(s, t) = 1$, if $t \neq s$ and $\delta(s, t) = 0$, if $t = s$.

A given grammar has full overlap with itself or any equivalent grammar, where we define equivalence as the ability to parse exactly the same set of strings. Conversely, two grammars have zero overlap if there is no string that both grammars can parse. For example, if *G1* defines postal address and *G2* defines email address then they have zero overlap as there are no strings that both can parse.

The degree of overlap is not symmetric, as can be seen if we consider two grammar fragments *G3* and *G4*, where the set of strings parsed by *G3* is a strict subset of those parsed by *G4* . In this case,

$$overlap(G3, G4) = 1$$

but

$$overlap(G4, G3) < 1.$$

Clearly, the degree of overlap could be determined by repeatedly generating and checking candidate strings—such an approach would be very computationally demanding.

Instead, we propose a method [2,3,9] based on estimating the number of edit operations needed to change an arbitrary string parsed by the first (source) grammar sequence into a string that would be parsed by the second (target) grammar. This will result the coverage of strings belonging to both grammars and is implemented through the grammar combination technique.

The former *maximal combination* technique [9] allows combination of grammars without filtering their cost. E.g, for the set of strings $S = \begin{Bmatrix} (a-b-c-d-e-f), \\ (a-b-a-c-d-e-f), \\ (b-c-d), \\ (a-d-e-f) \end{Bmatrix}$ the learned grammar is $G' = \{[a]-[b]-[a]-[c]-d-[e]-[f]\}$. Although this method can parse all the strings belonging to S, but it also allows for the parsing of strings which does not exist in S. Therefore our current method focuses on minimal combination; combination of grammars that have a difference of 1 and this have managed to avoid overgeneralization.

3.1 Incremental Evolutionary Grammar Learning Algorithm

As the fundamental of the study is to look at the pattern similarities within the strings, a parser that can tokenize and extract relevant symbols from each input is essential. This grammar derivation process will re-label the symbols with significant grammar markup (Fig. 2 and Fig. 3).

The incremental learning has allowed for the update of the available grammars when new pattern is triggered without needed to repeat the whole learning process (Fig.1). The employed minimal combination is governed by the cost between the source and the grammar; combination will only be executed when the change is one. This process contributes to the evolution of the whole grammar collection and the persistence of the parsing coverage.

A chosen grammar from the existing collection will be altered according to the grammar combination rules. Table 2 shows guidelines of grammar combination operations; insert, merge and create. These operations are based on the cost of changing the source to the target and the other way

Table 2: Grammar Combination Operation

Source Grammar	Target Grammar	Cost (Source,Target)	Cost (Target,Source)	Combination Operation	Combined grammar
a-b-c	a-b	0 1 0 - -	1 0 0 - -	Insert	a-b-[c]
a-b	a-b-c	1 0 0 - -	0 1 0 - -	Insert	a-b-[c]
a-b-c	a-B-c	0 0 0 - -	0 0 1 - -	Merge	a-B-c
a-B-c	a-b-c	0 0 0 - -	0 0 1 - -	Merge	a-B-c, B>b
a-F-c	a-G-c	0 0 1 - -	0 0 1 - -	Create	a-X-c , X:=F||G, $F \neq G$

ISBN: 978-989-95079-6-8

around. The *insert* operation is executed when the change based in the cost is a deletion or an insertion (Fig. 5). The combined grammar is always the altered version (with optional tags) of the longer grammar. When cost needs a substitute action and when generalization can be performed, *merge* operation is executed while *create* is executed when generalization is not suitable (Fig. 4). Note that the grammar combination is sensitive to the matched position of the grammar pairs. The notation '||' is only used as a convenience of the paper and not practised in the system development.

```
Algo. 1: Grammar Evolution
Input: S={s0,s1,…,sn}
Output: GT
Procedure:
initialise GT to empty set
gs = grammar derived from s[0] (Algo.3)
add gs to GT
for each string, s [i:1,...,n] in S[n]
  gt=grammar in GT that parses s with min cost C (Algo. 2)
  if C > 1 then
    Add gs to GT
  else
    if C = 1 and cost(gt,gs)=0 then
      replace gt with combine(gt,gs) (Algo.5)
    else if C = 0 and cost(gt,gs)=1 then
      replace gt with combine(gt,gs) (Algo. 5)
    else if C = 1 and cost(gt,gs)=1 then
      replace gt with combine(gs,gt) (Algo. 4)
    end if
  end if
```

Figure 1: Grammar Evolution Algorithm

```
Algo. 2: minimum parsing cost, C
Input: string,s, grammar gt[0... n]
Output: minGrammar and C
Procedure:
Initialise C to cost(s, gt[0] )
Initialise minGrammar to gt[0]
for each  [i:1,...n]
  if( C > cost(s, gt[i]) then
      C = cost(s, gt[i])
     minGrammar = gt[i]
  end if
end for
```

Figure 2: Maximum Parsing Cost Algorithm

```
Algo. 3: Grammar Derivation
Input:string, s
Input: terminal symbols representing S,
S ∈ {word, placename, streetending, number,...}
Output: gs, the grammar derived for s
Procedure:
initialise gs to empty string
initialise maxMem and mem to zero
for each token in s, s[i], [i:0,...n]
maxMem= fuzzy membership score of s[i] with t[0]
  for each terminal symbols, t[j], [j:1,...m]
    mem= fuzzy membership score of s[i] with t[j]
    if mem>maxMem and t[j] is less general than t[j-1] then
      append the terminal token into gs
    end if
  end for
end for
```

Figure 3: Grammar Derivation Algorithm

```
Algo. 4: Combine For Substitute Operation
Input: gs,gt
Output: combinedGrammar, subGrammar
Procedure:
initialise combinedGrammar and subGrammar to empty
string
Initialize n=lengthOf(gs)=lengthOf(gt)
for i=[0,…n]
  if gs[i] is equal to gt[i] then   combinedGrammar[i] = gs[i]
  else
     create  subGrammar= gs[i], subGrammar=  gt[i]
     append subGrammar into combinedGrammar[i[
end for
```

Figure 4: Combine (Substitute) Algorithm

```
Algo. 5: Combine For Insertion or Deletion Operation
Input: gs,gt
Output: combinedGrammar, subGrammar
Procedure:
initialize combinedGrammar, subGrammar  to empty string
if gs.length>gt.length then loop=gs.length
else loop=gt.length
for [i:0…loop]
   if gs[i] is equal to gt[i] then
      append gs[i] to combinedGrammar
  else
      if gs.length>gt.length then
         append gs[i] with optional tag to combinedGrammar
      else loop=gt.length
         append gt[i] with optional tag to combinedGrammar
  end if
end for
```

Figure 5: Combine (Insertion or Deletion) Algorithm

The impact of the grammar learning when the training set is prepared in various orders is being observed. Fig. 6 shows an example of a collection of grammars while Table 3 shows the grammar generated when the grammars are learned in the different orders as follows:

Order1: G_2-G_8-G_3-G_5-G_4-G_7-G_6-G_1.
Order 2: G_6-G_3-G_2-G_8-G_4-G_7-G_5-G_1
Order 3: G_2-G_5-G_3-G_8-G_4-G_7-G_6-G_1

Our observation has shown that the final set of evolved grammars can parse the same set of strings, with the same accuracy. Section 5 gives some remarks on this issue. However, further work is ongoing to investigate this in more detail from a theoretical and practical perspective.

$$G = \{G_1, G_2, ..., G_8\}$$
$$g_n = \{a, b, c, d, e, f, j, k\}, a > k > j$$
$$G_1 = w - n - w - se - pn - pc - cN$$
$$G_2 = w - w - n - w - se - pn - pc - cN$$
$$G_3 = w - w - acc - w - se - w - pn - pc - cN$$
$$G_4 = w - bt - w - se - pn - pc - cN$$
$$G_5 = w - w - n - w - se - w - pn - pc - cN$$
$$G_6 = w - w - acc - w - se - pn - pc - cN$$
$$G_7 = w - acc - w - se - pn - pc - cN$$
$$G_8 = w - w - bt - w - se - pn - pc - cN$$

Figure 6: Examples of training grammars

ISBN: 978-989-95079-6-8

Table 3: Evolved Grammars in Various Training Data Orders

Order 1	Order 2	Order 3
w-w-(bt||n)-w-se-pn-pc-cN ($G_2 \cup G_8$)	w-w-acc-w-se-[w]-pn-pc-cN ($G_6 \cup G_3$)	w-w-n-w-se-[w]-pn-pc-cN ($G_2 \cup G_5$)
w-w-(acc||n)-w-se-w-pn-pc-cN ($G_3 \cup G_5$)	w-w-(n||bt)-w-se-pn-pc-cN ($G_2 \cup G_8$)	w-w-acc-w-se-w-pn-pc-cN (G_3)
w-(bt||acc)-w-se-pn-pc-cN ($G_4 \cup G_7$)	w-(bt||acc)-w-se-pn-pc-cN ($G_4 \cup G_7$)	w-w-bt-w-se-pn-pc-cN (G_8)
w-w-acc-w-se-pn-pc-cN (G_6)	w-w-n-w-se-w-pn-pc-cN (G_5)	w-(bt||acc)-w-se-pn-pc-cN ($G_4 \cup G_7$)
w-n-w-se-pn-pc-cN (G_1)	w-n-w-se-pn-pc-cN (G_1)	w-w-acc-w-se-pn-pc-cN (G_6)
		w-n-w-se-pn-pc-cN (G_1)

4 Results and Discussion

The underlying ideas of the experiment are to examine approximate parsing, fuzzy overlap of grammar fragments, and their use in the incremental evolution of fuzzy grammars. To demonstrate the benefit of incremental evolution an experiment was carried out to compare the parsing scores of the grammars generated through fuzzy grammar learner (FG) versus the proposed incremental evolutionary learning grammar (IEG).

The term incremental is coined so that the grammar will always be able to parse all supplied string patterns (identified by grammars) including the triggered new string pattern. This simulates interaction with a human expert who would be able to correct errors—although clearly this disadvantages the FG method, which is not designed to undergo multiple training phases. On the other hand, the FG method is performed according to the normal supervised learning setting; the performance of the training dataset is measured by its parsing on the test dataset. No additional learning is allowed even the parsing score is lower than 1. Table 4 shows the terminal sets used in this experiment.

Table 4: Terminal Sets

Word (w)	Any token composed entirely of alphabetical characters
Number (n)	Any token composed of digits
Streetending (se)	One of {road, rd, avenue, ave, street, st, lane, ln} or similar
Placename (pn)	A handcrafted set of UK placenames
CountyName (cN)	A handcrafted set of UK district names
BusinessType (bt)	One of {café, restaurant, bar} or similar
Accommodation (acc)	One of {hotel, lodge} or similar
Alphanumeric (aN)	A token composed of alphabetical characters and digits

Two sets of address dataset in UK format are prepared; restaurant.xml and yellowpages.xml. The restaurant dataset contains a small set of restaurant addresses in the UK while the yellowpages dataset contains domestic addresses. Each dataset is split into 25% for training and 75% for testing. The restaurant dataset contains 85 entries while the yellowpages dataset contains 396 entries. Both have no repeated entries but naturally consist of similar or repeated string patterns. This experiment was repeated using three different orders of the training and test dataset, to test the validity of the results and the sensitivity and reliability of the algorithm. The same result was produced in the different data permutations. Future work on this study will focus on the formal theoretical proving of order independent learning to support this hypothesis.

Table 5 shows the comparison between the mean of maximum parsing score of the grammars learned. The average of exact parsing is a measure of the average of the maximum parsing score of the grammars on each string in the test dataset. Score of 100% means that all strings have been recognized by the grammar.

The IEG method generates new rules immediately and shows good incremental learning behavior as the grammar fragments evolve; meaning that the grammar fragments may be altered according to the change required to transform the source and target grammars so that they will be equally parsed with increased coverage. Thus, the IEG method will always give an optimal parsing score of 1 for each string it has encountered, as shown in Table 5. For each string in the dataset, there is either a grammar that can parse it partially or accurately; and a grammar can thoroughly parse a minimum of 1 string. This indicates that for every string in the dataset, there is at least a grammar that can parse it accurately (100% parsing). As opposed to the IEG method, the average of exact parsing in FG is much lower. Observation in the restaurant dataset shows that the minimum parsing score of the entries in FG is 0.667 and 0.75 in the yellowpages dataset. Both share a maximum of score=1.

As discussed in section 2, supervised learning method is constrained by its training stage; the training dataset should contain all the patterns expected to be seen in the test dataset. As shown in Table 5 the performance of the grammars generated in FG is lower than of IEG due to its insufficient training. The FG score in yellowpages dataset is higher than the restaurant dataset because the yellowpages dataset is larger than the restaurant dataset so during training a better coverage has been provided.

Table 5: Average of Exact Parsing

Setting	Dataset	Average of Exact Parsing (%)
FG	Restaurant	89.5
IEG	Restaurant	100
FG	Yellowpages	97
IEG	Yellowpages	100

This shows that the evolutionary learning method has successfully provided immediate and gradual learning to the grammars even when the training is found to be insufficient. The set of evolved grammars have successfully maintained the parsing coverage of the existing and newly combined

ISBN: 978-989-95079-6-8

patterns. The goal has been achieved where the adaptations are performed in light of new examples while preserving its current knowledge.

5 Summary and Conclusion

Incremental evolutionary grammar learning (IEG) is a method that can cope with newly found text pattern. It is an advantage over the expensive retrain-retest setting in traditional machine learning approaches when examples in the training phase are found to be insufficient. Empirical result shows that the IEG method has 100% parsing of the dataset compared to 89.5% and 97% achieved by the fuzzy grammar method. The incremental feature in IEG has also surpassed the genetic algorithm method both in the processing time and accuracy.

The underlying structure (grammar) of the text is first detected and corresponding grammar that represents the text is derived. The novel fuzzy grammar similarity is utilized to find similarity of the grammars. The evolution concept featured in this algorithm is reflected by the enrichment of the learnt patterns; where its incremental learning characteristic gradually updates its knowledge on the various patterns. Each time a new pattern is encountered, the adaptation is executed by a minimal adaptation which provides a *fit* coverage corresponding to the newly found and the altered grammar as well as all the strings that they belong. Over time, the set of evolved grammars become more concise, wider coverage yet sustaining its persistence.

Further work is ongoing to study the theoretical proving of the order-dependency effect on the evolved grammars, investigating a method to create more compact grammars and to expand the application of fuzzy grammar fragments to softer structures (e.g., identification of questions, negative comments, positive reviews, etc.).

References

[1] Martin, T. P. "Fuzzy sets in the fight against digital obesity," *Fuzzy Sets Syst.*, vol. 156, pp. 411–417, 2005.

[2] Petasis., G., (2004) eg-GRIDS: Context Free Grammatical Inference from Positive Examples using Genetic Search, International Colloquium on Grammatical Inference 2004, *Lecture Notes in Artificial Intelligence* 3264, pp. 223-234.

[3] Sakibara, Y., and Marumatsu. H., (2000) Learning Context-free Grammars from Partially Structured Examples, International Colloquium on Grammatical Inference 2000, *Lecture Notes in Artificial Intelligence* 1891, pp. 229-240.

[4] Smith, T.,C., and Witten, H., (1996) Learning Language Using Genetic Algorithm, *Lecture Notes in Computer Science* 1040 pp. 132-1445.

[5] Unold, O., and Cielecki, L.,(2007) Learning Context-Free Grammars from Partially Structured Examples: Juxtaposition of GCS with TBL, *Seventh International Conference on Hybrid Intelligent Systems*, pp. 348 – 352

[6] Mohd .S., N., Martin, T., Shen, Y. (2008) Incremental Evolutionary Grammar Fragments, *Proceedings of the UK Workshop on Computational Intelligence (UKCI 2008)*, pp. 89-94.

[7] Martin, T. and Azvine, B. (2006) Evolution of Fuzzy Grammars to Aid Instance Matching,. IEEE *International Symposium on Evolving Fuzzy Systems,* pp. 163 – 168.

[8] Martin, T, Shen., Y., Azvine, B. (2008) Incremental Evolution of Fuzzy Grammar Fragments to Enhance Instance Matching and Text Mining, *IEEE Transactions On Fuzzy Systems,* Vol. 16, No. 6, December 2008, pp. 1425-1438.

[9] Yuhua L., David M.L., Zuhair A. B., James D. O., and Keeley Crockett (2006) Sentence Similarity Based on Semantic Nets and Corpus Statistics,. *IEEE Transaction on Knowledge and Data Engineering*, Vol 18, No. 8 pp.1138-1150.

[10] Bosman, P. A. N. (2004) Learning Probabilistic Tree Grammars for Genetic Programming, Parallel Problem Solving from Nature - PPSN VIII, *Lecture Notes in Computer Science* Vol. 3242 pp. 192-201.

[11] Ravichandran, D., and Hovy, E.(2002) Learning Surface Text Patterns for a Question Answering System, *Proceedings of the 4th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 41-47.

[12] Dozier, C., et. al (2007) Fast Tagging of Medical Terms in Legal Text, *Proceedings Of The 11th International Conference On Artificial Intelligence And Law*, pp. 253-260.

[13] Serban, R., et al (2007). Extraction and Use of Linguistic Patterns for Modelling Medical Guidelines, *Artificial Intelligence in Medicine (2007) 39*, pp. 137-149.

[14] Viola, P. and Narasimhan, M., (2005) Learning to Extract Information from Semi-structured Text using a Discriminative Context Free Grammar,. *8th Annual International ACM SIGIR Conference*.

[15] Joshua, M. T, and Gilder, M. R., (2003) Extraction Of Protein Interaction Information From Unstructured Text Using A Context-Free Grammar, *Bioinformatics 19(16) Oxford University Press.* Vol. 19. No. 16. pp. 2046-2053.

[16] Tanabe, L., and Wilbur, W. J., (2002). Tagging Gene and Protein Names in Biomedical Text, *Bioinformatics Vol. 18, no. 8 2002*, pp. 1124-1132

[17] Mernik, M., Gerlic, G. , Zumer, V. and Bryant, B. R. (2003) Can a Parser be Generated from Examples, *Proceedings of the 2003 ACM Symposium on Applied Computing,* pp. 1063 –1067

ISBN: 978-989-95079-6-8

Managing imperfect temporal metadata in the Catalog services of Spatial Data Infrastructures compliant with INSPIRE

Gloria Bordogna[1] Paola Carrara[2] Marco Pagani[1] Monica Pepe[2] Anna Rampini[2]

1 CNR-IDPA,
Unità di Dalmine (Bg), Italy
2 CNR-IREA
Unità di Milano, Italy
Email: {gloria.bordogna, marco.pagani}@idpa.cnr.it, {carrara.p, pepe.m, rampini.a}@irea.cnr.it

Abstract—*In this paper we analyze the limitations of current recommendations of the INSPIRE (Infrastructure for Spatial Information in Europe) Directive as far as the temporal metadata definition for discovery purposes, and propose its extension so as to allow the representation and management of imperfect spatio-temporal metadata. We propose to extend the metadata in order to cope with the requirements of both metadata producers, who often are unable to specify precise values, and users who submit queries to catalog services for discovering interesting data, who may express soft selection conditions on metadata values. The proposal is illustrated and explained through an example taken from an active Spatial Data Infrastructure (SDI).*

Keywords— Geo-data, INSPIRE, metadata, Spatial Data Infrastructure, spatio-temporal imperfection.

1 Introduction

Infrastructures are complex systems in which a network of interconnected but autonomous components is used for the exchange and mobility of goods, persons, information. Their successful exploitation requires technologies, policies, investments in money and personnel, common standards and harmonised rules. Typical examples of infrastructures which are critical for society are transportation and water supply. In Information Technology, the term infrastructure could be related to communication channels through which information can be located, exchanged, accessed, and possibly elaborated. Since the last decade of the past Century, creation of infrastructures for spatial information emerged as an issue in some countries in order to prevent wasting of money in geo-data multiple creation, to favour their shares at multiple levels in the society, and to support decision making in various fields, such as the environment [1]. In 2007, the INSPIRE Directive of the European Parliament and of the Council entered into force [2] to trigger the creation of a European Spatial Data Infrastructure (ESDI) that delivers to the users integrated spatial information services*. These services should allow users to discover and possibly access spatial or geographical information from a wide range of sources, from the local to the global level, in an inter-operable way for a variety of uses. Discovery is performed through services that should follow INSPIRE standards and can be implemented through some products (either proprietary or not) that declare their compliance. Users' conditions, expressed through discovery service clients, are matched against archived metadata, describing associated geo-data, in order to give an answer to the above conditions. The matching result is a list (sometimes empty) of records of metadata satisfying the match. Discovery services currently provide an exact matching: users cannot express flexible selection conditions, enabling partial matching mechanisms between the ideal metadata and the archived metadata.

On the metadata side, their creation is usually in charge of data/services providers: metadata fields, their meaning, and their types/ranges of values are defined in INSPIRE through recommendation documents aimed at easing discovery [3], [4]. With respect to the temporal characterisation, current recommendations for metadata specification are inadequate for metadata providers: some needed fields are missing; the semantics of the recommended fields is ambiguous; they do not consider nor manage representation of uncertain or vague information due to incomplete knowledge.

In this paper we take into account this last drawback and propose to let imperfect temporal values be used by both metadata providers in describing the temporal validity of the geodata, and SDI users in expressing discovery conditions. Moreover we suggest mechanisms to introduce partial matching in discovery services.

2 Metadata extension proposal within SDI architecture of INSPIRE

2.1 SDI Architecture

An overview of the current understanding of the technical architecture of INSPIRE is depicted in figure 1; it has been adopted by the ESDI [5].

Discovery services are the element aimed at discovering information of interest to the users. They are connected to metadata catalogues, necessary due to the intrinsically distributed nature of an SDI, where information repositories

* Following the Directive, components of European SDIs include: metadata, spatial data themes (as described in Annexes I, II, III of the Directive), spatial data services; network services and technologies; agreements on data and service sharing, access and use; coordination and monitoring mechanisms, processes and procedures.

ISBN: 978-989-95079-6-8

are owned and maintained on the Web servers of different organisations. Services managing catalogues could have the twofold function of archives of metadata, which describe either geo-data or geo-services available in the SDI, and of discovery tools, able to match users' search conditions with archived metadata. Catalogues services should be fitted to let users express selection conditions that specify "what" is of interest through content keywords (e.g., landslides), "why" they are searched (e.g., to detect recurrence of landslides), "where" the interesting features should be located (e.g., in a bounding box surrounding the Alps), and "when" these features should have been observed (e.g., the date/s of the observations).

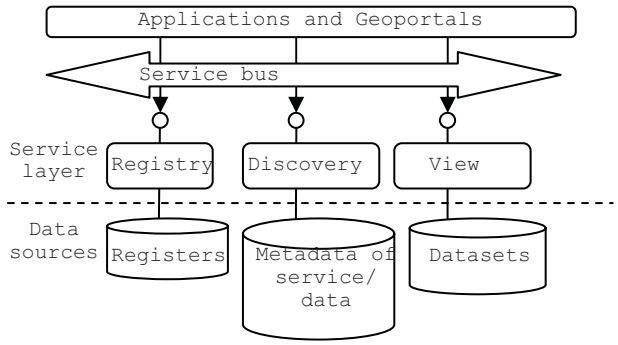


Figure 1: INSPIRE technical architecture overview; discovery service and metadata

2.2 Temporal metadata in INSPIRE SDIs

In order to ensure that the SDIs participating in the ESDI are compatible and usable in a trans-boundary, multicultural context such as Europe, the INSPIRE Directive requires that common Implementing Rules (IR) are adopted in a number of specific areas, in particular in metadata definition. The INSPIRE Metadata IR Version 3 of 26th October 2007 [4] defines four elements of temporal information, i.e.:

1. *Temporal extent*, the time period covered by the content of the resource
2. *Date of publication* of the resource
3. *Date of last revision* of the resource if the resource has been revised
4. *Date of creation* of the resource if it has not been revised

Only the first element (Temporal reference-5.1 Temporal extension) can be used to describe a temporal aspect of the content of geo-data itself, since the following three elements are referred to geo-data life cycle as a document. Nevertheless it plays an unavoidable role in numerous use cases, such as searching for events occurring in a particular period of time, looking for multiple observations of the same geographic area, evaluating if data can be used for a particular process, etc.

A specific study has been commissioned in order to improve temporal metadata specification [6]. It introduces some recommendations that rise attention to the role of 'Temporal extent' element; specifies the type and format of time values to be adopted either in expressing dates or time periods; suggests how to preserve precision in exchange or conversion processes.

2.3 Proposal of extension of temporal metadata

In our opinion and based on the experience as partners in the European Project IDE-Univers (http://www.ideunivers.eu), some further extensions should be adopted to improve temporal characterisation of geo-data in an SDI. In fact, the current specification of INSPIRE does not allow to specify the time relative to the occurrence of the event that is observed in one, several or a series of spatial data. Furthermore, current formats do not allow to specify imperfect values for metadata, both temporal, spatial, and thematic, which are the only kind of values that metadata producers can sometime provide, given that they often do not have enough information to specify precise values.

For example, if one has to create the metadata of a series of remote sensing images of a landslide that occurred in 1986 taken in distinct periods during its monitoring after its occurrence, the current specification just allows to indicate the time of the observations and not the time of the event, i.e. of the landslide occurrence. Nevertheless, this can be a very important information in an SDI context, because one can be interested in comparing the status of a landslide in different periods of time in the same place, as they appear in distinct images. One then, must be sure that the compared images refer to the same event, and not to distinct ones.

Moreover, in real situations, it often happens that the date or time of occurrence of an event is ill-known. This is due to several reasons: either because the event was not observed during its occurrence, and thus it can only be guessed by experts, or because it occurred long time ago so that its date can be deduced only imprecisely by the dating methods available (e.g., this is the case or archeological sites, geologic and paleontological findings).

In summary we propose:

- To introduce the possibility to include in metadata one or more events/processes/phenomena of reference for the geo-data: a satellite image can be an observation of fires in a region; a set of meteo records are measures of a rainfall; some thematic maps can be subsequent representations of urban growth, etc. A temporal extent element should be defined also for the reference event(s), of course.
- To allow the definition in metadata of soft temporal values in the temporal extent element of both geodata and related event(s); soft values should be allowed in the expression of single time points (by ex. the landslide occurred in the night of 15th December, 2005), of time intervals (by ex. snow precipitation of three days starting from 14th December 2008), and of subsequent either periodic or aperiodic dates/intervals (ex. fires occurring each Summer).

ISBN: 978-989-95079-6-8

3 Basic framework of the proposal

3.1 Extension of temporal metadata in INSPIRE

Notwithstanding we are aware that further extensions are possible, in this contribution we propose to include the following temporal metadata:

Instant in which the observation/event occurred:

- Date, time (e.g. the landslide occurred the 11-07-2008 at 8:00:00)

Period of validity, or duration of an observation/event:

- period (interval of dates, times) (e.g. the duration of a fire was from 11-07-2008 to 13-07-2008)

Sequence of instances of occurrences/events:

- multiple instants of time or dates (e.g. the dates of the distinct fires were 11-07-2008 , and 12-07-2008)

Sequence of periods of validity or duration of occurrences/events:

- multiple periods or durations (e.g. the times of the distinct fires were from 11-07-2008 to 13-07-2008 and from 20-08-2008 to 23-08-2008)

We also extend the time metadata format so as to allow the specification of ill-known instants, intervals and time series by means of fuzzy instant, fuzzy intervals and fuzzy time series associated with any kind of observation/event.

First of all we describe a formal framework in which all these fuzzy temporal indications are modeled, adopting the proposal of [7]. In order to express such fuzzy temporal indications into an eXtended Markup Language (XML)-type language, that is required for INSPIRE metadata, we adopt TimeML specification language [8].

The use of TimeML is motivated by the fact that it is a textual meta language, thus easy to read and to index by common Information Retrieval techniques, that can be employed in a catalog service context in order to represent the metadata contents for a successive search and discovery. It is enough flexible to allow the annotation (description) of the kind of event/observation and its temporal information, possibly imprecise and vague.

In the following subsections, let us first describe the representation within fuzzy set theory and possibility theory of time expressions, then introduce TimeML and specifically the tags we adopt, and finally, the partial matching mechanism we propose.

3.2 Modeling flexible time indications within the fuzzy set and possibility framework

The representation of temporal information requires the choice of a numeric scale to describe the order of events and phenomena; this scale is defined by an origin (time t=0) and a time unit, so that events occurred before t=0 have a negative time values and those occurred after have positive values. The choice of the time unit depends on the cycles of an observed periodic phenomenon, such as the movement of the sun. Nevertheless almost every time scale shares the basic unit *second*, and other units with lower granularity such as *minute*, *hour* and *day*. Besides these units of time, there are other natural language time indications such as *week*, *month*, *season*, *year*, *century* that are artificial notions, defined within a calendar such as the Julian and Gregorian ones. Other notions of time, related to human experience such as *now*, *soon*, *recent*, *often*, *end of the year*, are approximate time indications.

Several time models have been proposed [9], their common limitation is that they deal with crisp time indications.

A flexible framework has been defined within fuzzy set and possibility theory to express approximate hierarchical time indications close to natural language at distinct level of granularity [7]. It takes inspiration from the time hierarchy proposed in [10] and the time granularity defined as a mapping from the set of positive integers to the absolute time proposed in [11], and builds up a multi-granular hierarchical time structure in which also vague and imprecise time granules can be defined.

We assume this approach at the basis of our proposal. In this way it is possible to express temporal indications in several time units with distinct granularities, the less refined ones obtained by grouping units of the higher level granularity.

A *basic domain* G_0, consists of granules of time points below which the elements are not discernable. For example if G_0 is *hour* we cannot discern the *minutes* within an hour. Notice that, a granule of a domain G' (e.g. G'=*week*) that is not the basic domain G_0 (e.g. G_0=*hour*) can be defined by recursively grouping granules of its parent domains G=*day* and G_0 in the hierarchy. For example G'=*week*=7**day* =7*24*G_0=168*hours*.

The set of temporal specifications that we adopt are listed here following.

A *time point* indication is defined as a pair [t, G] in which t is an ordinal indicating the position with respect to the time origin on a domain G of granules. An example is: [t=2, *day*] that indicates the second day from the time origin; a fuzzy example is [t={0.8/3, 1./4, 0.7/5}, *day*] that means around the fourth day after the time origin.

A duration in time, i.e. a *time span*, is a pair [Δt, G] and can be denoted by either a set or a range of time points. A fuzzy time span example is [Δt ={0.8/3, 1./4, 0.7/5}, *year*] that means a duration of about 4 years.

A temporal distance from the origin, i.e. a *time distance*, is defined as a pair [d, G] in which d is a positive or negative value, indicating the distance in time granules on G from the origin. In this case [d=2, *day*] means two days after the origin. As t, also d can be a fuzzy set indicating a fuzzy time distance.

A *time interval* is a triple [t, Δd, G]; in a crisp case [t=1991, Δd=3, *year*] means 3 years from 1991.

A *composite span* is a union of spans $\cup[\Delta t_i, G_i]$, not necessarily adjacent and on the same basic domain G.

An *aperiodic time element* is a union of time intervals $\cup[t_i, \Delta d_i, G_i]$. The crisp example [t=1-11-2008, Δd=28, *day*] ∪ [t=30-11-2008, Δd=31, *day*] means 28 days from 1-11-2008 and 31 days from 30-11-2008.

A *periodic time element* is a union of a time interval and a time distance: $\cup[t_i, \Delta d_i, G_i], [d_k, G_k]$. By example ∪[t=1-8-2000, Δd=31, *day*], [d=1, *year*] means every August from

ISBN: 978-989-95079-6-8

year 2000. An example of vague periodic time element is ∪ [t=1-2000, Δd={0.2/1, 0.8/2, 1./3, 0,8/4}, *week*], [d=1, *year*] that means around the third week of every January from year 2000.

Since in the context of metadata compilation we may have time series that are related to finite repetitions of observations or events a *finite periodic time element* is defined as a composition of a periodic time element and a time point: $\cup[t_i, \Delta d_i, G_i], [d_k, G_k], [t, G]$ in which the time point t specifies the end of the repetition. An example of finite periodic time element is "*every autumn from 2000 to 2005*" that is formally expressed as: ∪[22-09-2000, 90, *day*], [1, *year*], [21-12-2005, *day*].

3.3 TimeML specification language for flexible time indications

TimeML is a markup language of the XML family for describing events, signals and temporal expressions into a text written in natural language. It is designed to address four situations:

1. Time stamping of events (identifying an event in time, instant or interval of validity).
2. Ordering events with respect to one another (relative ordering).
3. Reasoning with contextually underspecified temporal expressions (temporal functions such as 'last week' and 'two weeks before').
4. Reasoning about the persistence of events (how long does an event or the outcome of an event lasts).

The tags in TimeML[†] that we adopt to extend the INSPIRE metadata and that we model within the fuzzy framework previously described, are listed and illustrated in Table 1.

Table 1: TimeML tags adopted in this proposal

EVENT
Tag used to annotate the semantics of the event described. Syntactically, EVENT can be a verb (such as "raining"), but also a nominal, such as "fire"
MAKEINSTANCE
It indicates different instances (observations) of a given event. Different instances can have different attribute values, and every EVENT introduces at least one corresponding MAKEINSTANCE
TIMEX3
This tag is central to our objectives since it is primarily used to mark up explicit temporal expressions, such as times, dates, durations, etc. TIMEX3 allows marking up the following types of temporal indications specified by the attribute `type`: *Durations* such as "four weeks", "half a year"; *Calendar dates* (points in time equal or bigger than a day) both precise such as "13 August 2007" and imprecise or vague such as "few days ago", "end of July", "at the beginning of summer"; *Times of day* (smaller than a day) both precise such as "at 9.50.00 a.m." and imprecise or vague such as "before noon"; *Sets* (Recurring time expressions) such as "every month", "twice a week" The `value` attribute can assume XML datatypes based on the 2002 TIDES guideline, which extends the ISO 8601 standard for representing dates, times, and durations. E.g. "twelve weeks" becomes "P12W" and "21 February 2008 at 8.30.00 a.m." becomes "2008-2-21T8:30:00" The `mod` attribute allows specifying temporal modifiers that cannot be expressed either within `value` proper, or via links or temporal functions, such as "before", "after", "equal or less", "end".
TLINK
one of the TimeML link tags which encodes the relations that exist between two temporal elements (e.g., `BEGINS, HOLDS, INCLUDES, AFTER`)

For example the expression "*every autumn from 2000 to 2005*" is formulated in TimeML as follows:

```
<TIMEX3 tid="t10" type="SET" value="R6/2000-09-
22/P1Y0M0D">
every autumn from 2000 to 2005
</TIMEX3>
```

Finally, in TimeML it is possible to mark confidence values to be assigned to any tag and to any attribute of any tag. The confidence value associated with the `value` attribute of TIMEX3 expresses the uncertainty that the metadata provider has in assigning the temporal indication to an event or observation. For example, we can add the confidence annotation to TIMEX3 so as to specify the uncertain date of an observation that might have occurred the first or second of January 2000 as follows:

```
<TIMEX3 tid="t1" type="DATE" value="2000-01-01">
On January 1st, 2000
</TIMEX3>
<CONFIDENCE tagType="TIMEX3" tagID="t1"
confidenceValue="1."/>

<TIMEX3 tid="t2" type="DATE" value="2000-01-02">
On January 2nd, 2000
</TIMEX3>
<CONFIDENCE tagType="TIMEX3" tagID="t2"
confidenceValue="0.80"/>
```

4 Management of the proposed temporal metadata extension

The frameworks and assumptions of section 3 can be used to define the temporal fields associated with events and observations in the metadata of a SDI catalogue.

The metadata provider defines the time indications of events and observations by means of a metadata editor. We propose an example of the metadata of two thematic maps representing subsequent observations from satellite of the same event, i.e. the melting of the Lys Glacier (a glacier of the Italian Alps), during Summer 2007. In this example, the temporal indications of the occurrence of the observations

[†] TimeML vers. 1.2.1 <http://www.timeml.org>

ISBN: 978-989-95079-6-8

are known precisely. Following the extension we propose, in the metadata of the first map we could have fields such as:

Metadata 1
…
Event="Lys Glacier melting"
Occurrence="observation of Lys Glacier melting"
Time Position of Occurrence="1.7.2007"
…

In the metadata of the second map we could have:

Metadata 2
…
Event="Lys Glacier melting"
Occurrence="observation of Lys Glacier melting"
Time Position of Occurrence="3.9.2007"
…

Metadata are translated into TimeML sentences like in Table 2.

To allow partial matching with respect to flexible selection conditions specified by a user a parser translates the external TimeML definitions of temporal metadata into their internal fuzzy set representation μ_t. We can have fuzzy sets defined on distinct domains (G). The hierarchy of basic temporal concepts defining the temporal domains with distinct granularity, is represented by a graph that contains in each edge connecting node i to node j a mapping function $F_{i,j}$ that allows converting the current granule of the i-th node in terms of aggregation of granules of the j-th node (see Fig 2 which is a simplification to illustrate how the concept works).

This function $F_{i,j}$:G'→G, with G⊂G', and i,j identifiers of the i-th and j-th nodes, with i-th node defined on domain G' and j-th defined on G, associates a granule g'∈G'→$F_{i,j}$(g')∈G, where $F_{i,j}$(g') is a fuzzy set of granules on G. A temporal indication *t*, defined on G' (e.g., 2002 year) can be converted on another domain G (e.g., day) by repeatedly applying the mapping functions $F_{i,j}$ associated to the edges on the path from node G*'* to node G as proposed in [12].

Table 2: TimeML example

Lys Glacier
<EVENT eid="e10" class="OCCURRENCE">
is melting
</EVENT>
<MAKEINSTANCE eiid="ei1" eventID="e10"
pos="VERB" tense="PRESENT" aspect="PROGRESSIVE"
/>
<TIMEX3 tid="t2" type="DATE" value="2007-07-01">
On July 1st, 2007
</TIMEX3>
<TLINK eventInstanceID="ei1" relatedToTime="t2"
relType="DURING"/>
<MAKEINSTANCE eiid="ei2" eventID="e10"
pos="VERB" tense="PRESENT" aspect="PROGRESSIVE"
/>
<TIMEX3 tid="t3" type="DATE" value="2007-09-03">
On September 3rd, 2007
</TIMEX3>
<TLINK eventInstanceID="ei2" relatedToTime="t3"
relType="DURING"/>
…

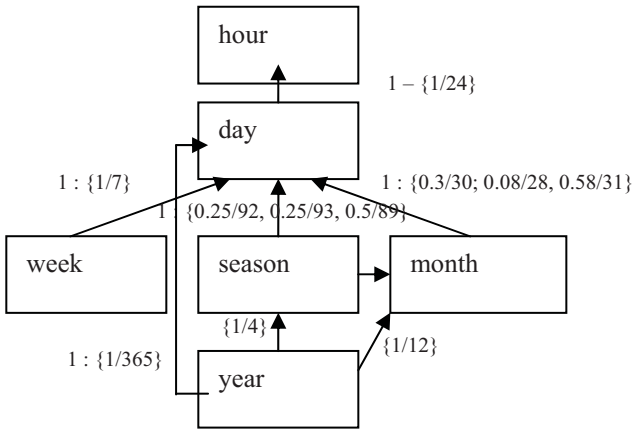


Figure 2: Simplified example of temporal graph: on each edge the fuzzy set defining the conversion function F of granules.

However, since there can be more than a single path P_1, ..P_k, connecting two nodes (e.g., P_1=*year*, *season*, *day* and P_2= *year, month, day*), multiple definitions of *t* on the domain G can be obtained t_{P1}, .. t_{Pk} In order to reconcile these definitions we choose the one obtained by considering the shortest path connecting the two nodes. If there exists more than one of such short paths with same length we generate the maximum bound of their definitions:

$$t_{P1 \oplus P2 \oplus Pk}(g) = \max(t_{P1}(g) .. t_{Pk}(g))$$

The reason for this choice is that the maximum bound comprehends all possible definitions of the temporal indication on the domain G.

For example, the sentences in Table 2 correspond to the two *time points* [t_1=1-7-2007, *day*], and [t_2=3-9-2007, *day*], with crisp pointwise membership functions.

On the other side, the user specifies her/his temporal soft selection conditions Q within a catalog service interface. They can be expressed by specifying soft constraints, i.e. with a desired membership function μ_Q, defined on a time line with a given granularity, chosen among one of the available in the temporal hierarchy (e.g. see fig. 3). An example of user selection condition for the Lys Glacier melting case reported in table 2 could be: *"Search for occurrences of glacier melting observations, occurring close to late Summer 2007"*.

Users' temporal query specifications are converted into the internal fuzzy set representation and, if necessary, transformed into the granularity of the data they have to be matched. The soft constraint of the example corresponds to a vague *time interval* with trapezoidal membership function μ_Q such as [t=15-6-2007, Δd={ 1-8-2007, 15-8-07, 23-9-2007, 10-10-2007}, *day*].

The internal fuzzy representation of the temporal metadata μ_t are then matched with the fuzzy representations of the soft query constraint μ_Q by applying a representation-based matching function as proposed in [13][14].

In this representation-based framework, both the metadata values μ_t, possibly uncertain, and the soft query condition μ_Q are interpreted as soft constraints and one can match

ISBN: 978-989-95079-6-8

them to obtain a degree of satisfaction $RSV(t,Q) \in [0,1]$, by computing either a measure of similarity [14] or a fuzzy inclusion measure [13] between the two fuzzy sets μ_Q and μ_t.

In the specific case of matching temporal constraints, it makes sense to allow the choice of both matching functions. One could be interested in selecting observations of an event taken in a date *close* to another date. This corresponds to select the similarity matching function [14]:

$$RSV(t,Q)=Similarity(\mu_t,\mu_q)=\frac{\sum_{i\in G} min(\mu_t(i),\mu_Q(i))}{\sum_{i\in G} max(\mu_t(i),\mu_Q(i))}$$

in which $\mu_t(i)$ and $\mu_Q(i)$ are the membership degrees of a time point i in a fuzzy metadata value t and in the query constraint Q.

Another case is when one wants to select observations that occurred *within* a period. This corresponds to select a matching function defined as a fuzzy inclusion [13]:

$$RSV(t,Q)=Inclusion(\mu_t,\mu_q)=\frac{\sum_{i\in G} min(\mu_t(i),\mu_Q(i))}{\sum_{i\in G}(\mu_t(i))}$$

Other matching functions could be defined corresponding with other temporal relations such as "*close before*" "*close after*" "*recent*" and so on.

The retrieved metadata can be ranked in decreasing order on the basis of *RSV* values, thus avoiding empty answers and suggesting an access order to the referred geo-data. In the example, both metadata are retrieved: Metadata 2 has a *RSV* score "1" being situated in late Summer, while Metadata 1 is also retrieved since it partially satisfies the query condition "*close*", thus meaning that it is associated to an observation of glacier melting that is in proximity of the limits of the temporal range 'late Summer'.

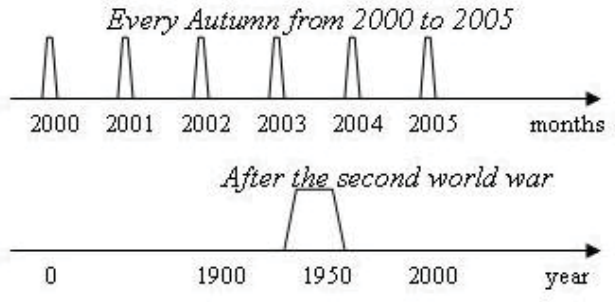


Figure 3: Examples of two soft temporal constraints defined on two timelines with distinct granularity (months and years, respectively). The constraint "*every Autumn from 2000 to 2005*" is defined as a fuzzy finite periodic time element, while "*after the second world war"* as a fuzzy time interval.

5 Conclusions

Spatial Data Infrastructures are becoming a common practice for discovering and accessing distributed heterogeneous geo-data. Nevertheless, the catalogue services on which SDIs are based are still founded on old database paradigms, that do not allow partial matching mechanisms, nor the representation and management of ill-defined metadata.

The present paper represents a first step in the proposal of a fuzzy framework to model catalogue services functionalities to manage ill-defined metadata and flexible queries. Specifically, we proposed a solution to express, represent, and manage temporal metadata, possibly imperfect, in a flexible way even if the revising policy within INSPIRE is long and needs to overcome several reviewing steps.

References

[1] Executive Order 12906, Coordinating Geographic Data acquisition and access: The National Spatial Data Infrastructure, edition of the U.S.A. Federal Register, Volume 59, Number 71, pp. 17671-17674, April 13, 1994

[2] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 (INSPIRE), online: <www.ec-gis.org/inspire/directive/l_10820070425en00010014.pdf>, 2007

[3] Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata, 2008

[4] Draft Implementing Rules for Metadata (version 3), European Commission, 2007-10-26, 2007

[5] INSPIRE Technical Architecture Overview, INSPIRE Cross Drafting Teams Report, 05-11-2007, online: < http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/INSPIRETechnicalArchitectureOverview_v1.2.pdf >, 2007

[6] M. Dekkers, Temporal Metadata for Discovery, M. Craglia (ed.), EUR 23209 EN, online: <http://inspire.jrc.ec.europa.eu/reports/TemporalMetadataforDiscovery>, 2008

[7] R. De Caluwe, F. Devis, P. Maesfranckx, G. De Trè, B. Van der Cruyssen, Semantics and Modelling of Flexible Time Indication, in Computing with words in Information/Intelligent Systems 1, Foundation, Zadeh L.A., Kacprzyk J., Physica Verlag, 1999

[8] TimeML, Markup Language for Temporal and Event expression, online:<http://www.timeml.org/site/index.html>

[9] V.J Tsotras., A. Kumar, Temporal database bibliography update, ACM SIGMOD Record, 25(1), 41-51, 1996

[10] R. Maiocchi, B. Pernici, F. Barbic, Automatic deduction of temporal indications, ACM Trans. On Database systems, 17(4), 647-668, 1992

[11] X.S. Wang, C. Bettini, A. Brodsky, S. Jajodia., Logical design for temporal databases with multiple granularities, ACM Trans. On database Systems, 22(2), 115-170, 1997

[12] R. De Caluwe, G. de Tré., B. Van der Cruyssen, F. Devos, P. Maesfranckx, Time management in fuzzy and uncertain object-orineted databases, in Knowledge management in fuzzy databases, O. Pons, M.A. Vila, J. Kacpzzyk (eds), Physica Verlag, 2000

[13] G. Bordogna, G. Pasi, A Flexible approach to evaluate soft conditions with unequal preferences in fuzzy databases to appear in Int. J. of Intel. Syst., 2007

[14] P. Bosc, L. Duval, O. Pivert., Value-based and representation-based Querying of Possibilistic Datases, in Recent Issues on Fuzzy Databases, G. Bordogna & G. Pasi eds, Studies in Fuzziness and soft computing n.53, Physica Verlag, 2000

ISBN: 978-989-95079-6-8

On Some Fuzzy Categories of Many-valued Topological Spaces

Ingrīda Uļjane
Department of Mathematics, University of Latvia
Institute of Mathematics and Computer Sciences, University of Latvia (IMCS UL)
Riga, Latvia
Email: uljane@mail.com

Abstract— *The aim of this paper is to construct some new fuzzy categories related to many-valued sets and many valued topologies and to consider their properties.*

Keywords— Fuzzy categories, GL-monoids, L-valued equalities, L-topologies and L-fuzzy topologies.

1 Introduction

Since the inception of the notion of a fuzzy set by Zadeh [19] and its generalization as an (L)-fuzzy set where L is a complete lattice by Goguen [3], the efforts of many researchers were forced to introduce the appropriate fuzzy counterparts of the classical mathematical concepts and to develop substantial corresponding theories. However from the categorical point of view in most cases the work was done in the context of ordinary, that is crisp categories. To explain our idea we give some examples

1. Following the works of Chang [2] and Goguen [4], an L-fuzzy topological space is a pair (X, τ) where X is a set and τ is a subfamily of L^X satisfying axioms which are natural analogues of classical topological axioms. The continuity of mappings between L-fuzzy topological spaces is defined just by "translating" the definition of continuity in fuzzy enviroment.

2. In [18] Mingsheng Ying introduced the concept of a fuzzifying topological space by semantical analysis of usual topological axioms, cf also [6]. Namely a fuzzifying topological space is a pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T} : L^X \to L$ is a mapping satisfying the axioms which are obtained by logical interpretation of the standard topological axioms. Continuity of mappings between two fuzzifying topological spaces is defined by logical interpretation of the axiom "the preimage of an open set is open".

3. In [11] A. Rosenfeld defined a fuzzy subgroup of a group $(G, \cdot, e)$ as a mapping $H : G \to [0, 1]$ such that $H(x \cdot y) \geq H(x) \wedge H(y)$ and $H(x) = H(x^{-1})$ for all $x, y \in G$. A homomorphism between fuzzy subgroups H_1 and H_2 is defined as a homomorhism between the underlying groups $f : G_1 \to G_2$ such that $H_1(x) \leq H_2(f(x))$ for every $x \in G_1$.

Thus in all these cases as well as in many other situations, the authors work in the following context: they consider a (usual) category whose objects are certain mathematical structures involving fuzzy sets as objects and mappings satisfying certain properties between these objects as morphisms.

On the other hand in [13] the concept of a fuzzy category was introduced. A fuzzy category is some category-type conglomerate whose objects and morphisms may be such only to a certain degree, see the next section for the precise definition. The foundations of the theory of fuzzy categories were developed in a series of subsequent papers, see e.g. [14], [10], [15], etc. These works contain also description of some concrete fuzzy categories related to algebra and topology, which are obtained by a certain fuzzification of classical categories.

The aim of this paper is to construct some new fuzzy categories related to many-valued sets and many valued topologies and to consider their properties. They can be regarded as fuzzy counterparts of the category $\mathbf{SET}(L)$ [7] and categories $\mathbf{TOP}(L)$ and $\mathbf{FTOP}(L)$ introduced in our work [16]. The fuzzy categories considered in this paper were first defined in our talk at the Czech-Latvian Seminar *"Advanced methods in Soft Computing",* in Trojanovice, Czech Republic, November 19-22, 2008.

2 Fuzzy categories

First we recall the concept of an (L)-fuzzy category (where L is a GL-monoid) in a special form appropriate for our merits.

Recall (see [7]) that a $GL-$monoid is an infinitely distributive lattice $(L, \leq, \wedge, \vee)$ [1] enriched with a monotone, commutative and associative binary operation $*$ such that

1. $a * 1 = a$ and $a * 0 = 0$ for all $a \in L$ where 0 and 1 are the bottom and the top elements of L respectively;

2. $a * \left(\bigvee\limits_{i \in \mathcal{I}} b_i \right) = \bigvee\limits_{i \in \mathcal{I}} (a * b_i)$ $\forall a \in L$, $\forall \{b_i : i \in \mathcal{I}\} \subseteq L$;

3. If $a \leq b$, then there exists $c \in L$ such that $a = b * c$. (The last property of a GL-monoid is called divisability.)

Such operation $*$ will be refered to as the conjunction in the GL-monoid. Important examples of GL-monoids are Heyting algebras (in this case $* = \wedge$) and Łukasiewicz MV-algebra (in this case L = [0,1] and $a * b = \max\{a + b - 1, 0\}$) [5].

It is known that every $GL-$monoid is residuated, i.e. there exists a further binary operation " $\mapsto$ " (residuation) on L linked to $*$ with the Galois condition:

$$a * b \leq c \Longleftrightarrow a \leq (b \mapsto c) \qquad \forall a, b, c \in L.$$

Explicitely implication is given by

$$a \mapsto b = \bigvee \{\lambda \in L \mid a * \lambda \leq b\}.$$

ISBN: 978-989-95079-6-8

Let $L = (L, \leq, \wedge, \vee, *)$ be a GL-monoid. *An (L)-fuzzy category* [13] is a pair $(\mathcal{C}, \mu)$ where $\mathcal{C}$ is an ordinary category with the class of objects $\mathcal{O}b(\mathcal{C})$, the class of morphisms $\mathcal{M}or(\mathcal{C})$, and where $\mu : \mathcal{M}or(\mathcal{C}) \to L$ is an L-subclass of the class of morphism such that:

(i) $\mu(g \circ f) \geq \mu(g) * \mu(f)$ whenever composition $g \circ f$ is defined in the category $\mathcal{C}$;

(ii) for each $X \in \mathcal{O}b(\mathcal{C})$ $\mu(id_X) = 1$ where id_X is the identity morphism.

3 Category SET(L) and fuzzy category $\mathcal{F}$-SET(L)

Category **SET**(L) was first defined in [7], see also [8]. The objects of the category **SET**(L) are L-valued sets, that is pairs (X, E) where X is a set and E is an L-valued equality, that is a mapping $E : X \times X \to L$ such that

1) $E(x, x) = 1$;

2) $E(x, y) = E(y, x)$;

3) $E(x, y) * E(y, z) \leq E(x, z)\ \forall x, y, z \in X$.

An L-subset A of an L-valued set (X, E) (that is a mapping $A : X \to L$) is called extensional if $A(x) * E(x, x') \leq A(x')$ for all $x, x' \in X$. The family of all extensional L-subsets of (X, E) will be denoted L^X_E.
The morphisms $f : (X, E_X) \to (Y, E_Y)$ in the category **SET**(L) are extensional mappings, that is mappings $f : X \to Y$ such that

$$E_X(x, x') \leq E_Y(f(x), f(x')) \quad \forall x, x' \in X.$$

Consider the category

$$\mathcal{C}(L) = (\mathcal{O}b(\mathbf{SET}(L)), \mathcal{M}or(\mathbf{SET})),$$

that is $\mathcal{C}(L)$ has objects from **SET**(L) and morphisms from **SET**, that is all mappings between the corresponding sets.
We define the measure μ of extensionality for a mapping $f : X \to Y$ by

$$\mu(f) = \bigwedge_{x,x' \in X} (E_X(x, x') \mapsto E_Y(f(x), f(x'))).$$

Theorem 3.1 *The triple $(\mathcal{O}b(\mathbf{SET}(L)), \mathcal{M}or(\mathbf{SET}), \mu)$ is an L-fuzzy category. It will be denoted by $\mathcal{F}$-**SET**(L) and called the L-fuzzification of the category* **SET**(L).

Proof: Let $id_X : (X, E) \to (X, E)$ be the identity mapping. Then, obviously, $\mu(id_X) = 1$. Hence, to prove the theorem, we have to show that if $f : (X, E_X) \to (Y, E_Y)$ and $g : (Y, E_Y) \to (Z, E_Z)$ are potential morphisms (that is mappings between the corresponding sets), then $\mu(g \circ f) \geq \mu(g) * \mu(f)$. Indeed,

$$\mu(g \circ f) =$$
$$= \bigwedge_{x,x' \in X} (E_X(x, x') \mapsto E_Z((g \circ f)(x), (g \circ f)(x'))) \geq$$
$$\bigwedge_{x,x' \in X} (E_X(x, x') \mapsto E_Y(f(x), f(x'))) *$$
$$\bigwedge_{x,x' \in X} (E_Y(f(x), f(x')) \mapsto E_Z(g(f(x)), g(f(x')))) \geq$$
$$\mu(f) * \bigwedge_{y,y' \in Y} (E_Y(y, y') \mapsto E_Z(g(y), g(y'))) =$$
$$\mu(f) * \mu(g).$$

□

4 Category TOP(L) and fuzzy category $\mathcal{F}$-TOP(L).

Definition 4.1 *Let $L = (L, \wedge, \vee, *)$ be a GL-monoid and (X, E) be an L-valued set. A family $\tau \subseteq L^X$ is called an L-topology on an L-valued set (X, E) if*

1. *$0_X \in \tau; 1_X \in \tau$ (where $0_X, 1_X$ are constant L-sets with values 0 and 1 respectively;*
2. *if $U, V \in \tau$, then $U \wedge V \in \tau$;*
3. *if $U_i \in \tau \quad \forall i \in \mathcal{I}$, then $\bigvee_{i \in \mathcal{I}} U_i \in \tau$;*
4. *if $U \in \tau$, then $U(x) * E(x, x') \leq U(x')\ \forall x, x' \in X$.*

The last condition means that all L-sets in τ are extensional and hence $\tau \subseteq L^X_E$.
The triple (X, E, τ) where τ is an L-topology on an L-valued set is called an L-valued L-topological space. Respectively, the elements $U \in \tau$ are called open L-sets in this L-valued L-topological space.

Definition 4.2 *A mapping $f : (X, E_X, \tau_X) \to (Y, E_Y, \tau_Y)$ is called continuous if*

1. *$E_X(x, x') \leq E_Y(f(x), f(x'))$ for all $x, x' \in X$,* that is f is an extensional mapping between the corresponding L-valued sets (X, E_X) and (Y, E_Y), and
2. *$f^{-1}(V) \in \tau_X$ whenever $V \in \tau_Y$.*

Obviously, if $f : (X, E_X, \tau_X) \to (Y, E_Y, \tau_Y)$ and

$$g : (Y, E_Y, \tau_Y) \to (Z, E_Z, \tau_Z)$$

are continuous, then the composition

$$g \circ f : (X, E_X, \tau_X) \to (Z, E_Z, \tau_Z)$$

is continuous and the identity mapping

$$id_X : (X, E_X, \tau_X) \to (X, E_X, \tau_X)$$

is continuous. Hence L-valued L-topological spaces and continuous mappings between them form a category **TOP**(L).
One can easily see that in case E is crisp, that is $E(x, x') = 0$ if $x \neq x'$ and $E(x, x) = 1$, the L-valued toplogical space is just an L-topologcal space in the sense of [2], [4] and the category of L-topological spaces is actually a complete subcategory of **TOP**(L).

Notice that, by lower-semicontinuity of conjunction, for extensional L-sets U_i, we have

$$\left(\bigvee_{i \in \mathcal{I}} U_i(x)\right) * E(x, x') \leq$$
$$\bigvee_{i \in \mathcal{I}} (U_i(x) * E(x, x')) \leq \bigvee_{i \in \mathcal{I}} U_i(x') \quad \forall x, x' \in X.$$

Thus the supremum of extensional L-subsets of an L-valued set is extensional itself. Therefore, in an analogy with classical topology we can define the interior $int(A)$ of an L-subset A of an L-valued L-topological space (X, E, τ) as the largest $(\geq)$ one of all open L-subsets of (X, E, τ) contained $(\leq)$ in A. Equivalently, it can be defined by the formula

$$int(A) = \bigvee \{U \in \tau \mid U \leq A\}.$$

ISBN: 978-989-95079-6-8

One can easily verify that the resulting operator

$$int : L^X \to L^X$$

satisfies all properties, analogous to the properties of the interior operator in classical topology.

The interior operator allows us to measure "the degree of openess" of an L-set A in an L-valued L-topological space by evaluating, to what extent an L-set A is contained in its interior. To realize this we need to extend an L-valued equality E from a set X to its extensional L-powerset L_E^X. This can be done as follows (see [17]):
Given $A, B \in L_E^X$ let

$$\mathcal{R}(A,B) = \bigwedge_{x,z\in X} ((E(x,z) * A(x)) \mapsto B(z)).$$

It is shown in [17] that $\mathcal{R} : L_E^X \times L_E^X \to L$ is an L-fuzzy order relation on L_E^X, that is it is refelexive ($\mathcal{R}(A,A) = 1$ for all $A \in L_E^X$) and transitive ($\mathcal{R}(A,B) * \mathcal{R}(B,C) \le \mathcal{R}(A,C)$ for all $A, B, C \in L_E^X$). By setting $\mathcal{E}(A,B) = \mathcal{R}(A,B) \wedge \mathcal{R}(B,A)$ an L-valued equality on L_E^X is obtained [17].
Now to fuzzify the category **TOP**(L) we consider the category

$$(\mathcal{O}b(\mathbf{TOP}(L)), \mathcal{M}or(\mathbf{SET}), \circ),$$

where objects are L-valued L-topological space (X, E, τ), but morphisms are from the category **SET**, that is mapings between corresponding sets. For a given mapping

$$f : (X, E_X, \tau_X) \to (Y, E_Y, \tau_Y)$$

we define the degree of extensionality by

$$\mu_1(f) = \bigwedge_{x,x'\in X} (E_X(x,x') \mapsto E_Y(f(x), f(x')),$$

the degree of continuity by

$$\mu_2(f) = \bigwedge_{V\in\tau_Y} \mathcal{R}(f^{-1}(V), int(f^{-1}(V)))$$

and finally the degree of being a morphism by

$$\mu(f) = \mu_1(f) \wedge \mu_2(f).$$

Theorem 4.3 *$(\mathcal{O}b(\mathbf{TOP}(L)), \mathcal{M}or(\mathbf{SET}), \circ, \mu)$ is a fuzzy category.*

Proof Let $f : (X, E_X, \tau_X) \to (Y, E_Y, \tau_Y)$ and

$$g : (Y, E_Y, \tau_Y) \to (Z, E_Z, \tau_Z).$$

The inequality

$$\mu_1(g \circ f) \ge \mu_1(g) * \mu_1(f)$$

was established in the previous section. To establish the inequality

$$\mu_2(g \circ f) \ge \mu_2(g) * \mu_2(f)$$

we are reasoning as follows:

Let $f : (X, E_X, \tau_X) \to (Y, E_Y, \tau_Y)$,
$g : (Y, E_Y, \tau_Y) \to (Z, E_Z, \tau_Z)$.
We fix $W \in \tau_Z$. Then by transitivity of $\mathcal{R}$

$$\mathcal{R}(f^{-1}(g^{-1}(W)), int_X(f^{-1}(g^{-1}(W)))) \ge \mathcal{R}(f^{-1}(g^{-1}(W)), f^{-1}(int_Y(g^{-1}(W))))* \mathcal{R}(f^{-1}(int_Y(g^{-1}(W))), int_X(f^{-1}(g^{-1}(W)))).$$

We estimate each member separately:

$$\mathcal{R}(f^{-1}(int_Y(g^{-1}(W))), int_X(f^{-1}(g^{-1}(W)))) \ge \mathcal{R}(f^{-1}(int_Y(g^{-1}(W))), int_X(f^{-1}(int_Y g^{-1}(W)))) \ge \mu_2(f).$$

$$\mathcal{R}(f^{-1}(g^{-1}(W)), f^{-1}(int_Y(g^{-1}(W)))) = \bigwedge_{x,x'} ((E(x,x') * g^{-1}(W)(f(x))) \mapsto int_Y(g^{-1}(W)(f(x')))) \ge$$

(by extentionality of f)

$$\ge \bigwedge_{x,x'} ((E(f(x),f(x')) * g^{-1}(W)(f(x))) \mapsto int_Y(g^{-1}(W)(f(x')))) \ge \bigwedge_{y,y'} \left((E(y,y') * g^{-1}(W)(y)) \mapsto int_Y(g^{-1}(W)(y'))\right) \ge \mu_2(g).$$

Thus

$$\mathcal{R}(f^{-1}(g^{-1}(W)), int_X(f^{-1}(g^{-1}(W)))) \ge \mu_2(g) * \mu_2(f).$$

Since this is valid for any $W \in \tau_Z$ we obtain the requested

$$\mu_2(g \circ f) \ge \mu_2(g) * \mu_2(f).$$

Finally referring to the properties of a GL-monoid, we get

$$\mu(g \circ f) = \mu_1(g \circ f) \wedge \mu_2(g \circ f) \ge (\mu_1(f) * \mu_1(g)) \wedge (\mu_2(f) * \mu_2(g)) \ge (\mu_1(f) \wedge \mu_2(f)) * (\mu_1(g) \wedge \mu_2(g)) = \mu(f) * \mu(g).$$

We conclude the proof by noticing that obviously $\mu(id_X) = 1$ for the identity morphism

$$id_X : (X, E_X, \tau_X) \to (X, E_X, \tau_X).$$

□

5 Category FTOP(L) and fuzzy category $\mathcal{F}$-FTOP(L)

The concept of an L-valued L-fuzzy topological space as a generalization of the concept of an L-fuzzy topological space [12], [9] and the corresponding category **FTOP**(L) were introduced in [17]. The objects of this category are triples $(X, E, \mathcal{T})$ where (X, E) is an L-valued set and $\mathcal{T} : L^X \to L$ is an L-fuzzy topology, that is an extensional subset of L_E^X such that

1) $\mathcal{T}(1) = \mathcal{T}(0) = 1$;

2) $\mathcal{T}(U \wedge V) \ge \mathcal{T}(U) \wedge \mathcal{T}(V) \quad \forall U, V \in L_E^X$;

3) $\mathcal{T}\left(\bigvee_{i\in I} U_i\right) \ge \bigwedge_{i\in\mathcal{I}} \mathcal{T}(U_i) \quad \forall \{U_i \mid i \in \mathcal{I}\} \subseteq L_E^X$.

and morphisms are extensional mappings

$$f : (X, E_X) \to (Y, E_Y)$$

such that $\mathcal{T}(f^{-1}(V)) \ge \mathcal{T}(V)$.
One can easily see that in case E is crisp, that is $E(x,x') = 0$

ISBN: 978-989-95079-6-8

if $x \neq x'$ and $E(x,x) = 1$, the L-valued L-fuzzy toplogical space is just an L-fuzzy topological space as it was defined in [12], [4] and the category of L-fuzzy topological spaces is actually a complete subcategory of $\mathbf{FTOP}(L)$.

Let $(X, E_X, \mathcal{T}_X)$, $(Y, E_Y, \mathcal{T}_Y)$ be L-valued L-fuzzy topological spaces and $f : X \to Y$ be a mapping of the corresponding underlying sets. The measure of the extensionality μ_1 of the mapping $f : (X, E_X) \to (Y, E_Y)$ is defined as above. The measure of continuity μ_2 of a mapping $f : (X, \mathcal{T}_X) \to (Y, \mathcal{T}_Y)$ is defined by

$$\mu_2(f) = \bigwedge_{V \in L_E^Y} \left(\mathcal{T}_Y(V) \mapsto \mathcal{T}_X(f^{-1}(V))\right).$$

Further, for a mapping $f : (X, E_X, \mathcal{T}_X) \to (Y, E_Y, \mathcal{T}_Y)$ we set $\mu(f) = \mu_1(f) \wedge \mu_2(f)$.

Theorem 5.1

$$\mathcal{F}-\mathbf{FTOP}(L) = (\mathcal{O}b(\mathbf{FTOP}(L)), \mathcal{M}or(\mathbf{SET}), \mu)$$

is an L-fuzzy category.

Proof: Since for the identity morphism

$$id_X : (X, E_X, \mathcal{T}_X) \to (X, E_X, \mathcal{T}_X)$$

obviously $\mu_1(id_X) = 1$ and $\mu_2(id_X) = 1$, we have $\mu(id_X) = 1$. Therefore to prove the theorem we have to establish that given two functions $f : (X, E_X, \mathcal{T}_X) \to (Y, E_Y, \mathcal{T}_Y)$ and $g : (Y, E_Y, \mathcal{T}_Y) \to (Z, E_Z, \mathcal{T}_Z)$, viewed as the mappings of the corresponding underlying sets it holds

$$\mu(g \circ f) \geq \mu(g) * \mu(f).$$

The inequality $\mu_1(g \circ f) \geq \mu_1(g) * \mu_1(f)$ was established in subsection 3. The inequality

$$\mu_2(g \circ f) \geq \mu_2(g) * \mu_2(f)$$

was established in [14], see also [15]. Now, referring to the properties of a GL-monoid, we get

$$\begin{gathered}\mu(g \circ f) = \mu_1(g \circ f) \wedge \mu_2(g \circ f) \geq \\ (\mu_1(g) * \mu_1(f)) \wedge (\mu_2(g) * \mu_2(f)) \geq \\ (\mu_1(g) \wedge \mu_2(g)) * (\mu_1(f) \wedge \mu_2(f)) = \mu(g) * \mu(f).\end{gathered}$$

□

References

[1] G. Birkhoff, *Lattice Theory*, 3rd ed., AMS Providence, RI, 1967.

[2] C.L. Chang, *Fuzzy topological spaces*, J. Math. Anal. Appl., vol. 24, (1968), pp. 182-190.

[3] J.A. Goguen, *L-fuzzy sets*, J. Math. Anal. Appl., vol. 18, (1967) 145-174.

[4] J.A. Goguen, *The fuzzy Tychonoff theorem*, J. Math. Anal. Appl., vol. 43, (1973) 737-742.

[5] J. Łukasiewicz, *Selected works: Studies in Logic and the Foundations of Mathematics*, Borkowski eds., North Holland, Amsterdam, 1970.

[6] U. Höhle, *Upper semicontinuous fuzzy sets and applications*, J. Math. Anal. Appl. vol.78, (1980), 659–673.

[7] U. Höhle, *M-valued sets and sheaves over integral commutative cl-monoids*, In: *Applications of Category Theory to Fuzzy Subsets* S.E. Rodabaugh, E.P. Klement and U. Höhle eds., Kluwer, Dodrecht, Boston, 1992, pp. 33 – 72.

[8] U. Höhle, *Commutative residuated l-monoids*, In: *Non-Classical Logics and their Applications to Fuzzy Subsets - A Handbook of the Mathematical Foundations of Fuzzy Set Theory*, U. Höhle and E.P. Klement eds., Kluwer, Dodrecht, Boston, 1994, pp. 53-106.

[9] T. Kubiak, On Fuzzy Topologies, Ph. D. Thesis, Adam Mickiewicz Univ., Poznań, Poland (1985).

[10] T. Kubiak, A.Šostak, *A fuzzification of the category of M-valued L-topological spaces*, Applied General Topology, vol. 5, (2004), 137–154.

[11] A. Rosenfeld, *Fuzzy groups*, J.Math.Anal.Appl., vol. 35, (1971), 512-517.

[12] A. Šostak, *On a fuzzy topological structure*, Suppl. Rend. Circ. Matem. Palermo, Ser II, vol. 11, (1985), 89-103.

[13] A. Šostak, *On a concept of a fuzzy category*, In: *14th Linz Seminar on Fuzzy Set Theory: Non-Classical Logics and Their Applications*, Linz, Austria, 1992, 63-66.

[14] A. Šostak, *Fuzzy Categories Related to Algebra and Topology* Tatra Mount. Math. Publ., vol. 20, (1999), 159 - 186.

[15] A. Šostak, *On some Fuzzy Categories Related to the Category L-TOP of L-topological spaces* Chapter 12 In: Topological and Algebraic Structures in Fuzzt Sets, E.P. Klement, S.E. Rodabaugh eds., Trends in Logic, vol. 20, Kluwer Academic Publ., 2003.

[16] I. Uļjane, A. Šostaks,*On a category of L-valued L-topological spaces*, Acta Univ. Latv., vol. 2005, pp. 99-107.

[17] I.Uļjane, *On the Order Type L-valued Relations on L-powersets*, Mathware and Soft Computing, Vol. XIV, n.3 (2007), pp. 183. - 199.

[18] Ying Mingsheng, *A new approach to fuzzy topology*, Part I, Fuzzy Sets and Syst., vol. 39, (1991), 303-321.

[19] L.A. Zadeh, *Fuzzy Sets*, Inform. and Control, vol. 8, (1965), 338-365.

ISBN: 978-989-95079-6-8

Valued Constraint Satisfaction Problems Applied to Functional Harmony

Nele Verbiest, Chris Cornelis and Yvan Saeys
Ghent University, Ghent, Belgium
Email: {Nele.Verbiest, Chris.Cornelis, Yvan.Saeys}@UGent.be

***Abstract*— *Harmonization with four voices is a musical problem which is subject to hard constraints, which absolutely need to be fulfilled, as well as to soft constraints, which preferably hold, but are not mandatory. In this paper, we model this problem as a valued constraint satisfaction problem (VCSP): costs are assigned to possible solutions based on the constraints they violate. We design an algorithm that finds a minimal-cost solution, thus solving the harmonization problem, and we present initial results obtained by this algorithm.*

Keywords— functional harmony, valued constraint satisfaction problems, soft constraints

1 Introduction

Ever since people have had computers at their use, they have attempted to make them acquire human skills like rational thought and creativity. While the prospect of computers actually creating art might still seem out of reach, in musical composition some successful efforts have already been made; in [1], Truchet and Codognet give an overview of musical problems that were solved using computers.

In this paper, we discuss the use of Valued Constraint Satisfaction Problems (VCSPs) in functional harmony with four voices (soprano, alto, tenor, bass). Solving harmonization problems is a basic skill that every composer has to acquire before attempting more serious work. These problems range from very simple to advanced, with many gradations in between. Their solution is governed by a set of strict rules that have to be verified, but also by a variety of "softer" guidelines with varying, subjective importance, imposed to make the result more interesting, more varied, ... Valued constraints can be constructed and enforced sensibly to model them.

The structure of this paper is as follows: in Section 2, we discuss related work, and in particular solvers that have already been proposed for this problem, and position our own contribution. In Section 3, we briefly explain what functional harmony is about, and which constraints govern harmonization problems. In Section 4, we give the necessary background on VCSPs, while in Section 5 we translate the musical composition problem and its constraints to this framework. In Section 6 we describe the algorithm designed for solving our VCSP. In Section 7 we present some results obtained by this algorithm, and in the last section we discuss its practical use and suggest future research directions.

2 Related work

Programs to compose music have been designed since the 50s. We are specifically interested in harmonization problems with four voices, a problem that can be stated as a constraint satisfaction problem [2]. To solve this problem, two main classes of techniques have been described in the literature: backtracking and genetic algorithms.

Backtracking algorithms tackle the problem in the following way. The harmonization of a piece of music proceeds in a left to right fashion, analyzing every subsequent note. For each note, one seeks a possible chord for that note, and if there is no solution for this note, the algorithm goes back to a preceding note to try other possibilities. The standard work in this class is due to Ebcioğlu [3]: he introduced an expert system for harmonizing chorales with four voices, based on a set with about 350 rules. Apart from this, there are also backtracking algorithms where possible chords for a note are ranked by musical suitability. At every step, one tries the best chord. If this chord does not lead to a solution, the chord is removed from the ranking and a solution with the second best chord is sought [4]. Existing backtracking solvers look for one solution and then stop.

Genetic Algorithms (GAs) use an evaluation function to judge the musical quality of a set of solutions, and proceed in an iterative way to optimize these solutions [5, 6], using evolutionary reproduction operators like mutation and crossover.

Both approaches have their advantages and disadvantages. Backtracking algorithms are very efficient in finding a single solution, but little can be said about its quality: any solution found by the algorithm is guaranteed to meet the imposed hard constraints, but soft constraints are generally not considered. Ranking possible chords for a note deals only partially with this problem; this tactic will find a good solution, but it is not guaranteed to be the best one. By virtue of the evaluation function, genetic algorithms are better at catering for soft constraints, but they also suffer from the fact that their solutions are often suboptimal.

The solver we describe in this paper combines the use of soft constraints (which can be seen as the analogon of the evaluation function in GAs) with the potential to find an optimal solution in the search space.

3 Functional harmony with four voices

We briefly describe the essentials of functional harmony, and refer the interested reader to [7, 8] for more details. The problem we consider is a simplification of the general harmonization problem: given a melody for a soprano, we aim to find corresponding melodies for alto, tenor and bass. We suppose that only quarter notes are used in all the melodies and we work in C major. In tonal music, one speaks about grades, which —for a particular tuning— coincide with a set of chords. The most important grades that are always considered are I, IV and V. Other grades that often appear are VI and sometimes II. In C major tuning, the first note is c which results in grade I being the chord built on c; the fourth note is f,

ISBN: 978-989-95079-6-8

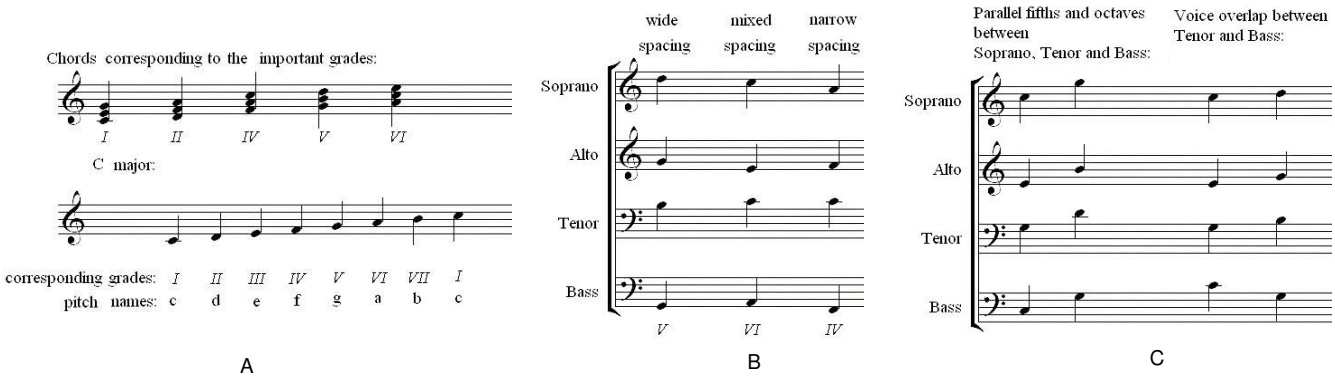


Figure 1: Essentials of functional harmony, illustrating the grades and chords of C Major (A), the concepts of narrow, mixed and wide spacing (B), and some examples of constructs that are not allowed (C).

so the chord corresponding to grade IV is the chord built on f. Fig. 1A illustrates the grades and chords of C major.

Now consider a note in the soprano. This note is in one or more of these chords. In harmonizing the soprano, one must choose a chord, and attribute the corresponding notes in the other voices to a note of this particular chord. The bass must always be the ground note (the lowest note) of the chord. For example, if IV is chosen, then the bass must be f. The soprano, alto and tenor have to be different from each other, so each note of the chord appears once, except for the ground note which appears twice.

We now introduce the concepts *wide* and *narrow* spacing. For every note in the soprano, we still have some freedom for the notes in alto, tenor and bass. We now restrict the possibilities a bit more. For each chord we choose, we have to decide whether we write it in wide or in narrow spacing. Narrow spacing means that between soprano and alto, and between alto and tenor there is no other note of the chord. Wide spacing means that there is always one note of the chord between soprano and alto, and between alto and tenor. This is illustrated in Fig. 1B. If the previous chord is written in narrow spacing, then the next chord must be narrow spaced as well. If the previous chord is wide spaced, then so must be the next chord.

The VIth grade chord is written like the other chords, except for the case when it is preceded by a Vth grade. In this case, the second note of the chord is written twice. This means that it has neither wide nor narrow spacing; we refer to this situation as *mixed* spacing. When we encounter a sequence V-VI it is possible to go from wide to narrow spacing or vice versa using mixed spacing on the VIth grade. Fig. 1B shows a good V-VI-IV sequence where we switch spacing.

Now we know how to write the chords we must consider the problem of choosing the right grade. To this end, a number of rules have been defined in the domain of functional harmony:

- The first and the last grade of a music piece must be I. The other grades depend on the previous grades.
- Some transitions between grades are not allowed, others are considered very good, while still others are not considered wrong but should be avoided as much as possible.
- Each of the different voices has its own range of notes that it can sing. The soprano can sing the highest notes, the bass sings the lower notes, and care has to be taken that these ranges are respected.
- It is forbidden to write parallel fifths or parallel octaves, to write the same grade before and after a bar or to overlap voices.

Some of these forbidden constructs are shown in Fig. 1C.

4 Valued constraint satisfaction problems

A constraint satisfaction problem (CSP) involves assigning values to variables that are subject to some constraints [9]. If there are constraints that are preferred but not essential, we can consider the problem of assigning values to the variables in such a way that the more important constraints are fulfilled and the less important ones are fulfilled to the extent possible. The latter yields a valued CSP (VCSP), which is characterized by a set of hard constraints (that must be fulfilled), and a set of so-called *valued constraints*. If the valued constraints are treated as mandatory, the problem is often unsolvable. If the valued constraints are ignored, we get solutions of bad quality. Our interest is therefore in the solution that best respects the set of constraints. To express what the best solution is, we assign costs to valued constraints. Every time a valued constraint is violated, we count the cost of this violation. In the end, we are interested in the solution with the lowest cost. We recall a formal definition of VCSP's.

Definition 1. *A Valued Constraint Satisfaction Problem (VCSP) is a quadruple* (X, D, C, V)*, where*
$X = \{X_1, ..., X_n\}$ *is a set of variables,*
$D = \{D_1, ..., D_n\}$ *is a set of finite domains:* D_i *is the set of possible values for* X_i*,*
$C = \{C_1, ..., C_r\}$ *is a set of hard constraints and*
$V = \{V_1, ..., V_s\}$ *is a set of valued constraints.*

A hard constraint C_i on the ordered set of variables $var(C_i)$ is the relation of the allowed combinations of values for the variables in $var(C_i)$; C_i is a mapping from the cartesian product of domains of variables in $var(C_i)$ to $\{0, 1\}$.

A valued constraint V_i on the ordered set of variables $var(V_i)$ specifies the cost of the combination of the values for the variables in $var(V_i)$; V_i is a mapping from the cartesian product of domains of variables in $var(V_i)$ to $[0, 1]$.

For example, if V_i pertains to the variables X_1, X_2 (i.e.

ISBN: 978-989-95079-6-8

$Var(V_i) = (X_1, X_2)$) and we assign values x_1 and x_2 to X_1 and X_2 respectively, the cost of this combination is given by $V_i(x_1, x_2)$. If C_i pertains to the variables X_1, X_2, X_3 (i.e. $Var(C_i) = (X_1, X_2, X_3)$) and we assign values x_1, x_2 and x_3 to these variables, the combination (x_1, x_2, x_3) is allowed if $C_i(x_1, x_2, x_3) = 1$ and forbidden if $C_i(x_1, x_2, x_3) = 0$.

5 Modeling the problem

We can now translate the musical problem of harmonizing a soprano to a purely mathematical problem. Suppose we have a soprano that lasts n times, with every time divided in 4 counts. All the times contain four quarter notes, except for the last time which contains one whole note. This makes a total of $l := (n-1) * 4 + 1$ notes. We call l the length of the problem. For every note, we want to store the notes for the bass, tenor, alto and soprano and we want to store the grade we use. This can be done using a $5 \times l$ matrix X.
The grades and notes are represented by a positive integer, and the central c is represented by the number 24. We assign integers to the other notes as follows: if the note is higher/lower than the central c, we count the number of half tones between the central c and this note and add/subtract it to 24. In this way, we fix the domain and variables of our VCSP. The variables are the entries $x_{i,j}, i = 1, ..., 5, j = 1, ..., l$ of the matrix X and the domain is the set of natural numbers $\mathbb{N}$.

5.1 Hard constraints

We will first translate the hard constraints to mathematical formulas. We only use grades I, II, IV, V and VI, which entails:

$$x_{5,j} \in \{1, 2, 4, 5, 6\}\ \forall j = 1, ..., l.$$

The first and last grade must be I, so

$$x_{5,1} = x_{5,l} = 1.$$

Every voice has a specific range of notes that it can sing, which can be expressed as:

$$\text{alto} : 19 \leq x_{2,j} \leq 38\ \forall j = 1, ..., l$$

$$\text{tenor} : 12 \leq x_{2,j} \leq 28\ \forall j = 1, ..., l$$

$$\text{bass} : 5 \leq x_{2,j} \leq 26\ \forall j = 1, ..., l$$

The grades before and after a bar must be different:

$$\forall j = 1, ..., l :$$

$$\text{if } j \bmod 4 \equiv 1 \text{ and } j \neq 1$$

$$\text{then } x_{5,j} \neq x_{5,j-1}$$

No parallel fifths or octaves are allowed:

$$\forall i, k, j : i, k = 1, 2, 3, 4, j = 1, ...l - 1 :$$

$$\text{if } (x_{i,j} - x_{k,j} \bmod 12 \equiv 0) \text{ and } k \neq i$$

$$\text{then } (x_{i,j+1} - x_{k,j+1} \bmod 12 \not\equiv 0)$$

and

$$\text{if } (x_{i,j} - x_{k,j} \bmod 12 \equiv 7) \text{ and } k \neq i$$

$$\text{then} (x_{i,j+1} - x_{k,j+1} \bmod 12 \not\equiv 7).$$

It is forbidden to write voice overlap (strictly between soprano, alto and tenor, not strictly between tenor and bass):

$$\forall j = 1, ..., l :$$

$$x_{1,j} > x_{2,j} > x_{3,j} \geq x_{4,j},$$

$$\forall j = 1, ..., l - 1 :$$

$$x_{2,j+1} < x_{1,j}$$

$$x_{3,j+1} < x_{2,j}$$

$$x_{4,j+1} \leq x_{3,j}.$$

The possibilities for the notes in the bass, tenor and alto and the forming of the chords (with wide and narrow spacing) are also hard constraints. We will not explicitly state these constraints here as mathematical formulas due to space constraints.

5.2 Valued constraints

We are now ready to introduce some valued constraints. There are five aspects of functional harmony we model using valued constraints. In the following subsections we propose techniques for calculating the cost of certain combinations. The results are five costs in $[0, 1]$ that correspond to the transition of one grade to another (c_1), the frequency of grades (c_2), the range of notes a voice can sing (c_3), the distance between two succeeding notes (c_4) and the use of contrary motion between bass and soprano (c_5). To allow a composer to determine the relative importance of each type of valued constraints, we attach weights $\{w_1, w_2, ..., w_5\}$ and define the total cost as $\sum_{i=1}^{5} w_i c_i$ where all weights sum to one. Assigning these weights depends on personal taste. Some composers might think it is important to have a lot of variation in the use of grades, while others might prefer to write music that is in the right range for the different voices. The weights we propose are as follows: $w_1 = 0.3, w_2 = 0.2, w_3 = 0.1, w_4 = 0.2, w_5 = 0.2$.

5.2.1 Grade transition constraints

As stated before, we only work with grades I, II, IV, V and VI. The most important grades are I, IV and V. V can be seen as a chord that builds up tension, I creates rest and IV is an advancing chord (to V). In music, it is important to have enough tension but there are points of rest needed as well.
II is a grade that can be used as a replacement grade for IV. It is also an advancing grade to V. VI is a grade that can replace I or IV, depending on the context. Generally speaking, when followed by V or preceded by I, VI is a replacement grade for IV, while when preceded by V or followed by IV, VI is a replacement grade for I. (This is also valid if we replace IV by II.) So depending on the context, VI can bring rest in music or can advance to tension. II and VI are used to bring more variation in music. For example, instead of writing I-IV-V-I (a very good sequence) all the time, we can also write I-II-V-VI. To express the aptitude of grade sequences, we introduce a valued constraint G. It is a mapping from the cartesian product of the possible grades to the interval $[0, 1]$ that represents the cost of the sequence:
$G : \{1, 2, 4, 5, 6\} \times \{1, 2, 4, 5, 6\} \rightarrow [0, 1]$.
To determine the cost of a sequence, we only look at two consecutive grades. This is sufficient since every bad sequence is

ISBN: 978-989-95079-6-8

the result of one or more separate bad successions in the sequence.
The assignment of the costs is an intuitive and subjective task. However, we try to bring some structure in it. The model sequence is I-IV-V-I. Therefore, we assign cost zero to every subsequence of length two that appears in it. All sequences that can be obtained from those subsequences by substituting a grade by its replacement grade (we call it *derivative sequences*) also have no cost.
Going from advancing to tension to rest is not very interesting. Therefore, we assign cost 1 to the sequence IV-I and all derivative sequences. Going from rest to tension without advancing should be avoided but is not particularly bad, so we assign cost 0.5 to the sequence I-V. Using the same sequence two times is not always interesting. We assign cost 0.3 to sequences I-I, V-V and derivatives, and cost 0.4 to sequences IV-IV and derivatives. An overview of the assignments is given in Table 1.

5.2.2 Grade frequency constraints

As stated before, in music it is important to maintain sufficient variation. This is not covered by the previous constraints: writing I-IV-V-I the whole time would give zero cost, but it would be boring to listen to. To express this we introduce additional valued constraints.
The algorithm designed in Section 6.1 has three options: it can generate a solution with only grades I, IV and V, only grades I, IV, V and VI, or using all grades I,II, IV, V and VI. The number of times a grade should ideally be used depends on this option, so we introduce three different valued constraints. First consider the option where only grades I, IV and V can be used. The best sequence here is I-IV-V-I. We want some variation so we sometimes write I-I, I-IV-I or I-V-I. Say that ideally, every three times we write I-IV-V, we should write I-I, I-IV-I or I-V-I once. The corresponding grade occurence percentages are given in Table 2.
We now define the valued constraint $H_1 : \{1, 4, 5\}^l \rightarrow [0, 1]$ as follows: for the whole piece of music, count how many times grades I, IV or V are used and calculate the percentages. Now for every grade, add the absolute difference between the ideal and real percentage, and divide the result by three to obtain a value in $[0, 1]$.
Since grade VI is a replacement grade for I, in the option where I, IV, V and VI can be used, we consider the same percentages as in the first option for IV and V, but we split the ideal percentage for grade I. Analogously, as grade II has the same function as grade IV, we split the ideal percentage for grade IV in the third option. The percentages are presented in Table 2. The valued constraints H_2 and H_3 are defined as before.
Again, these ideal percentages are constructed subjectively and intuitively. Different composers would propose different percentages, but they should be similar. Moreover, as for the previous valued constraints, the exact percentages are not very important, more significant are the gradations.

5.2.3 Voice range constraints

The problem we consider is to write a music piece for four voices. Of course, these voices cannot sing all existing notes. Theoretically, a bass can sing notes presented by integers from 5 to 26, yet there exist very good basses that can not reach the very low note presented by 5 or the very high note presented by 26. It is not wrong to write these extreme notes but it should be avoided, because even if the bass can sing these notes, they will not sound as good as the notes in the middle of their register. To express this mathematically, we introduce new valued constraints R_a, R_b and R_t as follows: for every note in the reach of a voice, R_b, R_t and R_a express how bad this note is for bass, tenor or alto respectively. Fig. 2 presents the values of R_a, R_b and R_t. The total cost is given by $(R_a + R_b + R_t)/3 \in [0, 1]$.

5.2.4 Distance constraints

In tonal music, it is important to have fluent melodies. Modeling this requires counterpoint techniques, which we do not consider here. Instead, we model the soundness of the melodies by considering the distances between the notes. Furthermore, it is hard to sing large intervals correctly. Therefore, we introduce a valued constraint D that limits the distance between two notes. We assign cost 1 to all intervals larger than an octave (12 half tones). Intervals smaller than an octave get cost $c_4 = d/12$, proportional to the distance d between the notes. An exception is made for the octave, which gets the same cost as a fifth because it sounds equally well and is equally easy to sing. The costs are presented in Table 3.

5.2.5 Contrary motion constraints

Contrary motion is the general movement of two melodic lines in opposite directions. That is, when one of the lines moves up, the other line moves down. In tonal music, contrary motion is important to maintain independence of melodic movement. To express this as a valued constraint we assign a cost

Table 1: The mapping G that assigns a cost to every allowed succession of two grades

GRADE1-GRADE2	COST=G(GRADE1,GRADE2)
II-V, V-I, I-II, IV-V	0
VI-V, VI-II, I-IV, V-VI	0
VI-IV, V-I, I-VI	0
I-I, V-V, VI-I, VI-VI	0.3
IV-II, IV-IV, II-II	0.4
I-V	0.5
IV-I, II-I, II-VI	1.0

Table 2: valued constraint H compares the count of grades to the ideal percentages.

GRADE	IDEAL PERCENTAGES		
	OPTION 1: I, IV AND V	OPTION 2: I, IV, V AND VI	OPTION 3: I, II, IV, V, VI
I	0.394	0.263	0.263
II	-	-	0.075
IV	0.303	0.303	0.228
V	0.303	0.303	0.303
VI	-	0.131	0.131

ISBN: 978-989-95079-6-8

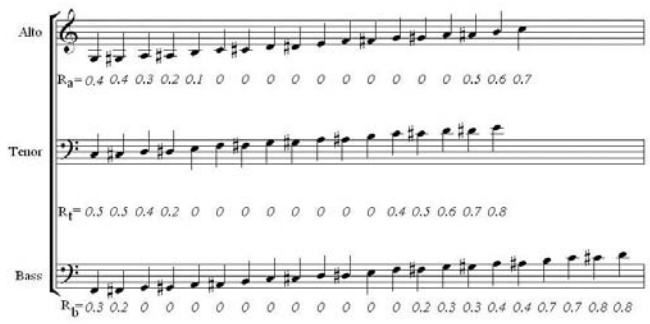


Figure 2: Costs that are incrurred when notes at the extreme ends of a voice's scale are used.

$c_5 = 1$ if there is no contrary motion between bass and soprano.

6 Solving the four voices harmonization VCSP

Now we translated the musical problem to a purely mathematical problem, we can design an algorithm that solves it. There exist good solvers for VCSPs [9] but we choose to design and implement a separate algorithm for the harmonization problem. The algorithm follows the thinking of a composer that solves the exercise: from left to right, try out all the possibilities and return when a solution gets stuck.

6.1 A backtracking algorithm

Algorithm 1 shows the outline of the main part of the algorithm. First note that it returns a single minimal-cost solution; if there are several such solutions, our algorithm returns the first one encountered. At each stage, max_cost and total_cost represent the cost of the optimal solution so far, and of the currently investigated solution, respectively. They are initialized in lines 1 and 2.

From Section 2, we know that the grade on the first note must be I. Either narrow or wide spacing can be used for the first chord; the former option is considered in lines 3–5, the second one in line 6–8.

The central part of the algorithm is in the recursive procedure Calculate_Solution. For every note i except the last one, all possible solutions are generated (line 2), that is, all permissible combinations of grades for note i, and corresponding notes for bass, tenor and alto. Assuming no hard constraints are violated (line 3), the cost of a solution w.r.t. the valued constraints is evaluated (line 4); all costs can be calculated directly, except from c_2 which is calculated considering the problem from the beginning until position i. c_2 is only calculated when $i > 4$. If the total cost for the current solution is smaller than the new costs added up to the total cost, one has to try the next possibility, because from here we will never find a solution that is cheaper than the solution we already found (line 5-6). Otherwise, one adds the costs except from c_2 up to the total cost and goes further looking for a solution (lines 8-9). c_2 is not added up to the total cost since this cost is calculated every step and it counts for the whole piece.

When we have reached the final note, we change the current solution and the maximum cost and we return in the recursion to find a cheaper solution (lines 16-17).

This can lead to a solution or to failure. In both cases we return to try the next posibilities. If there are no posibilities left, the algorithm does not go further in this branch of the recursion.

Table 3: valued constraint D is a measure for the soundness of an interval.

INTERVAL (IN HALF TONES)	COST c_4	INTERVAL (IN HALF TONES)	COST c_4
> 12	1.000	6	0.500
12	0.583	5	0.417
11	0.917	4	0.333
10	0.833	3	0.250
9	0.750	2	0.167
8	0.667	1	0.083
7	0.583	0	0.000

Algorithm 1: Algorithm that solves the problem of harmonizing a melody for a soprano, l is the length of the problem.

```
1  max_cost=Double.max;
2  total_cost=0;
3  Fill in the first chord, using narrow spacing;
4  total_cost=cost of this first chord (only w3c3);
5  Calculate_Solution(2,l);
6  Fill in the first chord, using wide spacing;
7  total_cost=cost of this first chord (only w3c3);
8  Calculate_Solution(2,l);
```

Procedure `Calculate_Solution`(**note position** *i*, **total length** *l*)

```
1  if (i < l) then
2      foreach possibility for note i do
3          if no hard constraints violated then
4              calculate costs ;
5              if total_cost+Σ_{i=1}^{5} w_i c_i > max_cost then
6                  go to next possibility;
7              else
8                  total_cost=total_cost+Σ_{i=1}^{5} w_i c_i − w_2 c_2;
9                  Calculate_Solution(i+1);
10             end
11         else
12             go to next possibility;
13         end
14     end
15 else
16     max_cost=total_cost+w_2 c_2;
17     save solution;
18 end
```

7 Results

In this section we present some test results. We tested the program for several problems but we will show only one representative example here.

We illustrate the suitability of our approach by showing four

ISBN: 978-989-95079-6-8

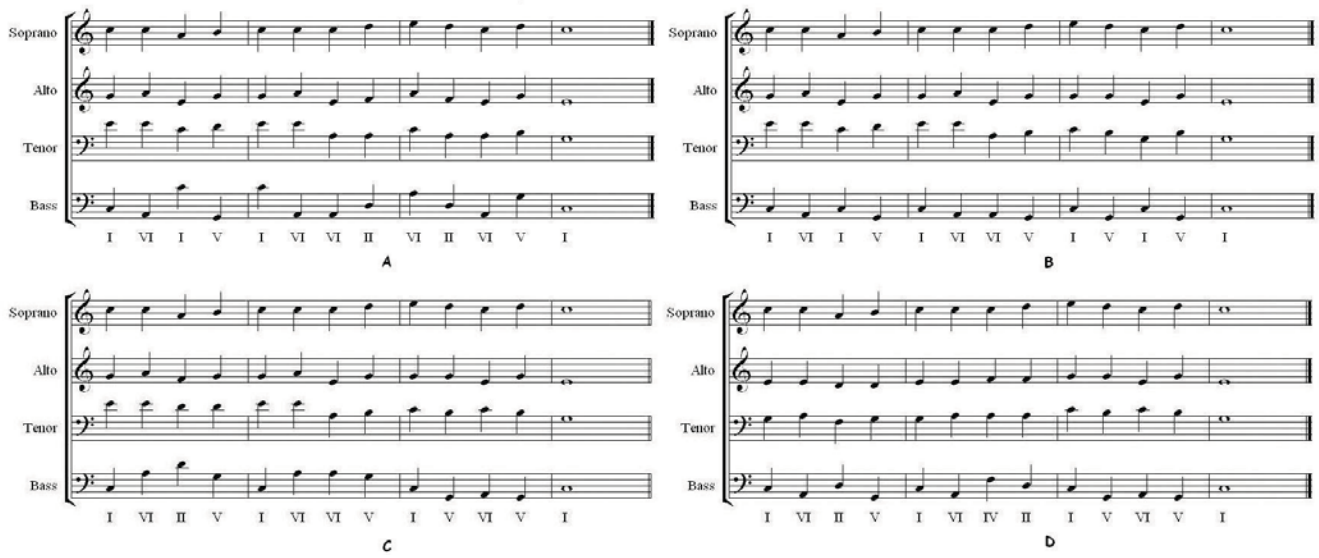


Figure 3: A representative harmonization exercise, showing four solutions: (A), (B), (C) show solutions with weights as proposed in Section 5. (A) has cost 3.14, (B) has cost 1.12 and (C) is the optimal solution with cost 0.65. (D) shows the optimal solution with weights $w_1 = 0.3, w_2 = 0.2, w_3 = 0.1, w_4 = 0.3$ and $w_5 = 0.1$. It has cost 0.47.

solutions with different costs. In Fig. 3A we show the solution with the highest cost (3.14). In this solution there are unacceptable distances in the bass. Furthermore, the melodies in the alto and tenor are not fluent, there is almost no contrary motion between soprano and bass and the grade sequence is bad.

Fig. 3B shows a solution with cost 1.12. This solution has a better bass line. The grade sequence is slightly better since we now have tension (grade V) in the second bar. This solution sometimes lacks contrary movement and there is no second grade used.

The optimal solution is shown in Fig. 3C. The grade sequence and grade frequencies are very well but there are many large distances between the bass notes. This can be explained by the weights we assigned to the costs: contrary motion is considered equally important as the distances between the notes. If we raise the weights corresponding to the distances between notes and we lower the weight corresponding to the use of contrary movement we obtain the result in Fig. 3D. Except from the fact that there is no tension in the second bar this is a good solution.

8 Practical Use And Further Work

The algorithm we presented is designed to solve harmonizing problems. It can be used as a didactic tool: teachers can use it to test quickly if problems have solutions, if these solutions are interesting enough and if they are not too difficult to solve. For instance, if a problem has only one solution where only grades I and IV are used, this problem cannot be solved well. Furthermore, students can use this program to compare their solutions to an optimal solution.

The model we presented can be extended in various ways. First of all, we can overcome the simplification of the harmonization problem by allowing modulations, chord inversions, non-chord tones, rhythm and ornaments, seventh chords and so on. This requires an extension of the domains and the introduction of new hard constraints. Furthermore, we can also introduce more soft constraints. For instance, some counterpoint rules can be expressed as soft constraints.

As illustrated in the previous section, the assignment of the weights is important. Fine-tuning of these parameters is needed and should be done more profoundly.

Acknowledgment

Chris Cornelis and Yvan Saeys thank the Research–Foundation Flanders for funding their research.

References

[1] C. Truchet and P. Codognet. Musical constraint satisfaction problems solved with adaptive search. *Soft Comput.*, 8(9):633–640, 2004.

[2] François Pachet and Pierre Roy. Musical harmonization with constraints: A survey. *Constraints*, 6(1):7–19, 2001.

[3] Kemal Ebcioğlu. An expert system for harmonizing chorales in the style of J. S. Bach. pages 294–333, 1992.

[4] Stéphanie Vanhove. Backtrackalgoritmen voor regelgebaseerde vierstemmige harmonisatie. Master's thesis, Ghent University, 2007.

[5] Somnuk Phon-amnuaisuk, Andrew Tuson, and Geraint Wiggins. *Evolving Musical Harmonisation*. 1999.

[6] Somnuk Phon-amnuaisuk and Geraint A. Wiggins. The four-part harmonisation problem: a comparison between genetic algorithms and a rule-based system. In *Proceedings of the AISB99 Symposium on Musical Creativity*, pages 28–34. AISB, 1999.

[7] Allan Forte. *Tonal Harmony in Concept and Practice*. 1962.

[8] Imogen Holst. *An ABC of music: a short practical guide to the basic essentials of rudiments, harmony, and form. With a foreword by Benjamin Britten*. 1963.

[9] Pedro Meseguer, Noureddine Bouhmala, Taoufik Bouzoubaa, Morten Irgens, and Martí Sánchez. Current approaches for solving over-constrained problems. *Constraints*, 8:9–39, 2003.

ISBN: 978-989-95079-6-8

Detecting and Reacting on Drifts and Shifts in On-Line Data Streams with Evolving Fuzzy Systems *

Edwin Lughofer[1] Plamen Angelov[2]

1. Department of Knowledge-Based Mathematical Systems, Johannes Kepler University
Altenbergerstrasse 69, A-4040 Linz, Austria
2. Department of Communication Systems, Infolab 21, Lancaster University
Lancaster, LA1 4WA, UK
Email: edwin.lughofer@jku.at, p.angelov@lancaster.ac.uk

Abstract— *In this paper, we present new approaches to handle* drift *and* shift *in on-line data streams using evolving fuzzy systems (EFS), which are characterized by the fact that their structure is not fixed and not pre-determined. When dealing with* drifts *and* shifts *in data streams one needs to take into account two major issues: a) automatic detection of, and b) automatic reaction to this. To address the first problem we propose an approach based on the concepts of* age *and* utility *of fuzzy rules/clusters. The second problem itself is composed of two sub-problems concerning the influence of the* drifts *and* shifts *on: 1) the antecedent parts (fuzzy set and rule structure) and 2) the consequent parts (parameters) of the fuzzy models. To address the latter sub-problem we propose an approach that introduces a gradual forgetting strategy in the local learning process. To address the former sub-problem we introduce two alternative methods: one that is based on the evolving density-based clustering, eClustering (used in* eTS*); and one that is based on the automatic adaptation of the learning rate of the evolving vector quantization approach (eVQ) (used in* FLEXFIS*). The paper is concluded with an empirical evaluation of the impact of the proposed approaches in (on-line) real-world data sets where* drifts *and* shifts *occur.*

Keywords— drifts and shifts in data streams, evolving fuzzy systems, detection and reaction to drifts and shifts, age of a cluster/fuzzy rule, gradual forgetting

1 Introduction

1.1 Motivation and State of the Art

Nowadays data-driven fuzzy systems enjoy a great attraction in many industrial applications, as opposed to expert-based fuzzy systems. They can be automatically generated from process data such as measurements, images (features) or signal streams. Furthermore, they are proven universal approximators [27], i.e. being able to model a given problem to (theoretically) any degree of accuracy. Moreover, they also allow an insight in the form of linguistically and visually interpretable rules to be gained [8].

During the last decade, the research field of 'evolving fuzzy systems' (EFS) emerged as an important part of the fuzzy systems research [4] [11], as they are capable to include new information on demand into the models and on-the-fly without necessarily using prior knowledge. EFS can also permanently learn from their environment and are applicable in fast on-line identification [4] and modelling processes as well as huge data bases which cannot be loaded into the virtual memory at once [18]. Often, there are no enough data in advance (off-line) to build reliable models with high predictive quality which can also require application of EFS as in [23].

Various approaches for EFS have been set up during the last years, one of the most popular and pioneering approach is the *eTS* family which comes with a regression [4], [3] and a classification variant [5]. Another approach for adaptation and evolution of clusters inspired by the evolving vector quantization (eVQ) [19] is the so-called *FLEXFIS* family [21], in particular *FLEXFIS* for regression [20] and for classification [22] (denoted as *FLEXFIS-Class*). A range of other alternative approaches includes ePL [16], SAFIS [12], and evolving fuzzy neural networks such as SOFNN [15] or GDFNN [28].

All these methods have a common denominator: they all are life-long learning approaches, which means that they incorporate all the data samples into the fuzzy models with equal weights in the same order as they are coming in during the on-line process. Hence, fuzzy models reflect a compact information of all the samples seen so far with equal importance. This is a beneficial to adapt the models, especially when a convergence to an optimality criterion or stable state of the model structure is achievable [20]. However, this benefit is only true in case of data streams which are generated from the same underlying data distribution respectively which do not show any *drift* or *shift* to other parts of the input/output space [26]. *Drift* (respectively *shift*) indicate the necessity of out-dating of previously learned relationships (in terms of structure and parameters) as these are not valid any longer and hence should be eliminated from the model. In order to cope with this problem, *drift* handling was already applied in other machine learning techniques, e.g. in connection with SVMs [14], ensemble classifiers [24] or instance-based (lazy) learning approaches [7] but to the best of our knowledge this concept was not yet applied to fuzzy systems.

*This work was partially supported by the Upper Austrian Technology and Research Promotion and by The Royal Academy, UK. This publication reflects only the authors' views.

ISBN: 978-989-95079-6-8

1.2 Our Approach

Hence, in this paper we are dealing with approaches for an appropriate handling of *drifts* and *shifts* by EFS. However, parts of the concepts, especially the detection of, and reaction to *drifts* in the consequent parts of the rules, can be applied to a wider range of EFS approaches including *eTS* and *FLEXFIS*. A more detailed description will be given in Section 2. For achieving an automatic approach, we propose 1.) the detection of a *drift* respectively *shift* based on the concepts of fuzzy rule *age* and *utility* function, and 2.) the reaction to a detected *drift* respectively *shift*, both in rule antecedent (for *eClustering* as applied in *eTS* and *eVQ* as applied in *FLEXFIS*) as well as consequent parts of Takagi-Sugeno fuzzy models [25].

The paper is concluded with an empirical evaluation of the impact of the proposed approaches on predictive accuracy of the evolving fuzzy models when applying to real-world data streams where *drifts* and *shifts* occur (Section 5). Note, that no benchmark data from the Internet or well-known data bases can be applied for empirical evaluation as these are usually all smooth in the sense that no *drift* occurs therein.

2 Problem Statement

In machine learning literature, they distinguish between different types of 'concept change' of the underlying distribution of (on-line) data streams: a) *drifts*, and b) *shifts*, see [26]. *Drift* refers to a *gradual* evolution of the concept over time. The concept *drift* concerns the way the data distribution slides smoothly through the data/feature space from one region to another. For instance, one may consider a data cluster moving from one position to another. This concept is closely related to the time-space representation of the data streams. While the concept of (data) density is represented in the data space domain, *drift* and *shift* are concepts in the joint data-time space domain.

An (artificial) example of a *drift* is demonstrated in Figure 1. There, the original data distribution (in a 2-D data space) is marked by circular samples, which changes over time into a data distribution marked by rectangular samples. If a conventional clustering process is applied by weighting all new incoming samples equally, the cluster center would end up exactly in the middle of the whole data bunch, averaging old and new data. Such a clustering approach is applied in EFS approaches to identify the local regions that can be used to form (the antecedent parts of the) fuzzy rules.

On the other hand, the concept *shift* refers to a sudden, abrupt change of the underlying concept to be learned. A *shift* in the input space opens up a data cloud in an unexplored region and hence usually automatically causes a new rule to be evolved by the incremental/evolving clustering algorithm. In Figure 2 a case of a *shift* in the output variable is shown: the original trajectory (consisting of dense data samples) on the right-hand side is marked with a light line, whereas the *shift* is represented by the dark data samples forming a trajectory above the other one. In case if only apply the usual adaptation

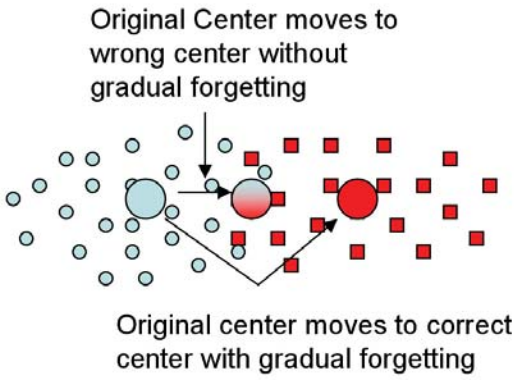


Figure 1: A *drift* in an evolving cluster (used for learning the antecedent parts of EFS), the distribution before the drift shown in circular, after the drift shown in rectangular samples.

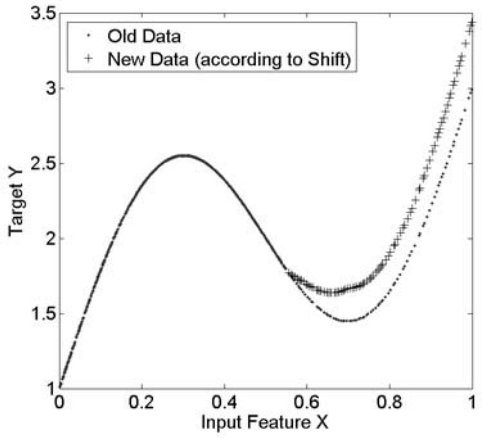


Figure 2: Example of a *shift* in the output variable; compare light dots (original data distribution) with dark dots (data distribution after the *shift*) on the right-hand side of the image (also marked by arrows)

of the consequent parameters with weighted recursive least squares without forgetting the old data due to the *shift* the approximation curve of the fuzzy model will end up exactly in-between these two.

In this paper, we demonstrate novel approaches for autonomous *drift* and *shift* detection and handling when learning fuzzy rule-based systems of Takagi-Sugeno type from on-line data streams in an evolving manner [4]. We focus on EFS approaches exploiting the Takagi-Sugeno model architecture [25].

3 Autonomous Detection of Drifts and Shifts in Data Streams by EFS

In this section the method for autonomous detection of *shifts* and *drifts* in data streams based on the *age* and *utility* of the cluster/fuzzy rule [1], [10] is described. This is an important step in the process of handling non-stationarity in data streams and for building au-

ISBN: 978-989-95079-6-8

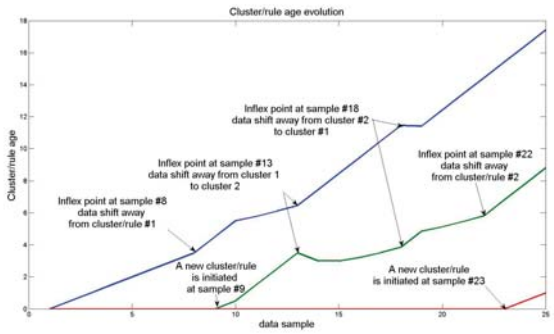


Figure 3: Evolution of the *age* of the clusters/fuzzy rules; *drifts* are related to the inflexion points; *shifts* are related to forming new clusters/fuzzy rules.

tonomous self-developing, self-learning, and evolving models and systems.

3.1 Detecting Drifts by Age of Clusters (Rules) [1]

In [1], the concept of cluster (respectively fuzzy rule) *age* was introduced. We extend this by including the membership degrees of the rules ($\Phi_{i,l}$ in case of the ith rule for the lth sample)

$$age^i = k - \frac{\sum_{l=1}^{n_i} I_l \Phi_{i,l}}{n_i} \quad (1)$$

where i is the rule index; n_i denotes the support of rule i; I_j denotes the time instance when the data sample was read; k is the current time instance. Since $\frac{\sum_{l=1}^{n_i} I_l}{n_i}$ can also be accumulated recursively, the *age* can be easily calculated when necessary. *Age* of the cluster/rule changes with each new data sample being read. If the newly read data sample does not fall into that cluster (supports that fuzzy rule with a low or even 0 value of $\Phi_{i,l}$ in case of fuzzy sets with finite support) the *age* grows by the rate of one sample at a time. That means, if a certain cluster (fuzzy rule) is not supported by any future samples after being initiated (say in the time instant, t_i) then its *age* at any time instant, k will be $k - t_i$. However, if any new data sample (between t_i and k) falls into that cluster (supports that fuzzy rule with a high value of $\Phi_{i,l}$), the *age* does not grow with the same rate, but with a smaller one and one can say that the cluster (fuzzy rule) is being *refreshed*.

The fuzzy rule *age* is important and closely linked to the data streams (which are sequences of data in time) and to the concept *drift*. We propose to analyze the *age* of clusters (respectively, fuzzy rules) on-line by using the gradient of the *ageing* curve as well as its second derivative which indicate a change in the slope of the *ageing* curve. When there is a **significant** change of the *ageing* which results in a significant change of the slope, then obviously the second derivative of the *age* curve will be indicative of this inflection points. An example is demonstrated in Figure 3, where the *drift* and *shift* are clearly marked on the *age* evolution curves.

3.2 Detecting Shifts by Utility Function

Shift in the data streams is a more significant, abrupt and sudden change of the data concept. Therefore, the distinction between *drift* and *shift* is the pace, the degree and speed of the changes, while both indicate a change in the data distributions with time. Essentially, the reaction to a detected *shift* has to be reflected in the structure evolution of the fuzzy rule base. While *drift* is related more to a smooth forgetting of parameters, learning and forgetting, *shift* is closely related to changes in the structural components.

Shift can be detected by the *utility* of the fuzzy rule. *Utility* is a parameter of the quality of the fuzzy rule that is defined [2] as the accumulated firing level of each rule given by Ψ_i summed over the life of each rule:

$$U_k^i = \frac{1}{k - I_k^i} \sum_{l=1}^{I_k^i} \Psi^l \quad (2)$$

$Utility$, U accumulates the weight of the rule contributions to the overall output during the life of the rule (from the current time instant back to the moment when this rule was generated). It is a measure of importance of the respective fuzzy rule comparing to the other rules (comparison is hidden in the relative nature of Ψ).

If the *utility* of a fuzzy rule is *low* then this rule becomes obsolete or not used very much. A *shift* away from or to a cluster/fuzzy rule can be detected by the derivative of the *utility* — if the value of the first derivative of the *utility* is *large* then there is a *shift* towards the cluster. The *shift* is away from a cluster if the derivative of the *utility* is *large* negative. Then this fuzzy rule is deemed obsolete (does not contribute significantly to the overall prediction/estimation/classification) and can be removed from the rule base [10].

4 Reactions to Drifts and Shifts in Data Streams by EFS

4.1 Reaction to Drift and Shift in the Antecedents

In this section, we introduce methods for addressing both *drifts* and *shifts* in data streams by adapting learning mechanisms for consequent parts and by evolving antecedent part. We demonstrate this on the example of the popular *eTS* method [4] and on *FLEXFIS* [20].

4.1.1 *By* eClustering

Reaction to a detected *shift* is by either a) forming a new rule around a new data sample which becomes an attraction point for the global data distribution, or b) replacement of a fuzzy rule which itself consists of; i) forming a new rule around the new point, and ii) removal of the rule which has lower density and is close to this newly added one [2] — see Figure 4. If locally optimal learning is being applied then removing a cluster and respectively a locally valid Kalman filter/RLS does not affect the overall learning significantly (only through the fuzzy weight Ψ [4]). If globally optimal learning is being applied, removal of a cluster (respectively fuzzy rules) does affect n columns and n rows of the covariance matrix directly and the remaining values of the covariance

ISBN: 978-989-95079-6-8

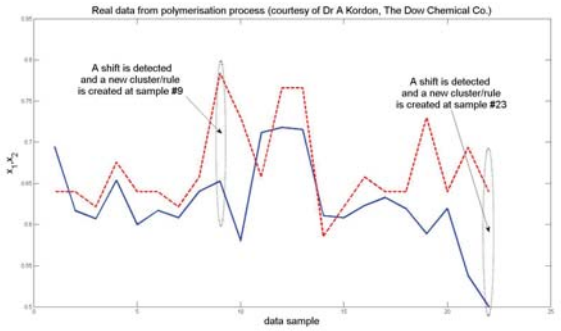


Figure 4: Shift in the time domain representation and two new rules evolved as a reaction to it.

matrix indirectly [4]. Deleting *old* fuzzy rules was applied in classification in [5] and in prediction [2].

4.1.2 *By* eVQ

When *shifts* in the data stream occur, usually new clusters are evolved automatically in *eVQ*, as they cause new data clouds in previously unexplored regions of the data space (in regions further away than a fraction of the space diagonal in the input space to which the vigilance parameter ρ is usually set, i.e. 0.1 to 0.3 of the space diagonal, see [19]. Synchronously, older clusters can be deleted by the concept of *utility* [9] [10] as discussed in Section 3.2. This does not affect the learning of consequent parameters for the other rules if local learning is applied.

For reacting to *drifts* in the antecedent parts of the rules we propose to re-adjust the parameter in the eVQ clustering algorithm η [19] which steers the learning gain. We define the tracking concepts for the ith rule throughout this section, which can be generalized to any rule in a straightforward manner. Currently, in *eVQ* the learning gain is defined by the following formula [18]:

$$\eta_i = \frac{0.5}{n_i} \tag{3}$$

If a *drift* occurs in a data stream, the centers and widths of the cluster(s) should change to the new data distribution (as shown in Figure 1). For re-activating the converged clusters, i.e. re-animating them for stronger movements in a drifting case, we suggest a sudden increase for the first sample in the *drift* phase, followed by a gradual decrease for the next samples in order to balance in the new sample distribution in the same manner as is done for original ones.

Here, we propose the following mechanisms for the learning gain η: first we transform the forgetting factor λ, used in the gradual out-weighting when doing consequent learning (see next section) and denoting the intensity of a *drift*, to a value in $[0, 1]$. Hereby, 0.9 (minimal value for λ) is mapped to 0.99, whereas 1 is mapped to 0, hence:

$$\lambda\_trans = -9.9\lambda + 9.9 \tag{4}$$

Then, when a *drift* occurs, we re-set the number of samples forming the ith cluster (n_i) (used in the denominator of the calculation of η_i) by

$$n_i = n_i - n_i * \lambda\_trans \tag{5}$$

This means that the stronger the drift is, the more n_i is decreased and hence the stronger the forgetting effect will be. In Figure 5 it is demonstrated how η_i develops (lines) in usual (non-drift) case (for the first 100 samples), then a drifting scenario is artificially caused with three intensities leading to the three λ values (in different line styles). After the *drift* indicator (at sample 100), it is decreased in usual way such that the jumped center can converge to the new data distribution.

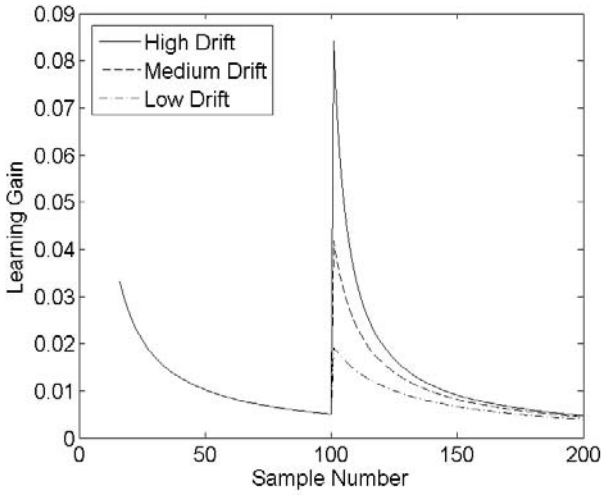


Figure 5: Abrupt increasing learning gain η during in case of a *drift* (after 100 samples) when applying different values for the forgetting factor λ.

4.2 *Reaction to Drifts and Shifts in the Consequents*

For the rule consequents *drifts* and *shifts* can be handled in one sweep as it is just a matter of the setting of the forgetting factor as we will see below. Whenever a *drift* in the output variable occurs (as shown in Figure 2) and is detected, it is necessary to apply a specific mechanism in the sample-wise incremental learning steps of the consequent parameters in Takagi-Sugeno fuzzy systems. When locally optimal fuzzily weighted Recursive Least Squares (wRLS) learning approach [4] is used this can easily be accommodated as detailed later on.

If we do not take *drift* into account all newer samples lying in the same local positions relative to the rules as the older samples are included with the same rule weights in the update process. Then the fuzzy model will end up with approximation curve as shown in the left image of Figure 6 with dotted lines. Obviously, the approximation ends up in the middle of the two trajectories (the newer after the *drift* shown in light font, the older shown in darker font), as trying to minimize the quadratic errors (least squares) of all samples to the curve. Hence, it is necessary to include a parameter in the update process, which forces older samples to be out-dated over time. Gradualism is important here in

ISBN: 978-989-95079-6-8

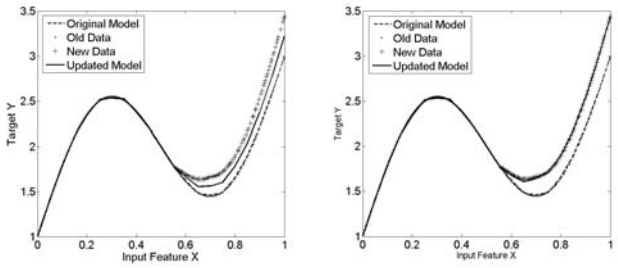


Figure 6: Left: The adapted model (dotted line) with the new incoming samples (dark dots) when applying conventional recursive weighted least squares approach; right: the adapted model (dotted line) by including a forgetting factor of 0.95 and using (7)-(9), the approximation surface lying exactly on the trajectory of the data samples denoting the novel data distribution

order to guarantee a smooth forgetting and to prevent abrupt changes in the approximation surface. For doing so, we re-define the least squares optimization function for the ith rule by

$$J_i = \sum_{k=1}^{N} \lambda^{N-k} \Psi_i(\vec{x}(k)) e_i^2(k) \longrightarrow \min_w \quad (6)$$

with $e_i(k) = y(k) - \hat{y}(k)$ the error of the ith rule in sample k. Assuming N the number of samples loaded so far, this function out-dates the sample processed i steps ago by λ^{N-k}. Usual values of λ lie between 0.9 and 1, where a value near 1 means a slow forgetting and a value near 0.9 a fast forgetting of former loaded data samples and the exact choice depends strongly on the strength of the *drift* (see below). Following a similar recursive deduction scheme as in conventional recursive least squares (RLS) [17], we obtain the following incremental update formulas for consequent parameters $\vec{w}_i$ for the ith rule:

$$\hat{\vec{w}}_i(k+1) = \hat{\vec{w}}_i(k) + \gamma(k)(y(k+1) - \vec{r}^T(k+1)\hat{\vec{w}}_i(k)) \quad (7)$$

$$\gamma(k) = \frac{P_i(k)\vec{r}(k+1)}{\frac{\lambda}{\Psi_i(\vec{x}(k+1))} + \vec{r}^T(k+1)P_i(k)\vec{r}(k+1)} \quad (8)$$

$$P_i(k+1) = (I - \gamma(k)\vec{r}^T(k+1))P_i(k)\frac{1}{\lambda} \quad (9)$$

with $P_i(k) = (R_i(k)^T Q_i(k) R_i(k))^{-1}$ the covariance matrix and $\vec{r}(k+1) = [1 \;\; x_1(k+1) \;\; x_2(k+1) \;\; \ldots \;\; x_p(k+1)]^T$ the regressor values of the $k+1$th data sample, Ψ_i the fulfillment degree of the ith rule, serving as weight in the recursive least squares algorithm.

The final question is how to set the parameter λ in order to guarantee an appropriate *drift* tracking. We propose a strategy to deduce it directly from the *age* curves analysis [10] since they are indicative of the speed of a *drift* (see Section 3). In Section 3, it was mentioned that the age of a rule always lies in $[0, k]$. Hence, we normalize the *age* of the ith rule to $[0, 1]$ by $age_i\_norm = \frac{age_i}{k}$ in order to achieve gradients of the normalized rule ages Δage_i_norm also lying in $[0, 1]$. Whenever the change of the gradient in the rule *age* curve is significant, recursive weighted RLS with forgetting (wRLSf) should be triggered. We use the following estimation for λ:

$$\lambda = 1 - 0.1\Delta^2 age_i\_norm \quad (10)$$

This guarantees a λ between 0.9 (strong forgetting) and 1 (no forgetting), according to the degree of the gradient change (1 = maximal change, 0 = no change). The forgetting factor is then kept for a while at this level (otherwise only one single sample would cause a gradual forgetting) and set back to 1, after a stable gradient phase is achieved (usually after around 20 to 30 samples showing a moderate value of $\Delta^2 age_i\_norm$). Setting back to 1 is necessary, as otherwise the forgetting will go on inside the new data distribution. This cause the drift phase in the antecedents to stop.

5 Evaluation

This section deals with the evaluation of the impact of reacting on *drifts* and *shifts* in case of data streams where actually *drifts* and *shifts* occur. This is done by implementing the approaches discussed throughout this paper. One application example is coming from a rolling mill, where the task was to identify a prediction model on-line for the resistance value of a steel plate at a rolling mill. The other application concerns a polymerization process in chemical industry.

5.1 On-Line Prediction Models at Rolling Mills

The task was to identify a prediction model for the resistance value of a steel plate at a rolling mill. This should be done in a first step with some off-line (precollected) measurement data in order to obtain a feeling about the achieved quality of the fuzzy model and then to refine the prediction model with newly recorded online data. The later step was possible as first a prediction for the resistance is given, influencing the whole process at the rolling mill, whereas a few seconds later (after the steel plate is passed), the real value for the resistance is measured, which can then be incorporated into the model adaptation process. In this sense, the correct measured value not the predicted (which might have been wrong)one is taken for learning. In fact, an improvement in terms of predictive power could be achieved when updating the fuzzy models trained in batch mode with 6000 samples during on-line operation mode with further 6600 samples. For details of the experimental setup and results see [20] (and also Table 1 below summarizing all the results).

Now, in this paper we want to examine whether a reaction onto drifts by a gradual forgetting of older samples during the on-line process may further improve the quality of the models. A justification of an application of reaction onto *drifts* is that the operation process at rolling mills is divided into different "stitches". One stitch represents one closed cycle in the rolling process. In the on-line mode the measurements come in continuously from stitch to stitch. However, for the current processed stitch, the previous stitch should play only little or even no role. However, the measurements from the

ISBN: 978-989-95079-6-8

Table 1: Comparison of prediction with and without applying gradual forgetting

Method	**MAE**	**Max MAE Too High / Max MAE Too Low / # MAEs > 20**
Analytical	7.84	63.47 / 87.37 / 259
Static fuzzy models	6.76	45.05 / 81.65 / 176
FLEXFIS	5.41	38.05 / 78.88 / 159
FLEXFIS with forg.	4.65	31.99 / 74 / 68

previous stitch are already included in the fuzzy models as updated by their samples. Thus, this means that older samples from the previous stitch should be forgotten when including samples from the current stitch. Another aspect is that here we do not need any drift/shift detection, as the drift/shift is indicated by the beginning of a new stitch. As no *drift* detection was carried out, λ was set to 0.97, which is a good compromise between fast forgetting (=strong locality of models) and low forgetting (=weak locality of models). The results are demonstrated in Table 1. Here, we also demonstrate the improvement of the predictive accuracy over analytical models by both, static and evolving fuzzy models. Three different types of errors are reported: the mean absolute error over all on-line samples (note that first a prediction is done and then the model updated with the same samples and based on feedback), the number of mean absolute error greater than 20, the maximal mean absolute error over all samples where the prediction was too low and the maximal mean absolute error over all samples where the prediction was too high. The latter value is the most important one as harming the steel plate is more dangerous as in case of predicting too low values. The results (Table 1) demonstrate the impact of the gradual forgetting.

Another interesting aspect is that the error on the single measurements starts to *drift* over time when gradual forgetting is not applied. This is underlined in the left image of Figure 7 which shows the single errors over the 6600 on-line samples: note the *drift* of the main error area away from the zero line at the end of the data stream.

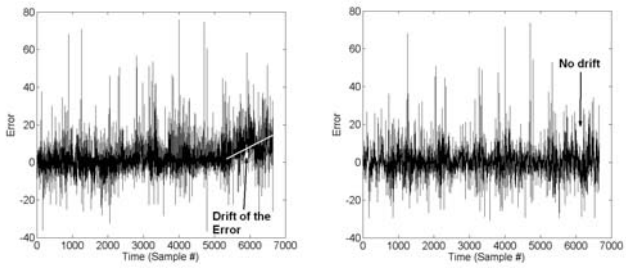


Figure 7: Left: The error curve for the 6600 on-line samples when no forgetting is applied: at the end the error starts drifting; right: no *drift*.

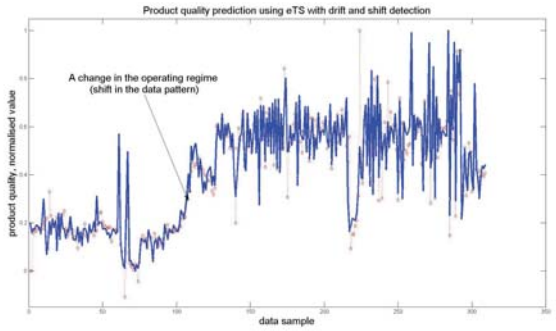


Figure 8: A comparison of the predicted and real data of product quality. Solid line - predictions by *eTS* with *shift* and *drift* detection; diamonds - real data.

5.2 *Another application example*

A case study based on real data (courtesy of Dr. Arthur Kordon, The Dow Chemical Co.) from the chemical industry is used as an illustration of the detection and reaction to *drift* and *shift* in *eTS* [6]. The *eTS* has been applied for prediction of the properties of a chemical composition by modelling the product composition in a distillation tower. The data set includes a change of the operating regime of the process which brings a challenge to the structure of the model (fuzzy rule based system). The data set includes also a number of other challenges, such as noise in the data, a large number of initial variables, etc. These problems cover a wide range of real issues in the industry. The process data is retrieved from physical ('hard') sensors used as inputs to the *eTS* applying hourly averages for every eight hours. The product composition (real output) is estimated by a laboratory analysis for comparison (it is given with diamonds in the Figure 8).

The estimation of the product composition contains noise due to the nature of the analysis. A significant operating condition change takes place after sample 127 (please see the Figure 8). The *eTS* was able to efficiently detect and react to this *shift* as well as to the *drifts* as depicted in Figures 3 and 4. The overall prediction is very good and compares favorably with the conventional models as detailed in the Table 2. Note, that the non-dimensional error index (NDEI) is defined as the ratio of the root mean square error over the standard deviation of the target data and should be ideally 0 while the variance accounted for (VAF) is defined as the ratio between the variance of the real data and the model output and is given out of a maximum of 100 (when the predictions coincide with the real data).

6 Conclusion

In this paper, we propose novel strategies and techniques for addressing concept *drift* and *shift* in on-line data streams. Therefore, two EFS approaches (*eTS* and *FLEXFIS*) are exploited as on-line modelling methodologies. These are extended by mechanisms which are 1.) able to detect *drifts* and *shifts* with fuzzy rule *age* and *util-*

ISBN: 978-989-95079-6-8

Table 2: Error measures when applying *eTS* with (Column #3) and without (Column #2) drift detection and reaction in chemical composition modelling

Measure	Without	With	Best (theor.) value
NDEI	0.3559	0.33465	0
VAF, %	87.319	88.807	100
correlation	0.9357	0.94285	1

ity, and 2.) to react on such occurrences appropriately. The latter is applied 1.) to the rule antecedent parts directly in the cluster space for reacting on *drifts* and *shifts* in the input space and 2.) to rule consequent parameters for reacting on *drifts* and *shifts* in the output variable, applicable to any EFS technique exploiting Takagi-Sugeno type fuzzy systems. Evaluation on real-world data sets shows that the novel techniques are able to improve the accuracy and stability of the fuzzy models, whenever the occurrence of *drifts* and *shifts* is present.

References

[1] P. Angelov and D. Filev. Simpl_eTS: A simplified method for learning evolving Takagi-Sugeno fuzzy models. In *Proceedings of FUZZ-IEEE 2005*, pages 1068–1073, Reno, Nevada, U.S.A., 2005.

[2] P. Angelov and X. Zhou. Evolving fuzzy-rule-based classifiers from data streams. *IEEE Transactions on Fuzzy Systems*, 16(6):1462–1475, 2008.

[3] P.P. Angelov. Evolving Takagi-Sugeno fuzzy systems from streaming data, eTS+. In P. Angelov, D. Filev, and N. Kasabov, editors, *Evolving Intelligent Systems: Methodology and Applications*, page to appear. John Wiley & Sons, New York, 2008.

[4] P.P. Angelov and D. Filev. An approach to online identification of Takagi-Sugeno fuzzy models. *IEEE Trans. on Systems, Man and Cybernetics, part B*, 34(1):484–498, 2004.

[5] P.P. Angelov, E. Lughofer, and X. Zhou. Evolving fuzzy classifiers using different model architectures. *Fuzzy Sets and Systems*, 159(23):3160–3182, 2008.

[6] P.P. Angelov and X.-W. Zhou. Evolving fuzzy systems from data streams in real-time. In *2006 International Symposium on Evolving Fuzzy Systems*, pages 29–35, 2006.

[7] J. Beringer and E. Hüllermeier. Efficient instance-based learning on data streams. *Intelligent Data Analysis*, 11(6):627–650, 2007.

[8] F. Hopner and F. Klawonn. Obtaining interpretable fuzzy models from fuzzy clustering and fuzzy regression. *Proc. 4th Intern. Conf. on Knowledge-based Intell. Eng. Syst. (KES)*, Brighton, UK, pages pp.162–165, 2000.

[9] P. Angelov. *Machine Learning (Collaborative Systems)*, patent (WO2008053161, priority date: 1 November 2006).

[10] P. Angelov and X.-W. Zhou. On Line Learning Fuzzy Rule-based System Structure from Data Streams. In *2008 IEEE International Conference on Fuzzy Systems within the IEEE World Congress on Computational Intelligence*, Hong Kong, June 1-6, 2008, pages 915–922, 2008.

[11] P.P. Angelov *Evolving Rule-based Models: A Tool for Flexible Systems Design*, Springer Physica Verlag, Heidelberg, Germany, February, 2002.

[12] N. Sundararajan H.-J. Rong, G.-B. Huang, and P. Saratchandran. Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction. *Fuzzy Sets and Syst.*, 157(9):1260–1275, 2006.

[13] N. K. Kasabov and Q. Song. DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE Trans. on Fuzzy Systems*, 10(2):144–154, 2002.

[14] R. Klinkenberg. Learning drifting concepts: Example selection vs. example weighting. *Intelligent Data Analysis*, vol. 8 (3), pages 281–300, 2004.

[15] G. Leng, T.M McGinnity, and G. Prasad. An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network. *Fuzzy Sets and Syst.*, 150(2):211–243, 2005.

[16] E. Lima, F. Gomide, and R. Ballini. Participatory Evolving Fuzzy Modeling. In P. Angelov, D. Filev, and N. Kasabov, editors, *2006 International Symposium on Evolving Fuzzy Systems*, IEEE Press, pages 36–41, 2006.

[17] L. Ljung. *System Identification: Theory for the User*. Prentice Hall PTR, Prentic Hall Inc., Upper Saddle River, New Jersey 07458, 1999.

[18] E. Lughofer. *Evolving Fuzzy Models — Incremental Learning, Interpretability and Stability Issues, Applications*. VDM Verlag Dr. Müller, Saarbrücken, 2008.

[19] E. Lughofer. Extensions of vector quantization for incremental clustering. *Pattern Recognition*, 41(3):995–1011, 2008.

[20] E. Lughofer. FLEXFIS: A robust incremental learning approach for evolving TS fuzzy models. *IEEE Trans. on Fuzzy Systems (sp. issue on EFS)*, 16(6):1393–1410, 2008.

[21] E. Lughofer. Towards robust evolving fuzzy systems. In P. Angelov, D. Filev, and N. Kasabov, editors, *Evolving Intelligent Systems: Methodology and Applications*, to appear. John Wiley & Sons, New York, 2009.

[22] E. Lughofer, P. Angelov, and X. Zhou. Evolving single- and multi-model fuzzy classifiers with FLEXFIS-Class. In *Proc. FUZZ-IEEE 2007*, pages 363–368, London, UK, 2007.

[23] E. Lughofer, J. E. Smith, M. A. Tahir, P. Caleb-Solly, C. Eitzinger, D. Sannen, and M. Nuttin. Human-Machine Interaction Issues in Quality Control Based on On-Line Image Classification. *IEEE Trans. on Systems, Man and Cybernetics part A: Systems and Humans*, 2009, to appear.

[24] S. Ramamurthy and R. Bhatnagar. Tracking recurrent concept drift in streaming data using ensemble classifiers. In *Proc. 6th Intern. Conf. on Machine Learning and Applic. (ICMLA), 2007*, pages 404–409, 2007.

[25] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. on Systems, Man and Cybernetics*, 15(1):116–132, 1985.

[26] A. Tsymbal. The problem of concept drift: definitions and related work. Technical Report TCD-CS-2004-15, Dept of Comp Science, Trinity College Dublin, Ireland, 2004.

[27] L.X. Wang. Fuzzy systems are universal approximators. In *Proc. 1st IEEE Conf. Fuzzy Systems*, pages 1163–1169, San Diego, CA, 1992.

[28] S. Wu, M.J. Er, and Y. Gao. A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks. *IEEE Trans. on Fuzzy Systems*, 9(4):578–594, 2001.

ISBN: 978-989-95079-6-8

How to learn fuzzy user preferences with variable objectives

Alan Eckhardt[1],[2] Peter Vojtáš[1],[2]

1.Department of Software Engineering, Charles University
Prague, Czech Republic
2.Institute of Computer Science, Czech Academy of Science
Prague, Czech Republic
Email: {eckhardt , vojtas}@ksi.mff.cuni.cz

***Abstract*— *This paper studies a possibility to learn a complex user preference model, based on CP-nets, from user ratings. This work is motivated by the need of user modelling in decision making support, for example in e-commerce. We extend our user model based on fuzzy logic to capture variation of preference objectives. The proposed method 2CP-regression is described and tested. 2CP-regression uses CP-nets idea behind and can be considered as learning of a simple CP-net from user ratings.*

***Keywords*—** user preferences, data mining, ceteris paribus

1 Introduction

The problem studied in this paper is based on the idea of helping user in making decisions. The main motivation lies in e-commerce, where user might benefit a recommendation of objects that might interest her without needing to process all objects manually. This recommendation should be rather automatic and transparent, because user typically does not want to fill in any complicated forms or invest much time in search.

When a user is buying a notebook, for example, she is considering the attributes such as price, manufacturer, size of RAM, harddisk, display. In traditional e-shop environment, it is possible to make some restrictions on these attributes and lower the number of notebooks the user has to process manually this way. But these restrictions are inflexible - when the user selects price from 1000$ to 2000$, a notebook with price 999$ would not appear in the result set. This is the reason for a flexible search, based on fuzzy user preferences.

In this paper we propose a method for handling more complicated user preferences. In [1] a phenomenon called ceteris paribus was described. Preference ceteris paribus means preference "all else being equal". When a user is comparing two notebooks, we can say that the size of RAM of 2GB is preferred to 1GB ceteris paribus, meaning that we assume that the remaining attributes of these notebooks are the same and their values are not important. A user model based on this idea was later proposed in [2] - a CP-network is a graph that captures preferences ceteris paribus or, in different view, it represents attributes, on which the preferences over the attribute RAM depend. And yet, as far as we know, there are no study of learning of CP-nets from user ratings. Our paper is a first subtle contribution to this field. We work with our user model based on fuzzy sets rather than CP-nets, still the proposal of learning the relations between attributes is strongly related to CP-nets.

In Section 2 our user model is described. Related work is studied in Section 3. In Section 4 is described in depth the proposed method for learning the relation between attributes. Then this approach is tested in Section 5 and we end with conclusion and future work in Section 6.

2 Two step user model

2.1 Notation

We will work with a set of objects X. Overall rating of an object is a fuzzy subset of X, i.e. a function $r(o) : X \to [0,1]$, where 0 means least preferred and 1 means most preferred. Every object has attributes $A_1, ..., A_N$ with domains $D_{A_1}, ..., D_{A_N}$, with one special attribute that serves as identification of object (ID). Let $X \subseteq \prod_{i=1}^{N} D_{A_i}$. We will use $X(a)$ when denoting a set of objects that have the attribute value a.

User model, in our view, is a method for representing user decision when considering user's preference of an object $o \in X$. Our user model consists of two steps. At the first step, attributes of o are normalised using fuzzy sets $f_i : D_{A_i} \to [0,1]$ (and here again, 0 is least preferred and 1 is most preferred value), so that the transformed object is in $[0,1]^N$. These fuzzy sets are also called objectives or preferences over attributes. At the second step, preference over attributes are aggregated into the rating of the whole object via a fuzzy aggregation function $@ : [0,1]^N \to [0,1]$. Aggregation function is also often called utility function.

We concentrate on learning of user models. User is expected to rate a small sample $S \subseteq X$ of objects ($r : S \to \{1,2,3,4,5\}$), where $|S| \ll |X|$. The size of training set is expected to be very small. We assume that these ratings were created by the user using some fuzzy sets $f_1, ..., f_N$ and an aggregation function $@$. Of course the user does not compute the result of a function, but her decision corresponds to this two step process.

Our model consist of $f_1, ..., f_N$ and of $@$ so we expect to construct these functions from user ratings of S. The learnt functions will be denoted as $\widehat{f_1}, ..., \widehat{f_N}$ and $\widehat{@}$. Learning of aggregation function is described in Section 2.2. This paper concentrates on learning of local preferences, which is described in Section 4.

2.2 Aggregation function learning

Aggregation function serves to combine preferences over attributes into a single rating that represents the preference of the object as a whole. There are many ways to aggregate local preferences, we are currently using a weighted average with

ISBN: 978-989-95079-6-8

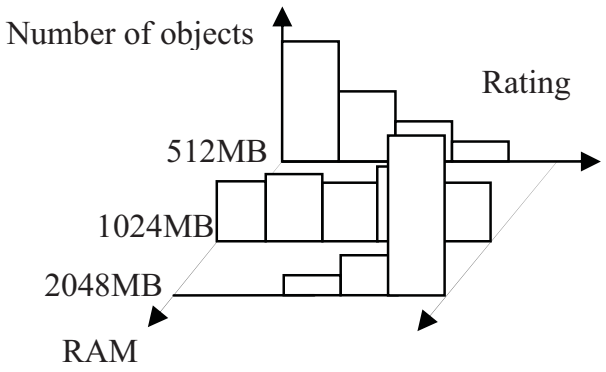


Figure 1: Uniform and expressive distributions of ratings.

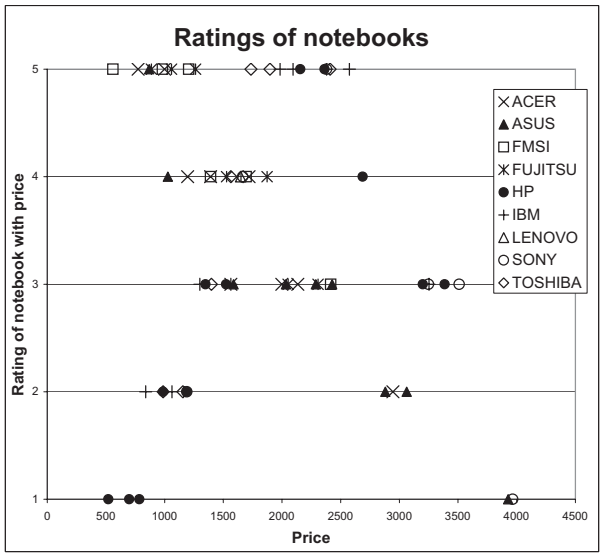


Figure 2: Distribution of ratings of notebooks with regard to price.

weights learnt from user ratings. Method called "Statistical" was described in [3], [4]. The distribution of ratings for attribute values $a_i \in A_i$ are considered when determining the weight of attribute A_i. In Figure 1 is an example of distribution of ratings across a domain of attribute RAM for a user that prefers larger sizes of RAM. If ratings are distributed near one point as in Figure 1 for 2048GB RAM, the attribute value is considered as decisive when doing the overall rating. However, if the distribution is rather uniform such as in Figure 1 for 1024MB RAM, then attribute value a_i does not play an important role in the overall rating. The measure of importance of an attribute value is computed with formula

$$imp(a_i) = 1/\sum_{o \in X(a_i)} |r(o) - avg_{o \in X(a_i)} r(o)|/|X(a_i)|$$

when taking only those objects o that have attribute value a_i. Then the importance of attribute A_i is computed as $W(A_i) = 1/(\sum_{a_i \in A_i} imp(a_i/|A_i|)$.

Our approach is not the only one - any data mining technique can be used here. Reader can simply imagine training a multilayer perceptron with normalised attribute values.

In Figure 2 is attribute price with ratings of notebooks. When a standard linear regression (described in Section 4.1) is used, attribute price would get a small weight. However, when notebooks of only a single manufacturer are used, then the error would be much lower and importance of price would be considerably higher. This may be visible for human, after a short inspection of the graph. Another visualisation of this situation is in Figure 3 with discretised attribute price. Ratings of all attribute values are distributed uniformly, so the importance of this attribute will be considered small.

3 Related work

User preferences are a wide field so there are plenty of research areas. For our case, user preference learning is the most important one. There are two totally different approaches to recommendation : content based and collaborative. For collaborative filtering, we may cite for example [5] but there are many other interesting works. Collaborative filtering is based on the idea that similar users tend to have similar preferences of an object. However, we do not consider this approach in this paper.

We concentrate here on content based, that means that the preferences are based on attributes of objects. For example, user considers the size of harddisk and other features when buying a notebook. This approach in our view corresponds more to how users decide in the real world. The difficulty here is that users are often inconsistent or take into account attributes that are not known or impossible to quantify, such as the design of notebook.

In this area, the main focus was on search of documents in the past [6, 7]. This is a specific area, because features of documents are of little structure. We are interested more in recommendation of more complicated objects like notebooks, with different types of attributes. Following works :[8, 9, 10, 11] deal with general, more structured, objects.

One of our main inspirations was CP-nets [2]. This user preference model captures complex user preferences in a graph representation, where a preference over one attribute may depend on the values of other attributes. There are many papers dealing with CP-nets, but there is none which describes a way to construct a CP-net automatically, by learning from ratings. Because preferences over attribute A_1 may depend on other attribute/s, e.g. A_2, user has to specify her preferences not only for all values of A_1 but she has to do it for every possible value of A_2. This means very much work and insight for the user, work that probably few people will undergo.

There is also another concept for representing user preferences other than ratings as in our case. These are preference relations, or specially fuzzy preference relations [12]. We do not consider this model, we rather follow approach identified by R.Fagin in [13] and consider user ratings as fuzzy sets.

4 Local preferences learning

As mentioned in Section 2, our user model is divided into two steps.

The first step is used for normalisation of every attribute to interval $[0, 1]$, so that the best object would have coordinates $[1, ..., 1]$. This can be viewed also as monotonisation of data - if for object o_1, the normalised value of its attribute value is lower in all attributes than object o_2, then o_2 is surely preferred to o_1 (if the normalisation is determined correctly).

ISBN: 978-989-95079-6-8

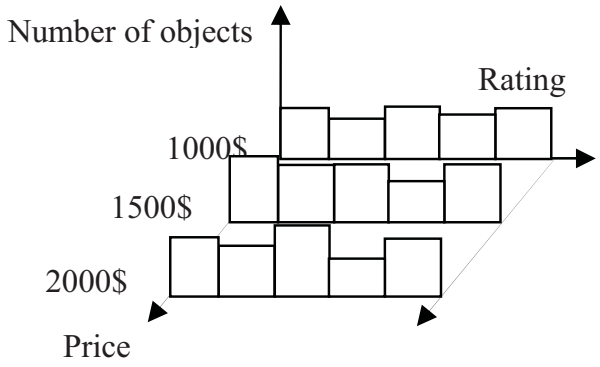


Figure 3: Uniform distributions of ratings across all domain.

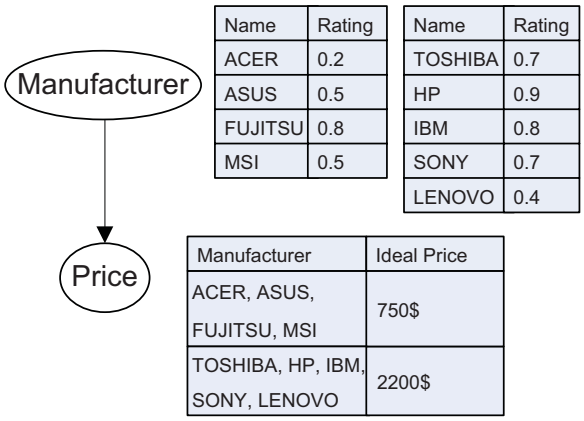


Figure 4: Example of a CP-net representing data about notebooks.

4.1 Linear regression

Linear regression is very useful method that serves to find a relation in a set of data in form of a linear function. It can be used for finding the preference of attribute values for numerical attributes, e.g. price. When given a set of notebooks' prices and ratings, linear regression will create a linear function $r(price) = \alpha * price + \beta$. Then we can normalise values of price using this function.

Traditional linear regression works on method of minimizing least squares of errors. This is not always useful, because it is dependent on the distribution of data points. For this reason, we proposed a method in [14] to accommodate better to the whole interval we study.

As a future work, we would like to implement a method for finding also more complicated functions, e.g. triangular fuzzy functions. The motivation is clear - linear functions are good for preferences of extreme values, e.g. lowest price, largest LCD etc. But if user prefers some value in the middle, like LCD of size 15", linear function would not capture this preference correctly.

4.2 2CP-regression

The main contribution of this paper is the proposal of a new method for acquisition of local preferences for numerical attributes, 2CP-regression. This method is motivated by preferences ceteris paribus [2] and CP-nets. Basic idea that represents ceteris paribus is that there are relations between attributes. For example, the manufacturer of a notebook influences the preferences of the price of that notebook. In our example, we present a user that prefers the price of 2200$ for manufacturers HP, IBM, Lenovo, Toshiba and Sony, and the price of 750$ for manufacturers Fujitsu-Siemens, Acer, Asus and MSI. A simple CP-net representing this scenario is in Figure 4, with ratings of manufacturers and ideal prices. Ideal price depends on the value of manufacturer, not on its preference. There are other examples and often the decisive attribute is a nominal one with rather small domain.

When constructing a fuzzy set for a numerical attribute A_1 (price in our case), we look at the values of other nominal attributes. For simplicity, let us consider only one, A_2. A_2 will be referred to as a "dividing attribute". The construction of the set is restricted to objects with value $a_2 \in A_2$; let us note this set $X(a_2) \subset X$. If dividing attribute A_2 really influences attribute A_1, then the values of A_1 should be "better" distributed in $X(a_2)$ than in whole X and the fuzzy set should better match the real preferences. The fuzzy set trained on $X(a_2)$ will be denoted as $\widehat{f_{a_2}^{A_1}}$. However, the size of training set may be radically decreased in this way. For that reason a general fuzzy set $\widehat{f^{A_1}}$ is constructed as in 4.1, which is used for values that did not occur during the construction of local preferences in the training set.

A method for construction of several fuzzy sets and a general one, all represented by linear functions, is described in the following pseudo code. Method `getObjectsWithA2(`a_2`)` returns the set $X(a_2)$, i.e. only those objects with attribute value a_2. Method `leaveOnlyA1AndRating` leaves only couples (A1, Rating) in the given set, which makes it suitable for linear regression that follows. Finally, the attribute value a_2 and the corresponding linear function $\widehat{f_{a_2}^{A_1}}$ are stored.

```
leaveOnlyA1AndRating(X);
f^{A_1} = buildClassifier(X);
for (∀a_2 ∈ A_2) {
    X(a_2) = getObjectsWithA2(a_2);
    if (|X(a_2)|< 2)
        continue;
    leaveOnlyA1AndRating(X(a_2));
    f_{a_2}^{A_1} = doLinearRegressionOn(X(a_2));
    storeCouple(a_2, f_{a_2}^{A_1});
}
```

Let us have an object o with unknown rating, with attribute values $a_1, ..., a_n$. When we want to normalise value a_1, we use the following method:

```
if(existsFuzzySetFor(a_2)){
    f_{a_2}^{A_1} = getFuzzySetFor(a_2);
    return f_{a_2}^{A_1}(a_1) ;
}
else
    return  f^{A_1}(a_1);
```

ISBN: 978-989-95079-6-8

It is possible that we did not encounter value a_2 in the training phase at all or only once. In such case, we use general function $\widehat{f^{A_1}}$ for normalisation of a_1.

4.3 A method for clustering results CP-regressions

In the previous section, we mentioned a problem that when using 2CP-regression, the number of training examples for linear regression decreases rapidly when the number of attribute values in the dividing attribute is high. In the following, we present a method for overcoming this problem.

Let us have a set of fuzzy sets $\widehat{f_{a_i}^{A_1}}$ for normalisation of attribute A_1 that depends on values of attribute A_2. Each fuzzy set is of the form $\widehat{f_{a_i}^{A_1}}(x) = \alpha * x + \beta$.

Now we start with fuzzy set $\widehat{f_{a_1}^{A_1}} = \alpha * x + \beta$ and match together those fuzzy sets, that have similar shape, ie. the values of α and β. After finding similar fuzzy sets, without the loss of generality let it be $\widehat{f_{a_2}^{A_1}}, ..., \widehat{f_{a_k}^{A_1}}$. Values $a_1, ..., a_k$ are grouped together and we construct a new linear function; we will use all objects $o \in X(a_1) \cup ... \cup X(a_k)$ for its construction. This new linear function should accommodate better to testing data, because it is constructed using a larger training set. When using this set, we treat it as $\widehat{f_{a_1,...,a_k}^{A_1}}$ so it is used for every object with attribute value in $a_1, ..., a_k$.

This method of using clusters for CP-regression was not fully implemented yet, but plan to implement it and test it in near future. We expect better performance especially for smaller training sets. In this case, we can do a bottom-up clustering that stops at a specified threshold of minimal training set size.

5 Experiments

5.1 Experiment settings

Our experiment was done on a set of 200 notebooks crawled from a web shop. There are five attributes: harddisk, display, price, producer and ram. We have created an artificial user U with preferences on the scale {1,2,3,4,5} for every notebook. Preferences were generated using a set of fuzzy sets f_i and an aggregation function @.

The user preferences on price were dependent on the actual value of the producer of the notebook. As it was mentioned in Section 4.2, for producers HP, IBM, Lenovo, Toshiba and Sony the best price was set to 2200$ and for Fujitsu-Siemens, Acer, Asus and MSI the best price was set to 750$. Hence the same price would produce different degree of preference for different manufacturers. Distribution of ratings of notebooks with regard to their price is in Figure 2. We can clearly see that the ratings are distributed more or less evenly across all the domain. However, a peak at about 2500$ appears when only the triangles are taken into account and another peak at about 1000$ is for crosses. Peaks do not correspond exactly to the artificial preferences; this is because the overall rating is influenced by other attributes, too. The overall trend is clearly visible for human - with increasing price crosses tend to go down and on the opposite triangles go up.

Testing was performed with a traditional cross-validation method. Because we are dealing with user ratings, it can not be expected that the user will rate many objects. That is why we limited the size of the training set (TSS in following text) to 75 ratings at maximum. Methods were tested on the rest of the set. When the user model is constructed from the first 10 notebooks, then it is tested on the remaining 190 notebooks. Then the next 10 notebooks are taken as training set and again the model is tested on the remaining 190. For TSS = 75 the model is tested on remaining 125 notebooks. Resulting errors are averaged for every TSS so the results should be reliable.

There are five methods tested: Mean, Statistical with Linear regression, Statistical with 2CP-regression, Support Vector Machine and Multilayer perceptron. Mean is a baseline. This method always returns the average rating from the training set. Multilayer perceptron and Support Vector Machine are traditional data mining methods. Implementation from Weka [15] was used.

Statistical method as aggregation function was used with two different normalisations. The first one is traditional linear regression; the second is our proposed method 2CP-regression.

No tuning of any method was performed - these methods are supposed to work on arbitrary data automatically without the need of adjusting any parameters.

5.2 Experiment results

We studied three types of error measures. First, most commonly used, is RMSE - root mean squared error. This captures the average error across the whole testing set. Second, Weighted RMSE, adds weight to each contribution to error. The weight associated is the same as the real rating, because good objects are of more concern than objects that are not preferred. In other words, it is worse if a good notebook does not appear on the first page of results than if a bad notebook does. Finally, we used tau coefficient to study the methods. Tau coefficient is a measure of correspondence between two ordered lists. Objects from testing set ordered by real user preferences are in the first list. The second list consists of the same objects, ordered by preferences proposed by the method. In this way, we can estimate the similarity of the ordering of objects by the method and by the real user. Tau coefficient ranges from -1 to 1; 1 means absolute correlation, e.g. the same lists, and -1 means the opposite correlation, e.g. reversed lists. Tau coefficient of 0 means that there is no relation between the two lists.

In Figure 5 are represented counts of unpredicted rating for each method. At TSS = 2 there were 40 notebooks that every method failed to predict and SVM failed even for 90 notebooks. At TSS = 5, there were 5 notebooks for all methods and 25 notebooks for SVM without prediction. So SVM performed worse than the rest of methods for small TSS.

In Figure 6 are results for RMSE. 2CP regression is an overall winner, only at TSS = 40 it was outperformed by Multilayer perceptron and at TSS = 2 by SVM, but SVM failed to predict far more notebooks.

The results are about the same for Weighted RMSE, as we can see in Figure 7.

Finally, results for Tau coefficient are in Figure 8. 2CP-regression was outperformed by Mean for the three smallest TSS, otherwise it was again the best method.

In all cases, 2CP-regression was better than simple Linear regression and Statistical with 2CP-regression outperformed even SVM and Multilayer perceptron.

ISBN: 978-989-95079-6-8

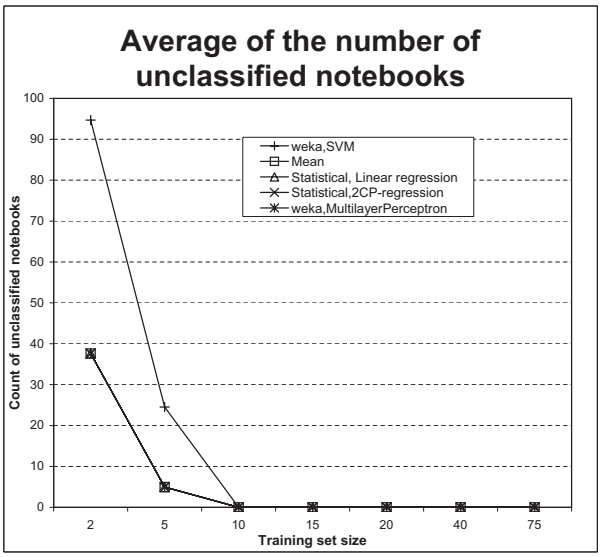


Figure 5: Count of unpredicted ratings of notebooks.

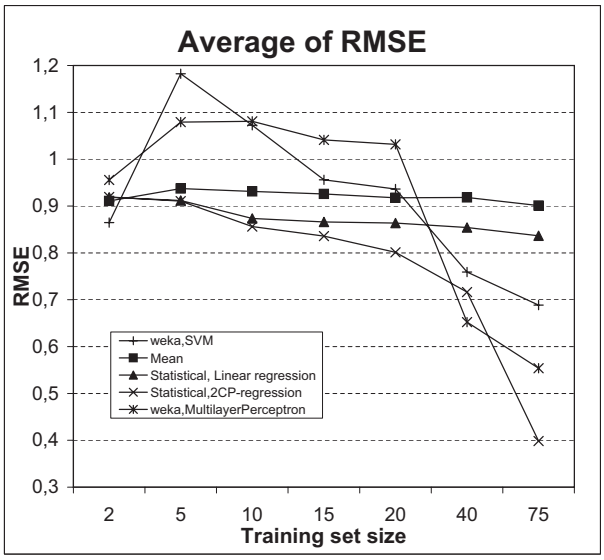


Figure 6: RMSE.

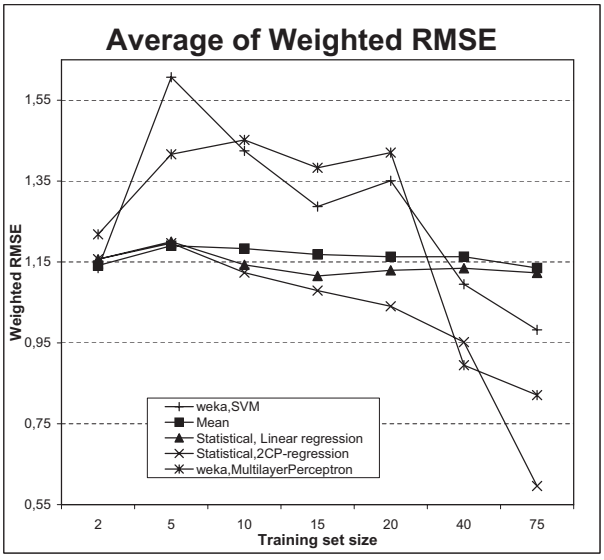


Figure 7: Weighted RMSE.

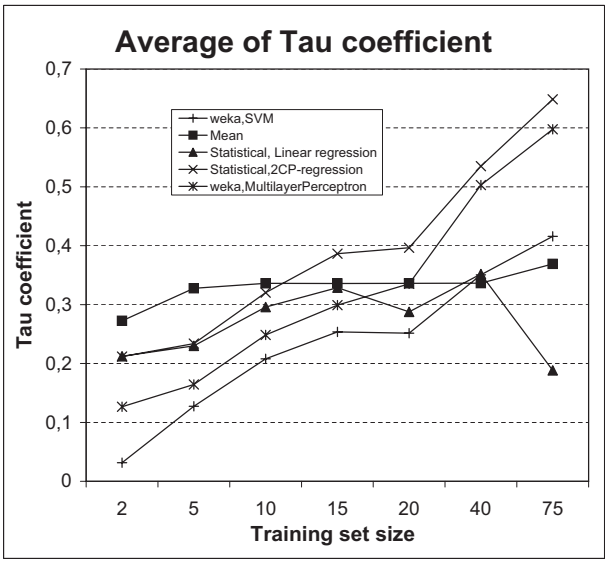


Figure 8: Tau coefficient.

6 Conclusions

We have proposed a method for finding preference dependence between two attributes, which was demonstrated on the example of a user for whom the ideal price depends on the manufacturer of the notebook. As far as we know this is the first contribution for learning ceteris paribus-network like structure from user ratings. We are aware that our model does not correspond to CP-nets completely, but the proposed part of our fuzzy model can be considered as a simple CP-net.

In our experiments, 2CP-regression performed very well, outperforming traditional data-mining techniques and also the normalisation with linear regression. Three different error measures were taken into account and in every one our proposed method was the best.

For future work, we would like to extend this approach for finding relations between more than two attributes, making a more general nCP-regression. We would also like to test it on real user preferences data, for example on NetFlix data [16].

Acknowledgment

The work on this paper was supported by Czech projects MSM 0021620838, 1ET 100300517 and GACR 201/09/H057.

References

[1] G. H. Wright. The logic of preference reconsidered. In *Theory and Decision*, volume 3, pages 140–169, 1972.

[2] Craig Boutilier, Ronen I. Brafman, Holger H. Hoos, and David Poole. Cp-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21:2004, 2004.

[3] Alan Eckhardt. Inductive models of user preferences for semantic web. In Jaroslav Pokorný, Václav Snášel, and Karel Richta, editors, *DATESO 2007*, volume 235 of *CEUR Workshop Proceedings*, pages 108–119. Matfyz Press, Praha, 2007.

[4] Alan Eckhardt, T. Horváth, D. Maruščák, R. Novotný, and Peter Vojtáš. Uncertainty issues in automating process connecting web and user. In P. C. G. da Costa, editor, *URSW '07 Uncertainty Reasoning for the Semantic Web - Volume 3*, pages 97–108. The 6th International Semantic Web Conference, 2007.

ISBN: 978-989-95079-6-8

[5] Al Mamunur Rashid, Istvan Albert, Dan Cosley, Shyong K. Lam, Sean M. McNee, Joseph A. Konstan, and John Riedl. Getting to know you: learning new user preferences in recommender systems. In *IUI '02: Proceedings of the 7th international conference on Intelligent user interfaces*, pages 127–134, New York, NY, USA, 2002. ACM.

[6] M. Ackerman, D. Billsus, S. Gaffney, S. Hettich, G. Khoo, D. Kim, R. Klefstad, A. Ludeman, J. Muramatsu, K. Omori, M. Pazzani, D. Semler, B. Starr, and P. Yap. Learning probabilistic user profiles: Applications to finding interesting web sites, notifying users of relevant changes to web pages, and locating grant opportunities. *AI Magazine*, 18:47–56, 1997.

[7] Thorsten Joachims. Optimizing search engines using clickthrough data. In *KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142, New York, NY, USA, 2002. ACM Press.

[8] Kim Cao-Van. *Supervised Ranking, from semantics to algorithms*. Ph.D. dissertation, Ghent University, 2003.

[9] Bruno Apolloni, Giacomo Zamponi, and Anna Maria Zanaboni. Learning fuzzy decision trees. *Neural Networks*, 11(5):885–895, 1998.

[10] Stefan Holland, Martin Ester, and Werner Kiessling. Preference mining: A novel approach on mining user preferences for personalized applications. In *Knowledge Discovery in Databases: PKDD 2003*, pages 204–216. Springer Berlin / Heidelberg, 2003.

[11] Sung Young Jung, Jeong-Hee Hong, and Taek-Soo Kim. A statistical model for user preference. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):834– 843, June 2005.

[12] Bernard de Baets and Jnos C. Fodor. Twenty years of fuzzy preference structures (19781997). *Decisions in Economics and Finance*, 20:45–66, 1997.

[13] R. Fagin, A. Lotem, and M. Naor. Optimal aggregation algorithms for middleware. In *In proceedings of Twentieth ACM Symposium on Principles of Database Systems, 2001 (PODS 2001)*, pages 102–113. ACM, 2001.

[14] Alan Eckhardt, T. Horváth, and Peter Vojtáš. Learning different user profile annotated rules for fuzzy preference top-k querying. In H. Prade and V. Subrahmanian, editors, *International Conference on Scalable Uncertainty Management*, volume 4772 of *Lecture Notes In Computer Science*, pages 116–130, Washington DC, USA, 2007. Springer Berlin / Heidelberg.

[15] Ian H. Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques, 2nd Edition*. Morgan Kaufmann, San Francisco, 2005.

[16] Netflix dataset. *http://www.netflixprize.com.*

ISBN: 978-989-95079-6-8

Multi-Dimensional Scaling applied to Hierarchical Rule Systems

Thomas R. Gabriel Kilian Thiel Michael R. Berthold

Nycomed Chair for Bioinformatics and Information Mining
Department of Computer and Information Science
University of Konstanz, Box 712, 78457 Konstanz, Germany
Email: {gabriel,thiel,berthold}@inf.uni-konstanz.de

Abstract *— This paper presents an approach for visualizing high-dimensional fuzzy rules arranged in a hierarchy together with the training patterns they cover. A standard multi-dimensional scaling method is used to map the rule centers of the top hierarchy level to one coherent picture. Rules of the underlying levels are projected relatively to their parent level(s). In addition to the rules, all patterns are mapped onto the two-dimensional projection in relation to the positions of the corresponding rule centers. Visualization is further extended by showing hierarchical relationships between overlapping rules of different levels, which are generated by a hierarchical rule learner. This delivers interesting insights into the rule hierarchy and offers better explorative properties. Additionally, rules can be highlighted interactively emphasizing the subsequent rules at all underlying levels together with the patterns they cover. We demonstrate that this technique allows investigation of interesting rules at different levels of granularity, which makes this approach applicable even for a large number of rules. The proposed technique is illustrated and discussed based on a number of hierarchical rule model visualizations generated from well-known benchmark data sets.*

Keywords— Multi-Dimensional Scaling, Fuzzy Rule Induction, Rule Hierarchy, Rule Visualization.

1 Introduction

Rule learning algorithms are widely used in data mining to automatically extract rules from data. In [7, 15] and [20], algorithms are described that construct hyperrectangles in feature space. The resulting set of rules encapsulates regions in feature space that contain patterns of the same class. Other approaches, which construct fuzzy rules instead of crisp rules, are presented, for example, in [1, 12, 18] and [19]. All of these approaches have in common that they tend to build very complex rule systems for large data sets originating from a complicated underlying system. In addition, high-dimensional feature spaces result in complex rules relying on many attributes further increasing the number of required rules to cover the solution space. An approach that aims to reduce the number of constraints on each rule individually is presented in [3]. The generated fuzzy rules only constrain a few of the available attributes and hence remain readable even in the case of high-dimensional spaces. The fuzzy rules generated by this method have been visualized by parallel coordinates [4, 10].

In [8], we described a method that attempts to tackle this inherent problem of interpretability of large rule models. We achieve this by constructing rule models with varying degrees of complexity. The method builds a rule hierarchy for a given data set. The rules are arranged in a hierarchy of different levels of precision. Lower levels of the model describe regions in input space with low evidence in the given data, whereas rules at higher levels describe more strongly supported concepts of the underlying data. The method is based on the fuzzy rule learning algorithm described in [3, 9], which builds a single layer of rules autonomously. We use the resulting rule system recursively to determine rules of low relevance, which are then used as a filter for the next training phase. The result is a hierarchy of rule systems with the desired properties of simplicity and interpretability on each level. Experimental results demonstrated that fuzzy models at higher hierarchical levels show a dramatic decrease in number of rules while still achieving a better or similar generalization performance than the fuzzy rule system generated by the original, non-hierarchical algorithm.

In this paper we present an approach that enables the visualization of hierarchically structured rules and data in one coherent plot. A standard multi-dimensional scaling (MDS) method is applied to project rules of a certain level from the original, high-dimensional space, into a lower, usually two-dimensional space. Rules of the underlying levels are projected relatively to their parent level(s), likewise all data points are projected in relation to the positions of the corresponding rule centers. Furthermore, the visualization shows hierarchical relationships between overlapping rules of different levels as generated by a hierarchical rule learner. Due to the hierarchical nature of the induced rule system, interactive exploration becomes possible across the entire rule model and provides interesting insights into the underlying concept.

The paper is organized as follows: In the next section we briefly describe the used hierarchical rule learning method, followed by a short introduction to multi-dimensional scaling methods, and how data and hierarchical rules can be visualized by applying an extended multi-dimensional scaling method. In the following section, we describe how rule hierarchies, together with their original data, can be explored within this visualization. To illustrate the proposed method, hierarchical rule systems are generated based on the well-known iris data and on the vehicle silhouettes data.

2 Learning Hierarchical Rule Systems

The rule induction algorithm used here is based on a method described in [3], which is based on an iterative algorithm. During each learning epoch, i. e. presentation of all training patterns, new fuzzy rules are induced when necessary and existing ones are adjusted whenever a conflict occurs. For each pattern three main steps are executed. First, if a new training pattern lies inside the support-region of an existing fuzzy rule of the correct class, its core-region is extended in order to cover the new pattern. Second, if the new pattern is not yet

ISBN: 978-989-95079-6-8

covered, a new fuzzy rule of the correct class is introduced. The new example is assigned to its core, whereas the support-region is initialized by infinity, that is, the new fuzzy rule covers the entire domain. Finally, if a new pattern is incorrectly covered by an existing fuzzy rule, the fuzzy rules' support-region is reduced to avoid the conflict. This heuristic for conflict avoidance aims to minimize the loss in volume. In [9], three different heuristics are compared to determine the loss in volume. As discussed in [3], the algorithm terminates after only a few iterations over the set of example patterns.

The final set of fuzzy rules can be described as an n-dimensional IF clause as antecedence and one assigned $class$ in the rule's conclusion:

$$\begin{array}{lllll}
\mathcal{R}_1^1 & \texttt{IF} & x_1 \texttt{ IS } \mu_{1,1}^1 & \wedge\cdots\wedge & x_n \texttt{ IS } \mu_{n,1}^1 \quad \texttt{THEN class } 1 \\
 & & \vdots & \vdots & \vdots \\
\mathcal{R}_{r_1}^1 & \texttt{IF} & x_1 \texttt{ IS } \mu_{1,r_1}^1 & \wedge\cdots\wedge & x_n \texttt{ IS } \mu_{n,r_1}^1 \quad \texttt{THEN class } 1 \\
 & & \vdots & \vdots & \vdots \\
\mathcal{R}_j^k & \texttt{IF} & x_1 \texttt{ IS } \mu_{1,j}^k & \wedge\cdots\wedge & x_n \texttt{ IS } \mu_{n,j}^k \quad \texttt{THEN class } k \\
 & & \vdots & \vdots & \vdots \\
\mathcal{R}_{r_c}^c & \texttt{IF} & x_1 \texttt{ IS } \mu_{1,r_c}^c & \wedge\cdots\wedge & x_n \texttt{ IS } \mu_{n,r_c}^c \quad \texttt{THEN class } c
\end{array}$$

where $\mathcal{R}_j^k$ represents rule j for class k. The rule base contains rules for c classes and r_k indicates the number of rules for class k ($1 \leq j \leq r_k$ and $1 \leq k \leq c$). The fuzzy sets $\mu_{i,j}^k : \mathbb{R} \mapsto [0,1]$ are defined for every feature i ($1 \leq i \leq n$), but in cases of unconstrained features the membership degree constantly remains 1. The overall degree of fulfillment of a specific rule for an input pattern $\vec{x} = (x_1, \ldots, x_n)$ can be computed using the fuzzy set operator for conjunction, called T-norm ($\top$):

$$\mu_j^k(\vec{x}) = \top_{i=1,\cdots,n} \left\{ \mu_{i,j}^k(x_i) \right\}.$$

The combined degree of membership for all rules of class k can be calculated using the fuzzy set operator for disjunction, called S-norm ($\perp$):

$$\mu^k(\vec{x}) = \perp_{j=1,\cdots,r_k} \left\{ \mu_j^k(\vec{x}) \right\}.$$

From these membership values the predicted class k_{best} for an input pattern $\vec{x}$ is derived as:

$$k_{\text{best}}(\vec{x}) = \arg\max_{k=1,\ldots,c} \left\{ \mu^k(\vec{x}) \right\}.$$

The algorithm uses trapezoidal membership functions, which can be described by four parameters $\langle a_i, b_i, c_i, d_i \rangle$, where a_i and d_i define the fuzzy rule's support-, and b_i and c_i its core-region for each attribute i of the input dimension. The core-region is defined as a rectangular area with an activation of 1, whereby the support-region decreases linearly to its boundaries with a degree of fulfillment of 0.

The resulting set of fuzzy rules can then be used to classify new patterns by computing the overall degree of membership for each class. The accumulated membership degrees over all input dimensions and across multiple rules are calculated by a fuzzy norm. For the purpose of this paper, we concentrate on the rules' core-regions only, that is, the part of each rule where the degree of membership is equal to 1 – resulting in crisp rules – is considered.

An extension of this rule learning algorithm is proposed in [8], which allows the generation of a hierarchy of rules. Therefore, the classical fuzzy rule algorithm is used to determine rules of low relevance recursively, which are then used as a filter for the next training phase. Training examples that cause the creation of small, less important rules are therefore excluded from the training phase of the next layer, resulting in a more general rule system, ignoring the withheld, small details in the training data. An example of a two-stage rule induction process is illustrated in Fig 1.

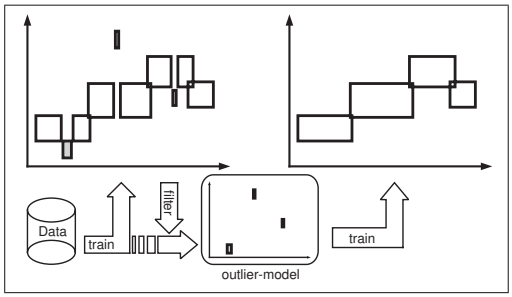


Figure 1: Two-stage outlier filtering as described in [2].

This process can be repeated to generate the desired hierarchy of rule systems with an increasing generality towards higher levels. The rule layers are arranged in a hierarchy of different levels of precision. Lower levels of the model describe regions in input space with low evidence in the training data, whereas rules at higher levels describe more strongly supported aspects of the underlying data.

3 Multi-Dimensional Scaling

To visualize objects of high-dimensional feature spaces, multi-dimensional scaling methods [6] can be applied to map them onto lower-dimensional spaces with usually two or three dimensions. In order to avoid a loss of proximity information, MDS methods try to preserve the pairwise distances between objects during the mapping process by minimizing an appropriate error function. The reduction of dimensions allows the visualization of high-dimensional objects in a lower-dimensional space, using traditional methods such as scatter plots or scatter matrices.

One technique to apply this kind of mapping by minimizing a particular error function is the well known Sammon algorithm [17]. For each object $\vec{X}_i$ in high-dimensional space $\mathbb{R}^H$, a spatial representation $\vec{x_i} \in \mathbb{R}^L$ in low-dimensional space (usually $L = 2$) has to be computed ($1 \leq i \leq N$ with N as the number of objects).

To approximate the distance information, the position of each object in the low-dimensional space has to be adjusted such that the pairwise distances between each two objects $(\vec{x_i}, \vec{x_j})$ in the low-dimensional space, $d_{ij} = d(\vec{x_i}, \vec{x_j})$, approximate the distances of the two corresponding objects $(\vec{X}_i, \vec{X}_j)$ in the high-dimensional space, $D_{ij} = D\left(\vec{X}_i, \vec{X}_j\right)$:

$$\forall_{i \neq j} : D_{ij} \approx d_{ij}, 1 \leq i, j \leq N.$$

Usually the Euclidean metric is used to measure these distances in the low-dimensional space, and often in the high-

ISBN: 978-989-95079-6-8

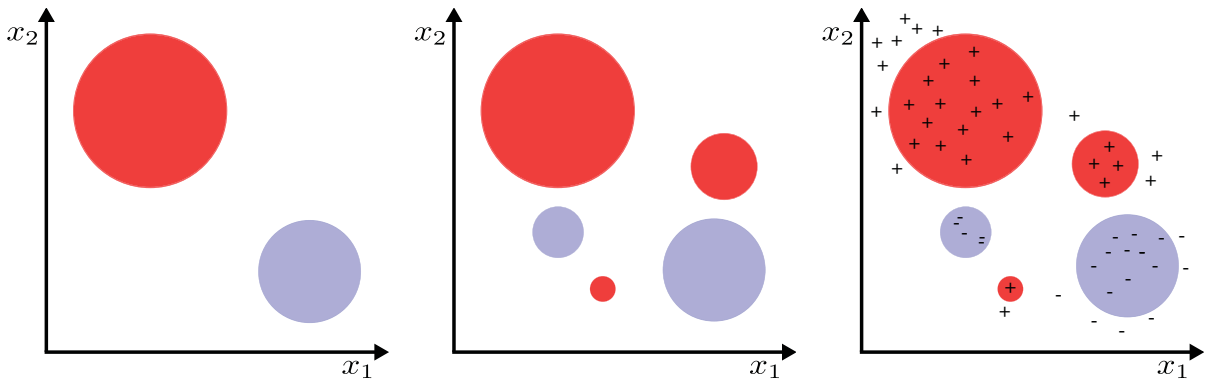


Figure 2: Shows a two-level rule hierarchy. (a) Two top-level rules assigned to two different classes, + (red) and − (blue), (b) three additional rules on the next level, (c) all rules together with their underlying data points.

dimensional space as well:

$$d_{ij}^2 = \sum_{l=1}^{L} (x_{i,l} - x_{j,l})^2.$$

Approximation of the pairwise distances in high- and low-dimensional space is formulated by Sammon as a minimization problem of a cost function, which aggregates the weighted squared differences of the distances in high- and low-dimensional space:

$$E = \sum_{i=1}^{N} \sum_{j>i}^{N} \omega_{ij} (d_{ij} - D_{ij})^2.$$

where the factors ω_{ij} are introduced to weight the distances individually and normalize the stress function E in order to be independent from the absolute values D_{ij}. The steepest gradient method is applied to minimize iteratively a cost function E for each object $\vec{x_i}$ at each step. Usually several iterations are needed by the algorithm to converge to a local cost minimum.

3.1 *Projective Multi-Dimensional Scaling*

As mentioned above, the original Sammon algorithm tries to preserve the pairwise distances of all objects. This means that the position of each object in the low-dimensional space is adjusted iteratively according to the position of the other objects in the high- and low-dimensional space. Subsequently this leads to a change of the positions of all objects in the low-dimensional space. With our projective multi-dimensional scaling approach, the positions of the objects in the low-dimensional space are adjusted according to a set of fixed projection objects.

First a set of objects in the high-dimensional space is selected. For each of these objects $\vec{X_p} \in \mathbb{R}^H$ a spatial representation $\vec{x_p} \in \mathbb{R}^L$ ($1 \leq p \leq M$ with $M \ll N$ as the number of fixed projection objects $\vec{x_p}$) has to be found by means of the standard MDS algorithm described above. Once these objects have been mapped onto the low-dimensional space, they are used as fixed objects and are not changed anymore. Furthermore, each of the other objects $\vec{X_i} \in \mathbb{R}^H$ are mapped onto the low-dimensional space according to the fixed projection objects. Therefore, the distance between a regular object $\vec{x_i} \in \mathbb{R}^L$ and a fixed projection object $\vec{x_p}$ in the low-dimensional space $d_{ip} = d(\vec{x_i}, \vec{x_p})$ has to be approximated to the distance between the two corresponding objects $(\vec{X_i}, \vec{X_p})$ in the high-dimensional space $D_{ip} = D\left(\vec{X_i}, \vec{X_p}\right)$, which translates to:

$$\forall_{i \neq p} : D_{ip} \approx d_{ip}, 1 \leq i \leq N, 1 \leq p \leq M.$$

Again, a cost function E_P is defined that aggregates the weighted squared differences of the distances in the high- and low-dimensional space:

$$E_P = \sum_{i=1}^{N} \sum_{p=1}^{M} \omega_{ip} (d_{ip} - D_{ip})^2.$$

The projective MDS method is useful if a set of objects has to be mapped according to an already existing set of mapped objects without modifying the mapping of the latter.

In the case of high-dimensional hierarchical data, projective MDS can be used to determine a spatial representation in several iterations. First the standard MDS method is applied to the top-level data. Once this data is mapped, projective MDS can be applied stepwise to the data of the lower levels using the data of higher levels as fixed projection objects. The low-dimensional representation of the already mapped data does not change anymore when applying projective MDS to the remaining levels.

3.2 *Hierarchical Rules in Multi-Dimensional Scaling*

It is more complex to visualize rule models on lower dimensions than simple data points. The main challenge is the mapping of the rules' antecedences, which are usually hyperrectangles in the original high-dimensional feature space. In order to visualize high-dimensional fuzzy rules by means of MDS, the center points of a rules' core-regions are mapped onto a low-dimensional space, as described in [11, 16]. For hierarchical fuzzy rule systems, the center points of the top-level rules are mapped onto the low-dimensional space using standard MDS. Subsequently, the rules of the underlying levels are mapped level by level via the projective MDS with the parent rule centers as fixed projection points. Finally, the data points are mapped according to all projected rules, whereas the neighborhood of rules and data points is approximated by the proposed MDS. The overlapping relation between rules of different levels may get lost in the visualization, but is again

ISBN: 978-989-95079-6-8

taken into account when rule systems are visually explored by interactive highlighting, see Sec. 4.2 for an example.

In addition to the rules' center points, the spread of a fuzzy rule has to be mapped onto the low-dimensional space in order to visualize the rules' sizes, possible overlaps, and the coverage according to the data points the rules are based on. In [11], we visualize the overlap of rules of flat rule systems. Dealing with hierarchical fuzzy rules, we only focus on the visualization of the spread and the number of covered data points of a rule as well as the rules' level and not on overlap with other rules.

The spread is visualized by a sphere around the mapped rule center with a radius $r_i = \rho(\lambda\omega_i + (1-\lambda)\sigma_i^2)$ according to the variance σ_i^2 and the number of the points ω_i covered by a rule $\mathcal{R}_i$, with λ as weighting coefficient $\lambda \in [0,1]$ and $\rho(\cdot)$ as scaling function. The number of vertices $v = 2^H$ of a hyperplane grows exponentially with the number of dimensions H whereas all vertices have the same distance to the center point. All vertices of a high-dimensional hyperplane are mapped by MDS, representing vertices of a low-dimensional polygon; all with the same distance to the polygons center point. With increasing dimensionality, the polygon spanned by the projected vertices converges to a sphere.

In Fig. 2 (a), two top-level rules are shown, covering data points of two different classes. The red rule covers data points of class $+$, while the blue covers data points of class $-$. Fig. 2 (b) illustrates rules of the primary and the secondary level, which consists of three smaller rules, two rules covering data points of class $+$ and two of class $-$. Fig. 2 (c) shows rules of all levels as well as the data points. It can be seen that the spheres representing the rules do not necessarily cover all the data points in the low-dimensional space as they do in the high-dimensional space. This is because the radius is based on the number of covered data instances of a rule as well as the variance of the covered data points in the high-dimensional space, but is not based on the position of the data points in the low-dimensional space.

4 Visualization Hierarchical Rules and Data

The following section illustrates the proposed approach on the well-known iris dataset before looking at a larger dataset, the vehicle silhouettes data. We compare the classical, non-hierarchical rule model to the hierarchical structured rules by visualizing all the rules together with their originating data instances in one coherent picture.

4.1 Iris Dataset

The first example shows a small two-level hierarchy trained on the iris data, which consists of 150 four-dimensional patterns assigned to three classes. Fig. 3 (a) shows the top-rule level containing three rules – one for each class, (b) the top and bottom-rule levels with additional 11 rules (some cover only a single or a few instances and therefore appear as small points), and (c) all data points and all rules together. The top-level rules are projected into the low-dimensional space based on each center point. The second level is then projected based on the first rule level. In the last step, data points are projected according to both hierarchical levels as shown in Fig. 3. The bottom picture shows rules and data instances of all three classes, which are almost perfectly separated from each other by three large rules. Smaller rules in-between are needed to cover artifacts and details in the data.

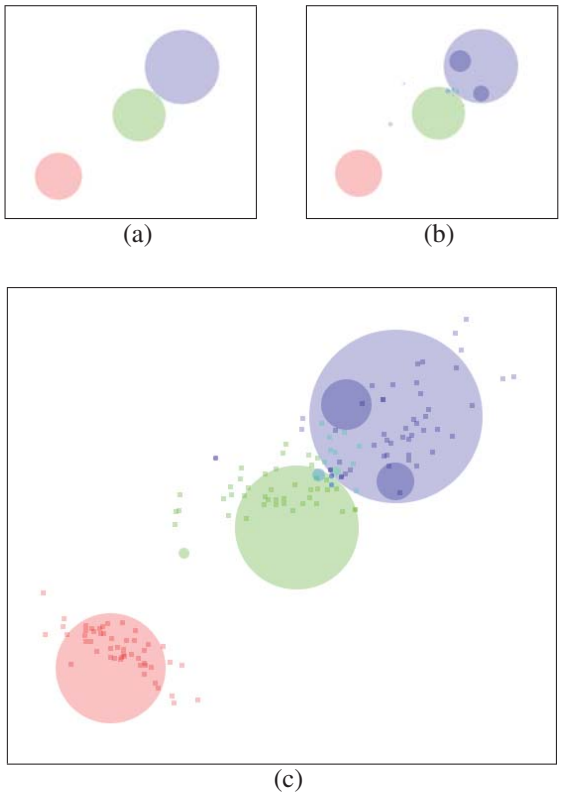


Figure 3: Two-level rule hierarchy trained on iris data: (a) three top-most rules, (b) additional 11 rules from the bottom-level, and all data points in (c).

4.2 Vehicle Silhouettes Dataset

The vehicle silhouettes dataset consists of 846 samples belonging to four car classes – Opel, Saab, bus, and van – represented in an 18-dimensional feature space. To demonstrate the usefulness of our proposed method, we trained a three-level fuzzy rule hierarchy on this data from the European StatLog–Project [14]. In a first experiment, a classical rule model with 222 rules without hierarchy information is trained on the vehicle data. As can be seen in Fig. 4, exploration is hardly pos-

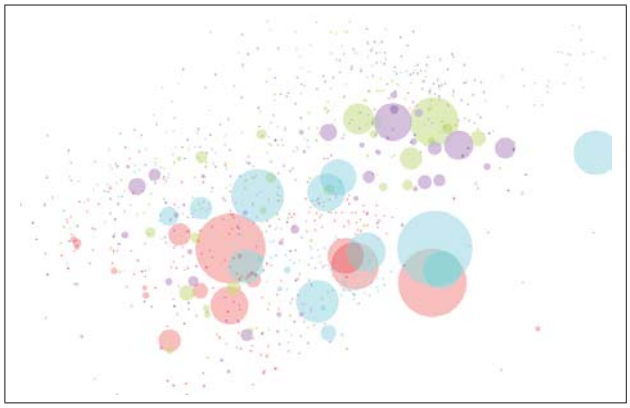


Figure 4: Classical, flat rule model without hierarchy information trained on the vehicle silhouettes dataset showing 222 rule for four classes.

ISBN: 978-989-95079-6-8

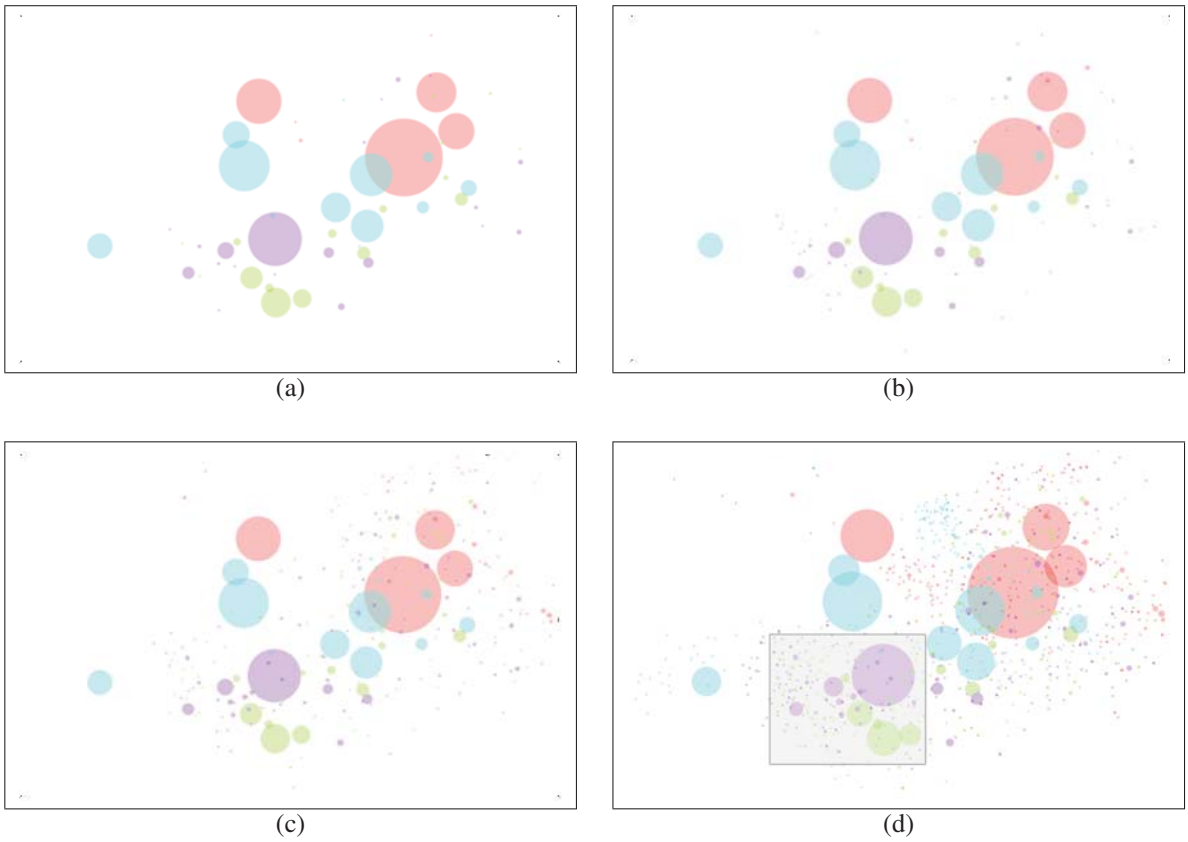


Figure 5: Stepwise rule hierarchy projection on vehicle data starting from top- to bottom-level (a)-(c), and (d) rules of all levels and data points.

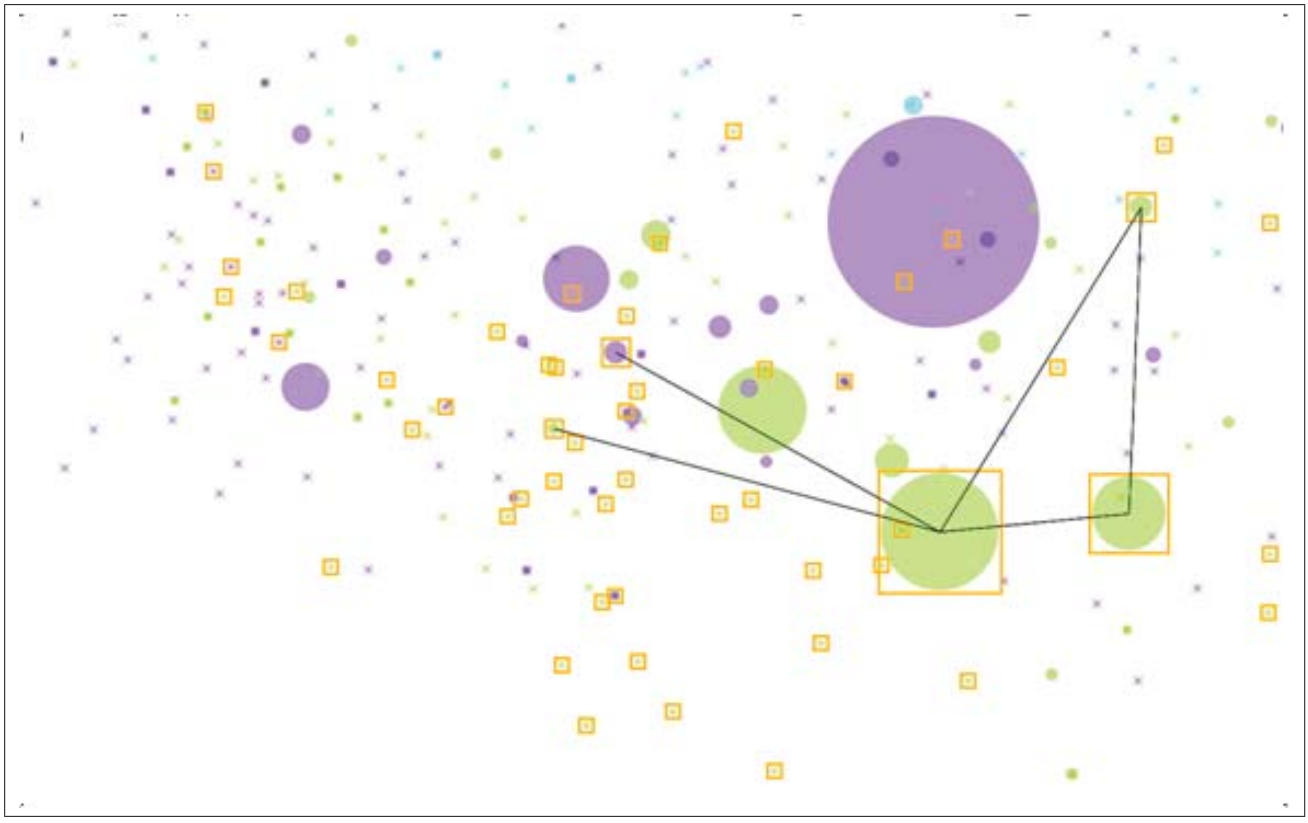


Figure 6: Exploration by zooming and highlighting (orange square) interesting rules and data points in the vehicle hierarchy.

ISBN: 978-989-95079-6-8

sible since overlapping information is only available between rules of the same single level [1].

The hierarchical rule induction algorithm is applied to generate a hierarchy of rule models in a second test. The three levels of the fuzzy rule hierarchy contain 55 rules in the top, 40 in the middle, and 174 at the bottom-most level. Fig. 5 shows each level together with its parent levels starting from (a) to (c), and (d) rules of all levels and data points. All levels are subsequently projected into the two-dimensional space always with respect to their parent level(s), as well as all data points that are projected relatively to all rule levels. Fig. 6 shows the surrounded area from Fig. 5 (d) enlarged. Selecting one rule highlights all overlapping rules in the levels above along with the data points they cover. All highlighted rules (orange square) that cover the same set of patterns are connected by a line to visually identify overlapping rules between different levels of the hierarchy. The figure shows five connected rules where a violet rule from a lower level completely overlaps with a large, green rule of a higher level. This is typically an indication for outliers and artifacts in the data expressed by smaller, more specific rules at lower levels generated by the rule learning algorithm at first place; whereas rules generated at higher levels of the hierarchy explain more general aspects of the data and are usually covered by larger rules. This visual line-up of rules of different levels allows further exploration by highlighting interesting data instances.

5 Conclusions

Hierarchically structured rules induced by a classical rule learning algorithm lead to a well-defined hierarchy of rules where levels further up explain more general aspects and rule models at lower levels concentrate on artifacts or outliers of the underlying concept. Combining this information with a mapping mechanism to visualize both this type of hierarchy and the data points enables interactive exploration across rule levels by focusing on overlapping regions. On the other hand, it highlights data points covered throughout the rule hierarchy. We demonstrated the explorative power of the proposed projective multi-dimensional scaling method based on the vehicle silhouettes dataset, delivering interesting insights into the underlying concept. Our approach is well suited for projecting other hierarchical structured data and rules driving interactive explorative data analysis.

References

[1] S. Abe, and M.S. Lan. A method for fuzzy rules extraction directly from numerical data and its application to pattern classification. *IEEE Transactions on Fuzzy Systems*. 3:18–28, 1995.

[2] M.R. Berthold. Learning fuzzy models and potential outliers. *In Computational Intelligence in Data Mining.* Springer, Berlin, 111–126, 2000.

[3] M.R. Berthold. Mixed Fuzzy Rule Formation. *International Journal of Approximate Reasoning (IJAR)*. Elsevier, 32:67–84, 2003.

[4] M.R. Berthold, and L.O. Hall. Visualizing Fuzzy Points in Parallel Coordinates. *IEEE Transactions on Fuzzy Systems*. IEEE Press, 11(3):369–374, 2003.

[5] M.R. Berthold, and R. Holve. Visualizing high dimensional fuzzy rules. *Proceedings of NAFIPS*. IEEE Press, 64–68, 2000.

[6] T.F. Cox, and M.A. Cox. Multidimensional Scaling. *Monographs on Statistics and Applied Probability*. Chapman and Hall, 1994.

[7] W. Duch, R. Setiono, and J.M. Zurada. Computational Intelligence Methods for Rule-based Data Understanding. *Proceedings of the IEEE*. IEEE Press, 92(5):771–805, 2004.

[8] Th.R. Gabriel, and M.R. Berthold. Constructing Hierarchical Rule Systems. In: M.R. Berthold, H.-J. Lenz, E. Bradley, R. Kruse, and C. Borgelt (Eds.): *Proc.* 5^{th} *International Symposium on Intelligent Data Analysis (IDA)*. Springer, Berlin, 76–87, 2003.

[9] Th.R. Gabriel, and M.R. Berthold. Influence of fuzzy norms and other heuristics on "Mixed Fuzzy Rule Formation". *International Journal of Approximate Reasoning (IJAR)*. Elsevier, 35:195–202, 2004.

[10] Th.R. Gabriel, A.S. Pintilie, and M.R. Berthold. Exploring Hierarchical Rule Systems in Parallel Coordinates. In: Famili et al. (Eds.): *Proc.* 6^{th} *International Symposium on Intelligent Data Analysis (IDA)*. Springer, Madrid, 97–108, 2005.

[11] Th.R. Gabriel, K. Thiel, and M.R. Berthold. Rule Visualization based on Multi-Dimensional Scaling. *IEEE International Conference on Fuzzy Systems*. IEEE Press, Vancouver, 66–71, 2006.

[12] C.M. Higgins, R.M. Goodman. Learning fuzzy rule-based neural networks for control. *Advances in Neural Information Processing Systems*. California, Morgan Kaufmann, 5:350–357, 1993.

[13] J. Meulman. A distance approach to nonlinear multivariate analysis. DSWO Press, Leiden, The Netherlands, 1986.

[14] D. Michie, D.J. Spiegelhalter, and C.C. Taylor, editors. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood Limited, 1994.

[15] S. Salzberg. A nearest hyperrectangle learning method. *Machine Learning*. 6:251–276, 1991.

[16] F. Rehm, F. Klawonn, and R. Kruse. Visualization of Fuzzy Classifiers. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*. World Scientic Publishing Company, 15(5):615–624, 2007.

[17] J.W. Sammon, Jr. A non-linear mapping for data structure analysis. *IEEE Transactions on Computers*. C18(5):401–409, May 1969.

[18] P.K. Simpson. Fuzzy min-max neural networks – part 1: Classification. *IEEE Transactions on Neural Networks*. 3:776–786, 1992.

[19] L.X. Wang, and J.M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*. 22:1313–1427, 1992.

[20] D. Wettschereck. A hybrid nearest-neighbour and nearest-hyperrectangle learning algorithm. *Proceedings of the European Conference on Machine Learning*. 323–335, 1994.

[1] In some rare cases, neighboring data points of different classes generate rules with overlapping core-regions.

ISBN: 978-989-95079-6-8

Rule Weight Optimization and Feature Selection in Fuzzy Systems with Sparsity Constraints *

Edwin Lughofer[1] Stefan Kindermann[2]

1. Department of Knowledge-Based Mathematical Systems, Johannes Kepler University
Altenbergerstrasse 69, A-4040 Linz, Austria
2. Department of RICAM, Johannes Kepler University
Altenbergerstrasse 69, A-4040 Linz, Austria
Email: edwin.lughofer@jku.at, stefan.kindermann@indmath.uni-linz.ac.at

Abstract *— In this paper, we are dealing with a novel data-driven learning method (*SparseFIS*) for Takagi-Sugeno fuzzy systems, extended by including rule weights. Our learning method consists of three phases: the first phase conducts a clustering process in the input/output feature space with iterative vector quantization. Hereby, the number of clusters = rules is pre-defined and denotes a kind of upper bound on a reasonable granularity. The second phase optimize the rule weights in the fuzzy systems with respect to least squares error measure by applying a sparsity-constrained steepest descent optimization procedure. This is done in a coherent optimization procedure together with elicitation of consequent parameters. Depending on the sparsity threshold, more or less rules weights can be forced towards 0, switching off some rules. In this sense, a rule selection is achieved. The third phase estimates the linear consequent parameters by a regularized sparsity constrained optimization procedure for each rule separately (local learning approach). Sparsity constraints are applied here in order to force linear parameters to be 0, triggering a feature selection mechanism per rule. In some cases, this may also yield a global feature selection, whenever the linear parameters of some features in each rule are near 0. The method is evaluated based on high-dimensional data from industrial processes and based on benchmark data sets from the internet and compared to well-known batch training methods in terms of accuracy and complexity of the fuzzy systems.*

Keywords— Takagi-Sugeno fuzzy systems, iterative vector quantization, rule weight optimization, sparsity constraints, rule selection, embedded feature selection

1 Introduction

1.1 Motivation and State of the Art

Nowadays data-driven learning of fuzzy systems enjoy a great attraction in many industrial applications, as, opposed to expert-based fuzzy systems, they can be fully automatically generated from process data such as measurements, images (features) or signal streams and hence save a lot of time and money for the development of system models. Their big advantage over other methods such as neural networks or genetic programming are that they are universal approximators [22] (i.e. being able to model a given problem to any degree of accuracy) and at the same time they still allow some insights in form of linguistically [5] and visually interpretable rules [19]. Due to these benefits, data-driven fuzzy systems are applied in many application areas such as fault detection, image classification or identification models in control systems and hence serve as important components in these systems.

During the recent 15 to 20 years a lot of data-driven fuzzy systems approaches were developed: some methods such as popular *genfis2* [23] (as implemented in MATLAB's fuzzy logic toolbox) and its successor *genfis3*, *FMCLUST* [2], exploit various clustering techniques (e.g. Gustafsson-Kessel [14] or variations of mountain and subtractive clustering [6]) for extraction clusters out of the data which are directly associated with the rules. A big family of design methods use genetic algorithms [9] for 1.) eliciting the optimal number of rules by achieving a reasonable tradeoff between accuracy and model complexity and 2.) for (fine-)tuning the parameters: these are also called genetic fuzzy systems [10] and enjoy a great attraction in the fuzzy community since more than 10 years [8].

Other methods use numerical optimization procedures for linear (consequent) and non-linear (antecedent) parameters in the fuzzy systems for minimizing the least squares errors (and variants) between estimated and predicted target values: a well-known method is the *ANFIS* approach [15] exploiting a neuro-fuzzy system architecture similar to a Takagi-Sugeno type system; another method demonstrated in [3] uses Levenberg-Marquardt algorithm for optimizing non-linear antecedent parameters in triangular and Gaussian fuzzy sets. *RENO* [4] exploits a generalized Gauss-Newton like method as a second-order approximation method with (regularized) smoothing constraints.

None of these methods perform an implicit feature selection approach, neither globally nor locally, which may be important in case of high-dimensional spaces for decreasing curse of dimensionality effect. Moreover, most of these methods do not have an integrated approach for a deterministic optimization of rule structure (number of rules) and consequent parts in a coherent procedure at the same time. Indeed, genetic fuzzy systems provide this aspect, however are non-deterministic and often lack efficiency in terms of computational performance.

*This work was supported by the Upper Austrian Technology and Research Promotion. This publication reflects only the author's view.

ISBN: 978-989-95079-6-8

1.2 Our Approach

In this paper, we present a novel technique for data-driven learning of Takagi-Sugeno fuzzy systems called *SparseFIS*, which is short for *Sparse Fuzzy Inference Systems*. This approach is characterized by mainly three issues (more details in Section 2):

1. it applies a top-down approach where the upper bound on the number of rules are allowed to be predefined (e.g. by an expert) and the non-linear parameters in the fuzzy sets of the antecedent parts are pre-estimated by an iterative vector quantization approach (Section 3).

2. it performs a rule selection strategy by optimizing rule weights together with the linear consequent parameters within one non-linear optimization procedure (Section 4). The application of a threshold operator therein (for forcing as much as possible rule weights towards zero) allows optimality in the least squares sense and readability at the same time.

3. it includes an embedded local (=for each rule separately) feature selection technique by applying a semi-smooth Newton method for a sparsity regularized estimation of the linear consequent parameters (Section 5). This may decrease the curse of dimensionality effect and improves interpretability of the achieved fuzzy systems.

Combining these three concepts together will guide us to the *SparseFIS* algorithm in Section 6, which is able to generate accurate and sparse fuzzy systems at the same time. Section 7 compares the novel method in terms of predictive accuracy and complexity with other well-known (and in MATLAB achievable) data-driven fuzzy system methods based on four data sets from the UCI respository and on noisy engine test bench data.

2 Problem Statement

Let us first define the Takagi-Sugeno fuzzy model with rule weights and by applying product t-norm with Gaussian fuzzy sets (in case of C rules):

$$y_{fuz} = \sum_{i=1}^{C} l_i(x)\Psi_i(x,\rho), \quad l_i(x) = \sum_{k=1}^{p} w_{i,k}x_k \qquad (1)$$

with

$$\mu_i(x) = \Pi_{j=1}^{p}\mu_{i,j}(x_j), \quad \mu_{i,j}(x_j) = e^{-\frac{(x_j - c_{i,j})^2}{2\sigma_{i,j}^2}}$$

and

$$\Psi_i(x,\rho) = \frac{\rho_i\mu_i(x)}{\sum_{j=1}^{C}\rho_j\mu_j(x)} \qquad (2)$$

Here $\rho = (\rho_1, \ldots \rho_C)$ are positive weights, which more or less steer the number of rules (as weights smaller than a low number $\epsilon > 0$ can be seen as not useful and hence omitted). For optimization purposes we use the least squares error measure, which, when applying TS fuzzy systems as above and in dependency of the unknown parameters, can be written as

$$J(w,c,\sigma,\rho) = \sum_{k=1}^{N}(y_k - \sum_{i=1}^{C} l_i(\vec{x}_k)\Psi_i(\vec{x}_k))^2 \qquad (3)$$

The training goals (=steps in the algorithm) in this paper can be summarized as follows:

1. Estimating the non-linear antecedent parameters in a preliminary phase due to a pre-defined upper bound on the number of rules.

2. Minimizing the least squares error by keeping the model complexity (in terms of the number of rules) as low as possible (rule selection from the initial number). This is done in a coherent procedure with optimization of the consequent parameter synchronously.

3. Minimizing the least squares error by keeping the number of significant linear consequent parameters in each rule as low as possible. This has the effect of a kind of local feature selection (which can turn into a global one if certain parameters belonging to the same input variable are not significant in any rule) and hence serves as an additional interpretability aspect as well as regularization property.

The last two steps guarantee a reasonable tradeoff between accuracy and complexity resp. readability of the achieved models, as we optimize the rules resp. consequent parameters according to an optimization criterion by applying synchronously constraints on the number of significant rules resp. significant features in the consequents. This also means that this technique does not need any computational intensive post-processing techniques (such as e.g. [20]) for reducing complexity. In the next three sections we describe how to achieve these three goals.

3 Estimating Antecedent Parts of the Rules

The first step is to yield a proper estimation of the non-linear premise parameters c and σ in the Gaussian fuzzy sets of the rules antecedent parts. For doing so, we apply a clustering approach, which is called iterative vector quantization as iterating over the whole data set a multiple times, and associate each cluster to one rule. The fuzzy sets of the rules antecedent parts are obtained by projection onto the one-dimensional axis: the centers of the fuzzy sets are the center components of a cluster in each dimension, the widths are achieved by calculating the variance in each dimension. In principle, any clustering method requiring a fixed number of cluster as input and delivering cluster prototypes = centers of fuzzy sets and rules as well as the range of influence of the clusters in each dimension can be applied here. In fact, the projection concept is also carried out in many other approaches such as genfis2 [6] or FMCLUST [2] and hence not new. The difference in this paper is that we are applying vector quantization [12] where the

ISBN: 978-989-95079-6-8

a-forementioned methods apply other clustering algorithms (such as subtractive clustering and Gustafson-Kessel).

First, we estimate an initial number of rules (=clusters) and send these to the iterative vector quantization algorithm for fine tuning the non-linear parameters. The number of initial rules are

- Either pre-defined according to user knowledge, interpretability or computational aspects (usually $<$ 100 rules).
- Or automatically extracted from the data by using an evolving clustering approach (*eVQ* [16]) which starts with 0 clusters and evolve new ones according to the distribution and characteristics of new data.

Then, we apply the iterative vector quantization (VQ) algorithm (please refer to [12]) plus estimate the ranges of influence of each cluster in each dimension after the end of VQ.

4 Rule Selection by Optimizing the Rule Weights

Once having the centers and widths of the predefined number (upper bound) of rules estimated, it is a big challenge to reduce the rules in order to enhance interpretability by synchronously not loosing significant accuracy. In this sense, we propose to optimize the rule weights as defined in (2) within a iterative optimization process using sparsity constraints. In a sense, we try to find a minimal least squares error solution subject to the constraint that a significant number of rule weights should be zero. In an optimization formulation a goal would be to minimize the least squares error (3) subject to the constraint that the number of weights ρ_i should be bounded:

$$\min J(w,c,\sigma,\rho) \quad \text{such that} \#\{\rho_i \ / \tag{4}$$

where $\#$ denotes the cardinality of a set and K is an appropriate constant. A minimizer of such a problem is expected to achieve a small error together with a small number of nonzero rule weights. However, there is an intrinsic difficulty with this formulation, as this is a combinatorial optimization procedure (note that we only bound the number of nonzero elements, there is no restriction, where the nonzero elements are). A method that solves such a problem would have to test all subsets of $\{1,\dots N\}$ as candidates for the nonzero entry index of ρ and hence requires exponential complexity in N. However, recently the surprising fact came up, that an almost optimal solution to (4) can be computed by l^1-minimization [7]: That is, instead of (4) the problem

$$\min J(w,c,\sigma,\rho) \quad \text{such that} \sum_{m=1}^{N} |\rho_i| \leq K, \tag{5}$$

is solved. This is a much more accessible approach, because the constraints are convex and there are efficient methods to solve it.

Our algorithm is a nonlinear version of the well-known iteration proposed by [11]. The iteration is basically a projected gradient descent algorithms. The gradient of least squares functional with respect to the jth rule weight ρ can be calculated to:

$$\frac{\partial y_{fuz}}{\partial \rho_j} = \sum_{n=1}^{C}\sum_{k=1}^{N} e_k l_n(x_k) \frac{\partial \Psi_n(x_k)}{\partial \rho_j} \tag{6}$$

with $l_n(x_k)$ the linear rule consequent functions in sample x_k, e_k the residual in sample x_k defined by:

$$e_k(c,\sigma,\rho) := y_k - \sum_{i=1}^{C} l_i(\vec{x}_k)\Psi_i(c,\sigma,\rho)(\vec{x}_k)) \tag{7}$$

and $\frac{\partial \Psi_n(x_k)}{\partial \rho_j}$ the partial derivative of the nth basis function with respect to ρ_j:

$$\frac{\partial \Psi_i}{\partial \rho_k} = \frac{\mu_i}{\sum_{j=1}^{C} \rho_j \mu_j(x)} \left(\delta_{i,k} - \frac{\rho_i \mu_k}{\sum_{j=1}^{C} \rho_j \mu_j(x)} \right) \tag{8}$$

where

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The idea of the projected gradient method is to apply a threshold operator on the rule weight update in each gradient descent iteration step T_α. Additionally, we also have to optimize the consequent parameters in each update step as needed in $l_n(x_k)$ and e_k. This is done by the local learning approach, estimating the parameters for each rule separately and independently. This has some robustness [18], computational, flexibility [1] and interpretability [24] advantages over global learning approach. In this sense, the whole rule weight optimization procedure becomes as in algorithm 1.

Algorithm 1 Rule Weight Optimization with Sparsity Constraints

1. Initialization of w^0 is done by using local learning with (11) on the initial number of rules (see Step 5 below); ρ^0 is set to a vector of ones (all rules contribute equally at the beginning).
2. w^k, ρ^k are given in iteration k.
3. Calculate $\frac{\partial \Psi_n}{\partial \rho_j}$ as given in (8).
4. Update the rule weights $j = 1,...,C$ componentwise separately by:

$$\rho_j^{k+1} = T_\alpha \left(\rho_j^k - \tau \sum_{n=1}^{C}\sum_{k=1}^{N} e_k l_n(x_k) \frac{\partial \Psi_n(x_k)}{\partial \rho_j} \right) \tag{9}$$

with τ a small constant defining the step size per iteration (usual setting $\tau = 0.1$) and T_α the soft-thresholding operator defined by:

$$T_\alpha = \max(\rho_j^{k+1}, \alpha) - \alpha \tag{10}$$

with α set to a small constant (usual guess is 0.1).

ISBN: 978-989-95079-6-8

5. Calculate $\Psi(c, \sigma, \rho^{k+1})$ for the new $\rho_j^{k+1}, j = 1, ..., C$ and generate the regression matrix $\sqrt{Q_i}R_i, i = 1, ..., C$ for each rule separately with R_i containing the original variables and a column of ones for the intercept and the weighting matrix Q_i defined by $diag(\Psi_i(\vec{x}(k))), k = 1, ..., N$.

6. Perform the weighted least squares approach for the linear consequent parameters whose analytical solution in closed form is:

$$\hat{\vec{w}_i} = (R_i^T Q_i R_i)^{-1} R_i^T Q_i \vec{y} \tag{11}$$

   Here, we also apply Tikhonov regularization [21] if required, i.e. when the condition of the matrix $R_i^T Q_i R_i$ is very high, indicating a matrix close to singular and hence an unstable solution in (11).

7. If the difference between two iterations $\|\rho^{k+1} - \rho^k\|$ is low enough or a pre-defined maximal number of iterations is reached, then stop, otherwise perform next iteration (goto Step 2).

After this optimization procedure, the rules with weights close to 0 are eliminated and the remaining rules C_{rem} are kept for further processing through the last step (described in the subsequent section). Note the in view of our derivation of this iteration, the output of this method should be on the one hand close to the data (this is done by minimizing the least squares functional), and have many rule weights as zero, (which can be discarded afterwards).

In order to visualize the update progress of the rule weights through Algorithm 1, Figure 1 shows the rule weights from three iterations (1, 7 and 13). The weights are sorted in order to get an idea how they are reduced through the optimization procedure by applying T_α. Clearly, after the first iteration almost all rule weights are bigger than zero, whereas after the 13th iterations more than 60 rules can be deleted.

5 Feature Selection by Learning of Consequent Parameters with Sparsity Constraints

Now the final step in our algorithm is to perform a final estimation of the linear consequent parameters $\vec{w}_i, i = 1, ...C_{rem}$ for all rules separately. Indeed, the outcome of Algorithm 1 contains not only the rule weights but also estimates the consequent parameters: these are optimized with standard weighted least squares method (and eventually plus a regularization term) in order to minimize least squares error. Here, we want to go a step further and make the consequent parameters also sparse, i.e. forcing as many of these as possible towards zero, again without loosing (significant) accuracy as optimized in an iterative optimization procedure called Semi-smooth Newton method. This evoke a local feature selection, as variables with low weights may be ignored. Furthermore, if through this selection scenario it turns out that certain variables have very low weights in all rules, it can be concluded that these ones are not necessary for the whole global fuzzy model at all.

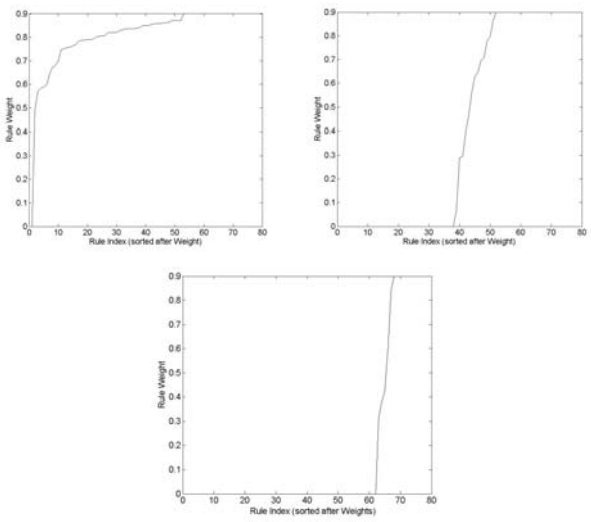


Figure 1: From left to right: sorted rule weights after iteration step 1, 7 and 13; note how the rule base is sparsed out from iteration to iteration (more and more rule weights become 0)

The consequent parameter $\vec{w}_i$ appear linearly in the Function F, in contrast to the rule weights. Although adding sparsity constraints make the optimization a nonlinear problem, because of the linear structure it seems favorable to use a Newton-type method for this problem. Despite the fact that a sparsity constraint problem is usually not differentiable, a generalized version of derivative (the slant derivative) can be defined and subsequently a Newton-type iteration (the Semi-smooth Newton method, [13]) can be applied. The main steps are a soft shrinkage operation on the residual and a Newton step, which makes use of active and inactive sets, see Algorithm 2. For the sake of simplicity, we define this algorithm for the ith rule. Local convergence for this iteration has been shown e.g. in [13].

Algorithm 2 Semi-Smooth Newton with Sparsity Constraints
Let $\sigma > 0$ and parameter k be an iteration index, and M_i the following covariance matrix for local learning

$$M_i = R_i^T Q_i R_i$$

then we apply the following steps:

1. Initialization is done by simply taking the latest estimated $\vec{w}$ (in the last iteration of Algorithm 1).
2. For the kth step we have already $\vec{w}_i^k$ evaluated
3. Step $k \to k + 1$:
   Calculate

$$\vec{u}_{ik} := \vec{w}_i^k - \sigma(R_i^T Q_i (R_i \vec{w}_i^k - \vec{y}))$$

4. Elicit in $\vec{u}_{ik}$ those indices which are greater or smaller than σ (with σ a small value, usually set to 0.1 and not sensitive):

$$\mathcal{A} := \{j : |(\vec{u}_i^k)_j| > \sigma\} \quad \mathcal{I} := \{j : |(\vec{u}_i^k)_j| \leq \sigma\}$$

ISBN: 978-989-95079-6-8

5. Calculate the residual by applying the sparsity constraint operator T_α as defined in (10):

$$\vec{r}_i^{\,k} = T_\alpha(\vec{w}_i^{\,k} - \sigma(R_i^T Q_i(R_i \vec{w}_i^{\,k} - \vec{y})))$$

6. Calculate the update: for all indices on the inactive set:

$$(update_I)_j = -(\vec{r}_i^{\,k})_j \quad \text{für all } j \in \mathcal{I}$$

7. Divide matrix M_i in active and inactive sets:

$$M_{\mathcal{A},\mathcal{A}} := (M_i)_{k,l}, \quad k,l \in calA$$

$$M_{\mathcal{A},\mathcal{I}} := (M_i)_{k,l}, \quad k \in calA, l \in calI$$

8. Calculate update on the inactive set:

$$(update_A)_j = (M_{\mathcal{A},\mathcal{A}})^{-1}\left(\frac{1}{\sigma}(-\vec{r}_i^{\,k})) - M_{\mathcal{A},\mathcal{I}} update_I\right)$$

9. Assign the consequent vector the current update vector one times for the active, one times for the inactive set:

$$\vec{w}_i^{\,k+1} = update$$

10. **If** the residual $\|\vec{r}_i^{\,k}\|$ is large and the iteration smaller than the maximal # of iterations, goto Step 2

An example of the impact of this algorithm on the development of the sparseness of the consequent parameters and hence local/global feature selection will be demonstrated in Section 7.

6 SparseFIS - The Algorithm

The whole *SpareFIS* method is simply obtained by combining the three algorithms described in the previous three sections to one, so applying iterative vector quantization first, then rule weight optimization and finally semi-smooth Newton method for sparse learning of consequent parameters. The explicit formulation is not made here due to space restrictions.

7 Evaluation

7.1 Experimental Setup

The evaluation of the novel method *SparseFIS* includes a comparison with other famous fuzzy batch learning algorithms such as *genfis2* [6] [23], *ANFIS* [15], *FMCLUST* [2] and the batch off-line version of the evolving fuzzy system approach *FLEXFIS* [17], which are all available as implementations in MATLAB. The evaluation will be based on five different high-dimensional data sets for regression: 1.) four data sets from the UCI repository [1] (clean, noise-free): auto-mpg, housing, forrest fires and concrete and 2.) one data set from an engine test bench, where measurement data were collected on-line and is affected by some white noise and disturbances.

The engine test bench data set includes various measurement channels together with their time delay (up to 10), the task was to build a high-performance k-step ahead prediction model for the emission channel NOX for two purposes: 1.) in order to save expenses for the sensor channel and 2.) to perform early recognition of faults (pipe leakage, sensor overheating, interface defects etc.) when using this model in a multi-channel fault detection approach.

A summary of the characteristics of all the data sets used in this evaluation is shown in Table 1. In all data set cases, the evaluation will be based on the accuracy of the models obtained by applying 10-fold cross-validation with a coupled best parameter grid search scenario. The later means that a parameter grid for the most essential parameter(s) in each method is defined consisting of a lower bound, an upper bound and a step-size and for each grid point cross-validation is performed. The best solution with respect to the CV-error (averaged mean absolute error over all folds or averaged maximal error over all folds) will be reported. For *SparseFIS* we will observe the impact of the sparsity constraints on local variable selection (per rule separately). The parameter grids were defined in the following way:

- *genfis2*: radius cluster is varied from 0.1 to 0.9 in steps of 0.05
- *ANFIS*: an initial model is constructed with genfis1 by specifying the number of fuzzy sets per input dimension varied from 2 to 5 in steps of 1
- *FMCLUST*: the number of rules is varied from 1 to 50 in steps of 1
- *FLEXFIS*: the vigilance parameter responsible for the number of clusters = rules is varied from 0.1 to 0.9 in steps of 0.05

For *SparseFIS*, we set the number of maximal allowed (initial) rules to 50 and τ to 0.05 for all experiments, requiring no parameter grid at all. All other parameters are set as mentioned in above algorithms.

7.2 Results

The obtained results are reported in Table 2. The rows represent the different methods and the columns the different data sets: the first number is the CV-error, i.e. the minimal averaged mean absolute error (averaged over the CV-folds, minimal over all parameter settings) plus-minus the standard deviation over these folds. The second number denotes the maximal error over all samples for the minimal averaged mean absolute error (so worst-case prediction error). The third number reports the complexity of the fuzzy models in terms of the number of rules (belonging to the model with minimal CV-error). The results on *SparseFIS* are reported two times, one time by using the full algorithm, one time without applying the sparsity constraint threshold T_α, so without forcing rule weights and consequent parameters towards zero (denoted as *SparseFIS uncon.*). From this table it can be realized that the full *SparseFIS* approach (with constraints) can outperform all other methods in three data sets (forrest fires, housing and NOX) in terms of mean absolute error (MAE), in the case of forest fires

[1] http://archive.ics.uci.edu/ml/

ISBN: 978-989-95079-6-8

Table 1: Some data sets from the UCI repository and their characteristics

	# Training Samples	# Input Variables	Source	Noise Level
Auto-MPG	398	8	UCI	None
Concrete	1030	8	UCI	None
Forrest Fires	517	12	UCI	None
Housing	506	13	UCI	None
NOX	667	181	Engine Test Bench	Medium to high

Table 2: Comparison of fuzzy modelling variants on four data sets from UCI repository and (noisy) NOX data from an engine test bench

Method	**Auto-MPG**			**Concrete**			**Forest Fires**		
	MAE±STD	Max	#rules	MAE±STD	Max	# rules	MAE±STD	Max	# rules
genfis2	2.23± 0.85	3.88	6	8.37± 1.70	11.52	3	19.64± 15.50	58.36	3
ANFIS	2.41± 0.84	4.07	16	11.25± 9.98	39.13	8	20.27± 15.60	58.38	2
FMCLUST	2.35± 0.91	3.99	20	7.75± 2.06	12.11	3	19.83± 15.52	58.93	2
FLEXFIS	**2.17± 0.73**	3.59	11	**7.73± 1.97**	11.96	8	19.69± 15.32	58.35	3
SparseFIS	2.20± 0.30	2.53	8	8.91± 4.55	21.01	6	**15.13± 6.76**	30.77	5
SparseFIS uncon.	2.48± 0.80	4.19	21	11.02± 5.95	25.55	14	20.13± 15.50	59.53	9

Method	**Housing**			**NOX**		
	MAE±STD	Max	#rules	MAE±STD	Max	# rules
genfis2	3.14± 1.31	6.53	4	12.97± 1.08	15.98	5
ANFIS	3.59± 1.39	6.56	4	13.72± 1.01	14.72	8
FMCLUST	2.84± 1.08	5.38	6	13.99± 1.05	15.55	6
FLEXFIS	2.98± 1.27	6.09	6	12.96± 0.98	14.31	8
SparseFIS	**2.84± 0.98**	5.65	6	**12.95± 0.97**	15.22	7
SparseFIS uncon.	3.24± 1.25	6.37	20	13.59± 0.83	14.76	27

even significantly. In the case of the auto-mpg data set, the new method performs similar than the other approaches, whereas for the concrete data set the *SparseFIS* is worse than the others. A similar interpretation of the results can be reported when inspecting the maximal error rate. The standard deviations shows us how sensitive the method is with respect to variations in the training data samples. Lower values represent less variations and hence more robust methods. The best performing methods for each data set with respect to the mean absolute error are shown in bold face. It is also worth mentioning that *SparseFIS* in its unconstrained version performs (mostly) weaker as the full approach; an explanation of this occurrence is that the short version tends to over-fit, as it does not reduce any rules or consequent parameters (and hence complexity).

Another important aspect when training fuzzy systems from data is a reasonable complexity of the obtained models in order to guide the models to more interpretable power [5]. Otherwise, the benefit of fuzzy models over other modelling techniques (such as neural networks, genetic programming etc.) is diminished. We measure the complexity of the final fuzzy models by the number of generated rules by the learning procedure applied onto the whole data set with the optimal parameter setting as obtained in the best parameter grid search and CV procedure. This number is reported in the third columns of each single part in Table 2. Here, we can realize that *SparseFIS* can compete with the other methods, although *genfis2* is mostly the best performer with respect to the number of generated rules. It is also remarkable that the numerical rule weight optimization concept as described in Algorithm 1 is able to reduce the number of rules from 50 to below 10 for all data sets (and this with a reasonable accuracy), confirming Figure 1.

Finally, in Figure 2 we inspect the impact of the sparsity constraints onto the linear consequent parameters as applied in Algorithm 2 (Step 4). Therefore, we show a plot of consequent parameters for all extracted rules (=rows) for the second (upper row) and the third features (lower row) in case of auto-mpg and using all input variables in *SparseFIS* method. While for the second feature four rules are sparsed out (compare left and right image in upper row), for the third feature all consequent parameters are forced to a small value around 0 which is not significant (note the range of this feature is 184). This means that feature #3 is not significant for modelling the target in the auto-mpg data set.

ISBN: 978-989-95079-6-8

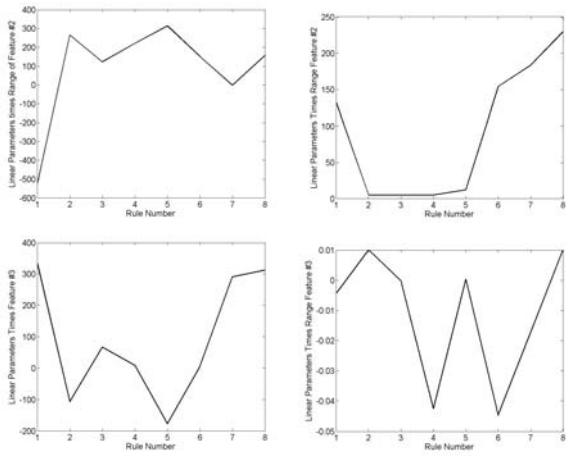


Figure 2: Left: Linear parameters times the ranges for features #2 (upper) and #3 (lower) when not applying sparsity constraints in consequent learning; right: the same as in left, but applying sparsity constraints

8 Conclusion

In this paper, we presented a novel data-driven learning technique for Takagi-Sugeno type fuzzy systems, called *SparseFIS*, which is able 1.) to perform rule selection together with consequent parameter estimation in a coherent optimization procedure and 2.) to gain more insight into the fuzzy models by yielding a local feature selection through sparsing out unimportant features in each rule. As a top-down method, it is possible to elicit an initial upper bound of number of (allowed) rules by expert-knowledge or by initial guess from the data. Tests on various multi-dimensional data sets from UCI repository and real-world applications show that the novel method can compete with other well-known learning methods in terms of both, accuracy and complexity and synchronously is able to put in more interpretation in from of local importance of the features.

References

[1] P.P. Angelov, E. Lughofer, and X. Zhou. Evolving fuzzy classifiers using different model architectures. *Fuzzy Sets and Systems*, 159(23):3160–3182, 2008.

[2] R. Babuska. *Fuzzy Modeling for Control*. Kluwer Academic Publishers, Boston, 1998.

[3] J. Botzheim, E. Lughofer, E.P. Klement, L. Kóczy, and T.D. Gedeon. Separated antecedent and consequent learning for takagi-sugeno fuzzy systems. In *Proceedings of FUZZ-IEEE 2006*, pages 2263–2269, Vancouver, Canada, 2006.

[4] M. Burger, J. Haslinger, U. Bodenhofer, and H. W. Engl. Regularized data-driven construction of fuzzy controllers. *J. Inverse Ill-Posed Probl.*, 10(4):319–344, 2002.

[5] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena. *Interpretability Issues in Fuzzy Modeling*. Springer Verlag, Berlin Heidelberg, 2003.

[6] S. Chiu. Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*, 2(3):267–278, 1994.

[7] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.

[8] O. Cordon, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena. Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy Sets and Systems*, 141(1):5–31, 2004.

[9] O. Cordon and F. Herrera. Identification of linguistic fuzzy models by means of genetic algorithms. *Fuzzy Model Identification. Selected Approaches*, pages 215–250, 2003.

[10] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena. *Genetic Fuzzy Systems — Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. World Scientific, 2001.

[11] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[12] R.M. Gray. Vector quantization. *IEEE ASSP Magazine*, pages 4–29, 1984.

[13] R. Griesse and D.A. Lorenz. A semismooth Newton method for Tikhonov functionals with sparsity constraints. *Inverse Problems*, 24(3):19 p., 2008.

[14] D. Gustafson and W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *Proc. IEEE CDC*, pages 761–766, San Diego, CA, USA, 1979.

[15] J.-S.R. Jang. ANFIS: Adaptive-network-based fuzzy inference systems. *IEEE Trans. Syst. Man Cybern.*, 23:665–685, 1993.

[16] E. Lughofer. Extensions of vector quantization for incremental clustering. *Pattern Recognition*, 41(3):995–1011, 2008.

[17] E. Lughofer. FLEXFIS: A robust incremental learning approach for evolving TS fuzzy models. *IEEE Trans. on Fuzzy Systems (special issue on evolving fuzzy systems)*, 16(6):1393–1410, 2008.

[18] E. Lughofer. Towards robust evolving fuzzy systems. In P. Angelov, D. Filev, and N. Kasabov, editors, *Evolving Intelligent Systems: Methodology and Applications*, page to appear. John Wiley & Sons, New York, 2009.

[19] D. Nauck and R. Kruse. Nefclass-x – a soft computing tool to build readable fuzzy classifiers. *BT Technology Journal*, 16(3):180–190, 1998.

[20] M. Setnes. Simplification and reduction of fuzzy rules. In J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, editors, *Interpretability Issues in Fuzzy Modeling*, volume 128 of *Studies in Fuzziness and Soft Computing*, pages 278–302. Springer, Berlin, 2003.

[21] A.N. Tikhonov and V.Y. Arsenin. *Solutions of ill-posed problems*. Winston & Sonst, Washington D.C., 1977.

[22] L.X. Wang. Fuzzy systems are universal approximators. In *Proc. 1st IEEE Conf. Fuzzy Systems*, pages 1163–1169, San Diego, CA, 1992.

[23] R.R. Yager and D.P. Filev. Learning of fuzzy rules by mountain clustering. In *Proc. of SPIE Conf. on Application of Fuzzy Logic Technology*, pages 246–254. Boston, MA, U.S.A., 1993.

[24] J. Yen, L. Wang, and C.W. Gillespie. Improving the interpretability of TSK fuzzy models by combining global learning and local learning. *IEEE Trans. on Fuzzy Systems*, 6(4):530–537, 1998.

ISBN: 978-989-95079-6-8

Increasing the Performance of Fuzzy Retrieval Using Impact Ordering

Carlos D. Barranco[1] Sven Helmer[2]

[1]Division of Computer Science, School of Engineering, Pablo de Olavide University
Seville, Spain
[2]School of Computer Science and Information Systems, Birkbeck, University of London
London, United Kingdom
Email: cbarranco@upo.es, sven@dcs.bbk.ac.uk

Abstract— *We propose an approach for indexing fuzzy data based on inverted files that speeds up retrieval considerably by stopping the traversal of postings lists early. This is possible because the entries in the postings lists are organized in a way that guarantees that there are no matching items beyond a certain point in a list. Consequently, we can reduce the number of false positives significantly, leading to an increase in retrieval performance. We have implemented our approach and evaluated it experimentally, comparing it to an approach that has previously been shown to be superior to other methods.*

Keywords— fuzzy databases, access methods, inverted files, physical design

1 Introduction

Being able to handle imprecise or uncertain data becomes ever more important in today's world. There are numerous applications that have to manage imperfect data in areas such as knowledge discovery, mediator systems, information retrieval, multimedia, and profiling (vague) users' preferences. One way to model imprecision and uncertainty is to employ concepts from fuzzy set theory. Fuzzy sets are making their way into the database world, evidenced by the fact that several proposals for fuzzy database models and systems exist [1, 2, 3, 4, 5] (for an overview see [6]).

We are focusing on a well-known model for fuzzy databases, namely the possibilistic database model [5]. This approach uses the concepts of possibility and necessity measures for flexible querying. While this is an elegant way to formulate queries, indexing possibility distributions is not straightforward. Bosc and Galibourg introduced an indexing principle based on α-cuts (i.e. the elements of a fuzzy set with a membership degree of at least α) [7]. The parameter α is tied directly to a threshold value determined by a user defining a query, which makes it unrealistic to materialize dozens or even hundreds of α-cuts in different indexes to be able to answer various queries efficiently. For that reason Bosc and Galibourg developed a filtering mechanism based on the supports and cores of the indexed possibility distributions[1]. However, this can introduce a large number of false positives, as the filter only acts as a "quick-and-dirty" test. All the candidates that pass the filter have to be checked for eligibility in a second step.

We propose an access method for indexing possibility distributions that cuts down the number of false positives significantly. Many different α-cuts are stored compactly in a single index structure by arranging the references to possibility distributions in a clever way. The idea is based on impact ordering of inverted files used in information retrieval (IR). In IR systems, document references are sorted in descending order according to the term frequency value to make sure that important documents for a query are processed first [8]. In our case references to possibility distributions are sorted according to the certainty degrees of their elements, making it possible to stop scanning lists of references at an early stage (because we know that no eligible data items can follow). By applying customized gap compression schemes to the inverted files, we can store the lists very compactly.

The remainder of this paper is organized as follows. In the section immediately following this introduction, we cover basic definitions of possibilistic database systems and how to query them. In Section 3 we describe our access method in more detail and explain how to answer queries based on the possibility measure and those based on the necessity measure. Section 4 includes a brief description of the approach that we compared ours to. This is followed by a section on the experimental evaluation, specifying the environment in which this evaluation was run and presenting the results of the evaluation. We wrap up the paper with a conclusion and outlook in Section 6.

2 Preliminaries

Before going into details about indexing, let us briefly define possibility distributions and their application in flexible querying as defined in [5].

2.1 Possibilistic Databases

We assume (without loss of generality) n data items $(o_1, o_2, \ldots, o_n)$ in our database, all of which have an attribute A with a discrete domain Ω. The value of attribute A of data item o_i is described by a possibility distribution function $\pi_{A(o_i)}$ on Ω.

The possibility distribution $\pi_{A(o_i)}$ is a set of possible values for the attribute A of the data item o_i (together with the certainty of each value) and is defined as:

$$\pi_{A(o_i)} : \Omega \mapsto [0, 1] \tag{1}$$

$\pi_{A(o_i)}(\omega_j) = 0$ (for $\omega_j \in \Omega$) means that it is *impossible* that the attribute A can take on the value ω_j. $\pi_{A(o_i)}(\omega_j) = 1$, on the other hand, means that it is *completely* possible that A can take on the value ω_j. If we want to express that it might

[1]Supports and cores are special α-cuts.

ISBN: 978-989-95079-6-8

ω_1	$\rightarrow$	$\alpha_{1,1}$	$r(o_1^{1,1}), r(o_2^{1,1}), r(o_3^{1,1}), \ldots$	$\alpha_{1,2}$	$r(o_1^{1,2}), r(o_2^{1,2}), r(o_3^{1,2}), \ldots$	$\ldots$	$\ldots$
ω_2	$\rightarrow$	$\alpha_{2,1}$	$r(o_1^{2,1}), r(o_2^{2,1}), r(o_3^{2,1}), \ldots$	$\alpha_{2,2}$	$r(o_1^{2,2}), r(o_2^{2,2}), r(o_3^{2,2}), \ldots$	$\ldots$	$\ldots$
$\vdots$							
ω_m	$\rightarrow$	$\alpha_{m,1}$	$r(o_1^{m,1}), r(o_2^{m,1}), r(o_3^{m,1}), \ldots$	$\alpha_{m,2}$	$r(o_1^{m,2}), r(o_2^{m,2}), r(o_3^{m,2}), \ldots$	$\ldots$	$\ldots$

Figure 1: Inverted file index for fuzzy retrieval

be possible for A to take on the value ω_j, then we can assign a value from the interval $(0, 1)$ to $\pi_{A(o_i)}(\omega_j)$, depending on how plausible we think it is that A takes on this value. Not that the possible values for A are mutually exclusive, i.e. A takes on a single value, we are just uncertain about which one is the correct value. For reasons of consistency, $\pi_{A(o_i)}$ should be normalized, i.e. $\exists\omega \in \Omega : \pi_{A(o_i)}(\omega) = 1$.

As already mentioned, we assume that Ω is discrete, since we are interested in indexing scalar data. In the case of applications using continuous numerical data, we either have to discretize the possibility distribution or use a different kind of index structure [9]).

2.2 Flexible Querying

We not only allow imprecision in the data, but also flexible querying, i.e. the query condition c determining the acceptable values of an attribute is described by a normalized fuzzy set $\mu_c : \Omega \mapsto [0, 1]$.

In order to be able to check whether a data item satisfies a query condition, we need to introduce the concept of a fuzzy measure. Let X be an element of the power set of Ω ($X \in \mathcal{P}(\Omega)A$). The *possibility* of X, $\Pi(X)$, is measured by looking at the elements of X:

$$\Pi(X) = \max_{\omega \in X} \pi(\omega) \tag{2}$$

Possibility theory uses two concepts to measure the likeliness of X: the possibility measure, as already described above, $\Pi(X)$ and the necessity measure $N(X)$. The *necessity* of X, $N(X)$, is defined by the unlikeliness of the complement of X:

$$N(X) = 1 - \Pi(\overline{X}) \tag{3}$$

What does this mean for an attribute value satisfying a query condition c? The possibility of doing so is measured with the help of the possibility measure:

$$\Pi(c|A(o_i)) = \max_{\omega \in \Omega} \min(\mu_c(\omega), \pi_{A(o_i)}(\omega)) \tag{4}$$

Informally speaking, $\Pi(c|A(o_i))$ returns the highest degree to which $A(o_i)$ can possibly satisfy c.

The necessity of an attribute value satisfying query condition c is defined as:

$$N(c|A(o_i)) = \min_{\omega \in \Omega} \max(\mu_c(\omega), 1 - \pi_{A(o_i)}(\omega)) \tag{5}$$

$N(c|A(o_i))$, on the other hand, returns the degree to which $A(o_i)$ certainly satisfies c.

Usually users are interested in a small subset of data items in the database. Selective queries can be formulated by providing an acceptance threshold α (e.g. return all data items whose attribute values possibly satisfy c to at least a degree of 0.8: $\{o_i|\Pi(c|A(o_i)) \geq 0.8\}$). Furthermore, using the necessity measure also leads to more selective queries than using the possibility measure.

3 Indexing Possibility Distributions

As mentioned in Section 1 we use an inverted file index to index possibility distributions. An inverted file consists of a *directory* containing all distinct values $\omega_1, \omega_2, \ldots, \omega_m$ of the domain Ω and a *list* (also called a *postings list*) for each value.[2] Within each postings list of the inverted file, we sort the references to data items o_j, denoted by $r(o_j)$, in descending order of the degree of membership of w_i in $\pi_{A(o_j)}$. Since we want to apply gap compression (more on this later in Section 3.3), we arrange the references in blocks, each block preceded by a value $\alpha_{i,k}$, meaning that this is the k-th block of the posting list associated with ω_i. All attributes of the data items referenced in a block satisfy $\pi_{A(o_j)}(\omega_i) \geq \alpha_{i,k}$. Figure 1 illustrates the layout described above; note that the data items have superscripts indicating which block they belong to. The threshold values $\alpha_{i,k}$ divide up the interval [0,1] into equally-sized partitions (i.e. each block covers an equally-sized interval). Within each block all references to data items (in the form of IDs) are sorted in ascending order.

Depending on the query type, we access the inverted file index slightly differently. We will first look at queries based on the possibility measure and then turn to those based on the necessity measure.

3.1 Possibility Measure Queries

If we are searching for all data items o_j whose attribute A satisfies condition c possibly to a degree of at least α, we only have to look up the values ω_i in the inverted file for which $\mu_c(\omega_i) \geq \alpha$ (see Equation (4)). In addition to that, only data items for which $\pi_{A(o_j)}(\omega_i) \geq \alpha$ will qualify. During a search, we scan a postings list until we reach the first block whose value $\alpha_{i,l}$ is smaller than the query threshold value α.[3] If α is between $\alpha_{i,l-1}$ and $\alpha_{i,l}$, then we still have to check the data items referenced in this block whether they are false positives or not (as some of them could satisfy the query predicate). However, only the data items in this last block can be false

[2] For an overview of traditional inverted files see [10].

[3] Obviously, we also stop when we reach the end of a postings list.

ISBN: 978-989-95079-6-8

positives and have to be checked, all other data items referenced in earlier blocks will satisfy the query predicate and can be retrieved without checking for false positives. If α is equal to $\alpha_{i,l-1}$ we do not have to check the items of block $\alpha_{i,l}$. Actually, in this case we do not have to check for false positives at all.

3.2 Necessity Measure Queries

Processing queries based on the necessity measure can be done in two different ways, depending on the cardinality of the domain Ω. For large domains we can simply use possibility measure queries as a filter (we will call this method 'simplified' in the following); for small domains we can use the index described above to determine the answer set (without accessing the data items themselves).

The inequality $N(X) \leq \Pi(X)$ always holds for necessity and possibility measures, i.e. $N(c|A(o_i)) \geq \alpha \Rightarrow \Pi(c|A(o_i)) \geq \alpha$. So instead of directly searching for the data items that satisfy the query condition c necessarily to the degree α, we search for data items satisfying c possibly. We retrieve these data items and check if they also satisfy c necessarily. A drawback of this technique is that we introduce false positives, due to the data items that satisfy c possibly but not necessarily.

Accessing all the candidates returned by the possibility measure query in the simplified processing will cause a lot of random I/O. However, we can use the inverted file index to help us in sorting out false positives returned by a possibility measure search. If we can find an ω_i for a data item o_j such that $\max(\mu_c(\omega_i), 1 - \pi_{A(o_j)}(\omega_i)) < \alpha$, then we know that $A(o_j)$ cannot satisfy c necessarily (more on this in just a moment, see also Equation 5). Having a value for $\mu_c(\omega_i)$ that is greater or equal to α will never result in $\max(\mu_c(\omega_i), 1-\pi_{A(o_j)}(\omega_i)) < \alpha$. So in order to sort out false positives, we have to access the postings lists of the values ω_i for which $\mu_c(\omega_i) < \alpha$. Unfortunately, this means we have to look up the complement of the values ω_i used for the possibility measure query and scan the corresponding postings lists.[4] (Consequently, this will only be efficient for relatively small domains.) During query processing, we keep scanning such a postings list in a blockwise fashion as long as $\alpha_{i,l} > 1 - \alpha$ holds and we stop as soon as this condition is not satisfied anymore.[5] Any references to data items found during this scan can be safely discarded from the candidates determined by the possibility measure query. The reasoning is the following: while $\alpha_{i,l} > 1 - \alpha$, we know that for all data items referenced in this block $\pi_{A(o_j)}(\omega_i) > 1 - \alpha$ ($\Leftrightarrow 1 - \pi_{A(o_j)}(\omega_i) < \alpha$) holds. Together with $\mu_c(\omega_i) < \alpha$, this means that $A(o_j)$ cannot satisfy c necessarily. Once $\alpha_{i,l}$ drops below $1 - \alpha$, we do not have enough information to determine whether the data items referenced in this block (and the following blocks) satisfy c necessarily or not and we stop scanning the list further. All the data items remaining in the candidate set after scanning the lists of the ω_i for which $\mu_c(\omega_i) < \alpha$ have to be fetched and checked for false positives.

[4] This is why we will also call this approach complement-based.

[5] Again, we also stop when reaching the end of a list.

3.3 Index Compression

The reason for arranging the references to data items in blocks is that we can use gap encoding on the ascending sequences of data item IDs. For example, instead of storing the sequence 103,110,114,116,121 we store 103,7,4,2,5. Compared to absolute ID numbers, relative gaps can be stored using less storage space. However, from time to time even gaps can become quite large, i.e. a large gap could need almost as much storage space as a regular ID. So that we do not have to allocate the same space for each gap, we use variable length encoding [11]. The main idea is to use few bits for small gaps and many bits for large gaps.

In order to keep the codewords byte-aligned for faster processing, we use Variable-Byte (VB) code [12]. Each byte contains a so-called continuation bit, which signals if the end of the current codeword has been reached or if we have to continue decoding. Usually the high bit is dedicated to this purpose, while the lower 7 bits of each byte encode gap information. If a gap fits into 7 bits, then we encode this gap in the lower 7 bits and set the continuation bit to 1. Otherwise, we encode the highest bits of the gap and set the continuation bit to 0. The remaining bits are encoded in the same manner. VB-encoding is a good compromise between compression ratio and processing speed. For example, encoding 5 in VB-code would result in **1**0000101, while encoding 824 would result in **0**0000110 **1**0111000 (where the bits in boldface are the continuation bits).

Instead of using floating-point numbers for the threshold levels $\alpha_{i,k}$ directly in the index, we store IDs of threshold levels using a one-byte integer, looking up the floating-point values in a table mapping the IDs to these values. Due to the fixed threshold levels we use throughout the whole index, we can do this. As we will see later in the evaluation section, 256 different threshold levels are more than enough. If more than 256 threshold levels should be needed, we can apply VB-encoding to the threshold levels as well.

4 Comparison to Existing Approach

We use the approach by Bosc and Galibourg [7] as a reference for comparison. There are two special α-cuts, the *core* $L_1(\mu_F)$ and the *support* $L_{>0}(\mu_F)$ of a fuzzy set F. Bosc and Galibourg have shown that the following implications hold (for all $\alpha > 0$):

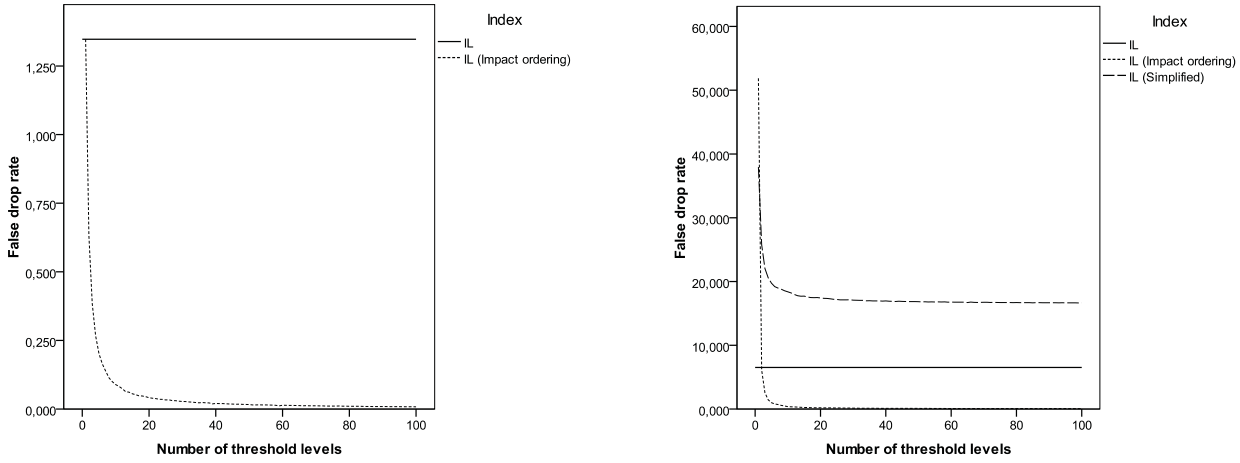
$$\Pi(c|A(o_i)) \geq \alpha \quad \Rightarrow \quad L_{>0}(\pi_{A(o_i)}) \cap L_\alpha(\mu_c) \neq \emptyset \qquad (6)$$

$$N(c|A(o_i)) \geq \alpha \quad \Rightarrow \quad L_1(\pi_{A(o_i)}) \subseteq L_\alpha(\mu_c) \qquad (7)$$

Given a query condition c we have to determine whether the α-cut of μ_c ($L_\alpha(\mu_c)$ intersects with the support of a data item's attribute value (for possibility queries) or if the core of the attribute value is a subset of $L_\alpha(\mu_c)$ (for necessity queries). If this is not the case, then we can safely discard a data item, as it will not satisfy the original query condition.

We implemented this technique using the same framework as for our approach, i.e. using compressed inverted files for indexing the cores and supports of possibility distributions. The index containing the cores stores the cardinality of each core together with the reference to the possibility distribution. In

ISBN: 978-989-95079-6-8



(a) Possibility measure

(b) Necessity measure

Figure 2: False drop rates

this way, superset queries (as described in Equation (7)) can be answered using an inverted file by filtering out all references whose cores have a higher cardinality than the query set. For a more detailed description on how to optimize this kind of data structure, see [13] (comparisons to other data structures can also be found in this paper; the inverted file approach, however, was the data structure with the best performance among all those that were tested).

5 Experimental Evaluation

5.1 Benchmark Environment

The benchmarks were run on a Pentium 4 (3.2 GHz) PC with 1GByte main memory running Windows XP. The index structures and a simulator for the storage system were implemented in Java 1.6. We decided to implement the simulator in Java, as this makes us platform-independent. The simulator allows us to avoid unwanted effects due to caching. We set the block size of the simulator to a page size of 4K, which is the page size of the underlying operating system.

We generated possibility distributions for the data items using the following parameters: the domain size ($|\Omega|$) is equal to 25 and the database size is equal to 100K data items (each data item is associated with a possibility distribution). The elements of a particular possibility distribution are taken randomly from the domain, using a uniform distribution. This means that we first determine the cardinality of the support of a possibility distribution, which is uniformly distributed in [1,22] (22 elements cover 90% of the domain). All the elements in the domain have the same probability of making it into the support. The membership degrees of the chosen elements are uniformly distributed in (0,1.0]. If a possibility distribution is not normalized, the highest degree is converted to 1.0.

In this paper we wanted to focus on typical fuzzy scalar domains, therefore we decided to keep the domain cardinality rather small. This is motivated by the fact that the elements of fuzzy domains usually represent categories. Normally these categories are determined by experts based on the criteria (1) that they should be easy to use by ordinary users and (2) that fuzzy relations (allowing flexible comparisons) can be defined on them. It is much easier to satisfy both criteria with a small domain cardinality; that is why typical fuzzy scalar domains are rather small.

Another important parameter is the clustering of the data items on disk pages. For the results shown in this paper we assume a "worst-case scenario" in terms of indexing, meaning that all the candidate items returned by the index are stored in consecutive blocks with 64.5 data items sharing a block on average. This will lessen the effect that false drops have on the retrieval performance, as we do not need a separate page access to fetch each of the false drops.

5.2 Results

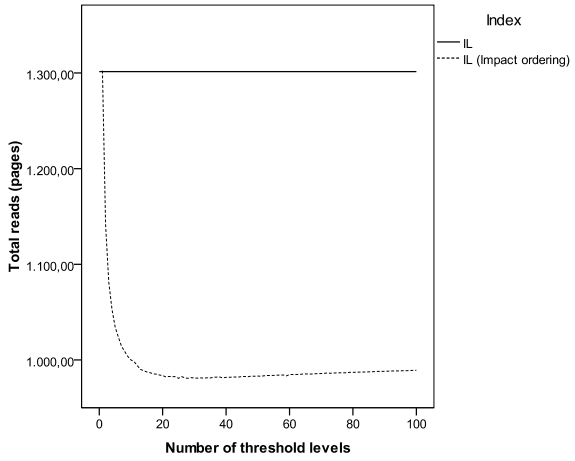
We were interested in several criteria, most importantly in the false drop ratio of the two different approaches and its impact on the query performance. To a lesser extent we also wanted to make sure that the overhead in terms of the index size is not prohibitive. Furthermore, we investigate the impact of the query threshold α on the performance of the indexes. In Sections 5.2.1 to 5.2.3 our competitor is included as a reference; the parameter we vary (number of threshold levels) only applies to our approach in these cases.

5.2.1 False Drop Rate

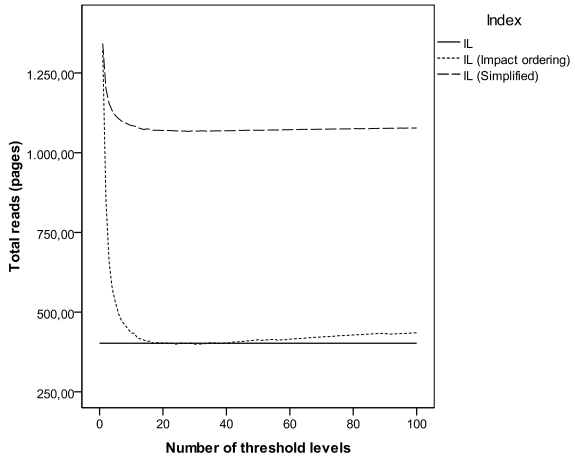
Figure 2 shows the results we obtained for measuring the false drop rate for varying the number of threshold levels in our index (Figure 2(a) depicts the results for possibility measure queries, while Figure 2(b) depicts those for necessity measure queries[6]). Each threshold level covers an equally-sized interval in [0,1]. The false drop rate is measured in the following way: (total number of IDs returned by index - number of true positives)/number of true positives. As can be clearly seen, adding more threshold levels to the data structure reduces the false drop ratio. However, the law of diminishing returns applies to this, i.e. with each added threshold level the improvement gain decreases. We can also notice that using a

[6]The method labeled 'simplified' applies possibility measures as a filter, while the other curve shows the results for the complement-based approach.

ISBN: 978-989-95079-6-8



(a) Possibility measure



(b) Necessity measure

Figure 3: Total number of page accesses

possibility measure as a filter for a necessity measure is not competitive in this scenario.

5.2.2 Retrieval Performance

While the false drop rate suggests to use a very high number of threshold levels, it is not as simple as that to determine the optimal number of threshold levels. By increasing the number of threshold levels we introduce an overhead in terms of index size and retrieval performance (larger indexes result in a longer query processing time). Let us first look at the retrieval performance of the access methods. Figure 3 shows the retrieval performance in total number of page accesses (part (a) for possibility measure queries and part (b) for necessity measure queries). The total number of page accesses includes the page accesses needed to navigate the index and the pages accessed for retrieving all candidate data items (false and true positives).

There is a sharp decrease in the number of page accesses when going from one threshold level to about twenty threshold levels. This is caused by the rapidly dropping false drop ratio. Then the curve peters out and slowly goes up again. In this part of the curve, the drop in performance, due to the aforementioned overhead, gains the upper hand over the slowly decreasing false drop rate. In our benchmark scenario we have identified 25 to be the optimal number of threshold levels for possibility measure queries and 28 to be the optimum for necessity measure queries. Again, the approach using the possibility measure as a filter for necessity measure queries is not competitive, so we will drop it from the graphs in the following sections.

5.2.3 Index Size

In Figure 4 we present the size of the index structures in terms of disk pages. The overhead added for introducing a larger number of threshold levels can be clearly seen. This is caused by a deterioration of the compression ratio. A few larger threshold blocks with smaller gaps between the IDs can be compressed better than a lot of smaller blocks with larger gaps. The stair-like appearance is caused by postings lists expanding by one page at roughly the same time (due to the uniform distribution of domain elements in the possibility distributions). At first glance it seems that our index structure needs more storage space than our competitor. However, as we only need about 20 to 30 threshold levels, this increase is only moderate and, more importantly, we can answer both, possibility measure and necessity measure, queries with the same index. For the other scheme separate indexes are needed (one storing the supports and one storing the cores of the possibility distributions).

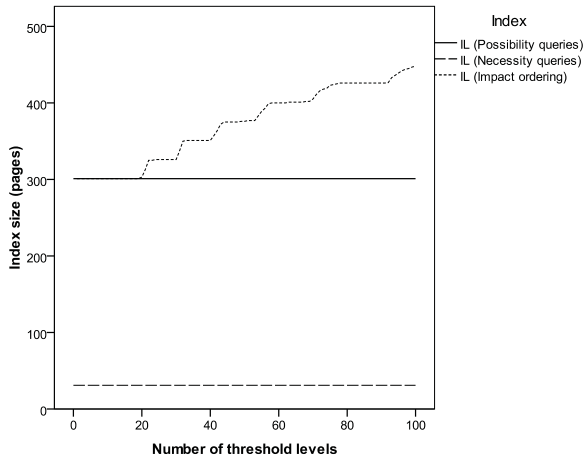
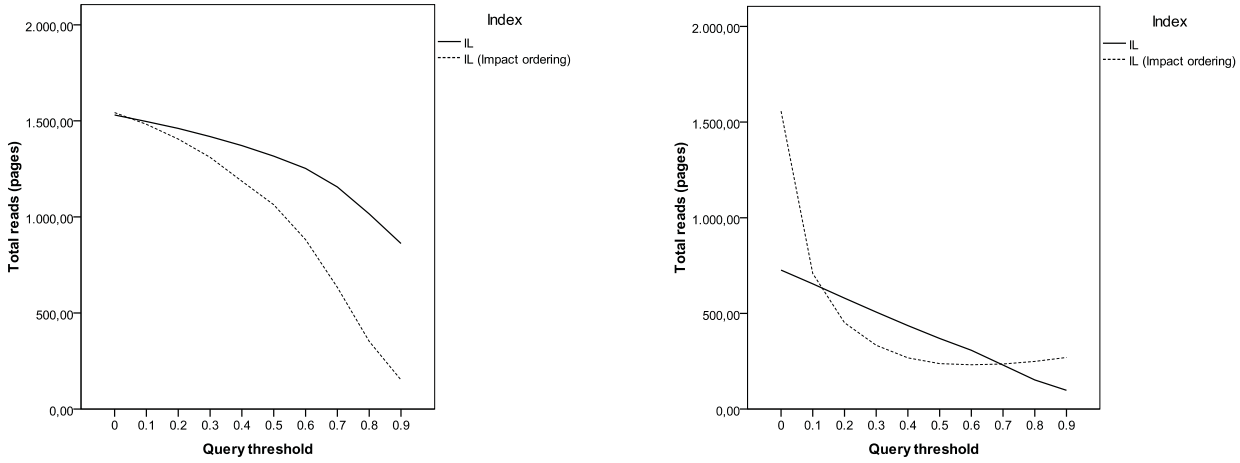


Figure 4: Index size

5.2.4 Impact of Query Threshold

Figure 5 shows the impact of the query threshold (α) on the performance. Generally, as the query threshold grows for possibility measure queries (Figure 5(a)), we have to check smaller and smaller parts of the query set (only elements that have a membership degree larger than or equal to α have to be considered). This in turn decreases the number of postings lists we have to traverse, leading to less page accesses. In addition to this, our approach can stop scanning postings lists earlier for higher query thresholds, leading to further improve-

ISBN: 978-989-95079-6-8



(a) Possibility measure

(b) Necessity measure

Figure 5: Impact of query threshold

ment in terms of the number of page accesses.

For necessity measure queries ((Figure 5(b)), things are more complicated. For the core-filtering approach by Bosc and Galibourg the number of postings lists we have to scan decreases with increasing threshold α. For our approach there are two effects at work. With increasing α the false drop rate drops, which leads to fewer page accesses due to false positives. However, with increasing threshold values the number of page accesses in the index increases. Although the number and lengths of postings list traversals for the possibility measure filter step goes down, the number and lengths of postings list scans for the subsequent complementary check on the necessity measure goes up at quite a fast rate. At some point this effect catches up with us and the number of accessed blocks increases again.

6 Conclusion and Outlook

We have shown how to build an index that stores threshold levels of possibility distributions in a compact way. Being able to represent possibility distributions using a much finer granularity than existing approaches helps us in reducing the number of false positives significantly. Even in an unfavorable environment, our index far outperforms existing approaches for possibility measure queries and was on a par for necessity measure queries. This kind of performance can already be achieved by introducing only a moderate number of threshold levels, meaning that there is practically no overhead, since we only need to maintain one index (while our main competitor needs one for possibility measure queries and one for necessity measure queries).

For future work we would like to analyze how to optimize the number and distribution of threshold levels given a fixed database. Even more interesting would be to optimize and update the index in a dynamic way, given a database with statistics on query workloads.

Acknowledgments

This work has been partially supported by the "Ministerio de Ciencia y Tecnología (MCYT)" (Spain) under grant TIN2007-68084-CO2-01, and the "Consejería de Innovación Ciencia y Empresa de Andalucía" (Spain) under research projects P06-TIC-01570 and P07-TIC-02611. We also thank the anonymous referees for improving the readability of the paper.

References

[1] B. Buckles and F. Petry. A fuzzy model for relational databases. *Fuzzy Sets and Systems*, 7(3):213–226, 1982.

[2] R. Caluwe de, editor. *Fuzzy and Uncertain Object-oriented Databases (Concepts and Models)*, volume 13 of *Advances in Fuzzy Systems - Applications and Theory*. World Scientific, 1997.

[3] R. George, A. Yazici, F.E. Petry, and B.P. Buckles. Uncertainty modeling in object-oriented geographical information systems. In *Third Int. Conf. for Database and Expert System Application (DEXA)*, Valencia, 1992.

[4] F. E. Petry and P. Bosc. *Fuzzy databases: principles and applications*. International Series in Intelligent Technologies. Kluwer Academic Publishers, 1996.

[5] H. Prade and C. Testemale. Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries. *Information Sciences*, 34:115–143, 1984.

[6] J. Galindo, A. Urrutia, and M. Piattini. *Fuzzy Databases: Modeling, Design and Implementation*. Idea Group Publishing, Hershey, PA, USA, 2006.

[7] P. Bosc and M. Galibourg. Indexing principles for a fuzzy database. *Information Systems*, 14(6):493–499, 1989.

[8] C.D. Manning, P. Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[9] C. D. Barranco, J. R. Campaña, and J. M. Medina. A B+-tree based indexing technique for fuzzy numerical data. *Fuzzy Sets and Systems*, 159(12):1431–1449, 2008.

[10] R. Sacks-Davis and J. Zobel. *Indexing Techniques for Advanced Database Systems*. Kluwer Academic Publishers, 1997.

[11] I.H. Witten, A. Moffat, and T.C. Bell. *Managing Gigabytes*. Morgan Kaufmann, San Francisco, 1999.

[12] A. Trotman. Compressing inverted files. *Information Retrieval*, 6(1):5–19, 2003.

[13] S. Helmer. Evaluating different approaches for indexing fuzzy sets. *Fuzzy Sets and Systems*, 140(1):167–182, 2003.

ISBN: 978-989-95079-6-8

Colour Image Segmentation using A-IFSs

Pedro Couto, Pedro Melo-Pinto[1] Humberto Bustince, Edurne Barrenechea, Miguel Pagola[2]

1.CITAB Research Centre, UTAD University
Vila Real, Portugal
2.Departamento de Automática y Computación, UPNA University
Pamplona, Spain
Email: {pcouto,pmelo}@utad.pt, {bustince,edurne.barrenechea,miguel.pagola}@unavarra.es

***Abstract*— *The problem of segmentation is an important research field and often a critical and decisive preprocessing step for many image processing applications. During the past years, the use of colour images instead of gray images is growing fast in many research areas. Moreover, the use of segmentation over colour images is advised when additional colour information conveyed is necessary to further discriminate the image regions. Recently, fuzzy logic theory, and Atanassov's intuitionistic fuzzy sets (A-IFSs), have been successfully applied to image segmentation and, in this paper we propose a general methodology for RGB colour image segmentation based on A-IFSs. The proposed methodology is based on a multilevel colour thresholding framework that uses Atanassov's intuitionistic index values for representing the uncertainty present in the assigning of the pixels to the different regions. This framework is to be applied to each RGB component separately to finally aggregate the results of the three components. The number of thresholds is auto established by the methodology based on the image entropy. Experimental results are presented.*

Keywords— Atanassov's Intuitionistic Fuzzy Sets, Color Image Segmentation, Fuzzy Logic, Multi-Thresholding.

1 Introduction

Color image segmentation deals with the subdivision of an image into disjoined regions according to the pixels colour with the goal of finding the image objects. As a result of the segmentation, the original set of pixels in the image is partitioned into a finite set of separated regions (objects).

Within the framework of fuzzy theory [1, 2, 3, 4], segmentation algorithms that use the concept of fuzzy entropy [1, 3, 5] are widely used in image processing applications. In previous works [6] we have proposed methodologies for image segmentation that use A-IFSs [7, 8, 9]. In those methodologies, Atanassov's intuitionistic fuzzy index values are used for representing the uncertainty/impresision on determining whether a pixel of the image belongs to an object of that image.

The general framework presented in [6] is the basis for colour segmentation framework proposed in this paper where it is extended to a multi-level thresholding framework since, often, images include several regions and, thus, it is necessary to compute more than one threshold in order to correctly segment the image.

In this methodology we use RGB images in such way that each RGB component is primarily segmented separately. After the three components are segmented we then aggregate the three segmented components into the final segmented image. For each component, the number of computed thresholds is automatically determined by the methodology on the basis of the images intuitionistic fuzzy entropy. This entropy is obtained trough the Atanassov's intuitionistic fuzzy index values.

The remainder of this paper is organized as follows: In Section 2 we describe the proposed methodology. Section 3 presents the experimental results obtained and, finally, in Section 4 some conclusions and some lines of future work are presented.

2 Proposed methodology

In this section we present the proposed methodology for colour image segmentation that besides its low computational cost is also able to auto determine the number of required thresholds based on the image intuitionistic fuzzy entropy.

Since this methodology deals with each RGB component separately, we will only describe the algorithm for one of the components. The exact same algorithm is to be applied to the other two components.

This methodology successively applies the bi-level thresholding algorithm proposed in [6], for the computation of one threshold, an unspecified number of times to an image, resulting in a set of thresholds for each RGB component. After the number of required thresholds needed for each RGB component are computed, the total number of thresholds is calculated and, on the basis of that number, the three component segmentations are combined in order to create a resulting gray segmented image. The combination of the components is made by adding all three intensities and afterwards normalizing them back to the original scale. Different ways of combining the three RGB thresholds can be explored.

Thus, in the following sub-sections, we will first describe the bi-level thresholding algorithm used and, in the last subsection the multi-level thresholding algorithm.

2.1 Bi-level Thresholding Algorithm

In this section we describe the general framework for image thresholding using A-IFSs and restricted dissimilarity functions proposed in [6]. The goal is to obtain the better threshold in order to separate the object from the background.

We will denote by (x, y) the coordinates of each pixel on the image Q, being $q(x, y)$ the gray level of the pixel (x, y) so that $0 \leq q(x, y) \leq L - 1$ for each $(x, y) \in Q$ where L is the image grayscale.

In this methodology Atanassov Intuitionistic Fuzzy Sets are used in the following way: In order to choose/construct the

ISBN: 978-989-95079-6-8

membership function of each pixel of the image to the associated fuzzy set, three numerical values are assigned to each one of them.

- A value for representing the expert knowledge of the membership of the pixel to the background. A membership function, constructed using dissimilarity functions, is used to obtain this value (see [10]).
- Dissimilarity functions are also used by the expert to construct a membership function to retrieve a value for representing the expert knowledge of the membership of the pixel to the object.
- The expert knowledge/ignorance, in determining the above mentioned membership functions, is represented by a third value obtained trough Atanassov's intuitionistic index.

In this sense, L fuzzy sets Q_{Bt} associated with the background and L fuzzy sets Q_{Ot} associated with the object are constructed. Each one of these fuzzy sets is associated with a gray level t of the grayscale L used. The membership functions of these sets are defined by means of restricted dissimilarity functions and, two of the possible expressions one might use are:

$$\mu_{Q_{Bt}}(q) = F\left(d\left(\frac{q}{L-1}, \frac{m_{\mathrm{B}}(t)}{L-1}\right)\right)$$

$$\mu_{Q_{Ot}}(q) = F\left(d\left(\frac{q}{L-1}, \frac{m_{\mathrm{O}}(t)}{L-1}\right)\right)$$

where

$$m_{\mathrm{B}}(t) = \frac{\sum_{q=0}^{t} qh(q)}{\sum_{q=0}^{t} h(q)} \quad (1)$$

$$m_{\mathrm{O}}(t) = \frac{\sum_{q=t+1}^{L-1} qh(q)}{\sum_{q=t+1}^{L-1} h(q)} \quad (2)$$

being $h(q)$ the number of pixels of the image with the gray level q, $F(x) = 1-0.5x$ and, the restricted dissimilarity function $d(x, y) = |x - y|$ (see [6, 11, 12]).

Due to these constructions, the membership functions constructed are always greater than or equal to 0.5 and, a pixel unequivocally belongs to one of the regions (object or background) if and only if its intensity q is equal to the mean of intensities of the that region ($m_O(t)$ or $m_B(t)$). When the difference between the pixel's intensity q and the mean of intensities of the region is maximal, then the value of its membership function to that region is minimal. Thus, the lesser the distance between a pixel's intensity q and the mean of intensities of the region, the greater the value of its membership to that region.

In this methodology, it is considered that the membership functions ($\mu_{Q_{Bt}}$ and $\mu_{Q_{Ot}}$) indicate the expert's degree of knowledge of the pixel belonging to the regions, meaning that, the Atanassov's intuitionistic fuzzy index (π) represents the ignorance/intuition present in the construction of these fuzzy sets.

Thus, the value represented by Atanassov's intuitionistic index indicates the knowledge/ignorance of the expert when assigning a pixel either to the background or the object. The Atanassov's intuitionistic index value associated with a pixel has the value of zero, when the expert is absolutely sure of that pixel belonging (either to the background or the object). The Atanassov's intuitionistic index value increases with respect to the ignorance/intuition of the expert as to whether the pixel belongs to the background or the object. If the expert doesn't know if a pixel belongs to the background or the object, its membership to both must be represented with the value 0.5 and, it is said that the expert used the greatest ignorance/intuition allowed in the construction of the membership functions, of the set associated with that pixel, to the background and the object resulting in a Atanassov's Intuitionistic Fuzzy Index maximum value. For this reason, A-IFSs (Atanassov's Intuitionistic Fuzzy Set [8, 13]) are used.

However, the ignorance/intuition should have the least possible influence on the choice of the membership degree. In this methodology the ignorance/intuition will have a maximum influence of 25 percent.

Under these conditions, within the set of all the possible expressions(see [6]), the following expression is used to calculate π:

$$\pi(q) = \wedge(1 - \mu_{Q_{Bt}}(q), 1 - \mu_{Q_{Ot}}(q)).$$

This π is now used to construct an A-IFS with each one of the fuzzy sets Q_{Bt} and Q_{Ot}.

$$\tilde{Q}_{Bt} = \{(q, \mu_{\tilde{Q}_{Bt}}(q), \nu_{\tilde{Q}_{Bt}}(q)) | q = 0, 1, \cdots, L-1\},$$

given by

$$\mu_{\tilde{Q}_{Bt}}(q) = \mu_{Q_{Bt}}(q)$$
$$\nu_{\tilde{Q}_{Bt}}(q) = 1 - \mu_{\tilde{Q}_{Bt}}(q) - \pi(q)$$

and

$$\tilde{Q}_{Ot} = \{(q, \mu_{\tilde{Q}_{Ot}}(q), \nu_{\tilde{Q}_{Ot}}(q)) | q = 0, 1, \cdots, L-1\},$$

given by

$$\mu_{\tilde{Q}_{Ot}}(q) = \mu_{Q_{Ot}}(q)$$
$$\nu_{\tilde{Q}_{Ot}}(q) = 1 - \mu_{\tilde{Q}_{Ot}}(q) - \pi(q)$$

At this stage, the Atanassov's intuitionistic fuzzy index is used to calculate the entropy IE of each one of the L Atanassov's intuitionistic fuzzy sets associated with the image.

The following expression is used to calculate the entropy IE, so that $0 \leq IE(\tilde{Q}_{Bt}) \leq 0.25$.

$$IE(\tilde{Q}_{Bt}) = \frac{1}{N \times M} \sum_{q=0}^{L-1} h(q) \cdot \pi(q) \quad (3)$$

where $N \times M$ are the image dimensions in pixels.

This way, the entropy on A-IFSs is interpreted as a measure of the degree of a A-IFS that a set has with respect to the fuzzyness of the said set (see [14]). Under these conditions, the entropy will be null when the set is a FSs and will be maximum when the set is totally intuitionistic.

ISBN: 978-989-95079-6-8

Finally, the gray level associated with the Atanassov's intuitionistic fuzzy set $\tilde{Q}_{Bt}$ of lowest entropy IE is selected as the best threshold.

2.2 *Multi-level Thresholding Algorithm*

In this approach, the goal is to obtain a set of thresholds in order to separate all the image objects. This method is based on a divide and conquer strategy and its main idea is to successively apply the algorithm proposed in [6] and presented in the last section, for the computation of one threshold, an unspecified number of times to an image. In Fig. 1 we illustrate the methodology computational process progress.

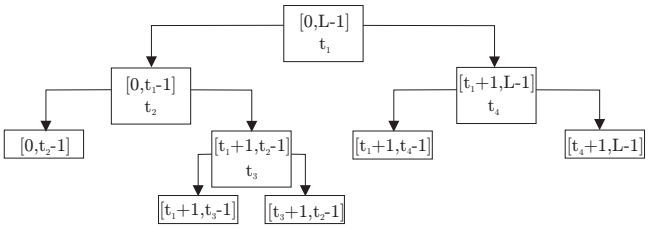


Figure 1: Computational process progress

First, the algorithm is applied to the original image resulting in the determination of the threshold value t_i. This threshold value is then used to create two sub-images: the sub-image with intensity values lower than t_i and the sub-image with intensity values greater than t_i. Then, the algorithm is to be applied to the sub-image which has the greater amplitude between gray-levels entropy values (difference between the intensity value of maximum entropy and the one of minimum entropy).

Hence, The proposed algorithm consecutively divides the resulting sub-images by means of the threshold value t_i obtained through the application of the algorithm for the computation of one threshold to each one of them.

The algorithms stopping criteria is based on the sub-images gray-levels entropy values in such way that the process stops when the difference between the entropy of the computed threshold value t_i and the maximum entropy of the sub-image that generated it, is inferior to the difference between the maximum and minimum entropies of the entire image divided by a constant K.

After experimental evaluation, the value of the constant K was set to 2.

The number of computed thresholds is closely related to the value of this constant K. This value is directly proportional to the number of computed thresholds since, the maximum difference of global entropy allowed for each sub-image to produce a new threshold is inversely proportional to the value of K. Thus, the value of the constant K can be chosen according to the purpose of the application where the algorithm is to be applied.

3 Experimental results

In order to illustrate the methodology we used the "Lena" image in Fig. 2 and Fig. 3.

In Fig. 2 the original "Lena" image and the segmentation result are presented. In Fig. 3, in the first column, we present

Figure 2: Original and segmented "Lena" image.

the image decomposed in its RGB components and, in the second column, each RGB component segmentation result is presented. For all the "Lena" segmented images the number of computed thresholds is also presented.

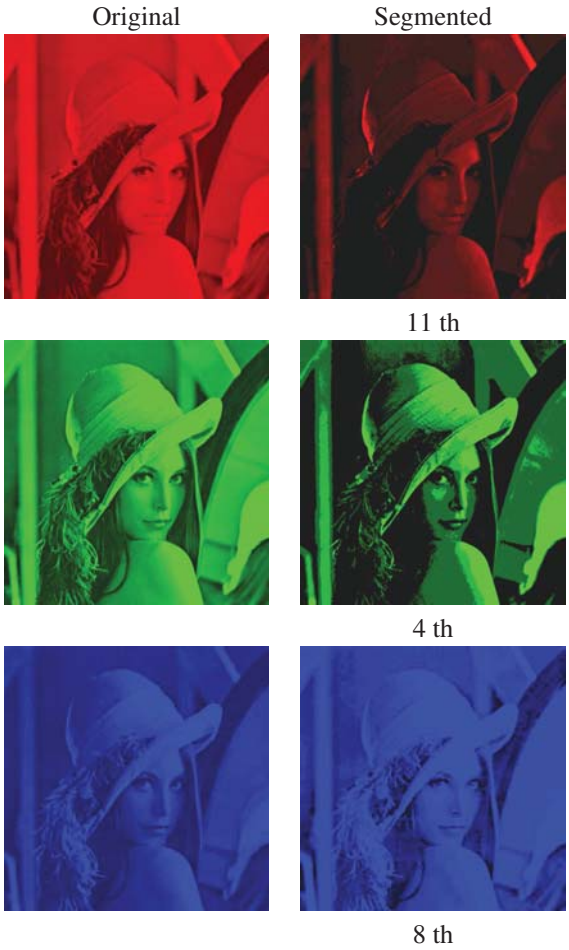


Figure 3: Original and segmented RGB components of "Lena" image.

Regarding the number of computed thresholds in each RGB component we can conclude that, as one can see looking at he original image, the red component is the one that has more information regarding the segmentation of the image and, therefore the red component was the one with more computed thresholds. We think that this is an indicator of the goodness of the methodology since it was able to accurately auto determine the required number of thresholds in each component

ISBN: 978-989-95079-6-8

separately.

Looking at Fig. 3, we can also see that the RGB components are highly related. We consider this to be the major drawback of segmenting images in the RGB colour space.

To test the performance of the proposed approach, more than a hundred images where segmented. However, in order to illustrate the obtained results, 5 images from the original set of images (taken form the Berkeley Segmentation Dataset: `http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/`) were selected and used as test images (see Fig. 4).

Figure 4: Original and segmented test images.

In Fig 4 we present the results from the application of the proposed methodology to complex images where, the value of the constant K was also set at 2 and, where we can see that the main features of the images are preserved.

4 Conclusions and Future Work

Since, an image usually contains more than one region, in this paper we presented a new extension of an existing method, that uses A-IFSs for multi-level colour image segmentation. The algorithm first segments each RGB component separately and finally combines them in order to obtain the final segmented image. The use of A-IFSs allowed us to use entropy in the same sense as fuzzy entropy is used in the fuzzy algorithms and, consequently, allowed us to endow the methodology with the capability of dealing with uncertainty.

The proposed methodology successively applies a bi-level thresholding algorithm an undetermined number of times to an image RGB component subdividing it into several disjoined regions. Moreover, the algorithm is able to auto determine the number of thresholds required to segment each RGB image component.

Considering the experimental results, we can say that, in general terms, the proposed algorithm provides suitable results and therefore can be considered when colour image segmentation applications are needed.

The proposed methodology is parameterizable through the setting of a constant K. The value of this constant is directly proportional to the optimal number of thresholds computed. Thus, the value of the constant K should always depend of the application needs. Future work is intended to study the incorporation of heuristics in the algorithm through the setting of the constant K value. Different combinations of the resulting thresholds of each one of the RGB components can be an object of future work in order to incorporate the perceptual differences among the three components.

Due to the high correlation between the three RGB components, further work is also intended to apply an extension of this algorithm to images in the HSV colour space. Being the HSV colour space a more intuitive color space, it's our belief that a more coherent segmentation will be achieved.

Acknowledgment

This work has been partially supported by *Acções integradas Luso-Espanholas* 2007, *Acciones Integradas Hispano-Portuguesas*, ref. HP2006-0128 (MEC), and POCI 2010 (*Programa Operacional Ciência e Investigação* 2010)

References

[1] J. Bezdek, J. Keller, R. Krisnapuram, and N. Pal. *Fuzzy Models and algorithms for pattern recognition and image processing, The Handbooks of Fuzzy Sets Series*. Kluwer Academic Publishers, Boston/London/Dordrecht, 1999.

[2] Z. Chi, H. Yan, and T. Pham. Optimal image thresholding. *Fuzzy algorithms: with application to image processing and pattern recognition*, pages 45–84, 1998.

[3] L. Huang and M. Wang. Image thresholding by minimizing the measure of fuzziness. *Pattern recognition*, 28(1):41–51, 1995.

[4] S. Zenzo, L. Cinque, and S. Levialdi. Image thresholding using fuzzy entropies. *IEEE Transactions on Systems, Man and Cybernetics*, 28(1):15–23, 1998.

[5] M. Forero. *Fuzzy Filters for Image Processing*, chapter Fuzzy thresholding and histogram analysis, pages 129–152. Springer, 2003.

[6] H. Bustince, M. Pagola, P. Melo-Pinto, E. Barrenechea, and P. Couto. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, in Studies in Fuzziness and Soft Computing*, chapter Image threshold computation by modelizing knowledge/unknowledge by means of A-IFSs, pages 627–644. Springer-Verlag, Berlin Heidelberg New York, 2007.

[7] K. Atanassov. Intuitionistic fuzzy sets. *VII ITKR's Session, Deposed in Central Sci. Techn. Library of Bulg. Acd. of Sci., Sofia*, pages 1684–1697, 1983.

ISBN: 978-989-95079-6-8

[8] K. Atanassov. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 20:87–96, 1986.

[9] P. Burillo and H. Bustince. Construction theorems for intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 84:271–281, 1996.

[10] E. Barrenechea. *Image Thresholding with Interval-valued Fuzzy Sets. Edge Detection. Contrast*. PhD thesis, Universidad Pública de Navarra, 2005.

[11] H. Bustince, E. Barrenechea, and M. Pagola. Relationship between restricted dissimilarity functions, restricted equivalence functions and e_n-functions: Image threshold invariant. *Pattern Recognition Letters*, In Press 2007.

[12] P. Couto. *Image segmentation using Atanassov intuitionistic fuzzy sets*. PhD thesis, Trás-os-Montes e Alto Douro University, Vila Real, Portugal, 2006.

[13] K. Atanassov. *Intuitionistic fuzzy sets. Theory and applications*. Springer-Verlag, Berlin Heidelberg New York, 1 edition, January 1999.

[14] P. Burillo and H. Bustince. Entropy on intuitionistic fuzzy sets and on intervalvalued fuzzy sets. *Fuzzy Sets and Systems*, 78:81–103, 1996.

ISBN: 978-989-95079-6-8

SaM: A Split and Merge Algorithm for Fuzzy Frequent Item Set Mining

Christian Borgelt[1] and Xiaomeng Wang[2]

1. European Center for Soft Computing, c/ Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Spain
2. Otto-von-Guericke-University of Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany
Email: christian.borgelt@softcomputing.es, xwang@iws.cs.uni-magdeburg.de

***Abstract*—*This paper presents SaM, a split and merge algorithm for frequent item set mining. Its distinguishing qualities are an exceptionally simple algorithm and data structure, which not only render it easy to implement, but also convenient to execute on external storage. Furthermore, it can easily be extended to allow for "fuzzy" frequent item set mining in the sense that missing items can be inserted into transactions with a user-specified penalty. In order to demonstrate its performance, we report experiments comparing it with the "fuzzy" frequent item set mining version of RElim (an algorithm we suggested in an earlier paper [15] and improved in the meantime).*

Keywords— data mining, frequent item set mining, fuzzy frequent item set, fault tolerant data mining

1 Introduction

Although frequent item set mining and association rule induction has been a focus of research in data mining for a long time now, leading to well-known algorithms like Apriori [1], Eclat [11] and FP-growth [7], there is still room for improvement. Recent research lines include filtering the found frequent item sets and association rules [16, 17], identifying temporal changes in discovered patterns [3, 4], and mining fault-tolerant or "fuzzy" frequent item sets [6, 10, 15].

In this paper we follow the last of these lines by presenting SaM, a split and merge algorithm for frequent item set mining, which can easily be extended to allow for "fuzzy" mining in the sense that missing items can be inserted into transactions with a user-specified penalty. Other distinguishing qualities of our method are its exceptionally simple processing scheme and data structure, which not only render it very easy to implement, but also convenient to execute on external storage.

2 Frequent Item Set Mining

Frequent item set mining is the following task: we are given a set B of *items*, called the *item base*, and a database T of *transactions*. An item may, for example, represent a product, and the item base may then represent the set of all products offered by a supermarket. The term *item set* refers to any subset of the item base B. Each transaction is an item set and may represent, in the supermarket setting, a set of products that has been bought by a customer. Since several customers may have bought the exact same set of products, the total of all transactions must be represented as a vector or a multiset (or, alternatively, each transaction must be enhanced by a *transaction identifier (tid)*). Note that the item base B is usually not given explicitly, but only implicitly as the union of all transactions. The *support* $s_T(I)$ of an item set $I \subseteq B$ is the number of transactions in the database T it is contained in. Given a user-specified *minimum support* $s_{\min} \in \mathbb{N}$, an item set I is called *frequent* (in T) iff $s_T(I) \geq s_{\min}$. The goal of frequent item set mining is to find all item sets $I \subseteq B$ that are frequent in the database T and thus, in the supermarket setting, to identify all sets of products that are frequently bought together.

A standard approach to find all frequent item sets w.r.t. a given database T and a support threshold $s_{\min}$, is a *depth-first search* in the subset lattice of the item base B. This approach can be seen as a simple *divide-and-conquer* scheme. For a chosen item i, the problem to find all frequent item sets is split into two subproblems: (1) find all frequent item sets containing i and (2) find all frequent item sets *not* containing i. Each subproblem is then further divided based on another item j: find all frequent item sets containing (1.1) both i and j, (1.2) i, but not j, (2.1) j, but not i, (2.2) neither i nor j etc.

All subproblems occurring in this recursion can be defined by a *conditional transaction database* and a *prefix*. The prefix is a set of items that has to be added to all frequent item sets that are discovered in the conditional database. Formally, all subproblems are tuples $S = (C, P)$, where C is a conditional database and $P \subseteq B$ is a prefix. The initial problem, with which the recursion is started, is $S = (T, \emptyset)$, where T is the given transaction database and the prefix is empty.

A subproblem $S_0 = (C_0, P_0)$ is processed as follows: choose an item $i \in B_0$, where B_0 is the set of items occurring in C_0. This choice is arbitrary, but usually follows some predefined order of the items. If $s_{C_0}(i) \geq s_{\min}$, then report the item set $P_0 \cup \{i\}$ as frequent with the support $s_{C_0}(i)$, and form the subproblem $S_1 = (C_1, P_1)$ with $P_1 = P_0 \cup \{i\}$. The conditional database C_1 comprises all transactions in C_0 that contain i, but with this item removed. This implies that all transactions are removed that do not contain any no other item than i. If C_1 is not empty, process S_1 recursively. In any case (that is, regardless of whether $s_{C_0}(i) \geq s_{\min}$ or not), form the subproblem $S_2 = (C_2, P_2)$ with $P_2 = P_0$. The conditional database C_2 comprises all transactions in C_0 (including those that do not contain i), but again with the item i removed. If C_2 is not empty, process S_2 recursively.

This recursive scheme is adopted by Eclat, FP-growth, RElim and several other frequent item set mining algorithms. They differ in how conditional transaction databases are represented: in a *horizontal representation*, a database is stored as a list (or array) of transactions, each of which lists the items contained in it. In a *vertical representation*, the items are first referred to with a list (or array) and for each item the transactions containing it are listed. However, this distinction is not pure, since there are combinations of the two forms of representing a database. Our SaM algorithm is, as far as we know, the first algorithm that is based on the general scheme outlined above and uses a purely horizontal representation.

ISBN: 978-989-95079-6-8

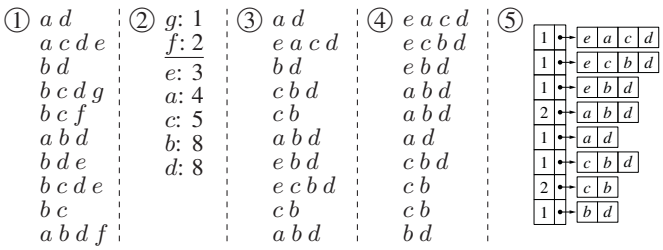


Figure 1: An example database: original form (1), item frequencies (2), transactions with sorted items (3), lexicographically sorted transactions (4), and the used data structure (5).

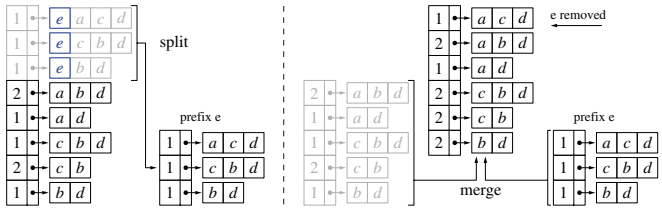


Figure 2: The basic operations: split (left) and merge (right).

The basic processing scheme can be improved with so-called *perfect extension pruning*: an item $i \notin I$ is called a *perfect extension* of an item set I, iff I and $I \cup \{i\}$ have the same support. Perfect extensions have the following properties: (1) if an item i is a perfect extension of an item set I, then it is also a perfect extension of any item set $J \supseteq I$ as long as $i \notin J$ and (2) if K is the set of all perfect extensions of an item set I, then all sets $I \cup J$ with $J \in 2^K$ (where 2^K denotes the power set of K) have the same support as I. These properties can be exploited by collecting in the recursion not only prefix items, but also, in a third element of a subproblem description, perfect extension items. They are also removed from the conditional databases and are only used to generate all supersets of the prefix that have the same support.

3 A Simple Split and Merge Algorithm

In this section we describe the basic form of our SaM (split and merge) algorithm. Preprocessing is very similar to many other frequent item set mining algorithms. The steps are illustrated in Figure 1 for a simple example transaction database: step 1 shows the transaction database in its original form. In step 2 the item frequencies are determined in order to discard infrequent items. With a minimum support of 3, items f and g are infrequent and thus eliminated. In step 3 the (frequent) items in each transaction are sorted according to their frequency, because processing the items in the order of increasing frequency usually leads to the shortest execution times. In step 4 the transactions are sorted lexicographically into descending order, with an item with higher frequency preceding an item with lower frequency. In step 5 the basic data structure is built by combining equal transactions and setting up an array, in which each element consists of two fields: an occurrence counter and a pointer to the sorted transaction.

The basic operations of the recursive processing, which follows the general divide-and-conquer scheme reviewed in Section 2, are illustrated in Figure 2: in the *split step* (left) the given array is split w.r.t. the leading item of the first transaction (item e in our example): all elements referring to transactions starting with this item are transferred to a new array. In this process the pointer (in)to the transaction is advanced by one item, so that the common leading item is "removed" from all transactions. Obviously, this new array represents the conditional database of the first subproblem (see Section 2), which is then processed recursively to find all frequent item sets containing the split item (provided this item is frequent).

The conditional database for frequent item sets *not* containing this item (second subproblem, see Section 2) is obtained with a simple *merge step* (right part of Figure 2). The new array and the rest of the original array are combined with a procedure that is almost identical to one phase of the well-known *mergesort* algorithm. Since both arrays are lexicographically sorted, one merging traversal suffices to create a lexicographically sorted merged array. The only difference to a *mergesort* phase is that equal transactions (or transaction suffixes) are combined: There is always only one instance of each transaction (suffix), while its number of occurrences is kept in a counter. In our example this results in the merged array having two elements less than the input arrays together: the transaction (suffixes) cbd and bd, which occur in both arrays, are combined and their occurrence counters are increased to 2.

Pseudo-code of SaM is shown in Figure 3: a single page of code suffices to describe the whole recursion in detail. The actual C code is even shorter, despite the fact that it contains additional functionality (like perfect extension pruning, Section 2), because certain operations can be written very concisely in C (especially when using pointer arithmetic).

4 Fuzzy Frequent Item Set Mining

There are many applications of frequent item set mining, in which the transactions do not contain all items that are actually present. However, standard algorithms are based on exact matching and therefore are not equipped to meet the needs arising in such applications. An example is the analysis of alarm sequences in telecommunication networks, where a core task is to find collections of alarms occurring frequently together, so-called *episodes*. One approach to accomplish this task is to slide a time window over the alarm sequence. Each window position then captures a specific slice of the alarm sequence [12]. The underlying idea is that in this way the problem of finding frequent episodes is reduced to that of finding frequent item sets in a database of transactions: each alarm can be seen as an item and the alarms in a time window as a transaction. The support of an episode is the number of window positions, so that the episode occurred in the window.

Unfortunately, alarms often get delayed, lost, or repeated due to noise, transmission errors, failing links etc. If alarms do not get through or are delayed, they are missing from the transaction (time window) its associated items (alarms) occur in. If we required exact containment of an item set in this case, the support of some item sets, which would be frequent if the items did not get lost, may be lower than the user-specified minimum. This leads to a possible loss of potentially interesting frequent item sets and to distorted support values.

To cope with such missing information, we rely on the notion of a "fuzzy" or approximate frequent item set. In contrast to research on fuzzy association rules (see, for example, [13]),

ISBN: 978-989-95079-6-8

```
function SaM (a: array of transactions,    (* conditional database *)
              p: set of items,             (* prefix of the cond. database a *)
              s_min: int) : int            (* min. support of an item set *)
var i: item;                               (* buffer for split item *)
    s: int;                                (* support of current split item *)
    n: int;                                (* number of frequent item sets *)
    b, c, d: array of transactions;        (* cond. and merged database *)
begin                                      (* – split and merge recursion – *)
  n := 0;                                  (* init. number of freq. item sets *)
  while a is not empty do                  (* while database is not empty *)
    b := empty; s := 0;                    (* init. split result, item support *)
    i := a[0].items[0];                    (* get the leading item *)
    while a is not empty                   (* of the first transaction and *)
    and   a[0].items[0] = i do             (* split database w.r.t. this item *)
      s := s + a[0].wgt;                   (* sum occurrences (support) *)
      remove i from a[0].items;            (* remove the split item *)
      if    a[0].items is not empty        (* if trans. is not empty *)
      then remove a[0] from a and append it to b;
      else remove a[0] from a; end;        (* move it to cond. db., *)
    end;                                   (* otherwise simply remove it *)
    c := b; d := empty;                    (* initialize the output array *)
    while a and b are both not empty do    (* merge step *)
      if    a[0].items > b[0].items        (* copy trans. from a *)
      then remove a[0] from a and append it to d;
      else if a[0].items < b[0].items      (* copy trans. from b *)
      then remove b[0] from b and append it to d;
      else b[0].wgt := b[0].wgt +a[0].wgt;
           remove b[0] from b and append it to d;
           remove a[0] from a;             (* combine weights and *)
      end;                                 (* move and remove trans.: *)
    end;                                   (* only one instance per trans. *)
    while a is not empty do                (* copy the rest of a *)
      remove a[0] from a and append it to d; end;
    while b is not empty do                (* copy the rest of b *)
      remove b[0] from b and append it to d; end;
    a := d;                                (* loop for second recursion *)
    if s ≥ s_min then                      (* if the split item is frequent: *)
      p := p ∪ {i};                        (* extend the prefix item set and *)
      report p with support s;             (* report the frequent item set *)
      n := n + 1 + SaM(c, p, s_min);
      p := p − {i};                        (* process cond. db. recursively, *)
    end;                                   (* sum the frequent item sets, *)
  end;                                     (* then restore the orig. prefix *)
  return n;                                (* return num. of freq. item sets *)
end;    (* function SaM() *)
```

Figure 3: Pseudo-code of the SaM algorithm.

where a fuzzy approach is used to handle quantitative items, we use the term "fuzzy" to refer to an item set that may not be present exactly in all supporting transactions, but only approximately. Related work in this direction suggested Apriori-like algorithms and mining with approximate matching was performed by counting the number of different items in the two item sets to be compared [6, 10]. However, here we adopt a more general scheme, based on an approximate matching approach exhibiting much greater flexibility. Our approach has two core ingredients: *edit costs* and *transaction weights* [15].

Edit costs: A convenient way of defining the distance between item sets is to consider the costs of a cheapest sequence of edit operations needed to transform one item set into the other [14]. Here we consider only insertions, since they are most easily implemented with our algorithm[1]. With the help of an *insertion cost* or *penalty* a flexible and general framework for approximately matching two item sets can be established. How one interprets such costs or penalties depends, of course, on the application. Note also that different items may be associated with different costs. For example, in telecommunication networks different alarms can have a different probability of getting lost: usually alarms raised in lower levels of the module hierarchy get lost more easily than alarms originating in higher levels. In such cases it is convenient to be able to associate the former with lower insertion costs than the latter. Insertions of a certain item may also be completely inhibited by assigning a very high insertion cost.

Transaction weights: Each transaction t is associated with a weight $w(t)$, the initial value of which is 1. If an item i is inserted into a transaction t, the transaction weight is "penalized" with a cost $c(i)$ associated with the item. Formally, this can be described as applying a combination function: the new weight of the transaction t after inserting an item $i \notin t$ is $w_{\{i\}}(t) = f(w(t), c(i))$, where f is a function that combines the weight $w(t)$ before editing and the insertion cost $c(i)$. The combination function f depends, of course, on the application and may be chosen from a wide range of possible functions. For example, any t-norm may be used. We choose multiplication here, that is, $w_{\{i\}}(t) = w(t) \cdot c(i)$, mainly for reasons of simplicity. Note, however, that with this choice lower values of $c(i)$ mean higher costs as they penalize the weight more, but that it has the advantage that it is easy to extend to inserting several items: $w_{\{i_1,\ldots,i_m\}}(t) = w(t) \cdot \prod_{k=1}^{m} c(i_k)$. It should be clear that it is $w_\emptyset(t) = 1$ due to the initial weighting $w(t) = 1$.

How many insertions into a transaction are allowed may be limited by a user-specified lower bound $w_{\min}$ for the transaction weight. If the weight of a transaction falls below this threshold, it is not considered in further mining steps. Of course, this weight may also be set to zero (unlimited insertions). As a consequence, the *fuzzy support* of an item set I w.r.t. a transaction database T can be defined as $s_T^{(\text{fuzzy})}(I) = \sum_{t \in T} \tau(w_{I-t}(t) \geq w_{\min}) \cdot w_{I-t}(t)$, where $\tau(\phi)$ is a kind of "truth function", which is 1 if ϕ is true and 0 otherwise.

Note that SaM is particularly well suited to handle item insertions, because its purely horizontal transaction representation makes it easy to incorporate transaction weights. With other algorithms, more effort is usually needed in order to extend them to approximate frequent item set mining.

For an implementation, it is beneficial to distinguish between unlimited item insertions ($w_{\min} = 0$) and limited item insertions ($w_{\min} > 0$). If $w_{\min} = 0$, it is possible to combine equal transactions (or transaction suffixes) without restriction: two equal transactions (or suffixes) t_1 and t_2 with weights w_1 and w_2, respectively, can be combined into one transaction (suffix) t with weight $w_1 + w_2$ even if $w_1 \neq w_2$. If another item i needs to be inserted into t_1 and t_2 in order to make them contain a given item set I, the distributive law (that is, $w_1 \cdot c(i) + w_2 \cdot c(i) = (w_1 + w_2) \cdot c(i)$) ensures that we still compute the correct support for the item set I. If, however, we have $w_{\min} > 0$ and, say, $w_1 > w_2$, then using $(w_1 + w_2) \cdot c(i)$ as the contribution of the combined transaction t to the support of the item set I may be wrong, because it may be that $w_1 \cdot c(i) \geq w_{\min}$, but $w_2 \cdot c(i) < w_{\min}$. Then the support contributed by the two transactions t_1 and t_2 would rather be $w_1 \cdot c(i)$. Effectively, transaction t_2 does not contribute, since its weight has fallen below the transaction weight threshold. Hence, with limited insertions, we may combine equal transactions (or suffixes) only if they have the same weight.

[1] Note that deletions are implicit in the mining process (as we search for *subsets* of the transactions). Only replacements are an additional case we do not consider here.

ISBN: 978-989-95079-6-8

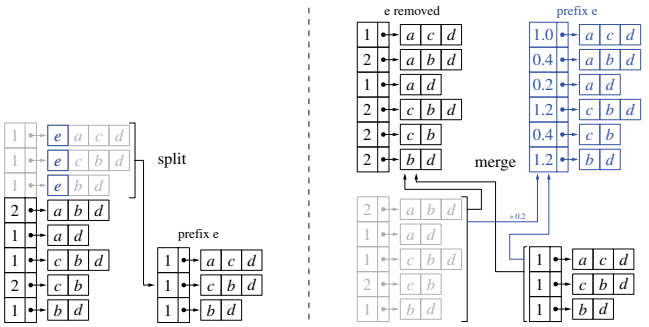


Figure 4: Unlimited item insertions, first recursion level.

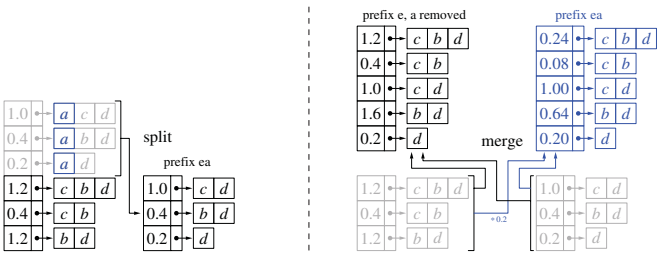


Figure 5: Unlimited item insertions, second recursion level.

5 Unlimited Item Insertions

If unlimited item insertions are possible ($w_{\min} = 0$), only a minor change of the data structure is needed: the integer occurrence counter for the transactions (or suffixes) has to be replaced by a real-valued transaction weight. In the processing, the split step stays the same (see Figure 4 on the left), but now it only yields an intermediate database with all transactions (or suffixes) that actually contain the split item under consideration (item e in the example). In order to form the full conditional database, we have to add those transactions that do not contain the split item, but can be made to contain it by inserting it. This is achieved in the merge step, in which two parallel merge operations are carried out now (see Figure 4 on the right). The first part (shown in black) is the merge that yields the conditional database for frequent item sets *not* containing the split item. The second part (shown in blue/grey) adds those transactions that do not contain the split item, weighted down with the insertion penalty, to the intermediate database created in the split step. Of course, this second part of the merge operation is only carried out, if $c(i) > 0$, where i is the split item, because otherwise no support would be contributed by the transactions not containing the item i and hence it would not be necessary to add them. In such a case the result of the split step would already yield the conditional database for frequent item sets containing the split item.

Note that in both merge operations equal transactions (or suffixes) can be combined regardless of their weights. As a consequence we have in Figure 4 entries like for the transaction (suffix) cbd, with a weight of 1.2, which stands for one occurrence with weight 1 and one occurrence with weight 0.2 (due to the penalty factor 0.2, needed due to the insertion of item e). As an additional illustration, Figure 5 shows the split and merge operations for the second recursion level.

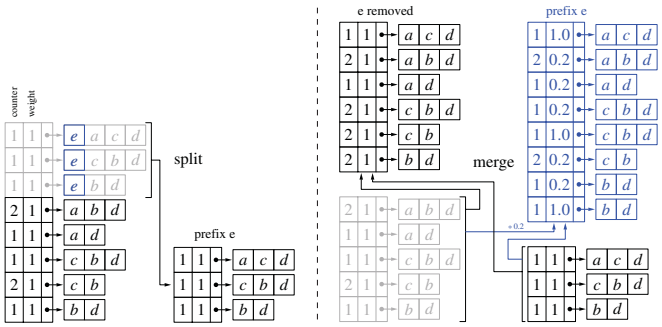


Figure 6: Limited item insertions, first recursion level.

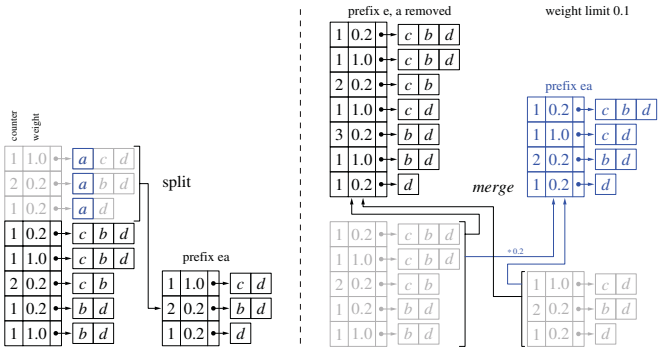


Figure 7: Limited item insertions, second recursion level.

6 Limited Item Insertions

If item insertions are limited by a transaction weight threshold ($w_{\min} > 0$), the transaction weight has to be represented explicitly and kept separate from the number of occurrences. Therefore the data structure must comprise, per transaction (suffix), (1) a pointer to the item array, (2) an integer occurrence counter, and (3) a real-valued transaction weight. The last field will be subject to a thresholding operation by $w_{\min}$, which eliminates all transactions with a weight less than $w_{\min}$. Hence there may now be array elements that refer to the same transaction (suffix), but differ in the transaction weight.

The processing scheme is illustrated in Figure 6. The split step is still essentially the same. However, the merge step differs due to the fact that equal transactions (or suffixes) can no longer be combined if their weight differs. As a consequence, there are now, in the result of the second merge operation (shown in blue) two array elements for cbd and two for bd, which carry different weights. This is necessary, because they may reach, due to item insertions, the transaction weight threshold at different times and thus cannot be combined.

That transactions are discarded due to the weight threshold rarely happens on the first level of the recursion. (This can occur only if the insertion penalty factor of the split item is smaller than the transaction weight threshold, which is equivalent to inhibiting insertions of this item altogether). Therefore, in order to illustrate how transactions are discarded, Figure 7 shows the second recursion level, where the conditional database with prefix e is processed. Here the second merge operation actually discards transactions if we set a transaction weight limit of 0.1: all transactions, which need two items (namely both e and a) to be inserted, are not copied.

ISBN: 978-989-95079-6-8

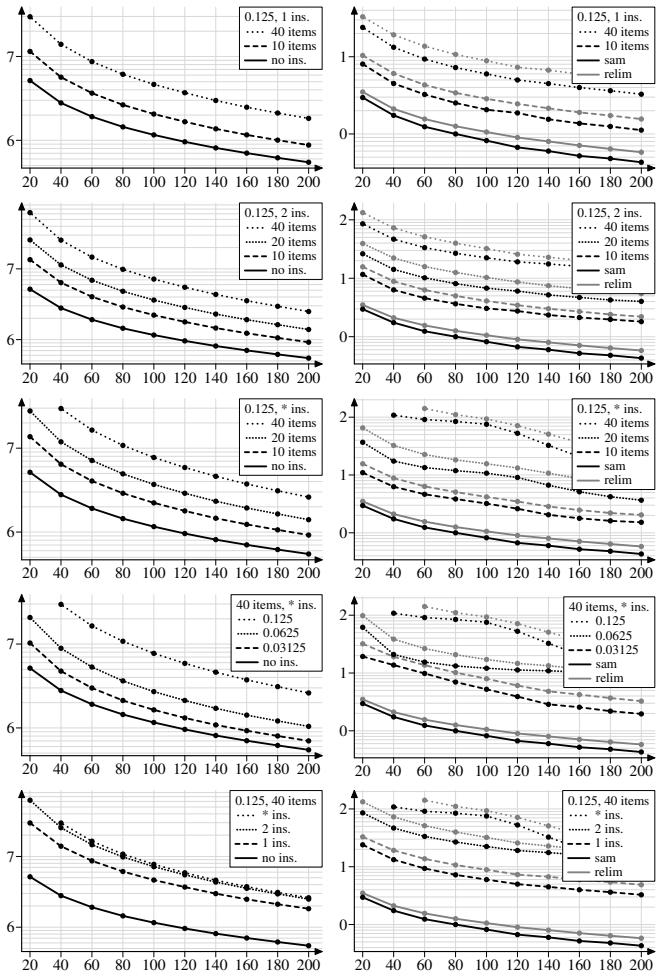


Figure 8: Experimental results on the Census (Adult) data; left: frequent item sets, right: execution times.

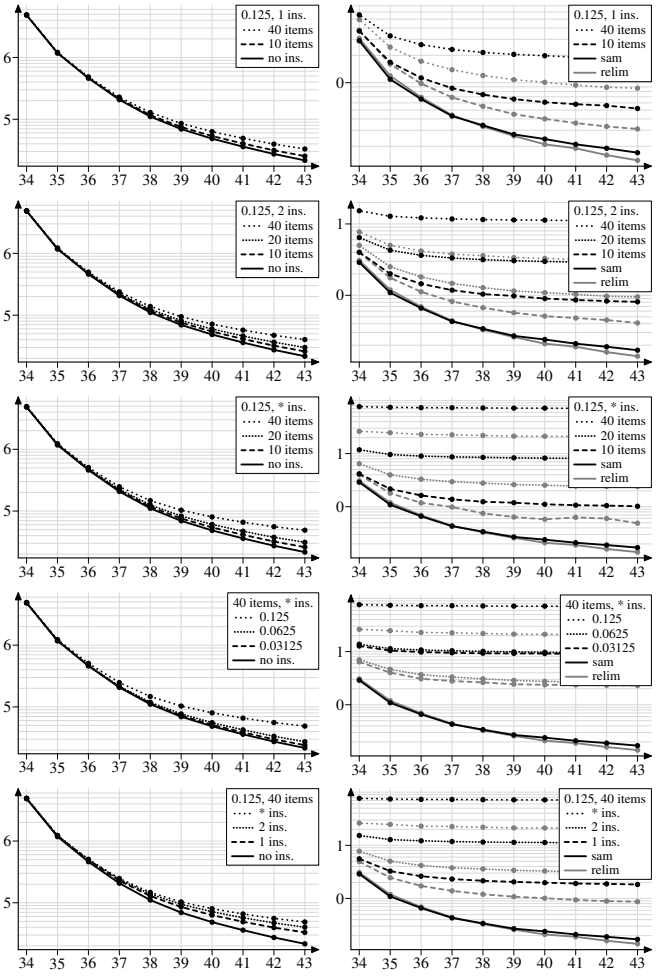


Figure 9: Experimental results on the BMS-Webview-1 data; left: frequent item sets, right: execution times.

7 Experiments

We ran experiments on several data sets, of which we chose two for this paper: Census (aka Adult, a data set derived from an extract of the US census bureau data of 1994, which was preprocessed by discretizing numeric attributes [2]) and BMS-Webview-1 (a web click stream from a leg-care company that no longer exists, which has been used in the KDD cup 2000 [8]). We chose these data sets, because Census is rather dense (a rather large fraction of all items occur in each transaction), while BMS-Webview-1 is rather sparse, and SaM and RElim [15] (the two algorithms of which we have implementations that can find approximate frequent item sets) exhibit a significantly different behavior on dense and sparse data sets.

The results are shown in Figure 8 for the census data set and in Figure 9 for the BMS-Webview-1 data set. In both figures the diagrams on the left show the decimal logarithm of the number of found frequent item sets, while the diagrams on the right show the decimal logarithm of the execution times (in seconds) for our implementations of SaM and RElim.[2] We tested insertion penalty factors of $\frac{1}{8} = 0.125$, $\frac{1}{16} = 0.0625$, and $\frac{1}{32} = 0.03125$, non-vanishing insertion penalty factors for 10, 20, and 40 items, and transaction weight thresholds that allowed for 1, 2 or an unlimited number of insertions.

As can be seen from the diagrams on the left of each figure, the two data sets react very differently to the possibility of inserting items into transactions. While the number of found frequent item sets rises steeply with all parameters for Census, it rises only very moderately for BMS-Webview-1, with the factor even leveling off for lower support values. As it seems, this effect is due, to a large degree, to the sparseness of BMS-Webview-1 (this still needs closer examination, though).

SaM fares considerably better on the dense data set (Census), beating RElim by basically the same margin (factor) in all parameter settings, while SaM is clearly outperformed by RElim on the sparse data set (BMS-Webview-1), even though the two algorithms are actually on par without item insertion (solid lines). On both data sets, the number of insertions that are allowed has the strongest influence: with two insertions execution times are about an order of magnitude larger than with only one insertion. However, the possibility to combine equal transactions with different weights still seems to keep the execution times for unlimited insertions within limits.

The number of items with a non-vanishing penalty factor and the value of the penalty factor itself seem to have a similar influence: doubling the number of items leads to roughly the same effect as keeping the number the same and doubling the

[2]Execution times were measured on an Intel Core 2 Quad Q9300 machine with 3 GB of main memory running openSuSE Linux 11.0 (32 bit) and gcc version 4.3.1.

ISBN: 978-989-95079-6-8

penalty factor. This is plausible, since there should not be much difference in having the possibility to insert twice the number of items or preserving twice the transaction weight per item insertion. Note, however, that doubling the penalty factor from from $\frac{1}{32}$ to $\frac{1}{16}$ has only a comparatively small effect on the BMS-Webview-1 data compared to doubling from $\frac{1}{16}$ to $\frac{1}{8}$. On the census data set the effects are a bit more in line.

Overall it should be noted that the execution times, although considerably increased over those obtained without item insertions, still remain within acceptable limits. Even with 40 items having an insertion penalty factor of $\frac{1}{8}$ and unlimited insertions, few execution times exceed 180 seconds ($\log_{10}(180) \approx 2.25$). In addition, we can observe the interesting effect on the BMS-Webview-1 data that at the highest parameter settings the execution times become almost independent of the minimum support threshold.

8 Conclusions

In this paper we presented a very simple split and merge algorithm for frequent item set mining, which, due to the fact that it uses a purely horizontal transaction representation, lends itself well to an extension to "fuzzy" or approximate frequent item set mining. In addition, it is a highly recommendable method if the data to mine cannot be loaded into main memory and thus the data has to be processed on external storage or in a (relational) database system. As our experimental results show, the SaM algorithm performs the task of "fuzzy" frequent item set mining excellently on the dense Census data, but shows certain weaknesses on the sparse BMS-Webview-1 data. However, our experiments provided some evidence (to be substantiated on more data sets) that "fuzzy" frequent item set mining is much more useful for dense data sets as more additional frequent item sets can be found. Hence SaM performs better in the (likely) more relevant case. Most importantly, however, one should note that with both SaM and RElim the execution times remain bearable.

Software: SaM and RElim sources in C can be found at:

http://www.borgelt.net/sam.html
http://www.borgelt.net/relim.html

References

[1] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules. *Proc. 20th Int. Conf. on very Large Databases (VLDB 1994, Santiago de Chile)*, 487–499. Morgan Kaufmann, San Mateo, CA, USA 1994

[2] C.L. Blake and C.J. Merz. *UCI Repository of Machine Learning Databases*. Dept. of Information and Computer Science, University of California at Irvine, CA, USA 1998

[3] M. Böttcher, M. Spott and D. Nauck. Detecting Temporally Redundant Association Rules. *Proc. 4th Int. Conf. on Machine Learning and Applications (ICMLA 2005, Los Angeles, CA)*, 397–403. IEEE Press, Piscataway, NJ, USA 2005

[4] M. Böttcher, M. Spott and D. Nauck. A Framework for Discovering and Analyzing Changing Customer Segments. *Advances in Data Mining — Theoretical Aspects and Applications (LNCS 4597)*, 255–268. Springer, Berlin, Germany 2007

[5] C. Borgelt. Keeping Things Simple: Finding Frequent Item Sets by Recursive Elimination. *Proc. Workshop Open Software for Data Mining (OSDM'05 at KDD'05, Chicago, IL)*, 66–70. ACM Press, New York, NY, USA 2005

[6] Y. Cheng, U. Fayyad, and P.S. Bradley. Efficient Discovery of Error-Tolerant Frequent Itemsets in High Dimensions. *Proc. 7th Int. Conf. on Knowledge Discovery and Data Mining (KDD'01, San Francisco, CA)*, 194–203. ACM Press, New York, NY, USA 2001

[7] J. Han, H. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. *Proc. Conf. on the Management of Data (SIGMOD'00, Dallas, TX)*, 1–12. ACM Press, New York, NY, USA 2000

[8] R. Kohavi, C.E. Bradley, B. Frasca, L. Mason, and Z. Zheng. KDD-Cup 2000 Organizers' Report: Peeling the Onion. SIGKDD Exploration 2(2):86–93. ACM Press, New York, NY, USA 2000

[9] J. Pei, J. Han, B. Mortazavi-Asl, and H. Zhu. Mining Access Patterns Efficiently from Web Logs. *Proc. Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'00, Kyoto, Japan)*, 396–407. 2000

[10] J. Pei, A.K.H. Tung, and J. Han. Fault-Tolerant Frequent Pattern Mining: Problems and Challenges. *Proc. ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMK'01, Santa Babara, CA)*. ACM Press, New York, NY, USA 2001

[11] M. Zaki, S. Parthasarathy, M. Ogihara, and W. Li. New Algorithms for Fast Discovery of Association Rules. *Proc. 3rd Int. Conf. on Knowledge Discovery and Data Mining (KDD'97, Newport Beach, CA)*, 283–296. AAAI Press, Menlo Park, CA, USA 1997

[12] H. Mannila, H. Toivonen, and A.I. Verkamo. *Discovery of Frequent Episodes in Event Sequences*. Report C-1997-15, University of Helsinki, Finland 1997

[13] C. Kuok, A. Fu, and M. Wong. Mining Fuzzy Association Rules in Databases. *SIGMOD Record* 27(1):41–46. 1998

[14] P. Moen. *Attribute, Event Sequence, and Event Type Similarity Notions for Data Mining*. Ph.D. Thesis/Report A-2000-1, Department of Computer Science, University of Helsinki, Finland 2000

[15] X. Wang, C. Borgelt, and R. Kruse. Mining Fuzzy Frequent Item Sets. *Proc. 11th Int. Fuzzy Systems Association World Congress (IFSA'05, Beijing, China)*, 528–533. Tsinghua University Press and Springer-Verlag, Beijing, China, and Heidelberg, Germany 2005

[16] G.I. Webb and S. Zhang. k-Optimal-Rule-Discovery. *Data Mining and Knowledge Discovery* 10(1):39–79. Springer, Amsterdam, Netherlands 2005

[17] G.I. Webb. Discovering Significant Patterns. *Machine Learning* 68(1):1–33. Springer, Amsterdam, Netherlands 2007

ISBN: 978-989-95079-6-8

Online Recognition of Fuzzy Time Series Patterns

Gernot Herbst, Steffen F. Bocklisch

Department of Electrical Engineering, Chemnitz University of Technology
D-09107 Chemnitz, Germany
Email: {gernot.herbst, steffen.bocklisch}@etit.tu-chemnitz.de

Abstract— *This article deals with the recognition of recurring multivariate time series patterns modelled sample-point-wise by parametric fuzzy sets. An efficient classification-based approach for the online recognition of incompleted developing patterns in streaming time series is being presented. Furthermore, means are introduced to enable users of the recognition system to restrict results to certain stages of a pattern's development, e. g. for forecasting purposes, all in a consistently fuzzy manner.*

Keywords— Fuzzy classification, fuzzy automata, multivariate time series, pattern recognition.

1 Introduction

Fuzzy sets have been successfully used for a long time for a model-free or model-based representation of time series or time series sequences, e. g. for the classification and prediction of nonlinear systems behaviour [1–4]. This article aims at extending the fuzzy time series classification approach of [2, 3] to an *always aware* online recognition system for nonstationary recurring patterns in multivariate streaming time series. *Always aware* refers to the ability of the recognition system presented in this paper to constantly expect to find an (incompleted) pattern in a time series in any of its possible stages of development, which will be achieved by classification of pattern subsequences. It will be shown how the computationally expensive classification results for all possible subsequences can be obtained at almost no additional cost compared to the classification of completed patterns by relating the recognition procedure to ideas from fuzzy automata.

2 Fuzzy Modelling of Time Series Patterns

2.1 A Multivariate Parametric Fuzzy Set

In this section a parametric membership function type will be presented, which will serve as the common basis for the classifiers developed in this article. For the univariate case, the membership function is defined in (1). Based upon a potential function, it has already been successfully used in numerous applications such as process or medical surveillance [1], time series modelling and prediction [2, 3] and classifier networks [5]. One important property of this function is its capability of modelling asymmetric fuzzy sets by individual parameters for the left- and right-hand function branches.

$$\mu(x) = \begin{cases} \dfrac{a}{1+\left(\dfrac{1}{b_l}-1\right)\left(\dfrac{r-x}{c_l}\right)^{d_l}} & : \quad x < r \\ \dfrac{a}{1+\left(\dfrac{1}{b_r}-1\right)\left(\dfrac{x-r}{c_r}\right)^{d_r}} & : \quad x \geq r \end{cases} \tag{1}$$

The maximum truth value a of the fuzzy set occurs at its modal point: $\mu(x = r) = a$. For normalised fuzzy sets, $a = 1$ holds, as will be the case throughout this paper. The six parameters $b_{l/r}$, $c_{l/r}$ and $d_{l/r}$ determine the extent and shape of the class. From Fig. 1, the effect of b ($0 < b < 1$) and c ($c > 0$) can be understood. Semantically, they correspond to the class borders. The d parameter ($d \geq 2$) influences the shape of μ's descent to zero, with increasing d leading to a sharper descent and $d \to \infty$ resulting in a rectangular shape. This is particularly interesting since crisp sets may be represented by (1) as well.

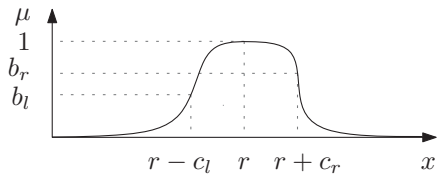


Figure 1: Normalised parametric membership function (1).

The parameters of this membership functions may be chosen based upon expert knowledge on the one hand, and automatically be computed [1] from a set of learning data on the other hand.[1] Both options benefit from the interpretability of these parameters.

A probably unique advantage of this function concept is that a multivariate membership function—based upon a conjunction of one-dimensional sets using a compensatory n-fold Hamacher intersection—may also be obtained in *parametric* form [1]. The n-fold operator is given by (2). The resulting multidimensional membership function is, for a simplified case of symmetric left- and right-hand branches, given by (3). Additionally (and omitted here for brevity), the class space may optionally be rotated in the underlying feature space for each class individually, leading to $(n-1)$ further rotation angle parameters, which can be obtained from a principal component analysis of the learning data. Fig. 2 depicts two examples of membership functions in a two-dimensional feature space.

$$\cap^{n}_{Ham}\,\mu_i = \frac{1}{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{1}{\mu_i}} \tag{2}$$

[1] The parameters of (1) are computed from data as follows: Firstly, the modal point r is being calculated by averaging the object values. The left- and right-hand side extents $c_{l/r}$ are chosen such that all learning objects, augmented by a so-called elementary fuzziness, lie within these class borders. Finally, the shape parameters b and d parameters are—if not specified manually—being computed to best fit the dispersion of the data.

ISBN: 978-989-95079-6-8

$$\mu(\boldsymbol{x}) = \frac{a}{1 + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{b_i} - 1\right) \cdot \left|\frac{x_i - r_i}{c_i}\right|^{d_i}} \tag{3}$$

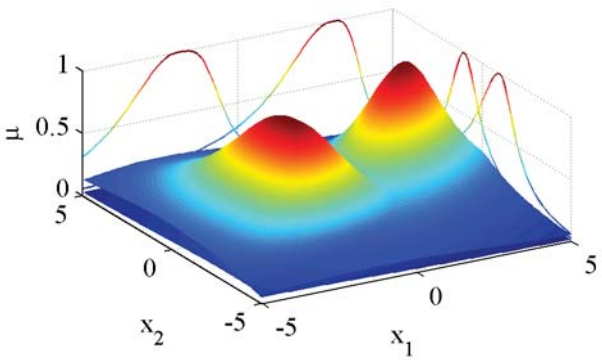


Figure 2: Two examples of two-dimensional fuzzy classes.

2.2 A Fuzzy Model for Time Series Patterns

The membership function concept introduced in section 2.1 forms the basis of a classifier model for multivariate time series sequences [2]. It allows to compute a continuous degree of similarity $\mu \in [0, 1]$ between a measured time series pattern and the fuzzy pattern model. Therefore pattern matching is being treated as a fuzzy classification task. The model described hereafter may be employed for equidistantly sampled univariate or multivariate time series.

Each point $\boldsymbol{x}(i)$ of a time series pattern (length: L sample points) is being described by a membership function $\mu_{P,i}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^n$, $i = 1, \ldots, L$. This model therefore exhibits a soft tolerance towards noise and other sources of imprecision. To compare a time series sequence against the model, i. e. to classify a pattern given by L points $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(L)$, the individual classification results of each point are combined to an overall degree of similarity $\mu \in [0, 1]$ by means of a fuzzy conjunction:

$$\mu = \mu_{P,1}(\boldsymbol{x}(1)) \cap \cdots \cap \mu_{P,L}(\boldsymbol{x}(L)) \tag{4}$$

If we recall the ideas of section 2.1 which led to the multivariate parametric membership function, it appears reasonable to reuse the same conjunction operator given by (2). In that manner, the overall classifier for the entire time series pattern could also be thought of as *one* fuzzy class created from the conjunction of $(L \cdot n)$ univariate fuzzy sets, forming an $(L \cdot n)$-dimensional membership function. Thusly classification of time series sequences would very well remain in line with the multivariate classification paradigm behind section 2.1. One key advantage of this model—especially in comparison to many *black box* models such as neural networks—is its support for partial classification, i. e. for the classification of any possible subsequence, simply by intersecting the underlying classification results for the sample points available. We will make use of this advantageous property later on.

Fig. 3 depicts an example of an univariate ($n = 1$) pattern (length $L = 286$ sample points) described by this fuzzy time series model, learned from a set of 14 instances of this pattern (*Coffee* dataset, available from [6]). By means of this figure we may also come to the interpretation of the pattern classifier as a corridor with soft boundaries. A measured pattern will, depending on its degree of similarity, more or less follow this corridor. The properties of the underlying membership function ensure that no pattern will be absolutely rejected ($\mu = 0$) by this classifier, thusly enabling a smooth, non-switching behaviour of the recognition system.

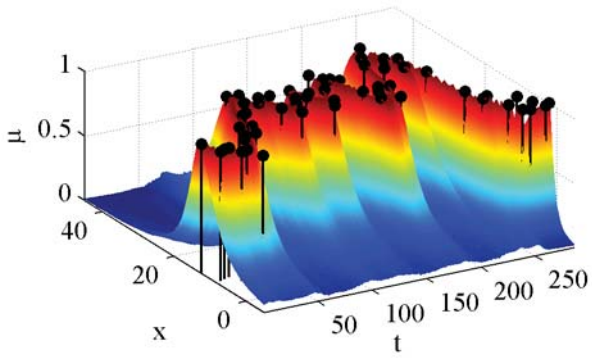


Figure 3: Fuzzy time series model along with a noisy pattern.

While the presented classifier aims at the recognition of recurring patterns that are similar in terms of absolute values, variations such as offsets or scaling may, at least to a certain extent, be covered by suitable preprocessing steps like normalisation. As in section 2.1, the time series pattern model may either be formulated based upon expert knowledge—eased by the interpretability of the individual parameters—or be result of a learning step using a set of pattern instances of the same phenomenon, employing the learning algorithm from section 2.1 for each point of the pattern. The latter is especially interesting in conjunction with clustering or motif mining algorithms. As there is no transform involved, the original properties of the pattern ensemble (energy content, statistics etc.) are mostly being preserved by this model. Furthermore, short-term forecasting of a time series with a partially elapsed pattern is made possible in a straightforward manner based upon the knowledge represented by the model.

3 Online Pattern Recognition

3.1 Task Description

The task of online pattern recognition—as we would like to understand it in the context of this work—can be described by means of Fig. 4. In a streaming univariate or multivariate time series $\boldsymbol{x}(t)$, we are looking for instances of a known pattern. *Online* pattern recognition firstly implies that an occurrence of the pattern has to be permanently expected. But in addition to that, we aim at detecting incompleted patterns, which would enable short-term predictions of the time series based on known patterns. For applications such as machine diagnosis, for instance, this facilitates preventive maintenance actions before a minor damage develops into a severe one.

At the current point in time t_{now} (as sketched in Fig. 4), the model containing the desired pattern has to be adjusted in its relative position τ such that the subsequence $\boldsymbol{x}([t_{\text{now}} - \tau] \ldots t_{\text{now}})$ optimally matches the partial model. τ then represents the time already elapsed in the pattern.

Summarisingly, an online pattern recognition system has to check if a pattern in a time series can be found at the current point in time, and if so, at which stage of development τ.

ISBN: 978-989-95079-6-8

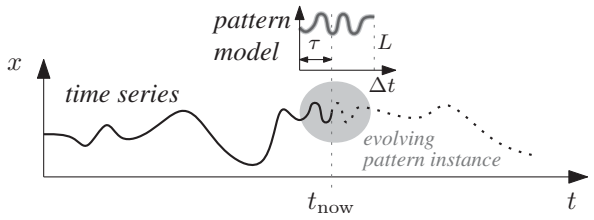


Figure 4: Online matching of evolving patterns.

3.2 Fuzzy Online Recognition

For reasons outlined above, we would like to treat the pattern matching problem as a fuzzy classification task, with time series subsequences (length τ, $\tau \leq L$) being input to a suitable classifier representing a pattern of length L. For each point in time t, this would yield classification results $\mu(t,\tau)$ for every possible value of τ. Semantically, $\mu(t_{\text{now}},\tau)$ represents the similarity of the time series subsequence $\boldsymbol{x}([t_{\text{now}}-\tau]\ldots t_{\text{now}})$ to the first part (length τ) of the pattern model/classifier.

In a truly fuzzy sense[2], *each* pattern would be found at *every* point in time in *all* possible stages of development τ, expressed by non-zero classification results $\mu(t,\tau) > 0\ \forall\tau$, even though $\mu(t,\tau)$ may, of course, be negligibly small. At any point in time t, fuzzy classification of evolving time series patterns therefore results in a *fuzzy set* $\mu(t,\tau)$—instead of just *one* truth value $\mu(t)$, as in a "conventional" fuzzy classification task, cf. Fig. 5. Representing the recognition result as a fuzzy set reflects the fuzziness of the recognition task. It is a joint fuzzy representation of the following information:

1. Could the pattern be found at the current point in time?
2. What is the pattern's state of development τ?

Both questions can—in real-world time series—not be answered precisely and unambigously. $\mu(t,\tau)$ commensurates with this vagueness. Furthermore, it may be interpreted as a fuzzy specification of τ, resembling the concept of fuzzy numbers.[3]

3.3 Representation as Fuzzy Automaton

In section 2.2, a suitable classifier model for patterns in equidistantly sampled time series has been introduced. As this model supports partial classification, it is well suited for the online pattern matching problem. In an online recognition system, however, obtaining $\mu(t,\tau)$ for every possible value of τ by classifying all respective subsequences $\boldsymbol{x}([t-\tau]\ldots t)$ would be a rather time-consuming procedure. More specifically, it appears incommensurate to recompute the whole fuzzy set $\mu(t,\tau)$ if only one new datum $\boldsymbol{x}(t)$ is being presented to the recognition system at time t. We therefore would like to reuse the classification result $\mu(t-1,\tau-1)$ of the subsequence

[2] As an unofficial definition, we would like to understand "true" fuzziness as a property of models and methods which do not (explicitely or implicitely) perform crisp decisions or cut-off operations. For instance, a "truly" fuzzy set has infinite support, thusly all elements x of the underlying universe of discourse belong to this set with $\mu(x) > 0$. As a counterexample, triangular fuzzy sets feature crisp and finite extents, and would therefore not be regarded as "truly" fuzzy models.

[3] Although not being a fuzzy number in the sense of Dubois and Prade [7], especially owing to the very likely multimodality, it matches the spirit of a fuzzy number as a fuzzified representation of a real-valued number (here: τ) quite well.

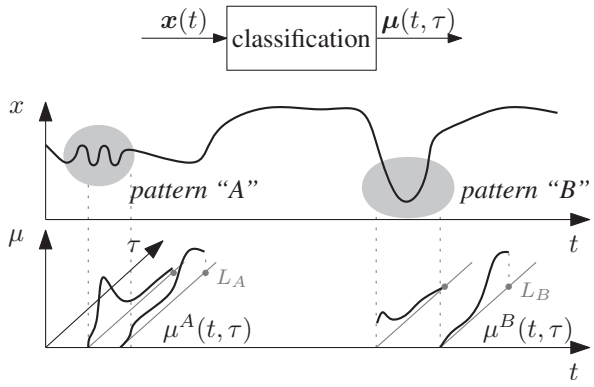


Figure 5: Fuzzy online recognition of two patterns "A" and "B" (length L_A and L_B). At each point in time t, the maximum truth values of the recognition results $\mu^{A/B}(t,\tau)$ w. r. t. τ point to the relative positions τ associated to the best match of the partially elapsed patterns. The fuzzy sets $\mu^{A/B}$ are shown at two different stages of development τ for each pattern.

$\boldsymbol{x}([(t-1)-(\tau-1)]\ldots t-1)$ when computing $\mu(t,\tau)$ for $\boldsymbol{x}([t-\tau]\ldots t)$.

For discrete time patterns of length L, the pattern recognition result $\mu(t,\tau)$ consists of L truth values ($\tau = 1,\ldots,L$). This result may be represented by means of a fuzzy state automaton [8], as sketched in Fig. 6. Being a fuzzy automaton, each state—corresponding to a certain value of τ—is active to a certain degree $\mu(t,\tau)$. At each point in time t, the automaton is triggered with a discrete event in the form of a new input value $\boldsymbol{x}(t)$, leading to new activation levels for each state. An activation level $\mu(t,\tau)$ depends on the current input value $\boldsymbol{x}(t)$ and the former activation level of the previous state:

$$\mu(t,\tau) = f(\mu(t-1,\tau-1),\boldsymbol{x}(t)) \tag{5}$$

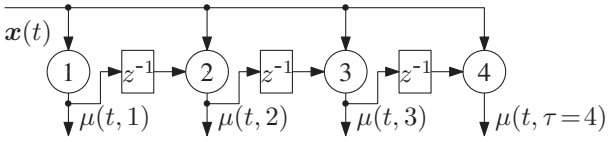


Figure 6: (Informal) fuzzy state automaton representation of a time series pattern recognition system, here for a pattern consisting of four sample points.

Recalling the internals of the time series pattern classifier from section 2.2, the overall classification result was gained from a compensatory Hamacher conjunction (cf. (4)) of individual classification results $\mu_{P,i}$ of time series sample points $\boldsymbol{x}_i$. Thusly $\mu(t-1,\tau-1)$ consists of a conjunction of $\tau-1$ truth values. $\mu(t,\tau)$ could therefore be obtained from a conjunction of $\mu(t-1,\tau-1)$ and the partial classification result $\mu_{P,\tau}(\boldsymbol{x}(t))$ for the τ^{th} point of the time series pattern. However, this conjunction should reflect the weight of the $\tau-1$ truth values that contributed to $\mu(t-1,\tau-1)$, so that $\mu_{P,\tau}(\boldsymbol{x}(t))$ does not outweigh all previous information. A suitable conjunction operator will be derived in the following section, which will enable us to rewrite (5) to such a weighted conjunction:

$$\mu(t,\tau) = \mu(t-1,\tau-1)\ {}^{(\tau-1)}\cap^{1}\ \mu_{P,\tau}(\boldsymbol{x}(t)) \tag{6}$$

ISBN: 978-989-95079-6-8

Equation (6) forms the basis for the mode of operation of the fuzzy automaton from Fig. 6, and will enable us to perform online recognition of time series patterns much more efficiently.

3.4 Weighted Compensatory Hamacher Conjunction

The compensatory Hamacher conjunction of n truth values μ_i $(i = 1, \ldots, n)$ has already been introduced in (2). By rearranging the components of (2) into two sets of truth values, we may formulate the compensatory Hamacher conjunction of $n = n_1 + n_2$ truth values as:

$$\begin{aligned}\mu_1 \cap \cdots \cap \mu_n &= \frac{1}{\frac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1}\frac{1}{\mu_i} + \sum_{j=n_1+1}^{n}\frac{1}{\mu_j}\right)} \\ &= \frac{1}{\frac{1}{n_1+n_2}\left(\frac{n_1}{\mu_1 \cap \cdots \cap \mu_{n_1}} + \frac{n_2}{\mu_{n_1+1} \cap \cdots \cap \mu_n}\right)} \\ &= (\mu_1 \cap \cdots \cap \mu_{n_1})^{\,n_1 \cap^{n_2}}\ (\mu_{n_1+1} \cap \cdots \cap \mu_n)\end{aligned} \tag{7}$$

Based upon (7) we may now define the (n_a, n_b)-weighted compensatory Hamacher conjunction of two truth values μ_a and μ_b, which was in a similar manner also derived in [9]:

$$\mu_a \,^{n_a \cap^{n_b}}\ \mu_b = \frac{1}{\frac{1}{n_a+n_b}\left(\frac{n_a}{\mu_a} + \frac{n_b}{\mu_b}\right)} \tag{8}$$

Thusly it becomes possible to compute the conjunction of new truth values with existing conjunction results without having to store the individual truth values leading to the existing results and while exactly retaining the weight of each truth value that contributed to the overall result.

Finally, we may define the vectorial weighted conjunction of two equal-sized truth vectors $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_b$ using weight vectors $\boldsymbol{n}_a$ and $\boldsymbol{n}_b$ containing the individual weights for the respective elements:

$$\boldsymbol{\mu}_a \,^{\mathbf{n_a} \cap^{\mathbf{n_b}}}\ \boldsymbol{\mu}_b = \begin{pmatrix} \vdots \\ \dfrac{1}{\frac{1}{n_{a,i}+n_{b,i}}\left(\frac{n_{a,i}}{\mu_{a,i}} + \frac{n_{b,i}}{\mu_{b,i}}\right)} \\ \vdots \end{pmatrix} \tag{9}$$

3.5 Update Equations for the Fuzzy Automaton

Recalling (6) and employing the weighted conjunction operator from (8), we may now formulate the update equation for each fuzzy state of the automaton representation of the recognition results $\mu(t, \tau)$ from section 3.3. Equation (6) holds for $\tau \geq 2$, as the first state of the automaton does not have a predecessor state to obtain $\mu(t-1, \tau-1)$. Since the weight of the non-existing zero-th state would be zero anyway, (6) can (for $\tau = 1$) be thought of to reduce to solely the classification result for the first sample point, cf. (10). All further results $\mu(t, \tau \geq 2)$ are computed by a weighted conjunction of antecedent recognition results and the classification results $\mu_{P,\tau}$ of the new sample point $\boldsymbol{x}(t)$ available at each point in time.

$$\mu(t, 1) = \mu_{P,1}(\boldsymbol{x}(t)) \tag{10}$$

To obtain a vectorial form of the update equations, we define $\boldsymbol{\mu}_P(t)$ as the vector containing the individual membership values of $\boldsymbol{x}(t)$ to all classes constituting the fuzzy time series pattern model, and $\boldsymbol{\mu}_\tau(t)$ as the state vector containing all levels of activation of the automaton, and thusly the overall pattern recognition results:

$$\boldsymbol{\mu}_P(\boldsymbol{x}(t)) = \begin{pmatrix} \mu_{P,1}(\boldsymbol{x}(t)) \\ \vdots \\ \mu_{P,L}(\boldsymbol{x}(t)) \end{pmatrix} \tag{11}$$

$$\boldsymbol{\mu}_\tau(t) = \begin{pmatrix} \mu_{\tau,1}(t) \\ \vdots \\ \mu_{\tau,L}(t) \end{pmatrix} = \begin{pmatrix} \mu(\tau, 1) \\ \vdots \\ \mu(\tau, L) \end{pmatrix} \tag{12}$$

Summarisingly, the recursive update equation for the recognition result (the fuzzy set $\mu(t, \tau)$) or its vectorial representation $\boldsymbol{\mu}_\tau(t)$ can be condensed from (5), (6) and (9) to (13):

$$\begin{aligned}\boldsymbol{\mu}_\tau(t) &= f(\boldsymbol{\mu}_\tau(t-1)), \boldsymbol{\mu}_P(\boldsymbol{x}(t))) \\ &= \left(\begin{pmatrix}1\\ \mathbf{0}\end{pmatrix} + \begin{pmatrix}\mathbf{0}^T & 0\\ \boldsymbol{I} & \mathbf{0}\end{pmatrix} \cdot \boldsymbol{\mu}_\tau(t-1)\right)^{\,\mathbf{n}_\tau \cap \mathbf{1}}\ \boldsymbol{\mu}_P(\boldsymbol{x}(t))\end{aligned} \tag{13}$$

$$\boldsymbol{n}_\tau = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ L-1 \end{pmatrix} \tag{14}$$

Incorporating the weight vector $\boldsymbol{n}_\tau$, this finally results in:

$$\boldsymbol{\mu}_\tau(t) = \begin{pmatrix} \dfrac{1}{\frac{1}{1}\cdot\left(\frac{0}{1} + \frac{1}{\mu_{P,1}(\boldsymbol{x}(t))}\right)} \\ \vdots \\ 1 \\ \dfrac{1}{\frac{1}{i}\cdot\left(\frac{i-1}{\mu_{\tau,i-1}(t-1)} + \frac{1}{\mu_{P,i}(\boldsymbol{x}(t))}\right)} \\ \vdots \\ 1 \\ \dfrac{1}{\frac{1}{L}\cdot\left(\frac{L-1}{\mu_{\tau,L-1}(t-1)} + \frac{1}{\mu_{P,L}(\boldsymbol{x}(t))}\right)} \end{pmatrix} \tag{15}$$

Apart from the $(L-1)$ weighted conjunctions of truth values that are required at each time step, the online recognition of a (partially elapsed) fuzzy pattern therefore does not require more underlying classification steps and computations of $\mu_{P,\tau}$ than would be needed just for the fuzzy classification and recognition of only completed patterns in streaming time series.

4 Post-Processing Recognition Results

4.1 Windowing of $\mu(t, \tau)$

The fuzzy set $\mu(t, \tau)$—describing the recognition results of a pattern found at the current point in time t in all possible stages of development τ—is a rather complex result containing a wealth of information. In many cases it will thusly have to be narrowed down to particularly interesting or relevant parts. In addition to that, shorter subsequences of the pattern will be found more often in a time series than the whole pattern. This results in a tendency to relatively often obtain high degrees

ISBN: 978-989-95079-6-8

of similary $\mu(t,\tau)$ for small values of τ, which will later on diminish if the time series does not develop further into an instance of the pattern. For longer patterns, a user will therefore always have to gauge and find a trade-off between the necessity of detecting patterns at an early stage of development τ (with a higher risk of a false or premature recognition result) and the importance of a reliable recognition result (which can ultimately only be achieved at the end of an evolving pattern). This trade-off cannot be universally formulated in a crisp manner, and neither does it seem advisable to do so.

These considerations lead us to the idea of letting the user of the recognition system focus on certain stages development τ of a pattern in a soft manner, such that the recognition system only reports matches of the evolving pattern for values of τ falling within certain (fuzzy) boundaries. Subsequently, we will call these boundaries the *fuzzy window of interest for* τ. We will represent this window of interest by means of a fuzzy set $\mu_w(\tau)$ defined over all possible values of τ, with $\mu_w(\tau) \to 0$ indicating low interest in a recognition result for this particular value of τ, and $\mu_w(\tau) \to 1$ indicating high interest.

To obtain a *windowed* pattern recognition result $\tilde{\mu}(t,\tau)$, the window of interest $\mu_w(\tau)$ and the recognition result $\mu(t,\tau)$ are combined as in (16) by a conjunction[4] of the two fuzzy sets. Semantically, this corresponds to a coincidence of a pattern being found at a certain stage of development and the user's subjective assessment of the importance of this particular stage.

$$\tilde{\mu}(t,\tau) = \mu(t,\tau) \cap \mu_w(\tau) \quad \forall\tau \tag{16}$$

Equation (16) in its general form includes some interesting special cases, such as the following: By definition, the truth values in $\mu(t,\tau)$ appear equally important to the user regardless of τ, so that higher values of μ will often be found for smaller values of τ, although the pattern has just begun to evolve. Following the windowing approach would also enable us to (for example: linearly) weigh a recognition result $\mu(t,\tau)$ according to the stage of development τ of the pattern by chosing a suitable $\mu_w(\tau)$. To achieve this, $\mu_w(\tau)$ has—in this special case—to be chosen as a fuzzy set with (linearly) increasing truth values, reaching $\mu_w(\tau = L) = 1$ at the end of the pattern (length L). If we, as an example, choose the algebraic product for the operator $\cap$ in (16), we would simply obtain (17) for a thusly weighted/windowed recognition result.

$$\tilde{\mu}_{linear}(t,\tau) = \mu(t,\tau) \cdot \frac{\tau}{L} \tag{17}$$

4.2 *Deriving Decisions*

In order to further process the (windowed) recognition results $\mu(t,\tau)$ or $\tilde{\mu}(t,\tau)$, e. g. when serving as a rationale for subsequent actions, it will often be necessary to come to a crisp decision regarding the similarity and stage of development of a pattern found in a time series. In these cases, the fuzzy set $\tilde{\mu}(t,\tau)$ has to be condensed to one or more crisp parameters.

Although many defuzzification methods for fuzzy sets are available [10], we would—for the sake of simplicity—like to restrict ourselves in this article to a maximum approach (first of maxima, FOM). The resulting crisp value τ^* from $\tilde{\mu}(t,\tau)$ points to the earliest stage of development of the pattern yielding the highest degree of similarity $\tilde{\mu}^*$ to the respective time series subsequence, both of which may be represented as a fuzzy singleton $\tilde{\mu}^*(t,\tau^*)$ for each point in time t.

5 Example

Fig. 7 shows a time series containing one instance of the pattern modelled and depicted in Fig. 3. For illustrative purposes, the instance has been highlighted, as the surrounding parts of the the time series are relatively similar in amplitude and shape.

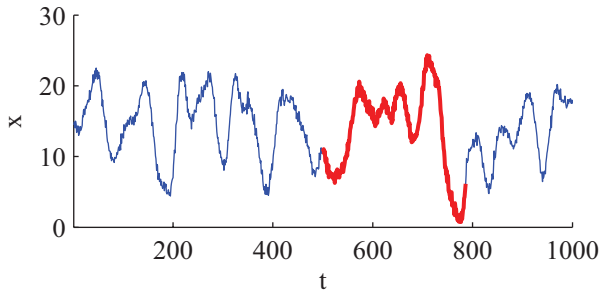


Figure 7: Univariate time series with a pattern instance.

Employing the fuzzy model from Fig. 3, we are now performing an online recognition of this pattern in the time series of Fig. 7. While advancing in time t through all points of the time series, the pattern is constantly being expected to be found in all[5] possible stages of development τ. The classification results for each point in time t—evolving fuzzy sets $\mu(t,\tau)$ defined over τ—are being shown in Fig. 8.

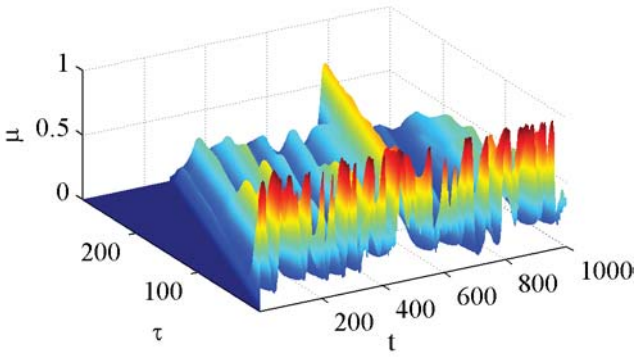


Figure 8: Results for an online recognition of the time series pattern of Fig. 3.

These recognition results $\mu(t,\tau)$ reveal high degrees of similarity of time series subsequences to the first part (i. e. small values of τ) of the pattern, and, in fact, it can be understood from Fig. 7 that the ascending first slope can (with slight variations) be found repeatedly throughout the time series. Apart from that, the progression of a detected instance is well visible in the course of $\mu(t,\tau)$. Only once a completed pattern is being found with distinctly high values for $\mu(t,\tau = L)$, which point to the pattern instance embedded in this time series.

To detect a pattern ahead of time ($\tau < L$), e. g. for forecasting purposes, the recognition result has to be well-founded, and therefore based upon a adequately large number of recognised

[4] As conjunction operator, a non-compensatory operator—such as all T-norm operators—should be chosen to ensure $\tilde{\mu}(t,\tau) \leq \mu(t,\tau)$.

[5] In the first part of the time series, the pattern may only be found at early stages of development $\tau < t$, i. e. $\mu(t,\tau > t) = 0$ must hold.

ISBN: 978-989-95079-6-8

sample points τ on the one hand, whilst τ being small enough to leave sufficient scope for actual forecasting. We will tackle both problems of "premature" pattern recognition and formulating a compromise for a suitable value of τ delivering usable recognition results by following the windowing approach for $\mu(t,\tau)$ presented in section 4.1. As an example, we would like to concentrate on recognition results for pattern instances being at least half-way, but not yet fully completed. Once again employing the unimodal parametric membership function from (1), we formulate a fuzzy window of interest $\mu_w(\tau)$ for this certain stage of the pattern's development, cf. Fig. 9.[6] Due to the fuzzy nature of the presented method, slight variations of $\mu_w(\tau)$ and its parameters will not alter the overall system behaviour and recognition results in a drastic, switching manner. This eases the design of $\mu_w(\tau)$, as no "critical" decisions have to be made when choosing its parameters.

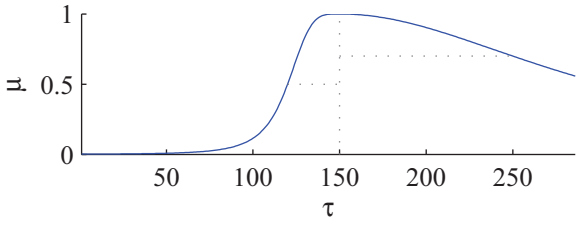


Figure 9: Fuzzy window of interest $\mu_w(\tau)$ for pattern recognition results $\mu(t,\tau)$ from Fig. 8.

For each t, the recognition result (fuzzy set) $\mu(t,\tau)$ was processed as given by (16) using a Hamacher product with $\mu_w(\tau)$. The windowed recognition results $\tilde{\mu}(t,\tau)$ are shown in Fig. 10. As a final step, a simple decision strategy was followed to generate crisp recognition results in singleton form for each point in time t. For each fuzzy set $\tilde{\mu}(t,\tau)$, the modal (maximum) value $\tilde{\mu}^*(t,\tau^*)$ was determined. Any result for τ^* with a corresponding truth value falling below a threshold $\tilde{\mu}^*(t,\tau^*) < 0.5$ was discarded. Results of this decision procedure are marked in Fig. 10. As can be seen in this figure, two candidate pattern instances are being identified at the desired stage of development τ, the first one of which vanishing quickly. The actual pattern instance embedded in the time series of Fig. 7 is being recognised reliably throughout the fuzzy window of interest.

6 Conclusions

This article dealt with the online recognition of patterns in multivariate time series. The vectorial recursive update equation for the recognition results derived in section 3.5 allows to obtain a fuzzy recognition result $\mu(t,\tau)$ for all possible stages of development τ of a pattern at almost the same computational cost as required for the fuzzy classification of only completed patterns. By formulating and applying a fuzzy window of interest, the recognition result $\mu(t,\tau)$ can be narrowed down to parts relevant for the respective application in a fuzzy manner. This also eases a subsequent crisp decision step.

Further work will concentrate on the application of the model and methods presented here to the recognition and prediction

[6] The parameter were chosen manually: $r = 150$, $b_l = 0.5$, $b_r = 0.7$, $c_l = 30$, $c_r = 100$, $d_l = 4$, $d_r = 2$. For $\tau < r$, $d_l = 4$ and $c_l = 30$ ensure a quick decrease of $\mu_w(\tau)$, whereas the right-hand side parameters b_r, c_r, d_r provide a gentle decrease of μ_w for $\tau > r$.

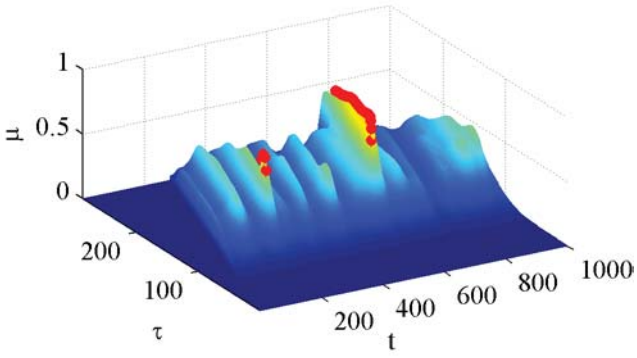


Figure 10: Windowed recognition results $\tilde{\mu}(t,\tau)$. Singleton results $\tilde{\mu}^*(t,\tau^*)$ of the decision procedure are marked with dots (slightly increased for better visibility).

of recurring patterns in energy load time series. Besides that, an enhancement of the decision procedure appears promising. A desirable goal would be a *fuzzy* decision process delivering *parametric* fuzzy information for the stage of development of a detected pattern. This could be more directly employed in subsequent (e. g. prediction) steps without discarding the fuzziness always inherent in a real-world recognition result.

References

[1] Steffen F. Bocklisch and Norman Bitterlich. *Fuzzy systems in computer science*, chapter Fuzzy pattern classification – methodology and application. Vieweg, Wiesbaden, 1994.

[2] Michael Päßler and Steffen F. Bocklisch. *Fuzzy Control: Theory and Practice*, chapter Fuzzy time series analysis, pages 331–345. Springer Verlag, 2000.

[3] Peter Bretschneider. *Ein Beitrag zur Vorhersage musterbasierter nichtlinearer stochastischer Signale*. PhD thesis, TU Ilmenau, 2002.

[4] Bernd Möller and Uwe Reuter. Prediction of uncertain structural responses using fuzzy time series. *Computers and Structures*, 86(10):1123–1139, 2008.

[5] Arne-Jens Hempel and Steffen F. Bocklisch. A conception for networks of fuzzy pattern classifiers. In *Proceedings 13th East West Fuzzy Colloquium*, Zittau, 2006.

[6] Eamonn Keogh, Xiaopeng Xi, Li Wei, and Chotirat Ann Ratanamahatana. The UCR time series classification/clustering homepage. http://www.cs.ucr.edu/~eamonn/time_series_data/, 2006.

[7] Didier Dubois and Henry Prade. Fuzzy numbers: An overview. In James C. Bezdek, editor, *Analysis of Fuzzy Information, Vol.I: Mathematics and Logic*, pages 3–39. CRC Press, Boca Raton, FL, 1987.

[8] William G. Wee and King-Sun Fu. A formulation of fuzzy automata and its application as a model of learning systems. *IEEE Transactions on Systems Science and Cybernetics*, 5(3):215–223, 1969.

[9] Ullrich Scheunert. *Fuzzy-Mengen-Verknüpfung und Fuzzy-Arithmethik zur Sensor-Daten-Fusion*. PhD thesis, Chemnitz University of Technology, 2001.

[10] Werner Van Leekwijck and Etienne E. Kerre. Defuzzification: criteria and classification. *Fuzzy Sets and Systems*, 108(2):159–178, 1999.

ISBN: 978-989-95079-6-8

The Collapsing Method: Does the Direction of Collapse Affect Accuracy?

Sarah Greenfield, Francisco Chiclana, Robert John[1]

1. Centre for Computational Intelligence, School of Computing, De Montfort University, Leicester LE1 9BH, UK
Email: {sarahg, chiclana, rij}@dmu.ac.uk

Abstract— *The Greenfield-Chiclana Collapsing Defuzzifier comes in many variants, depending on the order of slice collapse. The accuracy of the fundamental variants of forward, backward, inward and outward, and the composite variants of forward-backward and outward right-left is compared experimentally for the discretised interval type-2 fuzzy set.*

Keywords— Centroid, Collapsing, Defuzzification, Interval Type-2 Fuzzy Set, Representative Embedded Set.

1 Introduction

A fuzzy inferencing system (FIS) is a computerised system that uses fuzzy sets and rules to support decision making. *Type-2* FISs are being developed for an increasing number of applications such as [1], [2], [3]. There are five main stages to any FIS: fuzzification, antecedent computation, implication, aggregation and defuzzification. In the case of a type-2 FIS (where at least one fuzzy set is type-2), defuzzification consists of two parts — type-reduction and defuzzification proper, as shown in figure 1. Type-reduction is the procedure by which a type-2 fuzzy set is converted to a type-1 fuzzy set, known as the type-reduced set (TRS). The TRS is then easily defuzzified to give a crisp number.

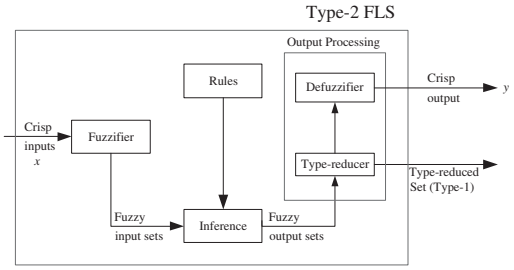


Figure 1: Type-2 FIS (from Mendel [4]).

The type-reduction stage of type-2 defuzzification is problematic owing to its computational complexity. This is the type-reduction algorithm originally described by Mendel ([4], pages 248-252):

1. All possible type-2 embedded sets ([4], definition 3-10, page 98) are enumerated.
2. For each embedded set the minimum secondary membership grade is found.
3. For each embedded set the domain value of the type-1 centroid of the type-2 embedded set is calculated.
4. For each embedded set the secondary grade is paired with the domain value to produce a set of ordered pairs (x,z). It is possible that for some values of x there will be more than one corresponding value of z.
5. For each domain value, the maximum secondary grade is selected. This creates a set of ordered pairs $(x,zMax)$ such that there is a one-to-one correspondence between x and $zMax$. This completes the type-reduction of the type-2 set to the type-1 TRS.

The resultant TRS, as with any type-1 fuzzy set, is readily defuzzified by finding its centroid.

Thus type-reduction involves the processing of *all* the embedded sets within the type-2 set. This is why we term the procedure 'exhaustive defuzzification'. Embedded sets are very numerous. For instance, when a prototype type-2 FIS performed an inference using sets which had been discretised into 51 slices across both the x and y-axes, the number of embedded sets in the aggregated set was calculated to be in the order of 2.9×10^{63}. Though individually easily processed, embedded sets in their totality give rise to a processing bottleneck simply by virtue of their high cardinality. Consequently, exhaustive defuzzification is an impractical technique.

A computationally simpler alternative to the exhaustive method is the *Greenfield-Chiclana Collapsing Defuzzifier*, introduced in [5]. This technique converts an interval type-2 fuzzy set into a type-1 fuzzy set which approximates to the *representative embedded set (RES)*, whose defuzzified value is equal to that of the original type-2 set. As a type-1 set, the RES may then be defuzzified straightforwardly.

In this paper we build on the work reported in [5]. The next section covers assumptions and definitions; section 3 presents an overview of the collapsing defuzzifier, after which section 4 introduces the theme of this article, the notion that *there are variants of the collapsing method.* Sections 5 and 6 are concerned with finding experimentally the most accurate variant.

2 Preliminaries

2.1 Assumptions

Discretisation The work presented here is concerned only with defuzzification of *discretised* type-2 fuzzy sets.

Interval Type-2 Fuzzy Set This paper is concerned with the *interval* type-2 fuzzy set.

Centroid Method of Defuzzification It is assumed that the centroid method of defuzzification ([6], page 336) is used.

2.2 Definitions

Definition 1 (Degree of Discretisation) *The* degree of discretisation *of a discretised fuzzy set is the separation of the slices.*

ISBN: 978-989-95079-6-8

Scalar Cardinality For type-1 fuzzy sets, Klir and Folger ([7], p17) define scalar cardinality as follows:

Definition 2 (Scalar Cardinality) *The* scalar cardinality *of a fuzzy set A defined on a finite universal set X is the summation of the membership grades of all the elements of X in A. Thus,*

$$| A |= \sum_{x \in X} \mu_A(x). \qquad ([7], p17)$$

To distinguish scalar cardinality from cardinality in the classical sense, we adopt the '$\| \|$' symbol for scalar cardinality.

3 Overview of the Collapsing Method

An interval type-2 set may be regarded as a blurred type-1 set. The collapsing method is a technique for deriving a type-1 fuzzy set from a type-2 fuzzy set, and may be thought of as a reversal of blurring. The type-1 set's membership function is calculated so that its defuzzified value approximates that of the type-2 fuzzy set. It is a simple matter to defuzzify the type-1 set, and to do so would be to find the defuzzified value of the original type-2 fuzzy set. The collapsing process approximates the output of the type-reducer followed by the type-1 defuzzifier, and in so doing reduces the computational complexity of type-2 defuzzification. We term this special type-1 set the *'representative embedded set approximation (RESA)'*. (Full details of the collapsing algorithm, including proof of the

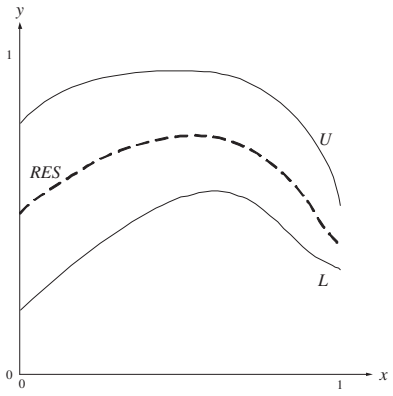


Figure 2: A Representative Embedded Set.

associated theorem, may be found at [5].) We formally state the Simple[1] Representative Embedded Set Approximation:

Theorem 1 (Simple Representative Embedded Set Approx.) *The membership function of the* embedded set *R derived by dynamically collapsing slices of a discretised type-2 interval fuzzy set $\tilde{F}$, having lower membership function L, and upper membership function U, is:*

$$\mu_R(x_i) = \mu_L(x_i) + r_i$$

with

$$r_i = \frac{\left(\|L\| + \sum_{j=1}^{i-1} r_j\right) b_i}{2\left(\|L\| + \sum_{j=1}^{i-1} r_j\right) + b_i}, \qquad (1)$$

and $b_i = \mu_U(x_i) - \mu_L(x_i)$, $r_0 = 0$.

This is an iterative formula. Collapsing proceeds vertical slice by vertical slice. The first slice is collapsed, the first y-value of the RESA calculated, the next slice is collapsed and the second y-value of the RESA calculated, and so on until all the slices have been collapsed. In this formula each b_i is the blur for vertical slice i, i.e. the difference between the upper membership function and the lower membership function for slice i. Each r_i is the amount by which the y-value of L must be increased to give the y-value of the RESA R.

4 Variants of the Collapsing Method

As we have seen, equation (1) is the formula for collapsing. This is in fact *a version* of collapsing — the most intuitive variant, whereby the slices are collapsed in the order of increasing domain value ($x = 0$ to $x = 1$). We term this *collapsing forward*. However slice collapse may be performed in any slice order giving slightly different RESAs. If the domain of the interval type-2 fuzzy set is discretised into s vertical slices, the number of permutations of these slices is $s!$ ([8], page 139). Therefore there must be $s!$ RESAs obtainable by varying the order of slice collapse. The question this piece of research addresses is, "Does the order in which the slices are collapsed affect the accuracy of the method?"

There are four fundamental variants, which we term *forward*, *backward*, *outward* and *inward*. Inward and outward may each be approached in two different ways. For the inward variant, slice collapse might start from the left or from the right. For the outward variant, the first slice is in the middle[2], but the second slice may be to the right or to the left. Added to these, there are composite variants, such as *forward-backward*, which is the mean of the defuzzified values found by collapsing forward and collapsing backward.

5 Experimental Comparison

5.1 Test Sets

We chose to test defuzzification in isolation from the rest of an FIS, on specially created test sets. Our methodology was to run different collapsing variants against each other to see which gave the most accurate results.

Symmetric Horizontal Test Set The lower membership function is the line $y = 0.2$; the upper membership function the line $y = 0.8$. The shape of this test set may be described as a horizontal stripe. The symmetry of this set tells us that its defuzzified value is 0.5.

Symmetric Triangular Test Set This is a normal test set. The lower and upper membership functions are both triangular in shape, both with vertices at $(0.4, 1)$. The symmetry of this set reveals its defuzzified value to be 0.4.

Asymmetric Gaussian Test Set This test set was deliberately designed to be asymmetrical, and hence a more realistic simulation of an FIS aggregated set. Both the lower and upper membership functions are Gaussian. As this set has no symmetry, exhaustive defuzzification (section 1) had to be employed to determine the actual defuzzified value, which, as

[1] In [5], we used the term 'simple' to describe an interval set in which each vertical slice consists of only two points, corresponding to L and U. The term is redundant in the context of this paper.

[2] We always employ an odd number of slices, giving a determinate middle slice.

ISBN: 978-989-95079-6-8

would be expected, varies slightly with the degree of discretisation.

5.2 Test Strategy

A preliminary set of tests was performed on the fundamental variants: forward, backward, inward, outward, and the composite variant forward-backward. As will be reported in the next section, the outward variant outperformed the others. This led on to further tests to discover the most accurate version of collapsing outward.

For the Gaussian test set, the only way of knowing its defuzzified value was to employ exhaustive defuzzification (section 1). This procedure only works properly for 21 slices or under. In contrast, the horizontal and triangular test sets reveal their defuzzified values by symmetry, so the number of slices used can be much higher, allowing finer discretisation.

6 Results and Conclusion

6.1 Preliminary Tests

Table 1 gives the results for the horizontal test set; table 2 gives the associated errors. Table 3 shows the triangular test set results, and table 4 the errors. The defuzzification results for the Gaussian test set are shown in table 5, with the errors in table 6. For all three test sets, the best performing variant was outward, followed by inward, then forward and backward. For the symmetrical sets (horizontal and triangular), the errors of collapsing forward were equal and opposite to those of collapsing backward. Therefore in these cases we would expect collapsing forward-backward to give exact results. This has been confirmed by experiments. For the Gaussian test set, backward performed more poorly than forward. In this case the composite of forward-backward performed worse than forward, though better than backward.

6.2 Further Tests

The outward variant may be performed in two ways, outward right and outward left (section 4). Collapsing right-left is the mean of collapsing right and collapsing left. The results and associated errors for the three versions of the outward variant as applied to the three test sets are shown in tables 7 to 9.

For the symmetrical horizontal test set, outward right and outward left gave rise to equal but opposite errors. For the composite outward right-left, these errors cancelled to zero.

The triangular test set, though symmetrical, is not placed symmetrically about $x = 0.5$. The errors of collapsing right and collapsing left were of equal sign and either equal or very close in quantity. When the errors were not equal, those of outward left were marginally smaller than those of outward right. Unsurprisingly, the performance of the composite of outward right-left fell between that of outward right and outward left.

For the Gaussian test set, the errors were all of negative sign. At all degrees of discretisation, outward left gave the best results, outward right the poorest, and outward right-left came in between.

For two of the three test sets outward left outperformed outward right. Our conjecture is that the position of the centroid is an important factor affecting which performs better out of outward right and outward left. This topic requires further research using a wider range of test sets, but for now we conclude that the optimum strategy is the composite outward right-left.

6.3 Why is Outward the Most Accurate Variant?

This explanation is based on the symmetrical horizontal test set. As each slice is collapsed, $\|L\| + \sum_{j=1}^{j=i-1} r_j$ in both the numerator and denominator of the collapsing formula (equation (1)) increases, which means that as the collapse progresses, the r_i for each collapsed slice i is a closer approximation to $\frac{1}{2}b_i$, i.e. half the 'blur' term. Thus with every successive collapsed slice, the RESA tends towards the midline of L and U, as shown in figure 3 for the forward and backward variants.

For the symmetrical horizontal test set, we take the RES to be the midline of L and U for two reasons. Firstly, by symmetry we would expect the RES to be a horizontal line. Secondly, as the number of slices is increased (either as the collapse progresses, or as the degree of discretisation is made finer), the RESA gets closer to the midline of L and U.

Therefore, as the slices are collapsed, the RESA approaches the RES. This means that the earlier slices in the RESA deviate more from the RES than the later ones. To get the best results, the collapse need to proceed symmetrically. Both the inward and outward variants meet this criterion; the inaccuracies are distributed symmetrically. However the greatest inaccuracy is associated with the first collapsed slice. To achieve maximum accuracy, the ideal place for this first slice to be positioned is centrally, as the effect on the defuzzified value obtained is then minimal. For this reason outward (figure 4) gives a more accurate defuzzified values than inward. We would expect the same reasoning to apply to all type-2 fuzzy test sets. However further investigation, using radically contrasting test sets, is planned.

References

[1] Hani Hagras. A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots. *IEEE Transactions on Fuzzy Systems*, 12:524–539, 2004.

[2] Robert I. John and Sarah Lake. Modelling nursing perceptions using type-2 fuzzy sets. In *Proc. EUROFUSE 2001*, pages 241–246, 2001.

[3] Turhan Ozen and Jonathan M. Garibaldi. Investigating adaptation in type-2 fuzzy logic systems applied to umbilical acid-base assessment. In *Proc. EUNITE 2003*, pages 289–294, Oulu, Finland, 2003.

[4] Jerry M. Mendel. *Uncertain rule-based fuzzy logic systems: Introduction and new directions*. Prentice-Hall PTR, 2001.

[5] Sarah Greenfield, Francisco Chiclana, Simon Coupland, and Robert I. John. The collapsing method of defuzzification for discretised interval type-2 fuzzy sets. 2008. Accepted for *Information Sciences Special Edition on Higher Order Fuzzy Sets*. DOI: 10.1016/j.ins.2008.07.011.

[6] George J. Klir and Bo Yuan. *Fuzzy sets and fuzzy logic*. Prentice-Hall P T R, 1995.

[7] George J. Klir and Tina A. Folger. *Fuzzy sets, uncertainty, and information*. Prentice-Hall International, 1992.

[8] ed. Carol Gibson. *The facts on file dictionary of mathematics*. Facts on File, New York and Oxford, second edition, 1988.

ISBN: 978-989-95079-6-8

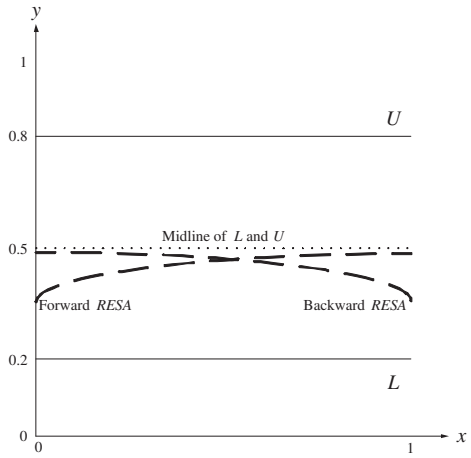


Figure 3: Forward RESA and Backward RESA.

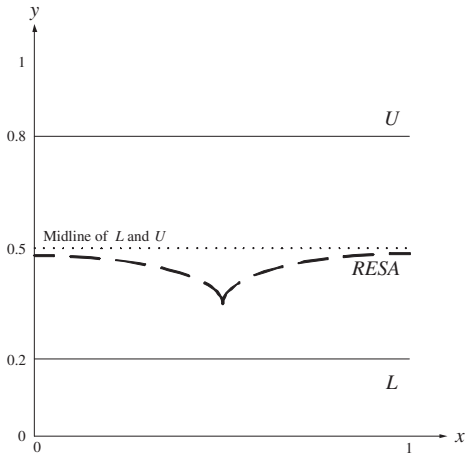


Figure 4: Outward RESA.

Table 1: Defuzzified Values Obtained by Collapsing the Symmetrical Horizontal Test Set.

Degree of Discretisation	Defuzzified Value	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward
0.1	0.5	0.5038320922	0.4961679078	0.4993086838	0.4995891494
0.05	0.5	0.5019998917	0.4980001083	0.4998049953	0.4998891777
0.02	0.5	0.5008177226	0.4991822774	0.4999665227	0.4999815068
0.01	0.5	0.5004115350	0.4995884650	0.4999914377	0.4999953154
0.005	0.5	0.5002064040	0.4997935960	0.4999978353	0.4999988213
0.002	0.5	0.5000827100	0.4999172900	0.4999996513	0.4999998107
0.001	0.5	0.5000413793	0.4999586207	0.4999999126	0.4999999526
0.0001	0.5	0.5000041401	0.4999958599	0.4999999991	0.4999999995
0.00001	0.5	0.5000004140	0.4999995860	0.5000000000	0.5000000000

Table 2: Errors Incurred in Collapsing the Symmetrical Horizontal Test Set.

Degree of Discretisation	Defuzzified Value	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward
0.1	0.5	0.0038320922	-0.0038320922	-0.0006913162	-0.0004108506
0.05	0.5	0.0019998917	-0.0019998917	-0.0001950047	-0.0001108223
0.02	0.5	0.0008177226	-0.0008177226	-0.0000334773	-0.0000184932
0.01	0.5	0.0004115350	-0.0004115350	-0.0000085623	-0.0000046846
0.005	0.5	0.0002064040	-0.0002064040	-0.0000021647	-0.0000011787
0.002	0.5	0.0000827100	-0.0000827100	-0.0000003487	-0.0000001893
0.001	0.5	0.0000413793	-0.0000413793	-0.0000000874	-0.0000000474
0.0001	0.5	0.0000041401	-0.0000041401	-0.0000000009	-0.0000000005
0.00001	0.5	0.0000004140	-0.0000004140	0.0000000000	0.0000000000

ISBN: 978-989-95079-6-8

Table 3: Defuzzified Values Obtained by Collapsing the Symmetrical Triangular Test Set.

Degree of Discretisation	Defuzzified Value	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward
0.1	0.4	0.4001359091	0.3998640909	0.4001131909	0.3998916916
0.05	0.4	0.4000597189	0.3999402811	0.4000498280	0.3999505451
0.02	0.4	0.4000230806	0.3999769194	0.4000195457	0.3999808751
0.01	0.4	0.4000115326	0.3999884674	0.4000098170	0.3999904744
0.005	0.4	0.4000057773	0.3999942227	0.4000049299	0.3999952381
0.002	0.4	0.4000023153	0.3999976847	0.4000019784	0.3999980943
0.001	0.4	0.4000011585	0.3999988415	0.4000009904	0.3999990469
0.0001	0.4	0.4000001159	0.3999998841	0.4000000992	0.3999999047
0.00001	0.4	0.4000000116	0.3999999884	0.4000000099	0.3999999905

Table 4: Errors Incurred in Collapsing the Symmetrical Triangular Test Set.

Degree of Discretisation	Defuzzified Value	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward
0.1	0.4	0.0001359091	-0.0001359091	0.0001131909	-0.0001083084
0.05	0.4	0.0000597189	-0.0000597189	0.0000498280	-0.0000494549
0.02	0.4	0.0000230806	-0.0000230806	0.0000195457	-0.0000191249
0.01	0.4	0.0000115326	-0.0000115326	0.0000098170	-0.0000095256
0.005	0.4	0.0000057773	-0.0000057773	0.0000049299	-0.0000047619
0.002	0.4	0.0000023153	-0.0000023153	0.0000019784	-0.0000019057
0.001	0.4	0.0000011585	-0.0000011585	0.0000009904	-0.0000009531
0.0001	0.4	0.0000001159	-0.0000001159	0.0000000992	-0.0000000953
0.00001	0.4	0.0000000116	-0.0000000116	0.0000000099	-0.0000000095

Table 5: Defuzzified Values Obtained by Collapsing the Gaussian Test Set.

Degree of Discretisation	Defuzzified Value (EM)	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward	Collapsing Forward-Backward
0.5	0.2899142309	0.2947300898	0.4090097593	0.2940174555	0.2884666838	0.3518699246
0.25	0.2906756945	0.2925398791	0.3712146394	0.2923170555	0.2900969651	0.3318772592
0.125	0.3043413255	0.3052741624	0.3526142975	0.3051864643	0.3041285835	0.3289442299
0.1	0.3074987724	0.3082433183	0.3477346996	0.3081777251	0.3073450280	0.3279890090
0.0625	0.3125118626	0.3129728510	0.3393585073	0.3129362993	0.3124323840	0.3261656791
0.05	0.3142610070	0.3146278507	0.3362363800	0.3145998182	0.3142020426	0.3254321154

Table 6: Errors Incurred in Collapsing the Gaussian Test Set.

Degree of Discretisation	Defuzzified Value (EM)	Collapsing Forward	Collapsing Backward	Collapsing Inward	Collapsing Outward	Collapsing Forward-Backward
0.5	0.2899142309	0.0048158589	0.1190955284	0.0041032246	-0.0014475471	0.0619556937
0.25	0.2906756945	0.0018641846	0.0805389449	0.0016413610	-0.0005787294	0.0412015647
0.125	0.3043413255	0.0009328369	0.0482729720	0.0008451388	-0.0002127420	0.0246029044
0.1	0.3074987724	0.0007445459	0.0402359272	0.0006789527	-0.0001537444	0.0204902366
0.0625	0.3125118626	0.0004609884	0.0268466447	0.0004244367	-0.0000794786	0.0136538165
0.05	0.3142610070	0.0003668437	0.0219753730	0.0003388112	-0.0000589644	0.0111711084

ISBN: 978-989-95079-6-8

Table 7: Defuzzified Values and Errors Obtained for the Symmetrical Horizontal Test Set, Collapsed Outward.

		Collapsing Defuzzified Values			Errors		
Degree of Disc.	Defuzzified Value	Outward Right	Outward Left	Outward Right-Left	Outward Right	Outward Left	Outward Right-Left
0.1	0.5	0.4995891494	0.5004108506	0.5000000000	-0.0004108506	0.0004108506	0.0000000000
0.05	0.5	0.4998891777	0.5001108223	0.5000000000	-0.0001108223	0.0001108223	0.0000000000
0.02	0.5	0.4999815068	0.5000184932	0.5000000000	-0.0000184932	0.0000184932	0.0000000000
0.01	0.5	0.4999953154	0.5000046846	0.5000000000	-0.0000046846	0.0000046846	0.0000000000
0.005	0.5	0.4999988213	0.5000011787	0.5000000000	-0.0000011787	0.0000011787	0.0000000000
0.002	0.5	0.4999998107	0.5000001893	0.5000000000	-0.0000001893	0.0000001893	0.0000000000
0.001	0.5	0.4999999526	0.5000000474	0.5000000000	-0.0000000474	0.0000000474	0.0000000000
0.0001	0.5	0.4999999995	0.5000000005	0.5000000000	-0.0000000005	0.0000000005	0.0000000000
0.00001	0.5	0.5000000000	0.5000000000	0.5000000000	0.0000000000	0.0000000000	0.0000000000

Table 8: Defuzzified Values and Errors Obtained for the Symmetrical Triangular Test Set, Collapsed Outward.

		Collapsing Defuzzified Values			Errors		
Degree of Disc.	Defuzzified Value	Outward Right	Outward Left	Outward Right-Left	Outward Right	Outward Left	Outward Right-Left
0.1	0.4	0.3998916916	0.3998916916	0.3998916916	-0.0001083084	-0.0001083084	-0.0001083084
0.05	0.4	0.3999505451	0.3999522908	0.3999514180	-0.0000494549	-0.0000477092	-0.0000485820
0.02	0.4	0.3999808751	0.3999812363	0.3999810557	-0.0000191249	-0.0000187637	-0.0000189443
0.01	0.4	0.3999904744	0.3999905679	0.3999905212	-0.0000095256	-0.0000094321	-0.0000094788
0.005	0.4	0.3999952381	0.3999952617	0.3999966681	-0.0000047619	-0.0000047383	-0.0000047501
0.002	0.4	0.3999980943	0.3999980981	0.3999980962	-0.0000019057	-0.0000019019	-0.0000019038
0.001	0.4	0.3999990469	0.3999990479	0.3999990474	-0.0000009531	-0.0000009521	-0.0000009526
0.0001	0.4	0.3999999047	0.3999999047	0.3999999047	-0.0000000953	-0.0000000953	-0.0000000953
0.00001	0.4	0.3999999905	0.3999999905	0.3999999905	-0.0000000095	-0.0000000095	-0.0000000095

Table 9: Defuzzified Values and Errors Obtained for the Gaussian Test Set, Collapsed Outward.

		Collapsing Defuzzified Values			Errors		
Degree of Disc.	Defuzzified Value	Outward Right	Outward Left	Outward Right-Left	Outward Right	Outward Left	Outward Right-Left
0.5	0.2899142309	0.2884666838	0.2890645675	0.2887656257	-0.0014475471	-0.0008496634	-0.0011486052
0.25	0.2906756945	0.2900969651	0.2902918203	0.2901943927	-0.0005787294	-0.0003838742	-0.0004813018
0.125	0.3043413255	0.3041285835	0.3041906758	0.3041285835	-0.0002127420	-0.0001506497	-0.0001816959
0.1	0.3074987724	0.3073450280	0.3073862653	0.3073656467	-0.0001537444	-0.0001125071	-0.0001331257
0.0625	0.3125118626	0.3124323840	0.3124493111	0.3124408476	-0.0000794786	-0.0000625515	-0.0000710150
0.05	0.3142610070	0.3142020426	0.3142130418	0.3142075422	-0.0000589644	-0.0000479652	-0.0000534648

ISBN: 978-989-95079-6-8

Data Integration, Approximate Categorisation and Fuzzy Associations

Trevor Martin[1,2] Yun Shen[1]

1 Artificial Intelligence Group, Department of Engineering Mathematics,
University of Bristol, University Walk, Bristol BS8 1TR, UK
2 BT Intelligent Systems Lab, BT Innovate, Orion Building
Adastral Park, Ipswich IP5 3RE, UK

Email: firstName.lastName@bristol.ac.uk

Abstract—*The use of hierarchical taxonomies to organise information (or sets of objects) is essential to the semantic web and is also fundamental to many aspects of web 2.0. In most cases, the seemingly crisp granulation of a taxonomy disguises the fact that categories are based on loosely defined concepts which are better modelled by allowing graded membership. Fuzzy categories may also arise when integrating information from multiple sources which do not conform to precisely the same taxonomy definitions. Knowledge of relations between categories can be summarised by association rules. In this paper, we outline a new method to calculate fuzzy confidences for association rules between fuzzy categories from different hierarchies. We illustrate with examples drawn from a system that integrates information from web-based sources.*

Keywords— fuzzy association rules, fuzzy data mining, fuzzy hierarchies, soft semantic web, mass assignment.

1 Introduction

The semantic web [1] is a co-ordinated attempt to develop "*common formats for integration and combination of data drawn from diverse sources*"* whilst Web2.0 is an emergent trend involving changes in web use, including user participation, content sharing, collaboration, etc. Both rely heavily on hierarchical taxonomies as a fundamental mechanism for information organisation. In the case of the semantic web, ontologies provide a formal definition of terms, relations, constraints, etc. as well as their correspondence to objects in the real world. For example, an ontology for wine retailers would define terms such as grape type, vintage, region, descriptions of flavour, etc. Web documents using these terms can be processed by machines, knowing that the "meaning" is defined by the ontology. The use of "tags" from the ontology enables automated reasoning as well as assisting human understanding.

In the case of Web 2.0, tagging and categorisation also play a central role. Photo and video sharing sites, wikis, blogs, and large parts of e-commerce rely on hierarchical categorisation of content. This may be implicit rather than explicit - a tag such as *sport* can encompass *athletics*, *swimming*, *football*, etc - but the hierarchical nature is unarguable. Rather than relying on formal, agreed ontologies, the meaning of a tag is essentially defined by the user who adds the tag, and a degree of shared understanding is necessary for the system to work.

The ability to categorise and summarise data is a key feature of human intelligence, enabling us to group multiple entities together into an (approximately) uniform whole and represent / reason about the group as a single concept. Many familiar ways to access information such as books, libraries, computer file structures, the web and other networks provide evidence that hierarchical categorisation is an efficient way to organise and access information.

There is rarely a single unique hierarchy - instead, a person chooses the most appropriate way to split up data according to their expertise / background knowledge and the problem at hand. For example, Fig 1 shows two possible divisions of customers. Here, as in most real-world hierarchies, groups of entities (or conceptual categories) are loosely defined, able to admit elements according to some scale of membership rather than according to an absolute yes/no test - indicating that fuzzy set theory [2] is appropriate.

An association exists when the extensions of two concepts overlap significantly, as indicated in Fig 1 by the directed link. Association rules (in their crisp form) are a well-established technique for knowledge discovery in databases,

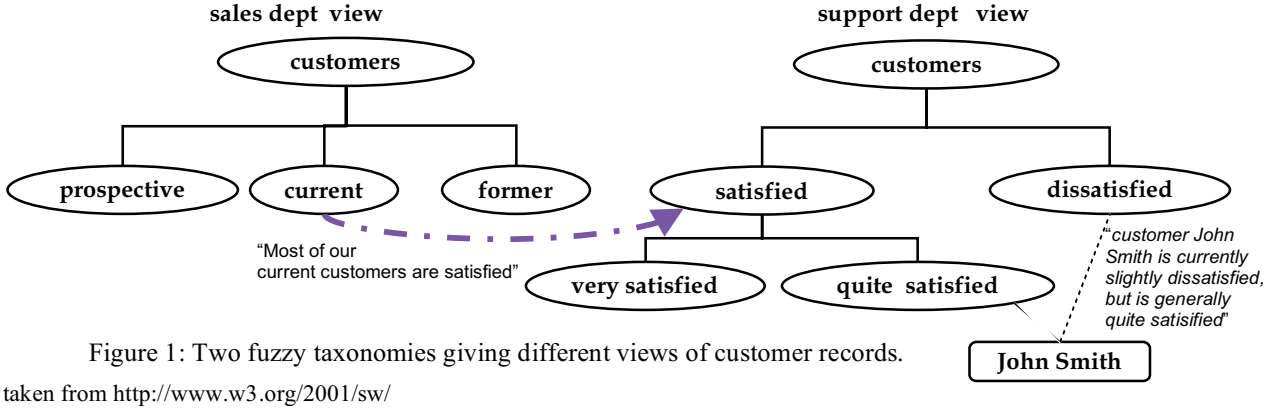


Figure 1: Two fuzzy taxonomies giving different views of customer records.

* taken from http://www.w3.org/2001/sw/

ISBN: 978-989-95079-6-8

enabling "interesting" relations to be discovered. There have been a number of proposals to develop fuzzy association rules, that is to discover the degree of association between fuzzy categories. Some of our recent work has used mass assignment theory [3-5] to find a point valued association strength between fuzzy categories [6], later extended to an interval-valued version [7]. In common with other work on fuzzy association rules, most of our previous work assumes there is a crisp value for the rule confidence.

The main contribution of this paper is a novel mass assignment-based method for calculating a *fuzzy* confidence in associations between fuzzy categories. It relies on a new method of converting fuzzy relations to mass assignments and a membership function for fuzzy confidence related to the movement of mass needed to produce that confidence value. The minimum and maximum values for the confidence can be found quickly, and memberships calculated based on the corresponding mass assignments. A full version, including proofs, will appear subsequently [8].

2 Background

The Smart Queries and Adaptive Data (SQuAD) project is concerned with adding structure (tags) to data and refining approximate knowledge from this data. It aims to assist in the extraction of useful information from diverse sources of semi-structured text data and consists of four main components:

- approximate fuzzy grammars [9] to tag small text fragments.
- SOFT (Structured Object Fusion Toolkit) [10] to help determine that entities from different sources are the same (a process also known as instance-matching, entity resolution)
- iPHI (intelligent Personal Hierarchies of Information) [11, 12] which aims to combine and integrate multiple sources of information and to configure access to the information based on an individual's personal categories.
- TRACK (Time-varying Relations in Approximately Categorised Knowledge). [7] The previous three processes leave us with a collection of objects organised into fuzzy categories, where the taxonomic structure reflects a view of the underlying data. Insight can be gained by considering relations between fuzzy categories, particularly from different category hierarchies. Of particular interest is the change over time in the degree of association between fuzzy categories - for example, in Fig 1, an executive in charge of the company would be interested to know whether the set of current customers (in the sales department's categorisation) is mostly the same as the fuzzy set of satisfied customers (in the support department's categorisation) and how the relation has changed over recent time, over the medium term and over the long term - particularly if there have been significant changes such as company practice, number/nature of competitors, etc during any of these periods.

2.1 Fuzzy Sets - the Conjunctive Interpretation

Many authors (e.g. [13]) have proposed fuzzy sets to model uncertain values in databases and knowledge based applications. The standard (disjunctive) interpretation of a fuzzy set in this context is as a possibility distribution - a single valued attribute which is not known exactly.

The conjunctive interpretation of a fuzzy set occurs when the attribute can have multiple values. These values are not repeated so it is a set, not a multi-set. For example, a person may be able to speak several languages; we could model this as a fuzzy set of languages, where membership would depend on the degree of fluency. This is formally a relation rather than a function on the underlying sets. We distinguish between the conjunctive interpretation - modelled by a monadic fuzzy relation – and the disjunctive interpretation – modelled by a possibility distribution by using the notation

$$F(a) = \{x/\mu(x) \mid x \in U\}$$

to denote a single valued attribute F of some object a (i.e. a possibility distribution over a universe U) and

$$R(a) = [x/\chi(x) \mid x \in U]$$

to denote a multi-valued attribute (relation). Fuzzy categories represent the latter case, since we have multiple values that satisfy the predicate to a greater or lesser degree.

2.2 Extending Association Rules to Fuzzy Categories

In creating association rules within transaction databases (e.g. [14], see also [15] for a clear overview), the standard approach is to consider a table in which columns correspond to items and each row is a transaction. A cell contains 1 if the item was bought, and 0 otherwise. The aim of association rule mining is to find links between disjoint subsets of items – for example, do customers generally buy biscuits and cheese when buying beer and wine? These disjoint subsets represent categories, as described earlier.

Let I denote the set of items, so that any transaction can be represented as $tr \subseteq I$, and consider X, the set of all transactions (strictly speaking, X is a multi-set but can be made into a set by adding a unique identifier to each transaction). We also specify two categories (itemsets) s and t, which are non-empty, non-overlapping subsets of I,

$$t \subset I \qquad s \subset I \quad \text{where} \quad s \cap t = \varnothing$$

and the sets of transactions containing s and t

$$S = \{x | x \in X \wedge s \subseteq x\}$$

$$T = \{x | x \in X \wedge t \subseteq x\}$$

An association rule is of the form $s \Rightarrow t$ and is interpreted as stating that when the items in s appear in a transaction, it is likely that the items in t will also appear i.e. it is not an implication in the formal logical sense. A slight abuse of notation allows us to use $S \Rightarrow T$ or $s \Rightarrow t$ as the rule.

Most authors use two measures to assess the significance of association rules. The support of a rule $s \Rightarrow t$ is the number (or relative number) of transactions in which both s and t appear, and the confidence of the rule is an estimate (based on the samples) of the conditional probability of t being contained in a transaction given that it contains s

$$Support(s,t) = |S \cap T| \qquad (1)$$

$$Conf(s,t) = \frac{|S \in T|}{|S|} \qquad (2)$$

Typically a threshold is chosen for the support, so that only frequently occurring sets of items s and t are considered; a second threshold filters out rules of low confidence.

ISBN: 978-989-95079-6-8

name	sales	salary
a	100	1000
b	80	400
c	50	800
d	20	700

good sales	*high salary*	*confidence*
sales≥80	*high≥400*	1
sales≥50	*high≥500*	0.667
sales>50	*high>500*	0.5
sales≥50	*high>800*	0.333

Figure 2 : A simple database of names (*a, b, c, d*), sales and salaries (top) and (bottom) confidences for the rule *good sales => high salary* arising from different crisp definitions of the terms *good sales* and *high salary*

For example, consider a database of sales employees, salaries and sales figures. A mining task might be to find out whether the good sales figures are achieved by the highly paid employees. Given the database table in Fig 2, we can obtain rule confidences ranging from 1/3 up to 1 by different crisp definitions of "*good sales*" and "*high salary*".. Although this is a contrived example, such sensitivity to the cut-off points adopted for crisp definitions is a good indication that a fuzzy approach is more in line with human understanding of the categories.

Various approaches to fuzzifying association rules have been proposed e.g. [15-17]. The standard extension to the fuzzy case is to treat the sets *S, T* as fuzzy and find the intersection and cardinality using a t-norm and sigma-count respectively.

$$Conf(S,T) = \frac{\in \mu_{S \cap T}(x)}{\in \mu_S(x)} \qquad (3)$$

In the example of Fig 2, a fuzzy approach would categorise employees according to simple membership functions for *good sales* (S) and *high salary* (T), which could lead to

$S = [a/1, b/0.8, c/0.5, d/0.2]$

and

$T = [a/1, b/0.4, c/0.8, d/0.7]$

and confidence 0.72 for the association $S=>T$ using (3).

As pointed out by [15], using min and the sigma count for cardinality can be unsatisfactory because it does not distinguish between several tuples with low memberships and few tuples with high memberships - for example,

$S = [x_1/1]$

$T = [x_2/1]$

leads to *Conf(S, T) = 0* but

$S_1 = [x_1/1, x_2/0.01, x_3/0.01, \ldots, x_{1000}/0.01]$

$T_1 = [x_1/0.01, x_2/1, x_3/0.01, \ldots, x_{1000}/0.01]$

leads to

$$Conf(S_1,T_1) = \frac{1000 \in 0.01}{1 + 999 \in 0.01} \in 0.91$$

which is extremely high for two almost disjoint sets (this example originally appeared in [18]). Using a fuzzy cardinality (i.e. a fuzzy set over the possible cardinality values) is also potentially problematic since the result is a possibility distribution over rational numbers, and the extension principle [19] gives a wider bound than it should, due to neglect of interactions between the numerator and denominator in Eq 2. For example, given

$S = [x1/1, x2/0.8]$

$T = [x1/1, x2/0.4].$

the fuzzy cardinalities are

$|S \cap T| = \{1/1, 2/0.4\}$,

$|S| = \{1/1, 2/0.8\}$

leading (by the extension principle) to a confidence of

$\{0.5/0.8, 1/1, 2/0.4\}$,

clearly incorrect as the confidence cannot be greater than 1.

2.3 Fuzzy Relations as Mass Assignments

As with previous work [6, 7] we start from the fact that a relation represents a conjunctive set of ordered *n*-tuples i.e. a conjunction of *n* ground clauses. For example, if *U* is the set of dice values then we could define a (crisp) predicate *differBy4or5* on $U \times U$ as the set of pairs

[(1,6), (1,5), (2,6), (5,1), (6,1), (6, 2)]

This is a conjunctive set, in that each pair satisfies the predicate. In a similar way, a fuzzy relation represents a set of *n*-tuples that satisfy a predicate to some degree. Thus *differByLargeAmount* could be represented by

[(1,6)/1, (1,5)/0.6, (2,6)/0.6, (5,1)/0.6, (6,1)/1, (6,2)/0.6]

The interpretation is not that a single pair satisfies this predicate, but that one set of pairs satisfies it (out of several possible sets of pairs).

A mass assignment [3-5] on a universe *U* is a distribution over the power set of *U*. Here, the mass assignment is on possible *relations* :

$R_1 = [(1,6),(6,1)]$

$R_2 = [(1,6),(1,5),(2,6),(5,1),(6,1),(6,2)]$

$m = \{\{R_1\} : 0.4, \{R_1, R_2\} : 0.6\}$

This is equivalent to treating the fuzzy relation as a fuzzy set of crisp relations:

$differByLargeAmount = \{R1/1, R2/0.6\}$

Similarly, a monadic fuzzy predicate *largeValue* defines a set of 1-tuples such as [6/1, 5/0.8, 4/0.3] which is written as a fuzzy set of crisp monadic relations:

$largeValue = \{[6]/1, [6,5]/0.8, [6,5,4]/0.3\}$

and has the mass assignment

$m_{largeValue} = \{\{[6]\} : 0.2, \{[6],[6,5]\} : 0.5, \{[6],[6,5],[6,5,4]\} : 0.3\}$

Our subsequent studies [7] show that this approach can sometimes overestimate the difference between full and nearly-full membership, which can lead to unreasonably large intervals calculated for the confidence of association rules. For example, under this interpretation, the monadic fuzzy relation $S = [a/1\ b/0.98]$ has the mass assignment

$m_S = \{\{[a]\} : 0.02, \{[a],[a,b]\} : 0.98\}$

ISBN: 978-989-95079-6-8

The normal mass assignment interpretation allows us to redistribute the mass on {*[a], [a,b]*} to either of the relations *[a]* or *[a,b]* which leads to the family of distributions:

$S = [a] : 1\text{-}x \quad , \quad [a,b] : x \quad$ where $0 \le x \le 0.98$

This flexibility in re-assigning mass means that for a source

$S = [a/1\ b/0.98]$ and target relation $T = [a/1\ b/0.98\ c/0.02]$

we get an interval [0.51, 1] which is surprisingly wide considering the two relations are so similar. NB this behaviour arises mostly in contrived cases, smaller intervals are calculated in the vast majority of "real" association rules that have arisen in our experimental studies such as [7].

2.4 Alternative Interpretation of Relations as Mass Assignments

The approach discussed above, which we will refer to as an open world approach, treats partial membership of a tuple x in a relation R (i.e. $0 < \cap_R(x) < 1$) as an upper bound for the mass that can be assigned to any set of tuples including x. This leads to a wide range of mass distributions that can be derived from the fuzzy relation R.

In the open world approach, for any tuple x where $\cap_R(x)<1$, the total mass that can be assigned to relations containing x is given by

$$0 \le \sum_{\substack{t=[x1,\ldots,xn] \\ x \in t}} m(t) \le \chi_R(x)$$

In the *largeValue* example above, consider the element $x=5$, which has $\cap_R(x) = 0.8$; the relations containing 5 are [5, 6] and [4, 5, 6] and we have

$$0 \le m_{largeValue}([5,6]) + m_{largeValue}([4,5,6]) \le 0.8$$

This gives a considerable degree of flexibility in assigning mass. The alternative interpretation used here - the closed world approach - regards partial membership of a tuple x in a relation R (i.e. $0 < \cap_R(x) < 1$) as *strictly equal* to the total mass assigned to the sets of tuples which include x, i.e.

$$\sum_{\substack{t=[x1,\ldots,xn] \\ x \in t}} m(t) = \cap_R(x)$$

This means there is no flexibility in the range of mass distributions that can be derived from the fuzzy relation R. However, there is flexibility in the mass assignments when R is combined with an assignment corresponding to another relation, for example in calculating association confidences as described later.

Under this interpretation, the monadic fuzzy relation *largeValue* discussed above has

$$m_{largeValue} = \{[6]:0.2,\ [6,5]:0.5,\ [6,5,4]:0.3\}$$

and clearly

$$m_{largeValue}([5,6]) + m_{largeValue}([4,5,6]) = 0.8$$

2.5 Closed world Mass-based Association Rules

For a source category

$$S = \left[x_1/\chi_S(x_1),\ x_2/\chi_S(x_2),\ldots,\ x_{|S|}/\chi_S(x_{|S|})\right]$$

and a target category

$$T = \left[x_1/\chi_T(x_1),\ x_2/\chi_T(x_2),\ldots,\ x_{|T|}/\chi_T(x_{|T|})\right]$$

we can define the corresponding mass assignments as follows. Let the set of distinct memberships in S be

$$\Lambda_S = \left\{\chi_S^{(1)}, \chi_S^{(2)}, \ldots, \chi_S^{(n_S)}\right\}$$

where

$$\chi_S^{(1)} > \chi_S^{(2)} > \ldots > \chi_S^{(n_S)}$$

and $n_S \le |S|$

Let

$$S_i = \left\{\left[x \,\middle|\, \chi_S(x) \ge \chi_S^{(i)}\right]\right\}$$

Then the mass assignment corresponding to S is

$$\left\{S_i : m_S(S_i)\right\}, \quad 1 \le i \le n_S$$

where

$$m_S(S_k) = \chi_S^{(k)} - \chi_S^{(k+1)} \qquad (7)$$

and we define

$$\chi_S^{(i)} = 0 \quad if \quad i > n_S$$

For example, the fuzzy category

$S = [a/1,\ b/0.8,\ c/0.5,\ d/0.2]$

has the corresponding mass assignment

$$M_S = \{[a]:0.2,\quad [a,b]:0.3,\quad [a,b,c]:0.3,\quad [a,b,c,d]:0.2\}$$

We can now calculate the confidence in the association between the categories S and T using mass assignment theory. In general, this is not a unique value as we are free to move mass (consistently) between the cells corresponding to S_i and T_j for each i, j.

For two mass assignments

$$M_S = \left\{S_i : m_S(S_i)\right\}, \quad 1 \le i \le n_S$$

$$M_T = \left\{T_j : m_T(T_j)\right\}, \quad 1 \le j \le n_T$$

the composite mass assignment is

$$M = M_S \oplus M_T = \left\{X : m(X)\right\}$$

where m is specified by the composite mass allocation function, subject to

$$\sum_{j=1}^{n_T} m_{ij} = m_S(S_i) \qquad \sum_{i=1}^{n_S} m_{ij} = m_T(T_j)$$

This can be visualised using a mass tableau (see [3]) as shown in Fig 3. Each row (column) represents a relation of the source (target) mass assignment. We label the rows $S_1, S_2, \ldots S_{nS}$ and columns $T_1, T_2, \ldots T_{nT}$, and assign mass m_{ij} to cell *(i, j)* subject to row and column constraints. The confidence in the association rule is given by

$$conf(M) = \frac{\sum_{i,j}\left(m_{ij} \times |S_i \cap T_j|\right)}{\sum_{i=1}^{n_S}\left(\sum_{j}^{n_T} m_{ij} \times |S_i|\right)} = \frac{n}{d}$$

$$where \quad n = \sum_{i,j}\left(m_{ij} \times |S_i \cap T_j|\right)$$

$$d = \sum_{i=1}^{n_S}\left(\sum_{j}^{n_T} m_{ij} \times |S_i|\right) = \sum_{i=1}^{n_S} |S_i| \times m_S^{(i)} \qquad (7)$$

ISBN: 978-989-95079-6-8

Clearly $n \geq 0$, $d > 0$ and d is a constant for a given source relation S, irrespective of T and M.
For example consider the fuzzy categories

$S = [a/1, b/0.8, c/0.5, d/0.2]$ and

$T = [a/1, b/0.4, c/0.8, d/0.7]$

One notable assignment is the least prejudiced distribution, obtained by taking the product of source and target masses for each cell as shown in Fig 3. This corresponds to the minimum entropy combination of the two mass assignments.

		0.2 [a]	0.1 [ac]	0.3 [acd]	0.4 [abcd]
0.2	[a]	[a] 0.04	[a] 0.02	[a] 0.06	[a] 0.08
0.3	[ab]	[a] 0.06	[a] 0.03	[a] 0.09	[ab] 0.12
0.3	[abc]	[a] 0.06	[ac] 0.03	[ac] 0.09	[abc] 0.12
0.2	[abcd]	[a] 0.04	[ac] 0.02	[acd] 0.06	[abcd] 0.08

Figure 3: mass tableau, showing intersections $S_i \in T_j$ and the least prejudiced mass distribution. The corresponding point valued rule confidence is 1.86 / 2.5 = 0.744

It is possible to show [8] that maximum rule confidence is obtained by moving mass towards the diagonal (top left to bottom right) and that minimum confidence is obtained by moving mass towards the bottom left - top right diagonal, as illustrated in figs 4 and 5.

		0.2 [a]	0.1 [ac]	0.3 [acd]	0.4 [abcd]
0.2	[a]	[a] 0.2	[a] 0	[a] 0	[a] 0
0.3	[ab]	[a] 0	[a] 0.1	[a] 0.2	[ab] 0
0.3	[abc]	[a] 0	[ac] 0	[ac] 0.1	[abc] 0.2
0.2	[abcd]	[a] 0	[ac] 0	[acd] 0	[abcd] 0.2

Figure 4 : mass tableau, showing intersections $S_i \in T_j$ and the mass distribution leading to minimum rule confidence 2.1 / 2.5 = 0.84

		0.2 [a]	0.1 [ac]	0.3 [acd]	0.4 [abcd]
0.2	[a]	[a] 0	[a] 0	[a] 0	[a] 0.2
0.3	[ab]	[a] 0	[a] 0.1	[a] 0.1	[ab] 0.2
0.3	[abc]	[a] 0	[ac] 0.1	[ac] 0.2	[abc] 0
0.2	[abcd]	[a] 0.2	[ac] 0	[acd] 0	[abcd] 0

Figure 5 : mass tableau, showing intersections $S_i \in T_j$ and the mass distribution leading to minimum rule confidence 1.5 / 2.5 = 0.6

2.6 Membership Function for Fuzzy Confidence

We define the membership function in terms of the mass which must be moved (relative to the least prejudiced distribution, where confidence has membership 1). Any other assignment of mass requires one or more elementary mass transfers relative to the LPD, and we are particularly interested in mass assignments corresponding to minimum and maximum confidence, M^{MIN} and M^{MAX}. We define a fuzzy interval C representing the confidence such that

$$\mu_C(conf(M)) = 1 - \frac{pos(M^{LPD} - M)}{N}$$

where $$N = \max\left(pos(M^{\max} - M^{LPD}), pos(M^{\min} - M^{LPD})\right)$$

Because the membership varies linearly with the amount of mass moved, it is triangular and can be calculated quickly by considering the end points. We note that it is possible for the membership function to be discontinuous at one end (i.e. to drop abruptly to zero). Fig 6 shows the membership function for the fuzzy confidence in the *good sales - high salary* example, with the calculations shown in figs 3, 4, and 5.

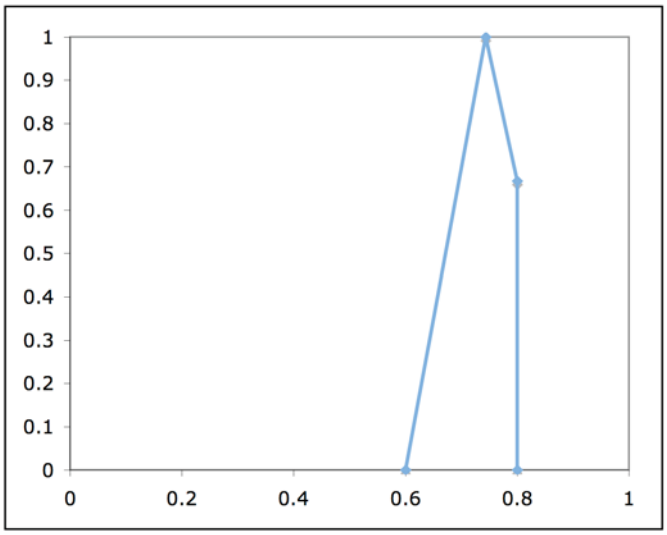


Figure 6 : Membership function for fuzzy confidence in the good sales - high salary example. Note the discontinuity at 0.8

3 Demonstrator Application

We have applied the algorithm to calculation of associations in an integrated taxonomic database of terrorist incidents, as described in [7]. Our primary source is the Worldwide Incidents Tracking System (WITS) [20] augmented by additional information from the MIPT Terrorism Knowledge Base (TKB) database[†]. WITS consists of incidents in which *"subnational or clandestine groups or individuals deliberately or recklessly attacked civilians or noncombatants (including military personnel and assets outside war zones and war-like settings)"*. We are not concerned with the correctness or otherwise of this definition, and treat the data simply as a testbed for SQuAD .

Once data has been integrated and categorised into the iPHI hierarchies, the methods outlined in this paper and [20]

[†] TKB ceased operation on 31 March 2008 and is now part of Global Terrorism Database (GTD) at www.start.umd.edu/data/gtd/

ISBN: 978-989-95079-6-8

have been applied to find significant associations between categories in different hierarchies (e.g. *geographic region*, *weapon type, perpetrator, casualty level* etc). Since the data is not static, we cannot assume that significant associations remain significant – indeed, valuable insight arises from changes in association levels relative to other associations, and trends in the strength of an association. 15,900 incidents between January 2005 and January 2006 were analysed.

Various associations can be extracted by consideration of fuzzy categories in different taxonomies. Although the vast majority of results from [7] led to reasonable intervals, there were a few cases in which intervals were quite large. As can be seen from Fig 7, much smaller intervals were obtained under the closed world method. The interval for closed world rules indicates the extremes of the fuzzy association confidence, with a symmetric triangular membership function (possibly truncated on one side). The interval for open world calculations has membership 1 throughout.

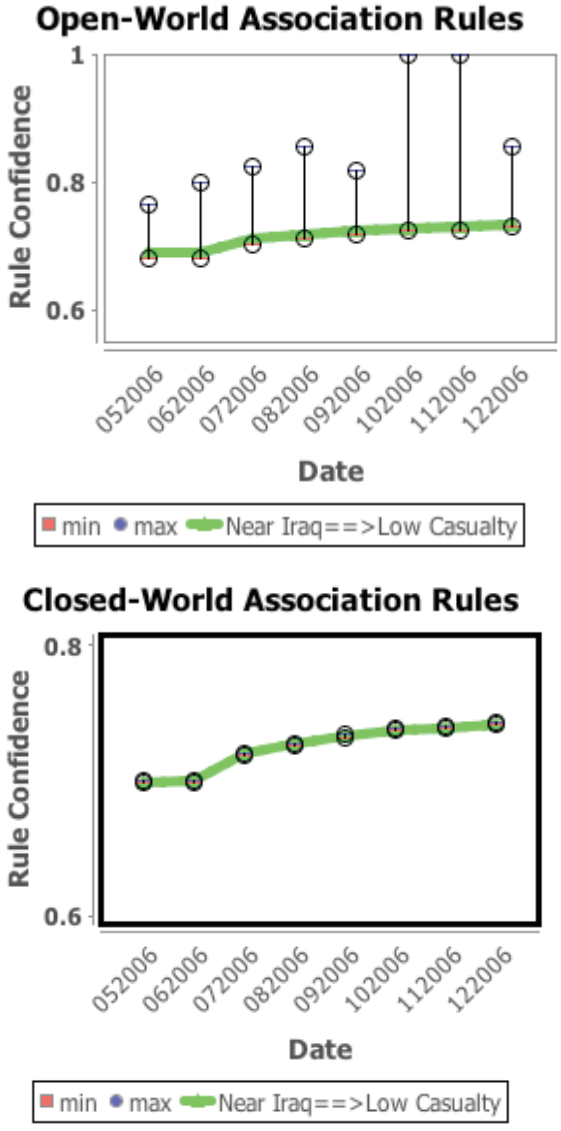


Figure 7 : Confidence intervals and point values for associations between the fuzzy categories *in/near Iraq* and *Medium casualty levels*. The fuzzy confidence (closed world) gives a much tighter interval than the open world calculation

4 Summary and Future Work

We have described a new method for calculating the confidence in an association rule between two (or more) fuzzy categories, based on mass assignment theory and yielding a fuzzy interval within which the confidence must lie. Combined with the point value confidence described in previous work, this enables us to order and plot association strengths with an indication of the degree of uncertainty.

Acknowledgment

This work was partly funded by BT and the Defence Technology Centre for Data and Information Fusion.

References

[1] Berners-Lee, T., J. Hendler, and O. Lassila, "The Semantic Web," in *Scientific American*, vol. 284, 2001, pp. 28-37.

[2] Zadeh, L. A., "Fuzzy Sets," *Inf. and Control*, 8, pp. 338-353, 1965.

[3] Baldwin, J. F., "Management of Fuzzy and Probabilistic Uncertainties for KB Systems," in *Encyclopedia of AI*, S. A. Shapiro, Ed., 2nd ed: John Wiley, 1992, pp. 528-537.

[4] Baldwin, J. F., "Mass Assignments and Fuzzy Sets for Fuzzy Databases," in *Adv. in the Shafer Dempster Theory of Evidence*, M. Fedrizzi, J. Kacprzyk, and R. R. Yager, Eds.: Wiley, 1994.

[5] Baldwin, J. F., T. P. Martin, and B. W. Pilsworth, *FRIL - Fuzzy and Evidential Reasoning in AI*. Wiley, 1995.

[6] Martin, T. P., B. Azvine, and Y. Shen, "A Mass Assignment Approach to Granular Association Rules for Multiple Taxonomies," presented at Uncertain Reasoning in the Semantic Web / Semantic Web 2007, 2007.

[7] Martin, T. P. and Y. Shen, "TRACK - Time-varying Relations in Approximately Categorised Knowledge," *Int. J. Comp Intelligence Research*, (to appear), 2009.

[8] Martin, T. P. and Y. Shen, "Fuzzy Association Rules to Summarise Multiple Taxonomies," in *Fuzziness and Scalability*, A. Laurent and Lesot. M-J, Eds.: 2009.

[9] Martin, T. P., Y. Shen, and B. Azvine, "Incremental Evolution of Fuzzy Grammar Fragments to Enhance Instance Matching and Text Mining," *IEEE Transactions on Fuzzy Systems*, 2008.

[10] Martin, T. P. and B. Azvine, "Soft Integration of Information with Semantic Gaps," in *Fuzzy Logic and the Semantic Web*, E. Sanchez, Ed.: Elsevier, 2005.

[11] Martin, T. P. and B. Azvine, "Acquisition of Soft Taxonomies for Intelligent Personal Hierarchies and the Soft Semantic Web," *BT Tech Jl*, 21, pp. 113-122, 2003.

[12] Martin, T. P., B. Azvine, and Y. Shen, "Intelligent Hierarchy Mapping: A Soft Computing Approach," *Information Tech and Intelligent Computing*, 2008.

[13] Bosc, P. and B. Bouchon-Meunier, "Databases and Fuzziness - Introduction," *Int J Intel Systems*, 9, 419, 1994.

[14] Agrawal, R. and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," VLDB, 1994.

[15] Dubois, D., E. Hullermeier, and H. Prade, "A systematic approach to the assessment of fuzzy association rules," *Data Mining and Knowl Discovery*, 13, 167-192, 2006.

[16] Bosc, P. and O. Pivert, "On Some Fuzzy Extensions of Association Rules," proc IFSA , 2001.

[17] Kacprzyk, J. and S.Zadrozny, "Linguistic Summarization of Data Sets Using Association Rules," proc Fuzz-IEEE, 2003.

[18] Martin-Bautista, M.J., M.A.Vila, H.L. Larsen, D. Sanchez, "Measuring Effectiveness in Fuzzy Information Retrieval," Flexible Query Answering Systems (FQAS), 2000.

[19] Zadeh, L. A., "The Concept of a Linguistic Variable and its Application to Approximate Reasoning (Part 1)," *Information Sciences*, vol. 8, pp. 199-249, 1975.

[20] WITS, "WITS - Worldwide Incidents Tracking System," National Counterterrorism Center Office of the Director of National Intelligence 2007.

ISBN: 978-989-95079-6-8

Around fuzziness. Some philosophical thoughts

Alejandro Sobrino M. Pereira-Fariña

Faculty of Philosophy, University of Santiago de Compostela
Praza de Mazarelos, s.n., 15782 Santiago de Compostela, Spain
Email: alejandro.sobrino@usc.es, martin.pereira@rai.usc.es

Abstract— *Fuzzy logic has a considerable success in the area of applications. However, this active growth in the implementation in applied electronic devices has not always been accompanied by a discussion about the theoretical grounds underlying those uses. This paper deals with some theoretical aspects that seem convenient to discuss for the purpose of clarifying the status of fuzzy logic, to get into its true nature and to appoint its achievements and limits. In order to carry out this goal, the paper is organized around three questions: 1) Fuzzy logic, is it a logic or a techno-logic?; 2) How fuzzy is fuzzy computing?; 3) Are there fuzzy objects?*

Keywords— Fuzziness, fuzzy computation, fuzzy objects, philosophy of soft computing, vagueness.

1 Introduction

Borderline cases are initially addressed by Ch. S. Peirce in 1902 [1]. In 1923, B. Russell positioned vagueness as an interesting subject in the philosophy arena [2]. Later, Hempel and Copi dealt with vagueness in separate articles published in the journal *Philosophy of Science* [3, 4]. Up till them, vague language was not considered a central topic for the analytic tradition, a school of thought characterized by an emphasis on clarity and argument, achieved by formal logic and analysis of language. M. Black introduced in 1937 [5] the 'consistency-profiles' as a way to represent the meaning of a vague predicate, defined by the standard use that a community make of it. Consistency-profiles are represented by curves, showing the degree of consistency in the application of a word to an object or concept. The notion involves three characteristics present in the modern forms to represent vagueness: the user of a language, a situation in which the user is trying to apply a symbol to an object and a way to represent the consistency of this application through a logic formalism.

Among the modern initiatives to deal with the formal representation of vagueness, understood as imprecision, the most successful has been Zadeh's fuzzy logic [6]. Fuzzy logic enables to symbolize vague propositions and to manage approximate inferences in a suitable way. As a result, it has had a tremendous growth, supported mainly by the success of their applications. However this active growth in the area of applications has not always been accompanied by a reflection on the theoretical grounds involved in the use of the underlying tool. As an applied logic it was not surprising that, in the beginning, Zadeh claimed that *"clearly the problems, the aims and the concerns of fuzzy logic are substantially different from those which animate the traditional logical systems. Thus, axiomatization, decidability, completeness, consistency, proof procedures and other issues which occupy the center of the stage in such systems are, at best, of peripheral importance in fuzzy logic"* [6, p.151]. And perhaps the same could be said of the metatheoretical discussions. In this paper we will deal with some theoretical problems that we believe that they are important for the purpose of clarifying the status of fuzzy logic, to get into its true nature and to draw its achievements and limits. The fundamental issues considered here take the form of three questions that, in turn, will name the next three sections of this work.

2 Fuzzy logic, in its true nature, is it a logic or rather than a *techno*-logic?

The reflection about the logical statute of fuzzy logic is involved with two traditional topics in Philosophy:

i. The classification of sciences.

ii. The fundamentation of sciences.

The classification of sciences is frequently revised in function of new emergent branches of knowledge, forcing to modify standard taxonomies. In the field of deductive sciences, fuzzy logic came as a new discipline that questioned some of the classical principles, prompting a debate about the nature of the logic and provoking a discussion about the status of the deviant logics, among which fuzzy logic was included [7].

Since its birth, fuzzy logic was valued in different ways by engineers (applied scientists) and by mathematicians (theoretical thinkers). Due to the success of the applications, fuzzy logic was welcomed in Engineering Schools. Nevertheless, it also caused rejection among mathematicians because, during a long time, it was a matter under construction, provisional and tentative.

Engineers use fuzzy logic according, at least, three criteria:

- Fuzzy logic allows us to represent complex scenarios that involve uncertainty using few rules. In the same vein, it permits to manage approximate inferences reaching good and quite stable solutions to the pretended conclusions.
- Although fuzzy logic is a (precise) logic about imprecision, it manages outputs that provides enough information to make good decisions. As Popper says: *"Although clarity is valuable in itself, exactness or precision is not: there can be no point in trying to be more precise than our problem demands"* [8, p.28]. So, in presence of imprecision, alternatives to the classical bivalent logic should be suggested.
- Systems with incorporated fuzzy rules are quite robust. Fuzzy logic admits a tolerance fringe between positive and negative outputs, although it is difficult to supply *a priori* guaranties about the stability of a fuzzy controller.

ISBN: 978-989-95079-6-8

These points agree with the Zadeh's *dictum*, known as 'Principle of Incompatibility': *"Stated informally, the essence of this principle is that as the complexity of a system increases, our ability to make a precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics"* [9, p. 28].

In summary, from the applied sciences point of view, as engineering, fuzzy logic is a useful discipline. Its success are the achievements of its applications and any further evidence is not requested about the logical properties of the formal tool. For engineers, fuzzy logic is a techno-logic, i.e., a logic to be implemented in electronic machines.

But beyond applications, fuzzy logic is a kind of logic and, therefore, a branch of mathematics. Mathematics is a theoretical science, not an applied one and its currency are precision, rigor and proof. So, frequently, fuzzy logic was not welcomed in mathematical circles while it was used as an *ad hoc* tool for solving quite particular problems in a large extent and specially in early applications. This partial and temporary knowledge was not well received in the field of mathematics, accustomed to dealing with eternal truths. Attending to its provisional character, mathematicians think that fuzzy logic is not a logic, but a kind of formalism for managing specific problems that involves imprecision; fuzzy logic, they say, does not manage any rationality, even not that it say to grasp: the imprecise or vague reasoning.

Attending to classical logic, the Verificability Principle allows us to substitute a proposition by its truth value. But in the field of fuzzy logic, this unanimous correspondence is problematic. A vague predicate can be represented, in a reasonable way, by similar linguistic labels or fuzzy numbers. As a matter of degree, fuzzy meaning is grasped in an imprecise way; hence, some margins of tolerance should be admitted in order to represent it. Accordingly, in fuzzy logic it is not possible to forget the world and focus only in the proofs. We need to reconcile both of them. Any attempt to give a fuzzy logic axiomatic approach hits with the absence of the stability of truth, with the lack of a perfect truth. Therefore, for mathematicians, fuzzy logic is not a true logic.

From the point of view of the fundamentation of sciences, logic is a main part of the scientific method. The hypothetico-deductive system is considered, frequently, as the scientific method *par excellence*. From Aristotle, Science is strongly related to deductive reasoning. In [10], it is defined a deductive science as a set S of sentences that satisfies four conditions:

1. *Reality postulate*: Sentences of S are about a particular subject matter, about real entities.

2. *Truth postulate*: Each sentence of S is true.

3. *Deductive postulate*: If something can be deduced from a subset of S, then it is also a member of S –closure property under logical consequence.

4. *Evidence postulate*:

   (a) *Definability*: The sentences of S are built from a finite set of terms, which have a meaning immediately obvious.

   (b) *Axiomatizability*: The truth of every sentence of S can be inferred logically from a certain finite subset S_1 of S whose truth is also obvious.

Postulates 1 and 2 require comments. It is known that both mathematics and physics are scientific theories. But while mathematics is an ideal science, physics is an experimental one. So, the objects of mathematics are not real objects, but ideal entities. Therefore, their statements can be absolutely true; timeless true. This forces to not cast doubts about the postulate 3. The simplification of the reality and the idealism related to the mathematical thinking lead to not questioning the postulate 4. But empirical statements are about real entities, and material objects change and vary over time. So, postulate 2 –the *truth postulate*– is not immediately obvious. Sometimes empirical sentences are true, sometimes are partially true (true in a degree, true with a probability,...) but they are always temporarily true; that is, true in absence of falsation. Postulate 2 is known as the Principle of Bivalence and it was traditionally rejected in other applications which are different from the logic of mathematical entities.

In twentieth century, with the birth of artificial intelligence, a new way to see deductive sciences came. The logic is not only applied to specific languages, such as the mathematical language –crisp in nature–, or to regimented natural language –specified artificially–, but also to statements that verbalize normal situations related to ordinary life. So, the *Reality Postulate* was extended to any real entity, including these with ill-defined borders and the *Principle of Bivalence* was replaced by the *Principle of Valence* [11], which states that every proposition has a truth value, not fixing how many or which. The relaxation of principles 1 and 2 casts doubt on the fourth: in real life it is difficult to think about meanings or truths immediately obvious. More attention is demanded by the third principle. Fuzzy logic is a deductive logic that grasps imprecise consequences. As a logic, it aims to conclude deductively, but as a theory of imprecise consequences, calls for margins of tolerance or imprecise intervals of plausible conclusions. Whether fuzzy logic verifies, or not, the deductive postulate depends on how the meanings of vague sentences are represented, as well as the approach to combine them into complex propositions. For example, Trillas et al. showed that the unique way to obtain consequences in the Tarski's sense from the Zadeh's compositional rule of inference is using the bold t-norm if the rule is represented by the Reichenbach's conditional definition [12].

From the beginning until now, fuzzy logic has evolved dramatically. Today it is still a formal tool with successful applications, but also a logic that seeks its theoretical consolidation. In this goal highlights –among others– the contributions made by the Spanish school of fuzzy logic, particularly those made by Esteva & Godo and Trillas. Esteva & Godo deal with fuzzy logic as a symbolic logic developed in the spirit of classical logic (syntax, semantics, axiomatization, truth-preserving deduction, completeness, etc. both in propositional and predicate logic) [13]. Hajek and Novak et al. contributed too to credit fuzzy logic in this vein [14, 15]. Hence, their work was directed to confer credit to fuzzy logic as a kind of logic. Trillas has been from a long time devoted to evolve a fuzzy metatheory of fuzzy logic; so, he has contributed to develop fuzzy logic as a fuzzy theory [12]. All of them, helped to

ISBN: 978-989-95079-6-8

credit fuzzy logic as an area of interest in the contemporary scientific panorama.

Despite these developments, fuzzy logic needs to assume new challenges. In the field of applications is a truism the success of the fuzzy controllers in many household devices. But to avoid falling in auto-complacency, it would be desirable to develop applications where vagueness or imprecision are the main component of a complex system; e.g., implementing fuzzy logic into devices in order to summarize perceptions [16]. In the theoretical perspective, fuzzy logic needs to debate about the ways to represent the meaning of a vague proposition, how to justify one choice or another among the various alternatives. Its credibility will be critically linked to this task. Another job will be to mix fuzzy logic with others logics while imprecision in not the only characteristic of human language or human decision-making. Context, time, intentionality, beliefs and its corresponding logics are jointly relevant to model human reasoning and, therefore, to build electronic devices that simulate human behavior.

The new fuzzy logic proposed by Zadeh under bright titles as 'Computing with Words and Perceptions', 'Precisiated Natural Languages', 'Protoforms' and so on [17], should make progress in this regard. Until now we have ashy headlines but low content.

3 How fuzzy is fuzzy computing?

'Turing machine' (TM) is perhaps the main reference in computer science. So, it is said that any function is computable if it is Turing computable. If a problem is computable, it admits an algorithmic definition; i.e., its solution can be reached by a finite sequence of instructions that ends in a finite time.

A TM requires that:

- Each cell of the tape contains only a symbol, although there can be an infinite number of cells.
- States are finite and discrete: the machine is in one state or another, but never in two states at the same time or in no state.

According to the discrete nature of the alphabet and the states, the computation performed by a TM is linear and precise. But although the representation and the computation of a problem are bivalent in nature, the output may be trivalent: in effect, a TM computation can conclude in three different states:

- Final state or success: The problem is decidable and it has a solution.
- Non-final state or failure: The problem is decidable but it does not have solution.
- Loop state or auto-referential state: The machine computes indefinitely and never stop. The problem is indecidable and it does not have any solution. Hence, a TM may have not solution because it stops on a not final state or because it does not stop.

There are so many TM as problems can be putted in an algorithmic form. But a nice achievement of the theory of computation is that any TM admits to be simulated by a special one: the binary TM. A binary TM works as follows: any symbol or state is represented by binary digit or a set of binary digits. So, an instruction as $10 \longrightarrow 10001l$ denote that the machine is in the internal state *1*, reads *0* in the tape, change to the state *1000*, writes *1* and it moves the tape one step to the left. A binary TM can simulate any other specific TM; so, it is an universal computing machine.

At this point, we wish to emphasize the capability of the binary TM to simulate the behavior of any other TM. But the TM not only support be simulated. It is also an excellent simulator; in particular, mimic goods solutions to problems which was not designed to. For example, several TM working simultaneously can imitate a parallel computation, although the way in which TM process is sequential and not parallel.

Fuzzy TM goes back to the late 1960s, when Zadeh spoke about the notion of fuzzy algorithm [18]. Santos gave the formal description of a fuzzy algorithm by fuzzy variants of Turing machines, Markov algorithms and finite automata [19]. He established the fundamentals of the fuzzy language too. Fuzzy TM are linked to classical TM through the support and crisp part of a fuzzy set. Fuzzy algorithms were drawn to perform imprecise computations.

What is meant by imprecise computations? At least, two answers are possible:

- In a soft sense, imprecise computations refer to problems whose working process can be expressed in an algorithm form through imprecise instructions; that is, rules that are or are not completely satisfied, but only fulfilled at some extent. From this perspective, a typical imprecise computation is a cooking recipe: add a pinch of salt, cook quite slowly, serve very cold,...
- In a hard sense, imprecise computations should refer to a genuinely form to compute vague instructions. If so, there would be a new model of computing, but up till today, the TM is the only machine for computing. Afterwards, we come back to this topic.

Imprecise computations in a soft sense can be simulated by a TM. A fuzzy instruction, as "add a pinch of salt', can be represented by a generalized membership function. Fuzzy logic provides a fuzzy arithmetic for managing fuzzy commands. So, a TM with a fuzzy arithmetic module can operate fuzzy computations and output a final state of partial success; that is, a solution qualified by a degree. The appearance of this fuzzy output may be continuous and not discrete. Discrete outputs passing serially, one after the other, in a small temporary sequence, may emerge as a continuous appearance –cartoons are a god metaphor: snapshot sequences of static pictures seem to have movement by themselves.

In a soft sense, a fuzzy computation can be performed by a classical TM, at least, in two ways:

- Weighting the symbols of the tape alphabet.
- Qualifying the states with degrees of preference.

So, from a theoretical point of view, fuzzy computation is classical computation augmented with a module that operates with degrees. "Fuzzy" refers to data high level representation not to a form of computing. Therefore, there is not a new or genuine way of computation [20]. However, from an applied

ISBN: 978-989-95079-6-8

perspective, the important thing is not to valuate the nature of the process, but its outcome. Today there is an enormous amount of computational processes that simulate approximate solutions to interesting problems. Still having the TM as a substrate, all of them have a very realist appearance, as it is showed by the software that control the smoothness of a train brake or the camera that stabilizes the frame for taking a good picture show. From this point of view, fuzzy computation is a fact regardless if the true nature of the computational process is fuzzy or not.

As we have said before, if it is possible to get good simulations of fuzzy instructions using classical methods, would still be desirable a true fuzzy computer? The answer of this question leads to discuss the differences and analogies between the real and the artificial, between the authentic and the simulated, directing, ultimately, the old philosophical question of what is 'being'. But this inquiry is out the limits of this paper.

Computing in the traditional sense involves inputs with numbers and symbols that are crisp. However, human beings use words to solve daily problems that attend to fuzzy rules; so to speak, humans use words to make computations. In the last years, some new idea has being emerging in this field. Fuzzy logic is emphasized to play a key role in granular computing, the methodological base for computing with words and perceptions. So, while in classical theory a sensor records a temperature by a number (30 C^o), in fuzzy theory a device should to convert the numerical input in the name of a perception: its hot. According to Ying [21], the start point of models for computing with words is to deal with fuzzy finite-state and fuzzy pushdown automata through extending their inputs to be strings of fuzzy subsets of input an alphabet. However, a study about the correlation between the strings or words and the symbols as inputs has not taken place.

In the context of translating human perceptions to computational perceptions and in the frame of a practical application, Triviño et al. [16] developed a methodology to generate a linguistic summary from data provided by sensors. A chain of transducers perform this task. A 'transducer' is a kind of automata capable of translating perceptions between two levels of granularity. In the project of a Linguistic Fuzzy Transducer the following actions are involved:

- Representing the granular word of the input by its associated vector of linguistic state variables.
- Representing the granular word of the output by its associated vector of linguistic state variables.
- Defining a set of rules that describes the evolution of the system in order to reach the outputs as a function of the values of the inputs.
- Generating a generic representation of the system by the methodology of Zadeh's protoform.

Besides fuzzy computation, there are other initiatives to deal with imprecision. The most remarkable is quantum computing. In quantum physics, vagueness or imprecision take the form of indetermination. But quantum indetermination and linguistic vagueness or imprecision are quite different concepts in nature. So, today quantum computing does not seem an adequate model for imprecise computing. Despite the different nature of vagueness and indetermination, the *continuum* and not deterministic nature of quantum computing encouraged to find alternative models in the general area of imprecision. Although today TM is the computational tool *par excellence*, other emerging tools, impelled by the problem of imprecision, are coming. If they can be an alternative to the classical way of computation is still an enigma, as they are in a starting point.

4 Are there fuzzy objects?

As previously noted, in 1923 Bertrand Russell published a seminal paper about vagueness. In that article, he provided a starting point to introduce the vagueness as a topic of discussion in the field of the analytic philosophy, that attempt to clarify, by analysis, the meaning of statements and concepts. After Russell, other philosophers like Hempel, Black, Dummet or Fine wrote about this subject [22].

Russell discussed the role of vagueness in the language and wrote: *"Vagueness and precision alike are characteristics which can only belong to a representation, of which language is an example. The have to do with the relation between a representation and that which it represents. Apart from representation, whether cognitive or mechanical, there can be no such thing as vagueness or precision; things are what they are, and there is an end of it"*[2, p. 2]. So, vagueness is confined to the realm of the language and excluded as a property of the objects.

Saying that there are only vague sentences and not vague objects seems to be a truism. Objects are what they are. Only the way to access them by the language is imperfect or imprecise in nature. In the vein of positivism, some philosophers thought the opposite: so, the first Wittgenstein said in the *Tractatus* that sentences, if they are to mean anything, must mirror reality in the same way that a picture does. But even in natural sciences, words picture objects in different ways: in that task, some words are more vague that others.

So, we can say unequivocally that the Everest is higher than Mont Blanc. But is it is intrinsically imprecise where the Mont Blanc starts or which area encompasses. A cloud is too a typical case of object with fuzzy boundaries. Quantum mechanics brought a new field of examples of vague objects. The 'spin' is a fundamental property of atomic nuclei, hadrons, and elementary particles. 'Having a spin direction x' is a property clearly defined and it is possible to know exactly when a particle is or not in a given spin direction. Making an experiment we can identify a number of electrons that have this property and the result obtained will be the same if the experiment is repeated. Nevertheless, we cannot notice if the particles that had the property in the first experiment are the same as the second. The particles fall or not under the extension of the predicate 'Having a spin direction x', but it is impossible to fix the identity of the resulting set.

We can refer to vague objects from three points of view:

- Objects are well defined but there is not enough information or appropriate knowledge for grasping them precisely.
- Objects are well defined but it is not possible to know their true being.
- Objects are ill defined because we cannot observe them

ISBN: 978-989-95079-6-8

directly and the instruments to watch them distort irremediably their nature.

The first two points are related to the representation of reality through language. The third refers to the essential nature of the objects. We approach them in an inverted order.

The mental experiment of Schrödinger's cat provides a good example of essential vagueness: suppose a cat into a box. Suppose you cannot see in it and the cat cannot see out. Into the box there are a rock, a Geiger counter and a vial of poison. Now, the rock is slightly radiactive, with an exactly 50% chance of emiting a subatomic particle during an hour. If the rock emits a particle, the Geiger counter activates a device that breaks the vial of poison, killing the cat. Quantum mechanics suggests that after an hour the cat is simultaneously alive and dead, so it is an indeterminate or vague object.

From a philosophical point of view, an essentialist perspective is approached by G. Evans, who address the impossibility of vague objects in the frame of the Kripke's theory of rigid designators. A term is a rigid designator if it always designates the same object in any world where the object exists. Proper nouns are rigid designators. Like this, we can imagine a world where the name 'cat' do not refer to a domestic, hairy, medium size, enemy of the mice,..., animal, but we cannot imagine that the name 'Lotfi Aliasker Zadeh' do not refer to Lotfi Aliasker Zadeh. If a word denotes always to the same object in every possible world, we can say that there are a strong correspondence between the word and the object: when we are talking about something, we can say that the words replace objects and the words have an unambiguous and precise meaning. Evans uses this preconception for showing the impossibility of vague objects.

In a short paper published in *Mind* in 1978, he introduced the question about the existence of vague objects [23]. The underlying question posed under the title: *Can there be vague objects?* This question adress the possibility of a vague identity relationship when both sides of the equal sign are occupied by terms that are rigid designators. Evans answers 'no' arguing the following proof (see table 1): Denote ∇ 'indeterminate',

Table 1: Evans' Theorem.

(1)	$\nabla(a = b)$	$a = b$ is indeterminate	(assumption)
(2)	$\lambda x[\nabla(a = x)]b$	b is indeterminely equal to a	(from 1)
(3)	$\neg\nabla(a = a)$	$a = a$ is determinate	axiom
(4)	$\neg\lambda x[\nabla(x = b)]a$	a is not indeterminely equal to b	(from 3)
(5)	$a \neq b$	a is not equal to b	(from 2 and 4)

Evans' argument shows that there are not vague entities if the terms are rigid designators: briefly, if it is indeterminate that *a* is equal to *b* and it is determinate that *a* and *a* are equals, whatever *a* be, *a* and *b* differ at least in one property; so, are different. But the Evans' argument entails a mistake: as D. Lewis point out in [24], Evans' proof start attaching vagueness to the identity relation involving rigid designators. But the own proof demonstrates that the premise is false; therefore, the starting point of the argument was inadequate. Evans' argument shows the inconvenience to judge the imprecision in a small subset of the language –that of rigid designators–, and the mistake to extend this conclusion to the overall language.

In general, language is imprecise in nature and its formal approach, a matter of degree. Furthermore, some words are essentially vague. In the sequel, we analyse the second way in which an object can be vague, according to the bottom clasification. There are objects of which we know all that it is required for knowing their meaning, for attributing them a valuation, but, in some cases, we cannot do that in a unique way. A 'Bald man' or a 'heap of sand' are some examples.

Sorites or *Falakros* is a well-known paradox since Greek philosophers that comes from the use of predicates as 'bald' or 'heap' in the argumentation. To analyse a *Sorites* argument with a classical look leads to paradoxical and unacceptable conclusions. Note that baldness can be mainly referred in terms of hairs and their distribution, that is, in terms of numerable instances. So, a man with 0 hairs is absolutely bald and a man with 10^{10} hairs, not. But counting does not ever clarify the status of a vague predicate. Adding to every man that is bald one hair does not modify its attribute and preserve bald. But the key point is how many hairs are neccesary to change a man from bald to not bald. If you choose a number, I can select another very close to the previous and to claim the truth with the same legitimacy. There is not a bald object that determines, in a unanimous way, a unique use of the word 'bald', but only different uses depending on the context. Different uses mean diverse representations, leading to variations –between some margins– in the conclusions reached by argumentations. Trillas and Urtubey proposed a reasonably cut point that divide when a person is going to be more bald that not bald [25]. But following a kind of metatheoretical game and applying a type-*sorites* reasoning to it, the solution comes irremeably affected by high-order vagueness: the objection that emerge is why not suggest a very similar, but distinct, cut point, to the one supplied. So, it seems quite difficult to solve problems of vagueness with precise measures.

Finally, there are objects well defined but there is not enough information or appropriate knowledge for grasping them precisely; that is, objects that are imprecisely verbalized. Linguistic imprecision should be analysed with fuzzy armamentarium. Fuzzy logic provides tools for managing problems that involve imprecise statements or approximate reasoning. In the field of fuzzy logic, there is not a debate about what is a vague object but how to represent words with vague meaning. From some philosophical traditions (phenomenology, e.g.), an object is how we describe it using words in context. Objects are represented by words, and sentences are represented by logic. If objects have a crisp shape, perhaps precise sentences can designate them and classical membership functions should be employed to formalize them. If the objects have loose frontiers, the way to speak about them are vague in nature and generalized membership functions should be used.

For fuzzy logic there are vague objects as there are fuzzy representations of them. There is no other consideration about its true nature. Fuzzy logic is not a physical theory, but a formal one. So, it does not pretend to reveal the essential nature of the objects, but knowing how people verbalize and represent them in a manageable way in order to solve problems. For fuzzy logic, fuzzy objects are mental or physical objects acceded by sentences uttered in a culture or context that point out the consensus required for an adequate formalization.

Fuzzy logic repairs on linguistic vagueness, not vague objects, anything they that be. Therefore, words or predicates

ISBN: 978-989-95079-6-8

that are instead of the objects are the objective of fuzzy logic. So, its function is to be faithful to how that ordinary people describe reality trough language, generating models to catch the meaning of the predicates according their use in context, generalising solutions whenever possible, making hypothesis on its explanatory character and contrasting the reached conclusions with those awaited. This is a dialectical process, typical of experimental sciences. As a formal tool, fuzzy logic must pursue eternal truths –even imprecise truths–, but like an experimental one, facts must show that truth is obtained only in a degree.

In addition to pursuing affordable challenges, as providing models for a large number of vague sentences in real scenarios, fuzzy logic must address others questions that are still elusive. One of them is how to justify and to manage a suitable specification of the context, basic for representing, in an adequate way, the meaning of vague predicates. There are some formal theories of context, but seems too complex to be manageable or too simple to account the complexity of the meaning. Other challenge is how to summarize a complex perception, where attention is frequently directed to an issue, with a unique membership function. Those are applied inquiries that, however, also require a theoretical debate.

5 Final Remarks

Philosophical discussion about fuzzy logic and its applications is convenient. The popularity of fuzzy logic is based, in a large extent, on the success of its applications. Applications show the working face of a theory, but its consolidation and credit come through the development of a meta-theory, that shed light over its foundations and limits. More work should be done in this direction.

Although vagueness is treated from fuzzy logic mostly by the side of applications, theoretical studies are useful if shed intriguing doubts or suggest illuminating challenges to applied developments. Theory without applications is lame, but applications without theory are blind.

So, to justify the election of a membership function consistent with the true context of a used word or to argue about alternatives for combining atomic sentences into a composite one are tasks that will help to approach more credible solutions to problems every time more generals; so, to transfer credit to fuzzy logic as logic.

The discussion about alternatives to classical computation and the characteristic of a genuine fuzzy TM serve to distinguish the real and the imitated behavior. Also it is interesting to emphasize the simulation power of the classical TM to get fuzzy outputs managing approximate algorithms with crisp implementations.

Finally, to consider how we represent objects permits us to focus on tools as fuzzy logic that manage the meaning of the words that verbalize not only objects, but objects in context.

Acknowledgments

The authors would like to acknowledge the valuable comments provided by the anonymous referees. Partially supported by HUM2007-66607-C04-02 grant from the Spanish Ministry of Education and Science, PEIC09-0196-3018 grant from the Autonomous Government of Castilla-La Mancha and by the Spanish Ministry of Science and Innovation (grant TIN2008-00040 and the FPU Fellow Program).

References

[1] J. M. Baldwin (ed.), editor. *Dictionary of Philosophy and Psychology*. MacMillan, 1902.

[2] B. Russell. Vagueness. *The Australasian Journal of Psychology and Philosophy*, 1:84–92, 1923.

[3] G. Hempel. Vagueness and logic. *Philosophy of Science*, 6, 1939.

[4] I. M. Copi. Borderline cases, vagueness and ambiguity. *Philosophy of Science*, 6, 1939.

[5] M. Black. Vagueness: An exercise in logical analysis. *Philosophy of Science*, 4:427–455, 1937.

[6] R.E. Bellman and L.A. Zadeh. Local and fuzzy logics. In J.M. Dunn and G. Epstein, editors, *Modern Uses of Multiple-Valued Logic*, pages 103–165. D. Reidel, 1977.

[7] S. Haack. *Deviant Logic, Fuzzy Logic: Beyond the Formalism*. The University of Chicago Press, Chicago, 1996.

[8] K. Popper. *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge Classics, London, 1963.

[9] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision process. *IEEE Transactions Systems, Man, and Cybernetics*, SMC-3:28–44, 1973.

[10] Aristotle. *Prior and posterior analytics*. Clarendom Press, Oxford, W. Ross edition, 1949.

[11] D. H. Sandford. Competing semantics of vagueness: many values vs. super truth. *Synthese*, 33(2-4):195–210, 1937.

[12] E. Trillas et al. *Introducción a la lógica borrosa*. Ariel, Barcelona, 1995.

[13] F. Esteva et al. On the standard and rational completeness of some axiomatic extensions of the monoidal t-norm logic. *Studia Logica*, 71:199–226, 2002.

[14] P. Hajek. *Metamathematics of fuzzy logic*. Kluwer, 1998.

[15] V. Novak, I. Perfilieva, I. Mockor, editor. *Mathematical Principles of fuzzy logic*. Kluwer, 2000.

[16] G. Triviño and A. van der Heide. Fuzzy languages and their relation to human and machine intelligence. In *Proc. of the Conf. IPMU*, 2008.

[17] L. A. Zadeh. A new frontier in computation - computation with information described in natural language. In *3rd International IEEE Conference Intelligent Systems*, pages 5–6, 2006.

[18] L. A. Zadeh. Fuzzy algorithms. *Information and Control*, 12(2):94–102, 1968.

[19] E. Santos. Fuzzy algorithms. *Information and Control*, 17:326–339, 1970.

[20] C. Moraga. Towards a fuzzy computability. *Mathware and Soft Computing*, 6:163–172, 1999.

[21] M. S. Ying. A formal model of computing with words. *IEEE Trans. Fuzzy Systems*, 10:640–652, 2002.

[22] R. Keefe and P. Smith, editors. *Vagueness: a reader*. The MIT Press, Cambridge, Massachusetts, 1999.

[23] G. Evans. Can there be *Vague* objects? *Analysis*, 38:208, 1978.

[24] D. Lewis. Vague identity: Evans misunderstood. *Analysis*, 48:128–130, 1988.

[25] E. Trillas and L. A. Urtubey. Towards the dissolution of the sorites paradox. *Applied Soft Computing (in press)*.

ISBN: 978-989-95079-6-8

Fuzziness as a Model of User Preference in Semantic Web Search

Veronika Vaneková[1] Peter Vojtáš[2]

1.Pavol Jozef Šafárik University in Košice, Slovakia
2.Charles University in Prague, Czech Republic
Email: veronika.vanekova@upjs.sk, peter.vojtas@mff.cuni.cz

Abstract— *Web 2.0 is based on personalized access to imprecise and incomplete information from heterogeneous sources. In this paper we present a web-based system enabling preference-based search. We use modified Fagin's model of fuzzy preferences based on aggregation of attribute preferences. Aggregation is generated by user ranking of objects. Our contributions are twofold. First we present a formal model and implementation suitable for efficient preferential search. Second we provide several fuzzy measures to evaluate experimental results of our system.*

Keywords— web search, fuzzy logic, fuzzy systems, user preference

1 Introduction

From its birth in 1965, fuzzy logic has spread to many areas. It has influenced logic, databases, neural networks, expert systems, natural language processing and other fields of both theoretical and practical research. Fuzzy logic allows us to model imprecise, vague or uncertain information and to specify vague requirements. Any theory or application can possibly be fuzzified. However, we should first consider the purpose, especially in case of applications. The important question is how much fuzziness we really deal with and where it comes from.

In this paper we propose one such application of fuzzy logic: we use fuzzy sets to represent a notion of user preference in the process of web search. Web search tasks are recently changing from finding a document to solving specific problem for specific user. A typical problem is to buy some item online, make a flight reservation, find a transportation line, apply for a job. Most of these problems actually require searching for objects from some *domain* (e.g. flights, books, notebooks, job offers). The search depends on various attributes of objects (e.g. price, destination, date of departure, airlines and class in the domain of flights). Fulltext search often fails to support such tasks.

Therefore we search data from one arbitrary, but fixed domain. Data is extracted from heterogeneous sources on the web and stored in a standard database or ontology. Our extraction tool does not create fuzzy data. Although the formal model can be easily extended to handle fuzzy data, we focus on the situations where the main source of fuzziness is not in the data itself, but in the imprecise way of human thinking and specifying requirements. Most web users do not know precisely what they want to find, but they are able to distinguish between better and worse objects. Hence we allow users to rate objects and we use inductive learning tool [1] to adjust user preferences. But users also have an opportunity to specify their preferences precisely by means of a special graphical interface. The best objects can be retrieved with top-k threshold algorithm (see [2, 3]).

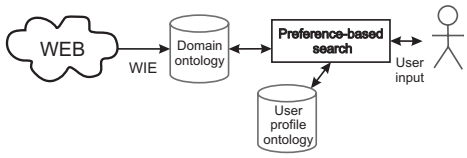


Figure 1: Overview of User-Dependent Search System.

This paper investigates how fuzzy logic can be used to make web search easier and more suitable for imprecise requirements. We compare our approach with other systems and formal models and point out the features necessary for Web 2.0 applications [4] – modeling vague preferences, dealing with huge amounts of data and ordering search results by degree of preference. We also analyze user data from our system with various measures of fuzzy similarity, equality, correlation and we compare the results.

2 Formal Models

Throughout the rest of this paper we will use the domain of job offers which is used also in the implementation. But the formal models described below are suitable for other domains as well. We tested the whole system or its parts in domain of cars, flats and investments.

Web search process is divided into two phases. The first phase is called *WIE* – web information extraction. Information is extracted from HTML pages, each of which contains data for a single object (job offer). Such pages are often generated by server side scripts, so that their structure and HTML tags are similar. The extraction tool [5] compares the page structure with *diff* algorithm and analyzes HTML tags with regular expressions. Differing parts of page sources contain relevant information, most often various attributes of domain objects. We use heuristics (like regular expression patterns for HTML structure) to avoid irrelevant information. Extracted crisp data is stored in domain ontology of job offers. Data repository is the only point that connects extraction phase with searching phase (see Fig. 1). Therefore it is easy to use other extraction tools or search in some existing standard database as well. It is also possible to have fuzzy domain data. Some extraction tools consider reliability of web sources, quality of source data or other conditions to determine fuzzy degree of extracted objects. Our formal model could be extended to consider this type of data as well. Full description of the extraction phase is beyond the scope of this paper, for more information see [6, 7, 5].

ISBN: 978-989-95079-6-8

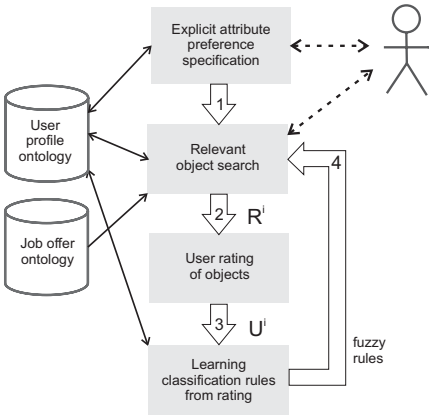


Figure 2: Flow Diagram of User-Dependent Search.

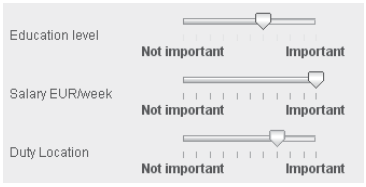


Figure 3: Interface for Attribute Importance.

The second phase, preference-based search, is detailed on Fig. 2. The task of job offer search can be repeated. The system reads user preferences (specified directly) or user feedback (previous search results rated in five-degree scale), updates stored preferences and finds new results. Every such *search cycle* provides more information about the user and helps the system to find more exact preferences.

In the rest of this section we will describe all phases of user-dependent search from Fig. 2 in more detail, together with underlying formal models. We especially emphasize the sources of fuzziness and their influence on the formal model.

2.1 Explicit Preference Specification

Current web portals often provide either a text field for fulltext search or some form to specify required attribute values. Job offer portals usually allow applicants to specify at least place, job position and contract type (i.e. permanent or temporary). But it is impossible to express things like "I prefer York to Birmingham" or "my decision depends on salary more than on location". Our system allows for vague user preferences and provides a suitable graphical interface.

First, the user is presented a list of all *attributes* from the domain of job offers that could influence his decision (like salary, benefits, required education level, required experience, place, job position, contract type, start date). The interface displays a slider marked with "Not important" and "Important" for each attribute (Fig. 3). User can drag the slider to specify how much he cares about the corresponding attribute. The settings from Fig. 3 will produce weighted average

$$@^U = \frac{6 \cdot \mu^U_{education\_level} + 10 \cdot \mu^U_{salary} + 7 \cdot \mu^U_{location}}{23}$$

Weighted averages and other types of aggregation will be described below.

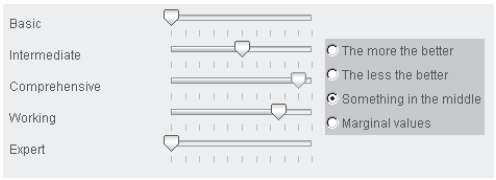


Figure 4: Interface for Attribute Preference.

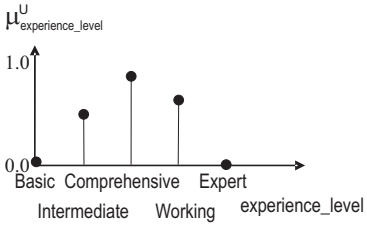


Figure 5: Preference to experience level, acquired from user input.

The next step is to specify preferred attribute values (only for attributes that are marked as "Important" to some non-zero degree). Now the graphical interface contains one slider for every attribute value. Fig. 4 shows the input interface for required experience level. The user prefers those job offers that require comprehensive experience (possibly because he has such experience in the profession).

Formally, user preference related to one particular attribute is a typical example of fuzzy subset. We consider a domain Δ of objects (job offers) and a finite set of domain attributes $\{A_1, \ldots, A_n\}$. Every attribute A_i is a function $A_i : \Delta \longrightarrow \Delta_{A_i}$, where Δ_{A_i} is a set of possible attribute values. Then *attribute preference* for user U and attribute A_i is defined as $\mu^U_{A_i} = \Delta_{A_i} \longrightarrow [0,1]$. Thus $\mu^U_{A_i}$ defines a fuzzy subset of Δ_{A_i}, which is usually a finite set with some natural element ordering.

Fuzzy set $\mu^U_{experience\_level}$ generated from user input on Fig. 4 is shown on Fig. 5. Job offers that require basic or expert experience are unacceptable, while job offers with comprehensive experience are the most preferred. Every attribute preference defines non-strict linear ordering of all objects based on their membership values. The ordering generated by the fuzzy set $\mu^U_{experience\_level}$ would not preserve original ordering of attribute values. Note that some distinct attribute values (e.g. basic and expert experience) have equal membership values. Such objects are *indiscernible* with respect to attribute $experience\_level$.

The most suitable membership functions to represent real user preferences are monotonic and trapezoidal. Some attributes are preferred similarly by all users (e.g. high salary is better than low), however, other attribute preferences may differ significantly (e.g. preference for required language level). These *preference types* can be seen on the right side of Fig. 4.

If we consider more than one attribute, we need to determine the final preference degree for every object, based on attribute preferences. User requirements are often conflicting so that the orderings are different, sometimes even reverse. Therefore we use *aggregation function* to acquire overall ordering. Aggregation function is n-ary monotone function $@^U : [0,1]^n \longrightarrow [0,1]$.

ISBN: 978-989-95079-6-8

Let us consider an object $x \in \Delta$ with attribute values $A_1(x), \ldots, A_n(x)$ and attribute preferences $\mu^U_{A_1}(A_1(x)) = p_1, \ldots, \mu^U_{A_n}(A_n(x)) = p_n$. The global preference degree of x will be $@^U(p_1, p_2, \ldots, p_n)$. A typical aggregation function is weighted average which reflects the idea that some attributes can be more important than the others. Our system acquires the weights from the interface shown on Fig. 3. We will describe other types of aggregation functions in Section 2.3.

2.2 Relevant Object Search

After storing user preferences the system proceeds to the next phase - relevant object search. It is a modification of Fagin's threshold algorithm [3] called *top-k* [2], based on the idea that users are interested only in a few best results from potentially huge dataset and that different users can have different attribute preferences.

Top-k uses a preprocessing phase to generate lists of objects ordered by attribute values. A typical structure for indexing is a B+ tree which allows us to follow records (attribute values) from highest to lowest or vice versa. User preference can define a different ordering of objects, but it can be simulated as well. If the user specified a trapezoidal membership function ("Something in the middle" preference), we start traversing values in both directions from the middle and we merge two non-increasing lists into one non-increasing list. Thus we gain a list of objects ordered from most preferred to least preferred. (The original threshold algorithm considered only one fixed ordering for each attribute.)

If our domain contains attributes $A_1, \ldots, A_n$, we prepare n ordered lists $L_1, \ldots, L_n$. Lists can be stored in heterogeneous repositories, or even acquired from web services. Top-k algorithm uses sorted access to the lists, reads attribute values and calculates preference values. One object can have different position in each ordered list, so some attribute values will be unknown in the meantime. After reading all attribute values of some object x, top-k algorithm calculates aggregation function to determine overall preference value of the object. A threshold value is determined to see if any object with some unknown values has a chance to be more preferred than x. If not, object x is written on the output.

This algorithm is very efficient for complex preferential search. The preprocessing phase causes that we can find k best objects without traversing the whole dataset. The output is an ordered list of k best objects (according to given preferences) together with overall preference values. Despite the fact that source data is crisp, we obtain fuzzy results similar to "object $x1$ is preferred by user U to degree $0,8$". This fuzziness originates from vague user preference representation and it is reflected in the results.

2.3 Learning Classification Rules from Ratings

Although the preference model is fuzzy, it is intended to represent real user preferences as precisely as possible. If they do not correspond, the user obtains different results, or results in wrong order. The graphical interface gives user a chance to rate the results in 5-degree scale (bad, poor, average, good, excellent) as can be seen on Fig. 6. Degrees can also be represented as percents, stars, smilies, etc. Then we use ordinal classification [8] to find a new aggregation function. As we showed in [9], it is sufficient to learn an aggregation function,

Figure 6: Interface to view and rate search results.

provided that types of attribute preferences did not change (e.g. from "The more the better" to "The less the better"). The new aggregation function has a form of *fuzzy rules*.

Fuzzy rules are well-known from fuzzy controllers [10], fuzzy inference systems [11] and many other application areas. A fuzzy rule consists of antecedent and consequent. *Antecedent* is typically a (fuzzy) conjunction or disjunction of fuzzy condition clauses. *Consequent* can be some action (in case of fuzzy controllers) or a fuzzy predicate (in case of fuzzy inference systems).

controller rule: IF temperature IS cold THEN heater IS high

inference system rule: IF wind IS strong AND barometer IS falling THEN forecast IS bad

Different types of rules look syntactically similar, but they differ in the semantics of consequent. In the first case the consequent "heater IS high" means that the heater device will be *set* to high level. In the latter case the consequent "forecast IS bad" means that the variable forecast will be assigned a fuzzy value bad, but no physical action will take place. Condition clauses contain crisp variables (temperature, wind, barometer) and fuzzy sets (cold, strong, falling) in both cases. Connectives AND, OR are fuzzy t-norms and t-conorms, IF-THEN implications are also fuzzified. There is usually a larger set of rules which are processed, their outputs are combined (aggregated) in some way into a single fuzzy set. It is then defuzzified to determine crisp output value. The most common defuzzification method is *centroid calculation*, which returns the geometric center of area under the fuzzy membership function.

Our usage is more similar to fuzzy inference systems because we use fuzzy rules to find a preference value, not to control some electronic device. It is somewhat simplified to reflect our preference model. The following example shows a set of classification rules. We chose fuzzy rules as aggregation method because they are human-understandable and very close to natural way of thinking.

IF $\mu^U_{salary}(salary(x)) \geq 0.7$ AND $\mu^U_{place}(place(x)) \geq 0.6$ AND $\mu^U_{position}(position(x)) \geq 0.5$ THEN $good^U(x) \geq 0.8$

IF $\mu^U_{experience\_level}(experience\_level(x)) \geq 0.3$ AND $\mu^U_{salary}(salary(x)) \geq 0.6$ THEN $good^U(x) \geq 0.6$

It is easy to see some differences compared to fuzzy rules mentioned above. Our fuzzy rules are induced from a small set of rated objects, so they are more specific. Condition clause $\mu^U_{salary}(salary(x)) \geq 0.7$ can be written more intuitively as $salary(x)$ IS good ≥ 0.7. We use arbitrary fuzzy sets $\mu^U_{A_i}$ instead of linguistic variables like "cold".

ISBN: 978-989-95079-6-8

Our condition clauses use α-cut defuzzification and thus our conjunctions and implications are crisp. If object x satisfies the antecedent, then the aggregation function $good^U$ will be greater or equal to the value specified in the consequent. If some object satisfies more antecedents, we choose the greatest output value. If it does not satisfy any rule antecedent, the result will be set to 0. A set of rules makes up a monotone aggregation function. An object y with all attribute preference values greater or equal to object x will be "good" to greater or equal degree than x.

3 Implementation

The system for user-dependent web search is implemented and available on http://x64.ics.upjs.sk:8080/nazou/. It is a web application created in Java with Wicket[1] framework. Java servlets run under Tomcat application server. The server also supports Sesame[2] database to store both our ontologies, i.e. job offer ontology and user preference ontology (see Fig. 2). User preference ontology schema is fixed, while domain ontology is arbitrary (see [12, 9] for details).

The application is divided into relatively independent tools: top-k search, inductive learning, and user preference acquisition called UPreA. UPreA provides user preferences for other tools, processes user inputs and manages all changes in user profile. It communicates with Sesame database through Java API to store and update user information.

4 System and Model Evaluation

Our goal is to experiment with the implemented system and to find out if our model is a suitable representation of real user preferences. As our model is fuzzy, we are interested in fuzzy degree of correspondence between our results and some "ideal" results. Analysis of experiments is further complicated by the fact that user preferences are evolving in time and user inputs can be inconsistent.

We anticipate vague user preferences and therefore we use inductive learning to adjust the aggregation function. Our preferences define an ordering of all objects from the domain, while user ratings define another ordering of some small subset of objects. Fig. 2 shows orderings that are generated in search cycles: R^i is the ordering of search results and U^i is the ordering specified by user ratings. We explore the correspondence of R^i and U^i. If these orderings are similar, the user must have assigned higher ratings to objects from the front of the result list and therefore our preference model reflected real user preferences well. Throughout the rest of this section we assume that R^i and U^i are permutations of objects $\{o_1, \ldots, o_n\}$. Also note that results R^0 are retrieved with user-defined aggregation function (weighted average), while other results R^i, $i \geq 1$ are retrieved with aggregation (fuzzy rules) learned from previous ratings.

The main problem is how to determine the correspondence of two fuzzy subsets. We present several measures below and compare the results. We analyse how R^i corresponds with U^i in six successive search cycles. In fact, some users stopped searching after the first or the second cycle, so the number of analyzed ordered lists is slightly decreasing. Data for this analysis (search results R^i and user ratings U^i) comes from real users who tested the implemented system.

Table 1: Correlations of result ordering and user evaluation.

Cycle	0.	1.	2.	3.	4.	5.
τ-correlation	0.73	0.77	0.79	0,77	0,81	0,81
Result lists	174	134	102	53	26	15

Table 2: Weighted similarity of result ordering and user evaluation.

Cycle	0.	1.	2.	3.	4.	5.
similarity	0.53	0.41	0.39	0,56	0,61	0,64

4.1 Correlation of Orderings

The first possibility is Kendall τ-correlation coefficient [13]. It is defined as $\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} - 1$, where n_c is the number of concordant pairs and n_d is the number of discordant pairs in both ratings. A pair of elements is discordant if their order is reversed. We modified this formula to $\tau = \frac{2P}{\frac{1}{2}n(n-1)}$, so that results belong to interval $[0, 1]$ to conform with our fuzzy model. However, in this case only results above 0.5 mean a positive correlation.

Kendall τ-correlation originally deals with strictly ordered lists. However, in our case many objects from R^i or U^i can share the same fuzzy value. We distinguish ties according to the (strict) order in which they are presented to user. This order in fact depends on indexing methods used in top-k search.

4.2 Weighted order similarity

If the order of some objects change, τ-correlation decreases evenly, regardless of the position where the discordance arose. We can also assume that the position of discordance is important – discordance on the 1^{st} and the 2^{nd} position is more significant than discordance on the 8^{th} and the 9^{th} position. This idea is reflected in *weighted order similarity*. We choose a decreasing vector of weights such that differences between adjacent values are also decreasing, e.g. $w = \langle 20, 15, 11, 8, 6, 5, 4, 3, 2, 1\rangle$ for $n = 10$. Weighted order coefficient is defined as $\mathcal{W} = 1 - \sum_{j=1}^{n} \left| w_{p_j} - w_{q_j} \right|$ where p_j is a sequence number of R^i_j and q_j is a sequence number of U^i_j. Note that this coefficient does not consider all pairs of elements as τ-correlation, only the sequence number of the same element in both orderings.

4.3 Fuzzy Equality

Another suitable measure is *fuzzy equality* [14], which is stronger condition than similarity. In this case we consider fuzzy aggregation values $@^U$, not only orders. We define fuzzy equality of fuzzy sets μ_A, μ_B as: $\equiv^\bullet (\mu_A, \mu_B) = \inf_{x\in\Delta}\{\vee^\bullet(\rightarrow^\bullet (\mu_A, \mu_B), \rightarrow^\bullet (\mu_B, \mu_A))\}$. It means that μ_A, μ_B are equal if $\mu_A(x)$ implies $\mu_B(x)$ and vice versa, for all x from the domain. We can choose different types of fuzzy implication and conjunction, namely Łukasiewicz, Gödel and product. The universal quantifier is always represented as infimum for all elements from the domain.

Finally, we compare results of these fuzzy measures on Fig. 7. τ-correlation and weighted similarity seem to be the most

[1] http://wicket.apache.org/
[2] http://www.openrdf.org/

ISBN: 978-989-95079-6-8

Table 3: Fuzzy equality of result ordering and user evaluation.

Cycle	0.	1.	2.	3.	4.	5.
$\equiv^{\bullet}_{L}$	0.19	0.41	0.32	0,35	0,42	0,42
$\equiv^{\bullet}_{G}$	0.23	0.22	0.14	0,20	0,22	0,24
$\equiv^{\bullet}_{p}$	0.17	0.28	0.16	0,23	0,31	0,27

Table 4: τ_R-correlation of result lists R^i, R^{i+1}.

i	0	1	2	3	4
$\lvert R^i \cap R^{i+1} \rvert$	4,50	6,21	5,95	4,88	5,52
$\tau_{\cap}$	0,65	0,80	0,76	0,63	0,76
τ_R	0,53	0,67	0,64	0,56	0,64

suitable measures. Fuzzy equalities impose too strict conditions on fuzzy values. We do not expect result lists and user ratings to be totally equal because we learn new preferences from their difference. Correlation increases most significantly between 1^{st} and 2^{nd} cycle where we start to use fuzzy rules. So experiments proved that fuzzy rules are suitable representation of user preference.

Another interesting point is that τ-correlation and weighted similarity have the same trends up to fifth cycle, while all fuzzy equalities have the same trends from second to fifth cycle. Łukasiewicz equality has the greatest results and Gödel equality has the smallest results because the corresponding fuzzy implications also have this feature.

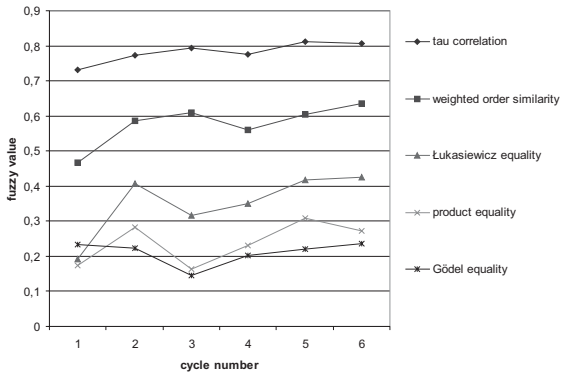


Figure 7: Comparison of different fuzzy measures.

4.4 Correlation of Subsequent Result Lists

We are also interested in the correlation of k-tuples R^i, R^{i+1}. If R^i and R^{i+1} have empty intersection, this correlation will be 0. In case of non-empty intersection, we can compute usual τ-correlation coefficient of the common elements $\tau_{\cap}$ and multiply the result with $\frac{\lvert R^i \cap R^{i+1} \rvert}{n}$. Thus we obtain τ_R. Average results for the first 5 cycles are shown in Table 4.

5 Related Work

In this section we present a short survey of similar preference models and searching methods. Many searching methods allow user to select desired attribute values, but they mostly produce a conjunctive query from user requirements. Our model is more general because aggregation functions are generalization of both (fuzzy) conjunctions and disjunctions.

A similar problem is addressed by *fuzzy multicriteria decision making* [15, 16]. It supports decision based on various attributes called *criteria*. Attribute preferences are usually represented as *fuzzy preference relations* – binary relations defined for every two-tuple of objects. Preference relation defines a degree to which one object is better than another object, based on the corresponding attribute. Other models of attribute preference mentioned in [16] are *utility functions* (identical with our fuzzy functions) and crisp *preference orderings* of objects. Aggregation is defined as *OWA operator* [17]. It is ordered weighted average associated with a normalized weight vector $w = (w_1, \ldots, w_n)^T$. Note that weight w_i is not connected with preference value of i^{th} attribute, but with i^{th} greatest preference value. This operator produces one "collective" preference relation from attribute preferences.

The *main difference* is that multicriteria decision is mainly designed for a small number of objects (alternatives). Each preference relation requires quadratic space w.r.t. number of objects. Decision making process consists of computing aggregated preference relation and transforming this relation to a global ranking of objects. Efficiency issues are left to implementations. Our model is closely connected with efficient top-k searching method which allows finding best objects without computing aggregation for all of them and sorting the results. Advantages of fuzzy multicriteria decision include support of sophisticated aggregation methods like Choquet integral and various modifications of OWA operator.

Another method to find interesting objects according to more attributes is based on *Skyline* [18]. Every attribute forms one dimension and every object is represented as a point in multi-dimensional space. Skyline is defined as a set of points which are not *dominated* by any other point (Pareto optimal), which means that no other point is equally good or better in all dimensions and strictly better at least in one dimension. A special Skyline clause can be integrated to relational database systems. Efficient methods of finding Skyline are proposed in [18, 19], including partitioning, indexing with B-trees and searching with top N algorithm (similar to top-k).

Compared to our system, Skyline produces a set of *unordered*, potentially interesting objects. Despite the similar searching methods, results can be different from our top-k results. First object o returned by top-k always belongs to Skyline set, but not necessarily the following objects, which can be in fact dominated by o, but still very close to it. Skyline also does not support attribute preferences other than simple "The more the better" or its reverse for ordinal attributes, thus it assumes the same attribute preferences for all users.

Another similar approach is *rank-aware querying* [20]. It uses general aggregation (scoring) functions to compute object ranking and returns a list of results ordered by ranking. Aggregations and new rank-join operators are integrated into relational database system. Rank-join operator performance is optimized by transforming a query into an execution plan and finding the most efficient plan. As in the case of Skyline operator, rank-aware querying does not support various attribute preferences or fuzzy rules. Simple preferences like "The less the better" can be simulated with scoring functions.

All-purpose fuzzy database system *FSQL Server* is proposed in [21]. It supports attribute preferences (called *linguistic labels*) for both ordered and unordered attribute do-

ISBN: 978-989-95079-6-8

mains. It is possible to define special fuzzy connectives to aggregate the results, but weighted average and fuzzy rules are not supported. In addition to fuzzy features needed in our system, FSQL supports fuzzy comparison, fuzzy constants, special values and α-cuts (fulfillment thresholds). FSQL is intended to support all common fuzzy features, while our system is designed for a single purpose of preferential web search.

From theoretical point of view, *rough set theory* [22] also has some common ideas with our model. It uses conditional attributes and decision attributes (similar to our user rating). Objects with the same attribute values belong to the same equivalence class and they are called *indiscernible* by those attributes. However, a set that contains one element from an equivalence class and does not contain another object from the same equivalence class, cannot be defined exactly (using only available attributes). Such set is called *rough set* and it can be defined with help of lower and upper approximation.

The main goal of rough set theory is to synthesize *decision rules* to determine the value of decision attribute from conditional attributes. The main difference is that a rough set *does not generate any ordering*, we can only distinguish objects inside the set from "boundary" objects.

6 Conclusions

Some approaches mentioned in the previous section are focused on theoretical support of decision process, some deal more with efficiency and optimization. We address both issues to achieve a clear-cut formal model and efficient implementation. Our main purpose is to enable user-specific web search in one arbitrary domain. All parts described in section 2 are important for personalized search: web information extraction, interface for explicit preference specification, top-k search, user rating and inductive learning. To the best of our knowledge, other fuzzy preference systems do not address all these phases to connect web and user.

So our approach has the following contributions:

- *Formal model* – fuzzy sets and aggregation as a model of user preference, suitable for preferential search.
- *Expressive power* – arbitrary fuzzy sets and fuzzy rules are strong tools to represent preferences.
- *Implementation* – web-based system implements specified functionality.
- *Model evaluation* – various fuzzy measures are used to analyze the results.

Acknowledgment

Partially supported by Czech projects 1ET100300517, MSM-0021620838 and Slovak project VEGA 1/0131/09.

References

[1] Tomáš Horváth. A model of user preference learning for content-based recommender systems. *Accepted for publication to COMPUTING AND INFORMATICS*.

[2] Peter Gurský and Peter Vojtáš. On top-k search with no random access using small memory. *Proceeding of ADBIS*, pages 97–111, 2008.

[3] Ronald Fagin, Amnon Lotem, and Moni Naor. Optimal aggregation algorithms for middleware. *PODS '01: Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 102–113, 2001.

[4] John Musser and Tim O'Reilly. *Web 2.0 Principles and Best Practices*. O'Reilly Radar Reports, 2006.

[5] D. Maruščák, R. Novotný, and P. Vojtáš. Unsupervised structured web data and attribute value extraction. *Proceedings of Znalosti*, 2009.

[6] Robert Baumgartner, Sergio Flesca, and Georg Gottlob. Visual web information extraction with lixto. pages 119–128, 2001.

[7] Eugene Agichtein. Web information extraction and user modeling: Towards closing the gap. *IEEE Data Eng. Bull.*, 29(4):37–44, 2006.

[8] Tomáš Horváth and Peter Vojtáš. Ordinal classification with monotonicity constraints. *ICDM 2006*, pages 217–225, 2006.

[9] Peter Gurský, Tomáš Horváth, Jozef Jirásek, Stanislav Krajči, Róbert Novotný, Veronika Vaneková, and Peter Vojtáš. Knowledge processing for web search - an integrated model. *Studies in Computational Intelligence (vol. 78)*, pages 95–104, 2007.

[10] Michio Sugeno. *Industrial Applications of Fuzzy Control*. Elsevier Science Inc., New York, NY, USA, 1985.

[11] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Hum.-Comput. Stud.*, 51(2):135–147, 1999.

[12] Peter Gurský, Veronika Vaneková, and Jana Pribolová. Fuzzy user preference model for top-k search. *Fuzzy Systems, 2008 (IEEE World Congress on Computational Intelligence)*, pages 1606–1612, 2008.

[13] Maurice Kendall and Jean D. Gibbons. *Rank Correlation Methods*. A Charles Griffin Title, 5 edition, 1990.

[14] Bernard De Baets and Radko Mesiar. Metrics and t-equalities. *Journal of Mathematical Analysis and Applications*, 267(2):531 – 547, 2002.

[15] Ronald R. Yager. Extending multicriteria decision making by mixing t-norms and owa operators: Research articles. *Int. J. Intell. Syst.*, 20(4):453–474, 2005.

[16] F. Chiclana, F. Herrera, and E. Herrera-viedma. Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations. *Fuzzy Sets and Systems*, 97:33–48, 1998.

[17] Ronald R. Yager and Janusz Kacprzyk, editors. *The ordered weighted averaging operators: theory and applications*. Kluwer Academic Publishers, 1997.

[18] Stephan Börzsönyi, Donald Kossmann, and Konrad Stocker. The skyline operator. *Proceedings of the 17th International Conference on Data Engineering*, pages 421–430, 2001.

[19] Jan Chomicki, Parke Godfrey, Jarek Gryz, and Dongming Liang. *Intelligent Information Processing and Web Mining*, volume 31/2005 of *Advances in Soft Computing*, chapter Skyline with Presorting: Theory and Optimizations, pages 595–604. Springer-Verlag, 2005.

[20] Ihab F. Ilyas, Rahul Shah, Walid G. Aref, Jeffrey Scott Vitter, and Ahmed K. Elmagarmid. Rank-aware query optimization. *SIGMOD '04: Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 203–214, 2004.

[21] Jose Galindo. *Fuzzy Databases: Modeling, Design, and Implementation*. IGI Publishing, 2006.

[22] Lech Polkowski and Andrzej Skowron. Rough sets: A tutorial. *Rough Fuzzy Hybridization: A New Trend in Decision-Making*, 2000.

ISBN: 978-989-95079-6-8

IFSA-EUSFLAT 2009

Multi-Goal Aggregation of Reduced Preference Relations Based on Fuzzy Interactions between Decision Goals

Rudolf Felix
F/L/S Fuzzy Logik Systeme GmbH, Joseph-von-Fraunhofer Straße 20, 44 227 Dortmund, Germany
Tel. +49 231 9 700 921, Fax. +49 231 9700 929, e-mail: felix@fuzzy.de

Abstract– *Many aggregation operators like fuzzy integrals and similar require a preference relation as input. If the number of decision alternatives increases it becomes almost impossible to provide the exponentially increasing number of preference statements which are needed because a preference relation is defined upon the power set of the set of decision alternatives. If, as in many real world applications, the number of decision goals increases, it is also almost impossible to supervise the consistency of preference statements with respect to the interactions between the decision goals. In this paper it is discussed in which way a decision making approach based on interactions between goals is applied in order to use preference information that is not defined upon the power set of the set decision alternatives but simply on the decision set itself. Instead of defining the preferences upon the power set of the decision alternatives with respect to all the goals, for every decision goal the preference of the decision alternatives for this goal is defined upon the set of decision alternatives by a linear preference ranking of the decision alternatives. Consequences for both decision modeling based on classical preference relations and based on weighted sums are discussed.*

Keywords– Aggregation complexity, decision making, interactions between goals, reduced preference relation, weighted sum.

1 Introduction

Many decision making approaches are limited with respect to the management of complexity [4], [5], [8], [9], [10]. The consequence of this limitation is that these models are rather not applicable for real world problems. Many aggregation approaches require a preference relation as input. For more complex decision problems such preference relations are difficult to obtain because a preference relation is defined upon the power set of the set of the decision alternatives. If the number of decision alternatives increases it becomes almost impossible to provide the exponentially increasing number of preference statements needed. If additionally, as in many real world applications, the number of decision goals increases, it is almost impossible to supervise the consistency of preference statements with respect to the interactions between the decision goals. In this paper it is discussed in which way a decision making approach based on interactions between goals is adapted in order to aggregate in complex decision situations requiring only reduced linear preference information. The required input preference information is not defined upon the power set of the decision alternatives. Instead of this, for every single decision goal a linear preference ranking of the decision alternatives with respect to that goal is required. Of course, the single goal preference rankings may rank the decision alternatives in different way for different goals as the goals usually are partly conflicting. The presented aggregation approach ascertains for each pair of goals their conflicts (and correlations) by computing the so called interactions between the goals from the initial input single goal rankings. The only additional information needed when the interactions between the goals have been calculated is a linear importance information of each goal which is expressed in terms of goal priorities. The goal priorities are numbers of [0,1] and are comparable with a fuzzy measure that ranks not the decision alternatives but the decision goals itself with respect to their importance (priority).

It turned out that the decision making based on interactions between goals is less limited because the complexity of both the input information required and the aggregation process is not higher than polynomial. Since the model has successfully been applied to many real world problems [3] it clearly helped to manage many complex aggregation processes in which the number was varying between 50 and 100. The number of decision alternatives was in some applications even more than several thousands.

In the subsequent sections first for better readability of the paper we repeat a brief description of the decision making approach based on interactions between decision goals. Then we show how the approach is applied in order to work with single goal preference rankings and a statement with respect to classical preference based decision modeling is made. Subsequently it is discussed under which conditions decision models based on weighted sums help and how they are related to decision situations with interacting decision goals. Finally, the consequences of the results are discussed.

2 Decision Making based on Interactions between Goals

In the following it is shown how an explicit modeling of interaction between decision goals that are defined as fuzzy sets of decision alternatives helps to manage complexity of the decision making and aggregation. This modeling of the decision making and aggregation process significantly differs from the related approaches and the way they manage complex decision situations. First the notion of positive and negative impact sets is introduced. Then different types of interaction between goals are defined. After this it is shown how interactions between goals are used in order to aggregate pairs of goals to the so called local decision sets. Then it is described how the local decision sets are used for the aggregation of a final decision set. The complexity of the different steps is discussed.

2.1 Positive and Negative Impact Sets

Before we define interactions between goals as fuzzy relations, we introduce the notion of the positive impact set and the negative impact set of a goal. A more detailed discussion can be found in [1],[2] and [3].

Def. 1a) Let A be a non-empty and finite set of decision alternatives, G a non-empty and finite set of goals, $A \cap G$

ISBN: 978-989-95079-6-8

$=\emptyset, a \in A, g \in G, \delta \in (0,1]$. For each goal g we define the two fuzzy sets S_g and D_g each from A into [0, 1] by:

1. Positive impact function of the goal g: $S_g(a) := \delta$, if a affects g positively with degree δ, $S_g(a) := 0$ else.
2. Negative impact function of the goal g: $D_g(a) := \delta$, if a affects g negatively with degree δ, $D_g(a) := 0$ else.

Def. 1b) Let S_g and D_g be defined as in Def. 1a). S_g is called the positive impact set of g and D_g the negative impact set of g.

The set S_g contains alternatives with a positive impact on the goal g and δ is the degree of the positive impact. The set D_g contains alternatives with a negative impact on the goal g and δ is the degree of the negative impact.

2.2 Interactions between Goals

Let $P(A)$ be the set of all fuzzy subsets of A. Let X, Y ∈ $P(A)$, x and y the membership functions of X and Y respectively. Assume now that we have a binary fuzzy inclusion $I: P(A) \times P(A) \to [0,1]$ and a fuzzy non-inclusion $N: P(A) \times P(A) \to [0,1]$, such that $N(X,Y) := 1 - I(X,Y)$. In such a case the degree of inclusions and non-inclusions between the impact sets of two goals indicate the degree of the existence of interaction between these two goals. The higher the degree of inclusion between the positive impact sets of two goals, the more cooperative the interaction between them. The higher the degree of inclusion between the positive impact set of one goal and the negative impact set of the second, the more competitive the interaction. The non-inclusions are evaluated in a similar way. The higher the degree of non-inclusion between the positive impact sets of two goals, the less cooperative the interaction between them. The higher the degree of non-inclusion between the positive impact set of one goal and the negative impact set of the second, the less competitive the relationship. The pair (S_g, D_g) represents the whole known impact of alternatives on the goal g. Then S_g is the fuzzy set of alternatives which satisfy the goal g. D_g is the fuzzy set of alternatives which are rather not recommendable from the point of view of satisfying the goal g.

Based on the inclusion and non-inclusion between the impact sets of the goals as described above, 8 basic fuzzy types of interaction between goals are defined. The different types of interaction describe the spectrum from a high confluence between goals (analogy) to a strict competition (trade-off) [1].

Def. 2) Let S_{g_1}, D_{g_1}, S_{g_2} and D_{g_2} be fuzzy sets given by the corresponding membership functions as defined in Def. 1). For simplicity we write S_1 instead of S_{g_1} etc.. Let $g_1, g_2 \in G$ where G is a set of goals. *T* is a t-norm.

The fuzzy types of interaction between two goals are defined as relations which are fuzzy subsets of $G \times G$ as follows*:*

1. g_1 *is independent of* g_2*:* <=> $T(N(S_1,S_2), N(S_1,D_2), N(S_2,D_1), N(D_1,D_2))$
2. g_1 *assists* g_2*:* <=> $T(I(S_1 S_2), N(S_1,D_2))$
3. g_1 *cooperates with* g_2*:* <=> $T(I(S_1,S_2), N(S_1,D_2), N(S_2,D_1))$
4. g_1 *is analogous to* g_2*:* <=> $T(I(S_1,S_2), N(S_1,D_2), N(S_2,D_1), I(D_1,D_2))$
5. g_1 *hinders* g_2*:* <=> $T(N(S_1,S_2), I(S_1,D_2))$
6. g_1 *competes with* g_2*:* <=> $T(N(S_1,S_2), I(S_1,D_2), I(S_2,D_1))$
7. g_1 *is in trade-off to* g_2*:* <=> $T(N(S_1,S_2), I(S_1,D_2), I(S_2,D_1), N(D_1,D_2))$
8. g_1 *is unspecified dependent from* g_2*:* <=> $T(I(S_1,S_2), I(S_1,D_2), I(S_2,D_1), I(D_1,D_2))$

The interactions between goals are crucial for an adequate orientation during the decision making process because they reflect the way the goals depend on each other and describe the pros and cons of the decision alternatives with respect to the goals. For example, for cooperative goals a conjunctive aggregation is appropriate. If the goals are rather competitive, then an aggregation based on an exclusive disjunction is appropriate.

Note that the complexity of the calculation of every type of interaction between two goals is O(card(A) * card(A)) = O((card(A))^2) [4].

2.3 Two Goals Aggregation based on the Type of their Interaction

The assumption, that cooperative types of interaction between goals imply conjunctive aggregation and conflicting types of interaction between goals rather lead to exclusive disjunctive aggregation, is easy to accept from the intuitive point of view. It is also easy to accept that in case of independent or unspecified dependent goals a disjunctive aggregation is appropriate. For a more detailed formal discussion see for instance [1],[2]. Knowing the type of interaction between two goals means to recognize for which goals rather a conjunctive aggregation is appropriate and for which goals rather a disjunctive or even exclusively disjunctive aggregation is appropriate. This knowledge then in connection with information about goal priorities is used in order to apply interaction dependent aggregation policies which describe the way of aggregation for each type of interaction. The aggregation policies define which kind of aggregation operation is the appropriate one for each pair of goals. The aggregation of two goals gi and gj leads to the so called local decision set $L_{i,j.}$. For each pair of goals there is a local decision set $L_{i,j.} \in P(A)$, where A is the set of decision alternatives (see Def 1 a)) and P(A) the power set upon A. For conflicting goals, for instance, the following aggregation policy which deduces the appropriate decision set is given:

if (g_1 is in trade-off to g_2) and (g_1 is slightly more important than g_2) then $L_{1,2} := S_1 / D_2$.

In case of very similar goals (analogous or cooperative goals) the priority information even is not necessary:
if (g_1 cooperates with g_2) then $L_{1,2} := S_1 \cap S_2$ because $S_1 \cap S_2$ surely satisfies both goals.
if (g_1 is independent of g_2) then $L_{1,2} := S_1 \cup S_2$ because

ISBN: 978-989-95079-6-8

$S_1 \cup S_2$ surely do not interact neither positively nor negatively and we may and want to pursue both goals.

In this way for every pair of goals g_i and g_j, i,j ∈ {1,…,n} decision sets are aggregated. The importance of goals is expressed by the so called priorities. A priority of a goal g_i is a real number $P_i \in [0,1]$. The comparison of the priorities is modeled based on the linear ordering of the real interval [0,1]. The statements like g_i *slightly more important than* g_j are defined as linguistic labels that simply express the extend of the difference between P_i and P_j.

2.4 Multiple Goal Aggregation as Final Aggregation based on the Local Decision Sets

The next step of the aggregation process is the final aggregation. The final aggregation is performed based on a sorting procedure of all local decision sets $L_{i,j}$. Again the priority information is used to build a semi-linear hierarchy of the local decision sets by sorting them. The sorting process sorts the local decision sets with respect to the priorities of the goals. Subsequently an intersection set of all local decision sets is built. If this intersection set is empty then the intersection of all local decision sets except the last one in the hierarchy is built. If the resulting intersection set again is empty then the second last local decision set is excluded from the intersection process. The process iterates until the intersection is not empty (or more generally speaking until its fuzzy cardinality is big enough with respect to a given threshold). The first nonempty intersection in the iteration process is the final decision set and the membership values of this set give a ranking of the decision alternatives that is the result of the aggregation process (for more details see [2]).

2.5 Complexity Analysis of the Aggregation Process

As already discussed for instance in [4] the complexity of the aggregation process is $O((card(A))^2 * (card(G))^2)$ and the complexity of the information required for the description of both the positive and the negative impact functions is $O(card(A) * card(G))$.

3 Application of the Aggregation in Case of Reduced Preference Relations

In preference based decision making the input preference information has to be defined upon the power set of the decision alternatives [6]. This means that the complexity of the input information required is exponential with respect to the cardinality of the set of decision alternatives. This also means that the input information required is very difficult to obtain. Especially if the number of decision goals increases and the goals are partly conflicting the required preference information has to be multidimensional and the provider of the preference information has to express all the multidimensional interactions between the goals through the preference relation. With increasing number of goals and interactions between them the complexity of the required input preference relation possesses the same complexity as the decision problem itself. But, if the complexity of the required input is the same as the solution of the underlying decision problem itself then the subsequent aggregation of the input does not really help to solve the problem and is rather obsolete.

Therefore we propose to reduce the complexity of the input preference relation. Instead of requiring a preference relation defined upon the power set of the set of the decision alternatives which expresses the multidimensionality of the impacts of the decision alternatives on the goals for every single decision goal a linear preference ranking of the decision alternatives with respect to that goal is required. This means that for every goal a preference ranking defined on the set of decision alternatives is required instead of a ranking defined on the power set of the alternatives. The multidimensionality of the goals is then computed from all the single goal preference rankings using the concept of interactions between the goals as defined in section X.2.

In the sequent we extend the definition Def. 1. The extension defines how a single goal preference ranking defined on the set of the decision alternatives is transformed into positive and negative impact sets:

Def. 1c) Let A be a non-empty and finite set of decision alternatives, G a non-empty, finite set of goals as defined in Def. 1) a),

$A \cap G = \emptyset, a \in A, g \in G, \delta \in (0,1]$.

Let $>p_g$ be a preference ranking defined upon A with respect to *g* defining a total order upon A with respect to g, such that

$a_{i1} >p_g a_{i2} >p_g a_{i3} >p_g \ldots >p_g a_{im}$, where m=card(A) and $\forall a_{ij}, a_{ik} \in A, a_{ij} >p_g a_{ik} :\Leftrightarrow a_{ij}$ is preferred to a_{ik} with respect to the goal g. The preference relation $>p_g$ is called the *reduced single goal preference relation* of the goal g.

For simplicity, instead of $>p_g$ we also equivalently write RSPR of the goal g. All the RSPRs for all $g \in G$ are called the *reduced preference relation* RPR for the whole set of goals *G*. In order to avoid complete redundancy within the RPR we additionally define that the RSPRs of all the goals are different.

Let us assume that there is a decision situation with *n* decision goals where *n=card(G)* and *m* decision alternatives where *m=card(A)*. In the subsequent we propose an additional extension of the original definition Def. 1b with the aim to transform the single goal preference relations RSPR of every goals $g \in G$ into the positive and negative impact sets *Sg* and *Dg*:

Def. 1d) Let again A be a non-empty and finite set of decision alternatives, G a non-empty, finite set of goals as defined in Def. 1) a), $A \cap B \neq \emptyset$, $a_i \in A$, *m=card(*A*), g*∈G, i, c∈ {1, …, *m*}. For any goal g we obtain both the positive and the negative impact sets Sg and Dg by defining the values of δ according to Def. 1a) and 1b) as follows:

Def. 1d1) For the positive impact set:

$Sg(a_i)=\delta:=1/i$ iff $i \in [1,c-1]$,
$Sg(a_i)=\delta:=0$ iff $i \in [c,m]$.

Def. 1d2) For the negative impact set:

$Dg(a_i)=\delta:=0$ iff $i \in [1,c-1]$,
$Sg(a_i) = \delta:= 1/(m-i+1)$ iff $i \in [c, m]$.

Using the definition Def. 1d1 and Def. 1d2 for any goal g∈G we obtain a transformation of the RPR into positive

ISBN: 978-989-95079-6-8

and negative impact sets of all the goals and can evaluate the interactions of the goals using Def. 2 that are implied by the RPR.

Compared to classical preference based decision models this transformation helps to reduce the complexity of the input preference information required without losing modeling power for complex real world problems. The advantage is that using Def. 2. the interactions between goals that are implied by the RPR expose the incompatibilities and compatibilities that may be hidden in the RPR. The exposed incompatibilities and compatibilities are used adequately during the further calculation of the decision sets. Note that the exposition is calculated with a polynomial number of calculation steps with degree 2. The only additional information required from the decision maker is the priority information for each goal which has to be expressed as a weight with a value between 0 and 1. Many real world applications show that despite the reduced input complexity there is no substantial loss of decision quality [3].

*Statement 1***:** *In particular this means that it is not necessary to have classical preference relations defined upon the power set of the decision alternatives in order to handle complex decision problems with both positively and negatively interacting decision goals.*

Another interesting question is how the decision making based on interactions between decision goals is formally related to aggregation methods based on weighted sums. In order to investigate this we introduce the notion of r-consistency of RPRs and will consider the question under which conditions the weighted sum aggregation may be appropriate from the point of view of the application of the decision making based on interactions between goals if we have an RPR as input. For this we define the following:

Def.3) Given a discrete and finite set A of decision alternatives. Given a discrete and finite set G of goals. Let $r \in (0,1]$. The reduced preference relation RPR is called r-consistent :$\Leftrightarrow \exists\, c \in \{1, \ldots, m\}$, $m=card(A)$ such that $\forall\, (g_i, g_j) \in G \times G$, $i,j \in \{1, \ldots, n\}$, $n=card(G)$, (g_i *cooperates with* g_j) $\geq r$.

Let us now consider an important consequence of the interaction between goals using the notion of r-consistency. This notion will imply a condition under which an aggregation based on a weighted sum may lead to an appropriate final decision.

For this let us assume that the quite sophisticated final aggregation process of the iterative intersections as described in section 2.4 is replaced by the following rather intuitive straight forward consideration of how to obtain an optimal final decision.

Again let us identify the local decision sets $L_{i,j}$ obtained after the application of the local decision policies for each pair of goals (g_i, g_j) as the first type of decision subsets of the set of all decision alternatives A which, according to the decision model are expected to contain an optimal decision alternative a_k. Thus we define the set of sets

T1:= { $L_{i,j}$ | i ,j $\in$ {1, .. n}, $n=card(G)$} and expect that the optimal decision alternative has a positive membership in at least one of the sets of T1. Since we want to consider multiple goals we may also expect that an optimal decision alternative may have a positive membership in at least one of the intersections of pairs of the local $L_{i,j}$. Thus we define T2:= { $L_{k,l} \cap L_{p,q}$ | k,l,p,q $\in$ {1, .. n}, $n=card(G)$}. In order to simplify the subsequent explanation we concentrate on the crisp case and replace all membership values > 0 in all $L_{i,j}$, and $L_{k,l} \cap L_{p,q}$ by the membership value 1. Now we define the system of these crisp sets GDS as follows:

Def.4) GDS:={∅,T1,T2}, T1,T2 are the sets of crisp sets that we construct as described above by replacing all membership values > 0 in all $L_{i,j}$, and $L_{k,l} \cap L_{p,q}$ by the membership value 1.

With this definition we are able to formulate the following theorem that describes a property of the crisp GDS in the case that the underlying decision situation stems from a reduced preference relation RPR to which the decision model based on interactions between decision goals is applied. This property will enable us to relate this decision making to the calculation of optimal decisions by concepts based on weighted sums that are strongly connected with the notion of a matroid [7] and optimal decisions obtained by Greedy algorithms.

Theorem 1: If the reduced preference relation RPR is r-consistent then the system ($\mathcal{P}$*(A),GDS) is a matroid.*

Sketch of the Proof:

The proof will show that the following matroid conditions [7] hold: 1. $\emptyset \in$ GDS, 2. $X \subseteq Y$, $Y \in$ GDS $\Rightarrow X \in$ GDS and 3. $X,Y \in$ GDS card(X)<card(Y) $\Rightarrow \exists x \in Y \backslash X$ and $X \cup \{x\} \in$ GDS.

Ad1.: Holds by definition of GDS. Ad2.: If X is one of the $L_{i,j}$s then Y must be one of the another $L_{k,l}$ with $i \neq k$, $j \neq l$. If X is one of the $L_{k,l} \cap L_{p,q}$ then Y must be $L_{k,l}$ or $L_{k,l}$. Ad3.: This condition holds because the RPR is r-consistent and therefore we have $\forall$ k,l,p,q $\in$ {1, .. n}, $n=card(G)$, card($L_{i,j}$)=card($L_{k,l} \cap L_{p,q}$)=c.

With this result we obtain the following corollary that relates decision making based on interactions between decision goals to calculating optimal decisions by concepts based on weighted sums.

Corollary 1: If the reduced preference relation RPR is r-consistent then there exists a greedy algorithm based on the maximization of the weighted sum $\Sigma_{i \in \{1,\ldots,n\}, k \in \{1,\ldots,m\}} P_i * S_i(a_k)$, *where* S_i *is defined as in Def. 1d1 and* $a_k \in A$ *being a decision alternative,* $n=card(G)$, $m=card(A)$. P_i *is the priority of the goal* g_i *as used for the calculation of both the local and the global decision sets (see sections 4.3 and 4.4).*

Sketch of the Proof:

The proof is an immediate result from the Theorem 1 and the theory of matroids used in the field of combinatorial optimization [7].

4 Discussion of the Consequences

As already mentioned, this corollary relates the decision making based on interactions between decision goals to the calculation of optimal decisions by concepts based on weighted sums. It shows that weighted sums both as aggregation and optimization concept are rather appropriate if the goals or criteria are cooperative e.g. if they interact

ISBN: 978-989-95079-6-8

positively (or at least do not interact at all being independent). In contrast to this the decision making based on interactions between decision goals is more general and it reflects both positive and negative interactions between the goals. This is important in the context of real world decision making and optimization problems which usually possess partly conflicting goal structures. If the aggregation is performed by weighted sums, like for instance Choquet integrals, then one can suppose that the aggregation will only work if the choices to be made are free of conflicts. The hypothesis formulated in [6] postulating that there exists a preference relation on subsets of decision alternatives which is compatible with a weak order of decision alternatives belonging to different subsets of decision alternatives seems not to be that natural as postulated (especially in the field of optimization problems where partly contradicting goals are daily business).

Even if for a given application field the postulated preference relation exists, in case the number of decision alternatives increases it becomes almost impossible to provide the exponentially increasing number of required preference statements. If additionally, as in many real world applications, the number of decision goals increases, it is almost impossible for a human decision maker to keep all the interactions between the goals in mind and supervise the consistency of the preference statements with respect to the interactions between these decision goals. We see that the adapted decision making approach based on interactions between goals helps reducing the complexity of the required preference information without reducing neither the quality of the decision results [3] nor the complexity of decision situations that are modeled [4]: The required input preference information is not defined upon the power set of the decision alternatives. Instead of defining the required preferences upon the power set of the decision alternatives with respect to all the goals, for every decision goal the preference of the decision alternatives for this goal is defined upon the set of decision alternatives by a simple linearly ordered preference ranking of the decision alternatives (for each particular goal). This means that it is not necessary to require from the decision maker that he or she keeps in mind all the interactions between the goals. We are able to calculate the interactions as implication of linearly ordered single goal preference rankings by applying Def. 2. accordingly as shown in section 3. The only additional information that we require from the decision maker are assertions about the importance of the goals, namely again only a simple (linear) ordering of the so-called priorities of goals which are real numbers from the interval [0,1] and express for every goal its importance in a particular decision situation (see sections 2.3 and 2.4).

5 Conclusions

In this paper it is discussed in which way a decision making approach based on interactions between goals is applied in order to use preference information that is not defined upon the power set of the set decision alternatives but simply on the decision set itself. Instead of defining the preferences upon the power set of the decision alternatives with respect to all the goals, for every decision goal the preference of the decision alternatives for this goal is defined upon the set of decision alternatives by a linear preference ranking of the decision alternatives. Although the input complexity is reduced, the decision quality is not. Therefore we conclude that classical preference relations defined upon the power set of the decision alternatives are not necessary in order to adequately handle complex decision problems. We have also discussed in which way the decision model based on interaction between decision goals can be related to decision making methods based on weighted sums only. Through a link to the theory of matroids it is shown that the model is more general in the sense that an adequate aggregation is possible even if the goals are partly conflicting whereas weighted sums are in this kind of situations rather restricted.

References

[1] Felix, R., "Relationships between goals in multiple attribute decision making", *Fuzzy Sets and Systems*, Vol.67, 47-52, 1994.

[2] Felix, R. (1998). “Decision-making with interacting goals”.In: *Handbook of Fuzzy Computation*, Ruspini, E., Bonissone, P.P, Pedrycz, W. (Eds.), IOP Publishing Ltd, 1998

[3] Felix, R. (2007). Real World Applications of a Fuzzy Decision Model Based on Relationships between Goals (DMRG). In: *Forging the New Frontiers, Fuzzy Pioneers I* (1965-2005), Springer Verlag in the series Studies in Fuzziness and Soft Computing, October 2007.

[4] Felix, R. (2008). Multicriterial Decision Making (MCDM): Management of Aggregation Complexity Through Fuzzy Interactions Between Goals or Criteria. In: *Proceedings of the 12th International IPMU Conference*, Málaga, Spain, 2008

[5] Modave, F., Dubois, D., Grabisch, M. and Prade, H., "A Choquet integral representation in multicriteria decision making", *AAAI Fall Symposium*, Boston, Ma, November 1997.

[6] Modave, F. and Grabisch, M., "Preference *representation by a Choquet integral: Commensurability hypothesis. In: Proceedings of the 7th International IPMU Conference*, Paris, France, 164-171, 1998.

[7] Oxley, J., “Matroid Theory”, Oxford University Press, 1992.

[8] Saaty, T. L., "The Analytic Hierarchy Process", Mc Graw-Hill, 1980.

[9] Torra, V., "Weighted OWA operators for synthesis of information", *Proceedings of the fifth IEEE International Conference on Fuzzy Systems*, New Orleans, USA, 966-971, 1996.

[10] Yager, R.R., "Families of OWA operators", *Fuzzy Sets and Systems*, Vol. 59, 125-148, 1993.

ISBN: 978-989-95079-6-8

A Fuzzy Real Option Valuation Approach to Corporate Investment Decisions

Elettra Agliardi[1] Maria Letizia Guerra[2] Luciano Stefanini[3]

1.Department of Economics, University of Bologna, Italy
2.Department Matemates, University of Bologna, Italy
3.Department of Economics and Quantitative Methods, University of Urbino, Italy
Email: elettra.agliardi@unibo.it, m.guerra@unibo.it, luciano.stefanini@uniurb.it

Abstract— Real options are a typical framework in economics that involves uncertainty. Very often, in fact, managers have vague ideas about the future expected cash flows, the cost of the project and many other variables that are fundamental in the process decision among many investments. The calculation of the value function of real options can take advantage of a model of uncertainty that include stochastic processes and fuzzy numbers.

A special version of the multiple population differential evolution algorithm is designed to compute the level-cuts of the fuzzy extension of the multidimensional real valued function of fuzzy numbers in the resulting optimization problems.

We perform some computational experiments about the option to defer investment, that is an American call option on the present value of the completed expected cash flows with the exercise price equal to the required outlay. We show that fuzziness may help for a more profitable decision.

Keywords— Fuzzy Numbers, Parametric Representation, Real Options, Sensitivity Analysis.

1 Introduction

Nowadays the most suitable valuation methodology for corporate investment decisions is the real options theory (ROT) because it takes into account management's flexibility to adapt ongoing projects in response to uncertain conditions. Since Myers' ([7]) innovative idea of viewing firm's future investment opportunities as real options – that is, the right but not the obligation to undertake some business decision at a cost during a certain period of time -, a vast literature has developed, which elaborate both theoretical and empirical methods for quantifying the values of various real (call or put) options embedded in investment opportunities. Dixit and Pindyck ([3]) develop a systematic treatment of ROT, providing the fundamentals of this method and also emphasizing the market implications of such valuation of investment decisions under uncertainty. Trigeorgis ([12]) provides a classification of real options that maps different categories of investments into the space of different types of financial options.

The value of real options depends on some basic variables:(i) the underlying asset, which is the current value of (gross) expected future operating cash flows, (ii) the exercise price, which is the cost of the project; (iii) the time to expiration of the option, that is the time up to which the project can be undertaken (either finite or infinite); (iv) the standard deviation of the value of the underlying risky asset; (v) the risk-free rate of interest over the life of the option. All the above-mentioned variables are uncertain and therefore various stochastic models have been introduced in ROT to deal with the uncertainty surrounding most corporate decisions. In general expected future cash flows are assumed to evolve according to a geometric Brownian motion, but it often happens that reality is more complex than a normal distribution.

The imprecision associated with the subjective judgement and estimation of future cash flows, which is typical of management's project decisions, needs to be incorporated in the treatment of uncertainty. By introducing fuzzy numbers, we are able to capture the somewhat vague and imprecise ideas the manager possesses about the future expected cash flows, the profitability of the project, the costs of the project, etc. To the best of our knowledge, such an approach has never been discussed in the literature, with the exception of Carlsson and Fuller [1], that interpret the possibility of making an investment decision in terms of a European option, while we use an American option. In addition, we elaborate a computing methodology which is more general and we can represent the shape of the value functions.

The paper is organized as follows. In section 2 we present some basic elements of fuzzy theory which will be used in the numerical implementation of real option models. In section 3 we describe the introduction of fuzziness in real options and section 4 collects some of the computational experiments that have been performed in order to capture how and how much fuzziness affects the decision. Finally, section 5 concludes.

2 Fuzzy Numbers and Parametric Representation

Fuzzy numbers (more details in [4]) are a very powerful and flexible way to describe uncertainty or possibilistic values for given variables for which a precise quantification is not possible or one is interested in evaluating the effects of variations around a specified value. In fact a fuzzy number models the different specifications of intervals around a given precise value; it is defined, informally, as a "cascade" of intervals, which start with a given number $a \in \mathbb{R}$ and grow increasing to a final interval which gives the most uncertain set of possible values. The levels of the cascade are usually parametrized by a parameter $\alpha \in [0,1]$ which represents the so called membership value (or possibilistic degree) of a given interval, with the convention that $\alpha = 1$ corresponds to the possibly exact value (the core of the fuzzy number) while $\alpha = 0$ corresponds to the highest uncertainty (the support of the fuzzy number). With the same convention on α we can say that $1 - \alpha$ is the level of uncertainty of the corresponding interval.

ISBN: 978-989-95079-6-8

A wide class of fuzzy numbers with the core at $a \in \mathbb{R}$ is obtained by considering its membership function $\mu : \mathbb{R} \longrightarrow [0,1]$ such that, denoting $[a^-, a^+]$ the interval representing the support,

$$\mu(x) = \begin{cases} L(x) & if \quad a^- \leq x \leq a \\ R(x) & if \quad a \leq x \leq a^+ \\ 0 & otherwise \end{cases} \quad \text{for } x \in \mathbb{R} \tag{1}$$

where $L(x)$ is the left branch, an increasing function with $L(a^-) = 0$, $L(a) = 1$ and $R(x)$ is the right branch, a decreasing function with $R(a) = 1$, $R(a^+) = 0$. A fuzzy number obtained by the form (1) is called LR-fuzzy number (with usual notation $u = \langle a^-, a, a^+ \rangle_{L,R}$).

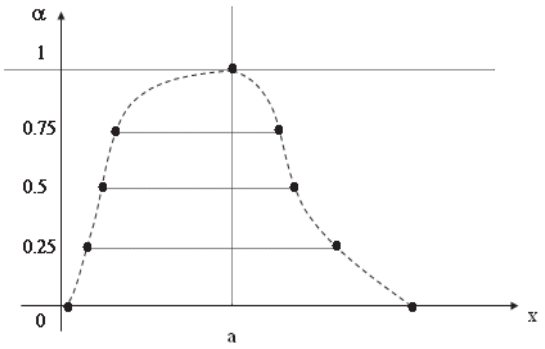


Figure 1: A fuzzy number as a "cascade" of intervals representing increasing uncertainty around the given value a.

For values of $\alpha \in]0,1]$, the $\alpha - cut$ is defined to be the compact interval $[u]_\alpha = \{x | \mu(x) \geq \alpha\}$ and the support is $[u]_0 = cl\{x | \mu(x) > 0\}$ ($cl(A)$ is the closure of set A). The *level-cuts* of a fuzzy number are "nested" closed intervals and this property is the basis for the LU representation (L for lower, U for upper).

Definition 1 *An LU-fuzzy quantity (number or interval) u is completely determined by any pair $u = (u^-, u^+)$ of functions $u^-, u^+ : [0,1] \longrightarrow \mathbb{R}$, defining the end-points of the $\alpha - cuts$, satisfying the three conditions:(i) $u^- : \alpha \longrightarrow u^-_\alpha \in \mathbb{R}$ is a bounded monotonic nondecreasing left-continuous function $\forall \alpha \in]0,1]$ and right-continuous for $\alpha = 0$;(ii) $u^+ : \alpha \longrightarrow u^+_\alpha \in \mathbb{R}$ is a bounded monotonic nonincreasing left-continuous function $\forall \alpha \in]0,1]$ and right-continuous for $\alpha = 0$;(iii) $u^-_\alpha \leq u^+_\alpha \ \forall \alpha \in [0,1]$. In particular, the $\alpha - cuts$ of a fuzzy number or interval are nonempty, compact intervals of the form $[u]_\alpha = [u^-_\alpha, u^+_\alpha] \subset \mathbb{R}$.*

The support of u is the interval $[u^-_0, u^+_0]$ and the core is $[u^-_1, u^+_1]$. We refer to the functions $u^-_{(.)}$ and $u^+_{(.)}$ as the lower and upper branches on u, respectively. The membership function can be written as $\mu_u(x) = \sup\{\alpha | x \in [u^-_\alpha, u^+_\alpha]\}$, where the left branch is the increasing inverse of $u^-_{(.)}$ on $[u^-_0, u^-_1]$ and the right is the decreasing inverse of $u^+_{(.)}$ on $[u^+_1, u^+_0]$.

To model the monotonic branches u^-_α and u^-_α we start with an increasing shape function p such that $p(0) = 0$ and $p(1) = 1$ and a decreasing function q such that $q(0) = 1$ and $q(1) = 0$, with the four numbers $u^-_0 \leq u^-_1 \leq u^+_1 \leq u^+_0$ defining the support $[u^-_0, u^+_0]$ and the core $[u^-_1, u^+_1]$ and we define

$$\begin{aligned} u^-_\alpha &= u^-_1 - (u^-_1 - u^-_0)p(\alpha) \text{ and} \\ u^+_\alpha &= u^+_1 - (u^+_1 - u^+_0)q(\alpha) \text{ for all } \alpha \in [0,1]\,. \end{aligned} \tag{2}$$

The two shape functions p and q, as suggested in [11], are selected in a family of parametrized monotonic functions where the parameters are related to the first derivatives of p and q in 0 and 1; there are many ways to define p and q as illustrated in [9]. For each decomposition we require (in the differentiable case) $4(N+1)$ parameters $u = (\alpha_i; u^-_i, \delta u^-_i, u^+_i, \delta u^+_i)_{i=0,1,...,N}$ satisfying the following conditions:

$$\begin{aligned} u^-_0 &\leq u^-_1 \leq ... \leq u^-_N \leq u^+_N \leq u^+_{N-1} \leq ... \leq u^+_0 \\ \delta u^-_i &\geq 0, \delta u^+_i \leq 0. \end{aligned} \tag{3}$$

and on each sub-interval $[\alpha_{i-1}, \alpha_i]$ we use the data $u^-_{i-1} \leq u^-_i \leq u^+_i \leq u^+_{i-1}$ and the slopes $\delta u^-_{i-1}, \delta u^-_i \geq 0$ and $\delta u^+_{i-1}, \delta u^+_i \leq 0$.

The simplest representation is obtained on the trivial decomposition of the interval $[0,1]$, with $N = 1$ (without internal points) and $\alpha_0 = 0, \alpha_1 = 1$. In this simple case, u can be represented by a vector of 8 components

$$u = (u^-_0, \delta u^-_0, u^+_0, \delta u^+_0; u^-_1, \delta u^-_1, u^+_1, \delta u^+_1). \tag{4}$$

In the search for the value of a real option, the fundamental step is the computation of fuzzy-valued functions. Given a function $y = f(x_1, x_2, ..., x_n)$ of n real (crisp) variables $x_1, x_2, ..., x_n$, its fuzzy extension is obtained to evaluate the effect of uncertainty on the x_j modelled by the corresponding fuzzy number u_j. If $v = f(u_1, u_2, ..., u_n)$ denotes the fuzzy extension of a continuous function f in n variables, then for each level α the resulting interval $[v^-_\alpha, v^+_\alpha]$ represents the propagation of uncertainty from x_j to y. In particular, if the uncertainty on the original variables is modelled by fuzzy numbers, the obtained v is yet a fuzzy number starting from a single value (at level $\alpha = 1$) to the most uncertain interval (at level $\alpha = 0$).

It is well known that the fuzzy extension of f to normal upper semicontinuous fuzzy intervals (with compact support) has the level-cutting property, i.e. the $\alpha - cuts$ $[v^-_\alpha, v^+_\alpha]$ of v are the images of the $\alpha - cuts$ of $(u_1, u_2, ..., u_n)$ and are obtained by solving the box-constrained optimization problems $(EP)_\alpha$:

$$\begin{cases} v^-_\alpha = \min \left\{ f(x_1, ..., x_n) | x_k \in [u^-_{k,\alpha}, u^+_{k,\alpha}],\ k = 1, ..., n \right\} \\ v^+_\alpha = \max \left\{ f(x_1, ..., x_n) | x_k \in [u^-_{k,\alpha}, u^+_{k,\alpha}],\ k = 1, ..., n \right\}. \end{cases} \tag{5}$$

With the exception of simple elementary cases for which the optimization problems above can be solved analytically, the direct application of *(EP)* may be difficult and computationally expensive. Usually, the $\alpha - cuts$ $[v^-_\alpha, v^+_\alpha]$ of v are computed at a prefixed given number of values α_j of interest (say from 10 to 100 points) and the membership function is approximated pointwise. As we will see, an advantage of the LU-parametrization is to obtain the extended fuzzy numbers $v = f(u_1, u_2, ..., u_n)$ in the same parametric form as for $u_1, u_2, ..., u_n$ with a possible important reduction in the computational effort and with a good approximation. As for the

ISBN: 978-989-95079-6-8

fuzzy numbers, we will consider the fuzzy extension of multivariate differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$.

For a vector of n fuzzy numbers $u = (u_1, u_2, ..., u_n)$ let

$$u_k = (u^-_{k,i}, \delta u^-_{k,i}, u^+_{k,i}, \delta u^+_{k,i})_{i=0,1,...,N} \text{ for } k = 1, 2, ..., n$$

be the LU representation of the $k-th$ component. Let $v = f(u_1, u_2, ..., u_n)$ and $v = (v^-_i, \delta v^-_i, v^+_i, \delta v^+_i)_{i=0,1,...,N}$ be its LU representation; the $\alpha - cuts$ of v are obtained by solving the box-constrained optimization problems (5). For each $\alpha = \alpha_i$, $i = 0, 1, ..., N$ the $\min$ and the $\max$ (5) can occur either at a point whose components $x_{k,i}$ are internal to the corresponding intervals $[u^-_{k,i}, u^+_{k,i}]$ or are coincident with one of the extreme values; denote by $\widehat{x}^-_i = (\widehat{x}^-_{1,i}, ..., \widehat{x}^-_{n,i})$ and $\widehat{x}^+_i = (\widehat{x}^+_{1,i}, ..., \widehat{x}^+_{n,i})$ the points where the $\min$ and the $\max$ take place; then

$$v^-_i = f(\widehat{x}^-_{1,i}, \widehat{x}^-_{2,i}, ..., \widehat{x}^-_{n,i}) \text{ and } v^+_i = f(\widehat{x}^+_{1,i}, \widehat{x}^+_{2,i}, ..., \widehat{x}^+_{n,i}) \tag{6}$$

and the slopes δv^-_i, δv^+_i are computed (as f is differentiable) by

$$\begin{aligned} \delta v^-_i &= \sum_{\substack{k=1 \\ \widehat{x}^-_{k,i} = u^-_{k,i}}}^{n} f'^-_k \delta u^-_{k,i} + \sum_{\substack{k=1 \\ \widehat{x}^-_{k,i} = u^0_{k,i}}}^{n} f'^-_k \delta u^+_{k,i} \\ \delta v^+_i &= \sum_{\substack{k=1 \\ \widehat{x}^0_{k,i} = u^-_{k,i}}}^{n} f'^+_k \delta u^-_{k,i} + \sum_{\substack{k=1 \\ \widehat{x}^0_{k,i} = u^0_{k,i}}}^{n} f'^+_k \delta u^+_{k,i} \end{aligned} \tag{7}$$

where $f'^-_k = \frac{\partial f(\widehat{x}^-_{1,i}, .., \widehat{x}^-_{n,i})}{\partial x_k}$ and $f'^+_k = \frac{\partial f(\widehat{x}^0_{1,i}, .., \widehat{x}^0_{n,i})}{\partial x_k}$.To solve the optimization problems (5), we use an implementation of a multiple population differential evolution algorithm extensively described and analyzed in [10].

- Dl ppleRi le fgkdfe kf PRSRh emRkd Rek

The *option to defer investment* is an American call option on the present value of the completed expected cash flows with the exercise price being equal to the required outlay. A project that can be postponed allows learning more about potential project outcomes before making a commitment. A seminal contribution on the option to defer is McDonald and Siegel [6] where the optimal time to invest and an explicit formula for the value of the option to invest are derived for an irreversible project whose net profits follow a geometric Brownian motion.

A firm is supposed to consider the following investment opportunity: at any time t the firm can pay some estimated cost K to install an investment project whose expected future net cash flows conditional on undertaking the project have an estimated present value Π. The installation of such project is irreversible. Let Π follow a geometric Brownian motion of the form:

$$d\Pi = \Pi(\mu dt + \sigma dW_t) \tag{8}$$

where $\mu < r$ is the appreciation rate, r is the risk-free interest rate and σ is the volatility ($\mu \in R, \sigma > 0$) and W is a standard Wiener process. For simplicity, let us assume that the time to expiration of this investment opportunity is infinite, which facilitates the derivation of a closed-form solution. If $V = V(\Pi)$ is the option value then it holds:

$$\frac{1}{2}\sigma^2 \Pi^2 V''(\Pi) + \mu \Pi V'(\Pi) - rV = 0$$

for $\Pi < \Pi^*$ with the initial condition $V(0) = 0$ and smooth-pasting $V(\Pi^*) = \Pi^* - K, V'(\Pi^*) = 1$. The solution is

$$\begin{cases} \Pi^* = K \frac{\phi}{\phi - 1} \\ V(\Pi) = (\Pi^* - K)\left(\frac{\Pi}{\Pi^*}\right)^{\phi} \end{cases} \tag{9}$$

with $\phi = \frac{1}{2} - \frac{\mu}{\sigma^2} + \left(\left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2r}{\sigma^2}\right)^{\frac{1}{2}} > 1$.

In our methodology fuzziness is present in three steps. The estimated present value Π of future net cash flows of the project follows the stochastic differential equation (8) and we model the uncertainty of its parameters across intervals of values. The intervals are built with differentiated levels of uncertainty; given a crisp value, the levels produce a shape that can be characterized by asymmetries or nonlinearities depending on subjective beliefs and available information of the decision maker. It follows that fuzzy parameters play the lead role in a sensitivity analysis that starts gradually from a null variation to the greatest variation of the uncertainty consistent with data. In particular W_t remains a standard Brownian motion (the theory and applications of fuzzy set-valued stochastic differential equations has received recently several significant contributions from the seminal paper of Feng [5]).

Fuzziness comes out also in the valuation function of the option (obtained with the extension principle) that depends not only on Π, σ and μ but also on r and K, which we assume to be fuzzy too.

Finally, fuzziness affects the crucial value Π^*: as soon as Π reaches the threshold value Π^*, the firm finds it optimal to invest (case of the option to defer investment) or disinvest and liquidate (case of the option to abandon). Thus, the decision is based on the threshold value, which depends on all the parameters of the model. In the valuation method based on fuzzy variables, $\{\Pi_t, t \geq 0\}$ is assumed to be a fuzzy stochastic process, which is specified by the following membership function

$$\mu_{\Pi_t(\omega)}(x) = max\{1 - \mid (x - \widehat{\Pi}_t(\omega))/\beta_t(\omega) \mid, 0\},$$

that is, the fuzzy random variable Π_t is of the triangular type, with centre $\widehat{\Pi}_t(\omega)$, and left-width and right-width $\beta(\omega)$. The assumption of fuzziness is related to the manager's subjective belief about the future profitability of the project. The choice of a triangle-type shape is not restrictive at all and is introduced for simplicity only. Observe that the fuzziness in the process increases as $\beta(\omega)$ becomes bigger. The α-cuts of $\Pi_t(\omega)(x)$ are $\Pi^{\pm}_{t,\alpha}(\omega) = \left[\Pi^-_{t,\alpha}(\omega), \Pi^+_{t,\alpha}(\omega)\right] = \left[\widehat{\Pi}_t(\omega) - (1-\alpha)\beta(\omega), \widehat{\Pi}_t(\omega) + (1-\alpha)\beta(\omega)\right]$. It is also reasonable to assume that K is a fuzzy number. In the case of an option to defer, K is the estimated liquidation value of the firm's stock of capital and is affected by depreciation, fluctuating market evaluation and taxation regimes. In the case of an option to abandon, K denotes the investment cost and has many components which can change during the waiting period, due to various unpredictable circumstances.

ISBN: 978-989-95079-6-8

The extension principle is then applied to obtain the fuzzy Π^* and $V(\Pi^*)$ from the exact solutions given in equation (9). In the formulae (6)-(7) the vector $\widehat{x}_i$ is equal to $(\widehat{\mu}_i, \widehat{\sigma}_i, \widehat{r}_i, \widehat{K}_i)$ and some of the partial derivatives that define the slopes of the representation are nothing else than the first order Greeks, in particular, $\frac{\partial f(\widehat{\mu}_i,\widehat{\sigma}_i,\widehat{r}_i,\widehat{K}_i)}{\partial \sigma}$ is the Vega and $\frac{\partial f(\widehat{\mu}_i,\widehat{\sigma}_i,\widehat{r}_i,\widehat{K}_i)}{\partial r}$ is the Rho.

The degree of the uncertainty and the way in which it is spread from the model, play a central role in the analysis of the real option. The nonlinearities entering in the definition of $V(\Pi)$ in (9) are the main cause of such effects and they can propagate or contract uncertainty. It is very important to perceive the magnitude and the type of these effects. In particular we are interested in the analysis of how the various kinds of uncertainties inserted into the parameters will produce the corresponding uncertainties in Π^*, $V^* = V(\Pi^*)$, Π^{**} and $V^{**} = V(\Pi^{**})$.

As soon as information (on μ, σ, r, K) is modelled by fuzzy numbers, Π^* and V^* also become fuzzy and are represented by $\alpha - cuts$ $\left[\Pi_\alpha^{*^-}, \Pi_\alpha^{*^0}\right]$ and $\left[V_\alpha^{*^-}, V_\alpha^{*^0}\right]$ for all degrees of possibility α. The maximal uncertainty corresponds to the supports at $\alpha = 0$, given by the intervals $\left[\Pi_0^{*^-}, \Pi_0^{*^0}\right]$ and $\left[V_0^{*^-}, V_0^{*^0}\right]$ for Π^* and V^* respectively.

Due to the nonlinearity of Π^* and V^*, the $\alpha - cuts$ are not necessarily symmetric and, for a given uncertainty on the input values μ, σ, r and K, they have different left and right variations. Let $\widehat{\Pi}^*$ and $\widehat{V}^*$ denote the values of Π_α^* and V_α^* corresponding to $\alpha = 1$. It is immediate to argue that V^* is symmetric if and only if $\Delta V_\alpha^{*^0} = \Delta V_\alpha^{*^-}, \forall \alpha \in [0, 1[$ where

$$\Delta V_\alpha^{*^0} = V_\alpha^{*^0} - \widehat{V}^*, \quad \Delta V_\alpha^{*^-} = \widehat{V}^* - V_\alpha^{*^-}.$$

The quantity $\Delta V_\alpha^{*^0}$ represents the possible increase in $\widehat{V}^*$ due to uncertainty and analogously, $\Delta V_\alpha^{*^-}$ measures the possible decrease. The same argument can be applied to Π_α^* and $\widehat{\Pi}^*$, defining the quantities $\Delta\Pi_\alpha^{*^0} = \Pi_\alpha^{*^0} - \widehat{\Pi}^*$ and $\Delta\Pi_\alpha^{*^-} = \widehat{\Pi}^* - \Pi_\alpha^{*^-}$.

. 6 f d gl kHkdf e HcRngRhdl Re ki

We test the fuzziness effect in the option to defer investment by running several computational experiments; the fuzzy version of the indicated parameters (say θ) are obtained as triangular symmetric fuzzy numbers, centered at the crisp values and with the support being the interval $[\theta - 0.1\theta, \theta + 0.1\theta]$, corresponding to a symmetric uncertainty of 10% of the value of the parameter. To analyze the effect of the uncertainty on the output variables, we give the plots of their membership functions and the tables including the values for $\alpha - levels$ with $\alpha = 1$ (the crisp level), $\alpha = 0.75$, $\alpha = 0.5$, $\alpha = 0.25$, $\alpha = 0$ (corresponding to the uncertainty of 10% in the parameters).

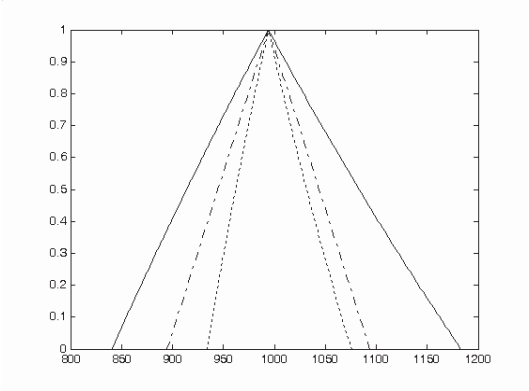
The robustness of the fuzzy model for the option to defer investment is tested with three sets of real data that we call, for short, Data1, Data2 and Data3, referring to three different industrial sectors. *Data1* refers to an investment decision in the human genome sciences project (HGSI) whose data are taken from the Human Genome project database (details in http://www.ornl.gov/sci/ techresources/ Human_Genome/ home.shtml). *Data2* refers to an investment decision in a big infrastructure, that is the Eurotunnel project (details in [2]). *Data3* deals with the case of an investment in new capacity in the public-utility sector, i.e. the electricity market (details in [8]). The values of the parameters μ, σ, r and K are the following:

	: HkH(	: HkH)	: HkH-
μ	0.01	0.025	0.03
σ	0.048	0.183	0.173
r	0.044	0.06	0.08
K	704.9	8865	600

In figures concerning the behavior of Π^* we report the three different cases that we will denote as: *Allfuzzy* (straight line) when the parameters μ, σ, r and K are fuzzy, *Kcrisp* (dotted line) when μ, σ, r are fuzzy and K is crisp and finally *Kfuzzy* (dashed line) when μ, σ, r are crisp and K is the unique source of uncertainty.

4.0.1 Results for Data1

Figure 2 shows that, as expected, the greatest uncertainty in Π^* occurs in the Allfuzzy case, when all the fuzzy quantities are considered to be fuzzy; but it is interesting to observe that in the Kcrisp case (dotted line) the generated uncertainty is less then in the Kfuzzy case (dashed line), i.e. the uncertainty in the values of only K produces more uncertainty on Π^* then the uncertainty in the values of μ, σ and r.



D5l1 hR) 4 Π^* for Data1

Table 1, Table 2 and Table 3 report values of the $\alpha - cut$ for Data1 in the Allfuzzy, Kcrisp and Kfuzzy case respectively.

Table 1

α	Π^-	Π^+
1.0	994.28	994.28
0.75	953.16	1037.56
0.5	913.97	1083.26
0.25	876.50	1131.69
0	840.58	1183.25

Table 2

Π^-	Π^+
994.28	994.28
977.60	1012.25
962.07	1031.67
947.56	1052.74
933.98	1075.68

Table 3

α	Π^-	Π^+
1.0	994.28	994.28
0.75	969.43	1019.14
0.5	944.57	1043.99
0.25	919.71	1068.86
0	894.86	1093.71

ISBN: 978-989-95079-6-8

Observe that in the Kfuzzy case (Table 3) the threshold value Π^* displays a symmetric shape in all analyzed projects because Π^* depends linearly on K. In the Allfuzzy and Kcrisp cases, instead, we can observe an asymmetric pattern, due to the nonlinear dependence of Π^* with respect to the other variables.

At level 0.5 the average values are 998.615 in Allfuzzy and 996.87 in Kcrisp, which are larger than the crisp value 994.28. Since on average the fuzzy threshold value is larger then without fuzziness, just considering the crisp value the decision to invest would be too early. Figure 3 shows the graphical behavior of the fuzzy function $V(\Pi)$ in the Allfuzzy case; the little crosses point the optimal values of Π corresponding to the levels of Π^* for $\alpha = 0, 0.25, 0.5, 0.75, 1$.

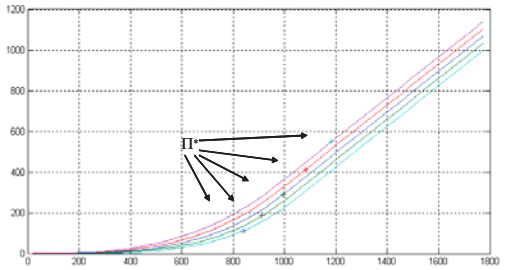


Figure 4: Data1 when parameters μ, σ, r and K are fuzzy.

It is evident that fuzziness implies a certain degree of freedom in the choice of Π^*. Figure 4 illustrates $V(\Pi)$ as a fuzzy function (a sequence of fuzzy numbers).

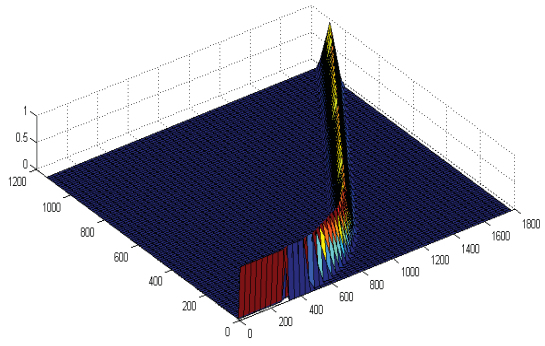


Figure 5: $V(\Pi)$ in the Allfuzzy case

4.0.2 *Results for Data2*

Figure 5 shows the behavior of Π^* in the three different cases: again the biggest uncertainty occurs in the Allfuzzy case, when all the quantities are fuzzy; but we observe that in the Kcrisp case (dotted line), when μ, σ, r are the sources of uncertainty and K is the unique crisp value, the uncertainty in Π^* is bigger then in the Kfuzzy case, i.e. the same level of uncertainty in K produces less uncertainty on Π^* then the uncertainty in the other parameters. With respect to Data1, there is here an inversion.

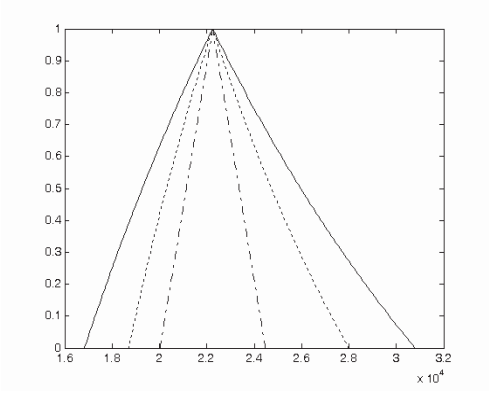


Figure 6: Π^* for Test 2

Table 4 and Table 5 report the values of the $\alpha - cut$ of Π^* in the Allfuzzy and Kcrisp case for Data2.

Table 4:

α	Π^-	Π^+
1.0	22249.62	22249.62
0.75	20686.18	23997.55
0.5	19277.60	25967.74
0.25	18000.33	28209.16
0	16835.43	30786.48

Table 5:

α	Π^-	Π^+
1.0	22249.62	22249.62
0.75	21216.59	23412.24
0.5	20292.21	24731.18
0.25	19459.82	26241.08
0	18706.04	27987.71

If we compute again the average values at level 0.5, they are 22622.67 in Allfuzzy and 22511.695 in Kcrisp, which are larger than the crisp value 22249.62. It follows that in the Data2 project it is confirmed the suggestion to wait for the decision to invest.

4.0.3 *Results for Data3*

The last project we consider for an option to defer investment is Data3; the relative values of Π^* are reported in Figure 6:

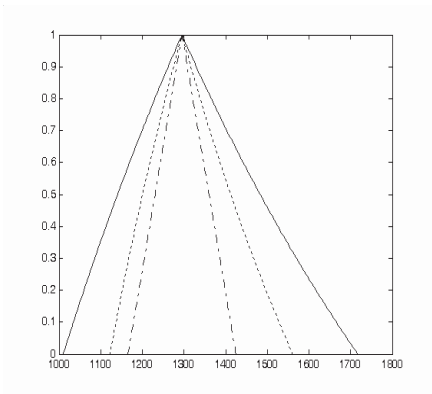


Figure 7: Π^* for Test4

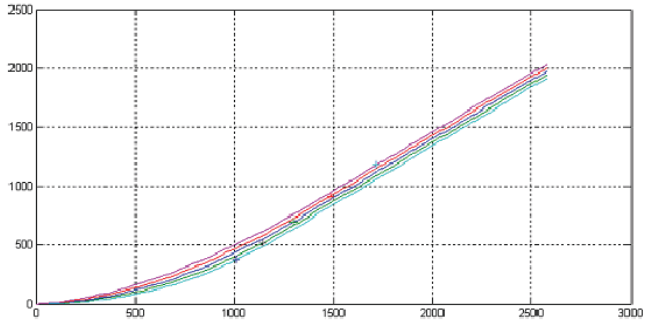
ISBN: 978-989-95079-6-8

Table 6, Table 7 report values of the $\alpha - cut$ for Data3 in the Allfuzzy and Kcrisp case respectively.

	GHIdR2	
α	Π^-	Π^+
1.0	1295.31	1295.31
0.75	1214.47	1384.31
0.5	1140.61	1482.98
0.25	1072.75	1593.18
0	1010.11	1717.32

GHIdR3	
Π^-	Π^+
1295.31	1295.31
1245.61	1350.55
1200.64	1412.36
1159.73	1482.03
1122.34	1561.20

The graphical representation of $V(\Pi)$ in Allfuzzy case is in Figure 7.



D5l1 hR34 Data3 in the Allfuzzy case

Some further considerations concerning the $\alpha - cut$ values in all the data set enable us to state that our model allows us to describe how the investment decision is actually affected by a perceived increase in "fuzziness". For a pessimistic (optimistic) firm an increase in fuzziness decreases (increases) the perceived value of the project in comparison with the crisp value. On average - for most decision makers- an increase in fuzziness has a positive impact on the investment opportunity, i.e. it increases the perceived value of the project. As a consequence, the decision to invest is delayed in comparison with the absence of fuzziness. However, for pessimistic decision-makers imprecise information about the project value becomes available over time, which makes waiting with investment less valuable. Thus, for pessimistic firms higher fuzziness erodes the subjective value of the investment opportunity. Notice that this result is in keeping with the literature on real options and ambiguity aversion it contrasts with the impact of volatility in the standard real option theory.

1 6feNlilfei

The powerful contribution of fuzzy modelling can be shown in many fields and, especially in human sciences like economics, it can provide rigorous models (a detailed motivation of its use is in Zadeh [13]). In tis paper, we model the uncertainty involved in real options theory through fuzzy numbers represented in the LU model; when including fuzziness, the decision rule moves away from the original one and the choice to delay or not the investment becomes a key feature of the fuzzy model.

Our model allows us to quantify how the threshold values change with fuzzy parameters. Practically, we solve the investment/disinvestment decision in terms of a problem of fuzzy stopping time - that is, a fuzzification of the classical optimal stopping time - and our computation experiments provide the sensitivity analysis of the decision variable with respect to the relevant parameters. Our main results are that the decision to invest is delayed in comparison with the case of absence of fuzziness, the decision to disinvest is anticipated in a fuzzy environment. The fuzzy model we use provides the set of values within which the decision is taken by the manager. Our empirical validation, which is based on case studies from real world, seems to confirm the validity of the model and to open up further ways of research in finance in a fuzzy environment.

F RSRhRe NRt

[1] C. Carlsson, R. Fullér, A Fuzzy Approach to Real Option Valuation, *Fuzzy Sets and Systems*, 139, (2003), 297-312.

[2] A. Damodaran, The Promise and Peril of Real Options, Working Paper, Stern School of Business, 1999.

[3] A. Dixit, R. Pindyck , *Investment under Uncertainty*, Princeton University Press, (1994), Princeton.

[4] D. Dubois, H. Prade (ed), *Fundamentals of Fuzzy Sets*, Kluwer, Boston, The Handbooks of Fuzzy Sets Series, 2000.

[5] Y. Feng, Fuzzy stochastic differential systems. *Fuzzy Sets and Systems*, 115, 2000, 351–363.

[6] R. McDonald, D. Siegel, The Value of Waiting to Invest, *The Quarterly Journal of Economics*, 101, 4 1986, 707-728.

[7] S. C. Myers, Determinants of Corporate Borrowing, *Journal of Financial Economics*, 1977, 147-76.

[8] F.A. Roques, N.S. Savva, Price Cap Regulation and Investment Incentives under Demand Uncertainty, Cambridge Working Papers in Economics 0636, 2006.

[9] L. Stefanini, L.Sorini, M.L.Guerra, Parametric Representations of Fuzzy Numbers and Applications to Fuzzy Calculus, *Fuzzy Sets and Systems*, 157/18, 2006, 2423-2455.

[10] L. Stefanini, Differential Evolution Methods for the Fuzzy Extension of Functions, Working Papers 0705, University of Urbino Carlo Bo, Department of Economics, 2007.

[11] L. Stefanini, L. Sorini, M.L. Guerra, Fuzzy Numbers and Fuzzy Arithmetics, in W. Pedrycz, A. Skowron, and V. Kreinovich (Eds.): *Handbook of Granular Computing*, John Wiley & Sons, 2008, 249-283.

[12] L. Trigeorgis, Real options and Interactions with Financial Flexibility, *Financial Management*, 1993, 202-24.

[13] L.A. Zadeh, Is there a need for fuzzy logic?, *Information Sciences*, 178, 2008, 2751-2779.

ISBN: 978-989-95079-6-8

A dynamic classification method for the discrimination of evolving data

Laurent Hartert Moamar Sayed Mouchaweh Patrice Billaudel

Université de Reims Champagne-Ardenne, CReSTIC,
Moulin de la Housse, BP 1039, 51687 Reims Cedex, France
Email: laurent.hartert@univ-reims.fr, moamar.sayed-mouchaweh@univ-reims.fr, patrice.billaudel@univ-reims.fr

Abstract—*Classes issued of evolving systems are dynamic and their characteristics vary over the time. Assigning a pattern to a class is achieved using a classifier. Therefore, the classifier parameters must be adapted online in order to take into account the temporal changes in the classes' characteristics. This adaptation is based only on the recent and useful information carried out by the new incoming classified patterns. In this paper, we propose to develop the classification method Fuzzy Pattern Matching (FPM) to be operant in the case of dynamic classes. This development is based on the use of an incremental algorithm to follow the accumulated gradual temporal changes of classes' characteristics after the classification of each new pattern. When these changes reach a suitable predefined threshold, the classifier parameters are adapted online using the recent and representative patterns.*

Keywords— Classification, dynamic patterns, evolving systems, pattern recognition.

1 Introduction

Dynamic systems evolve between different normal or faulting functioning modes in the course of time. In statistical Pattern Recognition (PR) [12, 9], observations about system functioning modes are divided into groups of similar patterns, called classes, using an unsupervised learning method [10, 4] or human experience. These patterns, with their class assignments, constitute the learning set. They are represented by a set of *d* features, or attributes, so they can be viewed as *d*-dimensional features vectors, or points, in the feature space. A supervised learning method [9, 16] uses the learning set to build a classifier that best separates the different classes in order to minimize the misclassification error. The model of each class can be represented by a membership function which determines the membership value of a pattern to a class. Then, new incoming patterns are assigned to the class for which they have the maximum membership value. In supervised learning methods, the membership function can be generated using Probability Density Function (PDF) estimation based methods or heuristic based ones.

Patterns describing the system functioning can be static or dynamic. A static pattern is represented by a point in the feature space while a dynamic pattern is represented by a multidimensional trajectory. In this case, the feature space has an added dimension which is the time [3]. Classes can also be static or dynamic. Static classes are represented by restricted areas formed by similar static patterns in the feature space. Hence, the way in which patterns occur is irrelevant to their membership values. Therefore, the classifier's parameters remain unchanged with the time. However, data issued from evolving processes are non stationary. In this case, classes become dynamic and their characteristics change in the course of time. Thus, the classes' membership functions must be adapted to take into account these temporal changes. This requires an adaptive classifier with a mechanism for adjusting its parameters over the time. Hence, some of the new incoming points reinforce and confirm the information contained in the previous data, but the other ones could bring new information (creation, drift, fusion, splitting of classes, etc.). This new information could concern a change in operating conditions, development of a fault or simply more significant changes in the process dynamic.

The general principle of dynamic PR methods [14, 3, 7, 13] is to observe the change of some statistical properties of classes, in order to decide in which state the system is: unchanged, gradually changed or completely changed. Thus, the classifier parameters, i.e. the membership functions, will be respectively unchanged, slightly adapted, or relearned from scratch.

In the literature, the membership functions are generally adapted using two approaches [2, 3, 14]. The first one acts directly on the classifier parameters, by substituting or adding some recent and representative patterns to the learning set [3, 14], according to the state in which the system is. This adaptation is based only on the most recent batch of patterns selected by one of the two following methods. The first method uses a time window, with a fixed or a variable size, which permits to reduce or limit the growing size of the database by accepting the *n* most recent patterns [14, 17]. The size of the time window must be well-chosen to obtain a compromise between a fast adaptation and a sufficient number of representative patterns. The second method is based on the use of a template containing a fixed number of selected patterns according to their age and usefulness [11]. Nevertheless, it is subjective and difficult to estimate the utility of patterns. The second approach to adapt the membership functions is based on the use of evolving neural networks [1, 2, 7]. In [2], a potential function based on the distance between data points is defined for the new points. The first data point potential is considered as equal to 1 and it establishes the first neuron (or rule) which is considered as the prototype (or centre) of the first cluster. Then, the next new data points may possess a potential close or greater than the one of the prototype neuron. This point can reinforce or confirm the information contained in the previous ones, or if the point is more informative than the data used as prototype, a new neuron (new rule) is added. In [1], the neural network is based on a multiprototype Gaussian modeling of non convex classes. The activation

ISBN: 978-989-95079-6-8

function of each hidden neuron determines the membership degree of an observation to one prototype of a class. With the first acquisition, the network is initialized there is creation of the first prototype constituting the first class. The prototype is parameterized by its centre and an initial covariance matrix. Then, according to the membership degree of new acquisitions, the prototype (the hidden neuron) can be adapted, eliminated or a new prototype can be created.

Dynamic classification methods are used to solve several real problems. The application of [3] concerns the credit-scoring which aims to decide whether a new customer is a good or a bad risk according to changes in his consumption. In [1], the authors aim to detect and to follow up the progressive evolution of the functioning mode from normal to faulty one of a thermal regulator. This evolution is due to the age of the system's components or to other temporal factors in its environment. In [13], the author looks to estimate a ratio between the number of good parts in a completed batch of microelectronic chips and the number of parts obtained from the same batch if there is no default. Due to high complexity and variability of modern microelectronics manufacturing industry, this ratio is affected daily by hundreds of material-related, equipment-related and human-related factors. In [7], dynamic traffic data streams are treated in order to reduce the waiting time of drivers at the road intersections.

We use the classification method Incremental Fuzzy Pattern Matching (IFPM) [15] which is an incremental version of FPM [6]. The membership functions are based on the estimation of the marginal class probability density functions using histograms. Then, the membership functions are adapted incrementally after the classification of each new pattern. IFPM provides good result in the case of static classes based on static patterns. This method is simple and has a low and constant classification time according to the size of the database. However, IFPM has no mechanism to forget the no more useful patterns.

In this paper, we propose to develop IFPM to be operant in the case of dynamic classes. This development is called Dynamic FPM (DFPM). The paper is structured as follows. The two first parts are devoted respectively to the functioning of IFPM and DFPM. In the third part, the limits of IFPM and results of DFPM are presented in the case of dynamic classes. The last part concludes the paper and presents the perspectives of our future work.

2 Incremental Fuzzy Pattern Matching

2.1 Principle

The functioning of Incremental FPM (IFPM) involves the learning, the classification, and the incremental update phases.

2.1.1 Learning phase

Let X_i be the set of learning points belonging to the class C_i. Each class C_i contains N_i points, or patterns, in the feature space, $\Re^d$, formed by d attributes. X is the learning set, containing N points x belonging to C classes, so $X = \bigcup_{i=1}^{C} X_i$. The learning phase consists in building, based on X, a decision rule characterized by a set of membership functions. These latter are based on histograms allowing to estimate the conditional probability density of each class C_i according to each attribute j. The number h of bins b_k, $k \in \{1,2,..., h\}$ of a histogram is determined experimentally. The choice of h conditions the performances of IFPM. The lower $\min^j$ and upper $\max^j$ borders of each attribute are determined manually by an expert. In the case of dynamic classes, these borders must be well defined to contain all patterns of the classes' evolutions. This constitutes a shortcome of IFPM when the classes are dynamic. The width Δ^j of a bin according to the attribute j is defined by:

$$\Delta^j = \frac{(\max^j - \min^j)}{h} \tag{1}$$

The histogram or the distribution of probability $\{p_i^j(b_{ik}^j), i \in \{1,2,...,c\}, j \in \{1,2,...,d\}, k \in \{1,2,...,h\}\}$ for the class C_i according to the attribute j is determined by calculating the probability $p_i^j(b_{ik}^j)$ of each bin b_{ik}^j:

$$p_i^j(b_{ik}^j) = \frac{n_{ik}^j}{N_i} \tag{2}$$

where n_{ik}^j is the number of learning points of the class C_i which are located in the bin b_{ik}^j and N_i is the total number of points of the class C_i. Then, the resulting distribution of probability is transformed into a distribution of possibility $\{\pi_i^j(b_{ik}^j), i \in \{1,2,...,c\},\ j \in \{1,2,...,d\}, k \in \{1,2,...,h\}\}$ by using the transformation of Dubois and Prade [8]:

$$\pi_i^j(b_{ik}^j) = \sum_{z=1}^{h} \min(\ p_i^j(b_{iz}^j),\ p_i^j(b_{ik}^j)) \tag{3}$$

The possibility measure has the advantage to take into account the imprecision and the uncertainty contained in the data [8]. Finally, the density of possibility Π_i^j of the class C_i according to the attribute j is obtained by a linear interpolation of the bins centers of the histogram of possibility.

2.1.2 Classification phase

A pattern x is assigned to a known class C_i using the following three steps:

- determination of the membership possibility value π_i^j of x^j to each class C_i according to each attribute j. This possibility is obtained by projection of x^j on the density Π_i^j of each class C_i according to each attribute j (Fig. 1).

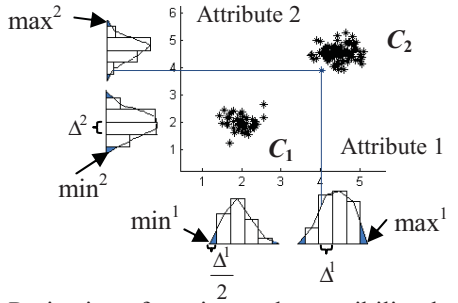


Figure 1: Projection of a point on the possibility densities. Δ^1 and Δ^2 are the bins' widths according respectively to attributes 1 and 2.

ISBN: 978-989-95079-6-8

- fusion, for each class C_i, of all membership possibility values $\pi_i^1, \pi_i^2, ..., \pi_i^d$ by the aggregation operator "minimum". The result of this fusion represents the membership possibility π_i of x to each class C_i,
- classification of the point x to the class for which it has the highest membership value.

2.1.3 *Incremental update phase*

Each new classified pattern carries information about the temporal changes of the class' characteristics. In order to take into account this new information, the membership functions must be relearned after the classification of each new pattern. An incremental learning algorithm is proposed in [15] to relearn the membership functions based on the most recently classified pattern in such a way that the previous learned information is reused. When a new point is classified in the class C_i, the number of points belonging to this class becomes $N_i + 1$, so the probability of each bin changes. If the point is located in the bin b_{ik}^j, the new probability $p_i'^j(b_{ik}^j)$ of this bin is [15]:

$$p_i'^j(b_{ik}^j) = \frac{n_{ik}^j}{N_i} \times \frac{N_i}{N_i+1} + \frac{1}{N_i+1} = p_i^j(b_{ik}^j) \times \frac{N_i}{N_i+1} + \frac{1}{N_i+1} \quad (4)$$

For each other bin, the new probability is:

$$p_i'^j(b_{iz}^j) = \frac{n_{iz}^j}{N_i} \times \frac{N_i}{N_i+1} = p_i^j(b_{iz}^j) \times \frac{N_i}{N_i+1}, z=1...h \ \& \ z \neq k \quad (5)$$

Using (4) and (5), the probability histograms can be updated after the classification of a new point without the need to recalculate them. We just need to know in which bin the new point is located. The histogram of possibility is then calculated by (3). IFPM has been developed to deal with data represented in a feature space of correlated attributes as the case of XOR database [5]. In the next part, the improvements realized on IFPM are presented.

3 Dynamic Fuzzy Pattern Matching

We have called Dynamic Fuzzy Pattern Matching (DFPM) the improved version of IFPM. DFPM integrates a mechanism to adjust the classifier's parameters when serious changes in the classes' characteristics are detected during a growing time window. The size of the time window depends of the application's dynamic. The next subsections detail the two phases of DFPM's functioning.

3.1 *Detection phase*

In the detection phase, three indicators are used to monitor the temporal changes of a system. The first indicator is a measure evaluating the bins' usefulness. This indicator is based on the accumulated temporal changes which have occurred in the probability distributions during a variable time sliding window. It considers the number of new points located in each bin of the probability histograms, as well as its initial probability before the assignment of the new pattern(s). Let $\Delta T = t_w - t_s$ be the growing size of the time window at the present time t_w. This time window starts at the instant t_s. Let $\{p_i^j(b_{ik}^j), k \in \{1,2,..., h\}\}$ be the probability histogram for the class C_i according to the attribute j at the instant t_s. The probability of each bin is calculated by (2). Let ΔN_i be the number of new classified points in the class C_i and let Δn_{ik}^j be the number of new points located in the bin b_{ik}^j at the present time t_w. The new probability of this bin is calculated by:

$$p_i'^j(b_{ik}^j) = \frac{n_{ik}^j + \Delta n_{ik}^j}{N_i + \Delta N_i} \quad (6)$$

where n_{ik}^j and N_i are respectively the number of points of the class C_i located in the bin b_{ik}^j and the total number of points of the class C_i before the classification of the new pattern(s). The accumulated temporal changes for each bin are evaluated as the difference between the bin's probabilities at the present time $p_i'^j(b_{ik}^j)$ and at the beginning $p_i^j(b_{ik}^j)$ of the growing time window:

$$\Delta p_i^j(b_{ik}^j) = p_i'^j(b_{ik}^j) - p_i^j(b_{ik}^j) = \frac{N_i.\Delta n_{ik}^j - n_{ik}^j.\Delta N_i}{N_i.(N_i + \Delta N_i)} \quad (7)$$

We define the usefulness measure as a gradual measure, $I_i^j(b_{ik}^j) : \{1,2,...,h\} \rightarrow [-1,1]$ for each bin b_{ik}^j of each class C_i according to each attribute j. This measure evaluates the usefulness of each bin based on the accumulated temporal changes in the probability distributions defined by (7):

$$I_i^j(b_{ik}^j) = \begin{cases} \dfrac{p_i'^j(b_{ik}^j) - p_i^j(b_{ik}^j)}{p_i^j(b_{ik}^j)} \in [-1,0] \ if \ \Delta p_i^j(b_{ik}^j) < 0 \\ \dfrac{p_i'^j(b_{ik}^j) - p_i^j(b_{ik}^j)}{p_i'^j(b_{ik}^j)} \in [0,1] \ if \ \Delta p_i^j(b_{ik}^j) > 0 \end{cases} \quad (8)$$

This measure is calculated after the classification of each new classified point. When the value of this indicator is negative for one bin, it indicates a decrease of its usefulness. This decrease is due to the fact that no new classified points, or just a few ones, are located in this bin. On the contrary, the positive values indicate an increase of the bin's usefulness due to the assignment of many new points in the bin, according to its initial value. It is interesting to take into account the old or initial probabilities. Indeed, if there is no important change in the probability distribution, it is normal that the bin with the highest probability receives more new points than the others. If most of the new classified points are located in the bins which possess a small initial probability, this indicates a serious change in the probability distribution and thus their usefulness indicators must have a more important positive value.

To detect an evolution, a second indicator, which is a residual, is calculated based on the usefulness' measure of all histograms' bins according to all attributes. The goal of this residual is to provide one indicator for each class, which evaluates the temporal accumulative changes in the probability distributions of this class. This residual is sensitive for these changes. Two thresholds *th*1 and *th*2 predefined by an expert according to the system dynamic are used to determine the size ΔT of the time window. The value of this residual is updated and compared after each pattern assignment to *th*1 and *th*2. If the residual value is

ISBN: 978-989-95079-6-8

inferior to *th*1, then no serious change in the class' characteristics is yet started. As long as the residual is inferior to *th*1, the measure of usefulness is based on the difference between the current bin probability and the one of the previous classified pattern. If the residual value is superior to *th*1 and inferior to *th*2, a serious change has begun to occur. The time window starts at this moment, $t=t_s$, and patterns are stored in a block, called the evolution block. During this window, the measure of usefulness is calculated using (8). Patterns are stored in the evolution block until the residual value reaches *th*2. At this moment the growing time window reaches its final size $t_w=t_f$. This residual is defined for each class C_i as follows:

$$0 \le R_i = \max\left(\frac{\sum_{k=1}^{h}\left|I_i^j(b_{ik}^j)\right|}{h}\right) \le 1, j \in \{1,..,d\} \quad (9)$$

When the residual value reaches *th2*, a serious change can be decided in a class so the classifier parameters (densities of possibility) must be adapted using only the patterns stored in the evolution block. The old patterns which have become obsolete are forgotten and an online update of the classifier's parameters can then be performed.

3.2 Adaptation phase

This phase consists in the adaptation of the probability histograms and the update of the lower and upper attributes' borders as well as the number of bins *h*. Indeed, in IFPM, these latter are defined initially using (1) and they remain unchanged. For dynamic classes, it is not realistic to define static attributes' borders. Thus, in DFPM these borders will be updated online with the classification of new points. If a new classified point involves the creation of a bin for a histogram, then a bin is added to the histogram according to this attribute so that *h* is incremented and the histograms' borders are updated:

$$\begin{cases} x^j < \min^j - \frac{\Delta^j}{2} \Rightarrow \min'^j = \min^j - \Delta^j \\ x^j > \max^j + \frac{\Delta^j}{2} \Rightarrow \max'^j = \max^j + \Delta^j \end{cases} \quad (10)$$

$$h' = h + 1 \quad (11)$$

The value of Δ^j is defined initially by (1). With this update *h* can be different according to each attribute. The adaptation of probability histograms is achieved using an incremental approach. Let $\left\{p_i^j(b_{ik}^j), k \in \{1,2,...,h\}\right\}$ be the probability histogram for the class C_i according to the attribute *j*. When the residual value of a class reaches *th*2, old patterns are deleted and only the new patterns of the evolution block are considered. So, the value of probability for each bin is calculated incrementally by:

$$p_i''^j(b_{ik}^j) = \frac{\Delta n_{ik}^j}{\Delta N_i} = \left(\frac{n_{ik}^j + \Delta n_{ik}^j}{N_i + \Delta N_i} - \frac{n_{ik}^j}{N_i} \times \frac{N_i}{N_i + \Delta N_i}\right) \times \frac{N_i + \Delta N_i}{\Delta N_i}$$

$$p_i''^j(b_{ik}^j) = \left(p_i'^j(b_{ik}^j) - (p_i^j(b_{ik}^j) \times \frac{N_i}{N_i + \Delta N_i})\right) \times \frac{N_i + \Delta N_i}{\Delta N_i}, k \in \{1,..h\} \quad (12)$$

This incremental adaptation permits to follow the evolution of classes online with a constant and low classification time. Then, the histogram of possibility is calculated using (3).
Fig. 2 illustrates the functioning phases of DFPM.

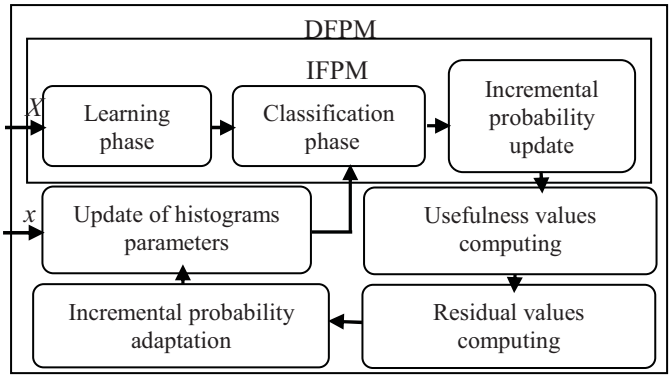


Figure 2: Functioning phases of DFPM.

4 Results with dynamic classes

Classes become dynamic if at least one of their characteristics changes with the time. For the sake of simplicity we have used a set of simulations described in a two dimension space. The obtained results can be extended for higher dimensions. The generated datasets follow a normal distribution of two independent variables (mean and standard deviation). We suppose that one unit of time corresponds to the classification of one point. This hypothesis is verified since DFPM has a constant and low classification time.

4.1 Class drift

Figure 3 shows a drift of a class in the course of time. The drift database has been created as follows:
- t=0: the possibility densities are already learned using 100 labeled patterns. The mean values of these data points are μ^1=3 and μ^2=3 (Fig. 3.a).
- t=1 to 100: 100 new incoming patterns, possessing the same initial class' characteristics (mean and standard deviation) occur in the class; thus there is no evolution or drift.
- t=101 to 200: a sudden change takes place in the class' mean values μ^j according to each attribute *j*, $j \in \{1,2\}$. This sudden change is followed by a slow drift of the mean value according to each attribute (Fig. 3.b):

$$\begin{cases} \mu^{1'}(t) = \mu^1 + 2 + \frac{4 \times (t-100)}{100} \\ \mu^{2'}(t) = \mu^2 + 2 + \frac{2 \times (t-100)}{100} \end{cases} \quad 101 \le t \le 200 \quad (13)$$

- t=201 to 300: 100 new patterns with the same characteristics as the ones of the drifted class occur. Fig 3.c presents the 0.1 membership level curve expected for the final drifted class. This curve contains all the points which have a membership possibility value to the class greater or equal to 0.1.

ISBN: 978-989-95079-6-8

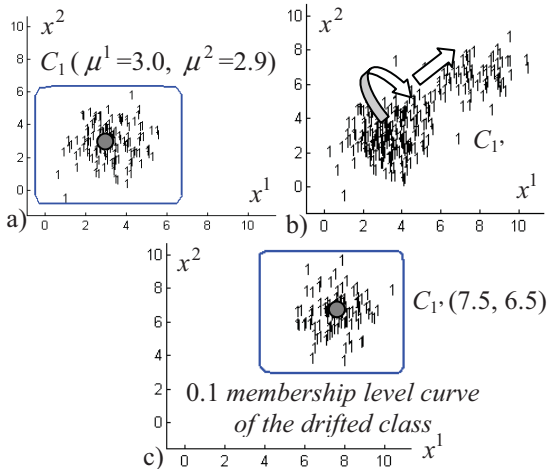


Figure 3: Evolution of the drift for the class C_1. The mean values of the class are presented before and after evolution between parentheses.

Fig. 4 presents the 0.1 membership level curve obtained by IFPM at the end of the class' drift ($t = 200$). We suppose here that each attribute borders were correctly determined in order to include the class drift.

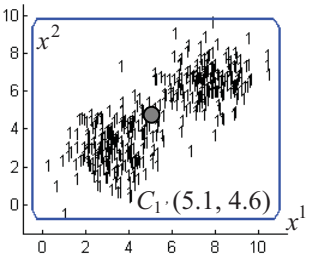


Figure 4: Classification of all patterns of the drifted class by IFPM.

All the patterns classified by IFPM are considered representative. Thus, the membership function, represented by the 0.1 membership level curve in Fig. 4, is not adapted to the drifted class with its new means values (see Fig. 3.c). This is due to the fact that all patterns were considered without eliminating any old and useless pattern.

The DFPM method is evaluated on the same drift of a class generated by (14). The residual values obtained progressively after the classification of each pattern in the class are represented in Fig. 5.

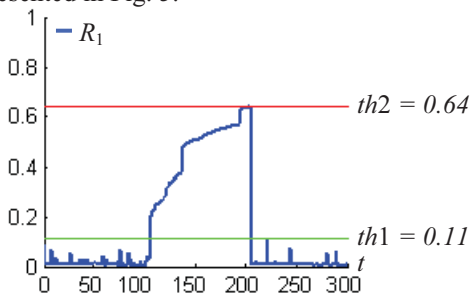


Figure 5: Residual values obtained by DFPM for the class drift.

Based on Fig. 5, we can conclude:

- between t=1 and t=100, the class' evolution has not yet begun.
- at t=105, the evolution of the class is detected. To detect this evolution the *th1* must be equal to 0.11.
- at t=204, the class drift is finished. To detect the end of this class drift the second threshold *th*2 must be equal to 0.64.
- finally, between t=205 and t=300 no evolution is observed so the residual values are small for the latest patterns and the class is stable. These patterns are classified in the class as usual.

Fig. 6 permits to see the membership level curve obtained by DFPM just before the evolution (Fig. 6.a) and the final result of classification after evolution when the adaptation of the histograms was realized (Fig. 6.b).

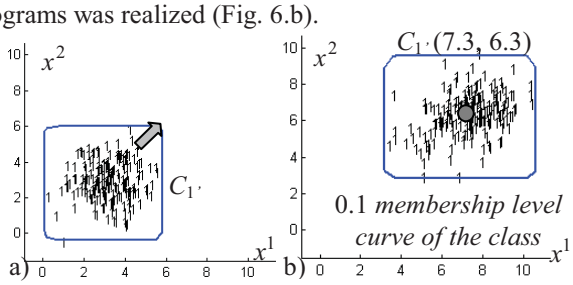


Figure 6: 0.1 membership level curve obtained by DFPM before evolution and after classification of all patterns.

The drift of the class has been well detected for both attributes. The classifier has been adapted with the representative patterns. However, the thresholds *th*1 and *th*2 must be well determined. A too big or too small *th*1 value leads to lose some representative patterns or to keep some obsolete ones, while a too big or a too small *th*2 value leads to detect too late or too early the class evolution.

The mean values of the class obtained by DFPM are close to those expected, contrary to those obtained by IFPM (Fig. 4). However, some patterns which represent the connection between the old class and the evolved one are still in the final class obtained by DFPM.

4.2 Rotation of class

The rotation of a class is presented in Fig. 7. The standard deviations of the class according to the two attributes are initiated at σ_1 =4 and σ_2 =0.25. The possibility densities are already learned using 150 learning patterns of the class (Fig. 7.a). Then, new incoming points evolve, between t=1 and t=150, to achieve a rotation of the class (Fig. 7.b).

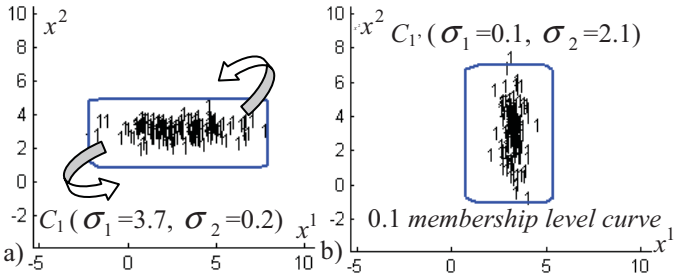


Figure 7: Rotation of the class C_1.

These changes are generated for the class according to each attribute j, $j \in \{1,2\}$, by:

$$\left.\begin{aligned} \mu_1^{1'} &= \mu_1^1 \times \sqrt{\frac{\sigma_2 \times t}{150}} \\ \mu_1^{2'} &= \mu_1^2 \times \sqrt{\frac{\sigma_1 \times t}{150}} \end{aligned}\right\} \quad 1 \le t \le 150 \qquad (14)$$

Fig. 8 presents the 0.1 membership level curve obtained by IFPM after the rotation of the class (t=150).

ISBN: 978-989-95079-6-8

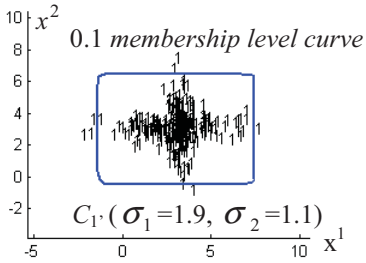


Figure 8: Final classification result obtained by IFPM for the patterns of the rotation's case.

All patterns have been classified in the class, but IFPM has not detected the rotation of the class. The membership level function is not adapted and useless patterns are still in the class. With DFPM, the residual values obtained after the classification of each pattern in the class are represented in Fig. 9.

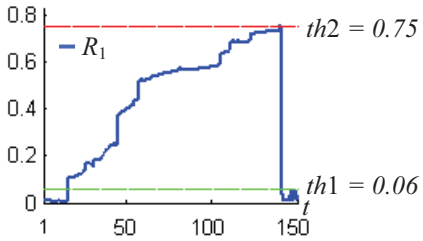


Figure 9: Residual values obtained by DFPM for the rotation of the class.

Based on Fig. 9, we can conclude that the use of $th1$=0.06 and $th2$=0.75 leads to well detect the class rotation.
Fig. 10 presents the membership level curve obtained by DFPM after classification of the evolving patterns when the adaptation of the histograms was realized.

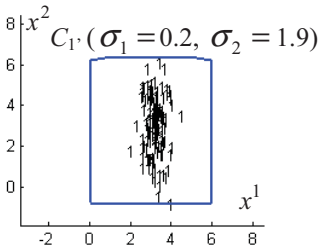


Figure 10: Classification result obtained by DFPM after adaptation of the class.

The classifier has been adapted with all representative patterns detected since the start of the class rotation. The new conserved useful patterns are the only patterns used to calculate the membership function. The resulting class corresponds well to the expected one, contrary to the class obtained by IFPM (Fig. 8).

5 Conclusions

In this paper, the Incremental version of the classification method Fuzzy Pattern Matching (IFPM) has been developed to discriminate dynamic classes. This development, called Dynamic FPM (DFPM), follows online the accumulated gradual changes in classes' probability distributions after the classification of each new pattern. Bins' usefulness values are calculated for each class and attribute. Based on these usefulness values, a residual value is calculated for each class in order to follow the evolution of each class. Then, the classes histograms are adapted in an incremental manner using the recent and useful patterns. This adaptation is achieved when the residual value reaches a suitable threshold according to the system dynamic. The DFPM method will be applied on two real applications. The first one concerns the following of a treatment's evolution for hemiplegic's patient according to its response to a medical treatment. The second application is a folding metal system which evolves from a normal mode to a faulty one according to temporal changes in the characteristics of the system and to the wear of its tools.

References

[1] H. Amadou Boubacar, *et al.,* AUDyC Neural Network using a new Gaussian Densities Merge Mechanism. *ICANNGA05*, pp. 155-158. Coimbra, Portugal, 2005.

[2] P.P. Angelov, A fuzzy controller with evolving structure. *Information Sciences* 161 (1-2), pp.21-35, 2004.

[3] L. Angstenberger, Dynamic Fuzzy Pattern Recognition. Dissertation, *Fakultät für Wirtschaftswissenschaften der Rheinisch-Westfälischen Technischen Hochschule*, Aachen, Germany, 2000.

[4] J.C. Bezdek, Pattern Recognition with fuzzy objective function algorithms, *Plenum Press,* 1981.

[5] M. S. Bouguelid, *et al.*, New results on diagnosis by pattern recognition, 4th International Conference on Informatics in Control, Automation and Robotic, pp. 167-172, Angers, France, 2007.

[6] M. Cayrol, *et al.,* Fuzzy Pattern Matching, Kybernetes, 11, pp. 103-116, 1982.

[7] L. Cohen, *et al.,* Incremental Info-Fuzzy Algorithm for Real Time Data Mining of Non-Stationary Data Streams, *TDM Workshop*, Brighton UK, 2004.

[8] D. Dubois, H. Prade, On possibility/probability transformations. *Fuzzy Logic*, pp. 103-112, 1993.

[9] R.O. Duda, *et al*., Pattern Classification 2nd edition. *Wiley-Interscience*, 2001.

[10] H. Frigui, R. Krishnapuram, Clustering by competitive agglomeration, *Pattern Recognition 30* (7), pp. 1109-1119, 1997.

[11] W.J. Gibb, *et al.*, Adaptive classification of myocardial electrogram waveforms, *IEEE Transactions on Biomedical Engineering*, N° 41, pp. 804-808, 1994.

[12] A.K. Jain, *et al*., Statistical Pattern Recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22 (1), pp. 4-37, 2000.

[13] M. Last, Online classification on nonstationary data streams, *Intelligent Data Analysis*, vol. 6 (2), pp. 129-147, 2002.

[14] G. Nakhaeizadeh, *et al.,* Dynamic Supervised Learning. Some Basic Issues and Application Aspects. *Classification and Knowledge Organization*. Springer Verlag, Berlin, pp. 123-135, 1997.

[15] M. Sayed Mouchaweh, *et al*., Incremental Learning in fuzzy pattern matching. *Fuzzy Sets and Systems* 132 (1), pp. 49-62, 2002.

[16] C.W. Therrien, Decision Estimation and Classification: An Introduction to Pattern Recognition and Related Topics. *John Wiley & Sons*, New York, 1989.

[17] G. Widmer, M. Kubat, Learning in the Presence of Context Drift and Hidden Contexts, *Machine Learning 23*, pp. 69-101, 1996.

ISBN: 978-989-95079-6-8

Possibility theory and formal concept analysis in information systems

Didier Dubois Henri Prade

IRIT, University of Toulouse III
118 Route de Narbonne, 31062 Toulouse Cedex 9, France
Email: dubois@irit.fr, prade@irit.fr

Abstract— *The setting of formal concept analysis presupposes the existence of a relation between objects and properties. Knowing that an unspecified object has a given property induces a formal possibility distribution that models the set of objects known to possess this property. This view expressed in a recent work by the authors of the present paper, has led to introduce the set-valued counterpart to the four set functions evaluating potential or actual, possibility or necessity that underlie bipolar possibility theory, and to study associated notions. This framework puts formal concept analysis in a new, enlarged perspective, further explored in this article. The "actual (or guaranteed) possibility" function induces the usual Galois connexion that defines the notion of a concept as the pair of its extent and its intent. A new Galois connexion, based on the necessity measure, partitions the relation in "orthogonal" subsets of objects having distinct properties. Besides, the formal similarity between the notion of division in relational algebra and the "actual possibility" function leads to define the fuzzy set of objects having* most *properties in a set, and other related notions induced by fuzzy extensions of division. Generally speaking, the possibilistic view of formal concept analysis still applies when properties are a matter of degree, as discussed in the paper. Lastly, cases where the object / property relation is incomplete due to missing information, or more generally pervaded with possibilistic uncertainty is also discussed.*

Keywords— possibility theory; formal concept analysis.

1 Introduction

Formal concept analysis [1] exploits the duality between objects and properties in a lattice theory setting, which has led to an original and practical view of the notion of a formal concept with application in data mining. A concept is then a pair made of a set of objects and a set of properties that are in mutual correspondence. These two sets are the extent and of the intent of the concept respectively. In this framework, properties are *binary*, and *complete* information is assumed about the relation linking objects and properties.

Fuzzy set theory [2] has emphasized the idea that properties are not always all-or-nothing notions, but are rather often a matter of degree. This has led to an extension of the original formal concept analysis setting by allowing intermediate truth values for the propositions "object x has property y" [3]. However, complete information is still assumed. Namely, for any pair (object, property), it it is known to what degree the object has the property. Besides, fuzzy sets have also been the starting point for the development of a new approach for the representation of uncertainty, named possibility theory [4, 5]. Fuzzy sets then have a disjunctive reading and represent states of incomplete information, i.e. (soft) restrictions on the mutually exclusive possible values of a single-valued variable.

The authors of the present paper have recently advocated the interest of a possibilistic reading of formal concept analysis where the possibility theory set-functions are shown to be meaningful in formal concept analysis [6]. Under this view, the set of objects known to possess a given property plays the role of a formal possibility distribution that restricts the possible value (identity) of an unspecified object only described as having the given property. This leads to an enlarged setting that we continue to investigate in this paper. Besides, this enlarged setting can be itself extended either by allowing properties to be non-Boolean, or by considering that the relation between objects and properties may be incompletely known.

The next section provides the background on a possibility theory-inspired view of formal concept analysis. Four basic operators that are the counterparts of the four basic set-functions in bipolar possibility theory are introduced, leading to consider another Galois connexion distinct from the usual one that gives birth to the notion of concept. Section 3 briefly describes the extension of the enlarged setting to fuzzy properties, i.e., when the relation linking objects and properties becomes fuzzy. Section 4 relates the operator underlying classical formal concept analysis and the notions of division (and quotient) in relational algebra. This provides a basis for defining the fuzzy set of objects having $most$ properties in a set (where $most$ is a fuzzy quantifier) and other related notions. Section 5 deals with the situation where information is incomplete or uncertain. Due to the lack of space, the paper only outlines new ideas, without developing them.

2 A possibility theory view of concept analysis

A formal information system is viewed here as a binary relation R between a set Obj of objects and a set $Prop$ of Boolean properties. Some authors speak of 'attribute' instead of 'property'. As we shall see in Section 5, this distinction only matters for attributes with non-binary domains. R is called *context* in formal concept analysis. If $X \subseteq Obj$, $\overline{X}$ is its complement $Obj \setminus X$. The notation $(x, y) \in R$ means that object x has property y. $R(x) = \{y \in Prop | (x, y) \in R\}$ is the set of properties of object x. Similarly, $R^{-1}(y) = \{x \in Obj | (x, y) \in R\}$ is the set of objects having property y. Its characteristic function induces a two-valued possibility distribution π:

$$\forall x \in Obj, \quad \pi(x) = \begin{cases} 1 & \text{if } x \in R^{-1}(y) \\ 0 & \text{otherwise,} \end{cases}$$

Intuitively speaking, if all we know about an unknown object is that it has property y then this object may be any x such that $\pi(x) = 1$ in context R. Thus, the relation R for a particular property y plays the role of a possibility distribution π encoding a set of possible values for x.

ISBN: 978-989-95079-6-8

2.1 Possibility theory

In possibility theory [5], two "measures" are associated with a possibility distribution π defined on a universe U as the characteristic (membership) function of a fuzzy set E representing the available information (in the above case $U = Obj$, and $E = R^{-1}(y)$ is an ordinary subset of U). Namely

i) a possibility measure Π (or "potential possibility"):

$$\Pi(A) = \max_{x \in A} \pi(x).$$

It estimates to what extent event A is *consistent with* the information represented by π. $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$ is the characteristic property of possibility measures [4].

ii) a dual measure of necessity N, expressing that an event is all the more necessarily (certainly) true as the opposite event is more impossible. N thus reflects an "actual necessity":

$$N(A) = 1 - \Pi(\overline{A}) = 1 - \max_{x \notin A} \pi(x),$$

where $\overline{A} = U \setminus A$. $N(A)$ estimates to what extent event A is *implied* by the information E represented by π (inasmuch as this information entails that any realization of $\overline{A}$ is more or less impossible). Necessity measures are characterized by the decomposability property $N(A \cap B) = \min(N(A), N(B))$.

Π and N are based on the $maximum$ of π over A and $\overline{A}$ respectively; two other set-functions [5] use the $minimum$:

iii) a measure of "actual (or guaranteed) possibility"

$$\Delta(A) = \min_{x \in A} \pi(x),$$

which estimates to what extent all elements in A are possible. Δ can be also termed "$sufficiency\ measure$" since $\Delta(A) = 1$ is enough for ensuring that all realizations of A are actually possible. Clearly, $\Delta \leq \Pi$. Note also that $\Delta(A)$ and $N(A)$ are unrelated. $\Delta(A \cup B) = \min(\Delta(A), \Delta(B))$ is the characteristic property of guaranteed possibility measures.

iv) a dual measure of "potential necessity or certainty"

$$\nabla(A) = 1 - \Delta(\overline{A}) = 1 - \min_{x \notin A}(\pi(x))$$

which estimates to what extent there exists at least one value in the complement of A that has a zero (or more generally a low) degree of possibility. This is clearly a $necessary$ condition for having "$x \in A$" somewhat certain. Property $\nabla(A \cap B) = \max(\nabla(A), \nabla(B))$ characterizes these measures.

2.2 An enlarged formal concept analysis setting

These four set functions make sense in the formal concept analysis. Namely, four remarkable sets can be defined:

$$R^{\Pi}(X) = \{y \in Prop | R^{-1}(y) \cap X \neq \emptyset\}$$
$$R^{N}(X) = \{y \in Prop | R^{-1}(y) \subseteq X\}$$
$$R^{\Delta}(X) = \{y \in Prop | R^{-1}(y) \supseteq X\}$$
$$R^{\nabla}(X) = \{y \in Prop | R^{-1}(y) \cup X \neq Obj\}$$

whose respective characteristic functions are $\Pi(X)$, $N(X)$, $\Delta(X)$, and $\nabla(X)$. Their meanings are as follows w. r. t. a subset of objects X in context R.

- $R^{\Pi}(X)$ is the set of properties that are associated with at least one object in X. Formally, we have

$$R^{\Pi}(X) = \cup_{x \in X} R(x).$$

$R^{\Pi}(X)$ is such that any object that satisfies one of them is possibly in X. In other words, if an object has none of the properties in $R^{\Pi}(X)$ then it cannot belong to X. Moreover, we have $R^{\Pi}(X_1 \cup X_2) = R^{\Pi}(X_1) \cup R^{\Pi}(X_2)$.

- $R^{N}(X)$ is the set of properties s. t. any object that satisfies one of them is necessarily in X. Having any property in $R^{N}(X)$ is a sufficient condition for belonging to X, and $R^{N}(X) = \overline{R^{\Pi}(\overline{X})} = Prop \setminus R^{\Pi}(\overline{X})$. Thus,

$$R^{N}(X) = \cap_{x \notin X} \overline{R(x)}.$$

and $R^{N}(X_1 \cap X_2) = R^{N}(X_1) \cap R^{N}(X_2)$.

- $R^{\Delta}(X)$, set of properties shared by all objects in X is

$$R^{\Delta}(X) = \cap_{x \in X} R(x).$$

In other words, satisfying all properties in $R^{\Delta}(X)$ is a necessary condition for an object to belong to X. $R^{\Delta}(X)$ is a partial conceptual characterization of objects in X: objects in X should have all the properties of $R^{\Delta}(X)$ and may have some others (that are not shared by all objects in X). It is worth noticing that $\overline{R^{\Pi}(X)}$ provides a negative conceptual characterization of objects in X since it gathers all the properties that are never satisfied by any object in X. Besides, it can be checked that $R^{N}(X) \cap R^{\Delta}(X)$ is the set of properties possessed by all objects in X and only by them. Moreover, we have $R^{\Delta}(X_1 \cup X_2) = R^{\Delta}(X_1) \cap R^{\Delta}(X_2)$.

- Note that $R^{\nabla}(X) = \overline{R^{\Delta}(\overline{X})} = Prop \setminus R^{\Delta}(\overline{X})$. Thus $R^{\nabla}(X)$ is the set of properties in $Prop$ that are not satisfied by at least one object in $\overline{X}$, i.e. $R^{\nabla}(X)$ is the set of properties that some object in $\overline{X}$ misses. In other words, in context R, for any property in $R^{\nabla}(X)$, there exists at least one object outside X that misses it. We have

$$R^{\nabla}(X) = \cup_{x \notin X} \overline{R(x)}.$$

and the following decomposability property holds $R^{\nabla}(X_1 \cap X_2) = R^{\nabla}(X_1) \cup R^{\nabla}(X_2)$.

Note that $R^{\Pi}(X)$ and $R^{N}(X)$ get larger when X increases, while $R^{\Delta}(X)$ and $R^{\nabla}(X)$ get smaller. The four modal-like operators R^{Π}, R^{N}, R^{Δ}, and R^{∇} have been considered by Düntsch and Orlowska [7] in the Boolean algebra setting, where R^{Δ} is called *sufficiency* operator, and its representation capabilities are studied. Taking inspiration as the previous authors from rough sets [8], Yao [9] also lays bare these four subsets. In both cases, the four operators were introduced without any mention of possibility theory.

Results in possibility theory have their counterparts in the enlarged formal concept analysis setting, as, e.g., [6]: If R^{-1} is s. t. $\forall y \in Prop, R^{-1}(y) \neq \emptyset$ and $R^{-1}(y) \neq Obj$, then

$$\forall X \subseteq Obj, \quad R^{N}(X) \cup R^{\Delta}(X) \subseteq R^{\Pi}(X) \cap R^{\nabla}(X). \quad (1)$$

Assuming that the property y is non trivial with respect to the set of objects Obj, i.e. $R^{-1}(y) \neq \emptyset$ (at least one object has property y) and $R^{-1}(y) \neq Obj$ (at least one object has not property y), then the four sets $R^{\Pi}(X)$, $R^{\Delta}(X)$, $R^{N}(X)$,

ISBN: 978-989-95079-6-8

Situation	$y \in R^{\Pi}(X)$	$y \in R^{\Delta}(X)$	$y \in R^{N}(X)$	$y \in R^{\nabla}(X)$
1. $X = \overline{R^{-1}(y)}$	0	0	0	0
2. $X \subset \overline{R^{-1}(y)}$	0	0	0	1
3. $\overline{R^{-1}(y)} \subset X$	1	0	0	0
4. $R^{-1}(y) \cap X \neq \emptyset$, $R^{-1}(y) \cap \overline{X} \neq \emptyset$	1	0	0	1
5. $R^{-1}(y) \subset X$	1	0	1	1
6. $X \subset R^{-1}(y)$	1	1	0	1
7. $R^{-1}(y) = X$	1	1	1	1

Figure 1: The seven possible positions of X and $R^{-1}(y)$

	objects							
	1	2	3	4	5	6	7	8
a					×	×	×	×
b					×	×		
c						×	×	×
d					×	×	×	×
e							×	
f					×	×		×
g	×	×	×	×				
h		×	×	×				
i				×				

Figure 2: R: a relation objects/properties a, b, c, ..., i

$R^{\nabla}(X)$ are necessary and sufficient for describing (and distinguishing between) the seven relative possible positions of X and $R^{-1}(y)$, as shown in Table 1, where 1 (resp. 0) stands for $y \in A$ (resp $y \notin A$) where A is the set $R^{*}(X)$ associated to the column (and $*$ is Π, Δ, N, or ∇). Note that the 9 = 16 - 7 remaining binary 4-tuples are ruled out by the constraints induced by (1), namely $R^{N}(X) \subseteq R^{\Pi}(X)$, $R^{\Delta}(X) \subseteq R^{\Pi}(X)$, $R^{\Delta}(X) \subseteq R^{\nabla}(X)$, $R^{N}(X) \subseteq R^{\nabla}(X)$. For instance, the "trivial" cases $R^{-1}(y) = \emptyset$ and $R^{-1}(y) = Obj$ (ruled out by the constraints) would be captured by distinct 4-tuples in Table 1, namely $(0 \quad 0 \quad 1 \quad 1)$ and $(1 \quad 1 \quad 0 \quad 0)$ respectively.

The above characterization of remarkable sets of properties w. r. t. a set of objects can be easily adapted for defining the corresponding sets of objects associated to a set of properties $Y \in Prop$: namely $R^{-1\Pi}(Y)$, $R^{-1N}(Y)$, $R^{-1\Delta}(Y)$, and $R^{-1\nabla}(Y)$. Their definitions can be easily obtained by swapping R and R^{-1} and exchanging the roles of the sets Obj and $Prop$. Namely,

$$R^{-1\Pi}(Y) = \{x \in Obj | R(x) \cap Y \neq \emptyset\} = \cup_{y \in Y} R^{-1}(y)$$
$$R^{-1N}(Y) = \{x \in Obj | R(x) \subseteq Y\} = \cap_{y \notin Y} \overline{R^{-1}(y)}$$
$$R^{-1\Delta}(Y) = \{x \in Obj | R(x) \supseteq Y\} = \cap_{y \in Y} R^{-1}(y)$$
$$R^{-1\nabla}(Y) = \{x \in Obj | R(x) \cup Y \neq Obj\} = \cup_{y \notin Y} \overline{R^{-1}(y)}.$$

Remark The above operators can be combined together. For instance, consider an object x_0. Let $R(x_0)$ be the set of its (known) properties. Compute $R^{-1\Delta}(R(x_0))$, the set of objects that share these properties. Then get $R^{\Pi}(R^{-1\Delta}(R(x_0)))$, which is the set of properties that are associated with at least one object sharing the properties of x_0. Viewing the table $Obj \times Prop$ as the information pertaining to a repertory of cases, and x_0 as a partially known extra object (not in Obj) for which one tries to guess other properties, the expression $R^{\Pi}(R^{-1\Delta}(R(x_0)))$ may be viewed as the result of a *case-based reasoning* procedure, i.e. a set of potential properties that x_0 may also have. Besides, its subset $R^{N}(R^{-1\Delta}(R(x_0)))$ is the set of properties that alone characterize the objects sharing the properties of x_0. Thus, if one of the properties in $R^{N}(R^{-1\Delta}(R(x_0)))$ is not already among the known properties of x_0, it may be considered as a serious candidate property for x_0.

2.3 Galois connexions

In formal concept analysis, the pair of set valued functions R^{Δ} and $R^{-1\Delta}$ induces a Galois connexion [10] between 2^{Obj} and 2^{Prop}. Then, a *formal concept* is a pair (X, Y) such that $X = \{x \in Obj | R(x) \supseteq Y\}$ and $Y = \{y \in Prop | R^{-1}(y) \supseteq X\}$, i.e. such that $X = R^{-1\Delta}(Y)$ and $Y = R^{\Delta}(X)$, X is called its *extent* and Y its *intent*. In other words, in a formal concept (X, Y), Y is the set of properties shared by all the objects in X, and X is the set of objects that possess all the properties in Y. Then $X \times Y \subseteq R$, i.e. $\forall x \in X, \forall y \in Y, (x, y) \in R$. A formal concept is a maximal pair that satisfies the latter condition (where maximality is taken in the sense of set inclusion).

Putting formal concept analysis in the perspective of possibility theory, it becomes then natural to also consider
- the pairs (X, Y) s. t. $X = R^{-1\Pi}(Y)$ and $Y = R^{\Pi}(X)$;
- the pairs (X, Y) s. t. t $X = R^{-1N}(Y)$ and $Y = R^{N}(X)$;
- the pairs (X, Y) s. t. $X = R^{-1\nabla}(Y)$ and $Y = R^{\nabla}(X)$.

First, observe that $X = R^{-1\nabla}(Y)$ and $Y = R^{\nabla}(X)$ holds if and only if $\overline{X} = R^{-1\Delta}(\overline{Y})$ and $\overline{Y} = R^{\Delta}(\overline{X})$ holds, i.e., if $(\overline{X}, \overline{Y})$ is a formal concept, due to the duality between operators R^{Δ} and R^{∇}. Similarly, $X = R^{-1\Pi}(Y)$ and $Y = R^{\Pi}(X)$ holds if and only if $\overline{X} = R^{-1N}(\overline{Y})$ and $\overline{Y} = R^{N}(\overline{X})$ holds. But, it can be easily seen that a pair (X, Y) such that $X = R^{-1N}(Y)$ and $Y = R^{N}(X)$, i.e. such that $X = \{x \in Obj | R(x) \subseteq Y\}$ and $Y = \{y \in Prop | R^{-1}(y) \subseteq X\}$ is not generallly a formal concept, as now exemplified. This Galois connexion has been introduced by [11] on a formal basis, but its practical meaning was apparently not really discussed.

Example 1 *We consider an example of relation R described by the table of Figure 2. This relation defines the links between eight objects $Obj = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and nine properties $Prop = \{a, b, c, d, e, f, g, h, i\}$. There is a "×" in the cell corresponding to an object x and to a property y if the object x has property y, in other words the "×"s describe the relation R (or context). An empty cell corresponds to the fact that $(x, y) \notin R$, i.e., it is known that object x has not property y. It can be checked that the pairs $(\{1, 2, 3, 4\}, \{g, h, i\})$, $(\{5, 6, 7, 8\}, \{a, b, c, d, e, f\})$ are pairs (X, Y) such that $X = R^{-1N}(Y)$ and $Y = R^{N}(X)$. These two pairs are not formal concepts. They are disjoint w. r. t. both Obj and $Prop$. Examples of formal concepts are $(\{2, 3, 4\}, \{g, h\})$, $(\{6, 7, 8\}, \{a, c, d\})$, or $(\{5, 6, 7, 8\}, \{a, d\})$. Note that these latter examples are here obtained by considering appropriate subsets in the previous pairs.*

A pair (X, Y) that satisfies $X = R^{-1N}(Y)$ and $Y = R^{N}(X)$

ISBN: 978-989-95079-6-8

properties \ objects	1	2	3	4	5	6	7	8
a					×	×	×	×
b					×	×		
c						×	×	×
d				×	×	×	×	×
e							×	
f					×	×		×
g	×	×	×	×				
h		×	×	×				
i				×				

Figure 3: R': relation R modified

is such that all the objects in X possess *at least one* property in Y and the properties in Y are only (possibly) possessed by the objects in X. While the intent of a formal concept is a *conjunction* of properties, the pairs (X, Y) forming an "N-block" correspond to sets of objects defined through *disjunctions* of properties. Finding such pairs, which may not exist, aims at decomposing the relation R into independent blocks without object or property in common, as in the Figure 2 example. When such a decomposition no longer holds, as in Figure 3, pairs (X, Y) such that $X = R^{-1N}(Y)$ and $Y = R^N(X)$ no longer exist, except for the trivial pair (Obj, $Prop$), as shown in the next example.

Example 2 *Let us now consider a modified version of relation R, say R′, depicted in the table of Figure 3, where object 4 has also the additional property d now. Then it can be checked that we still have* $R'^N(\{1,2,3,4\}) = \{g,h,i\}$*, but* $R'^{-1N}(\{g,h,i\}) = \{1,2,3\}$*, since* $R'(4) = \{d,g,h,i\} \not\subseteq \{g,h,i\}$*. Similary,* $R'^{-1N}(\{a,b,c,d,e,f\}) = \{5,6,7,8\}$*, but* $R'^N(\{5,6,7,8\}) = \{a,b,c,e,f\}$ *since* $R'^{-1}(d) = \{4,5,6,7,8\} \not\subseteq \{5,6,7,8\}$*. Thus the pairs* $(\{1,2,3,4\},\{g,h,i\})$*,* $(\{5,6,7,8\},\{a,b,c,d,e,f\})$ *are no longer pairs* (X, Y) *such that* $X = R'^{-1N}(Y)$ *and* $Y = R'^N(X)$ *in the new context R'.*

3 Handling fuzzy properties in the new setting

In the previous section, properties were supposed to be Boolean. Hence, when an object satisfies a property, it fully satisfies it: there is no intermediary degree of satisfaction since the property is not gradual. Thus, the relation linking objects and properties was all-or-nothing. When properties become a matter of intensity, i.e., when an object may have a property to some degree, the relation R between objects and properties becomes fuzzy. However, when relaxing the Booleanity assumption, we still assume that we have complete information. Namely, it is known to what extent α object x has property y for any pair (x, y), which is denoted $\mu_R(x, y) = \alpha$. Then, $\mu_{R^{-1}(y)}(x) = \alpha$ denotes the fact that object x satisfies property y at degree α where $\mu_{R^{-1}(y)}$ is the membership function of the fuzzy set of objects that constitutes the extension of $R^{-1}(y)$. Such an extension to standard concept analysis (based on the operator called here Δ) has been studied by Belohlavek [3]; see also [12].

Then, the four operators introduced in the previous section easily extend to the case where relation R is fuzzy. Namely,

$$\begin{array}{llll} \mu_{R^\Pi(X)}(y) & = \Pi_y(X) & = \max_{x\in X} & \mu_{R^{-1}(y)}(x) \\ \mu_{R^N(X)}(y) & = N_y(X) & = \min_{x\notin X} & 1-\mu_{R^{-1}(y)}(x) \\ \mu_{R^\Delta(X)}(y) & = \Delta_y(X) & = \min_{x\in X} & \mu_{R^{-1}(y)}(x) \\ \mu_{R^\nabla(X)}(y) & = \nabla_y(X) & = \max_{x\notin X} & 1-\mu_{R^{-1}(y)}(x) \end{array}$$

where Π_y, N_y, Δ_y and ∇_y are respectively potential possibility, actual necessity, actual possibility, and potential necessity measures, based on the gradual possibility distribution $\pi = \mu_{R^{-1}(y)}$. They thus enjoy the corresponding characteristic decomposability properties of these respective measures.

The following results are straightforward:

- $\mu_{R^N(X)}(y) = 1 - \mu_{R^\Pi(\overline{X})}(y)$;
- $\mu_{R^\nabla(X)}(y) = 1 - \mu_{R^\Delta(\overline{X})}(y)$
- if $\mu_{R^\Pi(X)}(y) = \alpha$ then $\exists x \in X, \mu_R(x, y) = \alpha$ and $\forall x \in X, \mu_R(x, y) \leq \alpha$
- if $\mu_{R^N(X)}(y) = \alpha$ then $\mu_R(x, y) > 1 - \alpha \Rightarrow x \in X$
- if $\mu_{R^\Delta(X)}(y) = \alpha$ then $x \in X \Rightarrow \mu_R(x, y) \geq \alpha$
- if $\mu_{R^\nabla(X)}(y) = \alpha$ then $\exists x \notin X, \mu_R(x, y) = 1 - \alpha$ and $\forall x \notin X, \mu_R(x, y) \geq 1 - \alpha$

The first two results extend duality relations to the graded case. The other ones express the meaning of each fuzzy set. Thus a property belongs to $R^\Pi(X)$ to degree α inasmuch as objects in X possess this property to at most degree α. Then $h(R^\Pi(X)) = \max_{y\in Y} \mu_{R^\Pi(X)}(y) = 0$ means that no object in X possesses a property in Y to any extent. A property belongs to $R^N(X)$ to degree α if any object possessing this property to a degree greater than $1 - \alpha$ necessarily belongs to X. In particular, any object possessing this property to some positive degree belongs to X, if $\alpha = 1$. A property belongs all the more to $R^\Delta(X)$ as any object in X possesses this property to a greater degree. Lastly, $\hbar(R^\nabla(X)) = \min_{y\in Y} \mu_{R^\nabla(X)}(y) = \alpha$ means that for any property y in Y objects outside X possess this property to at most to degree $1 - \alpha$. In particular, if $\hbar(R^\nabla(X)) = 1$, for any property in Y there exists an object outside X that misses it.

Moreover, we have the following counterpart to (1):

Proposition 1 *If* $R^{-1}(y)$ *is such that for* $y \in Prop$*,* $h(\mu_{R^{-1}(y)}) = 1$ *and* $\hbar(\mu_{R^{-1}(y)}) = 0$*, then*

$$\forall X \subseteq Obj, \quad \max(N_y(X), \Delta_y(X)) \leq \min(\Pi_y(X), \nabla_y(X)).$$

Lastly, it is clear that when nesting the operators as in the Remark at the end of Section 2.2, or when extending Galois connexions when R is fuzzy, we are led to extend again the above definitions to cases where X and Y become fuzzy sets themselves. From a possibility theory point of view, this means defining generalized measures for fuzzy events. Depending on the properties we want to preserve, several choices are possible here (see on this point [13], especially pages 51–64), which would lead to different extensions for the four operators. Choosing the appropriate extensions depends on the intended use and interpretation of the definition we want to generalize. We leave these questions for further research.

ISBN: 978-989-95079-6-8

4 Generalized quotient and fuzzy quantifiers

In a relational database, given an ordered set of attributes $\mathbf{A} = \{\mathcal{A}^1, ..., \mathcal{A}^n\}$, information is stored in a relational table $\mathcal{R}$ where each column corresponds to an attribute, and a row to an object belonging to *Obj*. Thus a cell in such a relational table corresponds to the value of an attribute for an object. Any row in the relational table $\mathcal{R}$ is also called a 'tuple'. A 'tuple' is thus an ordered set of attribute values pertaining to an object. Let $\mathcal{T}u(\mathcal{R})$ denote the set of tuples in $\mathcal{R}$. Quotients are relational algebra operations that aim at finding out the subrelational table $\mathcal{R} \div \mathcal{S}$ of a finite relational table $\mathcal{R}$, containing subtuples of $\mathcal{R}$ that have for complements in $\mathcal{R}$ *all* the tuples of a relational table $\mathcal{S}$. The quotient operation is defined by

Definition 1 *Relational quotient.*

$$\mathcal{R} \div \mathcal{S} = \{t, \forall s \in \mathcal{T}u(\mathcal{S}), (t, s) \in \mathcal{T}u(\mathcal{R})\}$$

where s denotes a tuple of $\mathcal{S}$ and t a subtuple of $\mathcal{R}$ such that (t, s) is a tuple of $\mathcal{R}$.

The definition of $R^{\Delta}(X)$ can be viewed as a particular case of such a division. Indeed $R^{\Delta}(X) = \{y \in Prop | R^{-1}(y) \supseteq X\} = \{y \in Prop | \forall x \in X, (x, y) \in R\}$. Besides, a relation $R \subseteq Prop \times Obj$ can be viewed as equivalent to a 2-attribute relational table $\mathcal{R}$ with $\mathbf{A} = \{Object - name, Property - name\}$, where (x, y) (resp. (y, x)) is a tuple in $\mathcal{R}$ (resp. $\mathcal{R}^{-1}$) if and only if $(x, y) \in R$. Then, it becomes clear that $R^{\Delta}(X) = \mathcal{R}^{-1} \div \mathcal{X}$, where $\mathcal{X}$ is the one-attribute relational table containing the object names. Similarly, $R^{-1\Delta}(Y) = \mathcal{R} \div \mathcal{Y}$ ($\mathcal{Y}$ is the one-attribute relational table associated to Y).

The fuzzy extensions of the basic operationsR^{Δ}, $R^{-1\Delta}$, which underly formal concept analysis, can thus be related to fuzzy division operations in fuzzy relational databases where tuples are weighted [14, 15]. As we shall see, this also provides a way for introducing fuzzily quantified conjunctions in order to require that tuples in R are associated with "at least k", or more generally '*most*' objects in X , rather than *all* elements in X as in formal concept analysis basic operations. In the following, we discuss these extensions in the setting of formal concept analysis.

First, the definition of $R^{\Delta}(X)$ extended to the case where R becomes fuzzy, as given in the previous section, namely $\mu_{R^{\Delta}(X)}(y) = min_{x \in X} \mu_{R^{-1}(y)}(x)$, is the exact counterpart to the fuzzy division $\mu_{\mathcal{R}^{-1} \div \mathcal{X}}(y) = \min_{x \in \mathcal{T}u(\mathcal{X})} \mu_R(x, y)$, for all pair (y, x) in $\mathcal{T}u(\mathcal{R}^{-1})$, when R is a fuzzy relation and X remains a classical set. The definition of a fuzzy division in fuzzy relation databases includes the more general case where X is also a fuzzy set. This can be done as well here, choosing an appropriate type of inclusion between fuzzy sets, i.e. an implication connective $\rightarrow$ in the expression:

$$\mu_{R^{\Delta}(X)}(y) = \min_{x \in Obj} \mu_X(x) \rightarrow \mu_{R^{-1}(y)}(x)$$

in relation with the intended meaning of having X fuzzy. For instance, taking Gödel implication ($a \rightarrow b = 1$ if $a \leq b$, and $a \rightarrow b = b$ if $a > b$) amounts to seeing $\mu_X(x)$'s as a significance threshold to which $\mu_R(x, y)$ is compared, while using Dienes implication ($a \rightarrow b = \max(1 - a, b)$) would be more in agreement with the idea of viewing $\mu_X(x)$ as a level of priority of x, just requiring the inclusion of important objects in $R^{-1}(y)$ (indeed the less important x, the greater $(1 - \mu_X(x))$, and the smaller the impact of x on the global evaluation, even when x totally fails to have property y (see [14, 15] for details).

Having both a fuzzy relation R and a fuzzy set X of objects may sound unrealistic in practice. In formal concept analysis, it is the starting point for natural weakening of the quantifier 'for all' into "at least k", or even into 'most'. The idea is to require that "at least k" (or more generally 'most') objects in X are the important objects that are the most in relation R with property y.

Let I be a fuzzy constraint on integers, defined by a membership function of the form: $\mu_I(0) = 1$and$\mu_I(i) \geq \mu_I(i+1)$. For instance, "at least k objects are important" is represented by $\mu_I(i) = 1$ if $0 \leq i \leq k$ and $\mu_I(i) = 0$ for $i \geq k + 1, n$ where $n = |X|$. Let us reorder the $\mu_{R^{-1}(y)}(x_k)$'s decreasingly, so that objects (x_k)'s that are more in relation R with y are the most important ones:

$$\mu_{R^{-1}(y)}(x_{\sigma(1)}) \geq \mu_{R^{-1}(y)}(x_{\sigma(2)}) \geq ... \geq \mu_{R^{-1}(y)}(x_{\sigma(n)}).$$

Then the extent to which property y is possessed by 'at least k' objects in X can be computed as

$$\mu_{R^{\Delta}(X),I}(y) = \min_i \ \max(\mu_{R^{-1}(y)}(x_{\sigma(i)}), 1 - \mu_I(i)),$$

I being defined as above. This expression, which involves an absolute fuzzy quantifier, may be easily modified in order to introduce relative quantifiers Q like 'most', having an increasing membership function in $[0, 1]$, by changing $1 - \mu_I(i + 1)$ into $\mu_Q(i/n)$ for $i = 0, n - 1$ and $\mu_Q(1) = 1$. It gives

$$\mu_{R^{\Delta}(X),Q}(y) = \min_i \ \max(\mu_{R^{-1}(y)}(x_{\sigma(i)}), \mu_Q(\frac{i-1}{n})).$$

Clearly, if Q means 'all', $\mu_Q(t) = 0$ for $t < 1$, then the above expression reduces to $\mu_{R^{\Delta}(X)}(y) = \min_{x \in X} \ \mu_{R^{-1}(y)}(x)$.

5 Incomplete and uncertain information

Until now, it has been assumed that we have *complete* information about the existing links between properties in $Prop$ and objects in Obj. Namely, $(x, y) \in R$ means that object x satisfies property y and $(x, y) \notin R$ means that object x does not satisfy property y, rather than "we do not know if $(x, y) \in R$ or not". Clearly, this assumption may be relaxed, while Boolean properties are still assumed: One may consider that there are pairs (x, y) for which it is not known at all if x has property y or not. This case has been considered in [16]. Information may be also uncertain, i.e., we are certain at level α that x has property y, or at level β that x has not property y.

In the most general case, properties are non Boolean (i.e. $\mu_R(x, y)$ is supposed to belong to $[0, 1]$), but the extent $\mu_R(x, y)$ to which an object x has a property y may be only fuzzily known under the form of a possibility distribution $\pi_{\mu_R}^{(x,y)}$ on $[0, 1]$ that restricts its possible values. Then one may not be even sure in general that some property y is possessed by an object x at least at some degree α). Since the information about $\mu_R(x, y)$ is now represented by a fuzzy set (on $[0, 1]$), and the four measures introduced in Section 3

ISBN: 978-989-95079-6-8

$\mu_{R^{\Pi}(X)}(y) = \Pi_y(X)$, $\mu_{R^{N}(X)}(y) = N_y(X)$, $\mu_{R^{\Delta}(X)}(y) = \Delta_y(X)$, $\mu_{R^{\nabla}(X)}(y) = \nabla_y(X)$ can themselves only be known under the form of induced possibility distributions (using the fuzzy set extension principle [2]). Let us take the example of $\mu_{R^{\Pi}(X)}(y) = \Pi_y(X) = \max_{x \in X} \mu_R(x, y)$. The induced possibility distribution is given by

$\pi_{\Pi_y(X)}(t) = \max_{i:\max_i t_i = t} \min_{x_i \in X} \pi_{\mu_R}^{(x_i,y)}(t_i)$ i.e.
$\pi_{\Pi_y(X)}(t) =$
$\max_i \min(\pi_{\mu_R}^{(x_i,y)}(t), \min_{j \neq i}(\max_{t_j \leq t} \pi_{\mu_R}^{(x_j,y)}(t_j))$.
Such a computation may be heavy in practice, but one may at least compute an upper bound of the possibility that a property y is associated to an object x with a degree equal to α, as $\pi^*_{\Pi_y(X)}(\alpha) = \max_{x \in X} \pi_{\mu_R}^{(x,y)}(\alpha)$. One may also compute the degree of membership of property y to the fuzzy set of properties that are possibly associated with at least one object in X at least to a degree ρ as $\Pi_{y,\rho}(X) = \max_{x \in X} \max_{\{t | t \geq \rho\}} \pi_{\mu_R}^{(x,y)}(t)$.

A maybe more promising approach for dealing with incomplete information in formal concept analysis is to first slightly modify the setting we start with by accommodating directly many-valued attributes instead of binary ones. We now outline this idea. Indeed when attribute domains are two-valued, they only give birth to a binary property (and its negation), while any non empty subset of a many-valued attribute domain (different from the domain itself) defines a non-trivial property. Take the example of the *color* attribute with domain $\{black, red, yellow, blue, green, ...\}$, $red\ or\ green$, or $red\ or\ yellow\ or\ blue$ are (imprecise) properties, beside the basic colors $black$, red, etc. Let Y now denote a set of attributes y, $dom(y)$ be the domain of y, P_y denote a non-empty subset of $dom(y)$. Let $\Gamma_y(x)$ represent the available information about the value of attribute y for object x. It is assumed that $\Gamma_y(x) \subseteq dom(y)$. Information is imprecise if $\Gamma_y(x)$ is not a singleton. For the moment, we suppose that information may be incomplete but not uncertain. $\Gamma_y(x) = \varnothing$ means that y does not apply to x, and $\Gamma_y(x) = dom(y)$ means that the value of y is unknown for x. $\Gamma_y(x) \neq \varnothing$ is now assumed.

Then $R^{-1N}(P_y) = \{x | \Gamma_y(x) \subseteq P_y\}$ is the set of objects that (certainly) have property P_y w. r. t. attribute y. Three other similar sets can be defined by reversing $\subseteq$, or replacing it by non-empty intersection, or non-covering union conditions, in the spirit of the basic definitions of Section 2. Let $R^{-1N\bigstar}$ be the relation that expresses that *certainly* objects have some (maybe imprecise) properties; it is defined on $X \times \cup_{y \in Y} \mathcal{P}(y)$, where $\mathcal{P}(y)$ denote the power set of $dom(y)$. Since $(\mathcal{P}(y), \subseteq)$ is a Boolean lattice, $(x, P_y) \in R^{-1N\bigstar}$ entails $(x, P'_y) \in R^{-1N\bigstar}$ as soon as $P_y \subseteq P'_y$ (if an object is red, it is also $red\ or\ green$). Clearly, $\{x | (x, red\ or\ green) \in R^{-1N\bigstar}\} = \{x | \Gamma_{color}(x) = red\} \cup \{x | \Gamma_{color}(x) = green\} \cup \{x | \Gamma_{color}(x) = red\ or\ green\}$. Moreover, we can also find out if there are only *possibly* $red\ or\ green$ objects, e.g. those that are known to be $red\ or\ blue$. If not, it means that there is no completion of the knowledge that can alter the extension of the set of $red\ or\ green$ objects in our example. More generally, we can look for concepts associated with sufficiently imprecise properties that remain stable under any knowledge completion. This can be extended to gradual uncertain knowledge by working with the α-cuts of the $\Gamma_y(x)$'s, i.e. pieces of information that are $(1-\alpha)$-certain. These are lines for further research.

6 Concluding remarks

Starting with a possibility-theoretic reading of concept analysis, we have reintroduced four operators, that enable us to describe all the different possible relations between a set of objects and a set of properties. Apart from retrieving the Galois connexion defining formal concepts, another Galois connexion based on the "actual necessity" operator is laid bare for decomposing the relation into independent blocks. The proposed setting extends to graded properties, leading to two kinds of Galois connected pair of fuzzy sets, whose meaning must be laid bare. Besides, the formal similarity between the actual possibility operator and relational algebra division operation, suggests a relaxation of the definition of concepts, computing the extent to which a property is *highly* possessed by *most* objects in a set. Lastly, extensions of the formal concept analysis setting to incomplete or uncertain information have been outlined. It is clear that many pending issues remain, such as e.g. the use of rough set reducts in this setting.

References

[1] B. Ganter and R. Wille. *Formal Concept Analysis, Mathematical Foundations*. Springer-Verlag, 1999.

[2] L.A. Zadeh. Fuzzy sets. *Inform. and Control*, 8, 338–353. 1965.

[3] R. Belohlavek. *Fuzzy Relational Systems: Foundations and Principles*, volume 20. Kluwer Academic/Plenum Press, 2002.

[4] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

[5] D.Dubois H.Prade Possibility theory: qualitative and quantitative aspects In *Quantified Representation of Uncertainty and Imprecision*, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, vol.1, 169–226. Kluwer Acad. Pu. 1998.

[6] D. Dubois, F. Dupin de Saint-Cyr and H. Prade. A possibility-theoretic view of formal concept analysis. *Fudamentae Informaticae*, 75: 195–213, 2007.

[7] I.Düntsch, E.Orlowska Mixing modal and sufficiency operators *Bull. Section of Logic, Polish Acad. of Sci.*, 28(2):99–106, 1999.

[8] Z. Pawlak. *Rough Sets. Theoretical Aspects of. Reasoning about Data*. Kluwer Acad. Publ., Dordrecht, 1991.

[9] Y.Y. Yao. A comparative study of formal concept analysis and rough set theory in data analysis. In *Rough Sets and Current Trends in Computing, 4th Int. Conf., RSCTC 2004*, Uppsala

[10] O.Ore Galois connexions *Tr. Ame. Mat. Soc.* 55:493-513, 1944

[11] A. Popescu. A general approach to fuzzy concepts. *Math. Log. Quart.*, 50(3):265–280, 2004.

[12] A. Burusco and R. Fuentes-Gonzalez. Construction of the l-fuzzy concept lattice. *Fuzzy Sets and Syst.*, 97,109–114, 1998.

[13] B. Bouchon-Meunier, D. Dubois, L. Godo, H. Prade. Fuzzy sets and possibility theory in approximate and plausible reasoning. In *Fuzzy Sets in Approximate Reasoning and Information Systems*, (J. C. Bezdek et al. eds.), Kluwer, 15-190, 1999.

[14] D. Dubois M. Nakata H. Prade Extended divisions for flexible queries in relational databases *Knowledge Management in Fuzzy Databases* O. Pons et al. eds., Physica-Verlag 105-21, 2000

[15] P. Bosc, O. Pivert, D. Rocacher Characterizing the result of the division of fuzzy relations *Int. J. Appr. Reas.*, 45, 511-30, 2007

[16] P. Burmeister and R. Holzer. Treating incomplete knowledge in formal concept analysis. *Formal Concept Analysis*, (B. Ganter et al. eds.), Springer Verlag, LNAI 3626, pp. 114-126, 2005.

ISBN: 978-989-95079-6-8

On the use of Measures of Separability of Classes to characterise the Domains of Competence of a Fuzzy Rule Based Classification System

Julián Luengo, Francisco Herrera[1]

1. Dept. of Computer Science and Artificial Intelligence,
CITIC-UGR (Research Center on Information and Communications Technology)
University of Granada, Granada, 18071, Spain
Email: {julianlm,herrera}@decsai.ugr.es

***Abstract*—**

In this work we study the behaviour of a Fuzzy Rule Based Classification System, and its relationship to a certain data complexity measures family. As Fuzzy Rule Based Classification System we have selected a recent proposal called Positive Definite Fuzzy Classifier, which is a Fuzzy System that uses Support Vector Machines for its training, obtaining accurate results and a low number of rules.

We have examined several data complexity metrics of separability of classes *over a wide range of data sets built from real data, and try to extract behaviour patterns from the results for this learning method. Using these data complexity measures and the accuracy results of the Positive Definite Fuzzy Classifier, we have built a rule set which describes both good or bad behaviours of this Fuzzy Rule Based Classification System.*

These rules use different values of such data complexity measures as antecedents, so we aim to predict the behaviour of the method from the data set complexity metrics prior to its application. Therefore, the rule set could characterise the domains of competence of this particular Fuzzy Rule Based Classification System.

***Keywords*—** Classification, Data complexity, Fuzzy Rule Based Systems, Support Vector Machines

1 Introduction

Fuzzy Rule Based Classification Systems (FRBCSs) [12, 14] are a very useful tool in the ambit of Data Mining, since they are capable of building a linguistic model clearly interpretable by human beings. There is a vast literature in the field of FRBCSs [14], which is very active at this time [9, 1, 16, 13].

The prediction capabilities of classifiers are strongly dependent on the problem's characteristics. An emergent field, that uses a set of complexity measures applied to the problem to describe its difficulty, has recently arisen. These measures quantify particular aspects of the problem which are considered complicated to the classification task [11]. Studies of data complexity metrics applied to particular classification's algorithms can be found in [11, 4, 3, 19].

In this work we are interested in analysing the relationship between FRBCSs and the complexity measures, considering a case of study using the Positive Definite Fuzzy Classifier (PDFC) proposed by Chen and Wang [5]. In particular we consider one type of data complexity measures based on the separability of classes.

To perform this study, we have created several binary classification data sets from real world problems, 438 ones, and computed the value of 3 metrics proposed by Ho and Basu [10]. We have analysed the intervals of the complexity measures values related to the created data sets, in which PDFC method performs well or badly, and then formulated a rule for such intervals. The rules try to describe the ranges where some information and conclusions about the behaviour of PDFC method can be stated.

This contribution is organised as follows. In Section 2 we describe the FRBCS we have used. In Section 3 the considered complexity measures are described. In Section 4 we include the experimental set-up, the results obtained and the rules extracted, along with their analysis. Finally, in Section 5 some concluding remarks are made.

2 Preliminaries: Fuzzy Rule Based Classification System

Any classification problem consists of l training patterns $x_p = (x_{p1}, \ldots, x_{pn})$, $p = 1, 2, \ldots, l$ from M classes where x_{pi} is the ith attribute value ($i = 1, 2, \ldots, n$) of the p-th training pattern.

Let us consider a two-class classification problem of assigning class label $y \in \{+1, -1\}$ to input feature vector x_p. In this work we consider the additive FRBCSs with constant `THEN`-parts.

As learning method we use the PDFC method [5], which uses a Support Vector Machine (SVM) approach to build up the model. In the following two Subsections, first we include the fuzzy inference model and a complete description of the PDFC algorithm.

2.1 Fuzzy Inference Model: PDFC Method

Consider a fuzzy model with m fuzzy rules of the form:

$$\text{Rule } j: \text{ If } A_j^1 \text{ AND } A_j^2 \text{ AND } \ldots \text{ AND } A_j^n \text{ THEN } b_j \qquad (1)$$

where A_j^k is a fuzzy set with membership function $a_j^k : \mathbb{R} \to [0, 1]$, $j = 1, \ldots, m$, $k = 1, \ldots, n$, $b_j \in \mathbb{R}$. If we choose product as the fuzzy conjunction operator, addition for fuzzy rule aggregation, and center of area defuzzification, then the model becomes a special form of the Takagi-Sugeno fuzzy model, and the input output mapping, $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$, of the model is defined as

$$\mathcal{F}(x_p) = \frac{\sum_{j=1}^{m} b_j \prod_{k=1}^{n} a_j^k(x_k)}{\sum_{j=1}^{m} \prod_{k=1}^{n} a_j^k(x_k)}. \qquad (2)$$

Equation (2) could not be well-defined on $\mathbb{R}$ if $\sum_{j=1}^{m} \prod_{k=1}^{n} a_j^k(x_k) = 0$ for some $x_p \in \mathbb{R}^n$, which could happen if the input space is not wholly covered by fuzzy rule "patches". To fix this problem, we add a fuzzy rule so that the

ISBN: 978-989-95079-6-8

denominator $\sum_{j=1}^{m}\prod_{k=1}^{n}a_j^k(x_k) > 0$ for all $x_p \in \mathbb{R}$. Thus the following rule is added:

$$\text{Rule } 0: \text{ If } A_j^1 \text{ AND } A_j^2 \text{ AND } \ldots \text{ AND } A_j^n \text{ THEN } b_0 \quad (3)$$

where $b_0 \in \mathbb{R}$, the membership functions $a_0^k(x_k) \equiv 1$ for $k = 1, \ldots, n$ and any $x_p \in \mathbb{R}^n$. Consequently, the input output mapping becomes

$$\mathcal{F}(x_p) = \frac{b_0 + \sum_{j=1}^{m} b_j \prod_{k=1}^{n} a_j^k(x_k)}{1 + \sum_{j=1}^{m}\prod_{k=1}^{n} a_j^k(x_k)}. \quad (4)$$

A classifier associates class labels with input features, i.e., it is essentially a mapping from the input space to the set of class labels. In binary case, thresholding is one of the simplest ways to transform $\mathcal{F}(x_p)$ to class labels $+1$ or -1.

Considering the FRBCS with $m+1$ fuzzy rules where Rule 0 is given by (3), then the system induces a binary fuzzy classifier, f , with decision rule

$$f(x_p) = \text{sign}(\mathcal{F}(x_p) + t) \quad (5)$$

where $t \in \mathbb{R}$ is a threshold. We can assume $t = 0$ without loss of generality.

The membership functions for a binary fuzzy classifier defined above could be any function from $\mathbb{R}$ to $[0, 1]$. We narrow our interests to a class of membership functions, $a_j^k : \mathbb{R} \to [0, 1]$, $j = 1, \ldots, m$, which are generated from a reference function a^k through location transformation [8], and the classifiers defined on them. In [5] well-known types of reference functions can be found, like the symmetric triangle and the gaussian function.

As consequence of the presented formulation, the decision rule of our binary fuzzy classifier can be written as:

$$f(x_p) = \text{sign}\left(\sum_{j=1}^{m} b_j K(x_p, z_j) + b_0\right) \quad (6)$$

where $z_j = [z_j^1, z_j^2, \ldots, z_j^n]^T \in \mathbb{R}$ contains the location parameters of a_j^k. $K : \mathbb{R}^n \times \mathbb{R}^n \to [0, 1]$ is a translation invariant kernel defined as

$$K(x_p, z_j) = \prod_{k=1}^{n} a^k(x_p^k - z_j^k) \quad (7)$$

which is actually a Mercer Kernel [6], if it has nonnegative Fourier transform. Again, from [5] some Mercer kernels can be built using the reference functions mentioned above.

Thus, the decision rule of a binary fuzzy classifier is

$$f(x_p) = \text{sign}\left(b_0 + \sum_{j=1}^{m} b_j \prod_{k=1}^{n} a_j^k(x_p^k)\right) \quad (8)$$

2.2 Learning method: SVM approach to build PDFC

Here, we assume that the reference functions are predetermined. So the remaining question is how to find a set of fuzzy rules ($\{z_1, \ldots, z_m\}$ and $\{b_0, \ldots, b_m\}$) from the given training so that the PDFC method has good generalization.

As given in (7), for a PDFC, a Mercer kernel can be constructed from the positive definite reference functions. The kernel implicitly defines a nonlinear mapping Φ that maps $\mathbb{X}$ into a kernel-induced feature space $\mathbb{F}$. Theorem 3.12 in [5] states that the decision rule of a PDFC can be viewed as a hyperplane in $\mathbb{F}$. It is well-known that the SVM algorithm finds a separating hyperplane with good generalization by reducing the empirical risk and, at the same time, controlling the hyperplane margin [21]. Thus we can use the SVM algorithm to find an optimal hyperplane in $\mathbb{F}$. Once we get such a hyperplane, fuzzy rules can be easily extracted. The whole procedure is described by the following algorithm 1.

Algorithm 1 SVM learning for PDFC

INPUTS: Positive definite reference functions $a^k(x_p)$ associated with n input variables, and a set of training samples
OUTPUTS: A set of fuzzy rules parameterized by z_j, b_j, and m. z_j contains the location parameters of the `IF`-part membership functions of the jth fuzzy rule, b_j is the `THEN`-part constant of the jth fuzzy rule, and $m+1$ is the number of fuzzy rules.
Steps:
1 Construct a Mercer kernel, K, from the given positivedefinite reference functions according to (7).
2 Construct an SVM to get a decision rule of the form

$$f(x) = sign\left(\sum_{i \in S} y_i \alpha_i K(x, x_i) + b\right),$$

being S the index set of the support vectors:
2.1 Assign some positive number to the cost C, and solve the quadratic program defined by the proper SVM to get the Lagrange multipliers α_i.
2.2 Find b (details can be found in, for example, [18]).
2.3 Extracting fuzzy rules from the decision rule of the SVM:
$b0 \leftarrow b$
$j \leftarrow 1$
for $i = 1$ **to** l **do**
  if $\alpha_i > 0$ **then**
    $z_j \leftarrow x_i$
    $b_j \leftarrow y_i \alpha_i$
    $j \leftarrow j + 1$
  end if
end for
$m \leftarrow j - 1$

In our study, we have considered the parameters values recommended by the authors. They are summarized as follows:

- $C = 100$ (weight of the classification error)
- $d = 0.25$ (parameter used by the reference functions)
- Type of reference functions: Gaussian

3 Data Complexity Measures Based on the Separability of Classes

In this section we describe the three metrics we have used in this contribution, with their correspondent acronym.

For our study, we will examine three measures of separability of classes from [10] which offer information for the PDFC method. They are described next.

- **N1**: fraction of points on class boundary. This method constructs a class-blind minimum spanning tree over the entire data set, and counts the number of points incident to an edge going across the two classes. The fraction of

ISBN: 978-989-95079-6-8

such points over all points in the data set is used as a measure. For two heavily interleaved classes, a majority of points are located next to the class boundary. However, the same can be true for a sparsely sampled linearly separable problem with margins narrower than the distances between points of the same class.

- **N2**: ratio of average intra/inter class Nearest Neighbour (NN) distance. For each input instance x_p, we calculate the distance to its nearest neighbour within the class (intraDist(x_p)) and the distance to nearest neighbour of any other class (interDist(x_p)). Then, the result is the ratio of the sum of the intra-class distances to the sum of the inter-class distances for each input example, i.e.,

$$N2 = \frac{\sum_{i=0}^{m} intraDist(x_i)}{\sum_{i=0}^{m} interDist(x_i)},$$

where m is the number of examples in the data set. This metric compares the within-class spread with the distances to the nearest neighbours of other classes. Low values of this metric suggest that the examples of the same class lay closely in the feature space. Large values indicate that the examples of the same class are disperse. It is sensitive to the classes of the closest neighbours to a point, and also to the difference in magnitude of the between-class distances and that of the within-class distances.

- **N3**: error rate of 1-NN classifier. This is simply the error rate of a nearest-neighbour classifier measured with the training set. The error rate is estimated by the leave-one-out method. The measure denotes how close the examples of different classes are. Low values of this metric indicate that there is a large gap in the class boundary.

4 Experimental Study: Analysis of the PDFC with Data Complexity Measures

In this Section we analyse the obtained results for the PDFC method. First, in Subsection 4.1 we present the experimental framework, with the data sets generation method, accuracy validation scheme, and the global average results of the PDFC method. Next we determine several rules based on PDFC's behaviour in Subsection 4.2. Finally we analyse the collective evaluation of the set of rules in Subsection 4.3.

4.1 Experimental Framework: Data Sets Generation

We evaluate the PDFC method on a set of 438 binary classification problems. These problems are generated from pairwise combinations of the classes of 20 problems from the University of California, Irvine (UCI) repository [2]. These are *iris, wine, new-thyroid, solar-flare, led7digit, zoo, yeast, tae, balanced, car, contraceptive, ecoli, hayes-roth, shuttle, australian, pima, monks, bupa, glass, haberman* and *vehicle*.

In order to do that, first we take each data set and extract the examples belonging to each class. Then we construct a new data set with the combination of the examples from two different classes. This will result in a new data set with only 2 classes and the examples which have two such classes as output. We perform this process for every possible pairwise combination of classes. However, if an obtained data set with this procedure proves to be linearly-separable, we discard it (since we could classify it with a linear classifier with no error). The complexity measure L1 from [10] indicates if a problem is linearly-separable if its value is zero, so every data set with a L1 value of zero will be discarded.

This method for generating binary data sets is limited by the proper combinatorics, and we can only obtain over 200 new data sets with the original 20 data sets with this first approach. In order to obtain more data sets, we group the classes two by two, that is, we create a new binary data set, and each of its two classes are the combination of two original classes each. For this second approach we have used *ecoli*, *glass* and *flare* data sets, since they have a high number of class labels. Again, those data sets with a L1 value of zero are discarded.

In order to measure the PDFC performance, we have applied a 10-fcv validation scheme. In Table 1 we show the global Training and Test accuracy obtained by the PDFC method.

Table 1: Global Average PDFC Training and Test Accuracy

PDFC Global % Accuracy Training	94.06%
PDFC Global % Accuracy Test	91.22%

4.2 Determination of Rules Based on the PDFC Behaviour

In the following we present the results of the execution over the 438 data sets summarized in Figures 1 to 3.

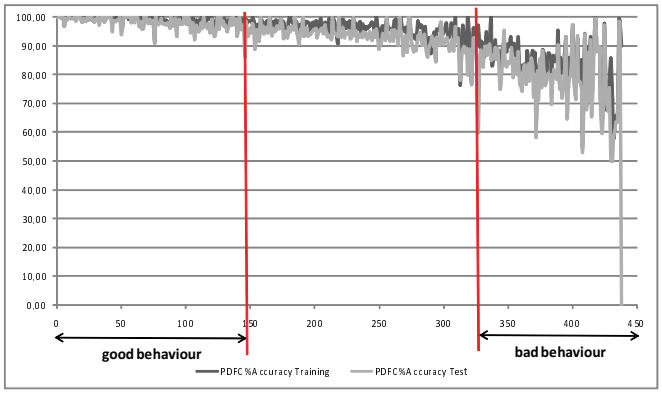


Figure 1: PDFC accuracy in Training/Test sorted by N1

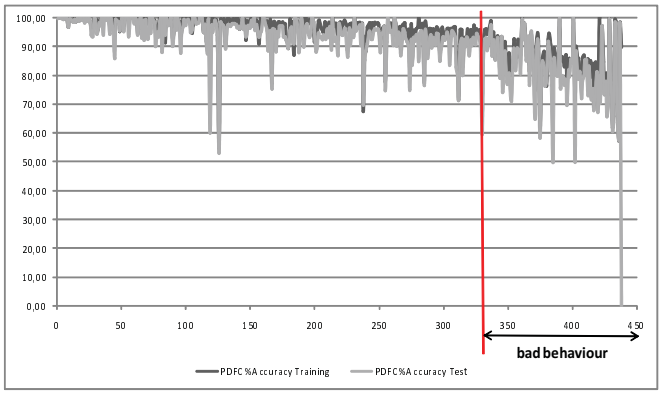


Figure 2: PDFC accuracy in Training/Test sorted by N2

For each complexity measure (N1, N2 and N3), the data sets are sorted by the ascending value of the corresponding complexity measure, and put altogether in a Figure. In the X

ISBN: 978-989-95079-6-8

Table 3: Rules with one metric obtained from the intervals

Id.	Rule	Support	%Training	Training Diff.	% Test	Test Diff.
R1+	If N1[X] $<$ 0.089 then *good behaviour*	33.11%	99.09%	4.91%	98.15%	6.55%
R2+	If 0 $<$ N3[X] $<$ 0.047 then *good behaviour*	28.31%	98.92%	4.86%	97.70%	6.48%
R1-	If N1[X] $\geq$ 0.25 then *bad behaviour*	25.57%	84.64%	-9.42%	79.01%	-12.21%
R2-	If N2[X] $>$ 0.5196 then *bad behaviour*	24.89%	86.14%	-7.92%	80.88%	-10.34%
R3-	If N3[X] $>$ 0.175 then *bad behaviour*	19.41%	82.64%	-11.42%	76.77%	-14.45%

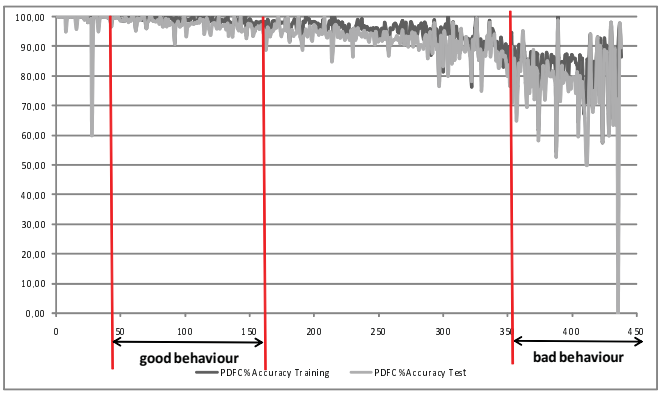


Figure 3: PDFC accuracy in Training/Test sorted by N3

axis we represent the data sets, not the complexity measure value, and the Y axis depicts the accuracy obtained both in training and test. The reason to do so is to give each data set the same space in the graphic representation. For those measures where we can find different *ad-hoc* intervals which present *good* or *bad behaviour* of the PDFC, we use a vertical line to delimit the interval of the region of interest.

- We understand for *good behaviour* an average high test accuracy in the interval, as well as the absence of over-fitting.
- By *bad behaviour* we refer to the presence of over-fitting and/or average low test accuracy in the interval.

In Table 2 we have summarized the intervals found ad-hoc from Figures 1 to 3.

Table 2: Significant intervals

Interval	PDFC Behaviour
$N1 < 0.089$	*good behaviour*
$0 < N3 < 0.047$	*good behaviour*
$N1 \geq 0.25$	*bad behaviour*
$N2 > 0.5196$	*bad behaviour*
$N3 > 0.175$	*bad behaviour*

From these ad-hoc intervals we construct several rules that model the performance of the FRBCS we have used. In Table 3 we have summarized the rules derived from Table 2. Given a particular data set X, we get the complexity measure of X with the notation $CM[X]$. Table 3 is organised with the following columns.

- The first column corresponds to the identifier of the rule for further references.
- The "Rule'" column presents the rule itself.
- The third column "Support" presents the percentage of data sets which verifies the antecedent of the rule.
- The column "% Training" shows the average accuracy in training of all the examples which are covered by the rule.
- The column "Training Diff." contains the difference between the training accuracy of the rule and the training accuracy across all 438 data sets.
- The column "% Test" shows the average accuracy in test of all the examples which are covered by the rule.
- The column "Test Diff." contains the difference between the test accuracy of the rule and the test accuracy across all 438 data sets.

As we can see in Table 3, the positive rules (denoted with a "+" symbol in their identifier) always show a positive difference with the global average, both in training and test accuracy. The negative ones (with a "-" symbol in their identifier) verify the opposite case. The support of the rules shows us that we can characterize a wide range of data sets and obtain significant differences in accuracy.

From this set of rules we can state that a low N1 value results in a good behaviour of the PDFC method. A low N3 value obtains the same results. In the other hand, a high value in the N1 metric produces a bad behaviour of the PDFC considered in our analysis. A high N3 value will also produce a bad behaviour of the PDFC method. With similar outcome, if N2 presents a high value, the PDFC method will obtain bad behaviour.

Although we have obtained some interesting rules, we can extend our study by considering the combination of these complexity metrics in order to obtain more precise and descriptive rules.

4.3 Collective Evaluation of the Set of Rules

The objective of this section is to analyse the good rules jointly, and the bad rules together as well. Thus we can arrive at a more general description, with wider support, of the behaviour of the PDFC with these joint rules. We perform the disjunctive combination of all the positive rules to obtain a single rule, and all the negative ones, so we obtain another rule.

ISBN: 978-989-95079-6-8

Table 4: Disjunction Rules from all simple rules

Id.	Rule	Support	%Training	Training Diff.	% Test	Test Diff.
PRD	If R1+ or R2+ then *good behaviour*	37.44%	98.99%	4.93%	97.90%	6.68%
NRD	If R1- or R2- or R3- then *bad behaviour*	31.51%	86.23%	-7.83%	81.24%	-9.98%
not characterised	If not PRD and not NRD then *good behaviour*	31.05%	96.06%	2.00%	93.27%	2.05%

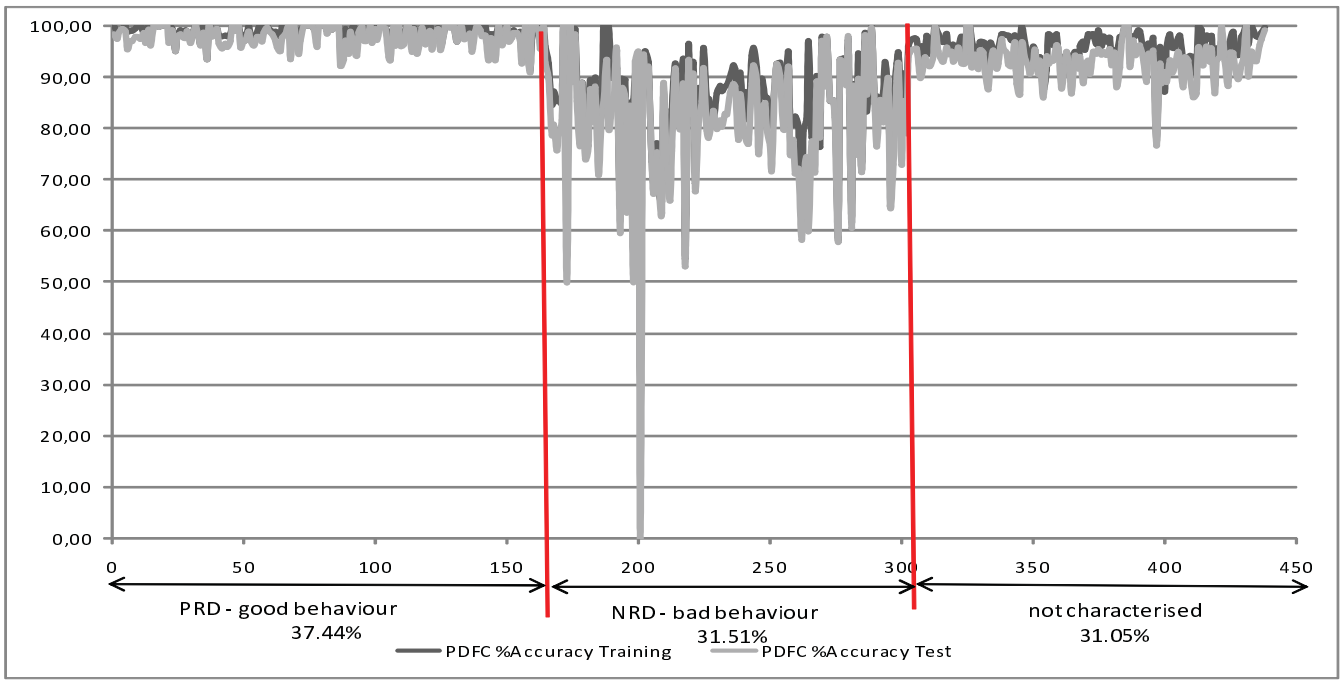


Figure 4: Three blocks representation for PRD, NRD and not covered data sets for PDFC

The new disjunctive rule will be activated if any of the component rules' antecedents are verified. In Table 4 we summarize both disjunctions, and a third rule representing those data sets which are not charaterised by either disjunction rules.

From the collective rules we can observe that the support has been increased from the single rules both for the Positive Rule Disjunction (PRD) and Negative Rule Disjunction (NRD). In the other hand, the Test and Training Accuracy Differences are similar to the single rules from Table 3. Since there are no data sets in PRD and NRD simultaneously, we can consider three blocks of data sets with their respective support, as depicted in Figure 4 (with no particular data set order within each block):

- The first block (the left-side one) represents the data sets covered by the PRD rule. They are the data sets recognized as being those in which the PDFC has good accuracy.
- The second block (the middle one) plots the data sets for the rule NRD, which are bad data sets for the PDFC method considered.
- The third and last block (the right-side one) contains the unclassified data sets by the previous two rules.

We can see that almost the 70% of the analysed data sets are covered by these two rules, and hence the *good behaviour* and *bad behaviour* consequents represent well the accuracy of PDFC methods.

5 Concluding Remarks

We have performed a study over a set of binary data sets with the PDFC method. We have computed some data complexity measures for the data sets in order to obtain intervals of such metrics in which the method's performance is significantly good or bad. We have constructed descriptive rules, and studied the interaction between the intervals and the proper rules.

We have obtained two rules which are simple and precise to describe both good and bad performance of the PDFC. Furthermore, we present the possibility of determining which data sets PDFC would performs well or badly prior to their execution, using the Data Complexity measures.

We must point out that this is a particular study for one specific method, the PDFC. On the other hand, this work presents a new challenge that could be extended to other FRBCSs, to analyse their domains of competence, and to develop new measures which could give more information on the behaviours of FRBCSs for pattern recognition.

Acknowledgment

This work has been supported by the Spanish Ministry of Science and Technology under Project TIN2008-06681-C06-01.

ISBN: 978-989-95079-6-8

J. Luengo holds a FPU scholarship from Spanish Ministry of Innovation and Science.

References

[1] P. Angelov, E. Lughofer and X. Zhou. Evolving fuzzy classifiers using different model architectures. *Fuzzy Sets and Systems*, 159(23):3160–3182, 2008.

[2] A. Asuncion and D.J. Newman. UCI Machine Learning Repository [http://www.ics.uci.edu/ mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, 2007.

[3] R. Baumgartner, R.L. Somorjai. Data complexity assessment in undersampled classification. *Pattern Recognition Letters*, 27:1383–1389, 2006.

[4] E. Bernadó-Mansilla, T.K. Ho. Domain of Competence of XCS Classifier System in Complexity Measurement Space. *IEEE Transactions on Evolutionary Computation*, 9(1):82–104, 2005.

[5] Y. Chen and J.Z. Wang. Support Vector Learning for Fuzzy Rule-Based Classification Systems. *IEEE Transactions on Fuzzy Systems*, 11(6):716–728,2003.

[6] N. Cristianini and J. Shawe-Taylor. *An introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[7] M. Dong, R. Kothari. Feature subset selection using a new definition of classificabilty. *Pattern Recognition Letters*, 24:1215–1225, 2004.

[8] D. Dubois and H. Prade. Operations on fuzzy numbers. *Int. J. Syst. Sci.*, 9(6):613–626, 1978.

[9] A. Fernández, S. García, M.J. del Jesus and F. Herrera. A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets and Systems*, 159(18):2378–2398, 2008.

[10] T.K. Ho, M. Basu. Complexity Measures of Supervised Classification Problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):289–300, 2002.

[11] M. Basu, T.K. Ho. *Data Complexity in Pattern Recognition*. Springer, 2006.

[12] H. Ishibuchi, T. Nakashima, M. Nii. *Classification and modeling with linguistic information granules: advanced approaches to linguistic data mining*. Springer, Berlin, 2004.

[13] M.Z. Jahromi and M. Taheri. A proposed method for learning rule weights in fuzzy rule-based classification systems. *Fuzzy Sets and Systems*, 159(4):449-459, 2008

[14] L. Kuncheva L. *Fuzzy classifier design*. Springer, Berlin, 2000.

[15] Y. Li, M. Dong, R. Kothari. Classifiability-Based Omnivariate Decision Trees. *IEEE Transactions on Neural Networks*, 16(6):1547–1560, 2005.

[16] E.G. Mansoori, M.J. Zolghadri and S.D. Katebi. SGERD: A steady-state genetic algorithm for extracting fuzzy classification rules from data. *IEEE Transactions on Fuzzy Systems*, 16(4):1061–1071, 2008.

[17] R.A. Mollineda, J.S. Sánchez, J.M. Sotoca. Data Characterization for Effective Prototype Selection. *First edition of the Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2005). Lecture Notes in Computer Science 3523, Springer-Verlag*, 27–34, 2005.

[18] J. Platt. Machines using Sequential Minimal Optimization. In *B. Schoelkopf, C. Burges and A. Smola (Eds.), Advances in Kernel Methods - Support Vector Learning*, 1998.

[19] J.S. Sánchez, R.A. Mollineda, J.M. Sotoca. An analysis of how training data complexity affects the nearest neighbor classifiers. *Pattern Analysis and Applications*, 10(3):189–201, 2007.

[20] S. Singh. Multiresolution Estimates of Classification Complexity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1534–1539, 2003.

[21] V. Vapnik. *Statistical Learning Theory*. New York: Wiley, 1998.

ISBN: 978-989-95079-6-8

Learning Concurrently Granularity, Membership Function Parameters and Rules of Mamdani Fuzzy Rule-based Systems

Michela Antonelli, Pietro Ducange, Beatrice Lazzerini, Francesco Marcelloni

Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni
University of Pisa
Via Diotisalvi 2, 56122 Pisa
Email: {michela.antonelli, p.ducange, b.lazzerini, f.marcelloni}@iet.unipi.it

Abstract—*In this paper we tackle the issue of generating Mamdani fuzzy rule-based systems with optimal trade-offs between complexity and accuracy by using a multi-objective genetic algorithm, which concurrently learns rule base, granularity of the input and output partitions and membership function parameters. To this aim, we exploit a chromosome composed of three parts, which codify, respectively, the rule base, and, for each variable, the number of fuzzy sets and the parameters of a piecewise linear transformation of the membership functions. We show the encouraging results obtained on a real world regression problem.*

Keywords— Accuracy-Interpretability Trade-off, Granularity Learning, Mamdani Fuzzy-Rule-Based Systems, Multi-objective Evolutionary Algorithms, Piecewise Linear Transformation.

1 Introduction

In the last years, the problem of finding the right trade-off between interpretability and accuracy of Mamdani fuzzy rule-based systems (MFRBSs) [1] has arisen a growing interest in the fuzzy community [2]. To this aim, Multi-Objective Evolutionary Algorithms (MOEAs) have been extensively used for tuning or learning the data base (DB) and the rule base (RB) of the MFRBSs [3][4]. All these approaches generate approximated Pareto fronts with non-dominated solutions in the interpretability-accuracy space. For example, in [5], authors use a predefined DB and evolve only the RB, while in [6] the membership function (MF) parameters are adapted to a specific context exploiting a predefined RB. Furthermore, in [7] and [8] the tuning of the DB is performed together with a rule selection.

The ideal approach would be to learn concurrently DB and RB. So far, only approximations of this ideal approach have been proposed such as to learn simultaneously the overall RB and, for the DB, only the MF parameters [9][10] or the granularities of the uniform partitions defined on the input and output variables [11][12].

On the other hand, both granularity and MF parameters are a critical factor in MFRBS generation [6][13], since both affect accuracy and interpretability. Indeed, the former fixes an upper bound to the number of rules in the MFRBS and the latter adapts the meaning of the linguistic values to the specific application context.

In this paper, we propose a multi-objective evolutionary approach to generate MFRBSs with different trade-offs between complexity and accuracy. The main novelty of our approach is that the RB, the granularity of the input and output partitions and the MF parameters of each fuzzy set are learnt concurrently during the evolutionary process. To deal with RBs defined on different granularities, we exploit the concept of virtual RBs introduced in [12]. Further, the learning of the MF parameters is performed by using a piecewise linear transformation [14][15], which allows us to obtain a high modelling capability with a limited number of parameters.

Chromosomes are composed of three parts, which codify the RB, and, for each linguistic variable, the number of fuzzy sets and the parameters of the piecewise linear transformation, respectively. In the evolutionary process, we adopt the mating operators proposed in [5] for the fist part of the chromosome, and standard crossover and mutation operators for the second and third parts, respectively.

Our approach has been tested on a real world regression problem, with nine input variables, and has provided Pareto fronts with solutions characterized by better trade-offs between accuracy and complexity than solutions belonging to Pareto fronts generated by learning only the rule base with a fixed DB, and concurrently the rule base and the MF parameters with fixed granularities, respectively.

2 Mamdani Fuzzy Systems

Let $\mathbf{X} = \{X_1, ..., X_f, ..., X_F\}$ be the set of input variables and X_{F+1} be the output variable. Let $U_f (f = 1, ..., F+1)$ be the universe of the f^{th} variable. Let $P_f = \{A_{f,1}, ..., A_{f,T_f}\}$ be a fuzzy partition of T_f fuzzy sets on variable X_f. An MFRBS is composed of *M* rules expressed as:

$$R_m: \text{IF } X_1 \textbf{ is } A_{1,j_{m,1}} \text{ AND } \ldots \text{ AND } X_F \textbf{ is } A_{F,j_{m,F}} \text{ THEN } X_{F+1} \textbf{ is } A_{F+1,j_{m,F+1}} \quad (m = 1, ..., M) \qquad (1)$$

where $j_{m,f} \in [1, T_f]$ identifies the index of the fuzzy set (among the T_f fuzzy sets of partition P_f), which has been selected for X_f in rule R_m.

To take the "don't care" condition into account [16] a new fuzzy set $A_{f,0}$ $(f = 1, ..., F)$ is added to all the F input partitions P_f. This fuzzy set is characterized by a membership function equal to 1 on the overall universe.

The terms $A_{f,0}$ allow generating rules which contain only a

ISBN: 978-989-95079-6-8

subset of the input variables. It follows that $j_{m,f} \in [0, T_f]$, $f = 1,..., F$, and $j_{m,F+1} \in [1, T_{F+1}]$.

An MFRBS can be completely described by the following matrix $J \in \mathbb{N}^{M \times (F+1)}$

$$J = \begin{bmatrix} j_{1,1} & \cdots & j_{1,F} & j_{1,F+1} \\ \cdots & \cdots & \cdots & \cdots \\ j_{m,1} & \cdots & j_{m,F} & j_{m,F+1} \\ \cdots & \cdots & \cdots & \cdots \\ j_{M,1} & \cdots & j_{M,F} & j_{M,F+1} \end{bmatrix}$$

where the generic element (m, f) indicates that fuzzy set $A_{f,j_{m,f}}$ has been selected for variable X_f in rule R_m.

We adopt the product and the weighted average method as AND logical operator and defuzzification method, respectively.

Given a set of *N* input observations $\mathbf{x}_n = [x_{n,1},..., x_{n,F}]$, with $x_{n,f} \in \Re$, and the set of the corresponding outputs $x_{n,F+1} \in \Re$, $n = 1,..., N$, we apply a multi-objective evolutionary algorithm which produces a set of MFRBSs with different trade-offs between accuracy and complexity by learning simultaneously the RB, the granularity of the partitions of each variable and the MF parameters. The choice of the appropriate trade-off between accuracy and complexity depends on the particular application.

3 Granularity and MF Parameter Learning

3.1 Granularity Learning

To determine simultaneously the granularity of partitions and the RB in the evolutionary process is not an easy task since the RB depends on the number of fuzzy sets used to partition the variables. In [12] we have introduced the concept of *virtual RB*. A virtual RB is composed of rules defined by considering the variables partitioned with a user-defined maximum number $T_{\max}$ of fuzzy sets (*virtual partition*). All the mating operators are applied to virtual RBs: the actual granularity is used only in the computation of the fitness. In practice, we generate virtual RBs and assess their quality using each time different "lens" depending on the actual number of fuzzy sets used to partition the single variables. Thus, we do not worry about the actual granularity in applying crossover and mutation operators.

To map the virtual RB defined on variables uniformly partitioned with $T_{\max}$ fuzzy sets into a concrete RB defined on variables uniformly partitioned with T_f fuzzy sets, we adopt a mapping strategy. Let X_f **is** $\hat{A}_{f,h}$, $h \in [0, T_{\max}]$, be a generic fuzzy proposition defined in a rule of the virtual RB. Then, the proposition will be mapped to X_f **is** $\tilde{A}_{f,s}$, with $s \in [0, T_f]$, where $\tilde{A}_{f,s}$ is the fuzzy set more similar to $\hat{A}_{f,h}$ among the T_f fuzzy sets $\hat{A}_{f,j}$ defined on X_f. For the sake of simplicity, we have trivially considered as similarity measure the distance between the centroids of the two fuzzy sets. If there are two fuzzy sets in $\tilde{P}_f = \{\tilde{A}_{f,1},..., \tilde{A}_{f,T_f}\}$ with centroids at the same distance from the centroid of $\hat{A}_{f,h}$, we choose randomly one of the two fuzzy sets.

Note that different rules of the virtual RB can be mapped to equal rules in the concrete RB. This occurs because distinct fuzzy sets defined on the partitions used in the virtual RB can be mapped to the same fuzzy set defined on the partitions used in the concrete RB. In the case of equal rules, only one of these rules is considered in the concrete RB. The original different rules are, however, maintained in the virtual RB. Indeed, when the virtual RB will be interpreted by using different "lens", all these rules can again be meaningful and contribute to increase the accuracy of the MFRBS. Thus, the concept of virtual RB allows us to explore the search space and concurrently exploiting the optimal solutions achieved during the evolutionary process.

3.2 MF Parameter Learning

We approach the problem of learning the MF parameters by using a piecewise linear transformation [14][15]. The transformation is described in Fig. 1 for a generic variable X_f. In the following, we assume that the interval ranges of the original and transformed variables are identical. Further, we consider triangular fuzzy sets $A_{f,j}$ defined by the tuple $(a_{f,j}, b_{f,j}, c_{f,j})$, where $a_{f,j}$ and $c_{f,j}$ correspond to the left and right extremes of the support of $A_{f,j}$, and $b_{f,j}$ to the core. Finally, given a generic partition $P_f = \{A_{f,1},..., A_{f,T_f}\}$, we assume that, for $j = 2 \ldots T_f - 1$, $b_{f,j} = c_{f,j-1}$ and $b_{f,j} = a_{f,j+1}$, and $a_{f,1} = b_{f,1}$ and $b_{f,T_f} = c_{f,T_f}$.

Before passing the input value x_f to the MFRBS, we apply the transformation $t(x_f)$. Thus, we have that:

$$A_{f,j}(x_f) = \tilde{A}_{f,j}(t(x_f)) = \tilde{A}_{f,j}(\tilde{x}_f)$$

where $\tilde{A}_{f,j}$ and $A_{f,j}$ are two generic fuzzy sets from the uniform and non-uniform fuzzy partitions, respectively. In those regions where *t* has a high value of the derivative (high slope of the lines), the fuzzy sets $A_{f,j}$ are narrower; otherwise, the fuzzy sets $A_{f,j}$ are wider.

As observed in [14], the transformation *t* must be non-decreasing. Further, the fuzzy sets $A_{f,j}$ of the non-uniform fuzzy partition induced by the transformation *t* are not necessarily of triangular shape. This could bring to fuzzy partitions difficultly interpretable. To preserve the shape of the MFs, we force the change of slopes in *t* to coincide with the cores of the fuzzy sets in the partitions.

Let $b_{f,1},..., b_{f,T_f}$ and $\tilde{b}_{f,1},..., \tilde{b}_{f,T_f}$ be the cores of $A_{f,1},..., A_{f,T_f}$ and $\tilde{A}_{f,1},..., \tilde{A}_{f,T_f}$, respectively. Transformation *t* can be defined as:

$$t(x_f) = \frac{\tilde{b}_{f,j} - \tilde{b}_{f,j-1}}{b_{f,j} - b_{f,j-1}}(x_f - b_{f,j-1}) + \tilde{b}_{f,j-1}, \quad b_{f,j-1} \le x_f < b_{f,j},$$

with *j*=2…T_f.

Once fixed the granularity of the partitions, $\tilde{b}_{f,1},..., \tilde{b}_{f,T_f}$ are fixed and therefore known. Further, $b_{f,1}$ and b_{f,T_f} coincide

ISBN: 978-989-95079-6-8

with the extremes of the universe U_f of X_f. Thus, $t(x_f)$ depends on T_f –2 parameters, that is, $t(x_f; b_{f,2},...,b_{f,T_f-1})$. Once fixed $b_{f,2},...,b_{f,T_f-1}$, the partition $P_f = \{A_{f,1},...,A_{f,T_f}\}$ can be obtained simply by transforming the three points $(\tilde{a}_{f,j}, \tilde{b}_{f,j}, \tilde{c}_{f,j})$, which describe the generic fuzzy set $\tilde{A}_{f,j}$, into $(a_{f,j}, b_{f,j}, c_{f,j})$ applying $t^{-1}(\tilde{x}_f)$.

We define the piecewise linear transformation on the maximum granularity. When we reduce the granularity, to maintain the original shape of the MFs, we do not apply the piecewise linear transformation to all the points of the universe, but only to the three points, which define the generic triangular MF. In practice, we transform $(\tilde{a}_{f,j}, \tilde{b}_{f,j}, \tilde{c}_{f,j})$ into $(a_{f,j}, b_{f,j}, c_{f,j})$ by applying $t^{-1}(\tilde{x}_f)$. Fig. 2 shows an example of this transformation for granularity 5 by using the piecewise linear transformation in Fig. 1, defined with granularity 7.

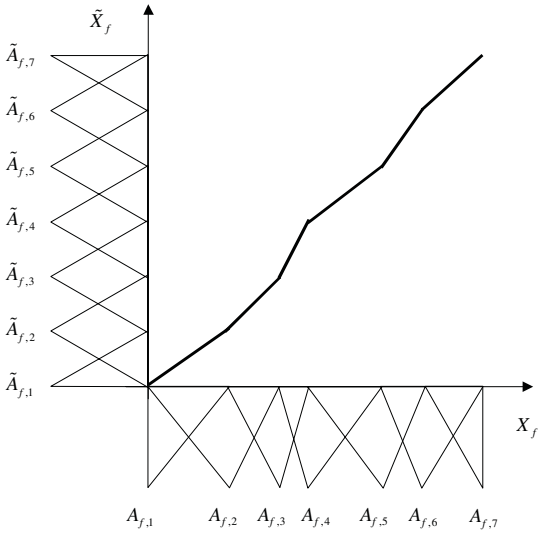


Figure 1: An example of piecewise linear transformation.

4 The Multi-Objective Evolutionary Approach

4.1 Chromosome coding

Each solution is codified by a chromosome C composed of three parts (C_1, C_2, C_3), which define the virtual RB, and the granularities and the piecewise linear transformations of all the variables, respectively. In particular, C_1 encodes the virtual RB by considering that each variable X_f is uniformly partitioned by using $T_{\max}$ fuzzy sets, that is, $P_f = \{A_{f,1},...,A_{f,T_{\max}}\}$. As described in [5], C_1 is composed of $M \cdot (F+1)$ natural numbers where M is the number of rules currently present in the virtual RB. The RB (defined as *concrete RB*) used to compute the fitness is obtained by means of the mapping strategy using the actual granularities fixed by C_2.

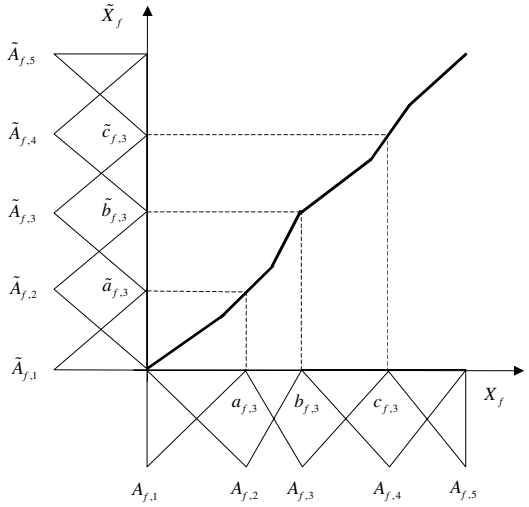


Figure 2: An example of piecewise linear transformation with granularity different from $T_{\max}$.

C_2 is a vector containing $F+1$ natural numbers: the f^{th} element of the vector contains the number $T_f \in [2, T_{\max}]$ of fuzzy sets which partition the linguistic variable X_f. $T_{\max}$ is fixed by the user and is the same for all the linguistic variables.

C_3 is a vector containing $F+1$ vectors of $T_{\max}-2$ real numbers: the f^{th} vector contains the $[b_{f,2},...,b_{f,T_{\max}-1}]$ points which define the piecewise linear transformation for the linguistic variable X_f. To preclude that the piecewise linear transformation can become decreasing, we force $b_{f,j}$ to vary in $\left[\tilde{b}_{f,j} - \frac{\tilde{b}_{f,j} - \tilde{b}_{f,j-1}}{2}, \tilde{b}_{f,j} + \frac{\tilde{b}_{f,j+1} - \tilde{b}_{f,j}}{2}\right]$, $\forall j \in [2, T_{\max}-1]$.

Each chromosome is associated with a bi-dimensional objective vector. The first element of the vector measures the complexity of the MFRBS as the number of propositions used in the antecedents of the rules contained in the concrete RB (the number of antecedents may be different, as explained in the next subsection, between the virtual and concrete RBs). The second element assesses the accuracy as the half of the mean square error (MSE) between the output of the MFRBS and the expected output.

4.2 Genetic operators

In order to generate the offspring populations, we exploit both crossover and mutation. We apply separately the one-point crossover to C_1 and C_2 and the BLX-α crossover, with α = 0.5, to C_3. Let s_1 and s_2 be two selected parent chromosomes. The common gene for C_1 is chosen by extracting randomly a number in $[M_{\min}, \rho_{\min}]$, where $M_{\min}$ is the minimum number of rules, which must be present in a rule base, and $\rho_{\min}$ is the minimum number of rules in s_1 and s_2. The common gene for C_2 is extracted randomly in $[1, F+1]$.

ISBN: 978-989-95079-6-8

As regards mutation, we apply two mutation operators for C_1. The first operator adds γ rules to the virtual RB, where γ is randomly chosen in $[1, \gamma_{\max}]$. The upper bound $\gamma_{\max}$ is fixed by the user. If $\gamma + M > M_{\max}$, then $\gamma = M_{\max} - M$. For each rule R_m added to the chromosome, we generate a random number $v \in [1, F]$, which indicates the number of input variables used in the antecedent of the rule. Then, we generate v natural random numbers between 1 and F to determine the input variables which compose the antecedent part of the rule. Finally, for each selected input variable f, we generate a random natural number $j_{m,f}$ between 1 and $T_{\max}$, which determines the fuzzy set $A_{f,j_{m,f}}$ to be used in the antecedent of rule R_m in the virtual RB. To select the consequent fuzzy set $A_{F+1,j_{m,F+1}}$, a random number between 1 and $T_{\max}$ is generated.

The second mutation operator randomly changes δ elements of the matrix J associated with the virtual RB. The number δ is randomly generated in $[1, \delta_{\max}]$. The upper bound $\delta_{\max}$ is fixed by the user. For each element to be modified, a number is randomly generated in [0, $T_{\max}$].

The mutation applied to C_2 randomly chooses a gene $f \in [1, F+1]$ and changes the value of this gene by randomly adding or subtracting 1. If the new value is lower than 2 and larger than $T_{\max}$, then mutation is not applied.

The mutation applied to C_3 first chooses randomly a variable $f \in [1, F+1]$, then extracts a random value $j \in [2, T_{\max} - 1]$ and changes the value of $b_{f,j}$ to a random value in the allowed interval $\left[\tilde{b}_{f,j} - \frac{\tilde{b}_{f,j} - \tilde{b}_{f,j-1}}{2}, \tilde{b}_{f,j} + \frac{\tilde{b}_{f,j+1} - \tilde{b}_{f,j}}{2}\right]$.

We experimentally verified that these mating operators ensure a good balancing between exploration and exploitation, thus allowing the multi-objective evolutionary algorithm described in the next subsection to create good approximations of the Pareto fronts.

4.3 Multi-objective evolutionary algorithm

We adopted the (2+2)M-PAES proposed in [5]. Unlike classical (2+2)PAES, which uses only mutation to generate new candidate solutions, (2+2)M-PAES exploits both crossover and mutation. Further, in (2+2)M-PAES, current solutions are randomly extracted at each iteration rather than maintained until they are not replaced by solutions with particular characteristics.

Fig. 3 shows a pseudo-code which describes the application scheme of the different operators to generate the offspring solutions o_1 and o_2 from the selected parents s_1 and s_2. Note that P_{c1}, P_{c2} and P_{c3} represent the probabilities of applying the crossover operator on C_1, C_2 and C_3, respectively. Similarly, P_{m1}, P_{m2} and P_{m3} represent the probabilities of applying the mutation operator on C_1, C_2 and C_3, respectively, and P_{add} and P_{inc} represent the probabilities of adding rules and of increasing the granularity of the selected variable, respectively. In Fig. 3 we also report the values of the probabilities we used in our experimentations.

At the beginning, we generate two solutions s_1 and s_2. The genes of C_2 in both the solutions are set to the maximum value $T_{\max}$, while the genes of C_1 and C_3 are randomly generated. At each iteration, the application of crossover and mutation operators produces two new candidate solutions from the current solutions s_1 and s_2. These candidate solutions are added to the archive only if they are dominated by no solution contained in the archive; possible solutions in the archive dominated by the candidate solutions are removed. Typically, the size of the archive is fixed at the beginning of the execution of the (2+2)M-PAES. In this case, when the archive is full and a new solution z has to be added to the archive, if z dominates no solution in the archive, then we insert z into the archive and remove the solution (possibly z itself) that belongs to the region with the highest crowding degree [5]. If the region contains more than one solution, then, the solution to be removed is randomly chosen.

```
P_c1 = P_c2 = P_c3 = 0.5;
P_m2 = P_m3 = 0.2;
P_add = 0.55;
P_inc = 0.85;
…
//Generate two new solutions
[s_1, s_2] = random_selection(archive)
o_1 = s_1
o_2 = s_2
if (rand() < P_c1)
    [o_1.C_1,o_2.C_1]=crossover_C_1 (s_1.C_1,s_2.C_1);
    P_m1 = 0.01;
else
    P_m1 = 1;
endif
if (rand() < P_c2)
    [o_1.C_2,o_2.C_2]=crossover_C_2 (s_1.C_2,s_2.C_2);
endif
if (rand() < P_c3)
    [o_1.C_3,o_2.C_3]=crossover_C_3 (s_1.C_3,s_2.C_3);
endif
loop i=1,2
    if (rand() < P_m1)
        if (rand() <P_add)
          o_i.C_1 = first_mutation_operator();
        else
          o_i.C_1 =second_mutation_operator();
        endif
    endif
    if (rand() < P_m2)
        if (rand <P_inc)
          o_i.C_2 = increase_granularity();
        else
          o_i.C_2 = decrease_granularity();
        endif
    endif
    if (rand() < P_m3)
       o_i.C_3 = transf_function_mutation();
    endif
endloop
```

Figure 3: Genetic operators application scheme.

ISBN: 978-989-95079-6-8

5 Experimental results

We tested our approach on a real world regression problem provided by the Bilkent University Function Approximation Repository [17]. The dataset contains the weather data of Ankara, a Turkish town, collected from 01/01/1994 to 28/05/1998. Here, the goal is to predict the mean temperature from the values of nine weather features, namely *max temperature*, *min temperature*, *dewpoint*, *precipitation*, *sea level pressure*, *standard pressure visibility*, *wind speed*, and *max wind speed*. The dataset contains 1609 instances of these features and corresponding mean temperatures. In order to asses the reliability of our approach, we performed a five-fold cross-validation and executed six trials, with different seeds, for each fold. We set $T_{\max}=5$, $M_{\min}=5$, $M_{\max}=30$, $\gamma_{\max}=5$, $\delta_{\max}=5$, the maximum number of evaluations to 300000 and the archive size to 64.

To assess the advantages of learning concurrently RB, granularities of partitions and MF parameters, we compared the results achieved by our approach with the ones obtained by applying the (2+2)M-PAES to learn only rules, that is, using only the first part of the chromosome and fixing the number of fuzzy sets in all partitions to $T_{\max}$, and to learn concurrently the RB and the MF parameters, using only the first and third parts of the chromosome with fixed granularities. We denote these two approaches as PAES-RB and PAES-SF, respectively, while the proposed approach is denoted as PAES-SFG.

Figures 4 and 5 show the average Pareto fronts achieved by the three algorithms on the training and test sets, respectively. The average Pareto front consists of the average values of the twenty most accurate solutions of each of the thirty Pareto front approximations (at least twenty solutions were always generated in all the trials for both datasets). In practice, the average Pareto fronts are obtained as follows. First, the solutions in the Pareto front approximations produced in each of the thirty trials are ordered for increasing MSE values. Then, only the twenty solutions with the lowest MSEs are retained for each Pareto front approximation. Finally, the average values, on the thirty Pareto front approximations, of complexity and MSE for these twenty solutions are computed. The choice of considering only the twenty solutions with the lowest MSEs was motivated by the observation that the other solutions are in general characterized by quite high MSEs which make these solutions impractical.

We note that the average Pareto front generated by PAES-SFG outperforms the other two average Pareto fronts both on the training and on the test sets. This proves that learning concurrently the RB, the granularities of the input and output partitions and the membership parameters allows generating FRBSs with better trade-offs between accuracy and complexity than the other two MOEAs used for comparison.

In Table 1 we show the average results corresponding to three representative points of the average Pareto fronts: the first (the most accurate), the median and the last (the least accurate) point. We refer to these average values as First, Median and Last, respectively. It is interesting to observe that the average MSEs of the most accurate solutions generated by PAES-SFG are 47% and 68%, and 33% and 56% lower than the corresponding solutions in the Pareto fronts generated by PAES-SF and PAES-RB with comparable complexity on the training and test sets, respectively.

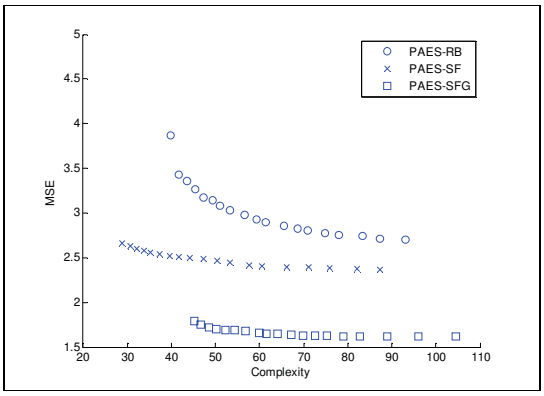


Figure 4. Average Pareto fronts on training set.

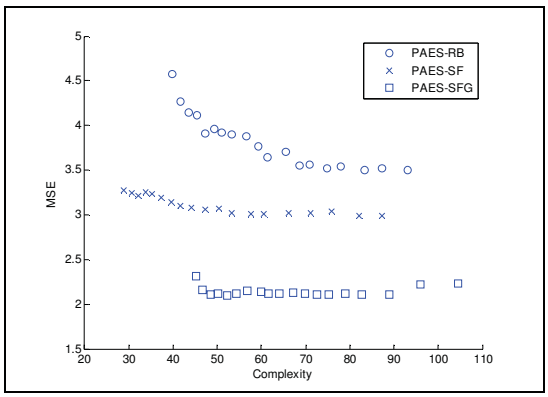


Figure 5. Average Pareto fronts on test set.

Table 1. Results obtained on three representative points of the average Pareto fronts

	MSE_{Tr}	MSE_{Ts}	Complexity	# Rules
	PAES-RB			
First	2.70±1.27	3.50±1.90	93.20±19.99	28.30±2.65
Median	2.93±1.40	3.77±2.17	59.33±17.02	22.80±4.48
Last	3.87±3.08	4.58±3.43	39.80±16.36	17.67±5.00
	PAES-SF			
First	2.36±1.60	2.99±2.18	87.23±17.33	24.20±4.24
Median	2.49±1.71	3.06±2.21	47.30±14.11	17.23±4.08
Last	2.66±1.88	3.28±2.51	28.80±11.10	12.97±3.69
	PAES-SFG			
First	1.61±0.34	2.24±1.11	104.53±27.86	25.77±4.59
Median	1.64±0.35	2.12±0.62	64.13±21.15	20.97±4.81
Last	1.79±0.62	2.31±1.11	45.10±15.76	17.33±4.19

Finally, Figures 6 and 7 show an example of complete Pareto fronts generated by the three algorithms in a single trial on the training and test sets, respectively. We have plotted only the solutions of the final archive with the MSEs lower than 5, since the other solutions in the archive are characterized by quite high MSEs and tend to visually flatten the plot. On the

ISBN: 978-989-95079-6-8

other hand, these solutions, though characterized by very low complexity, are not very interesting due to the high MSEs. We observe that the approximated Pareto fronts obtained by PAES-SFG are wide and quite dense. Further, except for some solutions in the low complexity zone, all the solutions in the Pareto front generated by PAES-SFG dominate the solutions at the same complexity of the other two fronts.

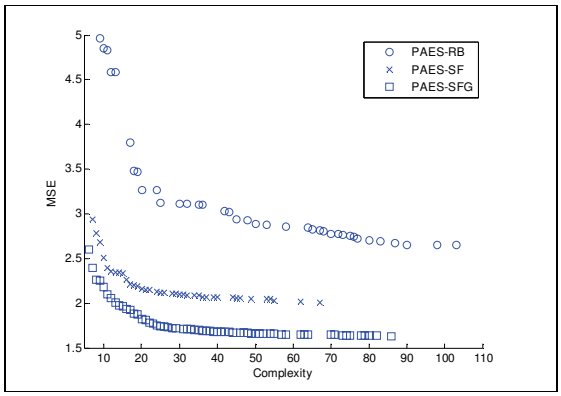


Figure 6. Examples of complete Pareto fronts on training set.

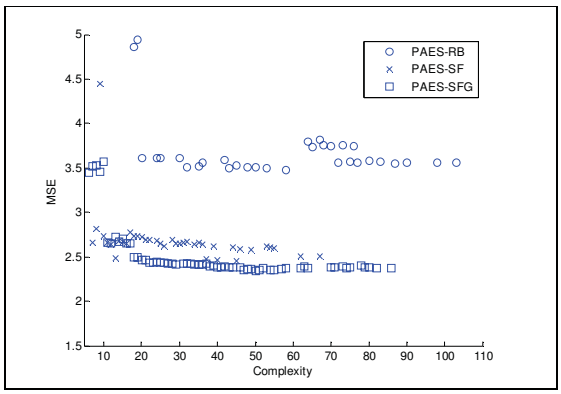


Figure 7. Examples of complete Pareto fronts on test set.

6 Conclusions

In this paper we have proposed a multi-objective evolutionary algorithm to generate a set of Mamdani rule-based fuzzy systems with optimal trade-offs between accuracy and complexity. We have exploited a modified version of the well known (2+2)PAES where the chromosome consists of three parts which codify the RB, and, for each variable, the number of fuzzy sets and the parameters of a piecewise linear transformation of the membership functions, respectively. To manage the dependence between granularity, and rule and membership parameters definition, we have adopted the following solution: the RB coded in the chromosome and the piecewise linear transformation are always defined on linguistic variables partitioned with a fixed maximum number of fuzzy sets. Only when accuracy and complexity have to be evaluated, the RB is actualized by using the real number of fuzzy sets determined by the second part of the chromosome and the DB by computing the corresponding piecewise linear transformation of the fuzzy sets. This approach has proved to be very efficient and effective, allowing both a good exploitation of the solutions and an accurate exploration of the search space.

The algorithm has been tested on a real world regression problem and has provided Pareto fronts with solutions characterized by better trade-offs between accuracy and complexity than solutions belonging to Pareto fronts generated by learning only the RB with a fixed BD, and concurrently the RB and the MF parameters with fixed granularities, respectively.

References

[1] E. H. Mamdani and S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, International Journal on Man-Machine Studies, 7: 1-13, 1975.

[2] J. Casillas et al. (eds.), Interpretability issues in fuzzy modeling, Springer, 2003.

[3] F. Herrera, Genetic fuzzy systems: taxonomy, current research trends and prospects, Evolutionary Intelligence, 1(1): 27-46, 2008.

[4] H. Ishibuchi, Multiobjective genetic fuzzy systems: review and future research direction, in Proc. of FUZZ-IEEE 2007 Int. Conf. on Fuzzy Systems, London, 2007.

[5] M. Cococcioni et al., A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems, Soft Computing, 11(11): 1013-1031, 2007.

[6] A. Botta et al., Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index, Soft Computing, 13(5): 437-449, 2009.

[7] R. Alcalá et al., A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 15(5): 539-557, 2007.

[8] M. J. Gacto, R. Alcalá, and F. Herrera, Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems, Soft Computing, 13(5): 419-436, 2009.

[9] P. Ducange et al., Knowledge Base Learning of Linguistic Fuzzy Rule-Based Systems in a Multi-objective Evolutionary Framework, Lecture Notes on Computer Science, Hybrid Intelligent Systems, pp. 747-754, 2008.

[10] P. Pulkkinen and H. Koivisto, Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms, International Journal of Approximate Reasoning, 48: 526-543, 2008.

[11] H. Ishibuchi and Y. Nojima, Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, International Journal of Approximate Reasoning, 44(1): 4-31, 2007.

[12] M. Antonelli et al., A multi-objective genetic approach to concurrently learn partition granularity and rule bases of Mamdani fuzzy systems, in Proc. of the 8th International Conference on Hybrid Intelligent Systems (HIS-2008), pp. 278–283, Barcelona, Spain, 2008.

[13] O. Cordón, F. Herrera, and P. Villar, Analysis and guidelines to obtain a good uniform fuzzy partition granularity for fuzzy rule-based systems using simulated annealing, International Journal of Approximate Reasoning 25(3):187-215, 2000.

[14] F. Klawonn, Reducing the number of parameters of a fuzzy system using scaling functions. Soft Computing, 10(9): 749-756, 2006.

[15] W. Pedrycz and F. Gomide, Fuzzy Systems Engineering: Toward Human-Centric Computing, Wiley-IEEE Press, 2007.

[16] H. Ishibuchi et al., Selecting fuzzy if-then rules for classification problems using genetic algorithms, IEEE Transactions on Fuzzy Systems, 3(3): 260-270, 1995.

[17] H. Altay Guvenir and I. Uysal, Bilkent University Function Approximation Repository, http://funapp.cs.bilkent.edu.tr, 2000.

ISBN: 978-989-95079-6-8

Semantically-driven flexible division in fuzzy object oriented models

Nicolás Marín[1] Carlos Molina[2] Olga Pons[1] M. Amparo Vila[1]

1. Department of Computer Science and AI, University of Granada, 18071, Granada, Spain
2.Department of Informatics, University of Jaén, Campus Las Lagunillas, A3, 23071, Jaén, Spain
Email: nicm@decsai.ugr.es, carlosmo@ujaen.es, {opc,vila}@decsai.ugr.es

Abstract— *Fuzzy Databases have been a fertile research area that has produced a wide variety of remarkable solutions for the storage and manipulation of imperfect information. Proposals can be found in the three most relevant data models: relational, object-oriented, and object-relational. Query capabilities have been especially studied in the context of the relational data model, while in object models many problems continue being a matter of research. In this paper, we focus on the resolution of "division queries" (in its relational sense) in an object oriented data model. As we will see, the presence of fuzzily described objects in the database make necessary to use suitable operators that take into account the resemblance that governs the comparison in the underlying reference universe. We also analyze the role of cardinality in this kind of queries.*

Keywords— Division, Quotient, Fuzzy, Object-Oriented, Object-Relational, Database.

1 Introduction

Fuzzy Set Theory[1] has proved to be an adequate tool in handling real world data when they are affected by imperfections of very different kinds. The application of this theory in order to extend conventional databases has led to the development of fuzzy database systems.

Object-oriented and object-relational database management systems allow the representation of schema when complex relationships make the use of Codd's relational model difficult. The object-oriented data model is more powerful from a modeling point of view, because it incorporates important features such as inheritance and encapsulation. As it was the case with the relational data model in the past, many researchers have recently tried to improve object orientation with the help of fuzzy concepts. As a result, fuzzy object-oriented database models (FOODBM) have appeared [2, 3, 4].

In [5, 6] a framework is presented which allows programmers to handle imprecision in the description of objects. An object-relational implementation of this model can be found in [7].

In this paper, we focus on the study of operators that can be used to solve a special kind of queries in fuzzy object-oriented databases: those queries that, in the relational context, are known as fuzzy divisions. Fuzzy division queries in the relational data model has been deeply studied during the last decades: from the initial works of Dubois et al.[8] till the last works of Bosc et al.[9], the interested reader can find a wide survey of remarkable proposals(e.g. [10, 11, 12]).

This work analyzes how to solve this kind of queries over a fuzzy object model, where the equality is substituted by resemblance as the basis of comparisons. Due to this fact, we have to suitably soften the division operator, so that resemblance can be appropriately taken into account.

The paper is organized as follows: section 2 briefly introduces the fuzzy object-oriented data model that is the basis of our proposal; section 3 describes the type of queries we want to solve in an object-oriented context; section 4 analyzes the problem of considering the resemblance when computing the inclusion associated to division queries; section 5 analyzes the role that the cardinality of the involved sets plays in the resolution of the query; finally, some conclusions are outlined in section 6.

2 A fuzzy object-oriented model

As we have previously mentioned in the introduction, this paper is devoted to the problem of how to solve division queries in a fuzzy object model like the one introduced in [13]. We briefly present in this section a summary of its main characteristics.

As in any object model, the state of an object is equated with a set of attribute values according to its class description. The approach described in [13] considers different types of attribute values in order to give support to the representation of fuzzily described objects. Together with the *precise* values, objects, and *crisp* collections of any conventional object-oriented data model, the fuzzy model permits to use more powerful values, as imprecise labels and fuzzy collections, in order to deal with imprecision in the state of the object.

Fig. 1 shows a description of different types of attribute values that we consider in our approach. That is, the object state can be composed by:

- Precise values: This category of values involves all the classical basic classes that usually appear in an object-oriented data model (e.g. numerical classes, string classes, etc.). Values in these domains are easily represented and compared using conventional built-in data types and the classical set of relational operators.
- Imprecise values: The case of imprecise (atomic) values is a bit more complex. In many cases, linguistic labels[14, 15, 16] are associated to this kind of values, but different types of imprecise values must be considered according to their semantics.
- Objects: the attribute value may be a reference to another object (constituting what is called a complex object).
- Collections: the attribute may be conformed by a set of values or, even, by a set of objects. Imprecision in this kind of attributes may appear at two levels:

ISBN: 978-989-95079-6-8

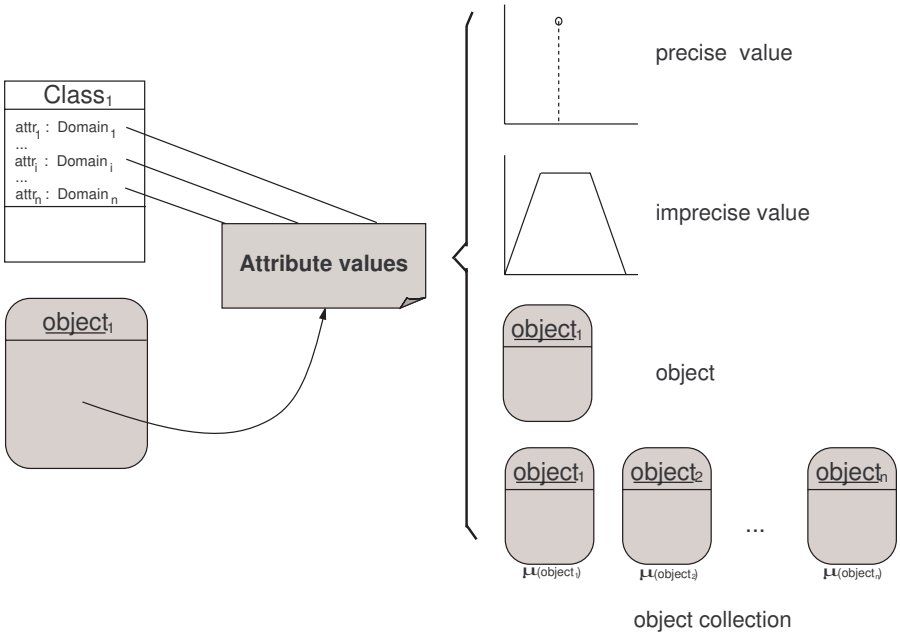


Figure 1: Different kinds of attribute values

– The set may be fuzzy. The semantics of membership degrees depends on the problem, but, in collections, the fuzzy set is considered to be conjunctive. For example, the set of languages a person can speak may be fuzzy if we take into account the degree of command for every language.
– The elements of the set may be fuzzy values or, in general, fuzzily described objects. For example, we can have a set of fuzzily described students.

An example of a fuzzy set of fuzzily described objects may be the collection of friends of a given person, if we use a degree to measure the friendship relation and the friends are described by means of fuzzy attribute values.

2.1 Generalization of equality

We do not only need a representation of fuzzily described objects in our model but also a way to manage this kind of objects. This includes a basic capability to compare the state of two objects belonging to a given class.

With fuzzily described objects like the ones presented in the previous paragraphs, this comparison has to be performed by a process which consists of two steps:

- computing degrees of resemblance for pairs of attribute values, and
- aggregating these resemblance degrees to obtain a general degree of object resemblance.

That is, if o_1 and o_2 are two objects belonging to class C which is characterized by type T_C whose structural component Str_C is an attribute set $\{x_1, x_2, \ldots, x_n\}$, then our goal is to find a resemblance degree between o_1 and o_2 by the aggregation of resemblance degrees of the pairs $(o_1.x_i, o_2.x_i)$.

$S_{x_i}(o_1, o_2)$ will stand for the resemblance degree observed between the x_i attribute values for objects o_1 and o_2, and $S(o_1, o_2)$ is the aggregated resemblance degree we want to calculate.

In this process, not all attributes need to have the same importance: each attribute x_i has an associated weight $p_{x_i} \in [0, 1]$ that represents its relative importance in the final decision.

Figure 2 summarizes the process of calculating resemblance degrees. In order to compare objects o_1 and o_2, we first compare their attribute values and obtain partial compatibility degrees $S_{x_i}(o_1, o_2)$. Then, we aggregate them to obtain a global resemblance opinion $S(o_1, o_2)$ according to the attribute importance established in the class.

As can be observed, the comparison of two objects involves a recursive procedure. For the sake of space, we omit here the operators for base cases and the analysis of the recursive aggregation procedure used in our model. Interested readers can found a detailed description in [17].

3 Division queries in fuzzy object-oriented databases

Codd defined a set of eight basic operators for his relational model. Some of them are implemented directly in SQL, but some other require particular implementation to fit into the SQL language. Relational division is one of the eight basic operations in Codd's relational algebra[18] that has these requirements.

Suppose that we have two relations t_a and t_b whose schema are, respectively, $T_A(X, Y)$ and $T_B(Y')$, where Y and Y' are compatible attributes. The division $t_a \div t_b$ obtains a new relation t_d with schema $T_D(X)$ and with the following set of tuples:

$$\{a | a \in domain(X) \wedge (\forall b, (b \in t_b) \rightarrow (< a, b > \in t_a)\} \quad (1)$$

Assume now that T_A and T_B are fuzzy relations, i.e., that their tuples are weighted by a number between 0 and 1. That

ISBN: 978-989-95079-6-8

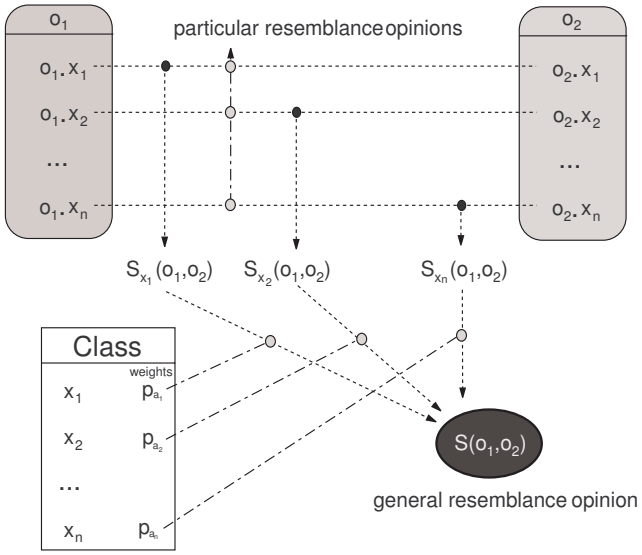


Figure 2: Obtaining the compatibility between two objects.

is, each tuple of t_a is affected by a membership degree, $\mu_{t_a}(< a, b >) \in [0, 1]$ and, similarly, each tuple of t_b has its corresponding $\mu_{t_b}(< b >) \in [0, 1]$.

The membership degree of the tuples of the division result can be obtained as follows [8]:

$$\mu_{t_d}(a) = inf_b\{\mu_{t_b}(b) \rightarrow \mu_{t_a}(a, b)\} \quad (2)$$

In the case of an object-oriented model, we have to consider (in the more general case) that attribute values are objects. Figure 3 depicts the tables with the new situation in an object-oriented context.

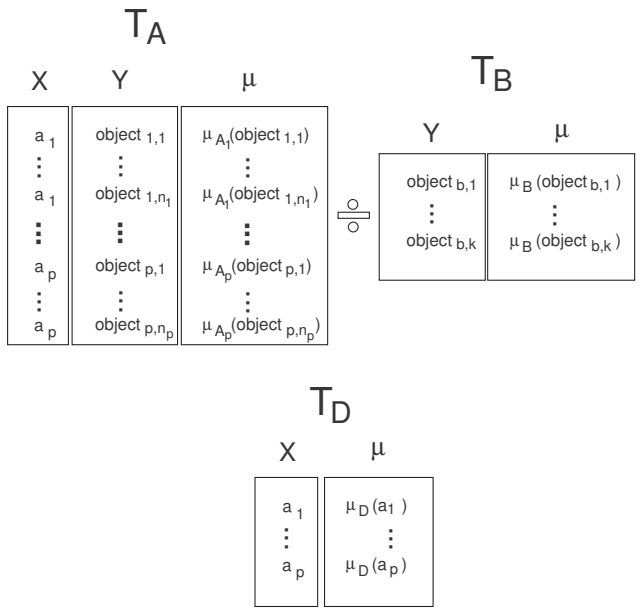


Figure 3: Tables involved in the division

Moreover, as in an object-oriented model attribute values do not necessarily have to be atomic, table T_A (and, similarly T_B) could have the schema showed in Fig. 4.

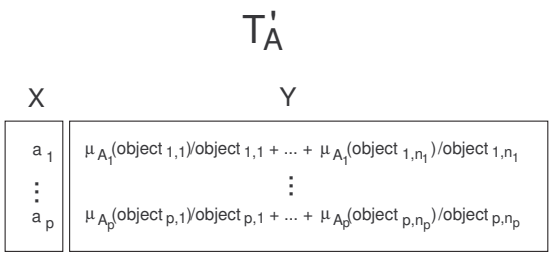


Figure 4: Object-oriented version of T_A

In any case, whatever the schema version of dividend and divisor tables are, the solution for division queries entails the resolution of the inclusion problem described in Fig. 5. Next sections are devoted to analyze this inclusion problem.

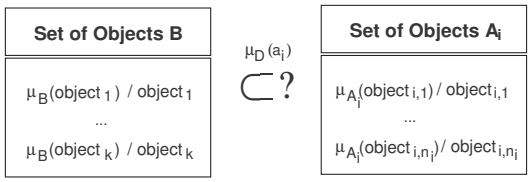


Figure 5: An inclusion problem

4 Two resemblance-based approaches

According to the previous paragraphs, in order to solve division queries in object-oriented models, we need to perform suitable computations of the above mentioned inclusion operator.

To compute the inclusion between any couple of fuzzy sets of objects (like in Fig. 5), we need to generalize the fuzzy inclusion operators, taking into account that the objects of the set may be fuzzily described.

Several proposals for the calculus of this inclusion degree can be found in the literature. In [19] the inclusion degree between two fuzzy sets A and B is calculated as follows:

$$N(B|A) = \min_{u \in U}\{I(\mu_A(u), \mu_B(u))\}, \quad (3)$$

where I stands for an implication operator, and μ_X is the membership function that describes the fuzzy set X. This degree coincides with the fuzzy division operator described in (2). The implication operator can be chosen in accordance to the properties we want the inclusion degree to fulfil [8, 10].

Nevertheless, independently of the chosen implication operator, this formulation supposes that both A and B are defined over a reference universe $\mathcal{U}$ made up of precise elements, where the classical equality is the basis of the comparisons. That is, the implication operator compares to what extent the presence of an element of the universe $\mathcal{U}$ in A forces its presence in B (i.e. we compare the membership degrees of the *same* object to both sets).

However, in the context of fuzzy databases, it frequently happens that the elements of the universe $\mathcal{U}$ are imprecise objects among which classical equality cannot be applied (as we

ISBN: 978-989-95079-6-8

commented in section 2. Instead, a similarity or a (more relaxed) resemblance relationship must be used. That is, for a given element in the set A, it is not clear which element of B has to be taken in order to compare the membership degrees.

In our example of object databases, the objects of the two set of Fig. 5 may be defined with imprecision and two *apparently* different objects (from the identity point of view) may be the *same* one (from the value equality point of view).

That is, if the reference universe $\mathcal{U}$ is formed by fuzzily described elements, a generalized version of (3) that takes into account resemblance among the involved objects must be used. We can consider two distinct ways of doing that, namely, making resemblance acts as a limit of the implication or making resemblance acts as a limit of membership to A_i.

4.1 Resemblance as a limit of the implication

If the reference universe $\mathcal{U}$ is formed by fuzzily described elements, (3) can be generalized as follows [20]:

Definition 1 (Resemblance driven inclusion degree) *Let A and B be two fuzzy sets defined over a finite reference universe $\mathcal{U}$, S be a resemblance relation defined over the elements of $\mathcal{U}$, and $\otimes$ be a t-norm. The inclusion degree of B in A driven by the resemblance relation S is calculated as follows:*

$$\Theta_S(A|B) = \min_{x\in\mathcal{U}} \max_{y\in\mathcal{U}} \theta_{B,A,S}(x,y) \tag{4}$$

where

$$\theta_{B,A,S}(x,y) = \otimes(I(\mu_B(x),\mu_A(y)),\mu_S(x,y)) \tag{5}$$

In (3), the inclusion degree is calculated taking into account the element of $\mathcal{U}$ that fulfils in a lower degree the implication condition between the membership degrees of this element to both sets. In (4), since the membership degrees of similar (but not necessarily equal from the identity point of view) elements can be compared, we restrict the implication using the resemblance degree of the two elements. In summary, for each element that belongs with a certain degree to the set A, we look for a *quite similar* object in $\mathcal{U}$ that belongs to the set B with an equal or higher degree.

Consider the example of Fig. 6.

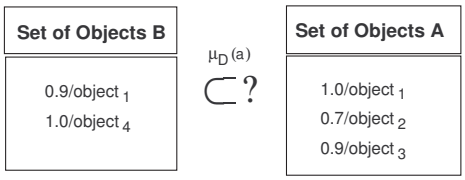


Figure 6: Example

In this example, $\{o_1, o_2, o_3, o_4\}$ is the reference universe $\mathcal{U}$ over which the resemblance relation S of table 1 is defined.

In this situation:

$\Theta_S(A|B)$= min {max {$\otimes$(I($\mu_B(o_1)$, $\mu_A(o_1)$), $\mu_S(o_1,o_1)$), ..., $\otimes$ (I($\mu_B(o_1)$, $\mu_A(o_4)$), $\mu_S(o_1,o_4)$)}, ..., max{$\otimes$(I($\mu_B(o_4)$, $\mu_A(o_1)$), $\mu_S(o_4,o_1)$), ..., $\otimes$(I($\mu_B(o_4)$, $\mu_A(o_4)$), $\mu_S(o_4,o_4)$)}}.

If we use the minimum as t-norm and (6) as implication operator, then:

$\Theta_S(A|B)$= min {max {1,0,0,0}, max {0,1,0,0}, max {0,0,1,0.7}, max {0,0,0.7,0}}= min{1,1,1,0.7}=0.7.

Table 1: Resemblance Relation

	o_1	o_2	o_3	o_4
o_1	1.0	0.0	0.0	0.0
o_2		1.0	0.0	0.0
o_3			1.0	0.7
o_4				1.0

$$I(x,y) = \begin{cases} 1, & \text{if } x \le y \\ y/x, & \text{otherwise} \end{cases} \tag{6}$$

4.2 Resemblance as a limit of relationship

As an alternative to the strategy used in the previous subsection, we can consider that the resemblance only restricts the membership degree of a certain object to the set A.

The assumption of this second version of flexible inclusion is to enlarge the set A into a superset A' obtained by composing A with the resemblance relation as suggested in [9]. The idea is to expand A in the sense that the objects of the reference class $\mathcal{U}$ *similar* to an object initially present in A, are added to A. This resemblance based tolerant inclusion can be defined as follows.

Definition 2 (Resemblance based tolerant inclusion degree) *Let A and B be two fuzzy sets defined over a finite reference universe $\mathcal{U}$, S be a resemblance relation defined over the elements of $\mathcal{U}$, and $\otimes$ be a t-norm. The tolerant inclusion degree of B in A according to the resemblance relation S is calculated as follows:*

$$\Psi_S(A|B) = \min_{x\in\mathcal{U}} \max_{y\in\mathcal{U}} \psi_{B,A,S}(x,y) \tag{7}$$

where

$$\psi_{B,A,S}(x,y) = I(\mu_B(x), \otimes(\mu_A(y),\mu_S(x,y)) \tag{8}$$

According to the example of Fig. 6 and the resemblance relation defined in Table 1:

$\Psi_S(A|B)$= min {max {I($\mu_B(o_1)$, $\otimes$($\mu_A(o_1)$,$\mu_S(o_1,o_1)$)), , ..., I($\mu_B(o_1)$, $\otimes$($\mu_A(o_4)$,$\mu_S(o_1,o_4)$))}, ..., max {I($\mu_B(o_4)$, $\otimes$($\mu_A(o_4)$,$\mu_S(o_4,o_1)$)), ...,I($\mu_B(o_4)$, $\otimes$($\mu_A(o_4)$,$\mu_S(o_4,o_4)$))}}.

If we use the minimum as t-norm and (6) as implication operator, then:

$\Psi_S(A|B)$= min {max {1,0,0,0}, max {1,1,1,1}, max {1,1,1,1}, max {0,0,0.7,0}}= min{1,1,1,0.7}=0.7.

4.3 Some notes on the two approaches

Although with our example, both flexible implementations of inclusion deliver the same value, normally the behaviour of these alternatives will differ.

An example where the second approach is more optimistic than the first one is described below.

Consider the naive example of Fig. 7.

In this example, $\{o_1, o_2\}$ is the reference universe $\mathcal{U}$ over which the resemblance relation S of table 2 is defined.

If we use the minimum as t-norm and (6) as implication operator, then:

$\Theta_S(A|B)$=0.6

$\Psi_S(A|B)$=1.0

ISBN: 978-989-95079-6-8

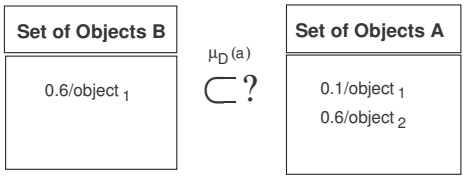


Figure 7: Example

Table 2: Resemblance Relation

	o_1	o_2
o_1	1.0	0.6
o_2		1.0

If we change the resemblance relation as in Table 3 and maintain the rest of data, then
$\Theta_S(A|B)$=0.9
$\Psi_S(A|B)$=1.0

Table 3: Resemblance Relation

	o_1	o_2
o_1	1.0	0.9
o_2		1.0

From a semantic point of view, with this configuration of the *Tolerant Inclusion*, in the computation of $\psi_{B,A,S}(x, y)$, the membership degree to A of a given object y bounds the needed resemblance between x and y. This effect can diminish if we use t-norms more restrictive than the minimum.

5 The role of cardinality

In order to finish our analysis of inclusion operators between fuzzy sets of fuzzily described objects, we now pay attention to cardinality, which may play a relevant role when computing the inclusion for two main reasons: On the first hand, if we soften the inclusion so that *similar* objects can fulfil the implication constraint, we have to be careful with the number of involved objects (the same object of A can be *matched* to more than one object of B); on the other hand, another point where the inclusion operators are able to be softened is to be tolerant with the number of objects of set B that we force to be included in A.

These two reasons lead us to new approaches to the inclusion operator which is supporting the division we want to solve.

5.1 Cardinality as a constraint of similarity

The use of resemblance relations instead of equality when computing inclusion degrees may make the obtained value be high, even if there is a great difference in terms of cardinality between the set of elements of B and the subset of elements of A matched to them during the computation of inclusion.

In some situations, cardinality may not be important. For example, imagine that A and B are two sets of *tools* and that we want to solve the question *Can a person do with tools of A all the tasks he can do with the tools of B?*. In this case, for each tool of B, we look for a tool of A with similar capabilities. In this context, it does not matter if the same tool of A is matched to more than one tool of B. That is, the number of *selected tools of A* is not relevant. However, if we need to assign this set of tools to a group of people that have to work independently, then, the number of *selected tools of A* is relevant and cardinality is important.

If we wish to distinguish between these situations, we need to weight the inclusion degree with a factor that takes into account the distance between the cardinalities of the fuzzy sets that are being compared.

Definition 3 (Cardinality Factor) *Let $\mathcal{U}$ be a reference universe. Let $|X|$ stands for the crisp cardinality of X. $CF : P(\mathcal{U}) \times P(\mathcal{U}) \to [0, 1]$ is a cardinality factor if:*

1. *If $|X| = |Y|$, $CF(X, Y) = 1$*
2. *If $|X| >= |Y|$, $CF(X, C) >= CF(Y, C)$*
3. *If $|X| >= |Y|$, $CF(C, Y) >= CF(C, X)$*

A simple example of CF is the following one:

$$CF(X,Y) = \begin{cases} 1, \text{ if } |X| >= |Y| \\ |X|/|Y|, \text{ otherwise} \end{cases} \quad (9)$$

This CF can be composed with a *relative fuzzy quantifier*[21] to adjust the behavior to the user needs.

Using (9) in the example of Fig. 8 and the resemblance relation of Table 4, we have:
$\Theta_S(A|B)$=1.0
$\Psi_S(A|B)$=1.0
but CF(matched elements of A,B)=0.5.

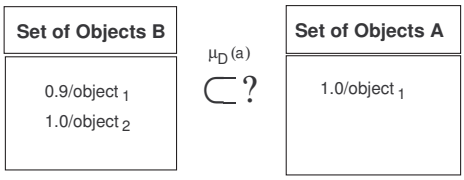


Figure 8: Example

Table 4: Resemblance Relation

	o_1	o_2
o_1	1.0	1.0
o_2		1.0

5.2 Cardinality as a relaxation of the inclusion

The other role cardinality can play in an inclusion operator is to soften the amount of elements of B that have to fulfil the inclusion relation. That is, with this kind of relaxation, the idea is to compute to what extent *almost all* elements of B are included in A. For example, in the case of tools, we want to solve the question *Can a person do with tools of A almost all the tasks he can do with the tools of B?*

There are many possible implementations of this idea in relation to what is called approximate division[8, 12]. One of the possible implementations of the corresponding approximate inclusion is to substitute the infimum in (3) by an OWA operator based on the desired *almost all* quantifier[22]. This way, we can adapt Θ and Ψ operators as follows.

ISBN: 978-989-95079-6-8

Definition 4 (Approximate Θ) *Let A and B be two fuzzy sets defined over a finite reference universe $\mathcal{U}$, S be a resemblance relation defined over the elements of $\mathcal{U}$, $\otimes$ be a t-norm, and $OWA_{\mathcal{Q}}$ an OWA operator induced by a regular non-decreasing quantifier $\mathcal{Q}$. The approximate inclusion degree of B in A driven by the resemblance relation S according to $OWA_{\mathcal{Q}}$ is calculated as follows:*

$$\Theta_{\mathcal{Q},S}(A|B) = OWA_{\mathcal{Q}}(\{\max_{y\in\mathcal{U}} \theta_{B,A,S}(x,y)\}_{x\in\mathcal{U}}) \qquad (10)$$

Definition 5 (Approximate Ψ) *Let A and B be two fuzzy sets defined over a finite reference universe $\mathcal{U}$, S be a resemblance relation defined over the elements of $\mathcal{U}$, $\otimes$ be a t-norm, and $\mathcal{Q}$ a relative fuzzy quantifier. The approximate tolerant inclusion degree of B in A based on the resemblance relation S according to $\mathcal{Q}$ is calculated as follows:*

$$\Psi_{\mathcal{Q},S}(A|B) = OWA_{\mathcal{Q}}(\{\max_{y\in\mathcal{U}} \psi_{B,A,S}(x,y)\}_{x\in\mathcal{U}}) \qquad (11)$$

6 Conclusions

In this paper, we have presented an approach to solve queries that are similar to the division queries of the relational data model but, now, in an object-oriented context. The fact that attribute values can be fuzzily described objects make us to extend the division operator by means of resemblance measures. We have considered two ways of incorporating such resemblance measures in the computation of inclusion, namely, restricting membership and restricting inclusion. Additionally, we have analyzed the role that cardinality can play in the division operator. Thus, we have considered to weight the inclusion operators with a cardinality factor that takes into account the number of elements of set A matched during the inclusion analysis. We have also propose two approximate inclusion operators based on the use of a quantifier for relaxing the division condition. With the described survey of operators, the user can semantically adapt the division operators to her/his needs in an object-oriented context.

Acknowledgment

Work partially supported by the projects *Representación y Manipulación de Objetos Imperfectos en Problemas de Integración de Datos: Una Aplicación a los Almacenes de Objetos de Aprendizaje*, Junta de Andalucía (TIC03175), and *Tratamiento de Información Difusa Temporal en un SGBD Relacional: Formulación Teórica, Implementación y Aplicaciones*, Estado Español (TIN2008-02066).

References

[1] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] Rita De Caluwe. *Fuzzy and Uncertain Object-Oriented Databases: Concepts and Models*, volume 13 of *Advances in Fuzzy Systems-Applications and Theory*. World Scientific, 1997.

[3] J. Lee, J-Y. Kuo, and N-L. Xue. A note on current approaches to extent fuzzy logic to object oriented modeling. *International Journal of Intelligent Systems*, 16:807–820, 2001.

[4] Zongmin Ma. *Advances In Fuzzy Object-oriented Databases: Modeling And Applications*. IGI Publishing, Hershey, PA, USA, 2004.

[5] N. Marín, O. Pons, and M. A. Vila. Fuzzy types: A new concept of type for managing vague structures. *International Journal of Intelligent Systems*, 15:1061–1085, 2000.

[6] F. Berzal, N. Marín, O. Pons, and M. A. Vila. Development of applications with fuzzy objects in modern programming platforms: Research articles. *Int. J. Intell. Syst.*, 20(11):1117–1136, 2005.

[7] L. Cuevas, N. Marín, O. Pons, and M. A. Vila. pg4db: A fuzzy object-relational system. *Fuzzy Sets Syst.*, 159(12):1500–1514, 2008.

[8] Didier Dubois and Henri Prade. Semantics of quotient operators in fuzzy relational databases. *Fuzzy Sets and Systems*, 78:89–93, 1996.

[9] P. Bosc, A. Hadjali, and O. Pivert. Exceptions and resemblance: Two keys for tolerant division operators. *Fuzzy Information Processing Society, 2008. NAFIPS 2008. Annual Meeting of the North American*, pages 1–6, May 2008.

[10] Patrick Bosc, Didier Dubois, Olivier Pivert, and Henri Prade. Flexible queries in relational databases—the example of the division operator. *Theor. Comput. Sci.*, 171(1-2):281–302, 1997.

[11] Xiaohui Tang and Guoqing Chen. A complete set of fuzzy relational algebraic operators in fuzzy relational databases. *Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference on*, 1:565–569 vol.1, July 2004.

[12] P. Bosc, O. Pivert, and D. Rocacher. About quotient and division of crisp and fuzzy relations. *J. Intell. Inf. Syst.*, 29(2):185–210, 2007.

[13] F. Berzal, N. Marín, O. Pons, and M. A. Vila. Managing fuzziness on conventional object-oriented platforms: Research articles. *Int. J. Intell. Syst.*, 22(7):781–803, 2007.

[14] L. A. Zadeh. The concept of linguistic variable and its application to aproximate reasoning I. *Information Sciences*, 8:199–251, 1975.

[15] L. A. Zadeh. The concept of linguistic variable and its application to aproximate reasoning II. *Information Sciences*, 8:301–357, 1975.

[16] L. A. Zadeh. The concept of linguistic variable and its application to aproximate reasoning III. *Information Sciences*, 9:43–80, 1975.

[17] Fernando Berzal, Juan C. Cubero, Nicolás Marín, María Amparo Vila Miranda, Janusz Kacprzyk, and Slawomir Zadrozny. A general framework for computing with words in object-oriented programming. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(Supplement-1):111–131, 2007.

[18] E. F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387, 1970.

[19] J-P. Rossazza, D. Dubois, and H. Prade. A hierarchical model of fuzzy classes. In *Fuzzy and Uncertain Object-Oriented Databases. Concepts and Models.*, volume 13, pages 21–61. Advances in Fuzzy Systems- Applications and Theory, 1998.

[20] N. Marín, J. M. Medina, O. Pons, D. Sánchez, and M. A. Vila. Complex object comparison in a fuzzy context. *Information and Software Technology*, 45(7):431 – 444, 2003. Software Engineering with Computational Intelligence S.I.

[21] Lotfi A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computers and Mathematics with Applications*, 9(1):149 – 184, 1983.

[22] R. R. Yager. Families of owa operators. *Fuzzy Sets and Systems*, (59):125–148, 1993.

ISBN: 978-989-95079-6-8

Rule Base and Inference System Cooperative Learning of Mamdani Fuzzy Systems with Multiobjective Genetic Algorithms

Antonio A. Márquez Francisco A. Márquez Antonio Peregrín

Information Technologies Department, University of Huelva
Huelva, Spain
Email: {amarquez, alfredo.marquez, peregrin}@dti.uhu.es

Abstract—*In this paper, we present an evolutionary multiobjective learning model achieving positive synergy between the Inference System and the Rule Base in order to obtain simpler, more compact and still accurate linguistic fuzzy models by learning fuzzy inference operators together with Rule Base. The Multiobjective Evolutionary Algorithm proposed generates a set of Fuzzy Rule Based Systems with different trade-offs between interpretability and accuracy in linguistic fuzzy modeling, allowing the designers select the one that involves the most adequate equilibrium for the desired application.*

Keywords— Linguistic fuzzy modeling, interpretability-accuracy trade-off, Multiobjective genetic algorithms, adaptive Inference System, adaptive defuzzification, rule learning.

1 Introduction

The main objective in system modelling is to develop reliable and understandable models. Interpretability and accuracy are usually contradictory requirements in the design of linguistic fuzzy models (FMs). Recent research into genetic fuzzy systems has focused on methods aimed at generating Fuzzy Rule-Based Systems (FRBS) with an appropriate trade-off between accuracy and interpretability [1, 2].

Two important tasks in the design of a linguistic FM for a particular application are the derivation of the linguistic Rule Base (RB) and the setup of the Inference System and defuzzification method. In the framework of the trade-off between interpretability and accuracy in fuzzy modeling, adaptive Inference Systems and defuzzification methods have acquired greater importance [3, 4].

Recently, the use of Multiobjective Evolutionary Algorithms (MOEA) has been applied to improve the aforementioned trade-off between interpretability and accuracy of linguistic fuzzy systems [5, 6, 7, 8, 9, 10, 11]. Some of them obtain the complete Pareto (the set of non-dominated solutions with different trade-offs) by selecting or learning the set of rules which best represents the example data, i.e., improving the system accuracy and decreasing the FRBS complexity. In [7, 8, 11] the authors also propose tuning the membership functions together with the rule selection to obtain simpler yet still accurate linguistic FMs.

Following these ideas on the advantage of using parametric operators and MOEAs to improve the trade-off between interpretability and accuracy, in [9] we presented a MOEA capable of learning the fuzzy inference operators (including inference and defuzzification) and of performing rule selection for Mamdani linguistic fuzzy systems. The proposed model aimed to achieve a positive synergy, that is, cooperation between the fuzzy operators and the RB to improve accuracy while at the same time simplifying the RB to improve interpretability.

Our main objective in this work is to include a new and highly important element in the learning process, the complete RB. Thus, we propose a MOEA capable of generating a set of FRBS (each with a high degree of cooperation between the RB and the Inference System) with varying optimal trade-offs between accuracy and complexity, so as to obtain compact and accurate linguistic fuzzy models by learning fuzzy operators and RB.

To do this, Section 2 describes the parametric fuzzy operators, Section 3 shows the RB learning used in this work, Section 4 is devoted to describing the MOEA learning proposal, Section 5 develops an experimental study, and finally, Section 6 presents some concluding remarks.

2 Adaptive Fuzzy Operators

In this section we describe the adaptive Inference System as well as the adaptive defuzzification method used in our learning proposal.

2.1 Adaptive Inference System

Linguistic FRBSs for system modeling use IF - THEN rules of the following form:

$$R_i: \text{If } X_{i1} \text{ is } A_{i1} \text{ and } \ldots \text{ and } X_{im} \text{ is } A_{im} \text{ then } Y \text{ is } B_i$$

with i = 1 to N, where N stands for the number of rules of the RB, X_{i1} to X_{im} and Y for the input and output variables respectively, and A_{i1} to A_{im} and B_i for the involved antecedents and consequent labels, respectively.

The expression of the Compositional Rule of Inference in fuzzy modeling with punctual fuzzification is the following: $\mu_{B'}(y) = I(C(\mu_{A1}(x_1), \ldots, \mu_{Am}(x_m)), \mu_B(y))$, where $\mu_{B'}(\cdot)$ is the membership function of the inferred consequent, $I(\cdot)$ is the implication operator, $C(\cdot)$ is the conjunction operator, $\mu_{Ai}(x_i)$ are the values of the matching degree of each input of the system with the membership functions of the rule antecedents, and $\mu_B(\cdot)$ is the consequent of the rule.

ISBN: 978-989-95079-6-8

The two components, the conjunction (C(·)) and the implication operator (I(·)) are suitable for parametrization in order for the Inference System to be adapted. Our previous studies in [3] show that models based on the adaptive conjunction is a more valuable option than those based on the adaptive implication operator. Hence, we selected the adaptive conjunction in this study in order to insert parameters in the Inference System.

Taking into account the aforementioned studies in [3], we have selected the Dubois adaptive t-norm with a separate connector for every rule, the expression for which is shown in (1).

$$T_{Dubois}(x, y, \alpha) = \frac{x \cdot y}{Max(x, y, \alpha)}, \quad (0 \leq \alpha \leq 1) \qquad (1)$$

This adaptive t-norm showed the highest accuracy in previous studies, compared with Frank and Dombi t-norms and is more efficiently computed. The use of an adaptive t-norm for the antecedent connection seeks better performance than traditional t-norms. Dubois t-norm performs between minimum ($\alpha = 0$) and algebraic product ($\alpha = 1$).

2.2 Adaptive Defuzzification Interface

There are various tendencies in the development of adaptive defuzzification methods reported in the literature. These employ one or more parameters in their expression for modifying the behaviour of the defuzzifier or, in most cases, to achieve higher accuracy.

Following the studies developed in [13], in this work we consider applying the defuzzification function to the fuzzy set inferred by each rule (getting a characteristic value) and computing them by a weighted average operator, because of its fine performance, efficiency and easier implementation. This way of working is named FITA (First Infer, Then Aggregate) [12].

We also consider the use of a product functional term of the matching degree between the input variables and the rule antecedent fuzzy sets (h_i), $f(h_i) = h_i \cdot \beta_i$ where β_i corresponds to one parameter for each rule R_i, i=1 to N, in the RB, as it is more efficiently computed and obtains similar results to other functions [13]. The adaptive defuzzification formula selected is shown in (2).

$$y_0 = \frac{\sum_{i}^{N} h_i \cdot \beta_i \cdot V_i}{\sum_{i}^{N} h_i \cdot \beta_i}, \qquad (2)$$

where V_i represents a characteristic value of the fuzzy set inferred from rule R_i, the Maximum Value or the Gravity Center (GC), the latter being the one selected in this paper.

The product functional term with a different parameter for each rule has the effect of weighted rules. This value associated with the rule indicates the importance of that rule for the inference process.

3 Rule Base Learning

The linguistic RB learning used in this work is based on the ad-hoc data driven methodology named COR [14]. This methodology manages a set of consequent label sets (one per rule). Instead of selecting the consequent with the best performance in each subspace as usual (Wang and Mendel [15]), the COR methodology considers the possibility of using another consequent, different from the best, which allows the FRBS to be more accurate thanks to having a RB with best cooperation. For this purpose, COR performs a combinatorial search among the candidate rules looking for the set of consequents which globally achieves the best accuracy.

COR consists of two stages:

1) *Construction of the search space*—This obtains a set of candidate consequents for each rule.

2) *Selection of the most cooperative fuzzy rule set*—This performs a combinatorial search among these sets seeking the combination of consequents with the best global accuracy.

In order to perform this combinatorial search, an *explicit enumeration* or an *approximate search technique* can be considered. In this work, we use a search technique because it is effective and quick.

4 Rule Base and Inference System Cooperative Learning with Multiobjective Algorithms

This Section describes the evolutionary multiobjective model proposed in this work. As was previously mentioned, our objective is to obtain a set of fuzzy systems with different trade-offs between accuracy and interpretability, using adaptive inference and defuzzification, and Rule Base learning (including rule selection). To do this, we exploit two specific MOEAs considering a threefold coding scheme (coding of rules and coding of the parameters of the Inference Systems and Defuzzification). We adopted two of the most representative second generation MOEAs, SPEA2 [16] and NSGA-II [17], as two general purpose MOEAs for performing the cooperative adaptation of the fuzzy operators and fuzzy rule learning.

4.1 SPEA2 and NSGA-II

The SPEA2 algorithm [16] (*Strength Pareto Evolutionary Algorithm for Multiobjective Optimization*) is one of the most well-known techniques for solving multiobjective problems. It is characterized by the following two aspects: a *fitness* assignment strategy, which takes into account both dominating and dominated solutions for each individual, and a density function, estimated by employing the nearest neighbourhood, which guides the search more efficiently.

NSGA-II algorithm [17] is another of the most well-known and frequently-used MOEAs for general multi-objective optimization in the literature. It is a parameterless

ISBN: 978-989-95079-6-8

approach with several interesting principles: a binary tournament selection based on fast non-dominated sorting, an elitist strategy and a crowding distance method to estimate the diversity of a solution.

4.2 Questions related to the MOEAs.

The evolutionary model uses a chromosome with threefold coding scheme (C_C+C_D+C_R) where:

- C_C encodes the α_i parameters of the conjunction connective. They are N real coded parameters (genes), one for each rule, R_i ,of the linguistic RB. Each gene can take any value in the interval [0, 1], that is, between the minimum and the algebraic product. This is represented by the C_C part of the chromosome shown in Figure 1.

- C_D encodes the β_i parameters of the defuzzification. They are N real coded parameters, one for each rule, of the linguistic RB. Each gene can take any value in the interval [0, 10]. This interval has been selected according to the study developed in [13]. It allows attenuation as well as enhancement of the matching degree. This is represented by the real part C_D shown in Figure 1.

- C_R encodes the learning Rule Base. It is an integer string of N genes, each one representing a candidate rule consequent of the initial RB. Furthermore, depending on whether a rule is selected or not, the value '-1' is assigned to the corresponding gene. This is represented by the integer part of the chromosome shown in Figure 1.

The initial population is randomly initialized in the fuzzy operators part with the exception of a single chromosome:

- C_C with the N genes is initiated to 0 in order to make Dubois t-norm equivalent to Minimum t-norm initially.

- C_D also with the N genes is initiated to 1 with the objective of beginning like the standard WCOA method.

The initial population in the fuzzy rule part, C_R, is initialized following these two exceptions:

- A single chromosome with the N rules obtained by the WM-method [15], that is, with all the genes initialized to correspondent consequent.

- Default chromosomes randomly initiated with all rules activated. In this case, in order to achieve solutions with a high accuracy we should not lose rules that could present a positive cooperation once their FM parameters have been evolved. The best way to do this is to start with solutions that select all the possible rules. This favors a progressive extraction of bad rules (those that do not improve with the tuning of parameters).

The crossover operator employed by the fuzzy operators part is BLX-0.5 [18] while the one used for the rule learning part is HUX [19].

Finally, four offspring are generated by combining the two from the C_R part with the two from the operators part (the two best replace their parents). The mutation operator changes a gene value at random in the C_R and operators part (one in each part) with probability 0.2.

In this work, to obtain an optimal set of FRBS with different trade-offs, the fitness, based on the interpretability (using the number of rules) and the accuracy (using the error measure), must be minimized.

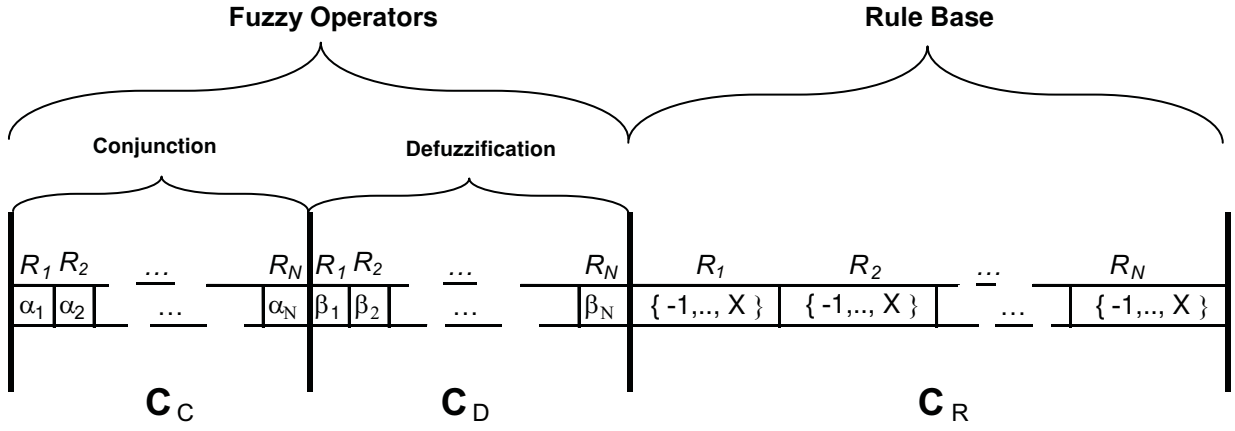


Fig. 1. Coding scheme for the MOEA with N rules

ISBN: 978-989-95079-6-8

5 Experimental Study

In order to analyse the practical behaviour of the proposed methods, we built several FMs in a real-world problem [20] of four input variables consisting in estimating the maintenance costs of medium voltage lines in a town. Methods considered for the experiments are briefly described in Table 1, where WM and COR methods are considered as reference. S denotes the methods that perform rule selection. If rule selection is performed after another method we give it a "+" denotation (eg. COR+S). However, if rule selection is performed jointly with another method we denote this in subscript (eg. COR_S). S–C–D means rule selection and fuzzy operators learning together. $SPEA2_{S\text{-}C\text{-}D}$, and $NSGA\text{-}II_{S\text{-}C\text{-}D}$ are the methods that learn the fuzzy operators and the rule selection together, while $SPEA2COR_{S\text{-}C\text{-}D}$ and $NSGA\text{-}IICOR_{S\text{-}C\text{-}D}$ are methods that also learn the RB as previously mentioned.

Table 1. Methods considered for comparisson

Ref.	Method	Description
[15]	WM	Wang & Mendel algorithm
[21]	WM + S	Wang & Mendel and then Rule Selection
[9]	WM + S–C–D	Wang & Mendel and then Rule Selection and Adaptive Fuzzy Operators
[14]	COR	COR
	COR + S	COR and then Rule Selection
	COR_S	COR with Rule Selection
	$COR_{S\text{–}C\text{–}D}$	COR with Rule Selection and Adaptive Fuzzy Operators
[9]	$SPEA2_{S\text{-}C\text{-}D}$	SPEA2 algorithm with Rule Selection and Adaptive Fuzzy Operators
-	$SPEA2COR_{S\text{-}C\text{-}D}$	SPEA2 algorithm with COR with Rule Selection and Adaptive Fuzzy Operators
[9]	$NSGA\text{-}II_{S\text{-}C\text{-}D}$	NSGA-II algorithm with Rule Selection and Adaptive Fuzzy Operators
-	$NSGA\text{-}IICOR_{S\text{-}C\text{-}D}$	NSGA-II algorithm with COR with Rule Selection and Adaptive Fuzzy Operators

5.1 Application Selected and Comparison Methodology

The application selected to test the evolutionary model is the aforementioned electrical distribution problem [20] that has a data set of 1059 cities with four input variables and a single output. The RB is composed of 65 linguistic rules achieved with the Wang and Mendel method [15]. The fuzzy partition used for inputs and output has 5 labels.

We considered a 5-fold cross-validation model, i.e., 5 random partitions of the data each with 20% (4 of them with 211 examples, and one of them with 212 examples), using the combination of 4 of them (80%) as training, and the remaining one as a test. We achieved a total of 30 trials for each evolutionary process, as the learning methods were run 6 times for each one of the data partitions. We show the average values of the medium square error (MSE) as a usual performance measure, computed considering the most accurate solution from each Pareto obtained with the multiobjective algorithm. This method of working was also employed in [9] in order to compare the single objective methods with the multiobjective ones based on considering the accuracy objective only, letting us see that the Pareto fronts are not only wide but also optimal, so similar solutions obtained with the WM + S–C–D or $COR_{S\text{–}C\text{–}D}$ must appear in the final Pareto. The MSE is computed with expression (3),

$$MSE\ (FM)_B = \frac{\frac{1}{2}\sum_{k=1}^{P}(y_k - FM(x_k))^2}{P} \qquad (3)$$

where *FM* denotes the fuzzy model the Inference System of which uses the Dubois t-norm as conjunction operator showed in expression (1), the inference operator is minimum t-norm, and the adaptive defuzzificación method is the one shown in expression (2). This measure uses a set of system evaluation data formed by P pairs of numerical data $Z_k = (x_k, y_k)$, k=1,..,P, with x_k being the values of the input variables, and y_k being the corresponding values of the associated output variables. The MOEAs population size was fixed at 200. The external population size of the $SPEA2_{S\text{-}C\text{-}D}$ and $SPEA2COR_{S\text{-}C\text{-}D}$ was 61.

5.2 Results and Analysis

To compare the results obtained we also used non-parametric tests, according to the recommendations made in [22]. The results obtained are shown in Table 2, where #R is the average number of rules, MSE_{tra} and MSE_{tst} are the average MSE for training and test respectively, and *Wilcoxon-test* is the result of applying a *Wilcoxon* signed-ranks test [23] (with 95% confidence), with the following interpretation: * represents the best average result (control algorithm); + means that the best result has better performance than that of the corresponding row, while sign (=) means it is similar to the best result. As we have mentioned, Demšar [20] recommends a set of simple, safe and robust non-parametric tests for statistical comparisons of algorithms, one of which is the Wilcoxon signed-ranks test [23]. This is analogous to the paired t-test in non-parametrical statistical procedures.

Table 2 only shows the best result for each MOEAs for accuracy. Analysing the results we can highlight the two following points:

- The learning of the RB allows remarkable improvement in accuracy: Looking at Table 2, we can observe that $SPEA2COR_{S\text{-}C\text{-}D}$ improves the accuracy of $SPEA2_{S\text{-}C\text{-}D}$ and $NSGA\text{-}IICOR_{S\text{-}C\text{-}D}$ improves $NSGA\text{-}II_{S\text{-}C\text{-}D}$. In spite of the fact that the number of rules shown in Table 2 is slightly larger, it must be taken into account that the results shown are the ones with the highest accuracy along the Pareto front. Looking at Table 3, we can observe similar values for accuracy in solutions with a lower number of rules from 40 to 32, so $NSGA\text{-}IICOR_{S\text{-}C\text{-}D}$ truly obtains better fuzzy systems (for accuracy and

ISBN: 978-989-95079-6-8

interpretability) than NSGA-II$_{S\text{-}C\text{-}D}$. Thus, the cooperation between the rules and the fuzzy operators improves the results of the evolutionary multiobjective proposal without rule learning.

- The solution with the best accuracy obtained with NSGA-II$_{COR_{S\text{-}C\text{-}D}}$ and SPEA2$_{COR_{S\text{-}C\text{-}D}}$ shows a similar accuracy to the single objective evolutionary model $COR_{S\text{–}C\text{–}D}$ (as shown in the non-parametric test) with a significant reduction in the number of rules, particularly for NSGA-II$_{COR_{S\text{-}C\text{-}D}}$. Consequently, the proposed method achieves more interpretable models with similar accuracy. We also notice that the best accuracy in Table 2 is obtained by the single objective model $COR_{S\text{–}C\text{–}D}$. The difference is small, but we can deduce that the evolutionary multiobjective methods are not achieving the most accurate solution. This fact suggested in [9] the design of more specific multiobjective algorithms in order to get even better solutions than the generic MOEAs SPEA2 and NSGA-II. Figure 2 shows the Pareto progress for each evolutionary algorithm (SPEA2$_{COR_{S\text{-}C\text{-}D}}$ and NSGA-II$_{COR_{S\text{-}C\text{-}D}}$) where we can observe the Pareto movement for each generation. Because to the adaptive fuzzy operators search space is large, we consider it may be necessary to focus the search process on the Pareto zone with highest accuracy, so that the same accuracy can be achieved as with single objective evolutionary algorithms based on accuracy.

Table 2. Results obtained

Method	#R	MSE$_{tra}$	Wilcoxon-test	MSE$_{test}$	Wilcoxon-test
WM	65	56135.75	+	56359.42	+
WM + S	40.9	41517.01	+	44064.67	+
WM + S–C–D	52.8	22640.95	+	26444.43	+
COR	65	50710.80	+	54584.76	+
COR + S	44.7	40763.48	+	43228.38	+
COR$_S$	43	39530.19	+	41060.99	+
COR$_{S\text{–}C\text{–}D}$	50	**20123.39**	*	**23323.72**	*
SPEA2$_{S\text{-}C\text{-}D}$	**38,60**	24021,41	+	29333,72	+
SPEA2$_{COR_{S\text{-}C\text{-}D}}$	41,10	21254,70	+	24079,32	=
NSGA-II$_{S\text{-}C\text{-}D}$	38,90	23364,63	+	28174,76	+
NSGA-II$_{COR_{S\text{-}C\text{-}D}}$	40,27	**20689,86**	=	**23346,34**	=

Table 3. A Pareto front example obtained from NSGA-II$_{COR_{S\text{-}C\text{-}D}}$

#R	MSE$_{tra}$	MSE$_{test}$	#R	MSE$_{tra}$	MSE$_{test}$	#R	MSE$_{tra}$	MSE$_{test}$
40	20036,37	22771,09	32	22354,40	24819,15	25	30133,70	42029,38
39	20198,77	22768,15	31	23266,76	25226,77	24	32481,97	43640,18
38	20388,66	22675,47	30	24347,59	27510,49	23	35058,65	43578,74
37	20783,97	23251,03	29	24994,05	26904,67	22	39128,61	53587,74
36	20999,33	23435,43	28	26190,94	29391,74	21	43276,59	58873,29
34	21719,71	24057,71	27	27445,85	30597,32	20	47428,26	69762,81
33	21901,55	24051,67	26	28706,22	33035,11	20	47428,26	69762,81

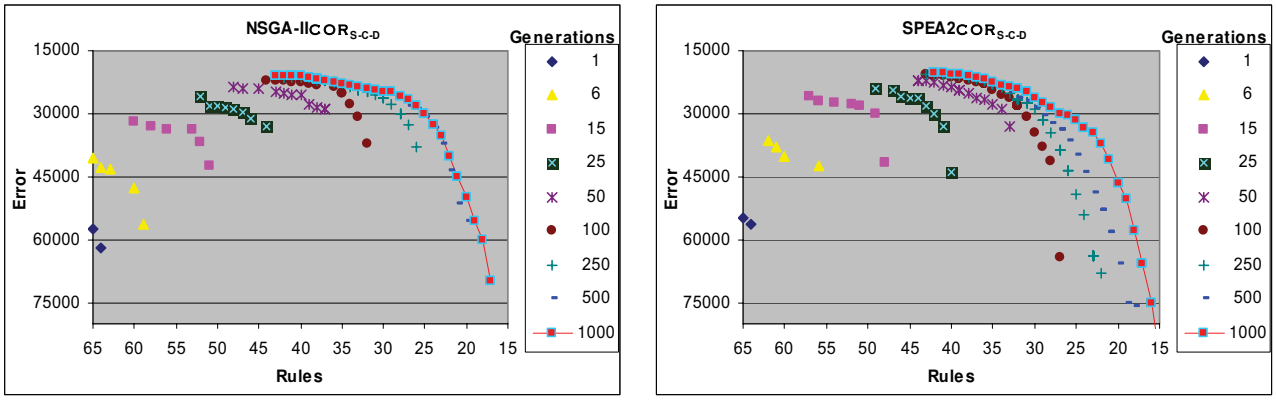


Fig. 2. Example of the Pareto front for NSGA-II$_{COR_{S\text{-}C\text{-}D}}$ and SPEA2$_{COR_{S\text{-}C\text{-}D}}$

.

ISBN: 978-989-95079-6-8

6 Conclusions

In the framework of the trade-off between accuracy and interpretability, the use of MOEAs gives a set of solutions with different levels of conciliation between both features. In this work we have proposed a multiobjective evolutionary learning model where the adaptive fuzzy operator parameters are learnt together with the RB. This fact allows both elements to cooperate, improving the accuracy as well as the interpretability.

The results obtained have shown that the use of MOEAs can represent a way to obtain a set of FRBSs in a single run with optimal trade-off between accuracy and interpretability. In terms of future work, some improvements may be developed in order to guide the search towards the desired Pareto zone with higher accuracy (right and central zone) where the FRBSs obtained are perhaps more interesting in more applications.

By focusing the search process we can reduce the effort of the search, and a better precision in the non-dominated solutions can be obtained, because the search effort is concentrated on a reduced zone of the Pareto, such that the density of the obtained solutions is higher. An improvement could be a change in the MOEAs used or a change in the non-dominated definitions in order to give more weight to the objective of accuracy.

Acknowledgment

Paper Supported by Projects TIN2008-06681-C06-06 and P07-TIC-03179.

References

[1] J. Casillas, O. Cordón, F. Herrera, et al. *Interpretability issues in fuzzy modeling*. Springer-Verlag , 2003.

[2] J. Casillas, O. Cordón, F. Herrera, et al. *Accuracy improvements in linguistic fuzzy modeling*. Springer-Verlag , 2003.

[3] J. Alcala-Fdez, F. Herrera, F.A. Márquez, et al. Increasing Fuzzy Rules Cooperation Based On Evolutionary Adaptive Inference Systems. *Int. J. of Intelligent Systems* 22(9): 1035-1064, 2007.

[4] F.A. Márquez, A. Peregrín, F. Herrera. Cooperative Evolutionary Learning of Fuzzy Rules and Parametric Aggregation Connectors for Mamdani Linguistic Fuzzy Systems. *IEEE Trans. on Fuzzy Syst*. 15(6): 1162-1178, 2007.

[5] H. Ishibuchi, T. Yamamoto. Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. *Fuzzy Sets and Syst*. 141: 59-88, 2004.

[6] K. Narukawata, Y. Nojima, H. Ishibuchi. Modification of evolutionary Multiobjective optimization algorithms for Multiobjective design of fuzzy rule-based classification systems. *Proc. 2005 IEEE Int. Conf. on Fuzzy Syst*. 809-814. Reno , 2005

[7] R. Alcalá, J. Alcala-Fdez, M.J. Gacto, et al. A Multi-Objective Evolutionary Algorithm for Rule Selection and Tuning on Fuzzy Rule-Based Systems. *Proc. of the 16th IEEE Int. Conf. on Fuzzy Syst. FUZZ-IEEE'07*, 1367-1372, London , 2007.

[8] R. Alcalá, M.J. Gacto, F. Herrera, et al. A Multi-Objective Genetic Algorithm for tuning and Rule Selection to obtain Accurate and Compact Linguistic Fuzzy Rule-Based Systems. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Syst.,* 15(5): 539-557, 2007.

[9] A.A. Márquez, F.A. Márquez, A. Peregrín. Cooperation between the Inference System and the Rule Base by using Multiobjective Genetic Algorithms. *Proceedings of Third Internacional of Hybrid Artificial Intelligence System,(HAIS'08)* 739-746 Burgos, Spain, 2008.

[10] M. Cococcioni, P. Ducange, B. Lazzerini, et al. A Pareto-based Multi-objective Evolutionary approach to the identification of Mamdani Fuzzy Systems. *Soft Computing* 11(11): 1013-1031, 2007

[11] M.J. Gacto, R. Alcalá, F. Herrera. Adaptation and Application of Multi-Objective Evolutionary Algorithms for Rule Reduction and Parameter Tuning of Fuzzy Rule-Based Systems. *Soft Computing* 13(5): 419-436, 2009.

[12] J.J. Buckley, Y. Hayashi. Can approximate reasoning be consistent?. *Fuzzy Sets and Syst.,* 65(1): 13-18, 1994.

[13] O. Cordón, F. Herrera, F.A. Márquez, et al. A Study on the Evolutionary Adaptive Defuzzification Methods in Fuzzy Modelling. *Int. J.of Hybrid Intelligent Syst*, 1(1): 36-48, 2004.

[14] J. Casillas, O. Cordón, F. Herrera. COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules. *IEEE Trans. Syst., Man, Cybern.—Part B: Cybern*. 32(4): 526-537, 2002.

[15] L.X. Wang, J.M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Trans. on Syst., Man, and Cybernetics* 22(6): 1414-1427, 1992.

[16] E. Zitzler, M. Laumanns, L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization. *Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems (EUROGEN'01*), 95-100, 2001.

[17] K. Deb, A. Pratap, S. Agarwal, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. on Evolutionary Computation* 6(2): 182-197 , 2002.

[18] L.J. Eshelman, J.D. Schaffer. Real-coded genetic algorithms and interval-schemata. *Found. of Genetic Algorithms* 2: 187-202, 1993.

[19] L.J. Eshelman. The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. *Found. of Genetic Algorithms* 1: 265-283 , 1991.

[20] O. Cordón, F. Herrera, L. Sánchez. Solving electrical distribution problems using hybrid evolutionary data analysis techniques. *Appl. Intell.,* 10, 5-24,1999.

[21] J. Casillas, O. Cordon, M.J. del Jesus, et al. Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Trans. on Fuzzy Syst*. 13(1):13–29, 2005.

[22] J. Demšar, Statistical comparisons of classifiers over multiple data sets, *J. Mach. Learn. Res*. 7: 1-30, 2006.

[23] D.J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Press, Boca Raton, FL, 2003.

ISBN: 978-989-95079-6-8

On the information provided by uncertainty measures in the classification of remote sensing images

Luísa M S Gonçalves [1,2] *Cidália C Fonte* [2,3] *Eduardo N B S Júlio* [4] *Mario Caetano* [5,6]

1 Polytechnic Institute of Leiria, Civil Engineering Department, Portugal
2 Institute for Systems and Computers Engineering at Coimbra, Portugal
3 Department of Mathematics, University of Coimbra, Portugal
4 ISISE, Civil Engineering Department, University of Coimbra, Portugal
5 Portuguese Geographic Institute (IGP), Remote Sensing Unit (RSU), Lisboa, Portugal
6 CEGI, Instituto Superior de Estatística e Gestão de Informação, ISEGI, Universidade Nova de Lisboa, Portugal
Email: luisag@estg.ipleiria.pt, cfonte@mat.uc.pt, ejulio@dec.uc.pt, mario.caetano@igeo.pt

Abstract—*This paper investigates the potential information provided to the user by the uncertainty measures applied to the possibility distributions associated with the spatial units of an IKONOS satellite image, generated by two fuzzy classifiers, based, respectively, on the Nearest Neighbour Classifier and the Minimum Distance to Means Classifier. The deviation of the geographic unit characteristics from the prototype of the class to which the geographic unit is assigned is evaluated with the Un non-specificity uncertainty measures proposed by [1] and the exaggeration uncertainty measure proposed by [2]. The classifications were evaluated using accuracy and uncertainty indexes to determine their compatibility. Both classifications generated medium to high levels of uncertainty for almost all classes, and the global accuracy indexes computed were 70% for the Nearest Neighbour Classifier and 53% for the Minimum Distance to Means Classifier. The results show that similar conclusions can be obtained with accuracy and uncertainty indexes and the latter, along with the analysis of the possibility distributions, may be used as indicators of the classification performance and may therefore be very useful tools. Since the uncertainty indexes may be computed to all spatial units, the spatial distribution of the uncertainty was also analysed. It's visualization shows that regions where less reliability is expected present a great amount of detail that may be potentially useful to the user.*

Keywords—Accuracy assessment, Minimum Distance to Mean Classifier, Nearest Neighbour Classifier, Non-specificity measures, Remote Sensing Images, Uncertainty.

1 Introduction

Some classifiers used in Remote Sensing allow the assignment of each spatial unit (pixel or object) to several classes through the computation of degrees of possibility, probability or membership associated with each class, as opposed to the traditional classifiers where each spatial unit is only assigned to one class [3, 4, 5, 6]. This additional information may be interpreted as degrees of membership of the spatial unit to the classes, and in this case are usually referred to as soft classifiers [7]. Even though the use of soft classifiers is increasing, most applications still require a hard classification into disjoint classes of interest. For these cases, the spatial units may be assigned to the class presenting the larger degree of possibility or probability. The additional information provided by these degrees of possibility of probability may be used as indicators of the classifier difficulty to assign only one class to the spatial unit and, together with the application of uncertainty measures, may provide valuable information to the user [2, 8, 9, 10].

In [1] it is shown that, for a fuzzy classifier which assigns to each spatial unit degrees of membership or possibility associated with the classes, the non-specificity measures may be used to estimate the classification accuracy. They can therefore be a useful tool, since, on one hand, they can give an estimation of the classification accuracy prior to the final accuracy evaluation, enabling the classification improvement before the final classification accuracy is assessed (which is a time consuming and expensive process), and on the other hand enable the spatialization of the uncertainty [11]. In this paper a very high resolution image is classified with two fuzzy classifiers and the uncertainty measures (two non-specificity measures and an exaggeration measure) are applied to both classifications, to determine the usefulness of the information provided by them. A comparison between the information provided by the non-specificity measures and the producer's and user's accuracy indexes is made to determine their compatibility.

To evaluate the usefulness of the information provided by the spatialization of uncertainty, an approach similar to the one proposed by [2] is considered, where it is proposed that, when the spatial units are assigned to the best degree of similarity, errors of omission and commission are committed. The former occurs because, when assigning the spatial unit only to the best class, its similarity to the other classes is omitted. The latter occurs because, when the spatial unit is assigned to the best class and the degree of similarity to that class is smaller that one, a commission error is introduced, since a higher similarity is considered. To estimate omission errors [2] used an entropy measure. Since the Shannon entropy should only be applied to probability measures, the U_n uncertainty measure was used instead. To estimate the commission errors the measure of exaggeration proposed in [2] was used. The results are shown and conclusions drawn.

2 Data and methods

2.1 Data

The study was conducted in a rural area with a smooth topographic relief, situated in a transition zone between the

ISBN: 978-989-95079-6-8

centre and south of Portugal featuring diverse landscapes representing Mediterranean environments. The area is occupied mainly by agriculture, pastures, forest and agro-forestry areas where the dominant forest species in the region are eucalyptus, coniferous and cork trees. An image obtained by the IKONOS sensor was used, with a spatial resolution of respectively 1m in the panchromatic mode and 4m in the multi-spectral mode (XS) and a dimension of 11 884 m by 14 432m. The geometric correction of the multi-spectral image consisted of its orthorectification. The average quadratic error obtained for the geometric correction was 1.39 m, inferior to half the pixel size, which guarantees an accurate geo-referencing.

2.2 Classification

Two fuzzy classifiers were used in this application to obtain the elementary entities that are the basic units of landscape, like crown trees and parts of buildings, called Surface Elements (SE), to produce a Surface Elements Map (SEM). Both classifiers compute degrees of similarity [12] between the values observed at each spatial unit and a set of ideal values, which are, in this case, the spectral response observed in each spatial unit and what is considered to be the ideal characteristics of the class in terms of spectral values, respectively. Even though both classifiers use the same theoretical tools to derive the degrees of similarity, the definition of the ideal characteristics of each class is different for each of them.

The first classification method used is an object–oriented supervised Fuzzy classifier based on the Nearest Neighbor Classifier (FNNC) available on the software eCognition. In the supervised classification methods, a training set has to be chosen for each class. The spectral responses of the pixels inside this set are the signature files used to characterize each class. This classification method assigns to each pixel a degree of membership to each class, depending on the distance between its spectral response and the closest spectral response of the pixels used in the training set for each class (see Fig. 1).

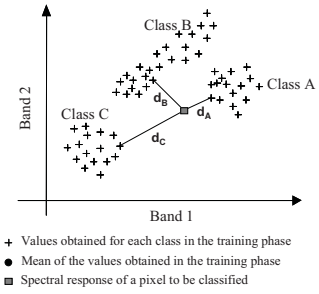


Figure 1: Nearest Neighbor Classifier.

The degrees of membership are computed using a membership function, where the ideal value, corresponding to a degree of membership equal to one, is obtained when the distance between the spectral response of the pixel to classify to the nearest reflectance value of the training set of the class is equal to zero (see Fig. 2).

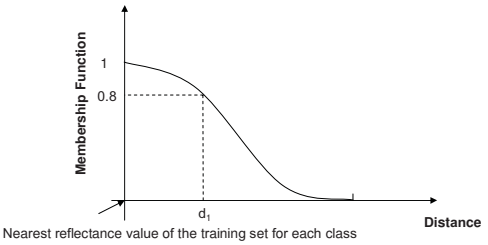


Figure 2: Membership function used in the fuzzy classifier based on the Nearest Neighbor Classifier.

The second classification method used is a pixel-based supervised Fuzzy classifier based on the underlying logic of Minimum-Distance-to-Means Classifier (FMDMC), available in the commercial software IDRISI. With this classification method, the image is classified based on the information contained in the signature files and a standard deviation unit (Z-score distance) introduced by the user. The fuzzy set membership is calculated based on a standardized Euclidean distance from each pixel reflectance, on each band, to the mean reflectance for each class signature (see Fig. 3), using a sigmoidal membership function (see Fig. 4).

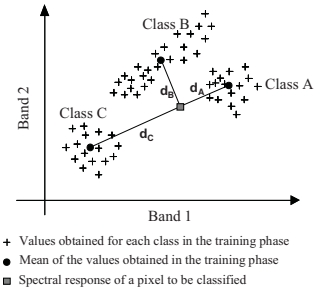


Figure 3: Minimum-Distance-to-Means Classifier.

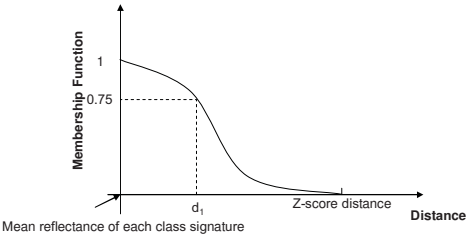


Figure 4: Membership function used in the fuzzy classifier based on the Minimum-Distance-to-Means Classifier

The underlying logic is that the mean of a signature represents the ideal point for the class, where fuzzy set membership is one. When distance increases, fuzzy set membership decreases, until it reaches the user-defined Z-score distance where fuzzy set membership decreases to zero. To determine the value to use for the standard deviation unit, the information of the training data set was used to study the spectral separability of the classes and determine their average separability measure.

The most representative surface elements of the study area are Eucalyptus Trees (ET), Coniferous Trees (CFT), Cork Trees (CKT), Shadows (S), Shallow Water (SW), Deep Water (DW), Herbaceous Vegetation (HV), Sparse Herbaceous Vegetation (SHV) and Non-Vegetated Area (NVA), and therefore these were the considered classes.

ISBN: 978-989-95079-6-8

2.3 Uncertainty measures

To evaluate the uncertainty of the possibilistic classifier the measures of nonspecificity were used. The uncertainty measure *NSp* proposed by [13] and given by

$$NSp(\Pi) = 1 - \sum_{i=1}^{n} \left[\Pi(x_i) - \Pi(x_{i+1}) \right] \frac{1}{i}$$

and a normalized version proposed by [1] of the U-uncertainty measure developed by [14], given by

$$U_n(\Pi) = \frac{[1 - \Pi(x_1)] \log_2 n + \sum_{i=2}^{n} [\Pi(x_i) - \Pi(x_{i+1})] \log_2 i}{\log_2 n}$$

In both equations Π is an ordered possibility distribution defined over a universal set X, that is, a possibility distribution such that $\Pi(x_1) \geq \Pi(x_2) \geq \ldots \geq \Pi(x_n)$, n is the number of elements of the universal set and $\Pi(x_{n+1})$ takes the value zero. A normalized version of the U-uncertainty measure was used to normalize the range of the measure, so that both non-specificity measures vary within the interval [0,1]. The non-specificity measures are appropriate to evaluate the uncertainty resulting from possibilistic classifications, since they quantify the ambiguity in specifying an exact solution [15, 16].

An additional uncertainty measure of exaggeration E, proposed by [2], was also used to quantify the commission errors at each spatial unit. This measure is given by

$$E = 1 - \Pi(x_1)$$

where $\Pi(x_1)$ is the largest degree of possibility.

2.4 Classification evaluation with the accuracy indexes

To evaluate the classifications accuracy a testing set was used. A random sampling of 900 pixels was selected considering the entire image scene. The sample unit was the pixel. The accuracy assessments was made with error matrixes, where the (i,j) entry is the number of pixels that is class i in the map and class j in the reference. The reference data were obtained from aerial images with larger resolution. The Global Accuracy was computed for both classifications. The User Accuracy (UA) and the Producer Accuracy (PA) indexes were also computed for all classes.

2.5 Comparison between uncertainty and accuracy indexes

To determine if the information given by the uncertainty measures may be used as indicators of the classifier's performance and if they are correlated with the results obtained with the classification quality assessment, two accuracy indexes were considered, namely the normalized user's and producer's accuracy; and one index for each uncertainty measure based on the mean, corresponding to the complement of the mean uncertainty per class ($I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$). Since high uncertainty values are expected to correspond to low accuracy values, and vice-versa, the complement of the uncertainty measures is considered to allow an easy comparison with the accuracy indexes. The correlation coefficient between the several accuracy and uncertainty indexes was also computed.

Even though uncertainty measures can be computed to the whole image, since the objective was to compare the results given by the uncertainty measures with the ones given by the error matrix, for this comparison only the sample pixels were used, so that the results were not influenced by the sample representativeness.

2.6 Spatial variation of uncertainty

Since there is a possibility distribution associated with all pixels or objects of the image, the uncertainty indexes may be computed for the whole image. This enables the visualization of the spatial distribution of uncertainty, and its spatial relation with the classes assigned to each spatial unit. As an estimator of the spatial units omission errors the U_n uncertainty measure was used, since it is more sensitive to dispersion than NSp (see [1]) and to estimate the commission errors the exaggeration E measure was used.

3 Results and discussion

3.1 Classification evaluation with accuracy indexes

The error matrix computed for the classification with the FNNC is presented in Fig. 5. The global accuracy obtained with this classification method was 70%.

	Error matrix of the classification with the fuzzy KNN									User's Accuracy (%)
	DW	SW	NVA	ET	S	HV	CKT	CFT	SHV	
DW	102									100.0
SW	3	92								96.8
NVA			76				2	1	8	87.4
ET				42	4	6	7	18	11	47.7
S	2				77		5			91.7
HV			2	1		90			7	90.0
CKT		1	11	8	18		50	6	26	41.7
CFT				3		43	10	42	9	39.3
SHV			23	3		22	16	3	73	52.1
Producer's Accuracy (%)	95.3	98.9	67.9	73.7	77.8	55.9	55.6	60.0	54.5	**69.77%**

Figure 5: Error matrix of the image classification with the method based on the Nearest Neighbour Classifier.

The error matrix shows that water classes (DW and SW) were well identified. Forestry species were often confused between each other and with other classes, such as Sparse Herbaceous Vegetation (SHV) and Herbaceous Vegetation (HV). Significant confusion was observed between Herbaceous Vegetation (HV) and Coniferous Trees (CFT). The class with the smaller value of PA is SHV (54.5%), which means it is the class with more omission error. And the class with smaller UA is CFT (39.3%) and therefore the class with more commission errors.

The global accuracy obtained with the classification with the FMDMC was 53% and the error matrix computed is presented in Fig. 6. The error matrix shows that Deep Water (DW) was the class best identified. Significant confusion was observed between Cork Trees (CKT), Eucalyptus Trees (ET) and Sparse Herbaceous Vegetation (SHV) and between ET and Coniferous Trees (CFT). The class with the smaller value of PA is SHV (14.8%), which means this is the class with more omission error. The pixels which are erroneously not included in this class are included mainly in the class Non Vegetated Areas (NVA), which is the class with smaller UA (27%) and therefore the class with more commission

ISBN: 978-989-95079-6-8

error, which receives pixels from all classes. These results show that this classifier presents great difficulty in classifying the class NVA.

	Error matrix of the classification with the fuzzy MDM									User's Accuracy (%)
	DW	SW	NVA	ET	S	HV	CKT	CFT	SHV	
DW	89				1					98.9
SW	3	54	1							93.1
NVA	16	40	104	8	19	79	15	15	89	27.0
ET				37		5	8	17	4	52.1
S	1				35		7			81.4
HV						65			1	98.5
CKT			7	13	3		50	11	18	49.0
CFT				5		16	6	27	3	47.4
SHV			3	1		2	6		20	62.5
Producer's Accuracy (%)	81.7	57.4	90.4	57.8	60.3	38.9	54.3	38.6	14.8	**53.21%**

Figure 6: Error matrix of the image classification with the method based on the Minimum Distance to Mean Classifier.

3.2 Comparison between uncertainty and accuracy indexes

For the classification made with the FNNC, the uncertainty indexes $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$ along with the user's and producer's accuracy, ordered with increasing values of $I_{1-\overline{NSp}}$, are shown in Fig. 7.

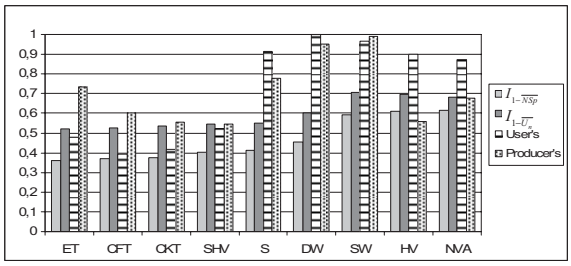


Figure 7: $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$ uncertainty indexes along with the user's and producer's accuracy of the classification based on the Nearest Neighbour Classifier.

It can be observed that for all classes the uncertainty indexes present low to medium values. To determine the reason for these relatively low values of certainty, the observation of the possibility distribution was made, and it was observed that the largest degree of possibility is in general very high, but also high levels of possibility are obtained for the second and sometimes even for the third classes, which explains the high levels of uncertainty (see Table 1). The analysis of the possibility distributions and the classes associated with the second degrees of possibility for the several points enable the identification of classes that might be easily confused, which are mainly the forest classes and SHV. Similar conclusions can be taken from the confusion matrix, with the exception that considerable confusion was also expected between CFT and HV, and this is not expected from the analysis of the degrees of possibility.

From this analysis, worst results are expected for the classes ET, CFT, CKT and SHV, which are exactly the classes with low levels of UA, $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$. This is an expected result, since in [1] it is shown that high correlation values are expected to occur between the uncertainty indexes and the UA. Similarly, the classes with higher values of $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$ are in general the ones presenting higher UA, except for the classes S and DW, which have slightly lower value of $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$.

Table 1: Mean values of the highest and second highest degrees of possibility of the possibility distributions associated with the sample points used to evaluate the accuracy of the classification with the FNNC, along with the most frequent second class.

Best class	$\overline{\Pi(x_1)}$	$\overline{\Pi(x_2)}$	Second class	% of pixels with the second class
SW	0.96	0.66	NVA	100
DW	0.99	0.87	NVA	51
NVA	0.84	0.37	SHV	52
ET	0.99	0.96	CKT	48
HV	0.93	0.55	SHV	50
SHV	0.94	0.83	CKT	52
CFT	0.98	0.94	ET	56
CKT	0.99	0.94	SHV	46
S	0.98	0.89	CKT	64

For the classification made with the FMDMC, the uncertainty indexes $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$ along with the user's and producer's accuracy, ordered with increasing values of $I_{1-\overline{NSp}}$, are shown in Fig. 8. It can be seen that the values of certainty given by both NSp and U_n are low to medium for all classes, presenting $I_{1-\overline{U_n}}$ slightly larger values than $I_{1-\overline{NSp}}$.

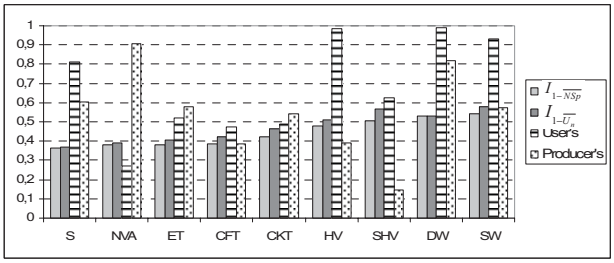


Figure 8: $I_{1-\overline{NSp}}$ and $I_{1-\overline{U_n}}$ uncertainty indexes along with the user's and producer's accuracy for the classification with the FMDMC.

To determine why such relatively high levels of uncertainty were obtained, an analysis of the possibility distributions was also made. This analysis showed that, for all classes, a high percentage of the pixels presented relatively low values for the highest degree of possibility, which explains the relatively low levels of certainty, and no point was classified with the highest degree of possibility equal to one (see Table 2). Further more, for all classes, the second highest degree of possibility for the great majority of testing sites was to the class Non Vegetated Areas (NVA) (except, of course, for the class NVA), in some cases with the degrees of possibility close to the higher degree, which means that there was great confusion between this class and all other classes.

ISBN: 978-989-95079-6-8

Table 2: Mean values of the highest and second highest degrees of possibility of the possibility distributions associated with the sample points used to evaluate the accuracy of the classification with the FMDMC, along with the most frequent second class.

Best class	$\overline{\Pi(x_1)}$	$\overline{\Pi(x_2)}$	Second class	% of pixels with the second class
SW	0.62	0.19	NVA	100
DW	0.53	0.003	NVA	26
NVA	0.41	0.07	HV	16
ET	0.48	0.14	NVA	88
HV	0.57	0.16	NVA	95
SHV	0.63	0.38	NVA	100
CFT	0.52	0.22	NVA	82
CKT	0.54	0.22	NVA	90
S	0.38	0.03	NVA	100

This indicates that one major problem of this classification was mainly due to the difficulty in discriminating NVA from the other classes. This same conclusion can be obtained from the accuracy analysis. The class NVA presents very different values for the user and producer accuracy, respectively 27% and 90%. This means that this class presents a very high percentage of commission errors, and therefore, a great amount of sites that should have been assigned to other classes were assigned to NVA, and were therefore absent from the other classes, such as SHV and HV, increasing their omission errors.

The information provided by the uncertainty measures is therefore in accordance with the one provided by the accuracy indexes, even though the analysis of the possibility distributions is necessary to reach the obtained conclusions.

The most evident discrepancies between the uncertainty and accuracy indexes obtained with the classification with the FMDMC are mainly for the class DW and S. These classes present higher values of accuracy, meaning that they were well identified according to the accuracy indexes, but considerably low levels of classification certainty, corresponding to lower values of the uncertainty indexes. This can be explained by the relatively low level of the first degree of possibility and the confusion between classes, which reinforces the assumption that uncertainty indexes capture the classifier difficulty in determining the correct class. Apparently some discrepancies occur between uncertainty and accuracy indexes for the SW and HV classes due to the fact that they present higher values of UA (93,1 and 98,5 respectively). However, the PA results are lower (57,4% and 38,9 % respectively) which reveal that these classes present higher omission errors and are in accordance with the lower values of $I_{1-\overline{U_n}}$ index. Two main aspects seem to be responsible for the results obtained with this classifier: 1) the larger degrees of possibility for each pixel are in general relatively low, which results in high uncertainty levels; 2) there is considerable confusion between all classes and the class NVA.

To explain the low degrees of possibility obtained with this classifier, a closer look to its classification approach is required. The computed degrees of possibility reflect the closeness of the spectral response of the testing set to an ideal value, which is the mean of the spectral responses obtained for the training set. The computed degrees of possibility are then obtained considering the distance between the spectral response at each testing site and that mean value, and therefore, if the obtained value is relatively distant from the mean, even if its spectral response is very close to some values obtained for the training set, a low degree of possibility will be assigned to it. This classifier and the obtained degrees of possibility, have therefore a limited capability to translate the information provided by the training set. For example, if a pixel has a spectral response equal to the spectral response of one of the pixels used in the training set, but which is relatively far from the mean, it will not have a degree of possibility of belonging to that class equal to one, as should be expected.

The fuzzy classifier based on the FNNC has a better behavior on this aspect, since it translates the information contained in the training set in a more reliable way. For example, all points with a spectral response equal to the spectral response of a point included in the training set will have a degree of possibility equal to one. Furthermore, if the spectral response of pixels is located inside the regions populated by the spectral responses of a particular class, high degrees of similarity are expected to occur to that class.

3.3 Spatial variation of uncertainty

Fig. 10 shows on a), b) and c) respectively an extract of the classification with the FNNC and the values of the U_n and E uncertainty measures. Fig. 10 d), e) and f) show the corresponding images obtained with the FMDMC.

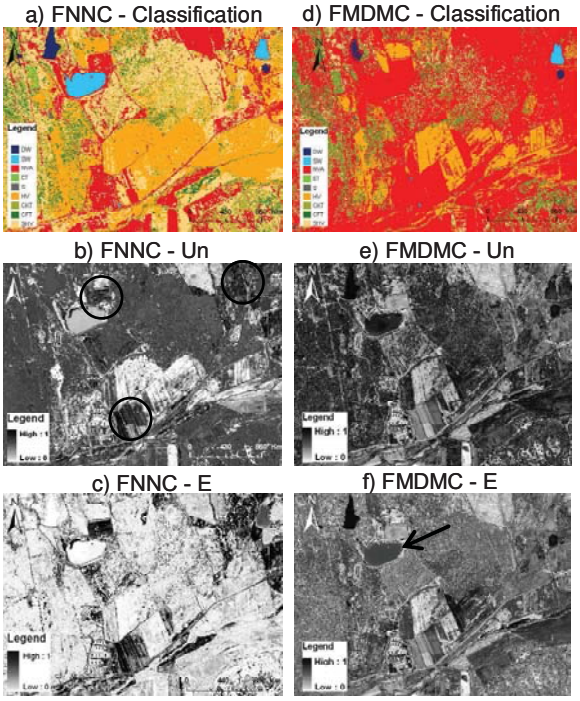


Figure 10: a), b) and c) show respectively the classification results, U_n and E uncertainty values obtained with the FNNC. Images d), e) and f) show the results corresponding to a), b) and c) obtained with the FMDMC.

It can be easily observed that the classification results are very different. Even though the accuracy evaluation and the

ISBN: 978-989-95079-6-8

uncertainty indexes already indicated that the classification with the FMDMC was worst, the visualization of the results clearly shows that it is bad. It can be seen that the FMDMC classified most of the image as NVA, while the FNNC manage to differentiate much more structures in the image.

Images shown in Fig. 10 b) and e) correspond to the spatial distribution of the U_n uncertainty measure of both classifications and may be used to illustrate the omission error committed when the pixel is assigned to the class corresponding to the largest degree of possibility. The regions with larger uncertainty are the ones where the degrees of possibility were more dispersed over other classes. Fig. 10b) shows that relatively large values of U_n were associated with regions that correspond to landscape units composed by different surface elements such as forest trees and Sparse Herbaceous Vegetation. Some region with higher U_n uncertainty can also be seen in the zones identified by the black circumferences. Fig. 10c) and f) show the commission errors, that is, the exaggeration committed when the best class is chosen. In general in Fig. 10c) low values are obtained for most of the image, higher values are mainly visible in some landscape boundaries and in the regions identified by the circumferences in Fig. 10b), which indicate, once again, that those regions may be problematic. Fig. 10f) shows that larger commission errors are obtained with the FMDMC than with the FNNC. One aspect clearly visible from Fig. 10e) and f) is that the spatial variation of uncertainty shows many structures invisible in the classification, such as the SW region shown in the region identified by the arrow.

4 Conclusions

The results obtained with the presented case study highlight that the classifiers FMDMC and FNNC presented possibility distributions with particular characteristics, which are a consequence of the method used to compute the degrees of possibility. The classification with the FNNC presented large possibility values for the best class, but also presented large values for the second and sometimes third class, and the classification with the FMDMC presented low values of the largest possibility. In both cases the presented characteristics increase uncertainty and are a signal of possible difficulties in the classification. In fact, neither of these classifications presented very good results, since Global Accuracies of only 70% and 53% were obtained.

It is also shown that when the classifications present relatively large and similar uncertainty values for all classes it is more difficult to take conclusions regarding the problematic classes only with these measures. Even though, an analysis of the possibility distributions may give valuable information. It is also very important to know how the degrees of possibility associated to the pixels are obtained, to understand what they represent.

In this study it was considered that the confusion matrix and the accuracy indexes were reliable, but in reality a human interpreted may have large difficulties in differentiating for example NVA from SHV, and this type of confusion between classes may influence greatly the accuracy indexes. In future analysis the uncertainty information present in the reference data, used to build the confusion matrixes, should be taken in consideration when comparing the results provided by the uncertainty and accuracy information.

The presented results also show that the spatialization of uncertainty may provide valuable information to the user. This will enable the identification of problematic regions, where the classification is probably less reliable. On the other hand, the spatial variation of uncertainty provides a great amount of information, which, if introduced in the classification itself, might be very useful.

References

[1] L. M. S. Gonçalves, *et al.*, Evaluation of soft possibilistic classifications with non-specificity uncertainty measures. *International Journal of Remote Sensing*. Accepted, 2009.

[2] A. Zhu, Measuring uncertainty in class assignment for natural resource maps under fuzzy logic. *Photogrammetric Engineering and Remote Sensing*, 63: 1195-1202, 1997.

[3] F. Wang, Improving remote sensing image analysis though fuzzy information representation. Photogrammetric Engineering and Remote Sensing, 56 (8): 1163-1169,1990.

[4] L. Bastin, Comparison of fuzzy c-means classification, linear mixture modeling and MLC probabilities as tools for unmixing coarse pixels. International Journal of Remote Sensing, 18 (17): 3629-3648, 1997.

[5] J. Zhang, G. M. Foody, Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches. *International Journal of Remote Sensing*, 22: 615-628, 2001.

[6] M. A. Ibrahim, *et al.*, Estimating and accommodating uncertainty through the soft classification of remote sensing data. *International Journal of Remote Sensing*, Vol.26: 2995-3007, 2005.

[7] G. M. Foody, Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, 17: 1317-1340, 1996.

[8] C. Ricotta, On possible measures for evaluating the degree of uncertainty of fuzzy thematic maps. *International Journal of Remote Sensing*, 26: 5573-5583, 2005.

[9] F. Maselli, *et al.*, Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. ISPRS Journal of Photogrammetry and Remote Sensing, 49 (2): 13-20,1994.

[10] W. G. Liu, *et al.*, Uncertainty and confidence in land cover classification using a hybrid classifier approach. Photogrammetric Engineering and Remote Sensing, 70 (8): 963-971, 2004.

[11] L. Bastin, P. Ficher, J. Wood, Visualizing uncertainty in multi-spectral remotely sensed imagery. Computers & Geosciences, 28:337-350, 2002.

[12] C. C. Fonte and W. Lodwick, Modelling the fuzzy spatial extent of geographical entities. In *Fuzzy Modeling with Spatial Information for Geographic Problems* (ed: Cobb, M., Petry, F., Robinson, V.). Springer-Verlag. 121-142. 2005.

[13] R. Yager, Measuring tranquility and anxiety in decision making: an application of fuzzy sets. *International Journal of General Systems*, 8: 139-146, 1982.

[14] M. Higashi, G. Klir, On measure of fuzziness and fuzzy complements. *International Journal of General Systems*, 8: 169-180, 1983.

[15] N. Pal, J. Bezdek, Quantifying different facets of fuzzy uncertainty. In *Fundamentals of Fuzzy Sets,* Dubois, D., Prade, H. (eds.), The Handbook of Fuzzy Sets Series, Kluwer Acad. Publ: 459-480., 2000.

[16] R. Yager, On the specificity of a possibility distribution. *Fuzzy Sets and Systems*, 50: 279-292, 1992.

ISBN: 978-989-95079-6-8

On interval and fuzzy calculations of economic uncertainty

Hans Schjær-Jacobsen

Copenhagen University College of Engineering
15 Lautrupvang, DK-2850 Ballerup, Denmark
Email: hsj@ihk.dk

Abstract—*This paper emphasizes that numerically correct calculation of economic uncertainty with intervals and fuzzy numbers requires implementation of global optimization techniques in contrast to straightforward application of interval arithmetic. In general, the latter approach produces incorrect results in the sense that the degree of uncertainty is severely overestimated with the risk of improper decision making. This is demonstrated by a simple case from managerial economics as well as a cost estimation case from a real life railway reconstruction project.*

Keywords—Economic uncertainty, fuzzy numbers, intervals, possibility, probability.

1 Introduction

Criticism has been raised towards probability theory as being a too normative framework to take all the aspects of uncertain judgement into account, Dubois and Prade [1]. In this paper we will focus on alternative methods of modelling of economic uncertainty like the interval representation and the fuzzy number representation.

The interval representation, Moore [2], [3] is particularly well suited to a situation where the knowledge of an uncertain parameter is limited to knowing its minimum and maximum value whereas nothing else is known. Based upon a mathematical theory of interval analysis this approach has shown to be useful in keeping track of worst and best cases in economic analyses and thus contribute to improved decision processes, see e.g. Schjær-Jacobsen [4], [5]. The introduction of fuzzy sets supports the concept of possibility rather than probability and translates natural language expressions into the mathematical formalism of possibility measures. It is generally recognized that possibility is distinct from probability. Probabilities can be interpreted as relative frequencies or, more generally, uncertain knowledge or belief of a statistical nature. In contrast, possibility relates to the degree of feasibility and ease of attainment or imprecise knowledge. Particularly, this paper will deal with some computational aspects of uncertainty representation by intervals and fuzzy numbers. It will be demonstrated that correct results are obtained in the general case by applying global optimization. It is the author's experience that these aspects are widely neglected, probably due to lack of communication and interaction between professional communities each one advocating either concept.

2 A simple case from managerial economics

In order to demonstrate the principles of numerical modelling of uncertainty by intervals and fuzzy numbers presented in this paper we introduce a simple yet instructive referential case from managerial economics. Consider a company selling one product at price p and quantity q into a market. For the sake of simplicity we may assume fixed cost to be zero. The turnover TR is

$$TR = TR(p,q) = p \cdot q, \tag{1}$$

and the variable cost VC is given by

$$VC = VC(q) = 20 \cdot q. \tag{2}$$

Then the profit π is

$$\pi = \pi(p,q) = TR(p,q) - VC(q) = p \cdot q - 20 \cdot q. \tag{3}$$

We want to find the price p* and quantity q* that gives the maximum profit π^*. It is seen from (3) that the profit is a monotone function of q and p which means that no maximum of π exists unless variables are constrained. If for example $p \leq 65$ and $q \leq 450$, then $p^* = 65$, $q^* = 450$, and $\pi^* = 20.250$.

Next we suppose that price and quantity are interdependent variables due to market conditions. Assuming that they are connected by the demand function

$$p = p(q) = 100 - 0{,}1 \cdot q \tag{4}$$

we get for the turnover

$$TR = TR(p) = -10 \cdot p^2 + 1.000 \cdot p, \tag{5}$$

the variable cost

$$VC = VC(p) = -200 \cdot p + 20.000, \tag{6}$$

and the profit margin

$$\pi = \pi(p) = -10 \cdot p^2 + 1.200 \cdot p - 20.000. \tag{7}$$

The profit of (7) is maximized by the optimality criterion that marginal cost MC be equal to marginal turnover MR which gives the results

ISBN: 978-989-95079-6-8

$$p^* = 60,\ q^* = 400, \text{ and } \pi^* = 16.000. \quad (8)$$

In the following, we shall refer to the above examples for explanations of computational details and special features.

3 Uncertainty modelling using intervals

2.1 Basics of interval analysis

Following Moore [3] and Caprani, Madsen, and Nielsen [6] we define a real interval number as an ordered pair [a; b] of real numbers with $a \le b$. It may also be defined as an ordinary set of real numbers x such that $a \le x \le b$, or

$$[a; b] = \{x \mid a \le x \le b\}. \quad (9)$$

If the basic arithmetic operations addition, subtraction, multiplication, and division are denoted by the symbol #, we can define operations on two intervals $\mathbf{I}_1 = [a_1; b_1]$ and $\mathbf{I}_2 = [a_2; b_2]$ based on the set-theoretic formulation:

$$\mathbf{I}_1 \# \mathbf{I}_2 = \{x \# y \mid a_1 \le x \le b_1,\ a_2 \le y \le b_2\}. \quad (10)$$

For basic operations on the intervals $\mathbf{I}_1$ and $\mathbf{I}_2$ we get the resulting interval $\mathbf{I} = [a; b]$ by the formulas

$$\begin{aligned}
&\mathbf{I} = \mathbf{I}_1 + \mathbf{I}_2 = [a_1+a_2;\ b_1+b_2],\\
&\mathbf{I} = \mathbf{I}_1 - \mathbf{I}_2 = [a_1-b_2;\ b_1-a_2],\\
&\mathbf{I} = \mathbf{I}_1 \cdot \mathbf{I}_2 = [\min(a_1 \cdot a_2,\ a_1 \cdot b_2,\ b_1 \cdot a_2,\ b_1 \cdot b_2);\\
&\qquad \max(a_1 \cdot a_2,\ a_1 \cdot b_2,\ b_1 \cdot a_2,\ b_1 \cdot b_2)], \quad (11)\\
&\mathbf{I} = \mathbf{I}_1 / \mathbf{I}_2 = [\min(a_1/a_2,\ a_1/b_2,\ b_1/a_2,\ b_1/b_2);\\
&\qquad \max(a_1/a_2,\ a_1/b_2,\ b_1/a_2,\ b_1/b_2)],\ 0 \notin [a_2; b_2].
\end{aligned}$$

It can be shown that the four basic interval operations are inclusion monotonic, commutative, and associative. However, the distributive rule is not valid in general. Instead, the so-called sub-distributivity holds, but only for addition and multiplication [6]

$$\mathbf{I}_1 \cdot (\mathbf{I}_2 + \mathbf{I}_3) \subseteq \mathbf{I}_1 \cdot \mathbf{I}_2 + \mathbf{I}_1 \cdot \mathbf{I}_3. \quad (12)$$

From a rational real valued function F of n real valued variables

$$F = F(x_1, x_2, \ldots, x_n) \quad (13)$$

we can create the interval extension function as an interval function $\mathbf{F}^I$ of n intervals

$$\mathbf{F}^I = \mathbf{F}^I(\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_n) \quad (14)$$

simply by replacing the real operators by interval operators and the real variables by intervals.

A rational function can be formulated in many ways whereas the same reformulations cannot be done for interval expressions due to the invalidity of the distributive rule. This implies that different formulations of a rational function will lead to different interval extension functions and thus to different interval results [6]. In the case of F being a monotonic function within the entire range of the input variables the minimum and maximum of $\mathbf{F}^I$ as an interval can simply be found among the function values F at the extreme points of the variables. In the general case of F being non-monotonic or variables appearing more then once, the calculation of $\mathbf{F}^I$ as an interval is non-trivial, which is demonstrated in the following example.

Example: Based on the real valued function $F = x \cdot (1 - x)$ the interval function $\mathbf{F}^I = \mathbf{I} \cdot (1 - \mathbf{I})$, $\mathbf{I} = [0; 1]$ is calculated. Straightforward application of formulas from (11) gives the result $\mathbf{F}^I = [0; 1]$ whereas the correct result is [0; 0,25].

In this paper the term "correct" is used to indicate the narrowest possible interval that can be calculated for an uncertain variable. Generally, to obtain this, iterative global optimization methods have to be used, see e.g. Hansen [7] and Kjøller *et al.* [8]. In order to obtain correct results (as in the above example) to an accuracy specified by the user, interval calculations in this paper are carried out using the Interval Solver 2000 program, Hyvönen and De Pascale [9], [10], as an add-in module to MS-Excel 2000. An overall absolute and relative precision of 10^{-6} has been applied.

Correct calculation of interval functions allows for strong statements about the uncertainties involved. *Firstly*, you can say that provided all uncertain input variables stay within their minimum and maximum values, the uncertain output function will stay within its minimum and maximum values. *Secondly*, the uncertain output function will not attain any value that is not a function value of some combination of the uncertain input values (within their minimum and maximum values).

2.2 Independent and interdependent variables

Independent variables using interval arithmetic

To calculate the uncertain profit we first use the formulas of interval arithmetic (11) by three different ways of calculation. As an example look at the independent and uncertain quantity and price

$$\mathbf{p} = [55; 65],\ \mathbf{q} = [350; 450], \quad (15)$$

Firstly, we use the turnover and variable cost as intermediate variables. By (1) we get for the uncertain turnover **TR** and by (2) for the uncertain variable cost **VC**

$$\begin{aligned}
&\mathbf{TR} = [55; 65] \cdot [350; 450] = [19.250; 29.250],\\
&\mathbf{VC} = 20 \cdot [350; 450] = [7.000; 9.000]. \quad (16)
\end{aligned}$$

Then by (3) we get for the uncertain profit

$$\boldsymbol{\pi} = \mathbf{TR} - \mathbf{VC} = [10.250; 22.250]. \quad (17)$$

However, the above calculation produces a too wide interval (17) for the profit. The reason for this is that in the expression (3) the variable q appears twice thus allowing the quantity q used to calculate **TR** to be different from the quantity used to calculate **VC**.

ISBN: 978-989-95079-6-8

Secondly, the rightmost form of (3) is used, giving

$$\boldsymbol{\pi} = [350; 450]\cdot[55; 65] - 20\cdot[350; 450] = [10.250; 22.250], \quad (18)$$

which is identical to (17) (and thus incorrect) because the arithmetic operations are identical.

Thirdly, (3) is rearranged before the interval calculations are carried out:

$$\pi = \pi(p,q) = q\cdot(p - 20). \quad (19)$$

We then get for the uncertain profit

$$\boldsymbol{\pi} = [350; 450]\cdot([55; 65] - 20) = [350; 450]\cdot[35; 45] = [12.250; 20.250], \quad (20)$$

which is a somewhat narrower interval than (17) and (18) because each variable is appearing only once in (19). Actually, (20) is the correct result. This can easily be verified simply by closer numerical inspection of the profit function (3) for various combinations of the variables p and q. The results are summarised in Table 1.

Independent variables using global optimization

Next, we calculate the uncertain profit by global optimization using Interval Solver 2000. With the same input variables (15) and intermediate variables **TR** and **VC** we get from (3)

$$\mathbf{TR} = [19.250; 29.250],\ \mathbf{VC} = [7.000; 9.000],\ \boldsymbol{\pi} = \mathbf{TR} - \mathbf{VC} = [10.250; 22.250], \quad (21)$$

which is identical to (17) and (18) for the same reason as mentioned above. With the same input variables and by way of formulas (18) and (19) we get the result

$$\mathbf{p} = [55; 65],\ \mathbf{q} = [350; 450],\ \boldsymbol{\pi} = [12.250; 20.250], \quad (22)$$

which is seen to be the correct result identical to (20). The results are summarised in Table 1.

Interdependent variables using interval arithmetic

To perform an uncertainty analysis around a given price p = 60 we set

$$\mathbf{p} = [55; 65]. \quad (23)$$

Obviously, since the quantity variable q can be eliminated by using the demand function (4), the number of variables is reduced from two to one, namely the price. Even in this simple case, as it turns out, the results obtained by interval arithmetic can be dramatically incorrect due to the fact that the profit function is now non-monotonic.

The profit is calculated in two different ways using interval arithmetic (11). *Firstly*, the profit is calculated by

Table 1. Summary of uncertainty analyses, independent price **p** = [55; 65] and quantity **q** = [350; 450]. Correct results shown in *italics*.

Calculation Method	$\boldsymbol{\pi} = \mathbf{TR} - \mathbf{VC}$	$\boldsymbol{\pi} = \mathbf{p}\cdot\mathbf{q} - 20\cdot\mathbf{q}$	$\boldsymbol{\pi} = \mathbf{q}\cdot(\mathbf{p} - 20)$
Interval Arithmetic	[10.250; 22.250]	[10.250; 22.250]	*[12.250; 20.250]*
Global Optimization	[10.250; 22.250]	*[12.250; 20.250]*	*[12.250; 20.250]*

intermediate variables **TR** and **VC** according to (5) and (6), respectively

$$\mathbf{TR} = -10\cdot[55; 65]^2 + 1.000\cdot[55; 65] = [12.750; 34.750] \quad (24)$$

and

$$\mathbf{VC} = -200\cdot[55; 65] + 20.000 = [7.000; 9.000], \quad (25)$$

which gives

$$\boldsymbol{\pi} = \mathbf{TR} - \mathbf{VC} = [3.750; 27.750]. \quad (26)$$

Secondly, the profit is calculated by the above formula

$$\boldsymbol{\pi} = -10\cdot\mathbf{p}^2 + 1.200\cdot\mathbf{p} - 20.000 \quad (27)$$

yielding

$$\boldsymbol{\pi} = -10\cdot[55; 65]^2 + 1.200\cdot[55; 65] - 20.000 = [3.750; 27.750]. \quad (28)$$

The results found above are identical but way out of order, since the correct result of the uncertain profit is [15.750; 16.000]. This is easily verified by closer numerical inspection of (7) with varying p.

Interdependent variables using global optimization

Global optimization is used to calculate the uncertain profit by means of (26) yielding [13.750; 17.750] and (27) yielding [15.750; 16.000]. It may be observed that correct results are produced in the latter case but not in the former, due to usage of intermediate variables **TR** and **VC**, which allows for a too wide resulting interval. The results are shown in Table 2.

Table 2. Summary of uncertainty analyses, interdependent price **p** = [55; 65] and quantity $\mathbf{q} = 1.000 - 10\cdot\mathbf{p}$. Correct result shown in *italics*.

Calculation Method	$\boldsymbol{\pi} = \mathbf{TR} - \mathbf{VC}$	$\boldsymbol{\pi} = -10\cdot\mathbf{p}^2 + 1.200\cdot\mathbf{p} - 20.000$
Interval Arithmetic	[3.750; 27.750]	[3.750; 27.750]
Global Optimization	[13.750; 17.750]	*[15.750; 16.000]*

ISBN: 978-989-95079-6-8

The differences in Tables 1 and 2 between correct intervals and the more wide intervals are considerable. When an uncertain variable is appearing more than once, interval arithmetic will usually produce too wide results and global optimization must to be used. With non-monotonic functions, global optimization is necessary. In all cases, intermediate variables should be avoided, since both interval arithmetic and global optimization will produce incorrect results.

4 Uncertainty modelling using fuzzy numbers

4.1 Fuzzy numbers and intervals

A fuzzy set A in X where X is a space of points (objects) with a generic element of X denoted by x, i.e. X = {x}, is characterized by a membership function $f_A(x)$ which associates with each point in X a real number in the interval [0; 1]. The value of the membership function $f_A(x)$ at x represents the "grade of membership" of x in A. Thus the closer the value of $f_A(x)$ to unity, the higher the grade of membership of x in A. Note that when A is an ordinary set, i.e. non-fuzzy, the membership function can take only two values 0 and 1. In other words, a fuzzy set is a set of ordered pairs (x, $f_A(x)$)

$$A = \{(x, f_A(x)) \mid x \in X\}. \tag{34}$$

It is also useful to define the ordinary (non-fuzzy) set A_α as the α-cut of A:

$$A_\alpha = \{x \in X \mid f_A(x) \geq \alpha, 0 \leq \alpha \leq 1\}. \tag{35}$$

In this paper we are mainly interested in the concept of fuzzy numbers as a means of representing uncertain or fuzzy information, Dubois and Prade [11], [12]. In addition to the simplest fuzzy number, namely the interval, we also make use of the triangular fuzzy number, Chiu and Park [13], [a; c; b] where a ≤ c ≤ b, that can be defined by its membership function:

$$\begin{aligned} f(x) &= (x-a)/(c-a), && a \leq x \leq c, \\ &= (b-x)/(b-c), && c \leq x \leq b, \\ &= 0, && \text{otherwise.} \end{aligned} \tag{36}$$

Mathematical operations on triangular fuzzy numbers can be facilitated by introducing the left L(α) and right R(α) representation of a fuzzy triangular number $\mathbf{F}^T$, refer to the α-cut (35):

$$\begin{aligned} &\mathbf{F}^T = [L(\alpha); R(\alpha)], \text{ where} \\ &L(\alpha) = a + (c-a)\alpha \text{ and } R(\alpha) = b + (c-b)\alpha, \\ &\alpha \in [0, 1]. \end{aligned} \tag{37}$$

Observe that in this notation a fuzzy number is written as an interval with upper and lower bounds depending on α. This means that addition, subtraction, multiplication, and division can be carried out by using interval methods for all values of α. Likewise, for any triangular function, the resulting triangular functional values can be calculated and represented by L and R functions using interval methods for all values of α.

Example: Based on the real valued function F = x·(1 – x) calculate the corresponding fuzzy function with triangular argument [0; 0,5; 1]. Correct results have been calculated with global optimization and are shown in Table 3. From the results in Table 3 it can be seen that the function F has a maximum of 0,250 at x = 0,5 corresponding to α = 1. Also note that the R(α) function has been correctly calculated to 0,250 for all values of α.

To obtain simpler representations and reduce the number of calculations, triple and quadruple representations of fuzzy variables corresponding to α-cuts 0 and 1 in (35) may be used. In the above example the result then is [0; 0,25; 0,25], where the extreme function values is obtained by global optimization on the interval [0; 1] and the interior point is obtained by conventional calculation at x = 0,5.

4.2 Simple case with triangular fuzzy numbers

Next we calculate the simple case from Section 2 using profit function (3) with independent triangular fuzzy input parameters **p** = [55; 60; 65], **q** = [350; 400; 450] defined by (36). The resulting membership function is shown in Table 4 for different values of α. It is easily seen that the value α = 0 corresponds to the correct profit interval previously found and α = 1 corresponds to the single point calculation of the profit. This is one of the important features of the triangular fuzzy number representation of uncertainty: It is easily communicated and understood that the ordinary single point calculation is extended to an interval around the single point representing the uncertain value of the input variable. The resulting triangular fuzzy profit is interpreted as follows: With the given uncertain input variables, the most possible value of the profit is 16.000 and profits outside the interval [12.250; 20.250] are impossible.

Calculations have also been carried out with interdependent triangular fuzzy input price parameter corresponding to **p** = [55; 60; 65] and the profit function (7), the results are shown in Table 5. For all values of α the correct maximum profit of 16.000 has been found. No profit values outside the interval [15.750; 16.000] are possible given the uncertain price input variable **p**.

Table 3. Fuzzy extension of x·(1-x) calculated with triangular argument [0; 0,5; 1].

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
L(α)	0,000	0,047	0,090	0,127	0,160	0,188	0,210	0,227	0,240	0,248	0,250
R(α)	0,250	0,250	0,250	0,250	0,250	0,250	0,250	0,250	0,250	0,250	0,250

ISBN: 978-989-95079-6-8

Table 4. Uncertain profit (3) by global optimization, triangular fuzzy input **p** = [55; 60; 65], **q** = [350; 400; 450].

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
L(α)	12.250	12.603	12.960	13.323	13.690	14.063	14.440	14.823	15.210	15.603	16.000
R(α)	20.250	19.803	19.360	18.923	18.490	18.063	17.640	17.223	16.810	16.403	16.000

Table 5. Uncertain profit (7) by global optimization, triangular fuzzy input **p** = [55; 60; 65].

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
L(α)	15.750	15.797	15.840	15.877	15.910	15.937	15.960	15.977	15.990	15.977	16.000
R(α)	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000	16.000

Table 6. Cost estimation for railway reconstruction case by triangular fuzzy numbers.

Var	Code	Item	[a; c; b]
	0,00	Management and specs.	[1.732; 1.780; 1.884]
$\mathbf{X_1}$	0,10	Project management	[524; 540; 575]
$\mathbf{X_2}$	0,20	Construction management etc.	[975; 1.000; 1.050]
$\mathbf{X_3}$	0,30	Design specifications etc.	[233; 240; 259]
$\mathbf{X_4}$	10,00	Environmental and soil eng.	[864; 888; 950]
$\mathbf{X_5}$	20,00	Traffic tasks	[48; 50; 53]
	30,00	Renewal of tracks	[8.905; 9.190; 9.664]
$\mathbf{X_6}$	30,10	New outbound main track	[975; 1.000; 1.050]
$\mathbf{X_7}$	30,20	Track renewal at platform 3/5	[5.432; 5.600; 5.880]
$\mathbf{X_8}$	30,30	New platform edge	[1.533; 1.580; 1.643]
$\mathbf{X_9}$	30,40	Track renewal depot, West	[285; 300; 321]
$\mathbf{X_{10}}$	30,50	Track layout design	[682; 710; 770]
$\mathbf{X_{11}}$	40,00	Platform and station	[538; 560; 602]
$\mathbf{X_{12}}$	50,00	Safety and signal installations	[5.035; 5.245; 5.586]
	60,00	Informatics incl. power supply	[2.374; 2.417; 2.626]
$\mathbf{X_{13}}$	60,10	Phase 2-4	[78; 80; 86]
$\mathbf{X_{14}}$	60,20	Sub project management	[249; 259; 275]
$\mathbf{X_{15}}$	60,30	Passenger information	[1.009; 1.030; 1.123]
$\mathbf{X_{16}}$	60,40	Electrical power supply	[1.038; 1.048; 1.142]
	70,00	Overhead line incl. pylons	[3.507; 3.624; 3.787]
$\mathbf{X_{17}}$	70,10	Overhead cables	[3.021; 3.122; 3.262]
$\mathbf{X_{18}}$	70,20	Layout and planning	[486; 502; 525]
$\mathbf{Y_1}$		Total cost before corrections	[23.003; 23.754; 25.152]
$\mathbf{X_{19}}$	A	Internal decision process	[1,006; 1,032; 1,098]
$\mathbf{X_{20}}$	B	Design specifications etc.	[1,009; 1,040; 1,100]
$\mathbf{X_{21}}$	C	Working process	[1,021; 1,042; 1,084]
Y		Total cost after corrections	[23.842; 26.565; 32.930]

Table 7. Total cost after corrections for railway reconstruction case by triangular fuzzy numbers.

α	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
L(α)	23.842	24.108	24.353	24.624	24.875	25.148	25.480	25.734	26.020	26.276	26.565
R(α)	32.930	32.244	31.592	30.895	30.269	29.626	28.983	28.387	27.743	27.161	26.565

ISBN: 978-989-95079-6-8

5 A railway reconstruction project

Consider the case of estimating the total cost incurred by a railway reconstruction project described by independent fuzzy input variables, namely 18 cost items $\mathbf{X}_1$,..., $\mathbf{X}_{18}$ and 3 correction factors $\mathbf{X}_{19}$,..., $\mathbf{X}_{21}$. This case has been treated previously by using the concept of imprecise stochastic variables, Schjær-Jacobsen [14]. The correction factors are introduced in order to account for overall influences not accounted for by the individual cost items. The total cost before corrections is the sum

$$\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_{18}. \quad (38)$$

The total cost after corrections $\mathbf{Y} = \mathbf{Y}(\mathbf{X})$ is a function of all 21 uncertain variables

$$\mathbf{Y} = (\mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_{18}) \cdot \mathbf{X}_{19} \cdot \mathbf{X}_{20} \cdot \mathbf{X}_{21}. \quad (39)$$

The uncertain input variables should be estimated by railway experts with relevant project experience. Subsequently, the total cost $\mathbf{Y}$ is calculated by means of global optimization. The total cost estimation results are shown in Table 6 in the case of triangular fuzzy input variables. The term "code" refers to the cost structure hierarchy. For example the item carrying code 0,00 is the sum of cost items at lower levels carrying the codes 0,10, 0,20, and 0,30. Thus the latter are input variables whereas the former is an output variable. The cost item carrying the code 10,00 is an input variable itself because it does not have cost items at lower levels involved. In Table 7 the membership function of total cost $\mathbf{Y}$ after corrections is shown.

For $\alpha = 1$ the single point calculation of 26.565 is shown corresponding to an ordinary calculation without uncertainty. This value is the most possible one and is attained when all input variables are attaining their most possible values. For $\alpha = 0$ the total cost after corrections is within the interval [23.842; 32.930] and any value outside this interval is impossible considering the uncertain input variables.

It should be mentioned here that comparative studies of approaches involving probabilistic methods (including Monte Carlo simulation), trapezoidal fuzzy numbers, and imprecise stochastic variables have been made in [14] and [15] and is further investigated in a research project concerning economic uncertainty in mega projects.

6 Conclusion

In order to calculate correct results with intervals and fuzzy numbers global optimization methods must be implemented in contrast to straightforward application of interval arithmetic. However, the experimental calculations reported in the paper also show that further care should be taken not to introduce intermediate variables, which may also result in excess width intervals. When calculating fuzzy extensions of non-monotonic functions global optimization must be used in order to produce correct results. Otherwise, additional and unnecessary uncertainty will arise eventually leading to wrong decisions.

The results of the paper indicate that correct calculations of intervals (and fuzzy number membership functions) allow for rather strong statements pertaining to economic uncertainties. For example, it can be said that provided all uncertain input variables stay within their limits the uncertain output variables will stay within their limits. Obviously, it is an advantage of the interval and triangular (and trapezoidal) fuzzy representation of uncertainty that the meaning is easily communicated and understood.

References

[1] D. Dubois and H. Prade. *Possibilty theory – An approach to computerized processing of uncertainty.* New York: Plenum Press, 1988.

[2] R.E. Moore. *Interval arithmetic and automatic error analysis in digital computing.* Ph.D. dissertation, Stanford University, USA, 1962.

[3] R.E. Moore. *Interval analysis.* New Jersey: Prentice-Hall, 1966.

[4] H. Schjær-Jacobsen. A new method for evaluating worst- and best-case economic consequences of technological development. *International Journal of Production Economics,* 46-47: 241-250, 1996.

[5] H. Schjær-Jacobsen. Handling economic risks and uncertainties of production technology investments. *Proceedings of the 14th International Conference on Production Research, Osaka, Japan,* 1: 346-349, 1997.

[6] O. Caprani, K. Madsen, H.B. Nielsen. *Introduction to interval analysis*. Lecture notes, Department of Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark, November 20, 2002.

[7] E. Hansen. *Global optimization using interval analysis*. Marcel Dekker, New York, USA, 1992.

[8] S. Kjøller, P. Kozine, K. Madsen, O. Stauning. *Non-linear global optimization using interval arithmetic and constraint propagation*. In: A. Törn and J. Zilinskas (Eds.), Models and algorithms for global optimization, Springer Verlag, 2007.

[9] E. Hyvönen and S. de Pascale. A new basis for spreadsheet computing: Interval Solver for Microsoft Excel. *Proceedings of the 11th Innovative Applications of Artificial Intelligence (IAAI-99)*, AAAI Press, Menlo Park, California, 1999.

[10] E. Hyvönen and S. de Pascale. *Interval Solver 2000 for Microsoft Excel*. User's Guide, Version 4.0. Program Release 4.0.0.2, Delisoft Ltd., Helsinki, Finland, 2000.

[11] D. Dubois and H. Prade. Operations on fuzzy numbers. *International Journal of System Science*, 9: 613-626, 1978.

[12] D. Dubois and H. Prade. Fuzzy real algebra: Some results. *Fuzzy Sets and Systems*, 2: 327-348, 1979.

[13] C.U. Chiu and C.S. Park. Fuzzy cash flow analysis using present worth criterion. *The Engineering Economist*, 39(2): 113-138, 1994.

[14] H. Schjær-Jacobsen. A unified approach to variability and imprecision in modelling of economic uncertainty. *Proceedings of the 15th International Working Seminar on Production Economics*, 1: 437-447, Innsbruck, Austria, March 3-7, 2008.

[15] H. Schjær-Jacobsen. Modeling of economic uncertainty, *Fuzzy Economic Review*, IX(2): 49-73, 2004.

ISBN: 978-989-95079-6-8

A method for locating the iliac crests based on the fuzzy Hough transform

Sergio Jaime-Castillo Juan Miguel Medina Antonio Garrido

Department of Computer Science and Artificial Intelligence, University of Granada
C/ Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain
Email: {sjaime,medina,a.garrido}@decsai.ugr.es

Abstract— *In this paper we present a method based on the fuzzy Hough transform to locate the iliac crests in posteroanterior radiographs. Firstly, a Canny edge detector is applied to the X-ray images in order to obtain an edge map. Then, the fuzzy Hough transform is used in combination with several constraints extracted from our domain knowledge. At the bottom of the radiograph we search for two circular arcs, as we can represent the shape of the iliac crests by means of these curves. Locating the iliac crests can be useful for a later processing of the X-rays in order to obtain a fully automated segmentation of the whole spine.*

Keywords— Computer-aided diagnosis, fuzzy Hough transform, medical imaging, X-rays.

1 Introduction

Computer-aided diagnosis is increasing its importance in healthcare. This kind of systems can help physicians to perform routine tasks that require a long training and a great amount of time, and can provide more accurate and consistent measures that may improve the diagnostic process.

Modern medical imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) provide images of high quality which can help physicians enormously in diagnostic process. However, these techniques are usually expensive and give patients high doses of radiation. Therefore, they are not suitable for all diagnoses and, although X-rays do not provide as much quality as the other ones, they are widely used in diagnosis of many diseases.

Much research has been done to segment anatomical structures in CT and MRI. Nevertheless, X-rays have not been as widely studied as the other ones due to its lower quality. In our work we use X-ray images because they are extensively used in diagnosing scoliosis, which is the problem we are focused on.

Scoliosis is a three-dimensional deformation of the spine that produces vertebral rotation and crushing, and lateral curvature. It is typically classified as congenital (caused by vertebral anomalies present at birth), idiopathic (sub-classified as infantile, juvenile, adolescent or adult according to the time when the onset occurred) or as a secondary symptom of another condition, such as cerebral palsy, spinal muscular atrophy or due to a physical trauma.

Diagnosis of scoliosis is performed by calculating Cobb angle ([1]) manually on X-rays. Automatic measure of Cobb angle could improve diagnosis since it can provide more reliable and consistent measures. In order to assess Cobb angle automatically we need to segment the spine accurately and, as we will see afterwards, the proposed method can be useful to do so.

The iliac crest is the long and curved upper border of the wing of the ilium. According to a study performed by Render in [2], the line connecting the iliac crests intersects the spine at the level of the fourth lumbar vertebra or the interspace between the fourth and the fifth lumbar vertebra in 78.5% of cases, while the point of intersection was lower than the fourth lumbar vertebra in 17.8% and at the level of the interspace between the third and the fourth lumbar vertebra in 3.7%. We are interested in locating these anatomical structures as we can use them as a reference to initialise a segmentation algorithm that is intended to segment the whole spine. In the same way, this reference point will be useful to label the vertebrae found by the segmentation algorithm, since we can use the results of Render.

In this work, we have used full-length standing posteroanterior X-rays. The pictures were acquired from a PACS (Picture Archiving and Communication System) as DICOM (Digital Imaging and COmmunication in Medicine) files with 5700×2400 pixels of resolution. The problem with this kind of images is the presence of noise and distortions due to its acquisition process, that makes difficult to detect and identify the anatomical structures present in the X-rays accurately. Other important problems are poor contrast and low visibility of some details in the images. In order to deal with these problems, first of all we have to preprocess the images to obtain new ones with less quantity of noise and, as far as possible, better visibility than the initial ones.

The rest of the paper is organised as follows: in section 2 we describe the foundations and the operation of the fuzzy Hough transform. In section 3 we present our method for locating the iliac crests in posteroanterior radiographs and, in section 4, we discuss the results achieved by it. Finally, in section 5, we comment some aspects of our future work related to this paper.

2 The fuzzy Hough transform

The Hough transform (HT) was originally proposed as a technique to detect straight lines by Hough in 1962 ([3]). It was extended to detect general curves with known analytical expressions by Duda and Hart in [4] and, later, Ballard proposed a new extension in [5] to detect any curve although its expression was not known. In this work we use the fuzzy Hough transform (FHT), another extension of the original algorithm developed by Han *et al.* in [6].

A straight line can be represented as $\rho = x\cos\theta + y\sin\theta$, where θ is the angle of the vector from the origin to the closest point of the line, ρ is the lowest distance between the line and the origin and (x, y) are the coordinates of a point in the image space. In that way, a line is defined by these two parameters ρ

ISBN: 978-989-95079-6-8

and θ. Similarly, a circle and other curves can be represented by its parameters. The main goal of the algorithm is to detect shapes by checking which points in the edge map belong to the curve defined by a certain point in the parameter space. This is done by mapping the image space to the parameter space, where each point defines a different curve. By doing so we simplify the problem to locate a peak in an accumulator array.

In the original HT, the value at a point of the parameter space is the number of points in the image space that occur strictly on the shape defined by this point of the parameter space. There is no difference between a point being far from the ideal line or being just a little off it. Because of this, with the HT is difficult to detect approximate shapes that may appear in noisy environments due to the noise itself, or due to the not accurate edge detection performed by the feature detector. Fuzzy theory can help to deal with the detection of those approximate shapes.

We have assumed that we can represent the shape of the iliac crest by two circular arcs. However, the iliac crest is not exactly a circular arc, although it is close to. Because of this, the results obtained by applying the conventional HT to the X-rays are not accurate, so we propose the use of the FHT in order to achieve better results. In section 4 we make a comparison between both techniques based on the results achieved by them.

The main idea in the FHT is that each edge point contributes more or less to one point in the parameter space depending on its distance from the ideal curve defined by this point. To do this, once the accumulator is computed, Han *et al.* proposed to convolve it with the membership function

$$g(d) = \begin{cases} ke^{-d^2/\sigma^2} & if\ d < R \\ 0 & otherwise \end{cases} . \qquad (1)$$

As they proved, when the membership function of fuzzy points is defined as a circle the transforms of those points lying inside it is bounded in the parameter space by r, which is the radius of that membership function. That is true for straight lines and circles and, as we are considering circular arcs in such a way that they are particular instances of circles with some restrictions, we shall use this smoothing function.

In (1) k, σ and R are constants that are chosen empirically and d is the distance between the two points of the image considered. In our case, they are set as $k = 1$, $R = 3$ and $\sigma = 1.3$ in the experiments.

3 The proposed method to locate the iliac crest

In this section we present our method for automated detection of the iliac crest, based upon the fuzzy Hough transform. The method starts with the acquisition of the X-rays and finishes with the location of the iliac crests. Fig. 1 shows the operation of the method. It involves basically the search for circular arcs at the bottom of the radiograph, where the iliac crests are supposed to be, by means of the FHT.

As we have mentioned beforehand, a preprocessing stage is necessary to deal with noise and distortions present in the images. In this work we have used a combination of unsharp masking (UM), adaptive histogram equalisation and median filter to achieve images more suitable to be analysed by a computer vision system.

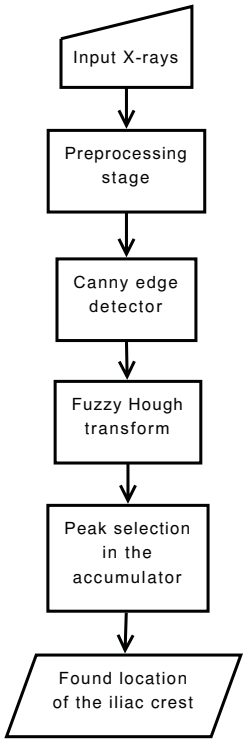


Figure 1: Flowchart of the proposed method to detect the iliac crest

Unsharp masking intends to improve the visibility of images by emphasising smoothed edges while the noise components are suppressed. The effect of the algorithm onto the images is edge sharpening, but usually it tends to reduce its contrast. Because of this, after the unsharp masking procedure we perform adaptive histogram equalisation to enhance it.

In unsharp masking, the high-frequency components of the image are extracted and then are added to the original image. First of all, a slightly blurred version I_{blur} of the image I_{orig} is obtained by convolving it with a Gaussian kernel and then is subtracted from I_{orig} to obtain I_{edge}. After that, I_{edge} is multiplied by a sharpness factor and the result, which contains the enhanced high-frequency components of the image, is then added to I_{orig}.

In our case, the unsharp masking is done by using

$$I_{sharp}(x,y) = I_{orig}(x,y) + 6 \cdot (I_{orig}(x,y) - I_{blur}(x,y)). \quad (2)$$

In (2) $I_{orig}(x,y)$ represents the original image and $I_{blur}(x,y)$ is a blurred version of the original image obtained by convolving $I_{orig}(x,y)$ with a Gaussian kernel of size 5×5 pixels and a variance of 5.

After the unsharp masking, we apply adaptive histogram equalisation to the image in order to enhance the contrast of the resultant image. Adaptive histogram equalisation divides the image into several regions and then distributes pixels values in each region of the image uniformly throughout the grey scale. It is performed in order to achieve a higher contrast for the images by means of the *adaptivehisteq* MATLAB function.

Finally, a median filter of size 11×11 pixels is applied to the image with the *medfilt2* MATLAB function to reduce the presence of impulse noise, naturally present in the X-rays.

ISBN: 978-989-95079-6-8

Once the image has been preprocessed, Canny edge detector is applied to it in order to obtain an edge map that will be used as a starting point for the FHT in the locating procedure. The Canny edge detector is one of the most widely used algorithms among all the features detectors. It was proposed by Canny in [7] in 1986. This method consists of several stages: first of all noise reduction is performed by means of a Gaussian filter; then it computes the gradient in every point of the image and discard those ones whose value is under a certain threshold; next, the pixels whose gradient magnitude is not maximum in the gradient direction are suppressed; finally a hysteresis thresholding is performed by using two thresholds to suppress spurious noisy pixels.

We have used the implementation available in MATLAB by means of the $edge$ function, and we have used the default parameters of the function.

In order to apply the FHT we need to identify the parameters that define the considered curve and the equation that relates them. A circular arc is defined by five parameters: the two coordinates of its centre, its radius, its orientation and its angular size. We could use the generalised Hough transform proposed in [5], but this would imply that we would have to use a five-dimensional accumulator. By doing so, the memory and computational requirements of the algorithm would rise enormously. Nevertheless, if we can fix one or more parameters, the requirements will remain reasonable.

In this case, after studying the characteristics of the iliac crests showed in the X-rays, we can assume that the radius of the circular arcs we use to represent them has to be between fifteen and forty pixels. We can also assume that an appropriate orientation for the circular arc could be 0 rad and a suitable angular size for it could be π rad. These assumptions we have made permit us to simplify the original problem reducing its computational requirements, which is one of the major drawbacks of the Hough transform. Instead of using a five-dimensional accumulator we can use a less computational expensive three-dimensional array.

In this work we have used a parametric representation of the circular arc, which allow us to use the parametric equation of the circle by restricting the permitted values for the parametric coordinate between 0 rad and π rad, instead of between 0 rad and 2π rad that we would use if we were searching for a whole circle. Moreover, the parametric representation facilitates the computation of the positions of the accumulator that have to be incremented. Thus, the equation used to describe the circular arc is

$$\begin{aligned} x(t) &= R \cdot \cos(t) + a \\ y(t) &= R \cdot \sin(t) + b \\ t &\in [0, \pi] \end{aligned} \tag{3}$$

where t is the parametric coordinate, R is the radius of the circular arc and (a, b) are the coordinates of the centre of the arc.

Next, we introduce the formal description of our implementation of the FHT:

1. Set the initial values of the accumulator A to 0.
2. For each edge point (x, y), do:
   (a) For each radius R between 15 and 40 pixels, do:
      i. For every $t \in [0, \pi]$, do:
         A. $a = x - R \cdot \cos(t)$
         B. $b = y - R \cdot \sin(t)$
         C. $A[a, b, R] = A[a, b, R] + 1$
3. Convolve the accumulator A with the smoothing function (1).

At this point we can apply the FHT, as we have just described it, to the edge map obtained in the previous step. By doing so, we get an accumulator array where the maximum values represent the most suitable arcs for the given image. We shall use the one hundred biggest ones of those peaks.

After the use of the FHT, several constraints coming from the domain knowledge are applied to the peaks found in the accumulator array to improve the performance of the algorithm.

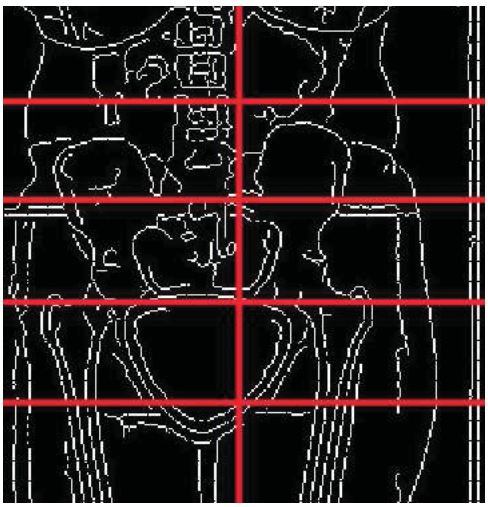


Figure 3: Canny edge map divided into regions to seek local maxima

We divide the lower part of the radiograph into two parts, and we seek the left iliac crest in the left half and the right iliac crest in the right half. Let us consider five regions of equal size within each considered division of the radiograph as we can see in Fig. 3. The search is performed separately in the left and in the right parts of the X-ray. In each region we search for the maximum local peak and record the parameters of the correspondent circular arc. Thus, we have ten local maxima, five for the left part and five for the right part, which represent ten circular arcs that are suitable locations for the iliac crests. We can make twenty five possible couples with those peaks by choosing one from the left and another one from the right.

We can assume that the locations of the left and the right iliac crests are, approximately, at the same height in the X-ray. Let us consider a tolerance factor of 15 pixels. In that case, we can discard those couples of peaks which do not satisfy this criterion. Finally, for the couples of peaks which satisfy the criterion, we add its values and we choose the one with the largest value. Thus, at the end we have a couple of peaks which corresponds to the location of the iliac crests obtained by the method. This process is called "peak selection in the accumulator" in Fig. 1.

We performed several experiments in order to choose the most appropriate values for the tolerance factor and the number of regions in terms of its performance.

ISBN: 978-989-95079-6-8

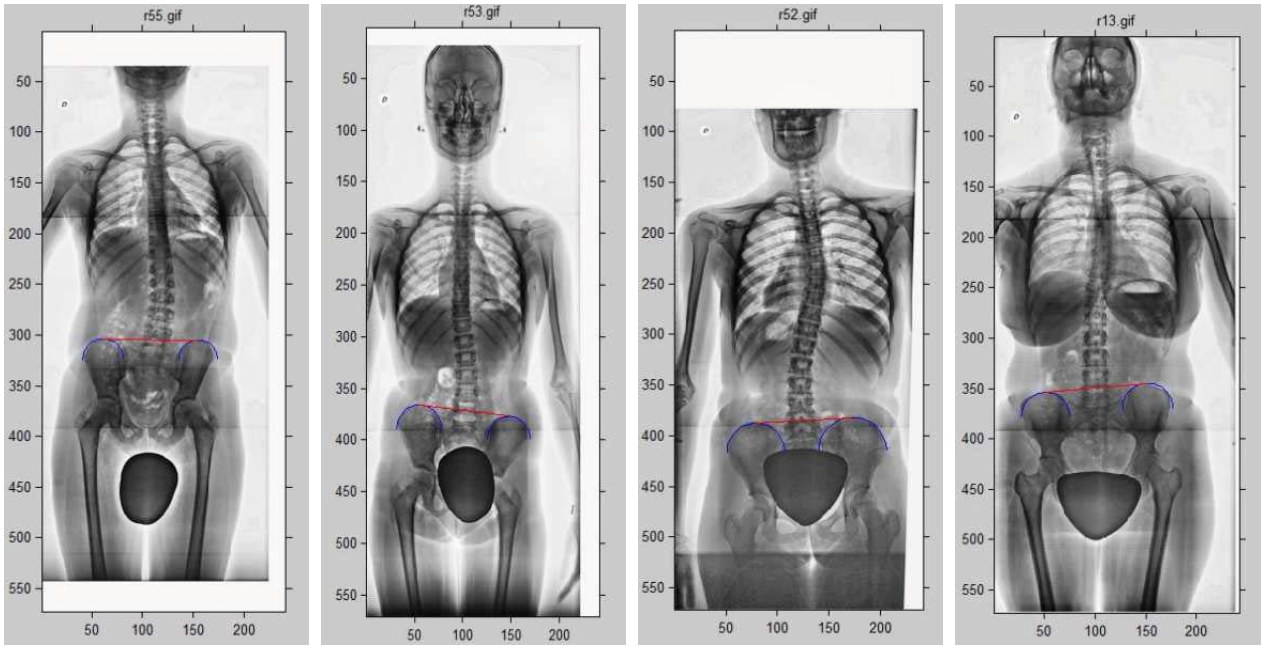
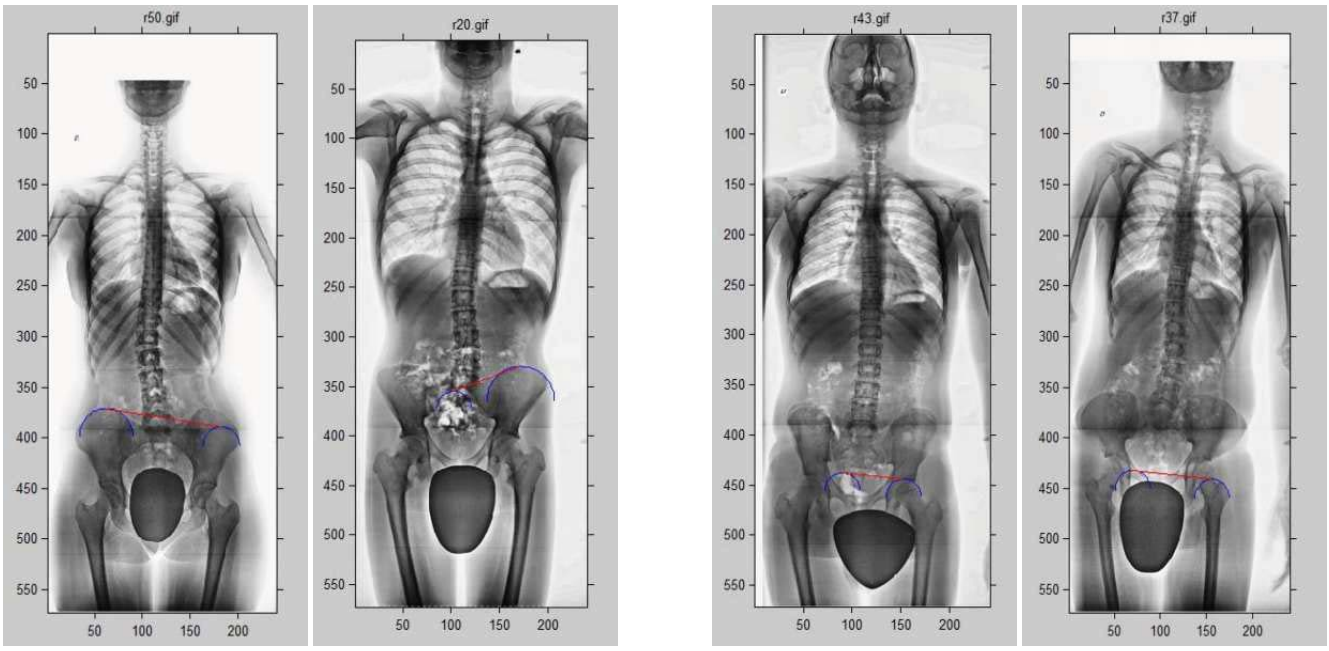


Figure 2: Some examples of the found locations of the iliac crests



(a) Only one iliac crest accurately located

(b) No iliac crest accurately located

Figure 4: Some examples which do not locate accurately the iliac crests

4 Results

In this section we discuss the results achieved by our method. We used a database with 45 X-ray images obtained from a PACS system in a local hospital. The images were in DICOM format with 5700×2400 pixels of resolution. Almost all the images were from patients suffering from scoliosis because we are focused on the study of this pathology by means of digital images. The resolution of the images was reduced to a tenth of its original one to reduce the time of processing before starting the location procedure. Experiments performed with several images suggest us that reduction in image size does not have a significant influence on the accuracy of our method.

It is important to remark that we are not interested in extract the exact shape of the iliac crest, but in the line drawn from the highest point of the left arc to the highest one of the right arc that, as we said beforehand, in an large number of cases will pass through the fourth lumbar vertebra or the interspace between the fourth and the fifth lumbar vertebrae. Therefore, accurate shape delimitation for the iliac crests is not essential.

The experiments we performed showed that in 28 out of 45 cases iliac crests were properly located, while in 3 images

ISBN: 978-989-95079-6-8

the method found the correct location for one iliac crest but failed to locate the another one. In Fig. 2 we can see some of the satisfactorily solved cases after performing the experiments. Moreover, Fig. 4 shows some cases that are not properly solved by our method.

The most important reasons that make our method fails in those cases are two: on the one hand, the presence of distortions in the area of the radiograph were the iliac crest is located, due to the position of the intestines which partially occlude it; on the other hand, the shape of some iliac crests is not completely round and this causes that our method fails because other rounder objects in the image, such as the femur heads, are located. In those cases other constraints should be included in our algorithm.

We also ran the same experiments using the method described here but using the standard HT instead of the FHT. Results obtained were satisfactory in 17 out of 45 images, whereas they were satisfactory for only one iliac crest in 14 out of 45. It is clear then that FHT can achieve better results in those images with a great presence of noise or in which the shapes sought are not clear.

5 Concluding remarks and future work

In this paper we have proposed a method for locating the iliac crests in X-ray images based on the fuzzy Hough transform. We can use the found location of the iliac crests to identify the fourth lumbar vertebra by drawing a line between the two crests and, that will be useful to initialise another algorithm designed to segment the spine. We are still working on improving accuracy and success rate of our method, as well as its efficiency, by incorporating additional constraints into the method.

It is important to remark that results achieved by means of FHT are clearly better than those obtained by standard HT. As shape of iliac crests is largely variable, it is difficult to find them by using a rigid template as HT does. For that reason FHT, which relaxes shape constraints by means of fuzzy logic, is more suitable in this problem.

Acknowledgement

This work has been partially supported by the Spanish "Ministerio de Ciencia y Tecnologa" (MCYT) under grants TIN2006-07262, and the "Consejería de Innovación, Ciencia y Empresa de Andalucía" (Spain) under research project P06-TIC-01570.

References

[1] J.R. Cobb. Outline for the study of scoliosis. *American Academy of Orthopaedic Surgeons, Instructional Course Lectures*, 5:261–275, 1948.

[2] C.A. Render. The reproducibility of the iliac crest as a marker of lumbar spine level. *Anaesthesia*, 51(11):1070–1071, 1996.

[3] P.V. Hough. Methods and means for recognizing complex pattern, 1962. US Patent 3,069,654.

[4] R.O. Duda and P.E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[5] D.H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[6] J.H. Han, L.T. Kóczy, and T. Poston. Fuzzy hough transform. *Pattern Recognition Letters*, 15(7):649–658, 1994.

[7] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

ISBN: 978-989-95079-6-8

Membership dependent stability analysis of TS fuzzy controlled systems using coupling attenuation

Jean-Yves Dieulot[1] Nesrine Elfelly[2]
[1]LAGIS UMR CNRS 8146, University of Lille
[2]LAGIS UMR CNRS 8146, Ecole Centrale de Lille
Cité Scientifique, 59650 Villeneuve d'Ascq, France
Email: dieulot@polytech-lille.fr, nesrine.elfelly@ed.univ-lille1.fr

Abstract— *The use of Linear Matrix Inequalities and Common Quadratic Lyapunov Functions is a powerful and commonplace tool for Takagi-Sugeno fuzzy controlled system analysis and synthesis. However, in practice, few practical and performing results are available when the subsystems exhibit different input matrices, because of the strong coupling between the subsystems/subcontrollers. In this paper, this coupling is demonstrated and a method is proposed which allows to synthesize, for a number of subsystems higher than 2, the local gains of a Parallel Distributed Controller. It is shown that the controller gains depend on the values of the input matrices and of the membership functions, and are thus able to relax classical stability conditions by embedding information on the fuzzy premises.*

Keywords— Fuzzy Control, Takagi-Sugeno Fuzzy Systems, Stability, Parallel Distributed Control.

1 Introduction

T-S fuzzy model is a fuzzy dynamic model which embeds a set of fuzzy rules in order to describe a global nonlinear system as a set of local linear models, these being smoothly connected by fuzzy membership functions [1]. As T-S fuzzy models are equivalent to polytopic linear models, systematic approaches to stability analysis and controller design can be developed using powerful conventional (linear and nonlinear) control theory with the help of optimization techniques such as Linear Matrix Inequalities solvers [2].

Control of T-S fuzzy system is most of the time achieved by the so-called Parallel Distributed Compensation method (PDC) for which the fuzzy controller shares the same fuzzy sets with the fuzzy model in the premise parts, and, thus, each control rule is distributively designed for the corresponding rule of a T-S fuzzy model. Since the consequent parts of T-S fuzzy models are described by linear state equations, linear control theory and corresponding stability tools can be used to design the fuzzy control local gains [3].

Among the mathematical tools used for assessing the stability of a T-S fuzzy system (with or without PDC control), the use of a Common Quadratic Lyapunov Function (CQLF) for all subsystems is the most popular, because it allows to derive easily explicit conditions [2] [4] [5]. It has been noted that Common Quadratic Lyapunov functions tend to be conservative, and even worse, might not exist for many complex highly nonlinear systems. Several works have allowed to relax stability conditions [4] [6] [7], and as an alternative, such tools as piecewise quadratic Lyapunov functions or fuzzy Lyapunov functions have been proposed (e.g. [8]). The drawbacks of these methods lie in the difficulty of synthesizing control gains - and not only analyze closed-loop stability with fixed controllers -, and that the results are quite independent of the membership functions, leading sometimes to poor overall performances. However, extended results have been obtained when considering additional bounds or shapes on membership functions considering linear consequents and PDC [9] [10] or nonlinear / non PDC controllers [11] [12] [13].

In short, there still remains a need for designing further tools to synthesize PDC control for T-S fuzzy systems, which appears to be very difficult when the input matrices of the systems of the fuzzy consequents are not identical or not proportional. In this case, the closed-loop system is not a parallel distribution of the individual closed-loop subsystems, because additional coupling terms appear, as the control gain designed for one subsystem is distributed over the remaining subsystems. Whereas most papers embed the general case, the examples are nearly always proposed for proportional or identical input matrices, or for a very limited number of input matrices.

When input matrices are different, it is not only necessary that all subsystems be stable, but also that any subsystem should be at least stable (or performing) under every other local controller, and, moreover, that all these closed-loop controlled subsystems should share a CQLF. A solution to this problem consists of solving a Riccati equation which has a positive definite solution that corresponds to a Lyapunov function, using some results from linear control theory [14]. In this case, the set of coupling terms is represented by a product of matrices involving a single uncertain matrix with a norm smaller than one, leading to a global Riccati equation [2].
Whereas this method is powerful and depends on the amplitude and variations of the membership functions, finding a global bounding matrix for the coupling terms is often not easy to work out, because these terms depend on the control gains themselves; in the case of control synthesis, when the gains are a priori unknown, this task can be impossible. The cancelation of such coupling terms has been tackled only for large-scale systems [15]. To sum up, practical and membership function-dependent of PDC control synthesis has not yet be performed when subsystems exhibit different input matrices as the explicit cancelation of coupling terms has not been proposed.

In this paper, the closed-loop T-S system under PDC control is rewritten in such a way that the cross-coupled terms are clearly

ISBN: 978-989-95079-6-8

seen as a weighted sum of the pairwise products involving the difference of two input matrices and their corresponding control gains. This allows, first, to determine easily a Riccati equation for each subsystem, which does not depend on the other subsystems control gains, and that takes the variation of fuzzy membership functions into account. Moreover, adding a little more conservativeness, allowing the control gains to be bounded, turns the stability conditions into a set of very simple Lyapunov equations, which might be a good help for control synthesis. Finally, it is shown that, when the number of subsystems is sufficient, the cross-coupled terms can be canceled by proposing nonlinear control gains.

2 Takagi-Sugeno fuzzy model and controller

2.1 Closed-loop model decomposition

The fuzzy model proposed by Takagi and Sugeno consists of a set of r fuzzy IF...THEN rules for which the consequents are linear models:

Plant Rule R_i: IF z_1 IS M_{i1} AND $\cdots$ AND z_g IS M_{ig} THEN $\dot{x} = A_i x + B_i u$;
where $x(t)$, $u(t)$ are state and input vectors, $z_i(t)$, M_{ij} are the premise variables and corresponding fuzzy rules, the final output of the fuzzy system being inferred as follows:

$$\dot{x} = \sum_{i=1}^{r} \mu_i \big(A_i x + B_i u\big), \tag{1}$$

where $\mu_i = \frac{\omega_i}{\sum_{i=1}^{r} \omega_i}$, ω_i is the grade of membership function of rule R_i.

For every subsystem S_i, a local controller can be defined as $u = K_i x$, where K_i is a control gain (possibly nonlinear). It is intuitive to build a fuzzy controller using a set of rules which share the same premises as the fuzzy model, hence distributing the local controllers within the global controllers according to their systems' weights.

Controller C_i: IF z_1 IS M_{i1} AND $\cdots$ AND z_g IS M_{ig} THEN $u = K_i x$, yielding:

$$u = \sum_{i=1}^{r} \mu_i K_i x. \tag{2}$$

Lemma 2.1. *Let the system $\dot{x} = \sum_{i=1}^{r} \mu_i (A_i x + B_i u)$ with PDC control $u = \sum_{i=1}^{r} \mu_i K_i x$ such that $A_i + B_i K_i = G_i$ and $\sum_{i=1}^{r} \mu_i \leq 1$, $\mu_i \geq 0$. The closed-loop system is:*

$$\dot{x} = \left(\sum_{i=1}^{r}\sum_{j=1}^{r} \mu_i \mu_j G_i + \sum_{i=1}^{r} A_i \Big(1 - \sum_{j=1}^{r} \mu_j\Big) + \sum_{i,j=1,j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x. \tag{3}$$

Proof. One has

$$\dot{x} = \sum_{i=1}^{r} \mu_i \left(A_i x + B_i \sum_{j=1}^{r} \mu_j K_j x \right)$$
$$= \sum_{i=1}^{r} \mu_i \left(A_i + \mu_i B_i K_i + B_i \sum_{j=1,j\neq i}^{r} \mu_j K_j \right) x.$$
$$\dot{x} = \sum_{i=1}^{r} \left(\mu_i^2 G_i + \mu_i A_i (1 - \mu_i) + \mu_i B_i \sum_{j=1,j\neq i}^{r} \mu_j K_j \right) x.$$

Moreover,

$$\sum_{i=1}^{r} \mu_i B_i \sum_{j=1,j\neq i}^{r} \mu_j K_j = \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j (G_i - A_i) + \sum_{i=1}^{r} \mu_i B_i \sum_{j=1,j\neq i}^{r} \mu_j K_j - \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j B_i K_i.$$

In this equation, one can rearrange the two last sums into a sum of pairwise terms:

$$\sum_{i,j=1,j\neq i}^{r} \mu_j B_j \mu_i K_i + \mu_i B_i \mu_j K_j - \mu_i \mu_j B_i K_i - \mu_j \mu_i B_j K_j = \sum_{i,j=1,j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i).$$

Hence,

$$\sum_{i=1}^{r} \mu_i B_i \sum_{j=1,j\neq i}^{r} \mu_j K_j = \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j (G_i - A_i) + \sum_{i,j=1,j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i).$$

One has now:

$$\dot{x} = \left(\sum_{i=1}^{r} \Big(\mu_i^2 G_i + \mu_i A_i (1 - \mu_i) \Big) + \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j (G_i - A_i) + \sum_{i,j=1,j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x.$$

As $\sum_{i=1}^{r} \mu_i^2 G_i + \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j G_i = \sum_{i=1}^{r} \sum_{j=1}^{r} \mu_i \mu_j G_i$, and

$$\sum_{i=1}^{r} \mu_i A_i (1 - \mu_i) - \sum_{i=1}^{r} \sum_{j=1,j\neq i}^{r} \mu_i \mu_j A_i = \sum_{i=1}^{r} \mu_i A_i \Big(1 - \sum_{j=1}^{r} \mu_j\Big).$$

We demonstrate the final result:

$$\dot{x} = \left(\sum_{i=1}^{r}\sum_{j=1}^{r} \mu_i \mu_j G_i + \sum_{i=1}^{r} \mu_i A_i \Big(1 - \sum_{j=1}^{r} \mu_j\Big) + \sum_{i,j=1,j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x.$$

□

ISBN: 978-989-95079-6-8

Remark: When $\sum_{i=1}^{r} \mu_i = 1$, formula (3) is reduced to:

$$\dot{x} = \left(\sum_{i=1}^{r} \mu_i G_i + \sum_{i,j=1, j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x.$$

2.2 Global Stability Verification

It should be recalled that a Linear Matrix Inequality (LMI) is a set of equations which can be put under the form $F_O + F_1 x_1 + \cdots F_n x_n \prec 0$, where $F_i^T = F_i$. A matrix F is said negative definite, which is noted $F \prec 0$ if $\forall x \neq 0, x^T F x < 0$.
A Bilinear Matrix Inequality involves bilinear terms and cannot be solved in a straightformard way as LMIs are.

Theorem 2.2. *The system $\dot{x} = \sum_{i=1}^{r} \mu_i (A_i x + B_i u)$, under PDC control $u = \sum_{i=1}^{r} \mu_i K_i x$, such that $A_i + B_i K_i = G_i$ and $A_i + B_i K_j = G_{ij}$, is stable if there exists a common positive definite matrix P such that:*

$$\begin{array}{l} \forall\, i = 1 \cdots r, P G_i + G_i^T P \prec 0, \\ \forall\, i < j, P\left(G_{ij} + G_{ji}\right) + \left(G_{ij} + G_{ji}\right)^T P \prec 0. \end{array} \quad (4)$$

Remark: Theorem (2.2) allows the determination of both the Lyapunov matrix and the controller gain, using a change of variable $N_i = K_i P^{-1}$, when being replaced in the stability conditions, leads to a set of LMIs in N_i and in P, the PDC controller being provided by $K_i = N_i P$. The conservativeness of the result comes from several reasons: of course, the global system can be stable without sharing a Common Quadratic Lyapunov function. The conditions are independent of the membership functions, and those roughly mean that system *i* is considered to behave "well" with the corresponding local controller $u = K_i x$, but also with any other local controller $u = K_j x, j \neq i$. Of course, it cannot be expected that a system with a controller designed for another plant has a "good" behavior, and hence, the PDC controller is designed according to the "worst" case among the pairs {Plant *i*, Controller *j*}.

3 Coupling terms attenuation

Theorem 3.1. *[14] First, we consider the linear uncertain system for which $\dot{x} = A + \sum_{i=1}^{r} D_i \delta_i E_i, \|\delta_i\| \leq 1$, and the elements of δ are Lebesgue measurable. Then the positive-definite matrix P is a common Lyapunov matrix for this system if there exists r positive scalars η_i such that:*

$$PA + A^T P + \sum_{i=1}^{r} \eta_i P D_i D_i^T P + \eta_i^- 1 E_i^T E_i \prec 0,$$

or simply, if $\forall i, \eta_i = 1$

$$PA + A^T P + \sum_{i=1}^{r} P D_i D_i^T P + E_i^T E_i \prec 0. \quad (5)$$

Remark: This Theorem was applied first by Tanaka [2] and then by numerous authors to the whole coupling term. Whereas this method provides for a rather non-conservative solution, it is clear that finding individual uncertain matrices might be a tedious task, because the rate of variation and thus the bounds on the uncertain matrix depend on the control gains themselves. It can thus be applied to analyze an existing solution (the gains are fixed) but not for gain synthesis considering models/controllers coupling. The following Theorem proposes a different application of this method to every individual component of the coupling terms.

Theorem 3.2. *Consider the system $\dot{x} = \sum_{i=1}^{r} \mu_i (A_i x + B_i u)$, under PDC controller:*

$$\dot{x} = \left(\sum_{i=1}^{r} \mu_i G_i + \sum_{i,j=1, j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x.$$

Let us suppose that: $\forall\, i$, there exists b_i such that:

$$\sum_{j/B_i \neq B_j} \mu_j (B_j - B_i) = b_i \delta_i, \text{ where } \|\delta_i\| \leq 1.$$

The matrices δ_i thus depend on membership functions and other input matrices μ_j and B_j; as μ_j may vary with time, δ_i is a matrix which may vary with time or with the state space x. The corresponding norm is the Euclidean norm.
The closed-loop system is quadratically stable if:

$$\forall\, i = 1 \cdots r, P G_i + G_i^T P + P b_i b_i^T P + K_i^T K_i \prec 0. \quad (6)$$

This can be turned into:

$$\forall\, i = 1 \cdots r, \begin{pmatrix} P G_i + G_i^T P & P b_i & K_i^T \\ b_i^T P & -I & 0 \\ K_i & 0 & -I \end{pmatrix} \prec 0. \quad (7)$$

Proof.

$$\begin{aligned} \dot{x} &= \sum_{i=1}^{r} \left(\mu_i G_i + \sum_{i,j=1, j\neq i}^{r} \mu_i \mu_j (B_i - B_j)(K_j - K_i) \right) x \\ &= \sum_{i=1}^{r} \mu_i \sum_{j=1, j\neq i}^{r} G_i + \mu_j (B_i - B_j) K_i. \end{aligned}$$

One has now: $G_i + \mu_j (B_i - B_j) K_i = G_i + b_i \delta_i K_i$, and one can apply the Theorem (3.1). □

Remark: Uncertain matrices δ_i do not depend anymore on the control gains but only on input matrices and membership functions which are known a priori. Their determination is thus quite easy and the membership functions are indeed embedded in the control synthesis. Note also that the corresponding *i* Riccati equations in (6) are decoupled, i.e. the *i*th equation only depends on the *i*th control gain, the influence of the other subsystems being lumped into the matrix $b_i \delta_i$. However, one can realize that equation (7) is a BMI in P and K_i, and thus less tractable for control gain synthesis, which motivates the following Corollary.

Corollary 3.3. *Let us suppose that:*

$$\forall\, i = 1 \cdots r,\ \exists\, Q_i \succ 0,\ K_i^T K_i - Q_i \prec 0.$$

Then, condition (7) can be expressed as:

$$\exists\, P \prec 0, \forall\, i = 1 \cdots r, P G_i + G_i^T P + Q_i' \prec 0, \quad (8)$$

ISBN: 978-989-95079-6-8

where $Q'_i = Pb_ib_i^TP + Q_i$, *with* $K_i^TK_i - Q_i \prec 0$,
which can be turned into:

$$\forall\, i = 1 \cdots r, \begin{cases} \begin{pmatrix} PG_i + G_i^TP + Q_i & Pb_i \\ b_i^TP & -I \end{pmatrix} \prec 0, \\ K_i^TK_i - Q_i \prec 0. \end{cases}$$

The Corollary simply reduces the search for a common Lyapunov matrix to a series of r Lyapunov equations and thus r LMIs. This is really a drastic improvement to other methods because, now, control gains can nearly be selected independently without the need to taking care of coupling terms, at the expense of a priori gain limitation.

4 Example

Let us take the 3 following systems:

$$A_1 = \begin{pmatrix} -1 & 2 \\ 0 & -2 \end{pmatrix}, B_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix};$$
$$A_2 = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, B_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix};$$
$$A_3 = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}, B_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix};$$

along with local gains:

$K_1 = (\ 2 \quad 1\ ), K_2 = (\ -2 \quad 1\ ), K_3 = (\ 2 \quad 0\ ).$

The premisses corresponding to systems 1, 2 and 3 are:
$\mu_1 = z$, $\mu_2 = 1 - z$ and $\mu_3 = z$ where $z \in [-1 \cdots 1]$.

For every subsystem i, it is quite easy to compute the matrices b_i such that $\sum_{B_i \neq B_j} \mu_i(B_j - B_i) = b_i\delta_i$, since the upper bound depends on the fuzzy variable z.
One finds:
${b_1}^T = (\ 1 \quad 0.25\ )$, ${b_2}^T = (\ 0.75 \quad 0.25\ )$,
${b_3}^T = (\ 1 \quad 1\ )$.

The application of Theorem (3.2) allows to find a common positive definite matrix $P = \begin{pmatrix} 1.28 & -0.37 \\ -0.37 & 0.87 \end{pmatrix}$ whereas it is impossible to find one by the classical method; it is easy to check that the gain K_2 is unable to stabilize matrix A_1 and the converse for K_1 and A_2. It is quite interesting to note that the result is quite tied to the value of the matrices b_i.
When all other variables keep the same values, but $b_2^T = (\ 1 \quad 1\ )$, then the Theorem (3.2) is no more applicable because a positive definite CQLF cannot be found. Thus, Theorem (3.2) is able to relax stability conditions, depending strongly on the membership functions and input matrices values.

5 Conclusion

In this paper, the stability of a Takagi Sugeno fuzzy system under the Parallel Distributed Compensation controller has been studied. This control strategy allocates the same weight to a local controller that the one in the fuzzy combination of the local submodels.
The influence of the coupling between any local subsystem and any local controller (different from the corresponding local controller designed from the local subsystem considered) in the closed-loop response has been highlighted, and it has been shown to be effective when the input matrices of the subsystems are different. It has been subsequently shown that a controller synthesis based on an analysis of each local subsystem controlled by any local compensator, would lead to conservative results.

A new approach has been proposed which, for every local subsystem, takes the coupling term coming from other subsystems into account, and proposes to choose the gain in order to cope with the effect of the coupling terms. This strategy allows to minimize the number of Linear Matrix Inequalities to be solved for controller synthesis and to take into account the shape of the membership functions.

References

[1] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15:116–132, 1985.

[2] K. Tanaka, T. Ikeda, and H. O. Wang. Robust stabilization of a class of uncertain nonlinear systems via fuzzy control: Quadratic stabilizability, H∞ control theory, and linear matrix inequalities. *IEEE Transactions on Fuzzy Systems*, 4:1–13, 1996.

[3] P. Gahinet, A. Nemirovski, A. Laub, and M. Chilali. *The LMI Control Toolbox*. The Mathworks, Inc., Natick, MA, 1995.

[4] K. Tanaka, T. Ikeda, and H. O. Wang. Fuzzy regulators and fuzzy observers: Relaxed stability conditions and LMI-based designs. *IEEE Transactions on Fuzzy Systems*, 6:250–265, 1998.

[5] J. Yoneyama. Robust H∞ control analysis and synthesis for Takagi-Sugeno general uncertain fuzzy systems. *Fuzzy Sets and Systems*, 157:2205–2223, 2006.

[6] C.H. Fang, Y.S. Liu, S.W. Kau, L. Hong, and C.H. Lee. A new LMI-based approach to relaxed quadratic stabilization of T-S fuzzy control systems. *IEEE Transactions on Fuzzy Systems*, 14:286–297, 2006.

[7] E. Kim and H. Lee. New approaches to relaxed quadratic stability condition of fuzzy control systems. *IEEE Transactions on Fuzzy Systems*, 8:523–534, 2000.

[8] G. Feng, C. L. Chen, D. Sun, and X. P. Guan. H∞ controller synthesis of fuzzy dynamic systems based on piecewise Lyapunov functions and bilinear matrix inequalities. *IEEE Transactions on Fuzzy Systems*, 13:94–103, 2005.

[9] A. Sala and C. Ariño. Relaxed stability and performance LMI conditions for Takagi-Sugeno fuzzy systems with polynomial constraints on membership function shapes. *IEEE Transactions on Fuzzy Systems*, 16:1328–1336, 2008.

[10] A. Sala and C. Ariño. Relaxed stability and performance conditions for Takagi-Sugeno fuzzy systems with knowledge on membership function overlap. *IEEE Transactions On Systems, Man, And Cybernetics-Part B*, 37:727–732, 2007.

[11] M. Abbaszadeh and H. J. Marquez. LMI optimization approach to robust H∞ observer design and static output feedback stabilization for discrete-time nonlinear uncertain systems. *IEEE Int. J. Robust Nonlinear Control*, 19:313–340, 2009.

[12] H.K. Lam and F.H.F. Leung. Stability analysis of fuzzy control systems subject to uncertain grades of membership. *IEEE Transactions on Systems, Man and Cybernetics - Part B*, 35:1322–1325, 2005.

ISBN: 978-989-95079-6-8

[13] H.K. Lam and F.H.F. Leung. LMI-based stability and performance conditions for continuous-time nonlinear systems in Takagi-Sugeno's form. *IEEE Transactions on Systems, Man and Cybernetics - Part B*, 37:1396–1406, 2007.

[14] P.P. Khargonekar, I.R. Petersen, and K. Zhou. Robust stabilization of uncertain linear systems: Relations between quadratic stabilizability and H1 control theory. *IEEE Transactions on Automatic Control*, 37:356–61, 1990.

[15] W.J. Wang and W.W. Lin. PDC Synthesis for T-S fuzzy large-scale systems. *IEEE Transactions on Fuzzy Systems*, 12:309–315, 2004.

ISBN: 978-989-95079-6-8

Characterization of L-fuzzy semi-filters and semi-ideals

Jorge Jiménez, Susana Montes[1], Branimir Šešelja, Andreja Tepavčević[2]
1. University of Oviedo, Spain
2. University of Novi Sad, Serbia
Email: {meana,montes}@uniovi.es, {seselja,andreja}@im.ns.ac.yu

Abstract— *Lattice-valued up-sets and down-sets of a poset are investigated under the cutworthy approach. It is proved that the collections of all lattice valued up-sets and down-sets of a given poset are complete lattices under fuzzy inclusion. Properties of these are investigated. For a given collection of crisp up-sets of a poset, necessary and sufficient conditions are given under which the family represents a collection of cuts of a lattice valued up-set. The corresponding conditions are also obtained for the analogous case of down-sets.*

Keywords— L-valued sets (L-fuzzy sets), L-valued up-sets, L-valued down-sets.

1 Introduction

Lattice valued sets and structures have been widely studied from Goguen's first paper [1] on this topic, and since Negoita and Ralescu published their book [2] back in 1975. They appear when the membership grades can be represented by elements of an arbitrary partially ordered set L, instead of just by numbers in the unit interval $[0, 1]$. Although a lot of studies have been done for fuzzy order structures (see, e.g., [3, 4, 5, 6] among many others), the relaxation of the requirements allow us to use lattice-valued sets in a wider context, since they can be more appropriate to model natural problems. It turns out that recently this approach to fuzziness attracts more and more interest, in particular when a structure has a residuated or similar lattice as a co-domain. Thus, in the last decade, lattice-valued mathematics have undergone a significant development. This have created bridges between them and algebraic theories, quantales and order-theoretic structures, various subdisciplines of topology and theoretical computer science (see, e.g., [7, 8, 9, 10, 11, 12, 13, 14]).

Due to the importance of up-sets (semi-filters) and down-sets (semi-ideals) in the classical representation theory, the main purpose of this paper is to study the generalization of these concepts for L-valued sets. Fuzzy semi-ideals (down-sets) and semi-filters (up-sets) were investigated in [15], refereing to a particular T-norm, therefore not in the framework of our cutworthy approach. Let us remember that a property generalized to a fuzzy structure is said to be cutworthy if the corresponding crisp property is preserved by the cut-structures. It is known that many properties of relations or algebraic structures are cutworthy if the co-domain of the investigated fuzzy structure is a complete lattice (e.g., properties of equivalence or ordering relations, then subalgebras of algebraic structures etc.). On the other hand, if this co-domain is a lattice with some additional operations like residuation, product etc., then this cutwortheness is not fulfilled. Since our aim is to connect a lattice valued structure with its cut-structures, the co-domain of all fuzzy structures here is a complete lattice, without additional operations.

In the present investigation, we start with a finite poset X and a complete lattice L. Then we introduce L-valued up-sets (semi-filters) and L-valued down-sets (semi-ideals) as isotone and anti-isotone mappings from X to L, respectively. It turns out that the cut-sets of these are precisely crisp up-sets and down-sets on X. Next we investigate collections of all L-valued up-sets and all L-valued down-sets on X. We prove that under the order induced by the one in L, these collections are complete lattices, fulfilling all identities satisfied by the lattice L. Moreover, we analyze the problem of existence of a representation of collections of crisp up-set on X by collections of cut-set of an L-valued up-set on X. We give necessary and sufficient conditions (fulfilled by the lattice L) under which such representation exists.

Our main motivation in this work was to present a mathematical characterization of these algebraic objets, namely semi-filters and semi-ideals, in the fuzzy environment. The justification to consider lattice valued fuzzy sets has been widely explained in the literature, since lattices are more richer structure and we can obtain non-comparable values of fuzzy sets. They can be applied, for instance, in image processing.

The work is structured in 4 sections. In Section 2 we recall the most relevant concepts concerning classical order sets, up-sets and down-sets, as well as lattice-valued sets. In Section 3 we introduce L-valued up-sets and down-sets and we relate these concepts. The first part of this section is devoted to a wide study of the properties and characterizations for L-valued up-sets. In the second part of the section, analogue problems are investigated for L-valued down-sets. In Section 4 we briefly address some conclusions and future work.

2 Preliminaries

In this section, some well-known definitions and preliminary results are recalled. They will be necessary in order to understand the new concepts introduced in this paper and their associated studies.

2.1 Order, up-sets and down-sets

Some necessary notions from the classical order theory are listed in the sequel, together with relevant properties. For more comprehensive presentation, see e.g., books [16, 17].

A **poset** is a nonempty set X equipped with an ordering relation $\leq$. A poset is usually denoted as an ordered pair $(X, \leq)$, or simply by the underlying set X. A sub-poset of $(X, \leq)$ is a poset on a subset Y of X in which the order is the one restricted from X, and usually denoted in the same way ($\leq$). An **up-set** (**semi-filter**) on a poset X is any subposet U, satisfying the following: for $x \in U, y \in X$, $x \leq y$

ISBN: 978-989-95079-6-8

implies $y \in U$. Dually, an **down-set** (**semi-ideal**) on X is any sub-poset D, satisfying: for $x \in D, y \in X$, $y \leq x$ implies $y \in D$. A **lattice** is a poset L in which for each pair of elements x, y there is a greatest lower bound (glb, infimum, meet) and a least upper bound (lub, supremum, join), denoted respectively by $x \wedge y$ and $x \vee y$. These are binary operations on L. A non-empty poset L is said to be a **complete** lattice if infimum and supremum exist for each subset of L. Complete lattice possesses the **top** (1) and the **bottom** element (0). On the other hand, a lattice L is **distributive**, if each operation is distributive with respect to the other. An example of distributive lattices is presented in the following lemma.

Lemma 1 *The collection of all up-sets (down-sets) of a poset X is a distributive lattice under inclusion.*

Given a lattice L, an element a in L is **completely meet-irreducible** if $a \neq 1$ and for every family $\{x_i \mid i \in I\}$ of elements from L, from $a = \bigwedge_{i \in I} x_i$ it follows that $a = x_i$ for some $i \in I$. When L is a lattice of finite length, a is just said to be **meet-irreducible**.

A **closure operator** on a lattice L is a function $C : L \to L$ such that, for all $p, q \in L$, it fulfills the three following requirements:

- $p \leq C(p)$,
- $p \leq q \to C(p) \leq C(q)$,
- $C(C(p)) = C(p)$.

If $p = C(p)$, then p is a **closed element** under the corresponding closure operator.

Lemma 2 *Let C be a closure operator on a lattice L. Then,*

1. *The subset of all closed elements of L is closed under meets in L.*
2. *The top element of L is a closed element under C.*

Lemma 3 *Let L be a complete lattice. Then,*

1. *For any closure operator C on L, the subset of all closed elements of L is a complete lattice under the order inherited from L.*
2. *If F is a subset of L closed under arbitrary meets, then the map such that to any element p in L associates the value $\bigwedge\{q \in F \mid p \leq q\}$ is a closure operator on L.*

By the definition of the supremum and the infimum in a complete lattice, we also have that $\bigwedge \emptyset = 1$ and $\bigvee \emptyset = 0$.

Lemma 4 *Let C be a closure operator on a complete lattice L. If we consider the relation $\sim$ on L defined by $x \sim y$ iff $C(x) = C(y)$, then*

1. *$\sim$ is an equivalence relation on L.*
2. *Each such equivalence class has the top element and this element is closed.*
3. *The set $L/\sim$ (denoted also by L/C) can be ordered: $[x]_\sim \leq [y]_\sim$ iff $C(x) \leq C(y)$ in L.*
4. *The poset $(L/\sim, \leq)$ is a lattice isomorphic with the poset of closed elements of L, under C.*

2.2 Lattice-valued sets

We present some notions from the theory of lattice-valued structures. More details about the relevant properties can be found in [18, 19].

Lattice-valued, **L-valued sets** or **L-fuzzy sets** are here considered to be mappings from a non-empty set X (domain) into a complete lattice L (co-domain) with the top and bottom elements 1 and 0, respectively. This concept can be seen as a generalization of the concept of fuzzy or valued set and it was introduced by Goguen (see [1]).

If $\alpha : X \to L$ is an L-fuzzy set on X then, for $p \in L$, the set

$$\alpha_p := \{x \in X \mid \alpha(x) \geq p\}$$

is called the **p-cut**, a **cut set** or simply a **cut** of α.

Example 1 *Let us consider the poset $(X, \leq)$ and the complete lattice $(L, \leq)$ represented in Figure 1.*

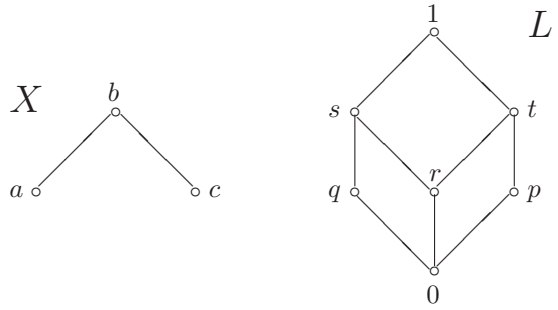


Figure 1: Hasse diagrams of X and L.

Some examples of L-valued sets are

X	a	b	c
α	r	t	p

X	a	b	c
ν	q	s	r

X	a	b	c
μ	r	0	p

X	a	b	c
λ	q	s	t

where we consider a tabular representation for the L-valued sets on X.

In the case of α, for instance, its p-cut is $\alpha_p = \{b, c\}$ and its t-cut is $\alpha_t = \{b\}$.

We can notice that the ordering in X does not have a role in this example. This ordering is used later in the paper.

Some characterizations of the collection of cuts of a lattice-valued set are presented in the next three results.

Proposition 1 *[18] Let $\mathcal{F}$ be a family of subsets of a nonempty set X which is closed under intersections and contains X. Let $\alpha : X \to \mathcal{F}$ be defined by*

$$\alpha(x) = \bigcap(p \in \mathcal{F} \mid x \in p).$$

Then, α is a $\mathcal{F}$-valued set on X with the codomain lattice $(\mathcal{F}, \supseteq)$ such that its family of p-cuts is $\mathcal{F}$ and for every $p \in \mathcal{F}$ it holds that $p = \alpha_p$.

Theorem 1 *[20] Let L be a fixed complete lattice. Necessary and sufficient conditions under which $\mathcal{F} \subseteq \mathcal{P}(X)$ is the collection of cut sets of an L-valued set with domain X are:*

ISBN: 978-989-95079-6-8

1. *$\mathcal{F}$ is closed under arbitrary intersections and contains X,*
2. *The dual poset of $\mathcal{F}$ under inclusion can be embedded into L, such that all infima and the top element are preserved under the embedding.*

Proposition 2 *[21] Let L be a lattice of finite length and let $\alpha : X \to L$ be an L-valued set. Then, all p-cuts of α are distinct if and only if all meet-irreducible elements of L belong to $\alpha(X)$.*

3 L-valued up-sets and down-sets

In classical order theory, the families of up-sets and down-sets, associated to any ordered set, play a central role in the representation theory. Once the L-valued sets are considered, a step forward would be to define and characterize up-sets and down-sets in the L-valued logic. That study is made in this section.

Classical up-sets and down-sets can be seen as sets which are "closed under going up" and "closed under going down", respectively. In fact, a subset is an up-set if and only if its characteristic function is order-preserving. Moreover, a subset is a down-set if and only if its characteristic functions is order-preserving, when the dual of the usual order is considered in $\{0, 1\}$. Thus, a natural generalization of these concepts for L-valued sets is presented in Definition 1.

Definition 1 *Let $(X, \leq)$ be a poset and let $(L, \leq)$[1] be a complete lattice in which 0 and 1 are the bottom and the top element, respectively.*

A function $\alpha : X \to L$ such that for all $x, y \in X$

$$x \leq y \implies \alpha(x) \leq \alpha(y)$$

*is an **L-valued up-set** or an **L-valued semi-filter** on X.*

Dually, a function $\mu : X \to L$ such that for all $x, y \in X$

$$x \leq y \implies \mu(y) \leq \mu(x)$$

*is an **L-valued down-set** or an **L-valued semi-ideal** on X.*

Example 2 *If we consider the L-valued sets introduced in Example 1, α and ν are L-valued up-sets, μ is an L-valued down-set and λ is neither L-valued up-set nor L-valued down-set.*

If $L = \{0, 1\}$, that is, if we are in a two-valued logic, it is immediate that any L-valued up-set is a classical up-set and the same happens for down-sets. Thus, these concepts generalize the classical ones.

The duality between these two concepts is showed in the following proposition, by means of complement operators [22].

Proposition 3 *Let $(X, \leq)$ be a poset and let $(L, \leq)$ be a complete lattice in which 0 and 1 are the bottom and the top element, respectively. If there exists a complement operator c on L, that is, a map $c : L \to L$ fulfilling*

1. *Boundary values: $c(0) = 1$ and $c(1) = 0$;*
2. *Non-increasing: $p \leq q \Rightarrow c(q) \leq c(p), \forall p, q \in L$;*
3. *Involution: $c(c(p)) = p, \forall p \in L$;*

then an L-valued set $\alpha : X \to L$ is an L-valued up-set if and only if, its c-complement α^c is an L-valued down-set, where α^c denotes the L-valued set on X defined by $\alpha^c(x) = c(\alpha(x))$.

3.1 Characterization of the family of L-valued up-sets

In this section we present a wide study of the concept of L-valued up-sets. This study is repeated later, in next section, for L-valued down-sets, due to the duality between these two classes of L-valued sets. Let us start with a characterization of the concept of L-valued up-set by means of α-cuts.

Theorem 2 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let α an L-valued set on X. The following two statements are equivalent:*

- *α is an L-valued up-set on X.*
- *The p-cut α_p of α is crisp up-set (semi-filter) on X, for any $p \in L$.*

The proof of this theorem is a immediate consequence of the definitions of p-cut and L-valued up-set.

Example 3 *Let us consider again the L-valued up-set α introduced in Examples 1 and 2. The family of its cuts is formed by the elements*

$$\alpha_0 = X, \alpha_p = \{b, c\}, \alpha_r = \{a, b\}, \alpha_t = \{b\}$$

$$\text{and } \alpha_q = \alpha_s = \alpha_1 = \emptyset.$$

As a consequence of the previous theorem, we know that all these crisp sets are up-sets. Conversely, since all the cuts are crisp up sets on X the previous theorem says that α is an L-valued up-set which is easy to check.

From now on, we denote by $\mathcal{U}_L(X)$ the subset of the L-valued power set of X formed by all L-valued up-sets on the poset X, that is,

$$\mathcal{U}_L(X) = \{\alpha : X \to L \mid \alpha \text{ is an } L\text{-valued up-set}\}.$$

This poset can be ordered naturally using the order induced by the one from the lattice L. Thus, for any $\alpha, \beta \in \mathcal{U}_L(X)$, we say that $\alpha \leq \beta$ if and only if for each $x \in X$ $\alpha(x) \leq \beta(x)$.

Moreover, meet and join exist for any subset of $\mathcal{U}_L(X)$, as it is stated in the following theorem.

Theorem 3 *The poset $(\mathcal{U}_L(X), \leq)$ is a complete lattice.*

In fact, the poset $(\mathcal{U}_L(X), \leq)$ is a complete sublattice of the lattice $(L^X, \leq)$. More specifically, it is a $\{0, 1\}$-complete sublattice with $0(x) = 0$ and $1(x) = 1$, for all $x \in X$.

Due to the definition of the order in $\mathcal{U}_L(X)$, we have the following obvious consequence.

Proposition 4 *If a lattice identity holds in L, then the same identity is satisfied in the lattice $(\mathcal{U}_L(X), \leq)$.*

[1] For simplicity, we consider the same notation for the order in X and in L. Of course these orders do not have to coincide. However, possible ambiguity is removed by the context, so that in any moment we are able to identify order represented by $\leq$.

ISBN: 978-989-95079-6-8

From Proposition 4, if L is e.g., a distributive lattice, then we also have that the lattice $(\mathcal{U}_L(X), \leq)$ is distributive.

Now, we prove a result which allows us to know when, starting from a given collection of crisp up-sets, it is possible to built an L-valued up-set.

Theorem 4 *Let $(X, \leq)$ be a poset, let $\mathcal{F} \subseteq \mathcal{P}(X)$ be a family of some up-sets of a poset X, and let $(L, \leq)$ be a complete lattice. Then, there is an L-valued up-set $\alpha : X \to L$ such that its family of cuts is equal to $\mathcal{F}$ if and only if the following two conditions hold:*

1. *$\mathcal{F}$ is closed under intersections and contains X;*
2. *there is a closure operator C on L, such that the poset $(L/C, \leq)$ is order isomorphic to $(\mathcal{F}, \supseteq)$.*

In the proof of this theorem we have to prove the equivalence among these conditions and the conditions imposed in Theorem 1.

Example 4 *Let us consider again the poset X, the complete lattice L and the L-valued sets α, ν introduced in Example 1. Let $\mathcal{F}$ be the family of up-sets of X defined by*

$$\mathcal{F} = \{X, \{a, b\}, \{b, c\}, \{b\}, \emptyset\}.$$

This family fulfills Conditions 1 and 2 in Theorem 4 and therefore, there exists an L-valued up-set whose family of cuts is equal to $\mathcal{F}$. This L-valued up-set is precisely the L-valued set α. Of course, it does not need to be unique. For instance, also ν is an L-valued up-set with $\mathcal{F}$ as its family of cuts.

In the previous theorem, we have characterized families of up-sets which are the cut-sets of an L-valued up-set. Now, we are going to characterize the L-valued up-set such that this family is formed by all the (crisp) up-sets.

Proposition 5 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let $\alpha : X \to L$ be an L-valued up-set. The following two statements are equivalent:*

- *The family of cut sets of α is formed by all the crisp up-sets of poset $(X, \leq)$.*
- *For every family $\{x_i \mid i \in I\}$ of elements from X and every $x \in X$, it holds that*

$$\alpha(x) \geq \bigwedge_{i \in I} \alpha(x_i) \Rightarrow \exists i \in I \mid x \geq x_i.$$

A main step in the proof of the previous proposition is to consider that the $\bigcup_{i \in I} {\uparrow} x_i$ is a crisp up-set of α, where $\{x_i \mid i \in I\}$ is the family of elements from X such that $\alpha(x) \geq \bigwedge_{i \in I} \alpha(x_i)$ for $x \in X$ and $x \not\geq x_i$ for all $i \in I$.

As a consequence of the previous proposition, we obtain two necessary conditions for a representation by means of all the crisp up-sets. They are presented in the two following corollaries.

Corollary 1 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let $\alpha : X \to L$ be an L-valued up-set. If the family of cut sets of α is formed by all the crisp up-sets of poset $(X, \leq)$, then α is an order-embedding.*

In the following result it is necessary to consider the subposet L^α of L associated to any L-valued set α on X and generated by taking all infima of all the subsets of $\alpha(X)$ (including infimum of the empty set - the top element of L). Thus, given an L-valued set α on X,

$$L^\alpha = \{p \in L \mid p = \wedge B \text{ with } B \subseteq \alpha(X)\}.$$

Let us note that it is immediate that L^α is a complete lattice.

Corollary 2 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let $\alpha : X \to L$ be an L-valued up-set. If the family of cut sets of α is formed by all the crisp up-sets of poset $(X, \leq)$, then $\alpha(x)$ is a completely meet-irreducible element in the lattice L^α for every $x \in X$ with $\alpha(x) \neq 1$.*

In order to illustrate Proposition 5 and Corollaries 1 and 2, we show an example of an L-valued set such that its family of cuts is formed by all the crisp up-sets of its domain.

Example 5 *Let us consider again the L-valued up-set α introduced in Example 1. In Example 4 we have obtained its family of cut sets. It is immediate that this is the family of all up-sets of X. Thus, by Corollaries 1 and 2, we know that α is an order-embedding and $\alpha(x)$ is a completely meet-irreducible element in the lattice $L^\alpha = \{0, p, r, t, 1\}$ for every $x \in X$.*

Now, we suppose that L is a lattice of finite length. Then, combining Proposition 2 and Proposition 5, we obtain the following result.

Theorem 5 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice of finite length and let $\alpha : X \to L$ be an L-valued up-set such that*

1. *α is an bijection to set $M(L)$ of all meet-irreducible elements of lattice L.*
2. *For every family $\{x_i \mid i \in I\}$ of elements from X, if $\alpha(x) \geq \bigwedge_{i \in I} \alpha(x_i)$ for $x \in X$, then $x \geq x_i$, for some $i \in I$.*

Then, the lattice L is isomorphic with the family of all up-sets of the poset X and therefore L is distributive.

Example 6 *Let us consider again the L-valued up-set α introduced in Example 1. Since $\alpha_s = \alpha_q = \emptyset$, it is not true that all its p-cuts are distinct (see Proposition 2) and therefore Condition (1) in the previous theorem does not hold. We can also notice that s is a meet irreducible element in lattice L and it is not a value of the function α. Therefore, we cannot use this example in order to illustrate Theorem 5.*

Let us consider a different poset $(X, \leq)$, a different complete lattice $(L, \leq)$ and a different L-valued up-set α, which are represented in Figure 2.

Given the L-valued set $\alpha : X \to L$, the p-cuts of α are all the up-sets of X:

$$\alpha_0 = \{0, a, b, c, 1\},\ \alpha_n = \{a, b, c, 1\},\ \alpha_p = \{a, b, 1\},$$

$$\alpha_q = \{a, c, 1\},\ \alpha_r = \{b, c, 1\},\ \alpha_s = \{a, 1\}\ \alpha_t = \{b, 1\},$$

$$\alpha_u = \{c, 1\},\ \alpha_v = \{1\},\ \alpha_1 = \emptyset,$$

which are all distinct.

Observe that using the mapping $E(l) = \alpha_l$ for all $l \in L$, that is, if we change l by α_l, the result is the set of all the upsets ordered by $\supseteq$.

ISBN: 978-989-95079-6-8

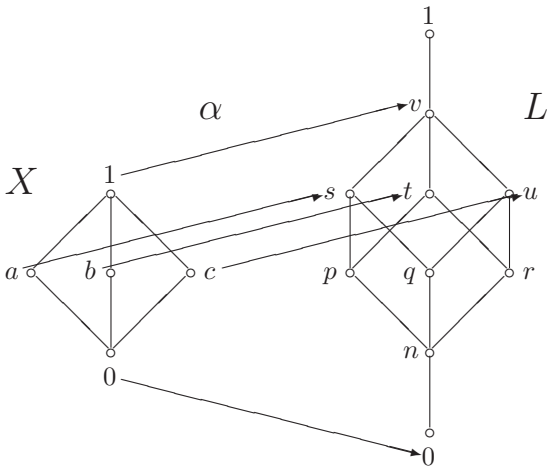


Figure 2: Hasse diagrams of X and L and L-valued set α.

Let us conclude this section with a version of the well-known Birkhoff Representation Theorem: on the one hand, by this theorem, we know that any finite distributive lattice L can be represented by the set of all up-sets on the set of all meet-irreducible elements of L. Now by Theorem 5, we can define an L-valued up-set whose cut sets are order isomorphic to L. Therefore, we can obtain a representation of any distributive lattice as a family of cut-sets of an L-valued up-set. This L-valued up-set is the embedding of $M(L)$ in L, that is, $\alpha : M(L) \to L$ with $\alpha(p) = p, \forall p \in M(L)$.

3.2 Characterization of the family of L-valued down-sets

All the previous studies made for L-valued up-sets could be repeated for L-valued down-sets. By the duality between these two concepts (see, for instance, Proposition 3), the obtained results are totally analogous.

Theorem 6 *Let $(X, \leq)$ be a poset and let $(L, \leq)$ be a complete lattice. A function $\mu : X \to L$ is an L-valued down-set on X if and only if any p-cut μ_p of μ is a crisp down-set (semi-ideal) on X, for every $p \in L$.*

Theorem 7 *Let $(X, \leq)$ be a poset and let $(L, \leq)$ be a complete lattice. If we consider the set $\mathcal{D}_L(X)$ formed by the collection of all L-valued down-sets on X and the induced order on $\mathcal{D}_L(X)$ is defined for any $\mu, \nu \in \mathcal{D}_L(X)$ as follows*

$$\mu \leq \nu \text{ if and only if } \mu(x) \leq \nu(x), \forall x \in X,$$

then the poset $(\mathcal{D}_L(X), \leq)$ is a complete lattice.

Proposition 6 *Let $(X, \leq)$ be a poset and let $(L, \leq)$ be a complete lattice. If a lattice identity holds in L, then the same identity is satisfied in the lattice $(\mathcal{D}_L(X), \leq)$. Thus, for a distributive lattice L, the lattice $(\mathcal{D}_L(X), \leq)$ is also distributive.*

Theorem 8 *Let $(X, \leq)$ be a poset, let $\mathcal{F} \subseteq \mathcal{P}(X)$ be a family of some down-sets of a poset X, and let $(L, \leq)$ be a complete lattice. Then, there is an L-valued down-set $\mu : X \to L$ such that its family of cuts is equal to $\mathcal{F}$ if and only if the following two conditions hold:*

1. *$\mathcal{F}$ is closed under intersections and contains X;*
2. *there is a closure operator C on L, such that the poset $(L/C, \leq)$ is order isomorphic to $(\mathcal{F}, \supseteq)$.*

Proposition 7 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let $\mu : X \to L$ be an L-valued down-set. The family of cut sets of μ is formed by all the crisp down-sets of poset $(X, \leq)$ if and only if for every family $\{x_i \mid i \in I\}$ of elements from X the condition $\mu(x) \geq \bigwedge_{i \in I} \mu(x_i)$ for $x \in X$ implies that $x \leq x_i$ for some $i \in I$.*

Corollary 3 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice and let $\mu : X \to L$ be an L-valued down-set. If the family of cut sets of μ is formed by all the crisp down-sets of poset $(X, \leq)$, then the two following statements hold:*

- *μ is an order-embedding.*
- *For any $x \in X$, $\mu(x)$ is a completely meet-irreducible element in the lattice L^{μ}.*

Theorem 9 *Let $(X, \leq)$ be a poset, let $(L, \leq)$ be a complete lattice of finite length and let $\mu : X \to L$ be an L-valued down-set such that*

1. *μ is an bijection to set $M(L)$ of all meet-irreducible elements of lattice L.*
2. *For every family $\{x_i \mid i \in I\}$ of elements from X, if $\mu(x) \geq \bigwedge_{i \in I} \mu(x_i)$ for $x \in X$, then $x \leq x_i$, for some $i \in I$.*

Then, the lattice L is distributive and L is isomorphic with the family of all down-sets of the poset X.

4 Conclusion

In this work, we have carried out an in-depth study of some particular classes of L-valued sets: up-sets and down-sets. We have started with the L-valued up-sets and we have proven that this is a cutworthy property. Moreover, we have established necessary and sufficient conditions under which for a given family of crisp up-sets, there exists an L-valued set such that its collection of cuts coincides with the family of crisp up-sets. In particular, we have done a deeper study for the L-valued up-sets such that its family of cuts is formed by all the crisp up-sets. In the case of lattices of finite length, we have arrived to the Birkhoff Representation Theorem. All these studies have been repeated for the case of L-valued down-sets and analogous results have been obtained.

In future works we would like to obtain some necessary and sufficient conditions not only for the existence of an L-valued up-set (resp. down-set) for a given family of up-sets (resp. down-sets), but for the uniqueness of such L-valued up-set (resp. down-set). The properties under which uniqueness is guaranteed were already studied for general fuzzy sets in [23]. Now, we try to adapt these results in two senses: L-valued sets instead of fuzzy sets (valued in the interval $[0, 1]$) and up-sets (resp. down-sets) instead of any sets.

Acknowledgements

The research reported on in this paper has been partially supported by Projects FEDER-MEC-MTM2004-01269 and FICYT-PC07-11 and also by Serbian Ministry of Science and

ISBN: 978-989-95079-6-8

Environment, Grant No. 144011 and by the Provincial Secretariat for Science and Technological Development, Autonomous Province of Vojvodina, grant "Lattice methods and applications".

References

[1] J.A. Goguen. *l*-fuzzy sets. *J. Math. Anal. Appl.*, 18:145–174, 1967.

[2] C.V. Negoita and D.A. Ralescu. *Applications of fuzzy sets to system analysis*. Basel, Birkhäuser Verlag, 1975.

[3] R. Bělohlávek. *Fuzzy Relational Systems*. Kluwer Academic Publishers, 2002.

[4] S. Díaz, B.De Baets, and S. Montes. Additive decomposition of fuzzy pre-orders. *Fuzzy Sets and Systems*, 158:830–842, 2007.

[5] S. Díaz, S. Montes, and B.De Baets. Transitivity bounds in additive fuzzy preference structures. *IEEE Trans. Fuzzy Systems*, 15:275–286, 2007.

[6] L.A. Zadeh. Similarity relations and fuzzy orderings. *Inform. Sci.*, pages 177–200, 1971.

[7] B. Borchardt, A. Maletti, B. Šešelja, A. Tepavčević, and H. Vogler. Cut sets as recognizable tree languages. *Fuzzy Sets and Systems*, 157:1560–1571, 2006.

[8] L.M. Brown and E. Ertürk. Fuzzy sets as texture spaces, i. representation theorems. *Fuzzy Sets and Systems*, 110:227–235, 2000.

[9] M. Chechik, B. Devereux, S. Easterbrook, A.Y.C. Lai, and V. Petrovykh. Efficient multiple-valued model-checking using lattice representations. *Lecture Notes in Computer Science*, 2154:451–465, 2001.

[10] Y.M. Liu and M.K. Luo. *Fuzzy Topology*. World Scientific, 1997.

[11] B. Šešelja. *l*-fuzzy covering relation. *Fuzzy Sets and Systems*, 158:2456–2465, 2007.

[12] B. Šešelja and A. Tepavčević. A note on natural equivalence relation on fuzzy power set. *Fuzzy Sets and Systems*, 148:201–210, 2004.

[13] B. Šešelja, A. Tepavčević, and H. Vogler. A note on cut-worthiness of recognizable tree series. *Fuzzy Sets and Systems*, 159:3087–3090, 2008.

[14] A. Tepavčević and G. Trajkovski. *l*-fuzzy lattices: an introduction. *Fuzzy Sets and Systems*, 123:209–216, 2001.

[15] R. Fuentes-González. Down and up operators associated to fuzzy relations and t-norms: A definition of fuzzy semi-ideals. *Fuzzy Sets and Systems*, 117:377–389, 2001.

[16] G. Birkhoff. *Lattice Theory*. Amer. Math. Soc. Providence R.I., 1967.

[17] B.A.Davey and H.A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 1992.

[18] B. Šešelja and A. Tepavčević. Completion of ordered structures by cuts of fuzzy sets, an overview. *Fuzzy Sets and Systems*, 136:1–19, 2003.

[19] B. Šešelja and A. Tepavčević. Representing ordered structures by fuzzy sets, an overview. *Fuzzy Sets and Systems*, 136:21–39, 2003.

[20] M. Gorjanac-Ranitović and A. Tepavčević. General form of lattice-valued fuzzy sets under the cutworthy approach. *Fuzzy Sets and Systems*, pages 1213–1216, 2007.

[21] B. Šešelja and A. Tepavčević. Representation of lattices by fuzzy sets. *Information Sciences*, 79:171–180, 1993.

[22] D. Sinha. A general theory of fuzzy arithmetic. *Fuzzy Sets and Systems*, 36:339–363, 1990.

[23] A. Jaballah and F.B. Saidi. Uniqueness results in the representation of families of sets by fuzzy sets. *Fuzzy Sets and Systems*, 157:964–975, 2006.

ISBN: 978-989-95079-6-8

Fuzzy Intersection and Difference Model for Topological Relations

Ahed ALBOODY[1] Florence SEDES[2] Jordi INGLADA[3]

1 Université Paul Sabatier (UPS)
Toulouse, 118 Route de Narbonne, F-31062-CEDEX 9, France
2 Institut de Recherche en Informatique de Toulouse (IRIT), Université Paul Sabatier (UPS),
Toulouse, 118 Route de Narbonne, F-31062-CEDEX 9, France
3 Centre National d'Etudes Spatiales (CNES)
Toulouse, 18 Avenue Edouard Belin, 31401- CEDEX 9, France
Email: albahed@yahoo.fr, Florence.Sedes@irit.fr, jordi.inglada@cnes.fr

Abstract—*Topological relations have played important roles in spatial query, analysis and reasoning in Geographic Information Systems (GIS) and spatial databases. The topological relations between crisp and fuzzy spatial objects based upon the 9-intersections topological model have been identified. However the formalization of the topological relations between fuzzy regions needs more investigation. The paper provides a theoretical framework for modelling topological relations between fuzzy regions based upon a new fuzzy topological model called the Fuzzy Intersection and Difference (FID) Model. A novel topological model is formalized based on Fuzzy Topological Space (FTS). In order to derive all fuzzy topological relations between two fuzzy spatial objects, the fuzzy spatial object (A) is decomposed in four components: the Interior, the Interior's Boundary, the Object's Boundary, and the Exterior's Boundary of A. By use of this definition of fuzzy spatial object, new 4*4-Intersection and Fuzzy Intersection and Difference (FID) models are proposed as a qualitative model for the identification of all topological relations between two simple fuzzy regions. These two new models are compared with other fuzzy models studied in the literature. Examples are provided to illustrate the use of these two models presented in this paper with results which can be applied for modeling GIS and geospatial databases.*

Keywords— Fuzzy Intersection and Difference Model, 4*4-Intersection Matrix, Fuzzy Objects, Topological Relations, GIS

1 Introduction

Topological relations have an important significance in GIS modelling since they are the basis for spatial modelling, spatial query, analysis and reasoning. How to identify the topological relations between spatial objects is a critical point in GIS modelling.

During recent years, topological relations have been much investigated in the crisp and fuzzy topological space. The well-known *4-intersection* approach described in [1, 2], as well as the *9-intersection* approach as discussed in [3], and the *Intersection and Difference* (*ID*) model studied in [4, 5, 6], were proposed to formalize topological relations between two simple regions in the *Crisp Topological Space* (*CTS*).

The *4-intersection* model is extended in [7] to deal with the topological relations between spatial objects with holes.

Geographical phenomena in GIS with uncertain boundaries can be modelled by *Regions with Broad Boundaries* (*BBRs*) as in [8]. A region with broad boundary is an extension of a region with a crisp boundary (refer to simple regions with holes as in [7]). Objects with broad boundaries as defined in [8] are spatial objects, whose crisp boundaries are replaced by an area expressing the boundary's uncertainty. The *9-intersection model* is extended in [9] to describe topological relations between *BBRs* by replacing the crisp boundary in the *9-intersection* with the *broad boundary*. Another method called *4-tuple representation* of topological relations between *BBRs* is used in [10] to infer new topological information. The *4-tuple representation* can distinguish the same topological relations as identified by the *extended 9-intersection*. The *4-tuple*, however, can be applied to the reasoning of topological relations between *BBRs* [10], because it uses the composition of topological relations between crisp regions to determine those between uncertain and vague regions.

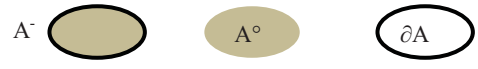
More recently fuzzy spatial objects have been emphasized since there are spatial features which are not always crisp. Fuzzy spatial objects are those with indeterminate boundaries. For fuzzy boundaries, that is, boundaries that are by nature not crisp, the broad boundary represents their minimum and maximum extent.

In order to derive the topological relations between fuzzy spatial objects, the *9-intersection approach* was updated into the *3*3-intersection approach* in the fuzzy topological space [11, 12]. Furthermore, in these two works, a *4*4-intersection matrix* was built up by using the topological properties of fuzzy sets, and then a *5*5-intersection matrix* can be built up based on certain conditions.

In the next section, we compare these models.

2 Related Work about Topological Relation Models and Spatial Objects

Crisp spatial objects have been formally defined in GIS. Point, line and polygon are three primitives in GIS. Fig. 1 represents the closure, interior and boundary of a closed disk as crisp spatial objects [2]. The *4-intersection* and *9-intersection matrix* are well-known approaches to identifying topological relation models between these two crisp spatial objects using the concept of the interior, boundary and exterior.



Closure Interior Boundary

Figure 1: Closure, Interior and Boundary of a Crisp Object

By using some topological invariants of the intersection such as the empty/non empty contents, the topological

ISBN: 978-989-95079-6-8

relations between two crisp spatial objects can be identified. This approach implies the following facts in *CTS*: (1) the interior, boundary and the exterior of a subset are topological invariants; (2) these topological invariants are mutually disjoint in *CTS*; and (3) the empty/non-empty contents of the intersections between these three topological parts of two subsets are topological invariants. Then the *4-intersection* and *9-intersection models* are defined as:

$$I_4(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B \\ \partial A \cap B^\circ & \partial A \cap \partial B \end{bmatrix}$$

$$I_9(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^- \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^\circ & A^- \cap \partial B & A^- \cap B^- \end{bmatrix}$$

Eight topological relations of the spatial reasoning system (Region Connection Calculus) *RCC8* (*DC*, *EC*, *EQ*, *PO*, *TPP*, *TPPi*, *NTPP* and *NTPPi*) have been identified between two simple regions by using these two models in [2, 3].

In *ID model* [4, 5, 6], a crisp spatial object is defined by its interior and boundary; two intersection sets are $A^\circ \cap B^\circ$ and $\partial A \cap \partial B$; two difference sets are A - B and B - A. This model can also distinguish the eight topological relations. This crisp topological model is represented by:

$$ID(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & A-B \\ B-A & \partial A \cap \partial B \end{bmatrix}$$

The main difference between the *4-intersection model* and *ID model* is that both intersection sets $A^\circ \cap \partial B$ and $\partial A \cap B^\circ$ of 4-intersection model are replaced by two differences A - B and B - A.

However, the fact (2) in CTS cannot hold in Fuzzy Topological Space (FTS). That means the interior; the boundary and the exterior of a fuzzy set may not be disjoint with each other. Therefore the *4-intersection, 9-intersection* and *ID models* cannot be directly applied for the identification of relations between two fuzzy sets.

The *9-intersection model* is extended in [8, 9] to spatial objects with broad boundary as simple fuzzy regions. It's expressed by the following matrix:

$$M = I_9(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \Delta B & A^\circ \cap B^- \\ \Delta A \cap B^\circ & \Delta A \cap \Delta B & \Delta A \cap B^- \\ A^- \cap B^\circ & A^- \cap \Delta B & A^- \cap B^- \end{bmatrix}$$

Using this model, *44 relations* between two simple fuzzy regions by using the *3*3-intersection matrix* are possible. For composite regions with broad boundaries, there are *14 additional* topological relations [8]. Fig. 2 represents a region with broad boundary as simple fuzzy spatial object.

Figure 2: Region with a Broad Boundary

In [14], they investigated a special space and formalized the 9-intersection in *Crisp FTS*. They proved that the *Crisp FTS* whose open sets are crisp is able to meet the above conditions.

The *9-intersection matrix* can be formalized as:

$$I_9(A,B)=\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^e \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^e \\ A^e \cap B^\circ & A^e \cap \partial B & A^e \cap B^e \end{bmatrix}$$

The above *3*3-intersection matrix* is derived based on the interior, boundary, and exterior of a simple fuzzy spatial object. By use of *9-intersection matrix*, *44 topological relations* are identified between two simple fuzzy regions (See Appendix 1. in [14]). Fig. 3 represents a region with indeterminate boundary as simple fuzzy spatial object.

Figure 3: Two Fuzzy Spatial Objects

In [14], it was shown that fuzzy spatial objects can be decomposed into four parts: the interior, the boundary of the boundary $\partial(\partial A)$, the interior of the boundary $(\partial A)^\circ$ and the exterior, which are mutually disjoint as in Fig. 4.

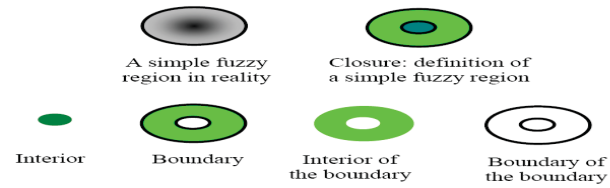


Figure 4: Interior, Boundary, Interior of the Boundary and Boundary of the Boundary of a Simple Fuzzy Region (After X. Tang and W. Kainz in [14])

Therefore, they introduced a *4*4-Intersection matrix* between two simple fuzzy spatial objects as follows:

$$I_{4*4}=\begin{bmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap (\partial B)^\circ & A^\circ \cap B^e \\ \partial(\partial A) \cap B^\circ & \partial(\partial A) \cap \partial B & \partial(\partial A) \cap (\partial B)^\circ & \partial(\partial A) \cap B^e \\ (\partial A)^\circ \cap B^\circ & (\partial A)^\circ \cap \partial B & (\partial A)^\circ \cap (\partial B)^\circ & (\partial A)^\circ \cap B^e \\ A^e \cap B^\circ & A^e \cap \partial B & A^e \cap (\partial B)^\circ & A^e \cap B^e \end{bmatrix}$$

Under certain conditions, *152 relations* are identified by using the *4*4-intersection* approach (See Appendix 2. in [14] for more details).

After investigation about the topological relations between two simple fuzzy regions compared with these models studied in the literature, we can see that some topological relations can't be identified by these models. Here are some relations presented in Fig. 5.

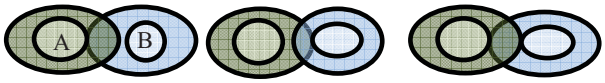


Figure 5: Some of Topological Relations to Identify

The question is how many topological relations there are exactly between two simple fuzzy objects? To answer this question, we will extend the *4-intersection* and the *ID models* with a new definition of the fuzzy boundary for fuzzy regions.

The disadvantage of the *3*3-intersection* and *4*4-intersection* models is that the *intersection operator* ($\cap$) is the most expensive one in terms of computation. In order to reduce the computational cost of the *3*3-intersection* and *4*4-intersection* models; and to reduce the computational complexity by avoiding spatial operations between

ISBN: 978-989-95079-6-8

topological components with different dimensions (1-D and 2-D), we will try to reduce the number of intersections by introducing the *difference operator* (-).

Finally, the motivation of the paper is trying to build a topological model for identification of all fuzzy topological relations between two fuzzy regions.

The structure of the paper is as follows. The fuzzy topological space is defined in section 3. Section 4 is our contribution which consists of a new definition of simple fuzzy regions, fuzzy boundaries and their properties. The novel major contribution of our study, proposed in section 5, is a new form of *4*4- intersection model* and *the Fuzzy Intersection and Difference (FID) model*. Section 6 shows the identification of fuzzy topological relations between two simple fuzzy regions by using two models (*4*4- intersection and FID*) based on empty/non-empty contents. We end with a discussion about results and a conclusion.

3 Fuzzy Topological Space (FTS)

Fuzzy topology is constructed based on fuzzy sets. It is an extension of general (crisp) topology.

Let A be a fuzzy subset of an ordinary (crisp) set X, and $\wp(X)=[0,1]^X$ be the fuzzy power set of X. $[0,1]^X$ can be viewed as a lattice in which a supremum (or *join*) is denoted by $\vee$ and an infimum (or *meet*) by $\wedge$, conventionally. They correspond to the *union* and the *intersection*, respectively. $\forall\delta\in\wp(X)$ if $(1)\,\Phi, X\subseteq\delta$, $(2)\,\forall A_i\in\delta,\ \vee A_i\in\delta,(3)$ $\forall U,V\in\delta,\ U\wedge V\in\delta$, then δ is called a *fuzzy topology* on *X* (*i*∈*I* is an index set). (*X, δ*) is called a *Fuzzy Topological Space (FTS)* as defined in [15, 16]. Every element of δ is called an open (fuzzy) set in (X, δ). A set A is a closed (fuzzy) set if its complement A^c is open. The union of all open sets contained in A is the interior of A, denoted by A°. The intersection of all closed sets containing A is called the closure of A, denoted by A^-. The exterior of A is the complement of A^- and is denoted by A^e. Obviously, it is an open set. The boundary ∂A of a subset A is the intersection of the closure of A with the closure of the complement of A.

The boundary of a subset may also have its interior and its boundary of the boundary. On the other hand, the interior and the closure of a subset also have their boundaries. For example, the boundary of the boundary of a fuzzy set A is the union of the boundary of the closure and the boundary of the interior of a fuzzy set [14].

Based on this information, one can define a maximum of four other different areas from that defined in [14] for each object: Interior, Boundary of the Object, Interior of the Boundary, and Exterior of the Boundary in the next section.

4 Definition of Simple Fuzzy Regions and a Fuzzy Boundary

In this section, we will develop a definition of simple fuzzy region, fuzzy boundary and their properties.

4.1 Definition of Simple Fuzzy Region

A crisp region is defined in *CTS*. Correspondingly, a fuzzy region should be defined in *FTS*. We now define a simple fuzzy region in *FTS*.

A simple fuzzy region is made up of two regions A_1 and A_2 with $A_1\subset A_2$ (see Fig. 6), where: (1) the interior of A is the interior of A_1 $A^\circ=(A_1)^\circ$ and A° is an open subset and connected; (2) the interior's boundary of A is the boundary of A_1 as $A^i=\partial(A^\circ)=\partial A_1$, and A^i is a closed subset and connected; (3) the boundary of A is ∂A defined as the interior of the difference between A_1 and A_2 as $\partial A=(A_2-A_1)^\circ$, and ∂A is an open subset and connected; (4) the exterior's boundary of A is the boundary of ∂A_2 as $A^e=\partial^e(\partial A)=\partial A_2$, and A^e is a closed subset and connected; and (5) the intersection of all closed sets containing A is called the closure of A, denoted by A^-. Fig. 6 shows the four components of simple fuzzy regions.

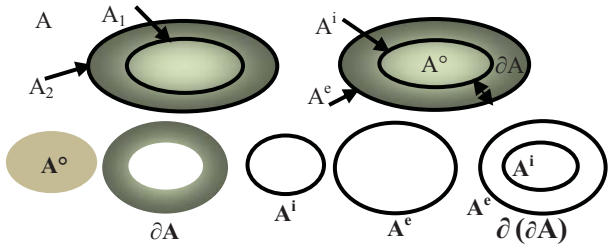


Figure 6: Interior, Boundary, Interior's Boundary, and Exterior's Boundary of a Simple Fuzzy Region

This definition is considered as the de decomposition of the boundary in [14, 15, 16] into disjoint subsets such as the interior boundary of the boundary, the exterior boundary of the boundary and the interior of the boundary with condition that is the interior of the boundary couldn't be a non-empty set.

We called the boundary (∂A) of A by the fuzzy boundary. This definition of fuzzy regions is very interesting to identify all topological relations between two simple fuzzy regions that other models [8, 9, 14] can't identify. We will prove that in the next section by developing a *4*4-intersection* matrix. In the next part, we will define the properties of the fuzzy boundary.

4.2 Properties of Fuzzy Boundary

Let A be a fuzzy set in *FTS* (X, δ). Based on the definition of simple fuzzy regions, we find the properties of the fuzzy boundary (∂A) as follows: (1) ∂A is an open subset of A; (2) ∂A = ¬ Ø is a non-empty subset; (3) the boundary of A (∂A) is the interior of the difference $\partial A=(A^e-A^i)^\circ$ between the exterior's boundary (A^e) and the interior's boundary (A^i) of A; (4) the union $A^i\cup\partial A\cup A^e$ of the interior's boundary (A^i), the boundary (∂A) and the exterior's boundary (A^e) of A is a closed subset of A; (5) the interior boundary of the boundary (∂A) of A $\partial^i(\partial A)=A^i=\partial(A^\circ)$ is the interior's boundary (A^i); and (6) the exterior boundary of the boundary (∂A) of A $\partial^e(\partial A)=A^e=\partial^e(A)$ is the exterior's boundary (A^e).

We also find that the intersections between A^o, A^i, ∂A, A^e are, respectively, always empty, and the union of these parts

ISBN: 978-989-95079-6-8

is equal to A as follows: $A^{\circ} \cap A^{i} = \emptyset$; $A^{i} \cap \partial A = \emptyset$; $\partial A \cap A^{e} = \emptyset$; $A^{\circ} \cap A^{i} \cap \partial A \cap A^{e} = \emptyset$; and $A^{\circ} \cup A^{i} \cup \partial A \cup A^{e} = A$.

It also can be easily proven by the above intersections that the interior (A^{o}), the interior's boundary (A^{i}), the boundary ∂A and the exterior's boundary (A^{e}) of a simple fuzzy region (A) are mutually disjoint.

In order to identify all possible topological relations, the condition of the mutual disjointness of these four parts of simple fuzzy regions is important to propose and construct a new method to form the intersection matrix. In the next section, a new *4*4-intersection matrix* and *fuzzy intersection and difference model* are proposed based upon this definition of simple fuzzy region.

5 Contributions: 4*4 - Intersection and Fuzzy Intersection and Difference (FID) Models

In this section, we will develop two models to identify the topological relations between two simple fuzzy regions.

Supposing there are two simple fuzzy objects A and B in the *FTS*, we adopt the interior, boundary, interior's boundary, and exterior's boundary to formalize two new topological models as in the next.

*5.1 Contribution 1: 4*4 - Intersection Model*

In the first contribution, the first model is a new *4*4 - intersection matrix* which uses the operator (∩) of intersection. Between these two simple fuzzy spatial regions A and B, the *4*4-intersection matrix* will be as presented in Table 1:

Table 1: 4*4 - Intersection Matrix

∩	B°	B^{i}	∂B	B^{e}
A°	A°∩ B°	A°∩ B^{i}	A°∩ ∂B	A°∩ B^{e}
A^{i}	A^{i} ∩B°	A^{i} ∩ B^{i}	A^{i} ∩∂B	A^{i} ∩ B^{e}
∂A	∂A ∩B°	∂A ∩ B^{i}	∂A ∩∂B	∂A ∩ B^{e}
A^{e}	A^{e} ∩B°	A^{e} ∩ B^{i}	A^{e} ∩ ∂B	A^{e} ∩ B^{e}

And the *4*4-inetrsection model* applied to simple fuzzy objects is expressed by the following expression:

$$I_{4*4}(A,B) = \begin{bmatrix} A^{\circ} \cap B^{\circ} & A^{\circ} \cap B^{i} & A^{\circ} \cap \partial B & A^{\circ} \cap B^{e} \\ A^{i} \cap B^{\circ} & A^{i} \cap B^{i} & A^{i} \cap \partial B & A^{i} \cap B^{e} \\ \partial A \cap B^{\circ} & \partial A \cap B^{i} & \partial A \cap \partial B & \partial A \cap B^{e} \\ A^{e} \cap B^{\circ} & A^{e} \cap B^{i} & A^{e} \cap \partial B & A^{e} \cap B^{e} \end{bmatrix}$$

This new *4*4- intersection matrix* (I_{4*4}) is considered as an extension of the *4-intersection* model [1, 2] for simple fuzzy spatial regions.

The *intersection operator* (∩) is perhaps the most expensive one in terms of computation. In order to reduce the computational cost of this *4*4-intersection* model, we will extend this model to the fuzzy intersection and difference model in the next part.

5.2 Contribution 2: Fuzzy Intersection and Difference (FID) Model

In this model, we will introduce the *difference operator* (-). In order to avoid spatial operations between topological components with different dimensions (A°, B°, ∂A, ∂B as 2-D; and A^{i}, B^{i}, A^{e}, B^{e} as 1-D), we will replace the intersection between the terms A°∩ B^{i}, A°∩ B^{e}, A^{i} ∩B°, A^{i} ∩∂B, ∂A ∩ B^{i}, ∂A ∩ B^{e}, A^{e} ∩B° and A^{e} ∩ ∂B in the *4*4-intersection* model by the differences as in the matrix (see Table 2).

The four intersections (A°∩B°, A°∩∂B, ∂A∩B°, ∂A∩∂B) with topological components with dimension 2-D and the four intersections (A^{i}∩B^{i}, A^{i}∩B^{e}, A^{e}∩B^{i}, A^{e}∩B^{e}) with topological components with dimension 1-D remain unchanged as similar as in [4, 5, 6] for the ID model.

Table 2: 4*4 - Intersection and Difference Matrix

∩	B°	B^{i}	∂B	B^{e}	-
A°	A°∩ B°	A^{i} - B^{i}	A°∩ ∂B	A^{i} - B^{e}	A°
A^{i}	B^{i} - A^{i}	A^{i} ∩ B^{i}	B^{i} - A^{i}	A^{i} ∩ B^{e}	A^{i}
∂A	∂A ∩B°	A^{e} - B^{i}	B^{e} – A^{i} ∂A ∩∂B	A^{e} - B^{e}	∂A
A^{e}	B^{i} - A^{e}	A^{i} - B^{i} A^{e} ∩ B^{i}	B^{i} – A^{e} B^{e} - A^{e}	A^{i} - B^{e} A^{e} ∩ B^{e}	A^{e}

By simplification and arrangement of this *4*4 - Intersection and Difference matrix*, we obtain two intersection matrices and two difference matrices (Table 3):

Table 3: Intersection and Difference Matrices

∩	B°	∂B
A°	A°∩ B°	A°∩ ∂B
∂A	∂A ∩B°	∂A ∩∂B

∩	B^{i}	B^{e}
A^{i}	A^{i}∩ B^{i}	A^{i}∩ B^{e}
A^{e}	A^{e} ∩ B^{i}	A^{e} ∩ B^{e}

-	B^{i}	B^{e}
A^{i}	A^{i} - B^{i}	A^{i} - B^{e}
A^{e}	A^{e} - B^{i}	A^{e} - B^{e}

-	A^{i}	A^{e}
B^{i}	B^{i} - A^{i}	B^{i} - A^{e}
B^{e}	B^{e} - A^{i}	B^{e} - A^{e}

At the end, the *Fuzzy Intersection and Difference (FID) model* is written as follows:

$$FID_{4*4}(A,B) = \begin{bmatrix} A^{\circ} \cap B^{\circ} & A^{\circ} \cap \partial B & A^{i} - B^{i} & A^{i} - B^{e} \\ \partial A \cap B^{\circ} & \partial A \cap \partial B & A^{e} - B^{i} & A^{e} - B^{e} \\ A^{i} \cap B^{i} & A^{i} \cap B^{e} & B^{i} - A^{i} & B^{i} - A^{e} \\ A^{e} \cap B^{i} & A^{e} \cap B^{e} & B^{e} - A^{i} & B^{e} - A^{e} \end{bmatrix}$$

The *FID* model combines two different operators (intersection and difference). The *FID* has two advantages: first, it reduces the computational complexity by avoiding spatial operations between topological components with different dimensions, e.g., A°∩ B^{i}, A°∩ B^{e}, A^{i} ∩B°, A^{i} ∩∂B, ∂A ∩ B^{i}, ∂A ∩ B^{e}, A^{e} ∩B° and A^{e} ∩ ∂B, with A°, B°, ∂A, ∂B as 2-D, and A^{i}, B^{i}, A^{e}, B^{e} as 1-D; and second, it reduces the computational cost due to only eight intersections in the matrix of *FID* model.

This *FID* model is considered as an extension of the *ID* model [4, 5, 6] for simple fuzzy spatial regions.

In general, there are 2^{16}= 65536 relations between two fuzzy regions by using the *4*4-intersection* matrix and *FID* model. For GIS applications, some conditions will limit the number of these relations as in [7, 8, 14]. However, how to find all possible topological relations between two simple fuzzy regions needs more investigation. It's done in the next section for the *4*4-intersection* matrix and *FID* model.

ISBN: 978-989-95079-6-8

6 Identification of Topological Relations based upon 4*4 - Intersection and FID Models

In this section, we focus on the identification of all fuzzy topological relations between two simple fuzzy regions by *4*4-intersection* and *FID* models.

Let A and B be two simple fuzzy regions. For relation identification, each intersection or difference in *4*4-intersection* and *FID matrices* takes value of either empty ($\emptyset$) or non-empty ($\neg\emptyset$). Every different set of *4*4-intersection* and *FID* matrices describes a different topological relation. Some values of these two matrices have no sense on the topological relation.

*6.1 Identification by 4*4 - Intersection Model*

For identification by *4*4-intersection model*, by respecting the definition in section 3, we scan all possible configurations for A and B in two different steps as following: (1) If the exterior's boundary of A intersects with the exterior's boundary of B ($A^e \cap B^e = \neg\emptyset$), then, we search all possible topological relations between A and B, we find *105 relations*; (2) If the exterior's boundary of A doesn't intersect with the exterior's boundary of B ($A^e \cap B^e = \emptyset$), then, we search all topological relations between A and B, we find *47 relations*. The total number of topological relations identified between A and B is *152 relations*. Some of these topological relations are not identified and determined in [8, 9, 14]. Here are some examples in Fig .7 and Fig .8.

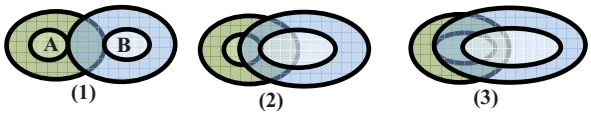


Figure 7: Examples for $A^e \cap B^e = \neg\emptyset$

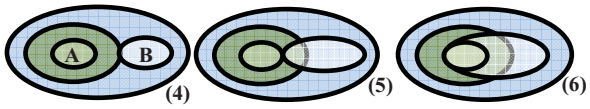


Figure 8: Examples for $A^e \cap B^e = \emptyset$

The *4*4-intersection matrices* correspondent for (1), (2) and (3) in Fig .7, and for (4), (5) and (6) in Fig .8 are given, respectively, by:

$$(1)\begin{bmatrix}\phi & \phi & \phi & \phi\\ \phi & \phi & \phi & \neg\phi\\ \phi & \phi & \neg\phi & \neg\phi\\ \phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}\ (2)\begin{bmatrix}\phi & \phi & \neg\phi & \neg\phi\\ \phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}\ (3)\begin{bmatrix}\neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}$$

$$(4)\begin{bmatrix}\phi & \phi & \neg\phi & \phi\\ \phi & \phi & \neg\phi & \phi\\ \phi & \phi & \neg\phi & \phi\\ \phi & \neg\phi & \neg\phi & \phi\end{bmatrix}\ (5)\begin{bmatrix}\phi & \phi & \neg\phi & \phi\\ \phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\end{bmatrix}\ (6)\begin{bmatrix}\neg\phi & \phi & \phi & \phi\\ \neg\phi & \neg\phi & \phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\end{bmatrix}$$

We note that the separateness between these relations can't be realized by other models studied in [8, 14]. In the next part, we will identify all relations by the *FID model*.

6.2 Identification by FID Model

For identification by *FID model*, by respecting the definition in section 3, we apply the same two steps in previous part (6.1) for $A^e \cap B^e = \neg\emptyset$ and $A^e \cap B^e = \emptyset$. 152 topological relations can be identified by using the *FID model*. These relations identified by *FID model* is the same relations by *4*4-intersection model*. We give some examples in Fig .9 and Fig .10.

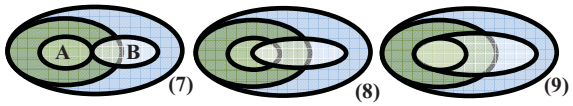


Figure 9: Examples for $A^e \cap B^e = \neg\emptyset$

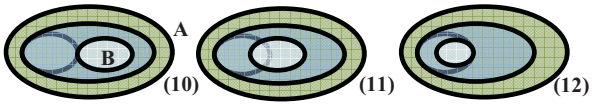


Figure 10: Examples for $A^e \cap B^e = \emptyset$

The *FID* matrices correspondent to (7), (8) and (9) in Fig .9, and to (10), (11) and (12) in Fig .10 are given, respectively, by:

$$(7)\begin{bmatrix}\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}\ (8)\begin{bmatrix}\neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}\ (9)\begin{bmatrix}\neg\phi & \phi & \phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\end{bmatrix}$$

$$(10)\begin{bmatrix}\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \phi & \phi & \neg\phi & \phi\end{bmatrix}\ (11)\begin{bmatrix}\neg\phi & \neg\phi & \neg\phi & \phi\\ \neg\phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \neg\phi & \phi\\ \phi & \phi & \neg\phi & \phi\end{bmatrix}\ (12)\begin{bmatrix}\neg\phi & \neg\phi & \neg\phi & \phi\\ \phi & \neg\phi & \neg\phi & \neg\phi\\ \neg\phi & \neg\phi & \phi & \phi\\ \phi & \phi & \neg\phi & \phi\end{bmatrix}$$

To find and extract these relations by the *4*4-intersection* and *FID models*, we have developed and implemented these two steps on MATLAB.

7 Applications and Discussion

For GIS applications, satellite images and spatial databases, these two models (*4*4-intersection* and *FID*) can determine the topological relations between simple fuzzy regions. For that, we need to know how to generate fuzzy spatial objects from satellite images and to find the four components of our definition (interior, interior's boundary, boundary and exterior's boundary) for satellite images and GIS objects. We can adopt processed data such as classification or segmentation results of satellite images. For example, Land Use and Land Cover (LULC), most of which is obtained from the classification results of satellite images, may be a good example of a fuzzy spatial object as in [17]. In principle, a fuzzy spatial object can also be generated by other methods as in [14, 18].

These two models can be used in order to evaluate the change detection process (for Land Cover changes)of geographical objects (Beach, Forest, Residential Area…) represented in GIS and satellite imagery (TM and SPOT images) databases as in [14, 17, 19, 20].

Comparing between the relations identified in [8, 9, 14] and our models (*4*4-intersection and FID*), we find that there are *100 new relations* which can't be discriminated and identified by other models. Some of these *100 new topological relations* identified by *4*4-intersection and FID* models are presented in the appendix of this paper.

8 Conclusion

In this paper, we have proposed a new definition of simple fuzzy spatial region by decomposing the spatial region into

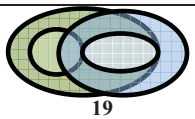
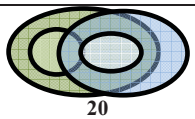
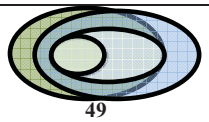
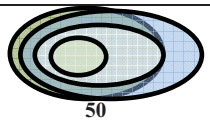
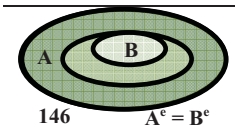
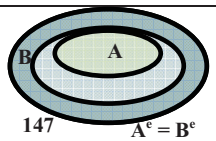
ISBN: 978-989-95079-6-8

four components: interior, interior's boundary, boundary and exterior's boundary. Based upon these four components, a *4*4-intersection matrix* is introduced to identify all topological relations between two simple fuzzy regions. Then, in order to reduce the computational complexity of the *4*4-intersection model,* the *Fuzzy Intersection and Difference (FID) model* is developed based on the *4*4-intersection matrix*. The main contribution of this work is these two models *4*4-intersection and FID. 152* fuzzy topological relations can be identified by using the *4*4-intersection and FID models*. Among these *152 relations*, *100 new relations* can't be discriminated or identified as different relations by other models [8, 9, 14].

In our future work, we will try to find more topological relations which can be identified if we adopt in the definition $A^e \cap A^i = \neg\emptyset$, $\partial A = \neg\emptyset$ and A^i don't intersect with the exterior of A. Then, we will try to classify these fuzzy topological relations identified by I4*4 and FID models to be grouped into the eight relations DC, EC, PO, TPP and TPPi, NTPP and NTPPi, and EQ of the spatial reasoning system *RCC8*.

Appendix

Some Topological Relations between Two Simple Fuzzy Regions by Using the *4*4-Intersection Model (I_{4*4} Matrix)* and the *Fuzzy Intersection and Difference (FID Matrix) Model*

Illustration	I_{4*4} Matrix	FID Matrix	Illustration	I_{4*4} Matrix	FID Matrix
 19	(19) $\begin{bmatrix} \phi & \phi & \neg\phi & \neg\phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$	(19) $\begin{bmatrix} \phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$	 20	(20) $\begin{bmatrix} \phi & \phi & \neg\phi & \neg\phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \phi & \phi & \neg\phi & \neg\phi \end{bmatrix}$	(20) $\begin{bmatrix} \phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \neg\phi & \neg\phi & \phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$
 49	(49) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$	(49) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \neg\phi & \phi & \neg\phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$	 50	(50) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$	(50) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \\ \phi & \phi & \neg\phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \neg\phi \end{bmatrix}$
 146 $A^e = B^e$	(146) $\begin{bmatrix} \neg\phi & \phi & \neg\phi & \phi \\ \phi & \neg\phi & \neg\phi & \phi \\ \phi & \phi & \neg\phi & \phi \\ \phi & \phi & \phi & \neg\phi \end{bmatrix}$	(146) $\begin{bmatrix} \neg\phi & \neg\phi & \neg\phi & \phi \\ \phi & \neg\phi & \neg\phi & \phi \\ \neg\phi & \phi & \phi & \phi \\ \phi & \neg\phi & \neg\phi & \phi \end{bmatrix}$	 147 $A^e = B^e$	(147) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \phi \\ \phi & \phi & \phi & \neg\phi \end{bmatrix}$	(147) $\begin{bmatrix} \neg\phi & \phi & \phi & \phi \\ \neg\phi & \neg\phi & \neg\phi & \phi \\ \neg\phi & \phi & \neg\phi & \phi \\ \phi & \neg\phi & \neg\phi & \phi \end{bmatrix}$

References

[1] M.J. Egenhofer, A formal definition of binary topological relationships, *The Third International Conference of Foundations of Data Organization and Algorithms (FODO)*, Paris, France, Springer-Verlag, pp. 452-472, 1989.

[2] M.J. Egenhofer, and R. Franzosa, Point-set topological spatial relations, *International Journal of Geographic Information Systems,* 5(2): 161-174, 1991.

[3] M.J. Egenhofer, and J.R. Herring, A mathematical framework for the definition of topological relationships, *Proceedings of the 4th International Symposium on Spatial Data Handling,* Zurich, Columbus, OH, International Geographical Union, pp. 803-813, 1990.

[4] M. Deng, T. Cheng, X. Chen and Z. Li, Multi-level topological relations between spatial regions based upon topological invariants, *Springer Science, GeoInformatica*, 11(2): 239-267, 2007.

[5] A. Alboody, J. Inglada and F. Sèdes, Enriching the spatial reasoning system RCC8, *16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS 2008)*, *The SIGSPATIAL Special,* 1(1): 14-20, 2009.

[6] A. Alboody, F. Sèdes and J. Inglada, Multi-level topological relations of the spatial reasoning system RCC-8, *The First International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA'09), IEEE Computer Society*, pp. 13–21, 2009.

[7] M.J. Egenhofer, E. Clementini and P. D. Felice, Topological relations between regions with holes, *International Journal of Geographic Information Systems*, 8(2): 129-144, 1994.

[8] E. Clementini, Objects with broad boundaries, *Book of S. Shekhar and H. Xiong: Encyclopedia of GIS*, Springer US, pp. 793-799, 2008.

[9] E. Clementini and P. Di Felice, An algebraic model for spatial objects with indeterminate boundaries, *P. Burrough, A. Frank (Eds.), Geographic Objects with Indeterminate Boundaries*, Taylor & Francis, London, pp. 155–169, 1996.

[10] S. Du, Q. Qin, Q. Wang and H. Ma, Reasoning about topological relations between regions with broad boundaries, *International Journal of Approximate Reasoning,* 47(2): 219-232, 2008.

[11] X. Tang, Y. Fang and W. Kainz, Fuzzy topological relations between fuzzy spatial objects, *LNCS in Springer Berlin, Fuzzy Systems and Knowledge Discovery*, 4223: 324-333, 2006.

[12] X. Tang and W. Kainz, Analysis of topological relations between fuzzy regions in a general fuzzy topological space, *Proceedings of Symposium on Geospatial Theory, Processing and Applications,* Canada, Ottawa, 2002.

[13] X. Tang, Y. Fang, and W. Kainz, Topological matrices for topological relations between fuzzy regions, *Proceedings of the 4th International Symposium on Multispectral Image Processing and Pattern Recognition (SPIE),* 6045: 604524, Wuhan, China, 2005.

[14] X. Tang and W. Kainz, Spatial object modelling in fuzzy topological spaces with applications to Land Cover change, *Ph.D Thesis, 9 January* 2004. www.itc.nl/library/Papers_2004/phd/xinming.pdf

[15] C. L. Chang, Fuzzy topological spaces, *Journal of Math. Anal. Appl*. 24: 182–190, 1968.

[16] X. Tang, W. Kainz and H. Zhang, Some topological invariants and a qualitative topological relation model between fuzzy regions, *4th International Conference on Fuzzy Systems and Knowledge Discovery, IEEE*, 1(24-27): 241- 246, 2007.

[17] A. Bassiri, A. Alesheikh and M. R. Malek, Spatio-temporal object modelling in fuzzy topological space, *The International Archives of the Photogrammetric, Remote Sensing and Spatial Information Sciences,* XXXVII (B2): 131-134, 2008.

[18] W. Shi and K. Liu, A fuzzy topology for computing the interior, boundary, and exterior of spatial objects quantitatively in GIS, *Computers & Geosciences, Elsevier,* 33: 898–915, 2007.

[19] T. Cheng, Fuzzy objects: their changes and uncertainties, *The Journal of the American Society for Photogrammetric Engineering and Remote Sensing*, 2 (1): 41 - 49, 2002.

[20] T. Cheng, P. Fisher and Z. Li, Double vagueness: uncertainty in multi-scale fuzzy assignments of duneness, *The Revue of Geo-Spatial Information Science*, *Springer*, 7(1): 58-66, 2004.

ISBN: 978-989-95079-6-8

T-norm-based fuzzy logics and logics for reasoning under vagueness

Thomas Vetterlein, Klaus-Peter Adlassnig

Institute for Medical Expert and Knowledge-Based Systems, Medical University of Vienna
Vienna, Austria
Email: {Thomas.Vetterlein,Klaus-Peter.Adlassnig}@meduniwien.ac.at

Abstract— *We contrast the concept underlying t-norm-based propositional fuzzy logics with the problem to whose solution fuzzy logics are frequently suggested as helpful – namely, to find a model of reasoning with vague information. We argue that fuzzy logics are useful as long as truth values can be identified with the meaning of the considered propositions. This, however, is rarely the case in practice; hence we see the need to broaden the concept underlying this important class of logics and try fresh approaches. In particular, we should flexibilise the formalism to allow that propositions do not arise in the same context, but are just known to be related in some way.*

We tackle the problem tentatively. We define a set of rules which, as we assume, are minimally required to enable us to argue about vague propositions whose content is not taken into account. Our choice of rules reflects the practical requirements of a certain expert system on which we work.

Although we deal here with fuzzy logic in a very direct sense, we arrive at calculi completely different from the t-norm-based ones. Without incorporating truth degrees explicitly, we are led to Belnap's logic, which can, but need not, be endowed with a semantics based on graded truth degrees. When formalising also truth degrees, we get a logic which can be based on what we call metric De-Morgan lattices.

Keywords— t-norm-based fuzzy logics, reasoning under vagueness, medical expert system, De-Morgan lattices, metric De-Morgan lattices

1 T-norm-based fuzzy logic for reasoning about vague information – a trap

Fuzzy logics are distinguished from classical logic by the incorporation of an extended set of truth values. In the standard case, the value 0 is used to express falsity, the value 1 is used to express trueness, and all remaining real values in between these two limit points are added in order to cope with the fact that objects may fulfil a property to an intermediate degree. We arrive naturally at the idea to evaluate propositions in the real unit interval, whose most basic feature is its linear order.

The connectives used in fuzzy logics consequently need to be interpreted by operations on $[0, 1]$. Typically, a conjunction is present, which is typically interpreted by a left-continuous t-norm. Sometimes, an involutive negation is present as well, conveniently interpreted by the standard negation $1 - \cdot$. A further connective, which logicians, in contrast to engineers, generally consider as the most basic one, is the implication, which is typically interpreted by the residuum belonging to the t-norm. Finally, the logic may or may not offer the possibility to express explicitly to which degree a proposition holds.

Based on this approach, more than just a few logics have been defined and intensively studied. Monographs of basic importance include [14, 13] as well as, as regards the explicitation of truth degrees, [18]. We remark that we certainly do not address all logics which have been called "fuzzy" in the literature; this would be impossible as it seems that nowadays any logic has a "fuzzy" counterpart. Here, we just speak about logics of the indicated type.

We wish to address in this note the peculiar relationship between t-norm-based fuzzy logics and a problem which is regularly mentioned in discussions on the fundamentals of fuzzy logics: how to formalise reasoning when the referred information is possibly vague. The discussion on the very nature of fuzzy logics is old. An aim has been to develop fuzzy logics, as they are, from clear, meaningful principles; see, among many papers, e.g. [9, 10, 21, 22]. Here, we want to approach once the subject from the other side, namely, from the point of view of a specific application: an appropriate formalisation of justifiable reasoning can simply be a practical need.

Let us first try to formulate our concern in an abstract way. We assume to be given a set of propositions whose content does not matter. This means that for the derivation of consequences, the meaning is not taken into account and can be assumed to be unknown. We just know that the propositions refer to the observable or unobservable properties of somebody or something, describable in, possibly scientific, natural language. Important for us, the propositions express the presence of some property which can be vague, where, as usual, vagueness is characterised by the possibility of borderline cases. Moreover, we assume to have some knowledge about the mutual relationships. These relationships may express that some property is more general than another one, or a causal implication based on experience; again, we do not require the relationships to hold necessarily strict. What we are finally interested in, is to find a formal framework which tells us how to derive new information from the one we have to our disposal. Since the outlined situation is very general, the framework must be very general as well; we wonder, so to say, about a minimal logic for reasoning under vagueness.

The described problem is not purely academic and in particular not part of any "ivory-tower" philosophical theory, but has a practical background. We work towards an appropriate formalisation of a medical expert system [7]. The system which we are to analyse is called Cadiag-2, the second generation of the expert systems Cadiag – "*C*omputer-*A*ssisted *Dia*gnosis", and aims at the differential diagnostic decision support in patient care [1, 2]. Cadiag-2 processes both vague and uncertain data; we restrict here to the first case only. The procession of uncertain data calls for a probabilistic logic and causes completely different problems than those discussed here; for probability theory, well-founded calculi exist, whose unpleas-

ISBN: 978-989-95079-6-8

ant feature, however, can be a too high complexity as well as incomprehensible inference rules.

As we intend to formalise propositions disregarding their content, our problem is clearly a case for a propositional logic. As mentioned, many propositional fuzzy logics have been proposed in the past, based on different ways how to endow the real unit interval with a structure. Moreover, a wide range of logics has been introduced related to the problem which we address here, in particular different versions of "logics of argumentation". For a comprehensive overview and a large collection of references, we recommend the handbook [12]. Here, we just mimic the first steps towards the generally much more sophisticated and often more specialised systems found in the literature. What we have in mind is to concentrate exclusively on the aspect of vagueness, to proceed in a way which can hardly be further generalised, and to see how the result relates to t-norm-based fuzzy logics.

The set of eligible propositional logics is much restricted by one basic requirement, dictated by the intended application: all constituents of the formal logic need to have a counterpart on the informal level. Specifically, there must exist a plausible way how to think about each connective which appears in the logic's language, most easily obtained by a clear correlation with a natural-language expression. Moreover, if a proposition is provable from others, there must exist a proof in a proof system such that each step is comprehensible as a plausible argument, rather than a pure manipulation of strings; in the best case, each step can easily be translated to an explanation in natural language, exhibiting the causal or logical relationship on which the argument is based.

A fuzzy logic of the above-indicated kind does not meet any of these requirements. To see the problem, let us assume that we actually can find a logic fitting to our needs. We actually feel that it is natural to assume so; after all, we wish to formalise possibly vague statements, and these are appropriately evaluated on a linear continuous scale. So let us see how the logic could look like. We need a conjunction $\wedge$ and a negation $\sim$; the interpretation by the infimum and the standard negation, respectively, will do in our case. We note that the interpretation of truth values is not our subject here, and our particular choices for the connectives provide just an example. Moreover, we need to express truth degrees explicitly; to this end, we add constants $\bar{r}$ for each rational $r \in [0,1]$. So far, we do not encounter problems.

The first serious problem comes with the implication. Let us tentatively add the connective $\rightarrow$ interpreted by the residuum belonging to $\wedge$. Note that we then arrive at the logic $\mathsf{RGL}_\sim$, the Gödel logic enriched with the standard negation and truth constants [8]. Now, there is a natural way how to think about a statement "$\alpha \rightarrow \beta$": we interpret it as "α implies β". However, this clarity disappears as soon as we nest implications on the left side, like in "$(\alpha \rightarrow \beta) \rightarrow \gamma$". If our reference, a set of propositions, has a priori the structure of a residuated lattice, we can say that $\alpha \rightarrow \beta$ denotes the weakest element which, together with α, implies β. But in our case, there is no such structure available; we recall that we do not wish to assume any a priori structure as we would have to analyse the propositions by content then.

We note that this critics is related to the discussion out of which relevance logics arose [3]. Furthermore, a discussion in which the role of the implication connective in fuzzy logics is opposed to the needs of certain applications, can be found in [4].

For us, there is a reasonable way how to proceed: to drop the implication as a generally applicable connective. To this end, we consider in [7] a logic in which the implication always appears at the only place at which it can be appropriately called an implication: on the outermost level. Namely, we consider pairs of implication-free formulas, with the intended meaning that the left one denotes a proposition which is stronger then the right one. Thus we leave the area of t-norm-based logics and enter the field of lattice word problems. Namely, our model is the following algebra K, a Kleene algebra with added constants: $([0,1]; \wedge, \vee, \sim, (\bar{r})_{r \in \mathbb{Q}[0,1]})$.

We arrive at a logic, which we call GZL, whose formal expressions possess straightforward interpretations. However, we encounter the second and even more serious problem when considering a proof system for GZL. Following the lines of [5], we have in [7] presented a proof system based on sequents-of-relations. Unfortunately, our requirement that proof steps should be comprehensible in an informal way, is far from being fulfilled. Consider the K-tautology

$$\alpha \wedge \sim\alpha \;\rightarrow\; \beta \vee \sim\beta. \tag{1}$$

In its proof, we have to make necessarily use of the possibility to use multisets of relations, namely,

$$\alpha \Rightarrow \beta \;\mid\; \beta \Rightarrow \alpha \tag{2}$$

will appear in the proof. However, α and β refer to arbitrary facts, and the tentative translation "α implies β, or β implies α" is nonsense. Note that the problem concerns the proved result as well; (1) cannot be interpreted as a statement which anybody would ever tell.

The deeper reason for this difficulty is the semantics. An element of $[0,1]$ is intended to be the truth degree of a proposition; but it is treated like its meaning. What we might think as being associated to a property, telling that the property does not fully apply, is already the property itself. As a consequence, when using $[0,1]$ as a model, we may be led to the situation that we compare something by strength what by content would never be comparable. A valid statement of the form "$\alpha \rightarrow \beta$" is not really translatable to "from α we can conclude β", but only that α is under all circumstances assigned a smaller truth value than β, and based on this interpretation, (2) becomes indeed meaningful and just expresses the linear order of the truth degrees. However, this interpretation is not what we want.

The interpretation in the linearly ordered set of reals may certainly be useful at other places. A requirement comes into play which we have frequently argued for: to put fuzzy logics on firm grounds, we need first to be aware of the nature of what we reason about. In case of t-norm-based fuzzy logic, we reason about a set of propositions which has the internal structure of a residuated lattice, as it is the case for universes of fuzzy sets. The same, by the way, applies to classical propositional logic, which exactly reasons about a collection of propositions endowed a priori with the structure of a Boolean algebra; the popular claims about a "general validity" of this logic are meaningless.

ISBN: 978-989-95079-6-8

For us, the only way out is to restrict the calculus for GZL to those inferences which are not in conflict with our intended interpretation: the sequent $\alpha \Rightarrow \beta$ should mean that α is a statement stronger than β. We can achieve this is by not allowing multisets of relations, but only single relations. The interesting observation in [7] is that by means of this restriction, we get – not exactly but, say – very close to the logic which is actually used for the expert system which we examine.

The observation that the logic which we need arises by a certain restriction of a t-norm-based fuzzy logic, might be considered interesting, but not really satisfying. It rather suggests that the conceptual differences between the logic underlying systems like Cadiag-2 or similar expert systems on the one hand, and fuzzy logics on the other hand, cannot be bridged.

2 A minimal logic for reasoning under vagueness, without explicit degrees

The problem how to formalise ways to argue about vague propositions of unspecified content and their mutual interrelations, calls for alternative solutions. Let us opt for the syntactical approach; we will assemble some inference rules which translate to argumentation steps in a straightforward way. We will then check if some semantics with a reasonable interpretation can be found ex post, taking all imaginable possibilities into account and in particular not restricting ourselves to structures known from fuzzy logics or fuzzy set theory.

We note that this procedure seems to be in sharp contrast to the guiding principles of mathematical modelling which we have defended earlier, namely the principle that prior to any formalisation, the structure of reference needs to be specified first, in a way that the meaning of all its constituents is clear. However, in the present case, we do not do metamathematics, we do not examine ways how structures of a certain type are generally examined in a sound way; we do mathematics. Namely, it is the way of reasoning itself which is our object of investigation, and we do not share the opinion that rules for proper argumentation are fixed and thus can be derived from some higher-level truth. Intuitively acceptable inference rules will rather constitute a structure over a set of atomic propositions, and we do not assume a canonical answer how it may look like. In any case, we examine a logic as a mathematical object, the notion "logic" just being a name for it.

As indicated, there is not really a canonical way to select rules. One may argue against certain rules shown below, or feel that there is something missing. A discussion would not be fruitful if no guidelines were provided. We keep with the application in medicine; the rules shown below are extracted from those essential for the medical expert system with whose formalisation we are concerned.

In this section, we consider the case that we do not deal with truth values explicitly. We define the propositional logic DML as follows.

Definition 2.1 The *propositions* of DML are built up from a set of symbols $\varphi_1, \varphi_2, \ldots$ and the two constants $\bar{0}, \bar{1}$ by means of the binary connectives $\wedge, \vee$ and the unary connective $\sim$; the set of propositions is denoted by $\mathcal{F}_L$. The *implications* of DML are ordered pairs of propositions, denoted by $\alpha \to \beta$, where $\alpha, \beta \in \mathcal{F}_L$; the set of implications is denoted by $\mathcal{F}_I$.

Moreover, a *sequent* is an ordered pair of a non-empty finite set of propositions and a single proposition, notated by $\gamma_1, \ldots, \gamma_k \Rightarrow \delta$. The axioms and rules of DML are the following, for any propositions α, β, γ and sequent Γ:

$$\bar{0} \Rightarrow \alpha \quad \alpha \Rightarrow \alpha \quad \alpha \Rightarrow \bar{1}$$

$$\frac{\Gamma \Rightarrow \alpha \quad \alpha \Rightarrow \beta}{\Gamma \Rightarrow \beta} \qquad \frac{\Gamma \Rightarrow \alpha}{\Gamma, \beta \Rightarrow \alpha}$$

$$\frac{\Gamma \Rightarrow \alpha \quad \Gamma \Rightarrow \beta}{\Gamma \Rightarrow \alpha \wedge \beta} \qquad \frac{\Gamma, \alpha, \beta \Rightarrow \gamma}{\Gamma, \alpha \wedge \beta \Rightarrow \gamma}$$

$$\frac{\Gamma, \alpha \Rightarrow \gamma \quad \Gamma, \beta \Rightarrow \gamma}{\Gamma, \alpha \vee \beta \Rightarrow \gamma} \qquad \frac{\Gamma \Rightarrow \alpha}{\Gamma \Rightarrow \alpha \vee \beta} \qquad \frac{\Gamma \Rightarrow \beta}{\Gamma \Rightarrow \alpha \vee \beta}$$

$$\frac{\alpha \Rightarrow \beta}{\sim\beta \Rightarrow \sim\alpha} \qquad \frac{\sim\alpha \Rightarrow \beta}{\sim\beta \Rightarrow \alpha} \qquad \frac{\alpha \Rightarrow \sim\beta}{\beta \Rightarrow \sim\alpha}$$

The notion of a proof of a sequent from a finite set of sequents is defined in the expected way. A *theory* of DML is a finite set of implications. An implication $\alpha \to \beta$ is called provable from $\mathcal{T} = \{\alpha_1 \to \beta_1, \ldots, \alpha_n \to \beta_n\}$ if there is a proof of $\alpha \Rightarrow \beta$ from $\{\alpha_1 \Rightarrow \beta_1, \ldots, \alpha_n \Rightarrow \beta_n\}$, in signs $\mathcal{T} \vdash \alpha \to \beta$.

As to be expected, a sequent $\gamma_1, \ldots, \gamma_k \Rightarrow \delta$ is meant to express that "γ_1 and ... and γ_k imply δ".

Note that this logic again does not contain the implication as a connective, and it does contain a negation. Moreover, the calculus is sound, but not complete, with respect to classical two-valued interpretations; again (1) is not derivable. For the lack of the implication connective, also the comparison with intuitionistic logic is (probably) not well possible.

But a semantics is easily found, since DML is Belnap's logic of De-Morgan lattices; DML differs only slightly from the calculus presented in [11]. A structure $(M; \wedge, \vee, \sim, 0, 1)$ is a De-Morgan lattice if (i) $(M; \wedge, \vee, 0, 1)$ is a distributive lattice and (ii) $\sim$ is an order-reversing and involutive unary operation. De-Morgan lattices are subalgebras of direct products of the algebra $(\mathcal{M}_4; \wedge, \vee, \sim, 0, 1)$, where $(\mathcal{M}_4; \wedge, \vee, 0, 1)$ is the four-element Boolean lattice and $\sim$ maps each of the two atoms to itself [16]. It follows that we can provide a semantics based on $\mathcal{M}_4$; we note that assigning one of the four truth values to a proposition φ is usually interpreted as that φ is known to be true, false, neither true nor false, both true and false, respectively.

Most remarkably, fuzziness does not appear. However, the algebra $\mathcal{M}_4$ possesses a natural "fuzzified" extension, and we may alternatively base DML on a kind of fuzzy semantics. To make the comparison possible, we consider the pair V_c and V_f; note that the algebra V_c is isomorphic to $\mathcal{M}_4$.

Definition 2.2 Let $V_c = \{(s,t) : s,t \in \{0,1\}\}$, endowed with the componentwise natural order and the operation $\sim$ defined by

$$\sim(s,t) = (1-t, 1-s) \tag{3}$$

for $s,t \in \{0,1\}$. A *crisp evaluation* of DML is a mapping $v\colon \mathcal{F}_L \to V_c$ preserving $\wedge, \vee, \sim$ and mapping $\bar{0}$ to $(0,0)$ and $\bar{1}$ to $(1,1)$. An implication $\alpha \to \beta$ is then said to be *satisfied* by v if $v(\alpha) \leq v(\beta)$. A theory $\mathcal{T}$ is said to *crisply entail* an implication $\alpha \to \beta$ if the latter is satisfied by all crisp evaluations satisfying every element of $\mathcal{T}$; we write $\mathcal{T} \models_c \alpha \to \beta$ in this case.

ISBN: 978-989-95079-6-8

Furthermore, let $V_f = \{(s,t)\colon\ s,t \in [0,1]\}$, endowed with the componentwise natural order and the operation $\sim$, which is again defined by (3), where however this time $s,t \in [0,1]$. We define *fuzzy evaluations*, *satisfaction*, and *fuzzy entailment* similarly as above.

Theorem 2.3 *Let $\mathcal{T}$ be a theory and $\alpha \to \beta$ an implication of* DML. *Then $\mathcal{T} \vdash \alpha \to \beta$ if and only if $\mathcal{T} \models_c \alpha \to \beta$ if and only if $\mathcal{T} \models_f \alpha \to \beta$.*

Proof. Completeness in the indicated sense, but with respect to arbitrary De-Morgan lattices, holds due to [11, Theorem 4.11, (A1)]. But any De-Morgan lattice is a subalgebra of a direct product of copies of V_c, or alternatively, of a direct product of copies of V_f. □

As regards the interpretation of the semantics which we have proposed for DML, the situation is surprising with respect to the fuzzy variant. There is a close connection of DML to a logic which has been proposed in the context of decision making, based on the observation that we often consider separately the arguments in favour and the arguments against a possible decision [19].

On the other hand, the fact that we can work, without the need to change the inference rules, with crisp truth values as well, is somewhat disillusioning. The reasoning is the same if we assume our propositions to refer to vague or crisp properties. The situation certainly changes when we include truth degrees, as to be done next.

3 A minimal logic for reasoning under vagueness, with explicit degrees

The logic DML discussed in the last section is intended to be useful for reasoning about relationships between statements involving vagueness. The vagueness, however, cannot be addressed directly; and the logic can actually equally well considered as a logic not concerning vagueness. Moreover, a possible non-strictness of the relationships themselves is not expressible.

In applications, it can be desirable to have the possibility to denote a vague property by one single symbol, to which the degree of presence is explicitly attached. In this chapter, we attempt to formulate a calculus similarly to DML, but with explicit reference to truth degrees.

This problem is much more involved. We believe that there are many possibilities, and that the decision which is the best is even more difficult than in the above case.

According to a common procedure, we could enrich the language by truth constants. We will not follow this way; apart from the fact that we have not succeeded to produce a reasonable result concerning a logic DML enriched with truth constants, the idea is actually not well in accordance with the guidelines formulated above: truth constants should not be mixed with the meaning of a proposition. So we will use the two sharp truth constants only, representing falsity and trueness.

Truth constants should appear on a separate level. We propose to make a graded implication the basic syntactical constituent:

$$\alpha \xrightarrow{t} \beta,$$

with the intended meaning that α implies β to a degree $\geq t \in [0,1]$, where α and β are propositions of DML. As an example, let α denote a crisp proposition like "having a body temperature of $37.8°C$", and let β denote "having fever"; then the statement would hold with, say, $t = 0.8$.

In general, α and β are meant to refer to any vague property. Then either t refers to the compatibility of α with β. Alternatively, we can mean that α and β are causally related; then the smaller t is, the less strict is this relationship. Finally, we may also deal with single properties. Namely, the expressions

$$\bar{1} \xrightarrow{t} \delta \quad \text{and} \quad \delta \xrightarrow{t} \bar{0} \tag{4}$$

may serve to express that δ holds to the degree t, or δ is refused with the degree t, respectively.

To formulate inference rules is not straightforward; in the present context, our aim can only be plausibility. The basic question is which truth degree is, by tendency, assumed after two successive inference steps which are both based on a non-strict relationship. To examine this problem is even more difficult than to make a reasonable choice with regard to the truth values themselves, a problem which has been studied numerous times, see e.g. [15, 17]. It would clearly be desirable to have methods at hand to examine also the present situation empirically, in analogy to the procedure followed in [15].

Only the case of implications of the form $\alpha \xrightarrow{1} \beta$ is clear; they are supposed to express strict relationships, and for them, the rules of DML should be applicable. Furthermore, the degrees are assumed to be lower bounds, hence an implication is the less expressive the smaller the indicated truth degree is. Statements of the form $\alpha \xrightarrow{0} \beta$ do not express anything.

So let us see how the set of rules for our refined logic could look like. For implications $\alpha \xrightarrow{1} \beta$, we will use the rules of DML. For implications of the form $\alpha \xrightarrow{t} \beta$, where $t < 1$, the rules of DML introducing $\wedge$ or $\vee$ are not generalisable though; the degree of the conclusion cannot be assumed to be calculable from the degrees of the assumptions. However, what we should be able to say is, if we replace α by a stronger proposition, or β by a weaker proposition, then the relationship between α and β should be characterised by a higher truth value, so that t will still be a lower bound. Next, assume that we have proved $\alpha \xrightarrow{s} \beta$ and $\beta \xrightarrow{t} \gamma$; then $\alpha \xrightarrow{u} \gamma$ will be derivable as well, and we have to offer a way to calculate the degree u from s and t. As mentioned above, a well-founded decision is impossible, hence just like in case of the design of a fuzzy logic, a pragmatic solution is needed here. We opt for the operation dual to the truncated addition: we take here the Łukasiewicz t-norm $\odot\colon\ [0,1]^2 \to [0,1],\ \ (a,b) \mapsto (a+b-1) \vee 0$.

We specify the propositional logic ArgL as follows.

Definition 3.1 The set $\mathcal{F}_L$ of *propositions* is defined like for DML. An *implication* of ArgL consists of ordered triples of two propositions and a rational value $t \in [0,1]$; we write $\alpha \xrightarrow{t} \beta$, where $\alpha, \beta \in \mathcal{F}_L$; the set of implications is denoted by $\mathcal{F}_I$.

Moreover, a *sequent* is an ordered triple consisting of a non-empty finite set of propositions, a single proposition, and a rational value $t \in [0,1]$; we write

$$\gamma_1, \ldots, \gamma_k \overset{t}{\Rightarrow} \delta.$$

ISBN: 978-989-95079-6-8

The *crisp* rules of ArgL are those of DML, the symbol $\Rightarrow$ being replaced at all places by $\stackrel{1}{\Rightarrow}$.

The *fuzzy* rules of ArgL are the following:

$$\frac{\Gamma \stackrel{s}{\Rightarrow} \alpha \quad \alpha \stackrel{t}{\Rightarrow} \beta}{\Gamma \stackrel{s \odot t}{\Rightarrow} \beta} \qquad \frac{\Gamma \stackrel{t}{\Rightarrow} \alpha}{\Gamma \stackrel{s}{\Rightarrow} \alpha}, \text{ where } s \leq t$$

$$\frac{\Gamma, \alpha \stackrel{t}{\Rightarrow} \delta}{\Gamma, \alpha \wedge \beta \stackrel{t}{\Rightarrow} \delta} \qquad \frac{\Gamma \stackrel{t}{\Rightarrow} \alpha}{\Gamma \stackrel{t}{\Rightarrow} \alpha \vee \beta}$$

The notion of a *proof*, a *theory*, the *provability* of an implication from a theory, is defined similarly like for DML.

To associate to this calculus a reasonable semantics, is the next challenge. Only one point seems to be certain – we are not led to fuzzy sets. The only remarkable fact is that a t-norm is involved; for connections between t-norms and a somewhat similar setting, see [6, 20].

Let us consider the following structures, so-to-say the algebraic counterpart of ArgL. Here, $\oplus : [0,1]^2 \to [0,1]$, $(a,b) \mapsto (a+b) \wedge 1$ is the t-conorm associated to $\odot$.

Definition 3.2 A structure $(A; \wedge, \vee, \sim, 0, 1, d)$ is called a *metric De-Morgan lattice* if $(A; \wedge, \vee, \sim, 0, 1)$ is a De-Morgan lattice and $d\colon A \times A \to [0,1]$ is such that (i) $d(a,b) = 0$ if and only if $a \leq b$ and (ii) $d(a,c) \leq d(a,b) \oplus d(b,c)$.

An *evaluation* of ArgL in a metric De-Morgan lattice A is a mapping $v : \mathcal{F}_L \to A$ preserving $\wedge, \vee, \sim$ and the constants. An implication $\alpha \stackrel{t}{\to} \beta$ is *satisfied* by an evaluation v if $d(v(\alpha), v(\beta)) \leq 1 - t$. Semantic entailment is defined as usual.

Let us consider the following instructive example. Let $(A; \wedge, \vee, \sim, 0, 1)$ be a Boolean algebra, and let $\mu\colon A \to [0,1]$ be a strictly positive submeasure on A, meaning that, for $a, b \in A$, (i) $\mu(0) = 0$, (ii) $\mu(a) > 0$ if $a > 0$, (iii) $a \leq b$ implies $\mu(a) \leq \mu(b)$, (iv) $\mu(1) = 1$, and (v) $\mu(a \vee b) \leq \mu(a) \oplus \mu(b)$. Furthermore, put $d(a,b) = \mu(a \wedge \sim b)$. Then we may check that $(A; \wedge, \vee, \sim, 0, 1, d)$ is a metric De-Morgan lattice.

Theorem 3.3 *Let $\mathcal{T}$ be a theory and $\alpha \stackrel{t}{\to} \beta$ an implication of* ArgL*, where $t \in (0,1]$. Then $\mathcal{T} \vdash \alpha \stackrel{t}{\to} \beta$ if and only if $\mathcal{T} \models \alpha \stackrel{t}{\to} \beta$.*

Proof. The "only if" part is easy; just define the satisfaction of sequents by identifying the set on the left side with its conjunction.

For the "if" part, assume that $\mathcal{T}$ does not prove the implication $\gamma \stackrel{s}{\to} \delta$.

Let A be the quotient of $\mathcal{F}_L$ w.r.t. the equivalence relation $\Leftrightarrow$, where $\alpha \Leftrightarrow \beta$ if $\alpha \stackrel{1}{\to} \beta$ and $\beta \stackrel{1}{\to} \alpha$ are provable from $\mathcal{T}$. Then A is naturally endowed with the structure of a De-Morgan lattice.

For $\alpha, \beta \in \mathcal{F}_L$, let $d([\alpha],[\beta]) = 1 - t$, where $t \in [0,1]$ is maximal such that $\mathcal{T} \vdash \alpha \stackrel{t}{\to} \beta$. Endowed with d, A is a metric De-Morgan lattice. By construction, the elements of $\mathcal{T}$ are satisfied by the natural embedding of $\mathcal{F}_L$ into A, but not $\gamma \stackrel{s}{\to} \delta$. □

Needless to comment, our formalism comes closer to measure theory than to fuzzy set theory.

From the interpretational point of view, the semantics based on metric De-Morgan lattices is to be clarified though. Note that only the special case of a Boolean algebra with a submeasure offers an intuitively well comprehensible picture.

4 Conclusion

T-norm-based propositional fuzzy logics are frequently discussed as a suitable tool to model reasoning under vagueness. We have stressed that this is the case as long as the propositions which are formalised share the same reference; namely, they must be modellable by a system of fuzzy sets over a common universe.

If propositions are arbitrary, we run into difficulties when trying to apply techniques of fuzzy logics. This is, for example, the case for the medical expert system Cadiag-2, whose knowledge base contains information on logical and causal relationships between entities which are processed regardless of their meaning. We have risen the question how to define, in this general setting, a minimal frame for what we could call formalised argumentations. As a proposal, we have designed two minimal, but for our needs fully sufficient, systems; we have done so purely syntactically, allowing only rules with a clear interpretation.

The first version concerns reasoning without explicit reference to truth degrees; what comes out is the De-Morgan logic, which allows for interpretations without any connection to fuzziness. The second version incorporates truth values, but the calculus which comes out, still is by no means related to t-norm-based fuzzy logic. The semantics which can be defined ex post are De-Morgan lattices endowed with a non-symmetric distance function.

Our calculi are qualified in that they allow to reproduce the inference mechanism of Cadiag-2. The further elaboration on details of the calculus is work to be done, as well as the analysis of the newly introduced notion of a metric De-Morgan lattice.

References

[1] K.-P. Adlassnig, G. Kolarz, W. Scheithauer, H. Effenberger, G. Grabner, Cadiag: Approaches to computer-assisted medical diagnosis, *Comput. Biol. Med.*, 15:315–335, 1985.

[2] K.-P. Adlassnig, G. Kolarz, W. Scheithauer, G. Grabner, Approach to a hospital-based application of a medical expert system, *Med. Inform.*, 11:205–223, 1986.

[3] A. R. Anderson, N. D. Belnap, J. M. Dunn, "Entailment. The logic of relevance and necessity", Vol. I and II, Princeton University Press, Princeton 1975 and 1992.

[4] L. Běhounek, On the difference between traditional and deductive fuzzy logic, *Fuzzy Sets Syst.* 159:1153–1164, 2008.

[5] M. Baaz, A. Ciabattoni, C. G. Fermüller, Cut-Elimination in a Sequents-of-Relations Calculus for Gödel Logic, in: Proceedings of the 31st International

ISBN: 978-989-95079-6-8

Symposium on Multiple-Valued Logic (Warsaw 2001), IEEE Computer Society; 181–186, 2001.

[6] B. De Baets, R. Mesiar, Pseudo-metrics and $\mathcal{T}$-equivalences, *J. Fuzzy Math.*, 5:471–481, 1997.

[7] A. Ciabattoni, T. Vetterlein, On the (fuzzy) logical content of Cadiag-2, submitted.

[8] F. Esteva, Ll. Godo, P. Hájek, M. Navara, Residuated fuzzy logics with an involutive negation, *Arch. Math. Logic*, 39:103–124, 2000.

[9] C. G. Fermüller, Theories of vagueness versus fuzzy logic: Can logicians learn from philosophers? *Neural Network World*, 13:455–466, 2003.

[10] C. G. Fermüller, Revisiting Giles's game, in: O. Majer, A.-V. Pietarinen, T. Tulenheimo (eds.), *Games: Unifying Logic, Language, and Philosophy*, Springer-Verlag, Berlin 2009; 209–228.

[11] J. M. Font, Belnap's four-valued logic and De Morgan lattices, *Log. J. IGPL*, 5:413–440, 1997.

[12] D. M. Gabbay, R. H. Johnson, H. J. Ohlbach, J. Woods, *Handbook of the logic of argument and inference. The turn towards the practical*, North-Holland, Amsterdam 2002.

[13] S. Gottwald, *A Treatise on Many-Valued Logics*, Research Studies Press, Baldock 2001.

[14] P. Hájek, *Metamathematics of Fuzzy Logic*, Kluwer, Dordrecht 1998.

[15] H. M. Hersh, A. Caramazza, A fuzzy set approach to modifiers and vagueness in natural language, *Journal of Experimental Psychology: General*, 105:254–276, 1976.

[16] J. A. Kalman, Lattices with involution, *Trans. Am. Math. Soc.*, 87:485–491, 1058.

[17] A. Matoušková, V. Novák, J. Kulka, Psycholinguistic Research of the Meaning of Linguistic Hedges, draft.

[18] V. Novák, I. Perfilieva, J. Močkoř, *Mathematical principles of fuzzy logic*, Kluwer, Dordrecht 1999.

[19] M. Öztürk, A. Tsoukiàs, Modelling uncertain positive and negative reasons in decision aiding, *Decision Support Systems* 43:1512–1526, 2007.

[20] T. Vetterlein, T-norms induced by metrics on Boolean algebras, *Soft Comput.*, 10:995–1000, 2006.

[21] T. Vetterlein, Fuzzy logic as a logic of the expressive strength of information, *Soft Comput.*, 12:479–485, 2008.

[22] T. Vetterlein, A way to interpret Łukasiewicz Logic and Basic Logic, *Studia Logica*, 90:407–423, 2008.

ISBN: 978-989-95079-6-8

On the representation theorem of multi-adjoint concept lattices*

Jesús Medina[1] Manuel Ojeda-Aciego[2]

1. Dept. Matemáticas, Universidad de Cádiz. Spain
2. Dept. Matemática Aplicada, Universidad de Málaga. Spain
Email: jesus.medina@uca.es, aciego@uma.es

Abstract— *Formal concept analysis has become an important and appealing research topic. There exist a number of different fuzzy extensions of formal concept analysis and of its representation theorem, which gives conditions for a complete lattice in order to be isomorphic to a concept lattice. In this paper we concentrate on the study of operational properties of the mappings α and β required in the representation theorem.*

Keywords— Formal concept analysis, multi-adjoint framework.

1 Introduction

Formal concept analysis [12] has become an important and appealing research topic both from a theoretical perspective [18, 29, 32] and from the applicative one. Regarding applications, we can find papers ranging from ontology merging [10, 27], to applications to the Semantic Web by using the notion of concept similarity [11], and from processing of medical records in the clinical domain [14] to the development of recommender systems [8].

Soon after the introduction of "classical" formal concept analysis, a number of different approaches for its generalization were introduced and, nowadays, there are works which extend the theory with ideas from fuzzy set theory [3, 21, 22] or fuzzy logic reasoning [2, 4, 9] or from rough set theory [20, 30, 33] or some integrated approaches such as fuzzy and rough [31], or rough and domain theory [19].

In this paper we concentrate on the fuzzy extensions of formal concept analysis, for which a number of different approaches have been presented. To the best of our knowledge, the first one was given in [6], although they did not advance much beyond the basic definitions, probably due to the fact that they did not use residuated implications. Later, in [3, 28] the authors independently used complete residuated lattices as structures for the truth degrees; for this approach, a representation theorem was proved directly in a fuzzy framework in [5], setting the basis of most of the subsequent direct proofs.

In [23, 24] as a new general approach to formal concept analysis multi-adjoint concept lattices were introduced, in which the philosophy of the multi-adjoint paradigm [15, 26] to formal concept analysis is applied. With the idea of providing a general framework in which the different approaches stated above could be conveniently accommodated, the authors worked in a general non-commutative environment; and this naturally led to the consideration of adjoint triples, also called implication triples [1] or bi-residuated structures [25] as the main building blocks of a multi-adjoint concept lattice.

*Partially supported by the Spanish Science Ministry under grant TIN 2006-15455-C03-01 and by Junta de Andalucía under grant P06-FQM-02049.

The representation (or fundamental) theorem gives conditions for a complete lattice in order to be isomorphic to a concept lattice. This theorem is proved in the classical case [12] and in the fuzzy paradigms [5, 13, 16, 23, 28]. As a consequence, to obtain the isomorphism it is necessary to search two mappings α and β that satisfy some properties, one of these relate the mappings with the relation. In this paper, we present a characterization of this last property, which is more efficient than the actual from the computationally point of view. Moreover, some other interesting properties of mappings α and β are introduced.

The structure of the paper is as follows: in Section 2 we recall the definition of the multi-adjoint concept lattices and, in particular, the mappings α and β required in the definition of lattice representing a multi-adjoint concept lattice. Then, in Section 3, we prove some new results concerning α and β. Finally, some concluding remarks are added.

2 Multi-adjoint concept lattices

In this section we will recall the more important definitions and results from [23]. The first definition introduces the basic building blocks of the multi-adjoint concept lattices, the *adjoint triples*, which are generalisations of the notion of adjoint pair under the hypothesis of having a non-commutative conjunctor.

The lack of commutativity of the conjunctor, directly provides two different ways of generalising the well-known adjoint property between a t-norm and its residuated implication, depending on which argument is fixed in the conjunction.

Definition 1 *Let $(P_1, \leq_1)$, $(P_2, \leq_2)$, $(P_3, \leq_3)$ be posets and $\&\colon P_1 \times P_2 \to P_3$, $\swarrow\colon P_3 \times P_2 \to P_1$, $\nwarrow\colon P_3 \times P_1 \to P_2$ be mappings, then $(\&, \swarrow, \nwarrow)$ is an* adjoint triple *with respect to P_1, P_2, P_3 if:*

1. *$\&$ is order-preserving in both arguments.*
2. *$\swarrow$ and $\nwarrow$ are order-preserving in the consequent and order-reversing in the antecedent.*
3. *$x \leq_1 z \swarrow y$ iff $x \,\&\, y \leq_3 z$ iff $y \leq_2 z \nwarrow x$, where $x \in P_1$, $y \in P_2$ and $z \in P_3$.*

Note that in the domain and codomain of the considered conjunctor we have three (in principle) different sorts, thus providing a more flexible language to a potential user. Furthermore, notice that no boundary condition is required, in difference to the usual definition of multi-adjoint lattice [26] or implication triple [1]. Nevertheless, some boundary conditions follow from the definition, specifically, from the adjoint property (condition (3) above) [23].

ISBN: 978-989-95079-6-8

Lemma 1 *If $(P_1, \leq_1)$, $(P_2, \leq_2)$, $(P_3, \leq_3)$ have bottom element and $(\&, \swarrow, \nwarrow)$ is an adjoint triple, then $(P_1, \leq_1)$ and $(P_2, \leq_2)$ have top element and for all $x \in P_1, y \in P_2$ and $z \in P_3$ the following properties hold:*

1. $\bot_1 \,\&\, y = \bot_3, \qquad x \,\&\, \bot_2 = \bot_3.$
2. $z \nwarrow \bot_1 = \top_2, \qquad z \swarrow \bot_2 = \top_1.$

In order to provide more flexibility into our language, we will allow the existence of several adjoint triples for a given triplet of posets. Notice, however, that since these frames will be used as the underlying structures of our generalization of concept lattice, it is reasonable to require the lattice structure on some of the posets in the definition of adjoint triple.

Definition 2 *A* multi-adjoint frame *$\mathcal{L}$ is a tuple*

$$(L_1, L_2, P, \preceq_1, \preceq_2, \leq, \&_1, \swarrow^1, \nwarrow_1, \ldots, \&_n, \swarrow^n, \nwarrow_n)$$

where $(L_1, \preceq_1)$ and $(L_2, \preceq_2)$ are complete lattices, $(P, \leq)$ is a poset and, for all $i = 1, \ldots, n$, $(\&_i, \swarrow^i, \nwarrow_i)$ is an adjoint triple with respect to L_1, L_2, P.

For short, a multi-adjoint frame will be denoted as $(L_1, L_2, P, \&_1, \ldots, \&_n)$.

Following the usual approach to formal concept analysis, given a frame, a *multi-adjoint context* is a tuple consisting of sets of objects and attributes and a fuzzy relation among them; in addition, the multi-adjoint approach also includes a function which assigns an adjoint triple to each object (or attribute). This feature is important in that it allows for defining subgroups of objects or attributes in terms of different degrees of preference, see [23]. Formally, the definition is the following:

Definition 3 *Let $(L_1, L_2, P, \&_1, \ldots, \&_n)$ be a multi-adjoint frame, a* context *is a tuple (A, B, R, σ) such that A and B are non-empty sets (usually interpreted as attributes and objects, respectively), R is a P-fuzzy relation $R\colon A \times B \to P$ and $\sigma\colon B \to \{1, \ldots, n\}$ is a mapping which associates any element in B with some particular adjoint triple in the frame.*[1]

Once we have fixed a multi-adjoint frame and a context for that frame, we can define the following mappings $^{\uparrow_\sigma}\colon L_2^B \longrightarrow L_1^A$ and $^{\downarrow^\sigma}\colon L_1^A \longrightarrow L_2^B$ which can be seen as generalisations of those given in [4, 17]:

$$g^{\uparrow_\sigma}(a) = \inf\{R(a,b) \swarrow^{\sigma(b)} g(b) \mid b \in B\} \quad (1)$$

$$f^{\downarrow^\sigma}(b) = \inf\{R(a,b) \nwarrow_{\sigma(b)} f(a) \mid a \in A\} \quad (2)$$

These two arrows, $(^{\uparrow_\sigma}, ^{\downarrow^\sigma})$, generate a Galois connection [23]. For the sake of self-containment, this concept is defined below:

Definition 4 *Let $(P_1, \leq_1)$ and $(P_2, \leq_2)$ be posets, and $^{\downarrow}\colon P_1 \to P_2$, $^{\uparrow}\colon P_2 \to P_1$ mappings, the pair $(^{\uparrow}, ^{\downarrow})$ forms a* Galois connection *between P_1 and P_2 whenever the following conditions hold:*

1. *$^{\uparrow}$ and $^{\downarrow}$ are order-reversing.*
2. *$x \leq_1 x^{\downarrow\uparrow}$ for all $x \in P_1$.*
3. *$y \leq_2 y^{\uparrow\downarrow}$ for all $y \in P_2$.*

Proposition 1 ([23]) *Let $(L_1, L_2, P, \&_1, \ldots, \&_n)$ be a multi-adjoint frame and (A, B, R, σ) be a context, then the pair $(^{\uparrow_\sigma}, ^{\downarrow^\sigma})$ is a Galois connection between L_1^A and L_2^B.*

As usual in the different frameworks of formal concept analysis, a *multi-adjoint concept* is a pair $\langle g, f\rangle$ satisfying that $g \in L_2^B$, $f \in L_1^A$ and that $g^{\uparrow_\sigma} = f$ and $f^{\downarrow^\sigma} = g$; with $(^{\uparrow_\sigma}, ^{\downarrow^\sigma})$ being the Galois connection defined above.

Definition 5 *The* multi-adjoint concept lattice *associated to a multi-adjoint frame $(L_1, L_2, P, \&_1, \ldots, \&_n)$ and a context (A, B, R, σ) is the set*

$$\mathcal{M} = \{\langle g, f\rangle \mid g \in L_2^B, f \in L_1^A \text{ and } g^{\uparrow_\sigma} = f, f^{\downarrow^\sigma} = g\}$$

where the ordering is defined by $\langle g_1, f_1\rangle \preceq \langle g_2, f_2\rangle$ if and only if $g_1 \preceq_2 g_2$ (equivalently $f_2 \preceq_1 f_1$).

The ordering just defined above actually provides $\mathcal{M}$ with the structure of a complete lattice [23]. This follows from proposition 1 (the arrows $(^{\uparrow_\sigma}, ^{\downarrow^\sigma})$ forms a Galois connection) and the theorem below.

Theorem 1 ([7]) *Let $(L_1, \preceq_1)$, $(L_2, \preceq_2)$ be complete lattices, let $(^{\uparrow}, ^{\downarrow})$ be a Galois connection between L_1, L_2 and consider $\mathcal{C} = \{\langle x, y\rangle \mid x^{\uparrow} = y, x = y^{\downarrow}; x \in L_1, y \in L_2\}$; then $(\mathcal{C}, \preceq)$ is a complete lattice, where*

$$\bigwedge_{i\in I}\langle x_i, y_i\rangle = \langle \bigwedge_{i\in I} x_i, (\bigvee_{i\in I} y_i)^{\downarrow\uparrow}\rangle;$$

$$\bigvee_{i\in I}\langle x_i, y_i\rangle = \langle (\bigvee_{i\in I} x_i)^{\uparrow\downarrow}, \bigwedge_{i\in I} y_i\rangle$$

and $\langle x_1, y_1\rangle \preceq \langle x_2, y_2\rangle$ if and only if $x_1 \preceq_1 x_2$.

From now on, we will fix a multi-adjoint frame $(L_1, L_2, P, \&_1, \ldots, \&_n)$ and context (A, B, R, σ). Moreover, to improve readability, we will write $(^{\uparrow}, ^{\downarrow})$ instead of $(^{\uparrow_\sigma}, ^{\downarrow^\sigma})$ and $\swarrow^b$, $\nwarrow_b$ instead of $\swarrow^{\sigma(b)}$, $\nwarrow_{\sigma(b)}$.

In the next section, we will present some new properties about the functions α and β involved in the representation (or fundamental) theorem for the multi-adjoint framework presented in [23]. In order to do this, we will recall some necessary definitions.

Definition 6 *Given a complete lattice L, a subset $K \subseteq L$ is* infimum-dense *(resp.* supremum-dense*) if and only if for all $x \in L$ there exists $K' \subseteq K$ such that $x = \inf(K')$ (resp. $x = \sup(K')$).*

A multi-adjoint concept lattice is said to be represented by a complete lattice provided there is a pair of functions, α and β, satisfying the conditions stated in the definition below:

Definition 7 *A multi-adjoint concept lattice*[2] *$(\mathcal{M}, \preceq)$ is* represented *by a complete lattice $(V, \sqsubseteq)$ if there exists a pair of mappings $\alpha\colon A \times L_1 \to V$ and $\beta\colon B \times L_2 \to V$ such that:*

[1] A similar theory could be developed by considering a mapping $\tau\colon A \to \{1, \ldots, n\}$ which associates any element in A with some particular adjoint triple in the frame.

[2] Recall that we are considering a multi-adjoint concept lattice on a fixed frame $(L_1, L_2, P, \&_1, \ldots, \&_n)$ and context (A, B, R, σ).

ISBN: 978-989-95079-6-8

1a) $\alpha[A \times L_1]$ *is infimum-dense;*

1b) $\beta[B \times L_2]$ *is supremum-dense; and*

2) For all $a \in A$, $b \in B$, $x \in L_1$, $y \in L_2$*:*

$$\beta(b, y) \sqsubseteq \alpha(a, x) \quad \textit{if and only if} \quad x \,\&_b\, y \leq R(a, b)$$

From the definition of representability above the following properties follow:

Proposition 2 *Given a complete lattice* $(V, \sqsubseteq)$ *which represents a multi-adjoint concept lattice* $(\mathcal{M}, \preceq)$*, and mappings* $f \in L_1^A$ *and* $g \in L_2^B$*, we have:*

1. β *is order-preserving in the second argument.*
2. α *is order-reversing in the second argument.*
3. $g^{\uparrow}(a) = \sup\{x \in L_1 \mid v_g \sqsubseteq \alpha(a, x)\}$*, where* $v_g = \sup\{\beta(b, g(b)) \mid b \in B\}$.
4. $f^{\downarrow}(b) = \sup\{y \in L_2 \mid \beta(b, y) \sqsubseteq v_f\}$*, where* $v_f = \inf\{\alpha(a, f(a)) \mid a \in A\}$.
5. *If* $g_v(b) = \sup\{y \in L_2 \mid \beta(b, y) \sqsubseteq v\}$*, then* $\sup\{\beta(b, g_v(b)) \mid b \in B\} = v$.
6. *If* $f_v(a) = \sup\{x \in L_1 \mid v \sqsubseteq \alpha(a, x)\}$*, then* $\sup\{\alpha(a, f_v(a)) \mid a \in A\} = v$.

Finally, the fundamental theorem for multi-adjoint concept lattices presented in [23] is the following.

Theorem 2 *A complete lattice* $(V, \sqsubseteq)$ *represents a multi-adjoint concept lattice* $(\mathcal{M}, \preceq)$ *if and only if* $(V, \sqsubseteq)$ *is isomorphic to* $(\mathcal{M}, \preceq)$.

3 New results about the mappings α and β

In this section, we introduce some new interesting properties about the mappings α and β. So, let us assume a complete lattice $(V, \sqsubseteq)$ which represents a multi-adjoint concept lattice $(\mathcal{M}, \preceq)$ and the mappings $\alpha\colon A \times L_1 \to V$, $\beta\colon B \times L_2 \to V$.

We will restate below the isomorphism constructed in fundamental theorem, based on both the α and β functions, since these expressions will be used later.

Proposition 3 ([23]) *If a complete lattice* $(V, \sqsubseteq)$ *represents a multi-adjoint concept lattice* $(\mathcal{M}, \preceq)$*, then there exists an isomorphism* $\varphi\colon \mathcal{M} \to V$ *and two mappings* $\beta\colon B \times L_2 \to V$, $\alpha\colon A \times L_1 \to V$*, such that:*

$$\begin{aligned} \varphi(\langle g, f\rangle) &= \sup\{\beta(b, g(b)) \mid b \in B\} \\ &= \inf\{\alpha(a, f(a)) \mid a \in A\} \end{aligned}$$

for all concept $\langle g, f\rangle \in \mathcal{M}$.

The following result shows continuity-related properties of α and β in their second argument.

Proposition 4 *The mappings* $\beta\colon B \times L_2 \to V$ *and* $\alpha\colon A \times L_1 \to V$ *satisfy that:*

1. *For all indexed set* $Y = \{y_i\}_{i \in I} \subseteq L_2$ *and* $b \in B$*:*

$$\beta(b, \sup\{y_i \mid i \in I\}) = \sup\{\beta(b, y_i) \mid i \in I\}$$

2. *For all indexed set* $X = \{x_i\}_{i \in I} \subseteq L_1$ *and* $a \in A$*:*

$$\alpha(a, \sup\{x_i \mid i \in I\}) = \inf\{\alpha(a, x_i) \mid i \in I\}$$

Proof: 1. Consider $b \in B$ and $Y = \{y_i\}_{i \in I} \subseteq L_2$, as $\alpha[A \times L_1]$ is infimum-dense and $\beta(b, \sup Y) \in V$, there exists an indexing set Λ such that $\beta(b, \sup Y) = \inf\{\alpha(a_j, x_j) \mid j \in \Lambda\}$; as a result $\beta(b, \sup Y) \sqsubseteq \alpha(a_j, x_j)$ for every $j \in \Lambda$. From proposition 2(1), we obtain that $\beta(b, y_i) \sqsubseteq \alpha(a_j, x_j)$, for every $i \in I$ and $j \in \Lambda$, and hence $\sup\{\beta(b, y_i) \mid i \in I\} \sqsubseteq \alpha(a_j, x_j)$ for every $j \in \Lambda$, then

$$\begin{aligned} \sup\{\beta(b, y_i) \mid i \in I\} &\sqsubseteq \inf\{\alpha(a_j, x_j) \mid j \in \Lambda\} \\ &= \beta(b, \sup Y) \end{aligned}$$

For the other inequality, let us consider $\sup\{\beta(b, y_i) \mid i \in I\}$ and, as $\alpha[A \times L_1]$ is infimum-dense, there exists an indexing set Λ' such that $\sup\{\beta(b, y_i) \mid i \in I\} = \inf\{\alpha(a_j, x_j) \mid j \in \Lambda'\}$. Now, for all $i \in I$ and $j \in \Lambda'$ we obtain that $\beta(b, y_i) \sqsubseteq \alpha(a_j, x_j)$, therefore, from Definition 7(2), $x_j \,\&_b\, y_i \leq R(a_j, b)$. Now, as $(\&_b, \swarrow^b, \nwarrow_b)$ is an adjoint triple, we have the following chain of equivalent statements:

$$\begin{aligned} x_j \,\&_b\, y_i &\leq R(a_j, b) && \text{for all } i \in I \\ y_i &\preceq_2 R(a_j, b) \nwarrow_b x_j && \text{for all } i \in I \\ \sup Y &\preceq_2 R(a_j, b) \nwarrow_b x_j \\ x_j \,\&_b \sup Y &\leq R(a_j, b) \end{aligned}$$

so, $\beta(b, \sup Y) \sqsubseteq \alpha(a_j, x_j)$ for every $j \in \Lambda'$, and thus

$$\begin{aligned} \beta(b, \sup Y) &\sqsubseteq \inf\{\alpha(a_j, x_j) \mid j \in \Lambda'\} \\ &= \sup\{\beta(b, y_i) \mid i \in I\} \end{aligned}$$

2. This proof is analogous using that $\beta[B \times L_2]$ is supremum-dense. □

We continue below by proving some boundary conditions fulfilled by α and β.

Proposition 5 *The two mappings* $\alpha\colon A \times L_1 \to V$ *and* $\beta\colon B \times L_2 \to V$ *are such that* $\alpha(a, \bot_1) = \top_V$ *and* $\beta(b, \bot_2) = \bot_V$ *for all* $b \in B$ *and* $a \in A$.

Proof: Given $a \in A$, let us prove that $\alpha(a, \bot_1) = \top_V$. Firstly, recall that lemma 1 implies that $\bot_1 \,\&_b\, y \leq R(a, b)$ for all $b \in B$ and $y \in L_2$; now, from Definition 7(2) we obtain that $\beta(b, y) \sqsubseteq \alpha(a, \bot_1)$ for all $b \in B$ and $y \in L_2$, that is, $\alpha(a, \bot_1)$ is an upper bound of the set of elements $\beta(b, y)$ for all $b \in B$ and $y \in L_2$. Now, as β is supremum-dense, there is an indexing set Λ such that $\top_V = \sup\{\beta(b_i, y_i) \mid i \in \Lambda\}$, therefore, we have that: $\top_V \sqsubseteq \alpha(a, \bot_1)$. Hence, $\top_V = \alpha(a, \bot_1)$.

The other equality follows similarly. □

From the propositions above, we obtain the following corollary which states the behaviour of α and β regarding suprema of any set (either empty or non-empty).

Corollary 1 *The mappings* $\beta\colon B \times L_2 \to V$, $\alpha\colon A \times L_1 \to V$ *satisfy that:*

ISBN: 978-989-95079-6-8

1. $\beta(b, \sup Y) = \sup\{\beta(b, y) \mid y \in Y\}$, *for all* $Y \subseteq L_2$ *and* $b \in B$.

2. $\alpha(a, \sup X) = \inf\{\alpha(a, x) \mid x \in X\}$, *for all* $X \subseteq L_1$ *and* $a \in A$.

As a consequence of the property above we have the following result, which gives us a more efficient form to write Property (2) in Definition 7 to check if a lattice is isomorphic to a concept lattice, that is, in order to apply Theorem 2.

Proposition 6 *Given* $a \in A$, $b \in B$, *the applications* $\beta_b: L_2 \to V$, $\alpha_a: L_1 \to V$ *have residuated mappings, that is, there exist* $\beta'_b: L_2 \to V$, $\alpha'_a: L_1 \to V$ *such that:*

$$\begin{array}{lll} \beta_b(y) \sqsubseteq v & \textit{if and only if} & y \preceq_2 \beta'_b(v) \\ v \sqsubseteq \alpha_a(x) & \textit{if and only if} & x \preceq_1 \alpha'_a(v) \end{array}$$

for all $x \in L_1$, $y \in L_2$ *and* $v \in V$.

Proof: If we define $\beta'_b(v) = \sup\{y \in L_2 \mid \beta_b(y) \sqsubseteq v\}$ and, similarly, $\alpha'_a(v) = \sup\{x \in L_1 \mid v \sqsubseteq \alpha_a(x)\}$, we obtain the result straightforward from Corollary 1. □

The following proposition states a necessary and sufficient condition for the mappings α and β to fulfill the second condition in the definition of representable lattice.

Proposition 7 *The mappings* α *and* β *satisfy Property (2) in Definition 7 if and only if, for all* $a \in A$, $b \in B$, $x \in L_1$, $y \in L_2$, *any of the following equalities holds:*

$$\begin{array}{lcl} \beta'_b(\alpha_a(x)) & = & R(a,b) \nwarrow x \\ \alpha_a(\beta'_b(y)) & = & R(a,b) \swarrow y \end{array}$$

Proof: Firstly, we assume that the mappings α and β satisfy Property (2) in Definition 7. The first equality is given from the following chain of equivalences, given $a \in A$, $b \in B$, $x \in L_1$, $y \in L_2$:

$$\begin{array}{lcl} y \preceq_2 \beta'_b(\alpha_a(x)) & \iff & \beta(b,y) \sqsubseteq \alpha(a,x) \\ & \iff & x \,\&_b\, y \leq R(a,b) \\ & \iff & y \preceq_2 R(a,b) \nwarrow x \end{array}$$

if we substitute y by $R(a,b) \nwarrow x$ in the first sentence and y by $\beta'_b(\alpha_a(x))$ in the last one. The second equality follows similarly.

Now, we assume that $\beta'_b(\alpha_a(x)) = R(a,b) \nwarrow x$, hence

$$\begin{array}{lcl} \beta(b,y) \sqsubseteq \alpha(a,x) & \iff & y \preceq_2 \beta'_b(\alpha_a(x)) \\ & \overset{(*)}{\iff} & y \preceq_2 R(a,b) \nwarrow x \\ & \iff & x \,\&_b\, y \leq R(a,b) \end{array}$$

where $(*)$ is given from the hypothesis. □

As a result of the previous proposition, we obtain a straightforward mechanism to obtain the mappings α and β in order to check whether a lattice is isomorphic to a concept lattice.

Finally, the following property shows that any subset of $A \times L_1$ or of $B \times L_2$ is related to a concept via α and φ, or β and φ, respectively.

Proposition 8 *Consider a multi-adjoint concept lattice* $(\mathcal{M}, \preceq)$ *represented by a complete lattice* $(V, \sqsubseteq)$ *and the mappings* $\alpha: A \times L_1 \to V$, $\beta: B \times L_2 \to V$, *then for each* $K \subseteq A \times L_1$, *there exists a unique concept* $\langle g, f\rangle \in \mathcal{M}$ *such that*

$$\inf\{\alpha(a,x) \mid (a,x) \in K\} \quad = \quad \varphi(\langle g, f\rangle)$$

Analogously, for each $K' \subseteq B \times L_2$, *there exists a unique concept* $\langle g, f\rangle \in \mathcal{M}$ *such that*

$$\sup\{\beta(b,y) \mid (b,y) \in K'\} \quad = \quad \varphi(\langle g, f\rangle)$$

Proof: Given $K \subseteq A \times L_1$, let us consider the sets $K_a = \{x \mid (a,x) \in K\}$, and the function $h: A \to L_1$ defined as $h(a) = \sup K_a$.

By Corollary 1, we have that, for all $a' \in A$, the following equality holds

$$\alpha(a', h(a')) = \inf\{\alpha(a', x) \mid x \in K_a\}$$

Therefore:

$$\begin{array}{l} \inf\{\alpha(a', h(a')) \mid a' \in A\} = \\ \qquad = \quad \inf\{\inf\{\alpha(a', x) \mid x \in K_{a'}\} \mid a' \in A\} \\ \qquad = \quad \inf\{\alpha(a', x) \mid (a', x) \in K\} \end{array}$$

Finally, we obtain the following chain of equalities:

$$\begin{array}{lcl} \inf\{\alpha(a,x) \mid (a,x) \in K\} & = & \inf\{\alpha(a', h(a')) \mid a' \in A\} \\ & \overset{(1)}{=} & \sup\{\beta(b', h^{\downarrow}(b')) \mid b' \in B\} \\ & \overset{(2)}{=} & \varphi(\langle h^{\downarrow}, h^{\downarrow\uparrow}\rangle) \end{array}$$

where (1) follows by Proposition 2, (2) by Proposition 3. This means that the concept whose existence is postulated in the statement is $\langle h^{\downarrow}, h^{\downarrow\uparrow}\rangle$.

Now, the uniqueness follows from the isomorphism φ:
If there would exist another concept $\langle g, f\rangle$ such that $\sup\{\beta(b, g(b)) \mid b \in B\} = \inf\{\alpha(a,x) \mid (a,x) \in K\}$, we would have:

$$\begin{array}{lcl} \varphi(\langle h^{\downarrow}, h^{\downarrow\uparrow}\rangle) & = & \sup\{\beta(b, h^{\downarrow}(b)) \mid b \in B\} \\ & = & \inf\{\alpha(a,x) \mid (a,x) \in K\} \\ & = & \sup\{\beta(b, g(b)) \mid b \in B\} \\ & = & \varphi(\langle g, f\rangle) \end{array}$$

Thus, $\langle h^{\downarrow}, h^{\downarrow\uparrow}\rangle = \langle g, f\rangle$.

The second statement follows similarly. □

4 Conclusions

The representation theorem is one of the most important results in the theory of formal concept analysis, since it provides conditions in order to determine whether a given lattice is isomorphic to some concept lattice. In this paper, an analytic expression for the mappings α and β involved in the representation theorem of t-concept lattices is provided, together with some interesting properties.

ISBN: 978-989-95079-6-8

References

[1] A. Abdel-Hamid and N. Morsi. Associatively tied implicacions. *Fuzzy Sets and Systems*, 136(3):291–311, 2003.

[2] C. Alcalde, A. Burusco, R. Fuentes-González, and I. Zubia. Treatment of L-fuzzy contexts with absent values. *Information Sciences*, 179:1–15, 2009.

[3] R. Belohlávek. Lattice generated by binary fuzzy relations (extended abstract). In *4th Intl. Conf. on Fuzzy Sets Theory and Applications*, page page 11, 1998.

[4] R. Belohlávek. Concept lattices and order in fuzzy logic. *Annals of Pure and Applied Logic*, 128:277–298, 2004.

[5] R. Belohlávek. Lattices of fixed points of fuzzy Galois connections. *Mathematical Logic Quartely*, 47(1):111–116, 2004.

[6] A. Burusco and R. Fuentes-González. The study of L-fuzzy concept lattice. *Mathware & Soft Computing*, 3:209–218, 1994.

[7] B. Davey and H. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, second edition, 2002.

[8] P. du Boucher-Ryana and D. Bridge. Collaborative recommending using formal concept analysis. *Knowledge-Based Systems*, 19(5):309–315, 2006.

[9] S.-Q. Fan, W.-X. Zhang, and W. Xu. Fuzzy inference based on fuzzy concept lattice. *Fuzzy Sets and Systems*, 157(24):3177–3187, 2006.

[10] A. Formica. Ontology-based concept similarity in formal concept analysis. *Information Sciences*, 176(18):2624–2641, 2006.

[11] A. Formica. Concept similarity in formal concept analysis: An information content approach. *Knowledge-Based Systems*, 21(1):80–87, 2008.

[12] B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundation*. Springer Verlag, 1999.

[13] G. Georgescu and A. Popescu. Concept lattices and similarity in non-commutative fuzzy logic. *Fundamenta Informaticae*, 53(1):23–54, 2002.

[14] G. Jiang, K. Ogasawara, A. Endoh, and T. Sakurai. Context-based ontology building support in clinical domains using formal concept analysis. *International Journal of Medical Informatics*, 71(1):71–81, 2003.

[15] P. Julian, G. Moreno, and J. Penabad. On fuzzy unfolding: A multi-adjoint approach. *Fuzzy Sets and Systems*, 154(1):16–33, 2005.

[16] S. Krajci. The basic theorem on generalized concept lattice. In V. Snásel and R. Belohlávek, editors, *International Workshop on Concept Lattices and their Applications, CLA 2004*, pages 25–33, 2004.

[17] S. Krajci. A generalized concept lattice. *Logic Journal of IGPL*, 13(5):543–550, 2005.

[18] S. O. Kuznetsov. Complexity of learning in concept lattices from positive and negative examples. *Discrete Applied Mathematics*, 142:111–125, 2004.

[19] Y. Lei and M. Luo. Rough concept lattices and domains. *Annals of Pure and Applied Logic*, 2009. Article in press (http://dx.doi.org/10.1016/j.apal.2008.09.028).

[20] M. Liu, M. Shao, W. Zhang, and C. Wu. Reduction method for concept lattices based on rough set theory and its application. *Computers & Mathematics with Applications*, 53(9):1390–1410, 2007.

[21] X. Liu, W. Wang, T. Chai, and W. Liu. Approaches to the representations and logic operations of fuzzy concepts in the framework of axiomatic fuzzy set theory I. *Information Sciences*, 177(4):1007–1026, 2007.

[22] X. Liu, W. Wang, T. Chai, and W. Liu. Approaches to the representations and logic operations of fuzzy concepts in the framework of axiomatic fuzzy set theory II. *Information Sciences*, 177(4):1027–1045, 2007.

[23] J. Medina, M. Ojeda-Aciego, and J. Ruiz-Calviño. Formal concept analysis via multi-adjoint concept lattices. *Fuzzy Sets and Systems*, 160(2):130–144, 2009.

[24] J. Medina, M. Ojeda-Aciego, and J. Ruiz-Calviño. On multi-adjoint concept lattices: denition and representation theorem. *Lect. Notes in Artificial Intelligence*, 4390:197–209, 2007.

[25] J. Medina, M. Ojeda-Aciego, A. Valverde, and P. Vojtáš. Towards biresiduated multi-adjoint logic programming. *Lect. Notes in Artificial Intelligence*, 3040:608–617, 2004.

[26] J. Medina, M. Ojeda-Aciego, and P. Vojtáš. Similarity-based unification: a multi-adjoint approach. *Fuzzy Sets and Systems*, 146:43–62, 2004.

[27] V. Phan-Luong. A framework for integrating information sources under lattice structure. *Information Fusion*, 9:278–292, 2008.

[28] S. Pollandt. *Fuzzy Begriffe*. Springer, Berlin, 1997.

[29] K.-S. Qu and Y.-H. Zhai. Generating complete set of implications for formal contexts. *Knowledge-Based Systems*, 21:429–433, 2008.

[30] M.-W. Shao, M. Liu, and W.-X. Zhang. Set approximations in fuzzy formal concept analysis. *Fuzzy Sets and Systems*, 158(23):2627–2640, 2007.

[31] L. Wang and X. Liu. Concept analysis via rough set and afs algebra. *Information Sciences*, 178(21):4125–4137, 2008.

[32] X. Wang and W. Zhang. Relations of attribute reduction between object and property oriented concept lattices. *Knowledge-Based Systems*, 21(5):398–403, 2008.

[33] Q. Wu and Z. Liu. Real formal concept analysis based on grey-rough set theory. *Knowledge-Based Systems*, 22(1):38–45, 2009.

ISBN: 978-989-95079-6-8

A Neuro-fuzzy System for Fraud Detection in Electricity Distribution

C. Muniz, M. Vellasco, R. Tanscheit, K. Figueiredo

Computational Intelligence Lab., Department of Electrical Engineering,
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, Brazil
Email: cyroassis@gmail.com, {marley, ricardo, karla}@ele.puc-rio.br

Abstract *— The volume of energy loss that Brazilian electrical utilities have to deal with has been ever increasing. Electricity distribution companies have suffered significant and increasing losses in the last years, due to theft, measurement errors and other irregularities. Therefore there is a great concern to identify the profile of irregular customers, in order to reduce the volume of such losses. This paper presents a combined approach of a neural networks committee and a neuro-fuzzy hierarchical system intended to increase the level of accuracy in the identification of irregularities among low voltage consumers. The data used to test the proposed system are from Light S.A., the distribution company of Rio de Janeiro. The results obtained presented a significant increase in the identification of irregular customers when compared to the current methodology employed by the company.*

Keywords— neural nets, hierarchical neuro-fuzzy systems, binary space partition, electricity distribution, fraud detection.

1 Introduction

Commercial electricity loss due to irregularities (energy theft and measurement errors) reach about R$ 5 billion a year in Brazil, which is around 5% of the total amount spent in energy consumption. A great part of that loss occurs in the area operated by Light S.A.: 3.8 million customers in 28 cities in the state of Rio de Janeiro. Electricity bills are, in some cases, up to 17% more expensive than they should be, so as to compensate for energy thefts and other fraudulent practices [1].

The company makes use of heuristics to indicate a set of low voltage customers who may practice some kind of irregularity. This set of customers, initially classified as suspects, is then analyzed by an expert, who selects a certain number to be directly inspected. The inspection is performed by an employee that visits each residence to confirm (or not) the irregularity. Based on this simple approach, Light S.A. has reached a low identification rate. It is evident that the adopted procedure is not effective and that a more accurate system must be developed.

Some researches have already tackled this problem, confirming the difficulty of correctly identifying irregular consumption. In [2], Rough Sets have been used to fraud detection among electrical energy consumers. The results obtained are promising but accuracy is still low (around 20%) due to the presence of noisy information in the company's database. In [3] and [4] a similar approach is presented, where only past consumption information is considered in the analysis. Due to the temporal nature of the data, time series analysis methods [5] are applied in order to get new invariant characteristics. However, other relevant variables, such as information about the measurement device and the average temperature are not considered.

This paper describes a new intelligent methodology, based on neural networks and fuzzy systems, that takes into consideration historical consumption as well as other important variables. The main objective is to increase the accuracy level when identifying irregular low voltage consumers, selected from a group of suspicious clients.

The proposed methodology is composed of two basic modules: Filtering and Classification, the former comprising an ensemble of five artificial neural networks, while the latter consists of a Neuro-fuzzy Hierarchical System [6]. In the Filtering module, each neural network indicates whether customers are suspected of irregularities or not. As in any ensemble-based methodology, the use of a committee of neural networks aims at improving accuracy by producing a consensus decision that is potentially more accurate than individual neural networks [7]. The Filtering Module was developed to better extract actual normal and irregular customers for training the Classification Module. This module is introduced in the expectation of improving even further the detection of irregularities. It should be noted that the database contains untrustworthy data, due to incorrect information filled in by some visiting employees.

The paper is organized in four additional sections: Section 2 describes the current methodology used by Light S.A. for detecting irregularities, Section 3 describes the proposed methodology, Section 4 presents the case studies considered, and, finally, Section 5 concludes the work.

2 Current Methodology for Fraud Detection

Currently, Light S.A. makes use of diverse methodologies for selecting low voltage customers that present irregularity indications. The *Quarterly Indicator* methodology consists of comparing the past three months' consumption. The *Annual Indicator* consists of evaluating clients' consumption during the last 24 months. The *Adjustment Factor* methodology compares last month consumption to that in the corresponding month of the preceding year. Finally, the *Tendency Factor* methodology consists of comparing clients' previous month's consumption with that expected in the present month. In all cases, if the reduction in consumption is higher than expected, the client is considered a suspect.

All customers identified by those methodologies form

ISBN: 978-989-95079-6-8

the set of suspects, which is analyzed by an expert; some are then recommended for direct inspection. Through this methodology, Light has attained an average *Positive Predictive Value* (PPV) of 25%, which is the proportion of proven irregular customers among all those who have been classified as suspects of some irregularity and have been inspected. Table 1 shows a typical confusion matrix [8] for a two-class problem (*normal* and *irregular* consumers).

Table 1: Suspected x Inspected Customers

System Output	Inspection Result	
	Irregular	Normal
Irregular	A	B
Normal	C	D

The PPV is computed by:

$$PPV = A / (A + B)$$

3 Proposed Methodology

The methodology proposed in this work consists of three basic modules: Pre-Processing and Normalization, Filtering and Classification.

The whole methodology is shown in Fig. 1. Fig. 1a shows the system's configuration used during the learning phase, which contains all three modules; Fig. 1b, on the other hand, presents a simpler configuration, used for the recall phase, which does not include the Filtering Module.

The first module, present in both training and recall phases and called Pre-processing and Normalization, is where all attributes are selected, normalized and coded.

It should be pointed out, however, that the available inspection database is, unfortunately, very noisy, containing unreliable information due to (intentional or not) incorrect indication of an irregular consumer as a normal one during inspection. Therefore, to create a more reliable database for training the Classification Module, a Filtering Module was conceived.

The Filtering Module makes use of a committee of five neural networks, which select *irregular* and *normal* consumers that better characterize the two different classes. The resulting filtered database is then employed for training the Classification Module, which eventually identifies a customer as *normal* or *irregular*. Therefore, future classifications of customers (recall phase) will employ only the Pre-processing and Classification modules. The Filtering Module is only necessary if the system needs to be retrained.

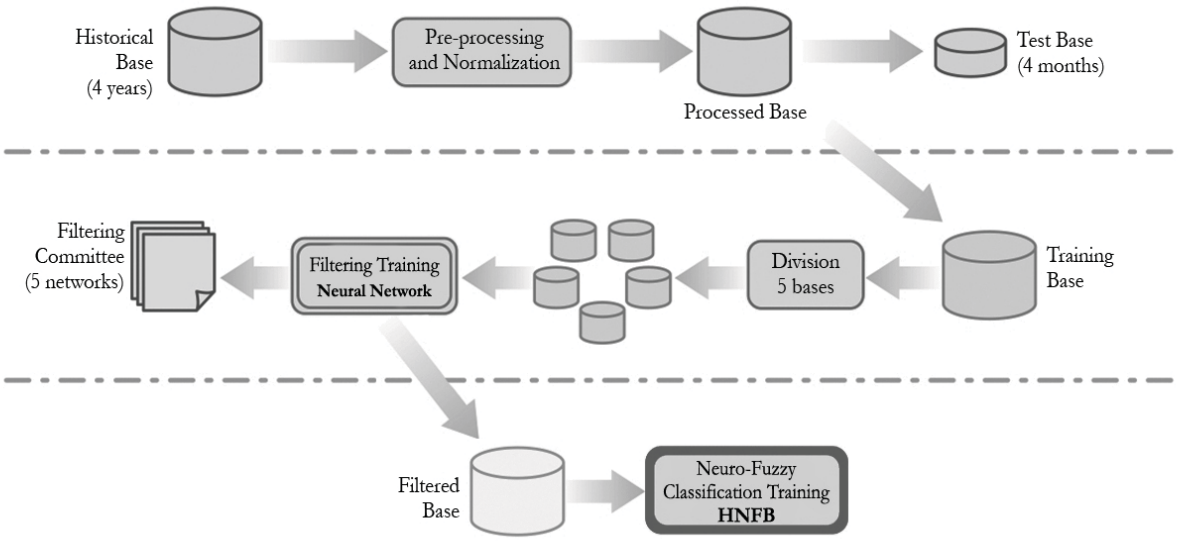


Figure 1(a): Complete methodology of the Learning Phase – Pre-processing, Filtering and Classification

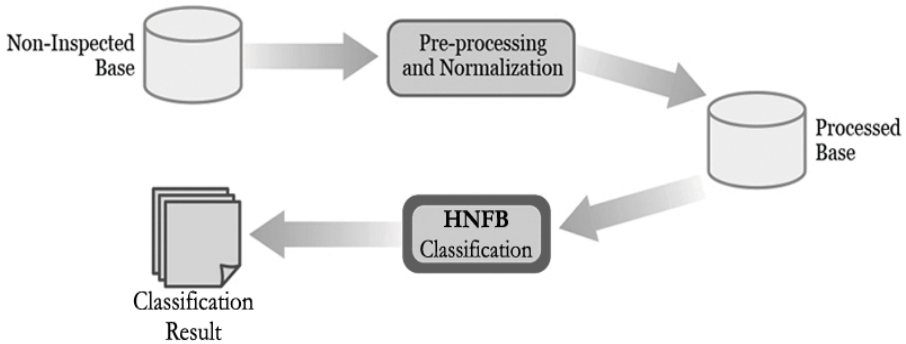


Figure 1(b): Methodology for the Recall Phase - Pre-processing and Classification modules.

ISBN: 978-989-95079-6-8

The following sub-sections present in more detail each of the modules.

3.1 Pre-Processing and Normalization Module

The data pre-processing module is composed of the following phases: selection of attributes, data cleaning and normalization. The database contains general information on both commercial/industrial and residential consumers in the so-called East Regional of Rio de Janeiro. It is well known that the behaviours of residential and non-residential clients are very distinct. Therefore, the database was divided into residential and non-residential clients, resulting in a specific Module for each class of clients.

After the evaluation of the entire available attributes, those which refer to customers' consumption for the last 4 years were selected, as well as some attributes that characterize the installation made at the consumer unit. Table 2 presents the attributes selected in this phase.

Table 2: Attributes selected from the database

Attribute	Description
Local	Consumer Unity Identification code
Inspection date	Date when the inspection was made
Related Month	Month used as a reference. Previous month to the inspection.
Origin	Main reason for the inspection
Measurer	Measuring device identifier
MonoBiTri	Indicates if the installation is mono, bi or triphasic
Consumption irregularity code	Last detected consumption irregularity
Reading irregularity code	Last confirmed reading irregularity
Month consumption	Consumer unity consumption during the related month
Previous year consumption	Registered consumption one year before the related month
Quarterly Indicator	Variation in the past three months' consumption, one by one
Adjustment Factor	Variation from the current month consumption to the same month from the previous year
Annual Indicator	Clients' consumption during the last 24 months
Tendency Factor	Clients' previous month's consumption with the consumption expected in the present month
Fraud indicator	Indicates whether the inspector found an irregularity or not

The first three attributes in the database identify, respectively, the installation where the inspection was made, the date of inspection and the month used as a reference for data acquisition. Those three attributes are presented only to identify consumer registers and are not used as inputs to the system. The five following attributes describe in detail the technical aspects of the installation and the customer's behavior. These are nominal attributes, coded as follows:

Origin – 60 different origins, divided into 4 classes (encoded in 2 binary inputs):

1. Intelligence: information extracted from the database
2. System: automatically identified information by the system of Light S.A
3. Report: complaints made by other customers or readers
4. Grouping: inspection without fraud suspicion

Measuring device – 90 models of measuring devices, grouped in 6 classes (encoded in 3 binary inputs):

1. MMD – direct mechanical measurer
2. OMMD – direct mechanical measurer (obsolete)
3. MMI – indirect mechanical measurer
4. OMMI – indirect mechanical measurer (obsolete)
5. MED – direct electronic measurer
6. OMED – direct electronic measurer (obsolete)

MonoBiTri – three different classes: monophasic, biphasic and triphasic (encoded by 2 binary inputs).

Reading Irregularity Code and *Consumption Irregularity Code* – coded in such a way that if one customer presents any kind of irregularity, his attribute will be 1; otherwise, it is zero.

The next six attributes – *consumption in a given month*, *previous year consumption* and the indicators explained below – define the consumption profile of the customer. Since these are numerical attributes, they have been normalized through:

$$value_{NORM} = \frac{v_{MAX} - value}{v_{MAX} - v_{MIN}}$$

where *value* is the attribute to be normalized, $value_{NORM}$ is the normalized attribute, v_{MAX} is the maximum attribute value registered in the last two years and v_{MIN} is the minimum attribute value registered in the last 2 years.

Four additional indicators have been computed from the monthly consumption information of each customer: *three-months moving average*, *six-months moving average*, *last 12 months average* and the *previous year average* (from the 24th previous month to the 12th previous month).

The *Fraud Indicator* attribute provides the result of the inspection, and indicates if any irregularity was found. This attribute was used as output (target) in training.

Finally, the following variables were also added to the set of attributes: *minimum* and *maximum monthly temperatures*, according to the customer's geographic area, and the *amount of consumption of all customers in the month* under analysis.

All data have been filtered (spurious data, redundant and incomplete were eliminated) and normalized [9]. Through normalization all attributes present the same interval of variation, which is important for training.

ISBN: 978-989-95079-6-8

3.2 Filtering Module

As already mentioned, the available database is rather noisy, due to some unreliable information in the Fraud Indicator attribute. In order to extract patterns from the database that are more reliable examples of actual *normal* and *irregular* customers, the Filtering Module was proposed. This consists of a committee of five neural network, all of them Multi-Layer Perceptrons (MLP) [10][11], with 21 inputs, one hidden layer and one neuron in the output layer, with sigmoidal threshold function (Fig. 2). The output neuron represents two classes: *irregular* customers (measurement or technical irregularities) and *normal* customers (those who do not show any irregularity).

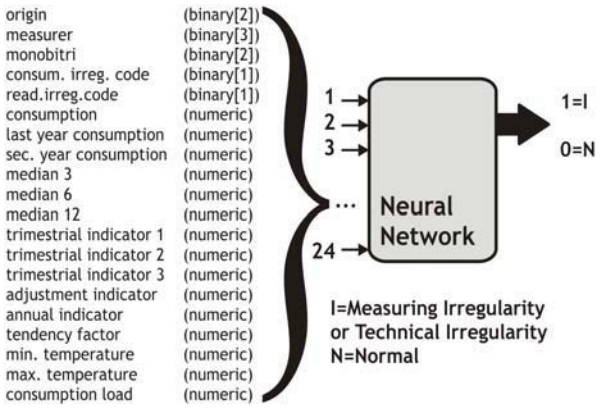


Figure 2: MLP Neural Network topology with one output

As the network output is a continuous value between zero and one, a threshold level of 0.5 was defined so as to distinguish *irregular* (output value ≥ 0.5) from *normal* customers (output value < 0.5). It should be mentioned that different threshold limits, between [0.3, 0.8] with steps of 0.05 were also evaluated, but the best performance has been achieved for 0.5.

The complete database is divided into five disjoint sets in order to train each of the networks that form the committee. After training, the complete database is processed by the neural networks. The selection of more reliable training patterns (filtering process) is carried out as follows: a pattern that is labeled as *irregular* in the dataset and is indicated by three or more votes in the committee evaluation as also *irregular* is taken as a positive true pattern, that is, a customer that actually characterizes an *irregular* pattern. On the other hand, patterns specified as *normal* in the database and indicated by at least three votes from the committee as being *normal* are called negative true patterns. These patterns are then considered as reliable examples of normal customers (see Fig. 3).

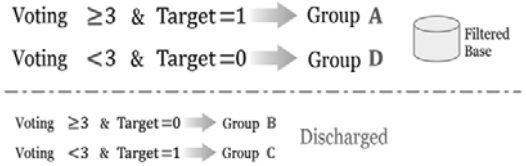


Figure 3: Limits for the Filtering Committee

A new filtered database is then formed, containing all positive and negative true patterns, which will then be used for training the Classification Module. The positively false registers (those normal patterns that receive three or more votes as being irregular from the committee) and the negatively false registers (irregular patterns that receive at least three votes as being normal) are discarded as noisy patterns and are not used to train the Classification Module.

3.3 Classification Module

As mentioned above, the main difference between this and the filtering module is the database used for training. Now only the filtered database, obtained from the filtering module, is used to adjust the classifier parameters. Once training is completed, the Classification Module is ready for use with new input patterns. In this phase, the Filtering Module is no longer necessary (see Fig. 1b), and will only be used again if retraining becomes necessary.

The structure used in this stage is a Hierarchical Neuro-fuzzy with Binary Space Partition (BSP) system [6] [12], briefly described below.

Hierarchical Neuro-Fuzzy Systems have been devised to overcome limitations of traditional neuro-fuzzy systems, which in general have a limited capacity for creating their own structure and rules. Additionally, most of the models employ grid partition of the input space, which, due to the rule explosion problem, are more adequate for applications with a smaller number of inputs. When a greater number of input variables are necessary, the system's performance deteriorates.

The Neuro-fuzzy Hierarchical BSP (HNFB) system used here makes use of basic cells. An HNFB cell is a neuro-fuzzy mini-system that performs fuzzy binary partitioning of the input space. The HNFB cell generates a crisp output after a defuzzification process. Fig. 4 illustrates the cell's functionality, where x represents the input variable, $\rho(x)$ and $\mu(x)$ are the membership functions *low* and *high*, respectively, which generate the antecedents of the two fuzzy rules, and y is the output. The linguistic interpretation of the mapping implemented by the HNFB cell is given by the following rules:

- If $x \in \rho$ then $y = d_1$
- If $x \in \mu$ then $y = d_2$

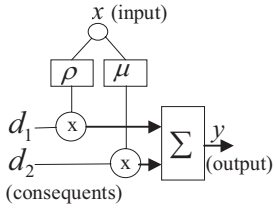


Figure 4: Neuro-Fuzzy BSP cell

Each rule corresponds to one of the two partitions generated by BSP. Each partition can in turn be subdivided into two parts by means of another HNFB

ISBN: 978-989-95079-6-8

cell. The profiles of membership functions $\rho(x)$ and $\mu(x)$ are complementary logistic functions.

The output y of an HNFB cell (defuzzification process) is given by the weighted average. Due to the fact that the membership function $\rho(x)$ is the complement to 1 of $\mu(x)$, the following equation applies:

$$y = \rho(x) * d_1 + \mu(x) * d_2 \quad \text{or} \quad y = \sum_{i=1}^{2} \alpha_i d_i$$

where α_i symbolizes the firing level of the rule in partition i and are given by $\alpha_1=\rho(x)$ and $\alpha_2=\mu(x)$. Each d_i corresponds to one of the possible consequents below:

- A singleton: d_i = constant.
- A linear combination of inputs: $d_i = \sum_{k=1}^{n} w_k x_k + w_0$

  where: x_k is the k-th input, w_k represent the weight associated with input x_k, n is the total number of inputs and w_0 is a constant value.
- The output of a generic cell j: $d_i = y_j$

The HNFB model may be described as a system that is made up of interconnections of HNFB cells, as illustrated in Fig. 5, along with the respective partitioning of the input space. In the system presented in Fig. 5, the initial partitions 1 and 2 ('BSP0' cell) have been subdivided; hence, the consequents of its rules are the outputs of BSP1 and BSP2, respectively. In turn, these subsystems have, as consequents, values d_{11}, y_{12}, d_{21} and d_{22}, respectively. Consequent y_{12} is the output of the 'BSP12' cell. The output of the system in Fig. 3(b) is:

$$y = \alpha_1(\alpha_{11}d_{11} + \alpha_{12}(\alpha_{121}d_{121} + \alpha_{122}d_{122})) + \alpha_2(\alpha_{21}d_{21} + \alpha_{22}d_{22})$$

Although each BSP cell divides the input space only into two fuzzy sets (*low* and *high*), the complete HNFB architecture divides the universe of each variable into as many partitions as necessary. The number of partitions is determined during the learning process. In Fig. 5, for instance, the upper left part of the input space (partition 12) has been further subdivided by the horizontal variable x_1, resulting in three fuzzy sets for the complete universe of discourse for this specific variable.

The training algorithm makes use of the gradient descent method for learning the structure of the model and, consequently, linguistic rules. The parameters that define the the membership functions of the antecedents and consequents are regarded as fuzzy weights of the neuro-fuzzy system.

A tuning parameter δ, called decomposition rate prevents the structure from growing indefinitely.

4 Experimental Results

This section presents the results obtained for Filtering, the Neural Networks Committee, and Classification, performed by the HNFB system.

4.1 Filtering

After processing the selected data three databases were obtained. These were composed of 4159 customers with non technical irregularity, 3754 customers with technical irregularity and 14405 normal customers.

Data were divided into Training, Validation and Testing databases. Training and Validation sets refer to the period between September 2002 and June 2006, with the exception of February and March 2006. The data for these two months, together with those for July and August 2006 were used to form the test set. These months were selected to evaluate the system's performance in different seasons (summer and winter in the southern hemisphere). Due to the unbalanced characteristic of the database, with many more examples of *normal* customers than the available number of *irregular* patterns (technical and non-technical irregularities), the training and validation sets were equalized to avoid a biased behaviour of the neural networks. The datasets actually used in the experiments consisted of 3754 customers with technical irregularities, 3754 randomly selected customers with non-technical irregularities and 7508 normal ones, also randomly chosen. For training, 75% of the balanced samples have been considered; the rest was left for validation.

Five different databases were created, each to train one of the five neural networks that form the filtering committee. For each neural network, 10 training processes were realized and the best performing network from the validation set has been selected to form the committee. The early stopping methodology was used during each training process, which specifies the optimal number of epochs to avoid over fitting and attain best generalization performance.

As the main objective of the proposed system was to increase the PPV of irregularity detection, two experiments were carried out: one considering the PPV as the error metric for the neural networks training, and another considering the general classification error as the metric used to select the best neural network. It was observed that the latter provided better results, since when the classification error for both classes is minimized, the PPV is indirectly maximized. Thus, in the following experimental results, the general classification error metric was used.

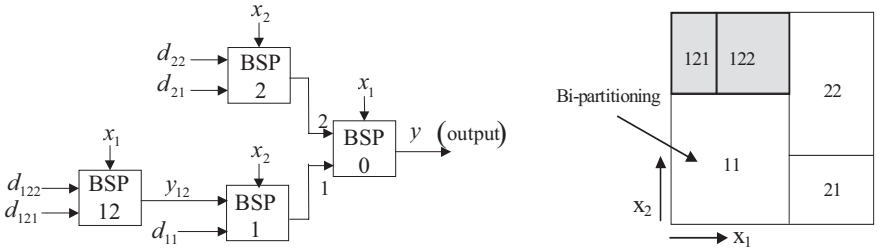


Figure 5: Example of an HNFB system and respective input space Partitioning

ISBN: 978-989-95079-6-8

After all neural networks had been trained and the committee members had been selected, a test was carried out with the data of customers investigated during the months of February, March, July and August 2006. That test base has 4663 registers of inspected customers.
Table 3 illustrates filtering results for one specific ZEI (Elementary Intervention Zones) of East Regional, in the same fashion of Table 1.

Table 3: Results with the filtering committee

	Committee	
	Irregular	Normal
Irregular	960	1164
Normal	572	1970
	PPV_{NN}:	45.2%
	Class. Error:	37.2%
	PPV_{Light}:	32.8%

As can be observed, by using only the Filtering Module the PPV has already increased from 32.8% to 45.2%.

4.2 Classification

The same inputs considered for filtering were initially considered for classification, performed by the HNFB system. After variable selection by the least-square estimator (LSE) [13], nine were effectively used as inputs to the system. The test database was identical to that used for testing in the filtering stage.

It can be seen from the classification results shown in Table 4 that the VPP has improved even further, reaching 51.2%. This attests the advantage of the proposed methodology.

Table 4: Results for classification

	Committee	
	Irregular	Normal
Irregular	991	943
Normal	541	2191
	PPV_{HNFB}:	51.2%
	Class. Error:	31.8%
	PPV_{Light}:	32.8%

In addition to performing classification, the HNFB also produces fuzzy rules as a result. In this experiment 45 rules have been generated, of which an example is:
If *Origin* is *low* and *minimum temperature* is *low* then class is *normal*.

5 Conclusions

This paper presented an intelligent system, based on a neural networks committee and on a Hierarchical Neuro-fuzzy BSP system, to identify consumers' frauds in an electrical distribution company of Rio de Janeiro, Brazil.

The system is formed by two modules, one for filtering the database and another for actual classification of a consumer. The use of this methodology greatly improved the current fraud detection rate.

Besides increasing the performance regarding the correct identification of fraudulent consumers, the system has the additional advantage of ranking irregular costumers as a result of the number of votes the neural networks committee provides to identify a customer as irregular.

The proposed methodology is being applied to other areas covered by the company. Additionally, a clustering algorithm will be applied to non-residential consumers in order to group them in classes with similar consumption behaviour.

References

[1] J. E. C. Junior, *Detecção de Fraudes em Unidades Consumidoras de Energia Elétrica Utilizando Rough Sets*, M.SC. Dissertation, UFMS (in Portuguese), 2005.

[2] J. E. Cabral, E. M. Gontijo, J. O. P. Pinto and J.R. Filho, Fraud Detection in Electrical Energy Consumers using Rough Sets, *IEEE International Conference on Systems, Man and Cybernetics*, 4: 3625-3629, 2004.

[3] T. Rauber, I. Drago, F. Varejão and R. Queiroga, Extracão e Seleção de Características na Identificação de Perdas Comerciais na Distribuição de Energia Elétrica, *XXV Congress of the Brazilian Computation Society* (in Portuguese), 2005.

[4] J. Rong, H. Tagaris, A. Lachsz and M. Jeffrey, Wavelet based Feature Extraction and Multiple Classifiers for Electricity Fraud Detection, *Transmission and Distribution Conference and Exhibition*, 3: 2251-2256, 2002.

[5] D. S. G. Pollock, *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*. Academic Press, New York, 1999.

[6] M. Vellasco, Pacheco, K. Figueiredo, F. Souza, Hierarchical Neuro-Fuzzy Systems - Part I, *Encyclopedia of Artificial Intelligence*, Juan R. Rabuñal, Julián Dorado & Alejandro Pazos (Eds.), Information Science Reference, 2008.

[7] R. Santos, M. Vellasco, F. Artola and S. Fontoura. Neural Net Ensembles for Lithology Recognition, *4th International Workshop on Multiple Classifier Systems*, LCNS 2709 – Multiple Classifier Systems, Springer-Verlag, 2003.

[8] F. Provost, T. Fawcett, and R. Kohavi, (1998), The Case Against Accuracy Estimation for Comparing Induction Algorithms, *15th International Conference on Machine Learning*, 445-453, 1998.

[9] M. Dash and H. Liu, Feature Selection for Classification, *Intelligent Data Analysis*, Elsevier, 1(3): 131 - 156, 1997.

[10] S. Haykin (1999), *Neural Networks - a Comprehensive Foundation*, Prentice-Hall, New Jersey, 1999.

[11] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[12] L.B. Gonçalves, M.M.B.R. Vellasco, M.A.C. Pacheco, F. J. de Souza, Inverted Hierarchical Neuro-Fuzzy BSP System: A Novel Neuro-Fuzzy Model for Pattern Classification and Rule Extraction in Databases, *IEEE Transactions on Systems, Man & Cybernetics, Part C: Applications and Review*, 36(2): 236-248, March 2006.

[13] F.-L. Chung and J.-C. Duan, On Multistage Fuzzy Neural Network Modeling, *IEEE Transactions on Fuzzy Systems*, 8(2): 125-142, 2000.

ISBN: 978-989-95079-6-8

A Fuzzy System for the Assessment of Human Reliability

D.M. Szwarcman, J. Domech, R. Tanscheit, M. Vellasco, M. Pacheco, N. P. Ribeiro

Computational Intelligence Lab., Department of Electrical Engineering,
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, Brazil
Email: {dilza, , ricardo, marley, marco, nick}@ele.puc-rio.br; jesus.more@estacio.br

Abstract *— This work presents a methodology for the characterization of human reliability based on fuzzy sets concepts, which has been implemented in an innovative decision support system, providing managers with an intelligent computational tool for reducing the possibility of human errors in industrial activities. Considering that such activities can be described as operational, maintenance or inspection processes, which are composed of a set of procedures, the methodology is carried out in two levels: the process level and the procedure level. The proposed system provides a human reliability index, which allows the identification of problems that may constitute causes of human errors, as well as the indication of possible strategies for the control of potentially adverse impacts of interactions that add uncertainty and complexity to processes.*

Keywords— fuzzy sets, human reliability, industrial process safety.

1 Introduction

Human reliability has received much attention in areas such as nuclear, aviation and petrochemical industries. Although resources have been historically applied mostly in equipment reliability and process optimization, it has been realized that priority should be given to the study of human reliability, focusing on the adaptation of the equipment and the working environment to the capabilities, limitations and needs of the human being. Human error usually arises from inadequacies of the system design, such as task complexity and error-likely situations. Errors are likely to occur when the task requirements exceed human limitations regarding perception, attention, remembering, etc. Some situational characteristics may predispose operators to errors, such as inadequate workspace and training procedures, as well as poor supervision. Errors may also reflect individual differences, related to human attributes such as abilities and attitudes. Important individual factors are susceptibility to stress and inexperience, for example, which may increase tenfold the possibility of occurrence of a human error.

Studies in human reliability are generally divided into two generations [1]. First-generation methods are characterized by comparing human performance to that of a machine, associating probabilities of success or of a fault to the operators' actions. Second-generation methods extend the analysis of human reliability to cognitive systems, by considering decision levels, diagnosis processes, dexterity, knowledge and organizational factors. These are probability-based methods, which makes it difficult to establish a precise model for human fault prediction, since a large quantity of data is needed for mapping all the uncertainties inherent to human behavior.

Probabilistic analysis is used for analyzing system reliability objectively and assumes that an equipment or human failure occurs at random. A failure of a single component may occur at random; a human error, however, does not necessarily occur in that way, since a human factor is composed of a large number of attributes (or performance shaping factors) and its functional structure is very complex. By using the probabilistic approach, where the equipment and procedure are qualified, it is assumed that the operator correctly implements all the procedure's provisions and thus isolates the human factor elements.

Other works [2] have contributed to human reliability research employing Fuzzy Set Theory and the concept of possibility of failure instead of probability of failure. Lian and Wang [3] used fuzzy relations to estimate the *fuzzy probability*. Nowakowski [4][5] observed that, when applying the fuzzy approach, probabilistic interpretations of human reliability are abandoned; instead, human reliability is defined in terms of possibility measures. Wang [6] used the concept of fuzziness to evaluate human performance in an inspection task. Onisawa [7] analyzed human reliability in the events that preceded the nuclear accident in Chernobyl and showed the importance of considering the possibility approach in human reliability analysis. More recently, Onisawa [8] created a model that integrates the subjectivity in specialists' opinions to reliability analysis.

This work presents a fuzzy-based decision support system for the analysis of human reliability in operation, maintenance and inspection activities in industrial and production processes, where the human error may have a great impact on safety and on the environment. As operational, maintenance and inspection processes consist of a set of procedures, the human reliability characterization method is carried out at both process and procedure levels.
This paper comprises four additional sections. Section 2 presents the methodology, describing in detail the human reliability characterization at both process and procedure levels. Section 3 presents a decision support system developed to provide managers with computational tools for the application of the human reliability characterization method, helping them to choose the most adequate strategic actions to be taken. Section 4 discusses a case study in the oil industry and Section 5 concludes the work.

2 Proposed Methodology

This work deals with a two-level analysis of human reliability in industrial activities. On a first level, each of the operation, maintenance and inspection processes are

ISBN: 978-989-95079-6-8

analyzed as a whole; then, on a second level, human reliability is considered for every procedure that composes each of the processes.

The following sub-sections describe in details the two-level analysis contemplated by the proposed methodology.

2.1 Process Level Characterization

This module is an improved version of the process level characterization first introduced in [9].

The characterization of human reliability in a process has the objective of establishing the degree of attendance of Performance Shaping Factors (PSFs), which can be of human, technical or environmental types. Such characterization may be carried out in any category: operation, maintenance or inspection.

The method for determining the degree of attendance of a set of attributes (see Fig. 1) begins with the selection of PSFs that affect the human being. After that, experts specify, through a questionnaire, the influence of each of these factors on human reliability. By aggregating the experts' opinions about the PSFs' influences, weighed by the degree of importance of each expert, a standard degree of influence is established for each PSF. By considering the degrees for all PSFs, a Quality Standard is obtained. In a second stage, workers' opinions on the attendance of each of the PSFs are collected, also through a questionnaire, producing the degree of attendance of each PSF [9].

For human PSFs that are related to cognitive characteristics, namely intelligence, communication ability, sociability, attention, stress and anticipation, the attendance degree is determined by psychometric tests applied to each operator. It is considered that, for such type of factors, the operators opinion would correspond to a self judgment, which could result in biased information about the attendance degree. Psychometric tests are likely to provide more reliable results, since they are constructed to determine individual psychosomatic characteristics in an unbiased manner. The enhanced process level characterization method is depicted in the diagram of Fig. 1, which highlights the Psychometrics tests module, one of the significant contributions of this work.

2.2 Procedure Level Characterization

The human reliability characterization at procedure level focuses on two main objectives: establish a failure possibility index for each procedure in an operation, maintenance or inspection process, and for each of the human actions that are performed to accomplish it; and secondly, establish the degree of attendance for each worker with respect to the cognitive demands of each action within the procedure.

The characterization method requires that experts specify the failure expectation for each action, the degree of dependency of each action on the preceding one, the influence that a failure in one action may have on subsequent ones and the cognitive demands of each action. All this information is acquired through the application of questionnaires, where experts give linguistic answers such as *low*, *medium*, or *high*. The experts' opinions are aggregated and weighed by the degree of importance of each expert so as to produce, for each action, a failure possibility index and standard demand degrees with respect to the cognitive factors considered. The overall failure index for the procedure can be obtained by combining the actions indexes. By submitting the workers to psychometric tests, the degree of attendance of each cognitive factor for all operators can be determined. The method is depicted in the diagram of Fig. 2.

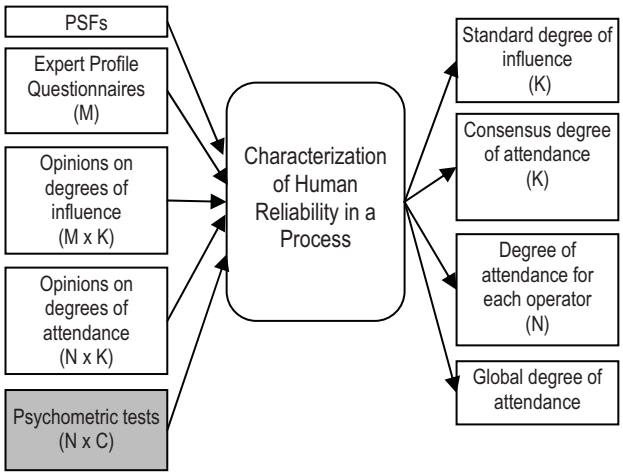


Figure 1: Characterization of Human Reliability in a Process, with K PSFs, C cognitive factors, M experts and N operators.

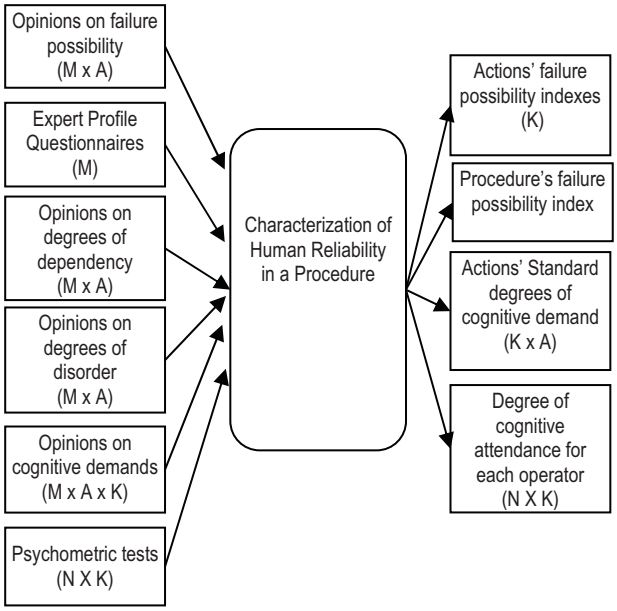


Figure 2: Characterization of Human Reliability in a Procedure, with A actions, K cognitive factors, M experts and N operators.

The procedure level characterization methodology consists of the following stages.

Establishment of a committee of decision-makers

This is one of the most important steps in the methodology, since the quality of information will depend on the experts' proficiency.

Establishment of the relative importance of each expert

This step is accomplished through an Expert Profile Identification Questionnaire (EPIQ), which consists of a set

ISBN: 978-989-95079-6-8

of questions for the evaluation of each expert's relative importance and thus assigning him a weight [10]. This weight will establish the influence of the expert's opinions on the final standard degrees and failure possibility indexes.

Choice of linguistic values for the evaluation of actions

This stage consists of choosing linguistic terms, or values, for the evaluation (by experts) of the expectation of failure of an action, its degree of dependency on the preceding one, its influence on subsequent ones and its cognitive demands. For instance, the terms used for the evaluation of the expectation of failure usually are: *very high, high, medium, low,* and *very low***.** All terms are associated to triangular fuzzy sets, defined by three parameters.

Acquirement of experts' opinions

This step consists of obtaining from the selected experts, through questionnaires, their opinions on the expectation of failure of each action, its degree of dependency on the preceding action, its influence on succeeding ones and its cognitive demands.

Application of psychometric tests to operators

Psychometric tests are applied to operators in order to obtain the degree of attendance of cognitive demands for each worker.

Fuzzy treatment of the data provided by experts and of psychometric tests results

Here, the individual prognoses from each expert for failure expectation, dependency, influence on future actions and cognitive demands are aggregated, generating a consensus for each evaluated attribute. The Hsu and Chen's model [11] is used to pool the expert's opinions: the similarity aggregation method (SAM) is used for combining the opinions of each expert. The opinion of an expert i is expressed by a fuzzy set denoted by $\tilde{A}_i$. The agreement degree (or similarity measure) $S(\tilde{A}_i, \tilde{A}_j)$ between two experts i and j can be determined by the proportion of the consistent area to the total area:

$$S(\tilde{A}i,\ \tilde{A}j) = \frac{\int_x \min(\mu_{\tilde{A}_i}(x), \mu_{\tilde{A}_j}(x))dx}{\int_x \max(\mu_{\tilde{A}_i}(x), \mu_{\tilde{A}_j}(x))dx} \quad (1)$$

Once all agreement degrees between experts are measured, an agreement matrix (AM) can be built, giving an insight into the agreement between the experts.

$$AM = \begin{bmatrix} 1 & S_{12} & S_{13} & ..... & S_{1n} \\ S_{21} & 1 & S_{23} & ..... & S_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ S_{n1} & S_{n2} & S_{n3} & ..... & 1 \end{bmatrix} \quad (2)$$

The average agreement degree AAD_i of expert E_i (i = 1 to n) is given by:

$$AAD_i = \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq i}}^{n} S_{ij} \quad (3)$$

The relative agreement degree RAD_i of expert E_i (i = 1 to n) is given by:

$$RAD_i = \frac{AAD_i}{\sum_{i=1}^{n} AAD_i} \quad (4)$$

Finally, by weighing the relative agreement degree of each expert k by the degree of importance Gi_k (obtained through a questionnaire), the consensus coefficient for that expert can be calculated by equation (5).

$$CC_k = \frac{RAD_k * Gi_k}{\sum_{i=1}^{n}(RAD_i * Gi_i)} \quad (5)$$

The aggregated opinions are given by:

$$\tilde{N} = \sum_{i=1}^{n} CC_i \times \tilde{A}_i \quad (6)$$

In order to obtain a single value for each of the evaluated attributes – representing the failure expectation index, the dependency and adverse influence degrees, and the cognitive demands for each action, the *max* defuzzification method is applied [12].

For the case of cognitive demand, after normalization, and considering all cognitive demands for all actions, a Cognitive Demand Standard is built.

The overall failure possibility index for an action i (P_i) is obtained from its failure expectation index (F_i), its dependency degree on the previous action ($D_{i,(i-1)}$) and the adverse influence degree caused by failure of preceding actions ($T_{j,i}$; where $j< i$):

$$P_i = (1-(1-F_i)*(1-D_{i-(i-1)})*\prod_{j<i}(1-T_{j,i})) \quad (7)$$

The overall failure possibility for a procedure is obtained from the actions indexes:

$$P_{procedure} = (1-\prod_{i}(1-P_i)) \quad (8)$$

Finally, a degree of attendance of the operators to the Cognitive Demand Standard of each action is obtained. For that matter, the degree of attendance (OP_k) to each cognitive demand is multiplied by its demand standard (DS_k) and a weighted average R is then calculated:

$$R = \frac{\sum_{k=1}^{n} DS_k * OP_k}{\sum_{k=1}^{n} DS_k} \quad (9)$$

3 Decision Support System

The decision support system for human reliability analysis was developed as an Intranet application, maintaining all information in a centralized database and being accessible to all professionals involved in a characterization project: managers, experts, operators and psychologists.

The system is totally configurable to any company or business area and can de used for characterization and analysis at process and procedure levels, allowing for:

- human reliability characterization trials or *projects*, as they are called in the system, for different units of the company and different periods of time;
- the complete configuration of the PSFs to be considered, as well as the corresponding questionnaires to be applied to experts and workers, so

ISBN: 978-989-95079-6-8

that questions can be set up in accordance with the personnel's profile and company's characteristics;
- the complete configuration of the sequence of actions that make up the procedures to be analyzed;
- the on-line answering of EPIQ, PSF and procedure evaluation questionnaires by experts;
- the on-line answering of PSF questionnaires by operators;
- the on-line input of the psychometric test results for all workers by an authorized psychologist;
- the association of reference files to each project, such as manuals or norms that must be followed on the given human reliability characterization trial.

The system provides two sets of reports that give managers effective support on strategic decisions regarding the improvement of human reliability, and, consequently, the minimization of process failure.

The first set of reports, called *Process Reports*, comprises results from the process level characterization, including:

- Degree of importance of *PSFs*:
  The opinions of all experts with respect to the degree of importance of all PSFs are presented in a comparative chart.
- Average degree of attendance of PSFs:
  The average degrees of attendance of all PSFs, as seen by each operator, are presented in a comparative chart. A degree close to 1 indicates that, in the environment being evaluated, the operator considers that, in average, the quality standard is being met. On the other hand, a degree closer to 0 indicates that the operator evaluates that, in average, the quality standard is not being attended to. In this case, the possibility of failure for this worker is high.
- Distance to quality standard for all PSFs, as seen by a given operator:
  The quality standard for PSFs, established by the experts, represents the PSF *demand*, while the attendance degrees specified by a worker represent the PSF *offer*. This report shows graphically and in tabular form the distances between offer and demand. The greater the distance between the attendance degree and the quality standard, the more efforts and investments should the company dedicate to the given PSF to increase its attendance level. Factors with distances larger than 0.5 should merit special attention (Fig. 3).
- Inclusion coefficients for all PSFs, as seen by a given operator:
  This report shows the inclusion degree of the fuzzy set *quality standard* in the fuzzy set *attendance*, given by each operator. An inclusion coefficient close to 1 indicates that, in the environment being evaluated, the particular operator assimilation is greater than the quality standard.

The second set of reports, called *Procedure Reports*, present results from the procedure level characterization, including:

- Failure expectation for actions:
  The experts' aggregated expectations with regard to the failure of each action are presented. Failure expectations are presented together with the corresponding aggregated certainty degrees, which indicate the overall degree of sureness of the experts with respect to their opinions.

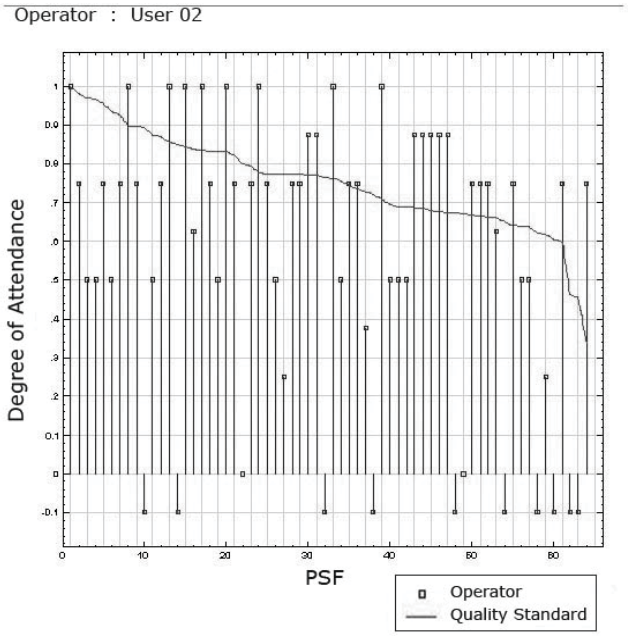


Figure 3: Distance to quality standard for all PSFs, as seen by a given operator.

- Dependency degree of actions:
  The experts' aggregated opinions on the dependency of each action on the previous one are presented. A degree close to 1 indicates that a failure on the previous action will most probably cause a failure on the current action. Actions with degrees higher than 0.6 deserve special attention. The dependency degrees are also presented together with the related aggregated certainty degrees.
- Influence of actions:
  In a matrix form, this report presents the experts' aggregated opinions about the possibility of a failure in each action having an influence on succeeding actions (Fig. 4). Degrees higher than 0.6 should be analyzed.

Action	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**1.00**	**0.71**	**0.75**	**0.85**	**0.76**	**0.76**	**0.72**	**0.76**	**0.76**	**0.76**	**0.76**	**0.76**	**0.83**	**0.65**	**0.83**	**0.86**	**0.83**	**0.75**
2		**1.00**	0.50	0.54	0.50	0.54	**0.69**	**0.76**	**0.92**	**0.76**	**0.83**	**0.76**	**0.72**	**0.65**	**0.81**	**0.75**	**0.81**	**0.75**
3			**1.00**	0.25	0.27	0.32	0.32	**0.83**	**0.66**	**0.72**	**0.65**	**0.72**	0.41	0.18	0.18	0.19	0.17	0.19
4				**1.00**	0.59	0.06	0.17	0.33	0.33	0.17	0.25	0.33	**0.69**	0.41	0.59	0.41	**0.67**	0.48
5					**1.00**	**0.81**	0.54	**0.76**	**0.75**	0.25	0.16	**0.86**	0.51	0.25	0.33	0.10	0.27	0.10
6						**1.00**	0.42	**0.76**	0.52	0.34	0.16	**0.76**	0.42	0.17	0.27	0.10	0.27	0.17
7							**1.00**	**0.86**	**0.83**	**0.83**	**0.86**	**0.76**	**0.72**	**0.67**	**0.76**	0.58	**0.66**	0.47
8								**1.00**	0.19	0.19	0.19	0.41	0.17	0.25	0.06	0.18	0.13	0.06
9									**1.00**	**0.83**	0.32	0.19	0.27	0.25	0.27	0.18	0.18	0.19
10										**1.00**	0.09	0.09	0.33	0.17	0.50	0.17	0.18	0.18
11											**1.00**	0.41	0.06	0.27	0.25	0.33	0.10	0.02
12												**1.00**	**0.81**	0.27	0.17	0.27	0.17	0.18
13													**1.00**	0.59	0.32	0.32	0.27	0.42
14														**1.00**	0.32	0.32	0.27	0.36
15															**1.00**	0.59	0.42	0.42
16																**1.00**	0.37	0.41
17																	**1.00**	0.52
18																		**1.00**

Figure 4: Influence of actions on subsequent ones

- Cognitive demands of actions:
  This report presents the degree with which a cognitive aspect is demanded for a worker to correctly execute an action (in the consensual opinion of all experts). The

ISBN: 978-989-95079-6-8

degree of attendance of each cognitive factor for each operator is also shown. It indicates on which cognitive aspects the worker should receive training. Cognitive demands higher than 0.6 deserve special attention.

- Procedure overall failure possibility:
  The overall failure possibility indexes for the procedure and its actions are shown.

4 Case Study

An oil refinery was chosen as the unit for the case study, namely REDUC (Refinery of Duque de Caxias). This is located at Campos Elísios district, Rio de Janeiro, Brazil, and produces petrol, lubricants and other oil derivatives. The object of this case study is Procedure 1 (part of the Refinery Operation Process [9]), composed of the actions listed in Fig. 5.

The influence on each action by each cognitive factor was evaluated by 17 experts, who answered the questionnaire. The cognitive factors studied were: Anticipation, Attention, Calculation, Emotional State, Group Identification, Intelligence, Memory, Perception, Personality and Stress.

Taking as a basis the API 770 norm, two PSFs questionnaires, one for operators and another for experts, were configured for the considered unit. Both questionnaires correlate one question to each PSF. By using this specific questionnaire, an expert can evaluate which are the most critical PSFs. The system generates the PSFs quality standard by putting together the answers of all experts and weighing them based on EPIQ. By answering his questionnaire, an operator evaluates each PSF attendance. To evaluate the attendance of PSFs associated to cognitive factors, a questionnaire is used by a specialized psychologist, who analyses, based on standard psychology tests, how each interviewed operator is responding to each cognitive factor.

In total, 51 workers of the Operation Process answered the PSFs questionnaire, giving opinions about their working conditions. The 17 available experts answered the questionnaire in order to determine which PSFs had higher need to be attended. They also answered the EPIQ and the questionnaires about actions of Procedure 1, that is, expectation of failure, degree of dependency, influence on subsequent actions and cognitive demands. After all operators and experts of REDUC answered their respective questionnaires, the final reports could be consulted in the system by *project* administrators and managers.

The reports obtained in this case study are described below.

Report of Failure Expectation of Actions

From the report presented in Fig. 6, it is possible to conclude that the actions which had the higher failure expectation were: Return of Line to The Operation, Release of Line to The Maintenance, Return of Tank to The Operation, Aliments Change, and Area Survey.

Action Dependency Ratio Report

This report, depicted in Fig. 6, demonstrates that the actions which had higher dependency ratio on the preceding action were: Return of Pot Flare (18), Return of Line to The Operation (16), and Product Shipments to CIAS, BR/GEI, PETROFLEX (8).

Action Influence Report (see Fig. 4)

Based on this report, it is possible to conclude that actions Demands (1) and Aliments Change (7) have more influence on subsequent actions than any other. Additionally, Product Shipments to CIAS, BR/GEI, PETROFLEX (8) and Loading of Trucks (12) are most affected than other actions.

	Action	Failure Expectation	Certainty Degree
1	Demands	0.0000	1.0000
2	Area Survey	0.3042	1.0000
3	Measurements and Sampling	0.2370	0.7865
4	Drainages	0.1043	0.7291
5	Pumps Starts	0.0269	0.9490
6	Pumps Stops	0.0269	0.8901
7	Aliments Change	0.4313	0.8901
8	PETROFLEX	0.2500	1.0000
9	Exchange Loads of Units	0.2298	1.0000
10	Exchange Receipt of Units	0.2298	0.9314
11	Gravitations	0.0269	0.9314
12	Loading of Trucks	0.1367	0.8901
13	Release of Equipments to Maintenance	0.2500	0.0784
14	Return of Tank to The Operation	0.4458	0.6542
15	Release of Line to The Maintenance	0.4844	0.6542
16	Return of Line to The Operation	0.4844	0.6542
17	Release of Pot Flare	0.2772	0.8901
18	Return of Pot Flare	0.2772	0.8901

Figure 5: Failure Expectation of Actions for Procedure 1.

Action	Dependency Ratio	Certainty Degree
1	--	1.0000
2	0.5334	1.0000
3	0.3855	0.7865
4	0.4215	0.7291
5	0.3336	0.9490
6	0.3067	0.8901
7	0.2840	0.8901
8	**0.6192**	1.0000
9	0.0303	1.0000
10	0.5510	0.9314
11	0.4388	0.9314
12	0.1043	0.8901
13	0.0378	0.0784
14	0.5055	0.6542
15	0.3623	0.6542
16	**0.7674**	0.6542
17	0.5334	0.8901
18	**0.9334**	0.8901

Figure 6: Dependency ratio between each action and the preceding one

Cognitive Demand Attendance Degree of Actions

ISBN: 978-989-95079-6-8

This report, shown in Fig. 7, is used for comparison between the cognitive demand of actions of Procedure 1 and the attendance of each worker. Therefore, a manager can infer whether an operator is capable of executing the actions he/she is assigned to. For instance, it can be seen in Fig. 7 that the given operator does not attend the Anticipation requirement of actions 9, 13 and 16.

Global Index of Failure Possibility of Actions

This report, shown in Fig. 8, indicates the possibility of failure of each action, considering every influence it experiences (such as influences and dependencies from other actions, failure expectation). The system allows the manager to choose the cut level of the influence to be considered. A higher cut level means that the preceding actions which have low influence will not be considered. As a result, in this case study, a 0.6 cut level returns the mid and final actions as the ones with the highest failure possibilities. In opposition, a 0.9 cut level returns a more disperse failure profile, since it only considers the preceding actions that have large influence on the action itself.

	Anticipation	Attention	Calculation	Emotive State	Group Id	Intelligence	Memory	Perception	Personality	Stress
1	0.600	**0.615**	0.262	**0.645**	0.523	0.466	**0.800**	0.458	**0.724**	0.472
2	**0.700**	**0.779**	0.047	**0.700**	0.600	**0.615**	**0.700**	**0.818**	**0.713**	0.577
3	0.600	**0.949**	**0.770**	**0.665**	0.600	**0.690**	**0.700**	**0.738**	**0.700**	**0.624**
4	0.516	**0.762**	0.431	0.521	0.581	0.566	0.100	**0.636**	0.455	**0.705**
5	**0.614**	**0.716**	0.554	**0.633**	0.414	0.600	0.509	**0.905**	0.570	0.494
6	**0.614**	**0.734**	0.197	0.509	0.413	0.600	0.414	**0.843**	**0.638**	0.410
7	**0.781**	**0.955**	**0.700**	**0.667**	**0.762**	**0.638**	**0.732**	**0.734**	**0.694**	**0.662**
8	**0.800**	**0.941**	**0.769**	**0.667**	**0.669**	**0.609**	**0.613**	**0.764**	**0.686**	**0.641**
9	**0.973**	**1.000**	**0.700**	**0.654**	**0.800**	**0.668**	**0.620**	**0.782**	**0.700**	0.600
10	**0.800**	0.000	0.539	**0.700**	0.000	**0.800**	**0.800**	**0.700**	**0.700**	0.600
11	**0.700**	**0.658**	**0.700**	**0.700**	0.000	0.570	0.558	**0.666**	**0.700**	0.600
12	**0.843**	**0.805**	**0.632**	**0.700**	**0.773**	0.600	0.600	**0.669**	**0.700**	**0.628**
13	**0.881**	**0.863**	0.600	**0.700**	**0.604**	**0.721**	**0.854**	**0.762**	**0.691**	0.403
14	**0.822**	**0.884**	0.390	0.533	**0.613**	**0.679**	**0.939**	**0.790**	**0.689**	0.500
15	**0.858**	**0.812**	**0.700**	**0.632**	**0.690**	**0.700**	**0.868**	**0.800**	**0.666**	0.500
16	**0.973**	**0.900**	**0.700**	**0.632**	**0.690**	**0.679**	**0.895**	**0.810**	**0.689**	0.500
17	**0.869**	**0.939**	**0.647**	0.579	**0.791**	**0.692**	**0.868**	**0.889**	**0.666**	0.513
18	**0.869**	**0.939**	**0.647**	**0.718**	**0.791**	**0.628**	**0.869**	**0.887**	**0.666**	**0.628**
Attendance										
User 35	**0.875**	**0.875**	--	--	**1.000**	0.500	--	--	--	0.125

Figure 7: Cognitive Demand Attendance

5 Conclusions

An innovative system for the evaluation of human reliability in industry has been presented. The approach taken makes use of fuzzy sets, so that experts and operators' opinions can be translated into mathematical terms. As a consequence, a Quality Standard (QS) can be established and the attendance to performance shaping factors can be evaluated.

This system can be used at both process and procedure (consisting of series of actions) levels and produces several reports that help managers to make decisions aimed at reducing the possibility of human errors. A case study considered a unit of a Petrobras (Brazilian Oil Company) refinery, where the system has been actually applied to.

Acknowledgment

The authors would like to thank Petrobras S.A. for the financial support to develop the decision support system.

Action	Global Index of Failure	
	(Cut Level 0,6)	(Cut Level 0,9)
1	0.0000	0.0000
2	0.5334	0.5334
3	0.4632	0.4632
4	0.4495	0.4495
5	0.3416	0.3416
6	0.5036	0.3131
7	0.6444	0.3807
8	0.9759	0.6554
9	0.9083	0.5796
10	0.9841	0.6108
11	0.9049	0.4480
12	0.9309	0.1592
13	0.9586	0.0761
14	0.8948	0.5223
15	0.8946	0.5236
16	0.9210	0.8264
17	0.9217	0.6403
18	0.9703	0.9452

Figure 8: Global Index of Failure Possibility of Actions.

References

[1] Y. Fujita, Human Reliability Analysis: a Human Point of View. *Reliability Engineering and System Safety* 38, 1992.

[2] Y. Nishiwaki, Human Factors and Fuzzy Set Theory for Safety Analysis. In: M.C. Cullingford, S.M. Shah, J.H. Gittus (Eds), *Implications of Probabilistic Risk Assesment*, Elsevier, 253-274, 1987.

[3] G.S. Liang, W.J. Wang, Evaluating Human Reliability using Fuzzy Relation, *Microelectronics and reliability*, 33(1): pp 63-80., 1993.

[4] M. Nowakowski, Possibility Distributions in the Linguistic Theory of Actions, *Int. Journal of Man-Machine Studies*, 12: 229-239, 1980.

[5] M. Nowakowski, The Human Operator: Reliability and Language of Action Analysis. In: W. Karwowski, A. Mital (Eds), *Applications of Fuzzy Set Theory in Human Factors*, Elsevier, 165-177, 1986.

[6] M. J. Wang, J. Sharit, and C.G. Drury, An Application of Fuzzy Set Theory for Evaluation of Human Performance on an Inspection Task, In: W. Karwowski, and A. Mital (Eds.) *Application of Fuzzy Set Theory in Human Factors*, Elsevier, 257-268, 1986.

[7] T. Onisawa, Y. Nishiwaki, Fuzzy Human Reliability Analysis on the Chernobyl Accident, *Fuzzy Sets and Systems*, 28: 115 -127, 1988.

[8] T. Onisawa, Subjective Analysis of System Reliability and its Analyzer, *Fuzzy Sets and Systems*, 83: 249-269, 1996.

[9] J. D. More, R. Tanscheit, M.M. Vellasco, M.A. Pacheco, D.M. Szwarcman, A Fuzzy Approach to the Study of Human Reliability in the Petroleum Industry. In: O. Castillo; P. Melin; O.M. Ross; R.S. Cruz; W. Pedrycz; J. Kacprzyk. (Org.). *Advances in Soft Computing - Theoretical Advances and Applications of Fuzzy Logic and Soft Computing*, Springer, 42: 415-424, 2007.

[10] J. D. More, *A Fuzzy Approach to the Evaluation of Human Reliability in Ultrasonic Nondestructive Examinations*, Ph.D. Thesis COPPE/UFRJ, Brazil, 2004.

[11] H. M. Hsu, C. T. Chen, Aggregation of Fuzzy Opinions under Group Decision Making, *Fuzzy Sets and Systems* 79: 279-285, 1996.

[12] T. J. Ross, *Fuzzy Logic with Engineering Applications*, 2nd Ed., John Wiley & Sons, 2004.

ISBN: 978-989-95079-6-8

Multivariate Segmentation of Time Series with Differential Evolution

Daniel Graves[1] Witold Pedrycz[2]

[1,2]Department of Electrical & Computer Engineering, University of Alberta, Edmoton, Canada
Email: dgraves@ualberta.ca[1], pedrycz@ee.ualberta.ca[2]

Abstract—*A new method of time series segmentation is developed using differential evolution. Traditional methods of time series segmentation focus on single variable segmentation and as such often determine sections of the time series with constant slope (i.e. linear). The problem of segmenting multivariate time series is significantly more involved since several time series have to be jointly segmented. Thus the concept of boundary becomes ill-defined since each time series may not be exactly synchronized and change identically in time. The problem is rectified by minimizing the mean of the variance of the slopes determined in each segment. Performance of the method is measured in terms of the classification rate and the accuracy of determination of boundaries. Experimental evidence shows the effectiveness of the method when applied to synthetic and real-world data compared with multivariate time series clustering approaches.*

Keywords—Multivariate segmentation, differential evolution, time series, fuzzy clustering.

1 Introduction

Many data come in the form of time series that need to be partitioned into contiguous blocks in time. Tools of Computational Intelligence have been applied to solve the time series partitioning problem are commonly referred to as time series segmentation including fuzzy clustering, hierarchical clustering, and genetic algorithms. Time series segmentation of a single variable has been studied extensively in the literature, c.f. [1,2,3,4,5]. However, there is a lack of research in the area of multivariate time series segmentation or the joint segmentation of several time series.

Only a few multivariate segmentation algorithms found in the literature cluster the time series by taking into account temporal information, e.g., [5,6,7,8]. Unfortunately clustering a time series to perform segmentation poses a few problems. An important issue with segmentation of a time series which has to be addressed is that segments must be contiguous in time. Several approaches have been considered to solve this problem. One method uses constrained clustering to find contiguous segments in time, i.e. constrained hierarchical clustering. Another approach uses time information as an auxiliary component exploited directly in the clustering process. A significant problem is to effectively use this auxiliary information in the clustering procedure leading to boundary identification. However, augmenting clustering with time information is not an appropriate vehicle for segmenting a time series since identification of boundaries requires use of time information as auxiliary information in clustering. A common problem produced by augmented clustering algorithms that account for time information is that the resulting partition does not necessarily consist of contiguous segments. Currently, the literature lacks an effective method of using the auxiliary information in clustering.

The focus of this research is in multivariate time series segmentation by a more direct segmentation method that avoids the problems produced by using clustering as a vehicle for segmentation by employing differential evolutionary optimization of an objective function. As well, evolutionary methods such as differential evolution are well suited for structural optimization. The reason for using differential evolution (DE) to segmentation a time series is that the segmented boundaries to be optimized cannot be *explicitly* expressed in the underlying objective function. This makes gradient-based optimization methods not feasible in this problem. The optimization calls for methods of structural optimization and various techniques of Evolutionary Computing are of interest. In this category of methods, DE has shown to be effective as far as quick convergence is concerned.

We focus on detecting slope changes in the time series in order to determine segments. The motivation behind approaching the problem of segmentation by detecting monotonicity is that often a time series can be represented quite effectively using only a handful of monotonic models for each segment. An example is with a set of ECG time series which are often characterized by sharp spikes in the middle. A potential application of segmentation that detects changes in slope is in signal compression, c.f. [2].

There are a number of other applications of multivariate time series segmentation. An example is presented in [9] for process monitoring, diagnosis and control of a medium and high-density polyethylene plant. Another application of multivariate time series segmentation is with music structural segmentation where the objective is to automatically determine a partition of structural elements in music, i.e. verse, chorus, bridge, by detecting significant transition points in the music.

The material is organized in five sections. A literature review is covered in Section 2. Section 3 includes a description of the segmentation algorithm and presents the essentials of differential evolution. Section 4 concentrates on the experimental evaluation of the multivariate segmentation algorithm. The proposed method is evaluated on synthetic data sets and on several real-world time series data sets coming from Time Series Data Library [10]. Section 5 presents the conclusions.

ISBN: 978-989-95079-6-8

2 Literature Review

Time series segmentation is a very important topic with temporal information processing as long time series need to be broken down into relevant segments. The problem of segmentation has not been solved with high boundary detection accuracy in the literature hence most of research conducted in the literature focuses on the segmentation aspect and leaves the labelling of segments to be done later.

Most time series segmentation approaches look at the problem of segmenting a time series of a single variable. There are three main categories of classical segmentation algorithms available in the literature:

- Bottom-up
- Top-down
- Sliding-window

Unfortunately high segmentation accuracy is not always achieved and there is no one method preferred than the others. The three algorithms are described in [1]. All three have an error threshold where they stop segmenting once the error is exceeded. The first bottom-up starts with each sample in its own segment much like agglomerative hierarchical clustering. Likewise top-down is similar to divisive hierarchical clustering in that the entire time series is broken down into smaller segments until the error for each segment is less than the error threshold. Sliding-window is a real-time segmentation algorithm that starts with the first sample and extends the segment with each new sample until an error threshold is exceeded.

Determining segments via linear approximation and evolutionary methods is common in the literature such as methods using genetic algorithms to approximate an ECG time series trough a set of its linear segments, cf. [2]. In [2], the fitness function minimizes the variability between the maximum and minimum slopes for each consecutive sample within the proposed segment. The number of segments is determined beforehand. The experiments are conducted on ECG signals which show fairly good match on the data, i.e. little information appears to be lost during segmentation (which also leads to dimensionality reduction).

Multivariate time series are common in practical problems. Many of the approaches discussed in the literature are not directly applicable to them. An inherent problem with multivariate time series segmentation is that it difficult to determine boundary points where each feature time series in the multivariate time series unanimously agree. For example, a question arises whether a significant change in one feature's value where all other features do not change significantly constitutes a boundary change point for a time series. Also not all features may change at exactly the same point in time; hence, multivariate segmentation approaches require greater boundary point flexibility to account for the latent changes among the feature values. As a result, some limited work has been done in the literature directly applicable to multivariate time series segmentation, c.f. [5,6,7,9].

The multivariate time series segmentation method developed in [9] approaches the problem as a constrained clustering problem with the constraint that a cluster (i.e. segment) must contain patterns at successive time points. The approach also constructs fuzzy sets for specifying the segments detected with fuzzy boundaries. Local Principal Component Analysis (PCA) models are used in clustering of the time series into segments based on Gath-Geva clustering. They apply the segmentation algorithm to process monitoring, diagnosis and control in particular monitoring a medium and high-density polyethylene plant, i.e. production of a versatile plastic. The data is multivariate and is successfully segmented via the proposed clustering algorithm.

A Gath-Geva-based clustering algorithm similar to the approach in [7,9] is introduced for application to multivariate time series segmentation. The approach described in [9] and [7] utilize a modified Gath-Geva clustering based on the probabilistic PCA (PPCA) [11] for each cluster. The method in [9] and [7] constrain the segment (cluster) fuzzy sets (membership matrix) to Gaussian membership functions. The method in [7] uses modified Gath-Geva clustering (a.k.a. fuzzy maximum likelihood clustering) to cluster a multivariate time series. The time information is included as an attribute in the time series. The segments are projected onto principal components based on a localized PPCA models for the rationale that the correlation between variables often changes with multivariate time series. The distance measure in the clustering is chosen to include two parts: Gaussian fuzzy sets determining distance in the time domain and a second term that measures the distance of the data from the PCA hyperplane model. The number of principal components to keep in the PPCA model and the number of segments (clusters) are two important parameters that need to be selected via some other method. The authors include some case studies in the literature for the application of the approach to multivariate time series segmentation although the experimental evidence is limited.

Speaker segmentation is discussed in [5]. The method exploits the BIC (Bayesian Information Criterion) procedure, known also as the minimum description length (MDL). The objective is to determine segments of different speakers. Some improvements are found in their approach over traditional BIC approaches. Multivariate segmentation in [6] is accomplished via constrained clustering for the purpose to determine the important structural components of a musical recording, i.e. intro, verse, chorus, solo, bridge, outro. They report some improvements in the classification rate and boundary f-value with their method over traditional approaches. A multivariate time series segmentation algorithm based on hierarchical Bayesian clustering was developed in [12] specifically designed for astronomical data that can be modelled as a Poisson counting process. Some experiments on synthetic data and real astronomical data are provided however they are limited. Authors in [8] describe multivariate segmentation approach for motion data using singular value decomposition (SVD) in the distance measure between multivariate motion time series. The experimental results are very limited.

ISBN: 978-989-95079-6-8

3 Multivariate Segmentation

3.1 *Problem Formulation*

The problem of multivariate segmentation involves finding the boundaries in a time series that denote changing behaviour in the time series. For instance a time series might be increase steadily and then at a time t_o the time series changes behaviour and starts decreasing steadily. Extending this to the multivariate case, each variable which is a time series in itself, exhibits changes in behaviour after the boundary at t_o. With single-variable time series segmentation, segments are contiguous homogenous partition series that are specified by a set of boundaries that divides segments from their adjacent ones. Multivariate segmentation is the problem of finding a set of boundaries that jointly divides each time series into a set of contiguous partitions of multivariate time series. With multivariate time series, segmentation is more complicated since each feature can be considered a separate time series and hence what might be considered a boundary in one may not be considered a boundary in the other. i.e., the transitions in each feature signal may start and end at different points in time providing conflicting clues as to where the segment starts and ends.

Given a multivariate time series $x_{t,k}$ for t=1…T with M features k=1…M, the time series can be segmented through slope analysis of each feature. The idea is to find segments with similar slopes, i.e. finding segments that are monotonic with relatively constant slopes. Assuming the number of segments c is known a priori, the boundaries for the segments are the free-parameters given by $[b_1, b_2, \dots, b_{c-1}]$ where $b_0 = 1$ and $b_c = T$. An example of a multivariate segmentation is given in figure 1.

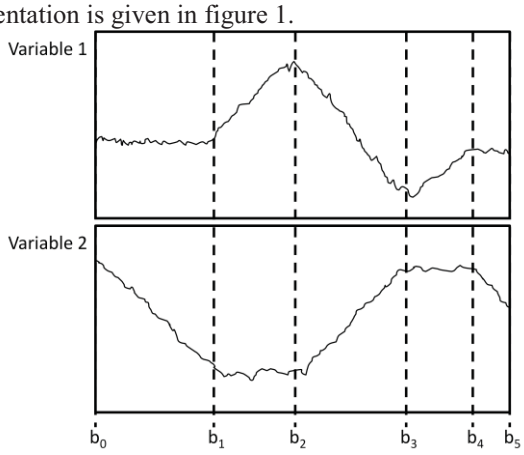


Figure 1: Multivariate segmentation illustration

3.2 *Differential Evolution*

Differential evolution is simple population-based evolutionary optimization technique that minimizes (or maximizes) an objective function much like genetic algorithms, c.f. [13]. Differential evolution uses mutation and crossover operations just as genetic algorithms; however, the mutation operation is based on population vector differences. The population of vectors are given by the set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_P\}$ where each vector $\mathbf{x}_i$ is real-valued and has dimensionality M. The objective function to be minimized is denoted by f(**x**). The vectors in the population assume values in the unit interval and P is the size of the population. The range of values for the population vectors is on the interval [0,1].

The essence of the differential evolution is expressed via the following pseudo code

1. Initialize Population of size P
2. Loop until stopping criterion met
    1. For each element $\mathbf{x}_i$ in the population
        1. Randomly select three parameter vectors: $\mathbf{x}_{r1}$, $\mathbf{x}_{r2}$, $\mathbf{x}_{r3}$ (not the same)
        2. Generate a mutant vector $\mathbf{v}_i=\mathbf{x}_{r1}+G[\mathbf{x}_{r2}-\mathbf{x}_{r3}]$ *[Mutation]*
        3. Generate trial vector $\mathbf{u}_i=\text{crossover}(\mathbf{v}_i,\mathbf{x}_i)$ by mixing vectors $\mathbf{v}_i$ and $\mathbf{x}_i$ *[Crossover]*
        4. If $f(\mathbf{u}_i) < f(\mathbf{x}_i)$ then replace $\mathbf{x}_i$ with $\mathbf{u}_i$ in population
3. Termination

The difference factor G controls the intensity of mutation (i.e. vector differences). The difference factor G can take on positive real numbers as well as zero, c.f. [4]. There is also a crossover operation. Many different kinds of crossover operations are possible. Here we use the one presented in [4]. This crossover operation combines two vectors $\mathbf{x}_i$ and $\mathbf{y}_i$ according to the pre-specified probability p_{cr}. The resulting crossover vector produced by crossover is give by the expression

$$u_{i,k} = \begin{cases} v_{i,k} & if\ r_k < p_{cr} \\ x_{i,k} & otherwise \end{cases} \quad (1)$$

where r_k is a uniform random sequence of numbers on the interval [0,1] for k=1 … M.

3.3 *Fitness Function*

The boundaries of the segments are determined via differential evolution. Here we elaborate on the pertinent algorithmic details.

A difference (describing slope) sequence is calculated for each variable k=1…M in the multivariate time series

$$D_{tk} = x_{t+1,k} - x_{t,k} \quad (2)$$

for t=1…T-1.

Given a set of boundaries $[b_1, b_2, \dots, b_{c-1}]$ where $b_0 = 1$ and $b_c = T$, the length of each segment is given as

$$L_i = b_i - b_{i-1} \quad (3)$$

The variance of the slopes of each variable are calculated via the expression

$$S_{ik}^2 = \sum_{t=b_{i-1}}^{b_i} \frac{(D_{tk} - \overline{D}_{ik})^2}{L_i - 1} \quad (4)$$

where $\overline{D}_{ik} = 1/L_i \sum_{t=b_{i-1}}^{b_i} D_{tk}$ is the mean slope of each variable k=1…M for each segment i=1...c. The mean of the standard deviations across the variables is calculated for each segment, i.e.

$$S_i = \frac{1}{M} \sum_{k=1}^{M} \sqrt{S_{ik}^2} \quad (5)$$

for i=1…c.

ISBN: 978-989-95079-6-8

Finally multivariate segmentation is accomplished by minimizing the following fitness function

$$F = \sum_{i=1}^{c} S_i L_i \tag{6}$$

where S_i is the standard deviation defined in (4) and L_i is the segment length defined in (2). The resulting segmentation favors segments with slopes that do not vary greatly; hence, the resulting segments are approximately linear.

3.4 *Population Encoding*

Each population vector is encoded in differential evolution as a vector of real-valued numbers denoted by $\mathbf{y} = [y_1, y_2, \ldots y_{c+1}]$. The set of boundaries $\mathbf{b} = [1, b_1, b_2, \ldots, b_{c-1}, T]$ is calculated from the vector $\mathbf{y}$ where the first and last elements ($b_0 = 1$ and $b_c = T$) can be discarded. The boundaries are calculated in the form:

1. $\mathbf{z} = sort - least - to - greatest(\mathbf{y})$
2. $\mathbf{a} = \frac{\mathbf{z} - z_1}{z_{c+1}}$
3. $\mathbf{b} = round(1 + (T-1)\mathbf{a})$

The elements of vector $\mathbf{y}$ is sorted from least to greatest and its sorted version is denoted as a new vector $\mathbf{z}$ so that the boundaries are given in order. The vector $\mathbf{z}$ is normalized to a vector of numbers on the range [0,1] called $\mathbf{a}$ by subtracting the smallest element in $\mathbf{z}$ given by z_1 since the vector is sorted and divided by the largest element z_{c+1}. The vector $\mathbf{a}$ maps directly to boundaries in the time series where 0 is the start of the time series and 1 is the end of the time series. The vector $\mathbf{b}$ is calculated by expanding the range of $\mathbf{a}$ to the set of integer time coordinates $\{1, T\}$. The reason for sorting the vector is that taking vector differences (the mutation operation in differential evolution) can result in negative values. Simply normalizing the vector by its sum as done in [2] would produce erroneous results since vector elements are not necessarily positive. Elements of $\mathbf{y}$ needed to be mapped to the interval [0,1] where the smallest value of $\mathbf{y}$ is mapped to the start of the time series.

4 Experiments

The objective function described in section 3.2 minimizes slope variance for each segment. The optimization procedure is unsupervised, i.e. does not use pre-labelled data during optimization. The following synthetic experiments are evaluated against an *a priori* segmentation provided by a human expert. The classification rate and boundary f-value criteria are used to evaluate the synthetic experiments. The classification rate is the percentage of correctly classified samples according to a pre-segmented time series (human labelled) and is indicative of the precision. The boundary f-value is a measure of the number of correct boundaries in the resulting segmentation. The boundaries discovered by the algorithm are denoted by vector $\mathbf{b}$ and the human labelled boundaries are given by the vector $\mathbf{h}$ where $\mathbf{b}$ and $\mathbf{h}$ have length denoted by $L_{\mathbf{b}}$ and $L_{\mathbf{h}}$ respectively. The boundary f-value is determined according to the following expression

$$f = \frac{2pr}{p + r} \tag{7}$$

where p is the boundary precision and r is the boundary recall calculated as follows

$$n = \sum_{i=1}^{L_{\mathbf{h}}} \begin{cases} 1 & if\ \theta > \left| \mathbf{h} - \min_{\forall j = 1 \ldots L_{\mathbf{b}}} \{|h_i - b_j|\} \right| \\ 0 & otherwise \end{cases} \tag{8}$$

$$p = \frac{n}{L_{\mathbf{b}}} \tag{9}$$

$$r = \frac{n}{L_{\mathbf{h}}} \tag{10}$$

where n is the number of correctly classified boundaries with respect to the provided human labelled boundaries and θ is a threshold value which determines the tolerance or region of acceptable difference between the human labelled boundary and machine labelled boundary. The following figure demonstrates the calculation of the f-value.

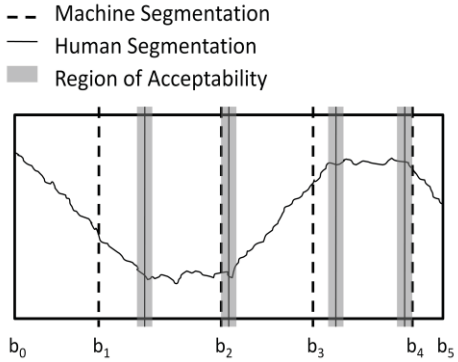


Figure 2: Multivariate segmentation illustration

The reconstruction error is the performance index for the real-world data sets since no human labels are present, i.e.

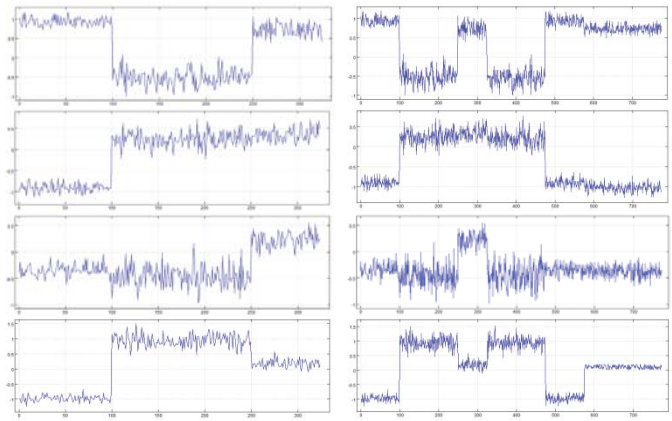
$$E = \frac{1}{MT} \sum_{k=1}^{M} \sum_{t=1}^{T} \left(\hat{x}_{t,k} - x_{t,k} \right)^2 \tag{11}$$

where $\hat{x}_{t,k}$ is the reconstructed time series. The time series $\hat{x}_{t,k}$ is reconstructed by approximating each segment via a linear line of the form $\hat{x}_{t,k} = m_k t + b_k$ from the first point in the segment to the last point in the segment for each variable k=1...M. The values m_k and b_k are the slope and intercept, respectively, of the line for each variable k.

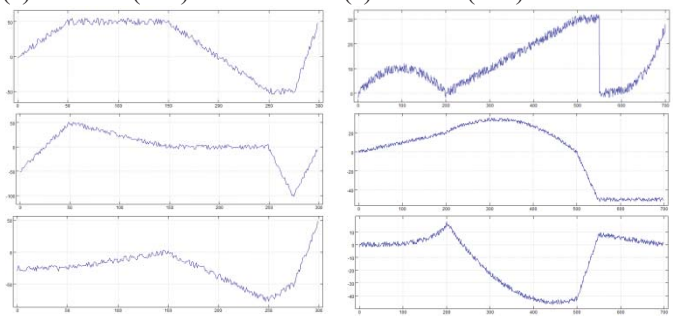
Four synthetic time series were segmented using the described multivariate time series segmentation algorithm. Each variable for time series DATA1, DATA2, DATA3, and DATA4 are given in the figure below where DATA1 and DATA3 have four feature time series and DATA2 and DATA4 have three. The length of each time series is 325, 300, 775, and 700 samples respectively. These time series were generated by creating a collection of linear segments and then adding white noise.

The experimental results on the synthetic data are shown in table 1. They are compared with FCM-DFS (Fuzzy C-Means clustering with Distinct Feature Sets) time series clustering described in [14] for segmentation. The value of theta was chosen to be relatively small, i.e. 3, since the synthetic data sets are relatively simple often with an underlying linear component mixed with white noise. Also, it was noted that generally the segmented boundaries were either very close to the human defined boundary (within 3 samples) or the boundary was not properly discovered by the segmentation algorithm.

ISBN: 978-989-95079-6-8



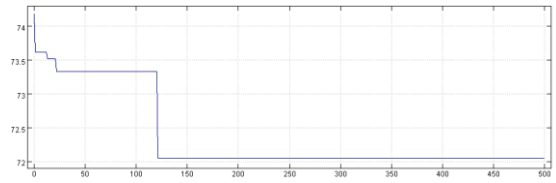
(a) DATA1 (c=3) (c)DATA3 (c=6)



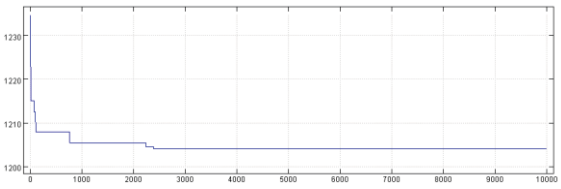
(b) DATA2 (c=5) (d) DATA4 (c=4)

Figure 3: Synthetic Data

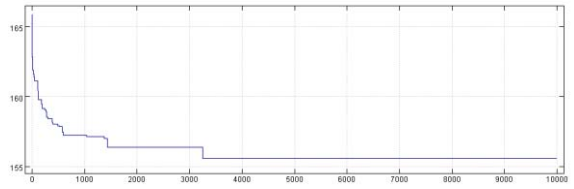
The following figure shows the convergence of differential evolution.



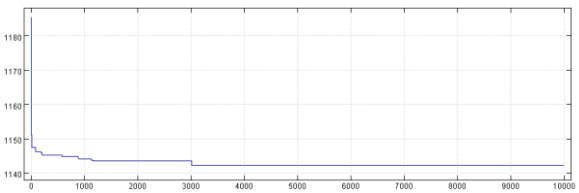
(a) DATA1



(b) DATA2



(c) DATA3



(d) DATA4

Figure 4: Fitness function in successive generations of differential evolution

Table 1: Synthetic experimental results

	P	G	p_{cr}	Num Iter.	F-value ($\theta = 3$)	Classification Rate
DATA1	100	1	0.25	500	100%	99.7%
DATA2	100	0.1	0.35	10000	50%	67.3%
DATA3	100	0.1	0.3	10000	100%	99.6%
DATA4	100	0.1	0.35	10000	66.7%	57.3%

The differential evolution parameters were fairly robust meaning that changes in the crossover rate and differential gain did not change performance dramatically. However, as the time series became more complex, i.e. DATA3, the differential evolution parameters had to be tuned to prevent the method from being trapped in local minima. From the practical point of view, the essential differential evolution parameters are differential gain, and crossover probably. It was found that increasing population size beyond 100 did not greatly improve the final boundaries. Initially the crossover probably was set to 0.5 and differential gain was set to 1. It was found through parameter adjustments that performance improved with slightly smaller crossover probability. A differential gain of 1 was sufficient on simple data sets; however, more complicated data sets such as DATA2, DATA3, and DATA4 required a smaller value for differential gain (i.e. less mutation) in order to obtain better classification rates and boundary f-value.

Table 2: Synthetic experimental results with FCM-DFS

	m	α	F-Value ($\theta = 3$)	Classification Rate
DATA1	2	0.5	66.7%	99.7%
DATA2	2	0.1	25.0%	80.3%
DATA3	2	0.4	71.4%	99.7%
DATA4	1.3	0.5	16.1%	81.2%

The accuracy of the boundaries discovered by FCM-DFS is not as high when compared with differential evolutionary multivariate segmentation on these synthetic data sets. The classification rate is higher for FCM-DFS but since the boundary f-value is much higher for the proposed method, the segmentation with DE is much more accurate with the determined boundary locations. The reason for the smaller classification rate with DE is that in some of the data sets, only one boundary is be determined correctly and its location is way-off the expected value. With FCM_DFS, several boundaries are not determined very well (accuracy is very low) but the locations are a bit closer to their expected values when a boundary is missed. The experiments on FCM-DFS were run for several different values of m {1.3, 1.5, 1.7, 2, 2.3, 2.5, 3} and several values of α {0.05, 0.1, 0.2, 0.4, 0.5, 0.6, 0.8, 0.9, 0.95}.

The segmentation approach was applied to several real valued time series [10]. The performance of the segmentation is evaluated against reconstruction error (11) as shown in Table 3.

The number of clusters was based on a visual estimate of the number of segments from the plot and then running several different values of c in close proximity to the estimated number of clusters. The results show that the segmentation algorithm does fairly well on some of the data sets with small reconstruction errors especially on power station and hog data sets.

ISBN: 978-989-95079-6-8

Table 3: Real-world time series experimental results

	P	G	p_{cr}	Num Iter.	c	Reconstruction Error
Lynx Pelts	100	0.1	0.9	4000	13	0.1253
Power Station	100	0.5	0.3	4000	12	0.0917
Hog	100	0.1	0.9	4000	16	0.0940
Housing	100	0.5	0.3	4000	17	0.1386
Housing Starts	100	2	0.3	4000	23	0.2185

5 Conclusions

The results on synthetic data show that multivariate time series segmentation show promise as a multivariate time series segmentation tool. The results are quite good on noisy synthetic data with boundary f-values reaching upto 100% and classification rates around 99%. Parameter tuning was important especially as the time series become more complicated and required smaller differential gain values, smaller crossover probability values and more iterations to achieve best results.

Future directions in research involve applying the multivariate segmentation algorithm to the problem of musical segmentation – a particularly difficult problem that automatically finds a partition in music that corresponds to the structural labels of a song, i.e. intro, verse, chorus, bridge, outro. Replacing the linear segment models with more flexible autoregressive models will be developed in future segmentation algorithms.

Acknowledgment

Support from the NSERC and Canada Research Chair (CRC) Program (W. Pedrycz) is gratefully acknowledged. This research is supported by a post-graduate PhD iCore fellowship (D. Graves).

References

[1] E. Keogh, S. Kasetty, On the need for time series data mining benchmark: a survey and empirical demonstration, *Proc. of the 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Edmonton, Alberta 102-111, 2002.

[2] A. Gacek, W. Pedrycz, A genetic segmentation of ECG signals, *IEEE Trans. On Biomedical Engineering*, 50(10): 1203-1208, 2003.

[3] F.L. Chung, T.C. Fu, V. Ng, R.W.P. Luk, An evolutionary approach to pattern-based time series segmentation, *IEEE Trans on Evolutionary Computation*, 8(5): 471-489, 2004.

[4] J.A. Crossman, H. Guo, Y.L. Murphey, J. Cardillo, Automotive signal fault diagnostics – part I: signal fault analysis, signal segmentation, feature extraction and quasi-optimal feature selection, *IEEE Trans on Vehicular Technology*, 52(4): 1063-1075, 2003.

[5] P. Delacourt, C.J. Wellekens, DISTBIC: a speaker-based segmentation for audio data indexing, *Speech communications*, 32: 111-126, 2000

[6] M. Levy, M. Sandler, Structural segmentation of music audio by constrained clustering. *IEEE Trans. on Audio, Speech, and Language Processing*, 16(2): 318-326, 2008.

[7] J. Abonyi, B. Feil, S. Nemeth, P. Arva, Modified Gath-Geva clustering for fuzzy segmentation of multivariate time-series, *Fuzzy Sets and Systems*, 149(1): 39-56, 2005.

[8] C. Li, Z. Peng, S.Q. Zheng, B. Prabhakaran, Segmentation and recognition of multi-attribute motion sequences, *Proc ACM Int. Conf. Multimedia*, New York, 836-843, Oct 2004.

[9] J. Abonyi, B. Feil, S. Nemeth, P. Arva, Fuzzy clustering based segmentation of time-series, *Proc. Intelligent Data Analysis*, 2810: 275-285, 2003.

[10] R.J. Hyndman, *Time Series Data Library*, http://www.robhyndman.info/TSDL (accessed on Jan 29, 2009).

[11] M.E. Tipping, M. Bishop, Mixtures of probabilistic principal component analyzers, *Neural Computation*, 11(2): 443-482, 1999.

[12] N. Dobigeon, J.Y. Tourneret, J.D. Scargle, Joint segmentation of multivariate astronomical time series Bayesian sampling with a hierarchical model, *IEEE Transactions on Signal Processing*, 55(2): 414-423, 2007.

[13] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization*, 11: 341-359, 1997.

[14] D. Graves, W. Pedrycz, Structural segmentation of music with fuzzy clustering, *In Proc. Acoustics Week in Canada*, Vancouver, BC, Canadian Acoustics, Oct 2008.

ISBN: 978-989-95079-6-8

Coevolutionary Genetic Fuzzy System to Assess Multiagent Bidding Strategies in Electricity Markets

Igor Walter[1] Fernando Gomide[2]
1.Brazilian Electricity Regulatory Agency – ANEEL
Brasília - DF, Brazil
2.Faculty of Electrical and Computer Engineering – FEEC
University of Campinas – Unicamp
Campinas - SP, Brazil
Email: igorwalter@aneel.gov.br, gomide@dca.fee.unicamp.br

Abstract— *In this paper we study a genetic fuzzy system approach to assess suitable bidding strategies for agents in online auction environments. Assessing efficient bidding strategies allows evaluation of auction models and verification whether the mechanism design achieves its goals. Day-ahead electricity auctions are particularly explored to give an experimental instance of the approach developed in this paper. Previous works have reported successful fuzzy bidding strategies developed by genetic fuzzy systems and coevolutionary algorithms. Here we review the coevolutionary algorithm and present recent results of the bidding strategies behavior. We analyze how the evolutionary strategies perform against each other in dynamic environments. Coevolutionary approaches in which coevolutionary agents interact through their fuzzy bidding strategies permit realistic and transparent representations of the behavior of the agents in auction-based markets. They also improve market representation and evaluation mechanisms. Experimental results show that coevolutionary agents can enhance their profits at the cost of increasing system hourly price paid by demand, an undesirable outcome from the perspective of the buyers.*

Keywords— Genetic fuzzy systems, electricity markets, auctions, multiagent systems, computational economics.

1 Introduction

The restructuring process of electricity markets has raised new challenges and opportunities, because currently there is no consensual market architecture. The progress of the power industry has mirrored the lack of insight about the mechanism design issues to be addressed. An electricity market is inherently a complex system populated by self interested, interacting economic agents. Tools rooted in the neoclassical economic theory have shown to be of limited value to study the behavior of economic agents in electricity markets. The approach considered in this paper is derived from agent-based computational economics (ACE) [15] and computational intelligence tools. The purpose is to find bidding strategies for a set of agents playing in a market using genetic fuzzy systems principles. The approach can be extended to other negotiation environments and contexts.

Evolutionary bidding strategies support decisions of market players and may uncover unknown and unexpected agents behaviors that help designers to simulate and analyze negotiation mechanisms. When coevolutionary agents extract benefits of strategic behavior, the mechanism designer can review the decision making process to correct eventual unfavorable decisions and fix mechanism flaws.

Recently a coevolutionary approach was suggested to assess bidding strategies for multiagent systems. The focus was on strategies encoded in fuzzy rule-based systems [18]. The aim was to learn models represented by coevolving knowledge bases and improve the performance of the agents when acting in competitive environments. In competitive environments data for learning and tuning of knowledge bases are rare and rule bases must evolve jointly with databases. An evolutionary algorithm whose operators use a variable length chromosome, a hierarchical relationship among individuals through fitness, and a scheme that successively explores and exploits the search space along generations has also been developed [17]. In this paper we first review the coevolutionary algorithm of [18], summarize recent results of coevolutionary bidding strategies, and analyze how evolutionary agent strategies react against each other in dynamic environments.

Electricity auctions are the main application example addressed in this paper due to its practical relevance and the experimental instance they provide. The example illustrates the design of bidding strategies to exploit the negotiation space and to take advantage of market power aiming at improved negotiation mechanisms.

This paper is organized as follows. After this introduction, a brief overview of strategic bidding and survey of related work are presented in section 2. Next, section 3 summarizes the approach addressed in this paper. Section 4 presents and discusses the experimental results. Section 5 concludes the paper and suggests issues for further research.

2 Evolutionary Techniques for Electricity Market Bidding Strategies

Currently, many power industries worldwide use auctions as a mechanism of resource allocation and system coordination. The design of auction mechanisms can give agents the ability to explore market imperfections for gaming.

A energy supplier (generator) competing in an electricity market has to decide upon how much energy to offer and at which price. In a perfectly competitive market, risk averse agents have an incentive to offer energy at a price equivalent to their marginal costs [7]. Electricity markets are, however, much more an oligopoly than a laissez-faire, with low or no demand elasticity in the short term, barriers to entry, and physical constraints. Thus, a energy supplier may have an incentive to offer energy at a price other than its marginal costs and to extract some surplus from such an imperfect market.

ISBN: 978-989-95079-6-8

The behavior of a generator bidding other than marginal costs in an effort to exploit market imperfections is defined as strategic bidding [6]. The most common strategy is to maximize the expected profit but other strategies may interest a supplier exercising its market power such as competing for being a base-load generator, increasing market share, increasing profit margin, etc.

2.1 Related Work

Evolutionary approaches for strategic bidding in electricity markets have been addressed in the literature in the realm of genetic algorithms, evolutionary strategies, and classifier systems. Multiagent system approaches have been proposed to model the electricity market as well, among them [8, 14].

The evolutionary strategy approach suggested in [19] evolves a single bid value, assumed to be valid for all hours of a day ahead market. In this sense, it does not give a bidding strategy, but a bid value that is narrowly valid for a specific situation.

The approach proposed in [1] is a classifier system capable of dealing with the grid constraints of the UK market and electrical restrictions that constraint generators to be *on* or *off* by the dispatch.

In [3] a coevolutionary approach based on a coevolutionary cooperative genetic algorithm (CCGA) is devised to analyze electricity market equilibrium. Results for a 3-firms market is presented. They show that the approach finds Cournot-Nash equilibrium and Pareto solutions. In a related work [4] the same coevolutionary CCGA scheme is used together with a Cournot model and a supply function equilibrium (SFE) model of a 5-firms experiment. In this case, the coevolutionary scheme converges to a Nash equilibria.

The coevolutionary approach to model the electricity market can be based on different coevolutionary methods such as the one addressed in [2] for duopolies. In this method, discretization of the state space imposes limitations.

The coevolutionary process proposed in [11] is more complex[1]. The intent is to coevolve agents strategies and market mechanisms. The authors use a genetic programming (GP) method to evolve a function the auctioneer uses to set the price between the ask and the bid price in a double auction framework. GP is used to represent the auctioneer and traders (buyers and sellers). Simulation results based on the Nicolaisen's experiment report a function with several terms after 10,000 generations. The authors claim that the function evolved was approximately equivalent to a discriminatory price k-double auction with $k = 0.5$.

3 Coevolutionary Fuzzy Bidding Strategies

The approach employed in this paper is to coevolve fuzzy bidding strategies. Previous work [17] has devised evolutionary fuzzy bidding strategies. In [18] a coevolutionary algorithm was introduced to study how the evolving strategies react against each other in dynamic bidding environment[2]. Contrary to alternative evolutionary approaches, the aim was to model bidding information and strategies within the framework of fuzzy set theory. By enabling a fuzzy system to learn trough an evolutionary algorithm, one expects to find effective and transparent bidding strategies. We assume that bidding strategies are encoded by fuzzy rule-based systems (FRBS). As opposed to trial and error methods, genetic algorithms are especially attractive to develop bidding strategies because they can optimize the knowledge base (dimension, membership functions and rules) for a given market configuration. Genetic algorithms are also more appropriate than conventional optimization techniques because rule base tuning may involve operators and high dimensionality search spaces.

Here a GFRBS (Genetic Fuzzy Rule-Based System[3]) scheme is adopted to simultaneously evolve the data base (granularity and membership functions) and the rule base of a FRBS with the aim to find the most profitable bidding strategy. Each population of the GFRBS is a coevolving specie that represents a market agent. The result is a set of fuzzy rule-based systems able to handle imprecise data typically found in auction environments.

The evolutionary approach detailed in [17] is used to evolve the data base employing a variable length chromosome to represent the rule base. In the rule base, both, the number of rules and size of each rule may change during the evolutionary process. A particularly suitable crossover operation detailed in [17] was developed to enhance system performance.

Many coevolutionary approaches emphasize explicit competition such as the host-parasite scheme to minimize sorting networks in [10]. The coevolutionary approach focused in this paper is based on the CCGA of Potter and De Jong [13]. As detailed below in Algorithm 1, our approach selects a representative agent for each coevolutionary specie. The species model bidding agents. The representatives are the ones who compete in the auctions. When representatives bid, their strategy may improve the profit of the remaining bidders. Thus, the remaining coevolutionary species may act in a tacit cooperation in this competitive game. Therefore, although CCGA was originally developed to coevolve subcomponents of a problem in an explicit cooperative framework, we use its main idea in a competitive environment: auction in electricity markets[4].

3.1 Genetic Algorithm

The genetic algorithm employed in this paper uses a variable length chromosome representation for both: data base and rule bases. The encoding schemes are similar to the ones adopted in [17]. Crossover operations in data base and rule base are synchronous, that is, crossover points of database and rule base induce exchange of chromosome portions to keep the rule structure and fuzzy terms meaningful in their offspring.

The first step to design a GFRBS must decide which part of the FRBS will be optimized by the genetic algorithm [5]. This decision usually means a trade off between granularity and search efficiency. In the approach adopted here, the following FRBS components are evolved: database (granularity and membership functions parameters) and rule base (number of "active" rules[5] and rule structure).

[1] A more complete report of this work can be found in [12].

[2] This and all previous papers are a result of the PhD thesis of the first author and has no links with the Regulatory Agency.

[3] For a taxonomy, a comprehensive survey of past and recent developments, as well as future trends of genetic fuzzy systems see [9].

[4] A similar approach was adopted in [3].

[5] An "active" rule is a rule that is processed during fuzzy inference while an "inactive" rule remains in the rule base genotype but it is not used during fuzzy inference.

ISBN: 978-989-95079-6-8

The GA also assumes the simultaneous evolution of the FRBS in the sense that each individual represents a complete fuzzy rule-based system as in the Pittsburgh approach. The FRBS is encoded in a chromosome Cr composed by three major components (Cr_1, Cr_2, and Cr_3) to represent the data and rule base.

More specifically, the chromosome component Cr_1 encodes granularity in a variable length chain of integers, and Cr_2 encodes membership functions parameters in a variable length chain of real numbers. The number of rules, rules size and the rules themselves define chromosome component Cr_3. Component Cr_3 has, in addition to two integers (number of rules and rules size), several variable chains of alleles of '0' or '1', where a '1' means that a linguistic term is used in the rule, otherwise it is '0'.

Algorithm 1 Coevolutionary GFRBS Algorithm

```
t ← 0
for each specie s do
   create random initial population Ω_0^s
end for
repeat
   for each specie s do
      for each specie r such that r ≠ s do
         choose a representative in Ω_t^r
      end for
      evaluate each individual in Ω_t^s through fitness function
      with the chosen representatives
      select parents in Ω_t^s based on relative fitness in Ω_t^s
      apply crossover and mutation on parents to produce
      offspring Ω_{t+1}^s
      t ← t + 1
   end for
until (stop criteria is satisfied)
```

As stated before, the coevolutionary algorithm detailed in Algorithm 1 is based on CCGA [13]. The specifics of the genetic operators are described next.

3.2 Genetic Operators

Due to the particular chromosome encoding structure used to represent the knowledge base, the nature of the genetic operators becomes an important issue to effectively evolve the FRBS. Since there are strong relationships among chromosome components, we need operators that work cooperatively on the chromosome.

Selection is performed using *roulette wheel*. We adopt an elitist model, but the best individual is not guaranteed to be selected for crossover, albeit kept intact in the offspring.

Crossover of the database follows the approach detailed in [17]. Summing up, two different crossover operators are used depending on whether the selected individuals have the same granularity or not. When the granularity is the same, a promising zone in the search space is found and must be appropriately exploited. In this case, the granularity of the offspring database (Cr_1) is kept the same and the membership functions parameters (Cr_2) combined following the max-min-arithmetic algorithm of [17]. When the selected pair has different granularity, a random crossover position p is chosen. Both, granularity (Cr_1) and the corresponding parameters of the membership functions (Cr_2) are recombined. Crossover operations of database and rule base are synchronous: crossover points of database and rule base induce exchange of chromosome portions that keep the rule structure and terms meaningful in their offspring.

Different operators can be used to mutate chromosomes, similarly as reported in [17]: a local variation is introduced to the granularity by adding or subtracting one with equal probability, mutation of membership function parameters uses the Michalewicz non-uniform mutation operator, and rule bases are mutated via the standard, bitwise reversing operation.

3.3 Electricity Market

The application example focused in this paper is electricity markets. The approach can, however, be extended to distinct negotiation applications. The coevolutionary GFRBS agents are thermal power suppliers that have to decide how to bid in auctions. The coevolutionary GFRBS species were evolved using a day-ahead electricity auction. We assume that the remaining agents are non evolutionary and conservative competitors. Conservative agents offer energy bidding all their capacity at their marginal costs [7]. The negotiation protocol is an uniform price sealed bid auction: the price paid for the energy is the same for all accepted bids and is equal to the last bid accepted. The auctioneer decides the hourly dispatch choosing the electricity generator agents outputs and the corresponding price to minimize overall energy cost. Cost minimization results in a merit order dispatch procedure, i.e., bids at lower prices come first. The performance of the best FRBS obtained were verified using test demand data for two weeks day-ahead auctions against conservative and evolutionary competitors. The results are presented and discussed in section 4.

The experiments use actual public data made available from the Brazilian Independent System Operator (ISO)[6]. Demand is supposed to be known to all the participant agents. The data used to evolve the GFRBS is the load of the week beginning on May 19th, 2002, for a portion of the national grid: the South sub-market. The fitness of the coevolutionary agents are the profit on the electricity auction of the decoded GFRBS for the training week. The best individual found by training is tested using the next two remaining weeks period that begins on May 26th 2002 [7].

3.4 Cost Function

A pool of power generator agents was setup based on actual, ISO's publicly available data. Electrical constraints and geoelectrical location are neglected in this work: all the plants are supposed to be in the same sub-market. There is some excess of supply and the running costs of producing electricity in coal, gas and oil plants have been modeled as quadratic functions of the power P_s supplied by the agent, expression (1). Nuclear plants are assumed to have linear cost functions.

$$F(P_s) = a + bP_s + cP_s^2 \quad \text{[GJ/h]} \tag{1}$$

The supplier cost function $C^j(.)$ is given by $F(P_s)$ multiplied by the fuel cost in \$/GJ. Hence costs are quadratic func-

[6] ONS: Operador Nacional do Sistema.

[7] These same data set was used in [17, 18]. Data set contains 504 samples; 168 used for training and 336 for testing.

ISBN: 978-989-95079-6-8

Table 1: Thermal power generator agents characteristics and cost functions.

Plant	Type	Capacity (MW)	Marginal cost ($/MWh)	$C^j(.)$
Ibirité	Gas	766.5	39.065	$3,632.08 + 31.966g + 0.00463g^2$
TermoRio	Gas	824.7	39.109	$3,904.05 + 31.912g + 0.00436g^2$
Argentina I	Gas	1,018	41.045	$4,459.61 + 32.775g + 0.00406g^2$
Argentina II	Gas	1,000	41.046	$4,379.82 + 32.774g + 0.00414g^2$

tions given by (2), where g_h^j is the amount supplied by agent j at hour h.

$$C^j(g_h^j) = \alpha + \beta g_h^j + \gamma {g_h^j}^2 \quad [\$/h] \tag{2}$$

3.5 Competitive Environment

We assume power demand D_h inelastic with price. Therefore, the auctioneer must assure, as commonly found in a single buyer auction, that for each hour h: $\sum_{j=1}^{T_h} g_h^j = D_h$, where g_h^j is the power supplied by agent j and T_h is the number of suppliers. Thus, allocation would cost the market $D_h \pi_h(D_h)$. The supplier agents profit is given by(3):

$$P_h^j = \pi_h g_h^j - C^j(g_h^j) \quad [\$/h] \tag{3}$$

Supplier agents (generators) must internalize all costs to a simple bid, a pair (q_h^j, p_h^j) of the amount offered (in MW) and its price ($/MWh), where the amount q_h^j is less than or equal to the agent capacity, G^j. Conservative agents bid a pair $(G^j, MC^j(G^j))$ where $MC^j(G^j)$ is the marginal cost at capacity G^j.

Table 1 shows the thermal generator agents characteristics for the agents we have chosen to coevolve. The complete data set of the thermal generation park can be found in [17]. The fourth column, the marginal cost ($/MWh), is the marginal cost at full capacity corresponding to conservative bids. The last column gives the cost functions of the thermal suppliers[8].

3.6 Fitness

Algorithm 2 Evaluation of Population Individuals

for each individual i such as $i \in \Omega_t^s$ **do**
  decode database and rule base of i as agent A_i
  add the agent A_i to the market
  simulate market (run auction) for the training period
  keep fitness F_{A_i} as the profit for the period
  remove A_i from the market
end for

The fitness of a GFRBS individual is defined as the profit of the corresponding agent during the training week. Algorithm 2 summarizes the evaluation of the fitness of each individual of the population.

4 Results

The experiments reported in this paper use a population of 40 individuals. Each individual encodes a single input-single output (SISO) FRBS: the input is the hourly load and the output the corresponding the bid price. The genetic operators described in section 3.2 were employed in all the experiments[9] with probability of crossover 0.5 and probability of mutation 0.01. We let the evolutionary process run for 500 or 1,000 generations[10]. While several experiments were performed with 2 to 5 species, some of them did not achieve stable behavior. For instance, for 5 coevolutionary species stability was not attained after 1,000 generations. Section 4.2 discusses briefly why stable behavior may not be necessarily expected, as the 5 coevolutionary species case shows.

Table 2: Comparison between coevolutionary and conservative strategies.

conservative strategy			
agent	hours	energy (MWh)	profit ($)
Ibirité	332	232,954	849,960
TermoRio	277	199,390	959,133
Argentina I	202	179,595	1,065,262
Argentina II	149	116,381	1,049,804
coevolutionary strategy: testing			
agent	hours	energy (MWh)	profit ($)
Ibirité	334	255,856	2,082,681
TermoRio	336	248,367	2,237,294
Argentina I	150	89,969	1,825,135
Argentina II	335	301,710	2,304,976
variation: coevolutionary / conservative			
agent	hours	energy (MWh)	profit ($)
Ibirité	+0.60%	+9.83%	+145.03%
TermoRio	+21.30%	+24.56%	+133.26%
Argentina I	-25.74%	-49.90%	+71.33%
Argentina II	+124.83%	+159.24%	+119.56%

4.1 Experiment: Four Coevolutionary Strategies

In [18] two thermal plants coevolved. They are the same used as evolutionary agents in [17]: Argentina I and Argentina II, respectively.

Here we report an experiment in which Argentina I and Argentina II are coevolved together with other two natural gas plants: TermoRio and Ibirité. The fitness of the best individual of each specie during 1,000 generations of the evolutionary process (training period) is depicted in Fig. 1. Notice that, after a short unstable period lasting about 100 genera-

[8] g_h^j is denoted by g for short.

[9] The choice of the evolutionary parameters is not subject to any optimization process.

[10] The training process can take about 2 hours of processing time for 2 species, corresponding to 500 generations, and above 6 hours for 4 species and 1,000 generations. Experiments were done using a Pentium 4,2 GHz 256 Mb RAM PC running GNU/Linux Fedora.

ISBN: 978-989-95079-6-8

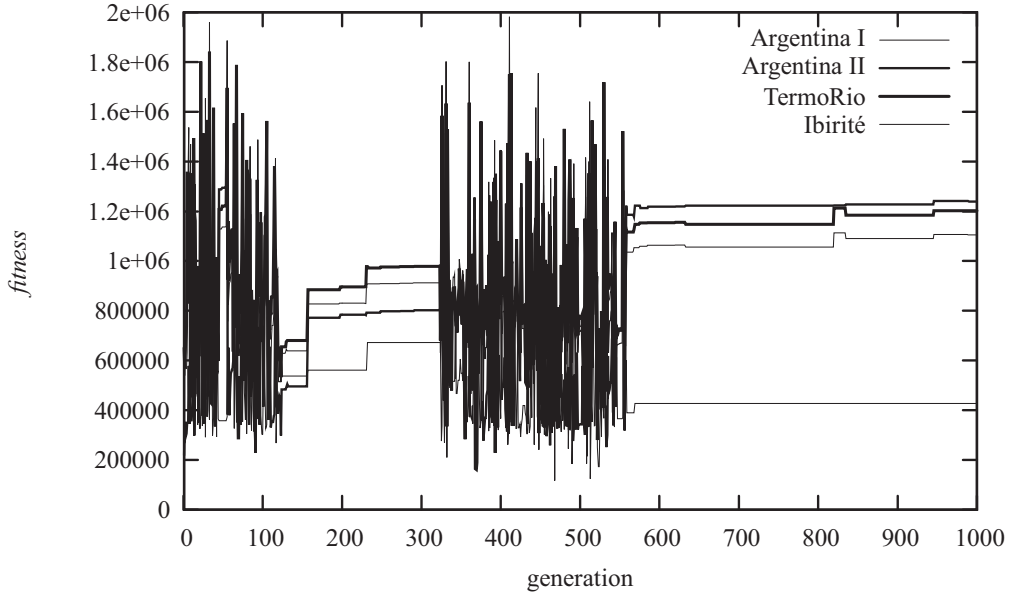


Figure 1: Fitness for coevolutionary species (4 thermal plants: Argentina I, Argentina II, TermoRio and Ibirité).

tions, the fitness of the best individual of each specie increases and stabilizes after about 200 generations and suddenly fitness behavior becomes unstable again. After approximately another 250 generations, the fitness becomes stable again, beginning at 550^{th} generation, and remains stable until the end of the evolutionary process window (1,000 generations). Notice also that the resulting outcome for the specie representing Argentina I becomes worse than in the one achieved during the first stable interval. This behavior suggests that the other three thermal plants may have learned how to exploit Argentina I agent weakness. All the remaining agents increase their profits. A close look at the fitness values shows that the species interact with each other during the whole evolutionary process.

After the end of the evolutionary process, the best individual of each specie is decoded into the corresponding bidding strategies to be played by the four agents. Table 2 summarizes the energy produced and profits obtained when the agents compete in the electricity market using the coevolutionary strategies. Table 2 shows how evolutionary agents outperform the corresponding conservative strategy of bidding their marginal costs at full capacity. The last four rows show the variation when using conservative and evolutionary strategies. We notice in Table 2 that Argentina I decreases energy production by half and still, with the hourly prices that result from the coevolutionary strategic bidding, increases its profit. This is a typical example of an undesirable behavior (from the perspective of the buyers) that mechanism designers wants to avoid before putting any market to work.

Clearly, the coevolutionary strategies affected the electric energy prices. The maximum hourly price of energy reached in the testing period was the same for both, purely conservative strategies or coevolutionary strategies, but coevolution made the average price 10.6% higher. The average price increase for the experiment reported in [18] was smaller, about 7.1%. Similarly as in [18], the individual hourly prices can be up to 53.2% higher for the coevolutionary strategies than purely conservative ones.

4.2 Multiagent Learning Equilibrium

It is important clarify that, while the results presented here and in [18] converge to a stable outcome, some of the experiments performed did not achieve stable outcome at all, remaining unstable for 500 or even 1,000 generations. Unstable outcomes typically occurred when there are many evolutionary agents (5 thermal generators in our experiments). While the dynamics of unstable behavior still is to be investigated, it does not seem to be a novelty, as it has been pointed out by Vidal in [16], who asserts that most multiagent learning systems do not necessarily converge to an equilibrium or stable behavior.

One reason for designers to use learning agents is because they do not know, at design time, the specific circumstances that the agents will face at run time. We will often see a multiagent system with learning agents when the designer can neither predict that an equilibrium solution will be found, nor which equilibria might emerge. The result is a form of closed loop feedback evolving system in which evolution and learning play a complementary role.

As stated in [16], the main reasons behind the difficulty to predict equilibrium solutions of a system include the existence of unpredictable environmental changes that affect the payoffs of the agents, and the fact that, in many systems, an agent only has access to its own set of payoffs. These reasons partially justify the difficulty a designer finds to predict equilibria, if any, a system could reach. However, the agents in a system may still play a game for which an equilibrium exists, even though the designer cannot predict it at design-time. Since in general the payoffs change, it is often the case that the agents are constantly modifying their strategies as an attempt to get better payoffs. This may result in unstable outcomes.

5 Conclusion

A coevolutionary genetic fuzzy systems approach to develop fuzzy bidding strategies was suggested in this paper. Coevo-

ISBN: 978-989-95079-6-8

lutionary approaches provide more realistic representation of agents in auction-based environments because they allow the bidding strategies to interact during the evolutionary process.

The results reported here for an electricity market example show that coevolution can improve agents profits at the expense of increasing the average electricity hourly price paid by demand, an outcome that is undesirable from the perspective of the buyers, and for auction based markets in general. Further research is needed to obtain design methods for bidding strategies in multiagent systems framework populated by intelligent agents. In particular, stability and equilibrium analysis of evolutionary learning in multiagent systems environments still remains a challenge.

Acknowledgment

The first author acknowledges Fábio Stacke for his helpful comments. The second author acknowledges the CNPq, the Brazilian National Research Council, for grant #304 857/2006-8. The authors are also grateful to the anonymous referees for the comments that helped to improve the paper.

Disclaimer

The results, interpretations and conclusions expressed in this work are of exclusive responsibility of its authors and should not be, in any hypothesis, attributed to ANEEL, the Brazilian Electricity Regulatory Agency, neither to its Board of Directors, nor to any Commission the author is affiliated to. The experiments presented in this paper have been accomplished based on general public available data. ANEEL is neither responsible for this work nor any consequence from its use.

References

[1] A. J. Bagnall. A multi-adaptive agent model of generator bidding in the UK market in electricity. In *Genetic and Evolutionary Computation Conference GECCO 2000*, pages 605–612. Morgan Kaufmann, 2000.

[2] T. D. H. Cau and E. Anderson. A co-evolutionary approach to modelling the behaviour of participants in competitive electricity markets. In *2002 IEEE Power Engineering Society Summer Meeting*, volume 3, pages 1534–1540, Chicago, IL, USA, July 2002.

[3] H. Chen, K. Wong, D. Nguyen, and C. Chung. Analyzing oligopolistic electricity market using coevolutionary computation. *IEEE Transactions on Power Systems*, 21(1):143–152, Feb. 2006.

[4] H. Chen, K. Wong, X. Wang, and C. Chung. A coevolutionary approach to modeling oligopolistic electricity markets. In *IEEE 2005 Power Engineering Society General Meeting*, volume 1, pages 230–236, June 2005.

[5] O. Cordón, F. Herrera, F. Hoffman, and L. Magdalena. *Genetic Fuzzy Systems: Evolutionary tuning and learning of fuzzy knowledge bases*, volume 19 of *Advances in Fuzzy Systems: Applications and Theory*. World Scientific, Singapore, 2001.

[6] A. K. David and F. S. Wen. Strategic bidding in competitive electricity markets: A literature survey. In *IEEE PES 2000 Summer Power Meeting*, volume 4, pages 2168–2173, Seattle, USA, July 2000. IEEE Power Engineering Society, IEEE.

[7] R. Green. Competition in generation: The economic foundations. *Proceedings of the IEEE*, 88(2):128–139, Feb. 2000.

[8] S. A. Harp, S. Brignone, B. F. Wollenberg, and T. Samad. SEPIA: A simulator for electric power industry agents. *IEEE Control Systems Magazine*, 20(4):53–69, Aug. 2000.

[9] F. Herrera. Genetic fuzzy systems: Taxonomy, current research trends and prospects. *Evolutionary Intelligence*, 1:27–46, 2008.

[10] W. D. Hillis. Co-evolving parasites improve simulated evolution as an optimization procedure. *Physica D*, 42:228–234, 1990.

[11] S. Phelps, S. Parsons, P. McBurney, and E. Sklar. Co-evolution of auction mechanims and trading strategies: Towards a novel approach to microeconomic desing. In *Proceedings of the 2nd Workshop on Evolutionary Computation and Multi-Agent Systems*, New York, NY, 2002.

[12] S. G. Phelps. *Evolutionary Mechanism Design*. PhD thesis, Univeristy of Liverpool, UK, July 2007.

[13] M. A. Potter and K. A. De Jong. Cooperative coevolution: An architecture for evolving coadapted subcomponents. *Evolutionary Computation*, 8(1):1–29, 2000.

[14] I. Praça, C. Ramos, Z. Vale, and M. Cordeiro. A new agent-based framework for the simulation of electricity markets. In *Proceedings of the IEEE/WIC International Conference on Intelligent Agents (IAT'03)*, pages 1931–1938, Halifax, Canada, 2003. IEEE.

[15] L. Tesfatsion and K. L. Judd, editors. *Handbook of Computational Economics: Agent-Based Computational Economics*, volume 2 of *Handbooks in Economics*. North Holland, 2006.

[16] J. M. Vidal. Learning in multiagent systems: An introduction from a game-theroretic perspective. *Lecture Notes on Artifical Intelligence*, 2636:202–215, 2003.

[17] I. Walter and F. Gomide. Design of coordination strategies in multiagent systems via genetic fuzzy systems. *Soft Computing*, 10(10):903–915, Aug. 2006. Special Issue: New Trends in the Design of Fuzzy Systems.

[18] I. Walter and F. Gomide. Coevolutionary fuzzy multiagent bidding strategies in competitive electricity markets. In *3^{rd} International Workshop on Genetic and Evolving Fuzzy Systems (GEFS 08)*, pages 53–58, Witten-Bommerholz, Germany, Mar. 2008. IEEE.

[19] G. Xiong, T. Hashiyama, and S. Okuma. An evolutionary computation for supplier bidding strategy in electricity auction market. In *IEEE Power Engineering Society Transmission and Distribution Conference*, volume 1, pages 83–88, Yokohama, Japan, 2002. IEEE.

ISBN: 978-989-95079-6-8

A Type-2 Fuzzy Portfolio Selection Problem Considering Possibility Measure and Crisp Possibilistic Mean Value

Takashi Hasuike, Hiroaki Ishii

Graduate School of Information Science and Technology, Osaka University, Japan
2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
Email: {thasuike, ishii}@ist.osaka-u.ac.jp

Abstract—*This paper considers a portfolio selection problem with type-2 fuzzy future returns involving ambiguous and subjectivity. Since this proposed problem is not well-defined due to fuzziness, introducing the fuzzy goal for the total future return and the degree of possibility, the main problem is transformed into the standard fuzzy programming problem including the secondary fuzzy numbers. Furthermore, using the hybrid solution approaches based on the linearity of the deterministic equivalent problem and the crisp possibilistic mean value, the efficient solution is constructed.*

Keywords— Portfolio selection, Type-2 fuzzy, Possibility measure, Crisp possibilistic mean value, Efficient solution method

1 Introduction

Portfolio selection problems are standard and most important problems in investment and financial research fields, and various studies have been so far performed. Furthermore, in recent investment fields, not only big companies and institutional investors but also individual investors called Day-Traders invest in stock, currency, land and property. Therefore, the role of investment theory called portfolio theory becomes more and more important. As for the research history on mathematical approach, Markowitz (Markowitz [18]) has proposed the mean-variance analysis model. It has been central to research activity in the real financial field and numerous researchers have contributed to the development of modern portfolio theory (cf. Elton and Gruber [4], Luenberger [17]). On the other hand, many researchers have proposed models of portfolio selection problems which extended Markowitz model; mean-absolute deviation model (Konno [13], Konno, et al. [14]), safety-first model [4], Value at Risk and conditional Value at Risk model (Rockafellar and Uryasev [19]), etc.. As a result, nowadays it is common practice to extend these classical economic models of financial investment to various types of portfolio models. In practice, after Markowitz's work, many researchers have been trying different mathematical approaches to develop the theory of portfolio selection.

Particularly, the prediction of future returns is one of the most important factors in theoretical and practical investment, and they have been treated as only random variables in many previous studies. Then, the expected returns and variances also have been assumed to be fixed values. However, investors receive effective or ineffective information from the real world and ambiguous factors usually exist in it. Furthermore, investors often have the subjective prediction for future returns which are not derived from the statistical analysis of historical data. Then, even if investors hold a lot of information from the investment field, it is difficult that the present or future random distribution of each asset is strictly set. Consequently, we need to consider not only random conditions but also ambiguous and subjective conditions for portfolio selection problems.

Recently, in the sense of mathematical programming, some researchers have proposed various types of portfolio models under randomness and fuzziness. These problems with probabilities and possibilities are generally called stochastic programming problems and fuzzy programming problems, respectively, and there are some basic studies using a stochastic programming approach, goal programming approach, etc.., and fuzzy programming approach to treat ambiguous factors as fuzzy sets (Inuiguchi and Ramik [8], Leon, et al. [15], Tanaka and Guo [20], Tanaka, et al. [21], Watada [23]). Furthermore, some researchers have proposed the mathematical programming problems with both randomness and fuzziness as fuzzy random variables (for instance, Katagiri et al. [11, 12]). In the studies [11, 12], fuzzy random variables were related with the ambiguity of the realization of a random variable and dealt with a fuzzy number that the center value occurs according to a random variable. Then, Yazenin considered some models for portfolio selection problems in the probabilistic-possibilistic environment, that profitabilities of financial assets are fuzzy random variables (Yazenin [25, 26]). On the other hand, future returns may be dealt with random variables derived form the statistical analysis, whose parameters are assumed to be fuzzy numbers due to the decision maker's subjectivity, i.e., random fuzzy variables which Liu (Liu [16]) defined. There are a few studies of random fuzzy programming problem (Katagiri et al. [9, 10], Huang [7]). Most recently, Hasuike et al. [6] proposed several portfolio selection models including random fuzzy variables and developed the analytical solution method.

However, in [6], the random distribution of each asset is assumed to be a normal distribution. From some practical studies with respect to the present practical market, it is not clear that price movements of assets occur according to normal distributions. In fact, considering the existence of various types of investors in the practical market and the subjectivity of investors, it is important that we need to develop a new portfolio selection problem to deal with a lot of subjectivity. In this paper, we assume future returns to be Type-2 fuzzy numbers which can be dealt with various types of membership functions, and propose a new Type-2 portfolio selection problem.

ISBN: 978-989-95079-6-8

In the sense of mathematical programming, since the proposed model is not formulated as a well-defined problem due to fuzziness, we need to set some certain optimization criterion so as to transform into well-defined problems. In this paper, introducing a fuzzy goal and the degree of possibility, we transform the main problem into the possibility maximization problem. However, this problem includes the fuzzy number and it is also not a well-defined. Therefore, in order to solve possibility maximization problem analytically, we introduce the crisp possibilistic mean value proposed by Carlsson and R. Fullér [1] and develop an efficient solution method to find a global optimal solution of deterministic equivalent problem.

This paper is organized in the following way. In Section 2, we introduce mathematical concepts of type-2 fuzzy set and crisp possibilistic mean value. In Section 3, we propose a type-2 fuzzy portfolio selection problem maximizing the total future return. Then, introducing the degree of possibility and the crisp possibilistic mean value, we transform the proposed model into the deterministic equivalent problem. In Section 4, in order to compare our proposed models with other models for portfolio selection problems, we provide a numerical example derived from current practical market data. Finally, in Section 5, we conclude this paper.

2 Mathematical concept

2.1 Type-2 fuzzy set

A type-2 fuzzy set is a set in which we also have uncertainty about the membership function, i.e., a type-2 fuzzy set is characterized by a fuzzy membership function whose grade for each element is a fuzzy set [0,1] (Castillo and Melin [3]). In the real world, there are many problems where the decision maker cannot determine the exact form of the membership function such as in time series prediction because of noise in the data. Therefore, it is important that the type-2 fuzzy set is introduced into real world problems.

Example 1 (Castillo and Melin [3])
Consider the case of a fuzzy set characterized by a Gaussian membership function with mean m and a standard deviation that can take values in $[\sigma_1, \sigma_2]$, i.e.,

$$\mu(x) = \exp\left\{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\},\ \sigma \in [\sigma_1, \sigma_2]$$

In this paper, with respect to future returns we treat the following *L*-shape fuzzy numbers as fuzzy numbers based on the type-2 fuzzy set:

$$\mu_{\tilde{r}_j}(\omega) = \max\left\{0, L\left(\frac{\omega - \bar{r}_j}{\alpha_j}\right)\right\},\ \alpha_j \in [\alpha_j^L, \alpha_j^U] \tag{1}$$

where $L(\omega)$ is a shape function from $\mathbb{R}$ to $\mathbb{R}$ satisfying the following conditions:

1. $L(-\omega) = L(\omega)$ for $\forall \omega \in \mathbb{R}$
2. $L(0) = 1$
3. $L(\cdot)$ is nonincreasing on $[0, \infty)$
4. Let $t_0 = \inf\{t > 0 | L(t) = 0\}$. Then $0 < t_0 < \infty$

The *L*-shape fuzzy number includes the more general membership function than the Gaussian in Example 1. Therefore, by using the *L*-shape fuzzy number, it is possible that we represent more versatile social problems.

2.2 Crisp possibilistic mean value

Carlsson and Fullér [1] introduced the notation of crisp possibilistic mean value of continuous possibility distribution, which are consistent with the extension principle. Let A a fuzzy number. Then, $[A]^\gamma$ denote the γ -level set of A as the following form:

$$[A]^\gamma = [a_L(\gamma),\ a_U(\gamma)]$$

Using this γ -level set of A, the crisp possibilistic mean value of A is introduced as follows:

$$E(A) = \int_0^1 \gamma\left(a_L(\gamma) + a_U(\gamma)\right) d\gamma$$

Example 2
In the case that a fuzzy number A is a trapezoidal fuzzy number with tolerance interval $[a, b]$, left spread α and right spread β. Then, the crisp possibilistic mean value of A is

$$\begin{aligned} E(A) &= \int_0^1 \gamma\left(a - (1-\gamma)\alpha + b + (1-\gamma)\beta\right) d\gamma \\ &= \frac{a+b}{2} + \frac{\beta - \alpha}{6} \end{aligned}$$

This crisp possibilistic mean value has been applied to various types of portfolio selection problems with fuzzy numbers (in detail, see [5]), and it becomes one of the most useful tool in the sense of fuzzy programming problem.

3 Type-2 fuzzy portfolio selection problem

In this paper, we deal with the following portfolio selection problem with future returns based on the type-2 fuzzy set maximizing total future returns.

$$\begin{aligned} &\text{Maximize} && \sum_{j=1}^n \tilde{r}_j x_j \\ &\text{subject to} && \sum_{j=1}^n x_j = 1, \quad 0 \le x_j \le p_j, \quad j = 1, 2, \cdots, n \end{aligned} \tag{2}$$

where $\boldsymbol{x} = (x_1, \cdots, x_n)$ is a composition of the portfolio and p_j is a limited upper rate of *j*th asset. In this problem, the objective function $\tilde{Z} = \sum_{j=1}^n \tilde{r}_j x_j$ is also a fuzzy number characterized by the following membership function:

ISBN: 978-989-95079-6-8

$$\mu_{\tilde{Z}}(\omega)=\max\left\{0,L\left(\frac{\omega-\sum_{j=1}^{n}\overline{r}_j x_j}{\sum_{j=1}^{n}\alpha_j x_j}\right)\right\},\ \alpha_j\in\left[\alpha_j^L,\alpha_j^U\right] \tag{3}$$

Due to the fuzziness of objective function, problem (2) is not a well-defined problem, and so we need to set some certain optimization criterion. Until now, researchers have proposed some solution approaches to solve fuzzy programming problem based on not only strict solution methods such as linear and nonlinear programming but also approximate solution methods such as GA and NN. In this paper, we focus on the strict solution method in the sense of mathematical programming, and so we introduce the possibility measure as follows:

$$\mathrm{Pos}\left\{\sum_{j=1}^{n}\tilde{r}_j x_j\geq f\right\}=\sup_{\omega}\left\{\mu_{\tilde{Z}}(\omega)\middle|\omega\geq f\right\}$$

where f is the target value to total future return. The possibility measure means that the total future return is more than f as much as possible in the aspiration level of each investor.

3.1 Possibility maximization model for the proposed type-2 fuzzy portfolio selection problem

On the other hand, in practical situations, the investor considers increasing the goal of total future return and that of possibility, simultaneously. Furthermore, considering many real decision cases and taking account of the vagueness of human judgment and flexibility for the execution of a plan, the investor often has subjective and ambiguous goals with respect to the target return f such as "Total future return $\sum_{j=1}^{n}\tilde{r}_j x_j$ is approximately larger than f_1 .". In this subsection, we propose the more flexible model considering the aspiration level to the goal for the total future return. We represent the subjective and ambiguous goals with respect to f as a fuzzy goal characterized by the following membership function:

$$\mu_{\tilde{G}}(f)=\begin{cases}1 & f_1\leq f\\ g(f) & f_0\leq p<f_1\\ 0 & f<f_0\end{cases} \tag{4}$$

where $g(f)$ is the strict increasing function. Furthermore, using a concept of possibility measure, we introduce the degree of possibility as follows:

$$\prod_{\tilde{Z}}(\tilde{G})=\sup_{f}\min\left\{\mu_{\tilde{Z}}(f),\mu_{\tilde{G}}(f)\right\} \tag{5}$$

Using this degree of possibility, we formulate a possibility maximization model for the proposed portfolio selection problem as the following form:

$$\begin{aligned}&\text{Maximize}\quad \prod_{\tilde{Z}}(\tilde{G})\\ &\text{subject to}\quad \sum_{j=1}^{n}x_j=1,\quad 0\leq x_j\leq p_j,\quad j=1,2,\cdots,n\end{aligned} \tag{6}$$

Then, this problem is equivalently transformed into the following problem introducing a parameter h:

$$\begin{aligned}&\text{Maximize}\quad h\\ &\text{subject to}\quad \prod_{\tilde{Z}}(\tilde{G})\geq h,\\ &\qquad\sum_{j=1}^{n}x_j=1,\quad 0\leq x_j\leq p_j,\quad j=1,2,\cdots,n\end{aligned} \tag{7}$$

In this problem, the possibility constraint $\prod_{\tilde{Z}}(\tilde{G})\geq h$ is transformed into the following form:

$$\begin{aligned}&\prod_{\tilde{Z}}(\tilde{G})\geq h\\ \Leftrightarrow\ &\sup_{f}\min\left\{\mu_{\tilde{Z}}(f),\mu_{\tilde{G}}(f)\right\}\geq h\\ \Leftrightarrow\ &\mu_{\tilde{Z}}(f)\geq h,\ \mu_{\tilde{G}}(f)\geq h\\ \Leftrightarrow\ &L\left(\frac{f-\sum_{j=1}^{n}\overline{r}_j x_j}{\sum_{j=1}^{n}\alpha_j x_j}\right)\geq h,\ f\geq g^{-1}(h)\\ \Leftrightarrow\ &f\leq\sum_{j=1}^{n}\overline{r}_j x_j+L^*(h)\sum_{j=1}^{n}\alpha_j x_j,\ f\geq g^{-1}(h)\\ \Leftrightarrow\ &\sum_{j=1}^{n}\overline{r}_j x_j+L^*(h)\sum_{j=1}^{n}\alpha_j x_j\geq g^{-1}(h)\end{aligned} \tag{8}$$

Using these inequalities, the main problem (6) is equivalently transformed into the following problem:

$$\begin{aligned}&\text{Maximize}\quad h\\ &\text{subject to}\quad \sum_{j=1}^{n}\overline{r}_j x_j+L^*(h)\sum_{j=1}^{n}\alpha_j x_j\geq g^{-1}(h),\\ &\qquad\sum_{j=1}^{n}x_j=1,\quad 0\leq x_j\leq p_j,\quad j=1,2,\cdots,n\end{aligned} \tag{9}$$

In this problem, if each spread of fuzzy number α_j is fixed, this problem is analytically solved by using the hybrid solution method of bisection algorithm on parameter h and the linear programming problem. However, in the case that α_j is not fixed but a fuzzy number, it is difficult that this problem is analytically solved by standard linear programming approaches due to including secondary fuzzy numbers.

3.2 Solution method based on the crisp possibilistic mean value

In order to solve problem (9) in the sense of mathematical programming, we introduce the crisp possibilistic mean value. Subsequently, we assume that each spread $\tilde{\alpha}_j$ is the following trapezoidal fuzzy number:

ISBN: 978-989-95079-6-8

$$\mu_{\tilde{\alpha}_j}(\omega)=\begin{cases}\dfrac{\omega-\alpha_j^L}{\underline{\alpha}_j-\alpha_j^L} & \left(\alpha_j^L\le\omega<\underline{\alpha}_j\right)\\ 1 & \left(\underline{\alpha}_j\le\omega\le\overline{\alpha}_j\right)\\ \dfrac{\alpha_j^U-\omega}{\alpha_j^U-\overline{\alpha}_j} & \left(\overline{\alpha}_j<\omega\le\alpha_j^U\right)\\ 0 & \text{otherwise}\end{cases},\ j=1,2,...,n \qquad (10)$$

Using these trapezoidal fuzzy numbers and the fuzzy extension principle, the membership function of fuzzy number $\tilde{\alpha}(\boldsymbol{x})=\sum_{j=1}^{n}\tilde{\alpha}_j x_j$ is also the following trapezoidal fuzzy number:

$$\mu_{\tilde{\alpha}(\boldsymbol{x})}(\omega)=\begin{cases}\dfrac{\omega-\sum_{j=1}^{n}\alpha_j^L x_j}{\sum_{j=1}^{n}\underline{\alpha}_j x_j-\sum_{j=1}^{n}\alpha_j^L x_j} & \left(\sum_{j=1}^{n}\alpha_j^L x_j\le\omega<\sum_{j=1}^{n}\underline{\alpha}_j x_j\right)\\ 1 & \left(\sum_{j=1}^{n}\underline{\alpha}_j x_j\le\omega\le\sum_{j=1}^{n}\overline{\alpha}_j x_j\right)\\ \dfrac{\sum_{j=1}^{n}\alpha_j^U x_j-\omega}{\sum_{j=1}^{n}\alpha_j^U x_j-\sum_{j=1}^{n}\overline{\alpha}_j x_j} & \left(\sum_{j=1}^{n}\overline{\alpha}_j x_j<\omega\le\sum_{j=1}^{n}\alpha_j^U x_j\right)\\ 0 & \text{otherwise}\end{cases},\ j=1,2,...,n \qquad (11)$$

Then, the γ -level set of $\tilde{\alpha}(\boldsymbol{x})$ is as follows:

$$\left[\tilde{\alpha}(\boldsymbol{x})\right]^{\gamma}=\left[\gamma\sum_{j=1}^{n}\underline{\alpha}_j x_j+(1-\gamma)\sum_{j=1}^{n}\alpha_j^L x_j,\ \gamma\sum_{j=1}^{n}\overline{\alpha}_j x_j+(1-\gamma)\sum_{j=1}^{n}\alpha_j^U x_j\right] \qquad (12)$$

Therefore, from the mathematical concept in Subsection 2.2, the crisp possibilistic mean value of $\tilde{\alpha}(\boldsymbol{x})$ is as follows:

$$\begin{aligned}E\left(\tilde{\alpha}(\boldsymbol{x})\right)&=\int_0^1\gamma\left(\gamma\sum_{j=1}^{n}\left(\underline{\alpha}_j+\overline{\alpha}_j-\alpha_j^L-\alpha_j^U\right)x_j+\sum_{j=1}^{n}\left(\alpha_j^L+\alpha_j^U\right)x_j\right)d\gamma\\&=\frac{1}{3}\sum_{j=1}^{n}\left(\underline{\alpha}_j+\overline{\alpha}_j\right)x_j+\frac{1}{6}\sum_{j=1}^{n}\left(\alpha_j^L+\alpha_j^U\right)x_j\end{aligned} \qquad (13)$$

Consequently, problem ( ) is transformed into the following problem from the viewpoint of crisp possibilistic mean value:

$$\begin{aligned}&\text{Maximize}\quad h\\&\text{subject to}\quad\sum_{j=1}^{n}\overline{r}_j x_j+\frac{L^*(h)}{3}\sum_{j=1}^{n}\left(\underline{\alpha}_j+\overline{\alpha}_j+\frac{1}{2}\left(\alpha_j^L+\alpha_j^U\right)\right)x_j\ge g^{-1}(h),\\&\qquad\sum_{j=1}^{n}x_j=1,\quad 0\le x_j\le p_j,\quad j=1,2,\cdots,n\end{aligned} \qquad (14)$$

In this problem, coefficients $\underline{\alpha}_j+\overline{\alpha}_j+\frac{1}{2}\left(\alpha_j^L+\alpha_j^U\right)$ are initially set as fixed values, and so this problem is similar to problem (9), which is one of the standard possibility programming problem. Then, it should be noted here that problem (14) is a nonconvex programming problem and it is not directly solved by the linear programming techniques or convex programming techniques. However, a decision variable h is involved only in first constraint. Therefore, we introduce the following subproblem involving a parameter q :

$$\begin{aligned}&\text{Maximize}\quad\sum_{j=1}^{n}\overline{r}_j x_j+\frac{L^*(q)}{3}\sum_{j=1}^{n}\left(\underline{\alpha}_j+\overline{\alpha}_j+\frac{1}{2}\left(\alpha_j^L+\alpha_j^U\right)\right)x_j\\&\text{subject to}\quad\sum_{j=1}^{n}x_j=1,\quad 0\le x_j\le p_j,\quad j=1,2,\cdots,n\end{aligned} \qquad (15)$$

In the case that parameter q is fixed, problem (15) is degenerated to a linear programming problem. Therefore, in a way similar to the standard possibility maximization model, we construct the following efficient solution method using the hybrid approach of bisection algorithm and linear programming based on the study [6].

Solution algorithm

STEP1: Elicit the membership function of a fuzzy goal for the total future return with respect to the objective function value.

STEP2: Set $q\leftarrow 1$ and solve problem (14). If the optimal objective value $Z(q)$ of problem (14) satisfies $Z(q)>g^{-1}(q)$, then terminate. In this case, the obtained current solution is an optimal solution of main problem.

STEP 3: Set $q\leftarrow 0$ and solve problem (14). If the optimal objective value $Z(q)$ of problem (14) satisfies $Z(q)\le g^{-1}(q)$, then terminate. In this case, there is no feasible solution and it is necessary to reset a fuzzy goal for the probability or the aspiration level f.

STEP 4: Set $U_q\leftarrow 1$ and $L_q\leftarrow 0$.

STEP 5: Set $\gamma\leftarrow\left(U_q+L_q\right)/2$.

STEP 6: Solve problem (14) and calculate the optimal objective value $Z(q)$ of problem (14). If $Z(q)>g^{-1}(q)$, then set $L_q\leftarrow q$ and return to Step 5. If $Z(q)\le g^{-1}(q)$, then set $U_q\leftarrow q$ and return to Step 5. If $Z(q)=g^{-1}(q)$, then terminate the algorithm. In this case, $\boldsymbol{x}^*(q)$ is equal to a global optimal solution of main problem.

4 Numerical example

In order to compare our proposed models with other models for portfolio selection problems, let us consider a numerical example based on the data of securities on the Tokyo Stock Exchange. In this paper, we compare our proposed model (14) in Section 3 with Carlsson et al. model [2] and Vercher et al. model [22]. These models are type-1 fuzzy portfolio models and include the possibilistic mean value and variance. These problems are formulated as the following form:

(Carlsson et al. model)

$$\begin{aligned}&\text{Maximize}\quad\sum_{j=1}^{n}\frac{1}{2}\left[a_j+b_j+\frac{1}{3}\left(\beta_j-\alpha_j\right)\right]x_j\\&\qquad-\frac{0.0123}{4}\left(\sum_{j=1}^{n}\frac{1}{2}\left[b_j-a_j+\frac{1}{3}\left(\alpha_j+\beta_j\right)\right]x_j\right)^2-\frac{0.0123}{72}\left(\sum_{j=1}^{n}\left(\alpha_j+\beta_j\right)x_j\right)^2\\&\text{subject to}\quad\sum_{j=1}^{n}x_j=1,\quad 0\le x_j\le p_j,\quad j=1,2,\cdots,n\end{aligned}$$

ISBN: 978-989-95079-6-8

(Vercher et al. model)

$$\text{Minimize} \quad \sum_{j=1}^{n}\frac{1}{2}\left[b_j - a_j + \frac{1}{3}(\alpha_j + \beta_j)\right]x_j$$

$$\text{subject to} \quad \sum_{j=1}^{n}\frac{1}{2}\left[a_j + b_j + \frac{1}{3}(\beta_j - \alpha_j)\right]x_j \geq \rho,$$

$$\sum_{j=1}^{n} x_j = 1, \quad 0 \leq x_j \leq p_j, \quad j = 1,2,\cdots,n$$

Let us consider ten securities shown in Table 1, whose mean values and standard deviations are based on historical data in the decade between 1995 and 2004. Then, we introduce the asset allocation rate x_j, $(j = 1,2,...,10)$ to each security, and its upper value is assumed to be 0.2.

Table 1: Sample data from Tokyo Stock Exchange

Returns	Sample mean	SD
R1	0.055	0.445
R2	0.046	0.289
R3	0.015	0.306
R4	0.114	0.208
R5	0.043	0.253
R6	0.034	0.269
R7	0.018	0.230
R8	0.171	0.297
R9	0.087	0.388
R10	0.090	0.318

With respect to our proposed type-2 fuzzy portfolio model, parameter $\overline{r}_j$ of the primary *L*-fuzzy number is assumed to be sample mean in Table 1. Then, in two previous models parameters, parameters of trapezoidal fuzzy number a_j, b_j, α_j, β_j, $(j = 1,2,...,10)$ based on study [22] are shown in Table 2 using the historical data in Table 1.

Table 2. Each parameter value based on the historical data

Returns	a_j	b_j	α_j	β_j
R1	-0.123	0.005	0.239	0.868
R2	-0.069	0.069	0.301	0.467
R3	-0.129	0.025	0.200	0.713
R4	0.005	0.177	0.198	0.235
R5	-0.082	0.114	0.217	0.323
R6	-0.052	0.108	0.290	0.382
R7	-0.056	0.067	0.236	0.369
R8	0.060	0.193	0.310	0.456
R9	-0.093	0.130	0.312	0.626
R10	0.009	0.236	0.563	0.264

The secondary membership functions for $\tilde{\alpha}_j$ of type-2 fuzzy numbers are assumed to be triangle fuzzy numbers, whose $\underline{\alpha}_j = \overline{\alpha}_j$ are variances based on the SD in Table1 and the spreads $\underline{\alpha}_j - \alpha_j^L$ and $\alpha_j^U - \overline{\alpha}_j$ become α_j and β_j in Table 2, respectively. Therefore, using these data and introducing the following fuzzy goals for our proposed model;

$$\mu_G(f) = \begin{cases} 1 & (0.15 \leq f) \\ \dfrac{f - 0.12}{0.03} & (0.12 \leq f < 0.15) \\ 0 & (f < 0.12) \end{cases},$$

we set the parameter ρ in the Vercher et al. model as $\rho = 0.06$, and obtain optimal portfolios for three models shown in Table 3.

Subsequently, we consider the case where an investor purchases securities at the end of 2004 according to each portfolio shown in Table 3. Then, the total return of three models at term ends of 2005, 2006 and 2007 become the following values shown in Table 4, respectively.

Table 3. Each optimal solution with respect to three models

	Proposed model	Carlsson model	Vercher model
R1	0.200	0	0
R2	0	0	0
R3	0	0	0
R4	0.200	0.200	0.200
R5	0	0	0.200
R6	0	0.200	0.200
R7	0	0	0.200
R8	0.200	0.200	0.200
R9	0.200	0.200	0.200
R10	0.200	0.200	0.200

Table 4. Total return to the portfolio of each model

Term end	Proposed	Carlsson	Vercher
2005	0.2524	0.2194	0.2852
2006	0.4050	0.3575	0.3257

From the result in Table 4, we find that our proposed model obtains the larger total return than the Carlsson et al. model dealt with the only crisp possibilistic mean value. Furthermore, from the total return at the term end of 2006, in the case that the investment period is longer, our proposed model obtains much larger total return.

5 Conclusion

In this paper, we have proposed a new portfolio selection problem based on the type-2 fuzzy set. Our proposed model has been initially not a well-defined problem due to primary and secondary fuzzy numbers. Therefore, in order to solve analytically in the sense of mathematical programming, we have introduced the possibility maximization model with fuzzy goal for the object and degree of possibility. Furthermore, by using the concept of crisp possibilistic mean value for the secondary fuzzy numbers, the main problem has been equivalently transformed into the parametric linear programming problem. Consequently, we have developed the efficient solution method to combine the bisection algorithm with the standard linear programming approach. Since this proposed model includes various types of investor's subjectivity in the form of type-2 fuzzy number, it is more versatile than previous models. Our proposed model is dealt with various practical situations in the investment such as the case putting the whole policy of investors including a lot of subjectivity by using the possibilistic mean value and the case

ISBN: 978-989-95079-6-8

indicating the sensible decision making to the investor with shaken investment policy due to uncertain but attractive information.

In this paper, as the criterion for the object, we introduced the possibility measure and the crisp possibilistic mean value. On the other hand, there are various types of criterions with respect to fuzzy object and constraints such as necessity and credibility measure, fuzzy interval approach, etc.. Therefore, as the future works, we need to consider more general and versatile type-2 fuzzy portfolio selection problems with such criterions.

Acknowledgment

The first author would like to thank to Japan Society for the Promotion of Science to support this work.

References

[1] C. Carlsson and R. Fullér, "On possibilistic mean value and variance of fuzzy numbers", *Fuzzy Sets and Systems* 122, pp. 315-326, 2001.

[2] C. Carlsson, R. Fullér and P. Majlender, "A possibilistic approach to selecting portfolios with highest utility score", *Fuzzy Sets and Systems* 131, pp. 13-21, 2002.

[3] O. Castillo and P. Melin, *Type-2 Fuzzy Logic: Theory and Applications*, Studies in Fuzziness and Soft Computing 223, Springer, 2008.

[4] E.J. Elton, M.J. Gruber, *Modern Portfolio Theory and Investment Analysis*, Wiley, New York, 1995.

[5] Y. Fang, K.K. Lai and S. Wang, *Fuzzy Portfolio Optimization; Theory and Method*, LNEMS 609, Springer, 2008.

[6] T. Hasuike, H. Katagiri, H. Ishii, Portfolio selection problems with random fuzzy variable returns, *Fuzzy Sets and Systems*, to appear.

[7] X. Hung, "Two new models for portfolio selection with stochastic returns taking fuzzy information", *European Journal of Operational Research* 180, pp. 396-405, 2007.

[8] M. Inuiguchi, J. Ramik, "Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem", *Fuzzy Sets and Systems* 111, pp. 3-28, 2000.

[9] H. Katagiri, T. Hasuike, H.Ishii, I. Nishizaki, "Random Fuzzy Programming Models based on Possibilistic Programming", *Proceedings of the 2008 IEEE International Conference on Systems, Man and Cybernetics*, pp. 1788-1793.

[10] H. Katagiri, H. Ishii and M. Sakawa, "Linear programming problems with random fuzzy variable coefficients", *Proceedings of 5th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty* 1, pp. 55-58, 2002.

[11] H. Katagiri, H. Ishii, M. Sakawa, "On fuzzy random linear knapsack problems", *Central European Journal of Operations Research* 12, pp. 59-70, 2004.

[12] H. Katagiri, M. Sakawa, H. Ishii, "A study on fuzzy random portfolio selection problems using possibility and necessity measures", *Scientiae Mathematicae Japonocae* 65, pp. 361-369, 2005

[13] H. Konno, "Piecewise linear risk functions and portfolio optimization", *Journal of Operations Research Society of Japan* 33, pp. 139-156, 1990.

[14] H. Konno, H. Shirakawa, H. Yamazaki, "A mean-absolute deviation-skewness portfolio optimization model", *Annals of Operations Research* 45, pp. 205-220, 1993.

[15] R T. Leon, V. Liern, E. Vercher, "Validity of infeasible portfolio selection problems: fuzzy approach", *European Journal of Operational Researches* 139, pp. 178-189, 2002.

[16] B. Liu, Theory and Practice of Uncertain Programming, Physica Verlag, 2002.

[17] D.G. Luenberger, *Investment Science*, Oxford Univ. Press, 1997.

[18] H.M. Markowitz, Portfolio Selection: Efficient Diversification of Investments, Wiley, New York, 1959

[19] R.T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk", *Journal of Risk* 2(3), pp. 1-21, 2000.

[20] H. Tanaka and P. Guo, "Portfolio selection based on upper and lower exponential possibility distributions", *European Journal of Operational Researches* 114, pp. 115-126, 1999.

[21] H. Tanaka, P. Guo and I.B. Turksen, "Portfolio selection based on fuzzy probabilities and possibility distributions", *Fuzzy Sets and Systems* 111, pp. 387-397, 2000.

[22] E. Vercher, J.D. Bermúdez and J.V. Segura, "Fuzzy portfolio optimization under downside risk measures", *Fuzzy Sets and Systems* 158, pp. 769-782, 2007.

[23] J. Watada, "Fuzzy portfolio selection and its applications to decision making", *Tatra Mountains Math. Pub*. 13, pp. 219-248, 1997.

[24] A.V. Yazenin, "Fuzzy and Stochastic Programming", *Fuzzy Sets and Systems*, 22, pp. 171-180, 1987.

[25] A.V. Yazenin, "Possibilistic-Probabilistic Models and Methods of Portfolio Optimization", *Studies in Computational Intelligence* 36, Springer-Verlag Berlin Heidelberg, Germany, pp. 241-259, 2007.

ISBN: 978-989-95079-6-8

Finding the Best Flexibility Strategies by Using an Integrated Method of FAHP and QFD

Müjde Erol Genevois Tuncay Gürbüz

Industrial Engineering Department, Galatasaray University

Ciragan Cad. No:36 Ortakoy, Istanbul, Turkey
Email: merol@gsu.edu.tr, tgurbuz@gsu.edu.tr

Abstract—*In this study, Turkish automotive sector needs for flexibility will be determined, and the relationships between flexibility levers and flexibility performance criteria will be evaluated with respect to the sectoral needs. The choice of the appropriate lever combinations will be made for the sector by using the results of this evaluation. The flexibility strategies will be ranked by using a fuzzy extended AHP approach. At the second phase of this study, the QFD tool will be used to find the best flexibility lever portfolio related to automotive sector strategies.*

Keywords— Flexibility Management, Fuzzy AHP, QFD, Turkish Automotive Sector.

1 Introduction

The markets in which manufacturers and service firms compete are increasingly influenced by intense foreign competition, rapid technological change, shorter product life-cycles and customers increasingly unwilling to settle for mass-produced items or services with limited value. The "new breed of customer" [1], who demands greater responsiveness to a dynamic set of requirements, and a new competitive environment, which exposes local companies to competition with the companies around the globe, form a new scenario that has challenged firms in most industries ([2]; [3]; [4]). In this new scenario, flexibility may be one of the most important capabilities needed for firms to achieve competitive advantage ([5]; [6]; [7]).

We defined flexibility as the capacity of adaptation under the double constraint of uncertainty and the urgency. This uncertainty can come from the providers (rupture, problems of transport), from the customers (variation of the request) or from the company itself (breakdowns of the equipment, problems of provisioning). The possible behaviors of the company face to these problems are called levers of flexibilities. After the mass production era of Ford and lean management era of Toyota, these days are witnessing the era of flexibility. Modern organization should keep as much as possible flexibility capability to be used in case of an incident.

This paper will briefly summarize the flexibility concept and flexibility levers as well as the existing situation of Turkish automotive industry by the flexibility point of view. It will introduce a model, based on Analytic Hierarchy Process (AHP) and Quality Function Deployment (QFD) methodology, which determine the best flexibility levers' portfolio to deal with consumers needs for flexibility in term of delivering a 5R product (right place-RP, right time-RT, right quantity-RQ, right product-RPr, right price-RPc). The Turkish automotive industry including the parts and components producers is considered in this study as a case industry. Automotive industry is considered as a complete industry covering all entities of the value chain.

In the second section of the paper, the flexibility management literature will be reviewed and Turkish automotive sector needs for flexibility will be determined. In the third section, the fuzzy extended AHP model will be introduced and the flexibility strategies will be ranked by using this approach. At the fourth phase of this study, the QFD tool will be used to find the best flexibility lever portfolio related to automotive sector strategies. The relationships between flexibility levers and flexibility performance criteria will be evaluated with respect to the sectoral needs. The choice of the appropriate lever combinations will be made for the sector by using the results of this evaluation. The study will be concluded with the presentation of the selected portfolio.

2 Flexibility Management

2.1 Literature Review

Flexibility becomes a strategic resource from the late 70s due to the uncertainty in the environment faced by most of the enterprises. This uncertainty is due to several factors, such as diversification in product range, reduced product life cycle, fragmentation and globalization of markets, sophistication of customers etc. In a broad view flexibility is defined as the capability of adaptation to change. The flexibility is a total capacity, which will be activated as a function of events, which are not controlled by the company. It has a passive vision and defensive usage of the capability. But flexibility may create some offensive strategies as well. By offering more diverse products than the competitors, by renewing the products more frequently or by customizing associated services, flexible pioneers can set the rules of game of the industry by creating uncertainty for their competitors. Therefore flexibility is defined as a strategic asset for the companies not only for adapting to the changes in the environment but also to change the environment in favor of itself.

ISBN: 978-989-95079-6-8

Some authors distinguish between internal and external flexibilities. For instance, Swamidass [8] distinguishes machine-level flexibility from plant-level flexibility. The former being "predominantly technology based" and the latter being derived from a combination of technology, infrastructure, design and engineering capabilities, and the competitive goals and objectives of a firm. Upton [9] defines internal flexibility as the operations strategy and the set of capabilities a firm nurtures to respond to its environment, and external flexibility as capabilities possessed by the firm and used to accommodate sources of variability to which the firm must respond and which are seen as flexible by the market. This external dimension fits the two major strategies proposed by Hyun and Ahn [10] for using flexibility: reactive and proactive. In the same vein, Gerwin [11] also suggests two major strategies for using flexibility: adaptive and redefinition. The adaptive strategy refers to the defensive or reactive use of flexible competencies to accommodate unknown uncertainty, while the redefinition strategy refers to the proactive use of flexible competencies to raise customer expectations, increase uncertainty for rivals and gain competitive edge.

The reasons companies are motivated to be flexible include the need to make design changes quickly, when competitors introduce new models and customers start switching supply sources; to focus on volume flexibility, when large customers reduce inventories and their demand rates become volatile; more flexible product mixes, when importers or domestic competitors start offering multiple quality and price levels; to respond quickly and supply the new products/services, when the customer tastes change quickly.

Especially, four types of flexibilities are important in point of view of the consumer: the volume (VF), the product mix (PMF), the new product (NPF) and the design flexibilities (DF). Volume flexibility can be defined as the ability to operate profitably at different production volumes [9]. Product mix flexibility is adopted to deal with uncertainty about the products that will be demanded by customers at a particular period ([12]; [13]). It is characterized by the ability to produce several products at one manufacturing facility without incurring a major cost penalty. New product flexibility is the ability of a system to add or substitute new products to the product mix. Finally, the design flexibility is the ability to change the design of a product very economically and quickly

The need for flexibility can be expressed as a satisfaction of the demand with delivery on time (I), at the right place (II), in required quantity (III) with the right product (IV) and the right price (V).

It may be noted that each of the above external elements are the causes beyond the control of any company. Further, none of these needs can be satisfied by the mass production strategy. Consequently, the classical methods as the affair with large production runs and economies of scale are a thing of the past, and are being replaced by a new concept - flexibility management.

To satisfy the need for various types of flexibility, Aggorwal [14] define three groups of levers or mechanisms. The first group is referred to as levers for internal flexibility. Each of these levers remains under the control of management at all times. The salient levers in this group are Planning/Scheduling Flexibility (1), Sequencing Flexibility (2), Routing Flexibility (3), Labor Flexibility (4), Machine/Equipment Flexibility (5), Design/Development Flexibility (6), Process/Technology Flexibility (7), Raw Materials Flexibility (8), Transport/Shipping Flexibility (9), Layout Flexibility (10), Expansion Flexibility (11), Financial Resources Flexibility (12).

The second group consists of levers for soft flexibilities. Each of these levers can enhance internal flexibility. In contrast to hard levers, they are never completely under management's control as each requires heightened participation and support of specialists and other employee groups. The levers constituting this group and their respective code numbers are Organizational Structure Flexibility (13), Decision Making Flexibility (14), Job Design Flexibility (15), Employee's Willingness for Change Flexibility (16), Managerial Perception Change Flexibility (17). These flexibility levers cannot be measured on a hard (quantitative) scale, as was the case with the levers of the first group; therefore they are designated soft levers flexibilities.

The third group of levers involves managerial manipulation of intangibles. Each of these levers can enhance one or more of the flexibilities either externally or internally. Again, their effectiveness cannot be measured on a hard scale, so they are called intangible levers (flexibilities). These are Reputation Building Flexibility (18), Flexibility in Enhancing Knowledge and Experience Base (19), Flexibility in Identifying Undiscovered and Unused Talents of Workforce (20), Flexibility in Development of Standing Agreements with Suppliers and Customers (21).

2.2 Changing Landscape in Turkish Automotive Industry

The present-day automotive industry of Turkey was first developed within the broader context of import substitution in the 1960s. High tariffs, quantitative restrictions and local content regulations were used to create a local vehicle industry. Access to the local market could only be achieved through local production. In Turkey, the government succeeded in attracting a number of assemblers (most notably Ford, Fiat and Renault), and by the early 1970s cars with a high degree of local content were being produced. While the assemblers were joint ventures between foreign companies and local groups by the late 1970s, a thriving components industry was created, much of which was locally owned. The industry continues to grow in the 1970s and 1980s mainly by protectionist policies. The rapid expansion of the automotive industry arose from a coming together of two distinct interests. However, the developing impact of the automotive industry in the 1990s is likely to be very different from that seen in the 1960s and 1970s. This is because the industry has globalized and changed its governance structures

Protectionist policies continued throughout the early 1990s. In 1995, by signing the Customs Union agreement with European Union, Turkey abolished high taxes over European origin vehicles. With customs union Turkish automotive industry changed its nature and it results Turkish

ISBN: 978-989-95079-6-8

car manufacturers to produce competitive products both in price and in quality. Another impact of the customs union over the industry is the shift of focus from local market to global markets. As a direct result of globalization, local producers have begun to manufacture vehicles for global markets. Due to global strategies of multinational vehicle producers, manufacturing facilities in Turkey are dedicated to manufacture specific models. This shift in the assemblers had an immediate impact over the suppliers that are faced to the obligations of the globalized environment. It become important in this uncertain environment to be armed with performer flexibility levers in order to deal with the rapid changes in the sector.

3 Proposed Methodology

3.1 Ranking Flexibility Strategies with Fuzzy Extended AHP Approach

In this paper the fuzzy extension of one of the most widely used MCDM methods namely FAHP is used to find the relative importance of the problems alternatives and criteria. The choice of the method used in this stage is arbitrary but the wide range and the ease of use for FAHP was determinant for us to choose it. And with fuzzy logic, we wanted to handle as precisely as possible the unprecised qualitative data collected from the experts.

The FAHP approach is introduced, with the use of Triangular Fuzzy Numbers-TFNs for pairwise comparison scale of FAHP according to the method of Chang's [15] fuzzy extent analysis and the correct normalization formula given later by Wang et al. [16].

In the conventional AHP, the pairwise comparisons for each level with respect to the goal are conducted using a nine-point scale proposed by Saaty [17]. According to Zadeh [18], it is very difficult for conventional quantification to define the complex situations, so the notion of a linguistic variable, whose values are words or sentences, is necessary. To assess the relative importance of the criteria and to evaluate the alternatives with respect to the problems criteria an assumed weighting set has been developed. The decision makers can use this linguistic rating set. The triangular fuzzy conversion scale of the linguistic values in the weighting set is shown in Table 1.

Table1. The triangular fuzzy conversion scale

Linguistic Values	***Triangular Fuzzy Numbers***
Very Low (VL) Very Poor (VP)	$\left(\frac{1}{5},\frac{1}{5},\frac{1}{3}\right)$
Low (L) Poor (P)	$\left(\frac{1}{5},\frac{1}{3},1\right)$
Exactly Equal	$(1,1,1)$
Medium (M) Fair (F)	$\left(\frac{1}{2},\frac{3}{2},\frac{5}{2}\right)$
High (H) Good (G)	$(1,3,5)$
Very High (VH) Very Good (VG)	$(3,5,5)$

Assume that $X=\{x_1,x_2,...,x_n\}$ is an object set, and $U=\{u_1,u_2,...,u_m\}$ is a goal set. According to the method of Chang's [24] fuzzy extent analysis, each object is taken and extent analysis is performed for each goal respectively. Therefore, *m* extent analysis values for each object can be obtained, with the following representation:

$$M_{g_i}^1, M_{g_i}^2,..., M_{g_i}^m \text{ for } i=1,2,...,n \quad (1)$$

Note that all the $M_{g_i}^j$, $j=1,2,\ldots,m$, are TFNs representing the performance of the object x_i with regard to each goal u_j. The steps of Chang's [19] extent analysis can be given as in the following:

- First, by fuzzy arithmetic operations, take the sum of each row of the fuzzy comparison matrix.

$$RS_i=\sum_{j=1}^{n}\tilde{a}_{ij}=\left(\sum_{j=1}^{n}l_{ij},\ \sum_{j=1}^{n}m_{ij},\ \sum_{j=1}^{n}u_{ij}\right),\quad i=1,\ldots,n \quad (2)$$

- Using fuzzy synthetic extent analysis, the value of fuzzy synthetic extent with respect to the i^{th} object x_i, $i=1,2,\ldots,n$ that represents the overall performance of the object across all goals can be determined by the following normalization formula given by Wang et al. [25]:

$$\tilde{S}_i=\frac{RS_i}{\sum_{j=1}^{n}RS_j}=\left(\frac{\sum_{j=1}^{n}l_{ij}}{\sum_{j=1}^{n}l_{ij}+\sum_{k=1,k\neq i}^{n}\sum_{j=1}^{n}u_{kj}},\ \frac{\sum_{j=1}^{n}m_{ij}}{\sum_{k=1}^{n}\sum_{j=1}^{n}m_{ij}},\ \frac{\sum_{j=1}^{n}u_{ij}}{\sum_{j=1}^{n}u_{ij}+\sum_{k=1,k\neq i}^{n}\sum_{j=1}^{n}l_{kj}}\right),\quad i=1,\ldots,n \quad (3)$$

- The degree of possibility of $M_1\geq M_2$ is defined as :

$$V(M_1\geq M_2)=\sup_{x\geq y}\left[\min\left(\mu_{M_1}(x),\mu_{M_2}(y)\right)\right] \quad (4)$$

- Compute the degree of possibility of $\tilde{S}_i\geq\tilde{S}_j$ by the following equation where $\tilde{S}_i=(l_i,m_i,u_i)$ and $\tilde{S}_j=(l_j,m_j,u_j)$:

$$V(\tilde{S}_i\geq\tilde{S}_j)=\begin{cases}1, & \text{if } m_i\geq m_j,\\ \dfrac{u_i-l_j}{(u_i-m_i)+(m_j-l_j)}, & \text{if } l_j\leq u_i,\quad i,j=1,\ldots,n;i\neq j\\ 0, & \text{otherwise,}\end{cases} \quad (5)$$

- To compare $\tilde{S}_i=(l_i,m_i,u_i)$ and $\tilde{S}_j=(l_j,m_j,u_j)$, we need both the values of $V(\tilde{S}_i\geq\tilde{S}_j)$ and $V(\tilde{S}_j\geq\tilde{S}_i)$. The degree of possibility for a convex fuzzy number $\tilde{S}$ to be greater than *k* convex fuzzy numbers $\tilde{S}_i$, *i*=1,…,k can be defined by:

$$V(M\geq M_1,\ldots,M_k)=V[(M\geq M_1)\wedge\ldots\wedge(M\geq M_k)]=\min V(M\geq M_i)$$
$$\text{for } i=1,2,...,k$$

- Assume that for the alternative A_i,

$$d'(A_i)=\min V(S_i\geq S_j),\qquad \text{for } j=1,2,...,n;\ j\neq i \quad (6)$$

- Then the weight vector is given by,

$$W'=(d'(A_1),d'(A_2),...,d'(A_n))^T \quad (7)$$

- Via normalization, the normalized weight vectors are,

$$W=(d(A_1),d(A_2),...,d(A_n))^T \quad (8)$$

ISBN: 978-989-95079-6-8

To illustrate our model, a decision-making group is formed that consists of the 5 experts from the Turkish automotive industry. After a detailed discussion on every criterion, four main criteria of flexibility types have been identified for the hierarchical structure. The AHP hierarchy scheme is constructed and shown in Figure 1.

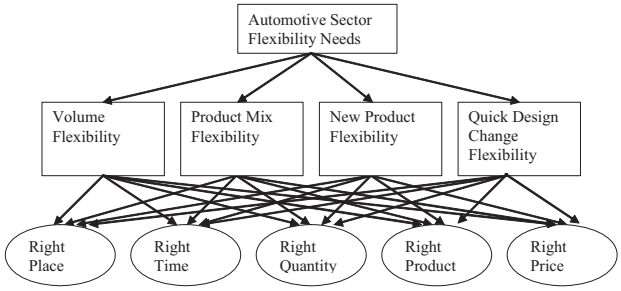


Figure 1. Hierarchical Structure of Flexibility Needs

The aim of the evaluation is to rank the importance of consumers needs for flexibility in the sector. The normalized weight vectors of alternatives with respect to criteria are calculated and shown in Table 2.

Table 2. The normalized weight vectors of the alternatives

	VOLUME	MİX	NEW PRODUCT	QUICK DESIGN CHANGE	
	0,752	0,1	0,045	0,103	
RP	0,108	0,129	0,131	0,083	0,109
RT	0,148	0,144	0,154	0,203	0,154
RQ	0,221	0,094	0,096	0,144	0,195
RPr	0,276	0,228	0,489	0,252	0,278
RPc	0,247	0,404	0,131	0,318	0,265

The need for volume flexibility is highly important in the automotive sector. The consumer demand in this sector is significantly sensible to the global economic indicators. On the other hand, the right price is very important when we are faced to a mix product flexibility problem. In the same way, placing the right product in the market is extremely important when we have to satisfy a new product flexibility need. Finally, the 5R strategies are ranked for their relative importance as follows: $RPr \succ RPc \succ RQ \succ RT \succ RP$.

3.2 Selection of the best flexibility lever portfolio using QFD

Quality function deployment (QFD) is ‘‘an overall concept that provides a means of translating customer requirements into the appropriate technical requirements for each stage of product development and production (i.e., marketing strategies, planning, product design and engineering, prototype evaluation, production process development, production, sales)’’ [19]. QFD belongs to the sphere of quality management methods, offering us a linear and structured guideline for converting the customer’s needs into specifications for, and characteristics of new products and services. The method involves developing four matrixes, or ‘houses’, that we enter by degrees as a project for a given product or production process is developed on increasingly specific levels [20]. In the present article, our attention focuses on the Planning Matrix, or (HOQ) [21] (Fig. 2).

The HOQ provides the specifications for product design (or engineering characteristics) in terms of their relative importance and of target values that have to be reached in design and production. In a sense, the HOQ is the hub of the whole QFD method: its construction enables us to proceed from the customer’s requirements to the design specifications ([22]; [23]). This paper describes the HOQ and its process following the approaches suggested by Brown [24], and Griffin and Hauser [25]. Step 1: Identify the WHATs. The expected benefits in a product or service in the customer’s own words are customer needs and are usually called customer attributes (CA) or ‘‘WHATs’’, area (A) in Fig. 2. In assigning priorities to WHATs, it is necessary to balance efforts in order to accomplish those needs that add value to the customer. The priorities are usually indicated in the area designated as (B) in Fig. 2. Step 2: Determination of HOWs. Engineering characteristics are specified as the ‘‘HOWs’’ of the HOQ and also called measurable requirements. HOWs are identified by a multidisciplinary team [26] and positioned on the area marked as (C) on the matrix diagram, Fig. 2. Step 3: Preparation of the relationship matrix (D). A team judges which WHATs impact which HOWs and to what degree. Step 4: Elaboration of the correlation matrix. The physical relationships among the technical requirements are specified on an array known as ‘‘the roof matrix’’ and identified as (E) in Fig. 2. Step 5: Action plan. The weights of the HOWs, identified as area (F), are placed at the base of the quality matrix. These weights are one of the main outputs of the HOQ, and are determined by

$$\text{Weight (HOW)}_i = V(\text{HOW})_{i1} * \text{imp}(\text{WHAT}_1) + \ldots + V(\text{HOW})_{in} * \text{imp}(\text{WHAT}_n)$$

where $V(\text{HOW})_{in}$ is the correlation value of HOW_i with WHAT_n, and imp (WHAT_n) represents the importance or priority of WHAT_n.

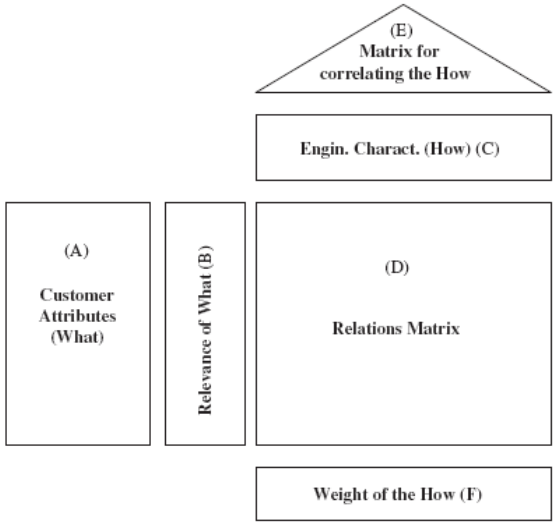


Figure 2. House of quality.

Various quantitative methods have been suggested to use in QFD to improve its reliability and objectiveness, noticeably the methods of management science/operational research (MS/OR), marketing research, and fuzzy logic.

ISBN: 978-989-95079-6-8

			Column Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
			Max Relationship Value in Column	3	3	3	9	3	9	9	3	9	3	9	9	3	3	3	3	3	3	3	3	9
			Requirement Weight	214,47	93,371	98,853	471,84	183,76	276,98	388,36	109,98	312,18	157,28	258,78	296,75	194,62	194,62	155,67	102,97	109,71	155,67	247,58	162,94	467
			Relative Weight	4,61	2,01	2,12	10,14	3,95	5,95	8,35	2,36	6,71	3,38	5,56	6,38	4,18	4,18	3,35	2,21	2,36	3,35	5,32	3,50	10,04
			Difficulty (0=Easy to Accomplish, 10=Extremely Difficult)	4	5	7	6	5	7	9	6	5	8	8	6	6	6	7	5	5	7	6	4	6
			Minimize (▼), Maximize (▲), or Target (x)	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
			Target or Limit Value																					
Row Number	Max Relationship Value in Row	Relative Weight	Flexibility Levers / Flexibility Requirements	Planning/Scheduling Flexibility	Sequencing Flexibility	Routing Flexibility	Labor Flexibility	Machine/Equipment Flexibility	Design/Development Flexibility	Process/Technology Flexibility	Raw Materials Flexibility	Transport/Shipping Flexibility	Layout Flexibility	Expansion Flexibility	Financial Resources Flexibility	Organizational Structure Flexibility	Decision Making Flexibility	Job Design Flexibility	Employee's Willingness for Change Flexibility	Managerial Perception Change Flexibility	Reputation Building Flexibility	Flexibility in Enhancing Knowledge and Experience Base	Flexibility in Identifying Undiscovered and Unused Talents of Workforce	Standing Agreements with Suppliers and Customers
1	9	10,86	Right Place									9				1	1	1		1	1	1		3
2	3	15,35	Right Time	3	3	1	3	1				3	1			1	1	1		1	1	1		3
3	9	19,48	Right Quantity	3	1		9	3		3		3	3	9	3	3	3	1	1		1	3		3
4	9	27,83	Right Product	3	1	3	9	3	9	9	3	3	3	3		3	3	3	3	3	3	3	3	9
5	9	26,48	Right Price	1				1	1	3	1	1			9	1	1	1			1	3	3	3

Figure 3. The HOQ applied to the flexibility levers portfolio choice

Works applying MS/OR methods to QFD include analytic hierarchy process (AHP) to prioritize customer needs (whats) ([27]; [28]; [29]), AHP and benchmarking integrated to rate ‘‘whats’’ [30], AHP and QFD for combining customers’ requirements and preferences [31], AHP and two multi-attribute decision-making methods for rating ‘‘whats’’ [32], AHP, multiattribute utility theory, and linear programming methods for rating ‘‘whats’’ [33], value functions to capture ‘‘whats’’ [34], AHP to study the ranking sensitivity of ‘‘whats’’ in QFD [35].

In this paper, we are using the F-extended AHP method which is described in the previous section for ranking the WHAT column of the HOQ (fig. 3). These are the flexibility levers in the HOW column. The experts have judged which flexibility levers impact which flexibility needs and to what degree. The physical relationships among the technical requirements are also specified but not given here. Finally, the table 3 indicates the importance weights of the flexibility levers.

4 Conclusions

In this paper, a decision making model, based on Analytic Hierarchy Process (AHP) and Quality Function Deployment (QFD) methodology, which determine the best flexibility levers’ portfolio to deal with consumers needs for flexibility in term of delivering a 5R product is developed. Thus, the need for flexibility is calculated with the F-extended AHP method. The experts have judged which flexibility levers impact which flexibility needs and to what degree.

The need for volume flexibility is highly important in the automotive sector. On the other hand, the right price is very important when we are faced to a mix product flexibility problem. In the same way, placing the right product in the market is extremely important when we have to satisfy a new product flexibility need.

Facing continuous change in the automotive industry, using the flexible human resources factors and being able to develop the flexible agreements with suppliers and customers are very important. On the other hand, an automotive firm must have a flexible technology to adapt any change and be highly flexible in his transport and shipping design.

For the future works, the study will develop a Fuzzy-QFD for better represent the vagueness of the experts’ opinions on the flexibility levers. In a next time, the three phases of QFD will be included for a global design strategy.

Table 3. The importance weight of the flexibility levers

Row Number	Flexibility Levers	Relative Weight (Relative Importance)
4	Labor Flexibility	10,14%
21	Flexibility in Development of Standing Agreements with Suppliers and Customers	10,04%
7	Process/Technology Flexibility	8,35%
9	Transport/Shipping Flexibility	6,71%
12	Financial Resources Flexibility	6,38%
6	Design/Development Flexibility	5,95%
11	Expansion Flexibility	5,56%
19	Flexibility in Enhancing Knowledge and Experience Base	5,32%
1	Planning/Scheduling Flexibility	4,61%
13	Organizational Structure Flexibility	4,18%
14	Decision Making Flexibility	4,18%
5	Machine/Equipment Flexibility	3,95%
20	Flexibility in Identifying Undiscovered and Unused Talents of Workforce	3,50%
10	Layout Flexibility	3,38%
15	Job Design Flexibility	3,35%
18	Reputation Building Flexibility	3,35%
8	Raw Materials Flexibility	2,36%
17	Managerial Perception Change Flexibility	2,36%
16	Employee's Willingness for Change Flexibility	2,21%
3	Routing Flexibility	2,12%
2	Sequencing Flexibility	2,01%

ISBN: 978-989-95079-6-8

Acknowledgment

This research has been financially supported by Galatasaray University Research Fund.

References

[1] Handfield, R.B. and Nichols, E.L. *Supply Chain Redesign: Transforming Supply Chains into Integrated Value Systems*, Prentice-Hall, Upper Saddle River, NJ. 2002.

[2] Mehta, S.N., Cisco fractures its own fairy tale, *Fortune*, May, p. 104, 2001.

[3] Monckza, R.M. and Morgan, J., Competitive supply strategies for the 21st century, *Purchasing*, 128(1): 48-59, 2000.

[4] Pagell, M., Understanding the factors that enable and inhibit the integration of operations, purchasing and logistics, *Journal of Operations Management*, 22(5): 459-87, 2004.

[5] Holweg, M., The three dimensions of responsiveness, *International Journal of Operations & Production Management*, 25(7/8): 603-22, 2005.

[6] Storey, J., Emberson, C. and Reade, D., The barriers to customer responsive supply chain management, *International Journal of Operations & Production Management*, 25(3): 242-60, 2005.

[7] Reichhart, A. and Holweg, M., Creating the customer-responsive supply chain: a reconciliation of concepts, *International Journal of Production Planning & Control*, 27(11): 1144-72, 2007.

[8] Swamidass, P., Manufacturing flexibility: operations management association, *Monograph no. 2, National Association of Manufacturers*, Washington, DC., 1988.

[9] Upton, D.M., The management of manufacturing flexibility, *California Management Review*, 36(2): 72-89, 1994.

[10] Hyun, J. and Ahn, B.H., A unifying framework for manufacturing flexibility, *Manufacturing Review*, 5(4): 251-60, 1992.

[11] Gerwin, D., Manufacturing flexibility: a strategic perspective, *Management Science*, 39(4): 395-410, 1993.

[12] Gerwin D., Manufacturing flexibility: a strategic perspective, *Management Science,* 39(4):395–410, 1993.

[13] Bengtsson J, Olhager J., The impact of the product mix on the value of flexibility, *Omega*, 30:265–73, 2002.

[14] Aggorval, S., Flexibility Management: the Ultimate Strategy, *Industrial Management,* Jan.-Fev., 1997.

[15] Chang, D. Y., Applications of the Extent Analysis Method on Fuzzy AHP, *European Journal of Operational Research,* 95: 649–655, 1996.

[16] Wang, Y.-M., Luo, Y., Hua, Z., On the Extent Analysis Method for Fuzzy AHP and its Applications*, European Journal of Operational Research,* 186: 735–747, 2008.

[17] Saaty, T. L., *The Analytic Hierarchy Process*, Mc. Graw-Hill, New York, 1980.

[18] Zadeh, L. A., Fuzzy sets, *Information and Control*, 8: 338-353, 1965.

[19] Sullivan, L.P., Quality function deployment, *Quality Progress*, 19(6): 39–50, 1986.

[20] Akao, Y., *Quality Function Deployment: Integrating Customer Requirements into Product Design*, Productivity Press 1990.

[21] Hauser, J.R., Clausing, D. The house of quality, *Harvard Business Review*, 63–73, 1988.

[22] Schmidt, R. The implementation of simultaneous engineering in the stage of product concept development: A process orientated improvement of quality function deployment, *European Journal of Operational Research* 100(2): 293–314, 1997.

[23] Fariborz, Y.P., Rafael, A.C., Quality function deployment for the good of soccer, *European Journal of Operational Research*, 137: 642–656, 2002.

[24] Brown, P.G., QFD: echoing the voice of the customer, *AT&T Technical Journal*, 18–32, 1991.

[25] Griffin A., Hauser, J., The voice of the customer, Technical report, Working paper 92-106, *Marketing Science Institute*, Cambridge, MA. 1992.

[26] Hauser, J.R., How Puritan7Bennett used the house of quality, *Sloan Management Review*, 34(3): 61–70, 1993.

[27] Armacost, R.T., Componation, P.J., Mullens, M.A. and Swart, W.W., An AHP framework for prioritizing custom requirements in QFD: an industrialized housing application, *IIE Transactions*, 26(4): 72–79, 1994.

[28] Chuang, P.T., Combining the analytic hierarchy process and quality function deployment for a location decision from a requirement perspective, *International Journal of Advanced Manufacturing Technology,* 18(11): 842–849, 2001.

[29] Madu, C.N., Kuei, C.H., Aheto, J., Winokur, D., Integrating total quality management in the adoption of new technologies, *Benchmarking: An International Journal*, 1(3): 52–66, 1994.

[30] Lu, M.H., Madu, C.N., Kuei, C. and Winokur, D., Integrating QFD, AHP and benchmarking in strategic marketing, *Journal of Business & Industrial Marketing*, 9(1): 41–50, 1994.

[31] Lin, B., Fite, D., Managing a sea of quality information at Ark-La-Tex Aquatics, *National Productivity Review* 15(1): 79–85, 1995.

[32] Chan, L.K., Wu, M.L., Prioritizing the technical measures in quality function deployment, *Quality Engineering*, 10(3): 467–479, 1998.

[33] Ho, E.S.S.A., Lai, Y.J. and Chang, S.I. An integrated group decision-making approach to quality function deployment, *IIE Transactions*, 31: 553–567, 1999.

[34] Askin, R.G., Dawson, D.W., Maximizing customer satisfaction by optimal specification of engineering characteristics, *IIE Transactions*, 32(1): 9–20, 2000.

[35] Xie, M., Goh, T.N., Wang, H., A study of the sensitivity of ‘‘customer voice’’ in QFD analysis, *International Journal of Industrial Engineering*, 5(4): 301–307, 1998.

ISBN: 978-989-95079-6-8

Continuous OWA operator and its calculation

Yasuo Narukawa[1] Vicenç Torra[2]

1.Toho Gakuen, 3-1-10, Naka, Kunitachi, Tokyo, 186-0004, Japan
1.The Institute for Research in Business Administration, Waseda University,
1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo,169-8050, Japan
2.IIIA - CSIC,
Campus UAB s/n, 08193 Bellaterra, Catalonia, Spain
Email: narukawa@d4.dion.ne.jp, vtorra@iiia.csic.es

Abstract*— *In this paper, we study the OWA operator on the real line, which corresponds to the Continuous OWA operator (COWA). After defining it, we introduce some properties and fundamental formulas for its computation. Among them, we give as an example, a differential equation for the COWA operator. **Keywords:** Fuzzy measures, Order weighted averaging operator, Choquet integral, Continuous OWA operator

1 Introduction

Aggregation operators [1, 3, 8] are used to combine information to obtain a datum of better quality. In recent years there is an increasing interest in these topics for their application in decision problems and artificial intelligence.

Among aggregation operators, one of the most well known and useful one is the Ordered Weighted Averaging operator (OWA) introduced by Yager [12, 13, 14]. The OWA is regarded as a Choquet integral with respect to a fuzzy measure[6, 9].

In classical statistics, when the amount of data is large, it is usual to approximate a discrete distribution, such as a binomial distribution, by a continuous distribution. A continuous distribution, such as the normal distribution, is a Lebesgue integral on the real line and is based on lots of results of classical integral theory. At present, there are a very few theoretical results about the Choquet integral on the real line. This paper is the first step for a theory of Choquet integral on the real line.

The structure of this paper is as follows. In Section 2 we review fuzzy measures and the OWA operator, and introduce a few results related to the OWA operator.

In Section 3, we define the OWA operator on the real line, called the Continuous OWA operator (COWA) operator and introduce some fundamental formulas and expressions for their computation.

In Section 4, we show an example of differential equation, which will be a hint to find a weighting function for the COWA operator. The paper finishes with some conclusions.

2 Preliminaries

In this section, we define fuzzy measures, the Choquet integral and the OWA operator, and show their basic properties.

To introduce both a discrete space and a non-discrete space in a unified way, we use the terms in general topology in this section. Let X be a locally compact Hausdorff space and $\mathcal{B}$ be a class of Borel sets, that is, the smallest $\sigma-$algebra which includes the class of all closed sets. We say that $(X, \mathcal{B})$ is a measurable space.

Example 1 *We consider two examples of Hausdorff spaces:*

(1) *The set of all real numbers R is a locally compact Hausdorff space. If $X = R$, $\mathcal{B}$ is the smallest $\sigma-$ algebra which includes the class of all closed intervals.*

(2) *Let $X := \{1, 2, \ldots, N\}$. X is a compact Hausdorff space with a discrete topology. Then we have $\mathcal{B} = 2^X$.*

Definition 1 [7] Let $(X, \mathcal{B})$ be a measurable space. A fuzzy measure (or a non-additive measure) μ is a real valued set function, $\mu : \mathcal{B} \longrightarrow [0, 1]$ with the following properties;

(1) $\mu(\emptyset) = 0$

(2) $\mu(A) \leq \mu(B)$ whenever $A \subset B$, $A, B \in \mathcal{B}$.

We say that the triplet $(X, \mathcal{B}, \mu)$ is a fuzzy measure space if μ is a fuzzy measure.

A fuzzy measure is said to be continuous if $A_n \uparrow A$ implies $\mu(A_n) \uparrow \mu(A)$ and $A_n \downarrow A$ implies $\mu(A_n) \downarrow \mu(A)$.

Definition 2 Let $(X, \mathcal{B})$ be a measurable space. A function $f : X \to R$ is said to be measurable if $\{x | f(x) \geq \alpha\} \in \mathcal{B}$ for all $\alpha \in R$.

Example 2 *Let f be a continuous function. Then, for all $\alpha \in R$, $\{f \geq \alpha\}$ is a closed set. Therefore, f is measurable.*

$\mathcal{F}(X)$ denotes the class of non-negative measurable functions, that is,

$$\mathcal{F}(X) = \{f | f : X \to R^+, f : \text{measurable}\}$$

Definition 3 [2, 5] Let $(X, \mathcal{B}, \mu)$ be a fuzzy measure space. The Choquet integral of $f \in \mathcal{F}(X)$ with respect to μ is defined by

$$(C) \int f d\mu = \int_0^\infty \mu_f(r) dr,$$

where $\mu_f(r) = \mu(\{x | f(x) \geq r\})$.

Let $A \subset X$. The Choquet integral restricted on A is defined by

$$(C) \int_A f d\mu := (C) \int f \cdot 1_A d\mu.$$

ISBN: 978-989-95079-6-8

Definition 4 *Let $D \subset R^N$. An aggregation operator Ag is a function $Ag : D \to R$ with the following properties;*

(1) *(Unanimity or idempotency)*

$Ag(a, \ldots, a) = a$ *if* $(a, \ldots, a) \in D$

(2) *(Monotonicity)*

If $a_i \leq b_i$ *for all* $i = 1, \ldots, n$, $\mathbf{a} = (a_1, \ldots, a_N), \mathbf{b} = (b_1, \ldots, b_N)$ $\mathbf{a}, \mathbf{b} \in D$, *then* $Ag(\mathbf{a}) \leq Ag(\mathbf{b})$.

Yager introduced the Ordered Weighted Averaging operator in [12].

Definition 5 *[12] Given a weighting vector* $\mathbf{w}$ *with weights* $(w_1, \ldots, w_N)$, *the Ordered Weighted Averaging operator is defined as follows:*

$$OWA_{\mathbf{w}}(\mathbf{a}) = \sum_{i=1}^{N} w_i a_{\sigma(i)}$$

where σ *defines a permutation of* $\{1, \ldots, N\}$ *such that* $a_{\sigma(i)} \geq a_{\sigma(i+1)}$, $\mathbf{a} = (a_1, \ldots, a_n)$.

A fuzzy measure μ on $\mathcal{B}$ is said to be symmetric [4] if $\mu(A) = \mu(B)$ for $|A| = |B|$, $A, B \in \mathcal{B}$. Symmetric fuzzy measures on $\{1, \ldots, N\}$ can be represented in terms of N weights w_i for $i = 1, \ldots, N$ so that $\mu(A) = \sum_{i=1}^{|A|} w_i$. Using a symmetric fuzzy measure, we can represent any OWA operator as a Choquet integral.

Proposition 6 *Let* $X := \{1, 2, \ldots, N\}$; *then, for every* $OWA_{\mathbf{w}}$, *there exists a symmetric fuzzy measure satisfying* $\mu(\{1\}) := w_1$ *and* $\mu(\{1, \ldots, i\}) := w_1 + \cdots + w_i$ *for* $i = 1, 2, \ldots, N$, *such that*

$$OWA_{\mathbf{w}}(\mathbf{a}) = (C) \int \mathbf{a} d\mu$$

for $\mathbf{a} \in R_+^N$.

3 Continuous OWA operator

In the following we consider aggregation operators on the real line. Let λ be a Lebesgue measure on $[0, 1]$, that is, $\lambda([a, b]) = b - a$ for $[a, b] \subset [0, 1]$.

Definition 7 *Let* $\mathcal{F}_b([0, 1])$ *be a class of bounded measurable function on* $[0, 1]$. *A continuous aggregation operator Ag on the real line is a functional* $Ag : \mathcal{F}_b([0, 1]) \to R$ *with the following properties;*

(1) *(Unanimity or idempotency)*

$Ag(a) = a$ *if* $a(x) = a$ *for all* $x \in [0, 1]$

(2) *(Monotonicity)*

If $a(x) \leq b(x)$ *for all* $x \in [0, 1]$, *then* $Ag(a) \leq Ag(b)$.

(3) *(Continuity) Let* a_n, $a \in \mathcal{F}_b([0, 1])$ *for* $n = 1, 2, 3, \ldots$ *and* $\lim_{n\to\infty} a_n = a$. *Then,* $\lim_{n\to\infty} Ag(a_n) = Ag(a)$.

Let $D \subset R^N$. For every $a := (a_1, \ldots, a_n) \in D$, we can define a function $f \in \mathcal{F}_b([0, 1])$ by $f(x) := a_k$ if $(k-1)/n \leq x < k/n$. Therefore the definition above is one of the generalization of aggregation operators on D.

Since the Choquet integral with respect to a continuous fuzzy measure satisfies all the conditions above, the Choquet integral with respect to a continuous fuzzy measure is a continuous aggregation operator.

Using the Choquet integral we can define the continuous OWA (COWA) operator.

Definition 8 *Let* μ *be a fuzzy measure on* $([0, 1], \mathcal{B})$. μ *is said to be symmetric, if* $\lambda(A) = \lambda(B)$ *implies* $\mu(A) = \mu(B)$.

Definition 9 *Let* $a \in \mathcal{F}_{\lfloor}([0, 1])$. *The continuous OWA operator is defined by*

$$COWA_{\mu}(\mathbf{a}) = (C) \int \mathbf{a} d\mu$$

where μ *is a symmetric fuzzy measure.*

Let μ be a symmetric fuzzy measure on $([0, 1], \mathcal{B})$. Suppose that $\lambda(A) < \lambda(B)$. Then there exists $B' \in \mathcal{B}$ such that $\lambda(B) = \lambda(B')$ and $A \subset B'$. Then we have

$$\mu(A) < \mu(B') = \mu(B).$$

Therefore we have the next proposition.

Proposition 10 *Let* μ *be a symmetric fuzzy measure on* $([0, 1], \mathcal{B})$. *Then there exists a monotone function* $\varphi : [0, 1] \to [0, 1]$ *such that* $\mu = \varphi \circ \lambda$.

It follows from the proposition above that we can consider the Choquet integral with respect to $\varphi \circ \lambda$ as the COWA operator. We will write $COWA_{\varphi}$ instead of $COWA_{\varphi\circ\lambda}$ and we will say that φ is the weight for the COWA operator.

Let $f : [0, 1] \to R$ be monotone increasing with $f(0) = 0$ and differentiable. We define the sequence of functions $\{f_k\}$ by $f_1 = f$, $f_{k+1} = \int_0^x f_k d\lambda$ for $x \in [0, 1]$, $k = 1, 2, \ldots$. Then we have

$$\begin{aligned}(C) \int_{[0,1]} f d\lambda^n &= \int_0^{\infty} \lambda^n (f \cdot 1_{[0,x] \geq \alpha}) d\alpha \\ &= \int_0^{f(x)} (x - f^{-1}(\alpha))^n d\alpha\end{aligned}$$

Let $t := x - f^{-1}(\alpha)$, Since we have $d\alpha = -f'(x - t)dt$, $t = x$ if $\alpha = 0$ and $t = 0$ if $\alpha = f(x)$, then

$$\begin{aligned}(C) \int_{[0,1]} f d\lambda^n &= \int_x^0 t^n \cdot (-f'(x - t)) dt \\ &= \int_0^x t^n f'(x - t) dt\end{aligned}$$

Next let $s := x - t$, we have

$$\begin{aligned}(C) \int_{[0,1]} f d\lambda^n &= \int_0^x (x - s)^n f'(s) ds \\ &= [(x - s)^n f(s)]_0^x + n \int_0^x (x - s)^{n-1} f_1(s) ds\end{aligned}$$

ISBN: 978-989-95079-6-8

Since $f(x) = 0$, we have

$$(C)\int_{[0,1]} f d\lambda^n = n\int_0^x (x-s)^{n-1} f_1(s) ds.$$

It follows from integration by part again that

$$\int_0^x (x-s)^{n-1} f_1(s) ds = (n-1)\int_0^x (x-s)^{n-2} f_2(s) ds.$$

Repeating above calculation, we have the next lemma.

Lemma 11 *Let $f : [0,1] \to R$ be monotone increasing with $f(0) = 0$ and differentiable. We define the sequence of functions $\{f_k\}$ by $f_1 = f$, $f_{k+1} = \int_0^x f_k d\lambda$ for $x \in [0,1]$, $k = 1, 2, \ldots$. Then we have*

$$(C)\int_{[0,x]} f d\lambda^n = n! f_n(x)$$

for $x \in [0,1]$.

Example 3 *Let $f(t) = t$, we have $f_1 = \frac{1}{2}x^2, \cdots, f_n = \frac{1}{(n+1)!}x^{n+1}$.*

$$(C)\int_{[0,x]} t d\lambda^n(t) = \frac{1}{n+1}x^{n+1}$$

for $x \in [0,1]$.

Let the weight w be ∞-order differentiable. Then we can express w by

$$w(x) := \sum_{k=1}^{\infty} a_k x^k.$$

Since the Choquet integral is linear with respect to the fuzzy measures. We have

$$(C)\int_{[0,x]} f dw \circ \lambda = \sum_{k=1}^{\infty} k! a_k f_k(x)$$

for $x \in [0,1]$. Therefore we have the next theorem.

Theorem 12 *Let $f : [0,1] \to R$ be monotone increasing with $f(0) = 0$ and differentiable. We define the sequence of functions $\{f_k\}$ by $f_1 = f$, $f_{k+1} = \int_0^x f_k d\lambda$ for $x \in [0,1]$, $k = 1, 2, \ldots$.*

$$COWA_{\mathbf{w}}(f) = \sum_{k=1}^{\infty} k! a_k f_k(1)$$

for $x \in [0,1]$.

4 Differential equation for COWA operator

In this section we consider the definition of a weight for the COWA. As we will see below, a differential equation helps in this definition.

Let us define a weight $w(x) := \dfrac{e^x - 1}{e - 1}$. Then we have

$$w(x) = \frac{1}{e-1}\sum_{k=1}^{\infty} \frac{1}{k!} x^k.$$

Since $a_k = \dfrac{1}{(e-1)k!}$, we have

$$(C)\int_{[0,x]} f dw \circ \lambda = \frac{1}{e-1}\sum_{k=1}^{\infty} k! f_k(x)$$

for $x \in [0,1]$.

Then we have the next differential equation for a weighting function.

$$\frac{d}{dx}(C)\int_{[0,x]} f dw \circ \lambda = f(x) + \frac{1}{e-1}\sum_{k=1}^{\infty} k! f_k(x)$$

for $x \in [0,1]$. Therefore we have the next differential equation.

$$\frac{d}{dx}(C)\int_{[0,x]} f dw \circ \lambda = f(x) + (C)\int_{[0,x]} f dw \circ \lambda$$

for $x \in [0,1]$.

Proposition 13 *Let f be a monotone increasing function with $f(0) = 0$ and let f be differentiable. If the weighting function satisfies the next equation:*

$$\frac{d}{dx}(C)\int_{[0,x]} f dw \circ \lambda = f(x) + (C)\int_{[0,x]} f dw \circ \lambda$$

for $x \in [0,1]$, then we have

$$w(x) = \frac{e^x - 1}{e - 1}$$

5 Conclusion

We defined the OWA operator on the real line (COWA operator), and show some fundamental formulas for its calculation. We give an example of differential equation to find a weighting function for COWA operator. This methods will be applicable to the Weighted OWA (WOWA) operator introduced by Torra [10, 11], which is one of the generalization of OWA operator. We expect to have some new results related to the WOWA operator in the near future.

Acknowledgements

Partial support by the Spanish MEC (projects ARES – CONSOLIDER INGENIO 2010 CSD2007-00004 – and eAEGIS – TSI2007-65406-C03-02) is acknowledged.

References

[1] Calvo, T., Mayor, G., Mesiar, R. (eds.) (2002) Aggregation Operators, Physica-Verlag.

[2] Choquet, G. (1955) Theory of capacities. *Ann. Inst. Fourier, Grenoble.* 5,131-295.

[3] Grabisch, M., Murofushi, T., Sugeno, M. (eds.) (2000) Fuzzy Measures and Integrals: Theory and Applications, Physica-Verlag.

[4] Miranda, P., Grabisch, M. (2002) p-symmetric fuzzy measures, Proc. of the IPMU 2002 Conference, 545-552, Annecy, France.

ISBN: 978-989-95079-6-8

[5] Murofushi, T., Sugeno, M. (1989) An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure, Fuzzy Sets and Systems 29 201-227.

[6] Ralescu, Anca L. , Ralescu, Dan A. (1997) Extensions of fuzzy aggregation, Fuzzy Sets and Systems, 86, 321-330.

[7] Sugeno, M. (1974) *Theory of fuzzy integrals and its applications,* Doctoral Thesis, Tokyo Institute of Technology.

[8] Torra, V., Narukawa, Y. (2007) Modeling decisions: information fusion and aggregation operators, Springer.

[9] Torra, V. Narukawa, Y. (2008) Choquet Stieltjes Integral, Losonczi's Means and OWA Operators. In V. Torra, Y. Narukawa (Eds.): 5th International Conference, MDAI 2008, Sabadell, Spain, October 30-31, 2008. Proceedings. Lecture Notes in Computer Science 5285 Springer, 62-73.

[10] Torra, V. (1996) Weighted OWA operators for synthesis of information, Fifth IEEE International Conference on Fuzzy Systems (IEEE-FUZZ'96) (ISBN 0-7803-3645-3), 966-971, New Orleans, USA.

[11] Torra, V. (1997) The weighted OWA operator, Int. J. of Intel. Syst. 12 153-166.

[12] Yager, R. R. (1988) On ordered weighted averaging aggregation operators in multi-criteria decision making, IEEE Trans. on Systems, Man and Cybernetics, 18 183-190.

[13] Yager, R. R., Filev, D. P. (1994) Parameterized and-like and or-like OWA operators, Int. J. of General Systems 22 297-316.

[14] Yager, R.R. (1993) Families of OWA operators, Fuzzy Sets and Systems, 59 125-148.

ISBN: 978-989-95079-6-8

A Process Algebra Approach to Fuzzy Reasoning

Liliana D'Errico Michele Loreti

Dipartimento di Sistemi e Informatica
Università di Firenze, Italy
Email: {liliana.derrico, michele.loreti}@unifi.it

***Abstract*— *Fuzzy systems address the imprecision of the input and output variables, which formally describe notions like "rather warm" or "pretty cold", while provide a behaviour that depends on fuzzy data. This class of systems are classically represented by means of Fuzzy Inference Systems (FIS), a computing framework based on the concepts of fuzzy if-then rules and fuzzy reasoning. Even if FIS are largely used, these lack in compositionality. Moreover, the analysis of modeled behaviuors needs complex analytic tools. In this paper we propose a process algebraic approach to specification and analysis of fuzzy behaviours. Indeed, we introduce a Fuzzy variant of CCS (*Calculus of Communicating Processes*), that permits compositionally describing fuzzy behaviours. Moreover, we also show how standard process algebra formal tools, like modal logics and behavioural equivalences, can be used for supporting fuzzy reasoning.*

***Keywords*—** Fuzzy Systems, Process Algebras, Compositional Fuzzy Reasoning

1 Introduction

Human perception of real world abounds with concepts without strictly defined constraints (examples are fat, very, more, slowly, old, etc). Such concepts can be described by means of Fuzzy Sets [1, 2]: classes of objects in which transit from membership to not membership gradually takes place.

Fuzzy Sets are widely used in control systems where the system behaviour can depend on data without precise values. In a *Fuzzy System* input values are described by means of variables that model data (temperature, speed) representing them as fuzzy sets, each of which identifies a different range of values (e.g. *cold*, *warm*, *hot*,...).

These systems are classically represented by means of Fuzzy Inference Systems (FIS) [3], a computing framework based on the concepts of fuzzy if-then rules, fuzzy set theory and fuzzy reasoning. A Fuzzy If-Then-Rule is a rule of the form *If* x *is* A *then* y *is* B where A and B are linguistic values defined by fuzzy sets on universes of discourse X and Y, respectively. Fuzzy reasoning is an inference procedure used to derive conclusions from a set of fuzzy If-Then-Rules and one or more conditions.

Defuzzification is the essential "process" to translate the fuzzy result in a crisp result. The most frequently used defuzzification strategy is the centroid of area defined as:

$$z_{coa} = \frac{\int_z \mu_{C'}(z)\, z\, dz}{\int_z \mu_{C'}(z)\, dz}$$

where $\mu_{C'}(z)$ is the aggregated output membership function. Other defuzzification strategies arise for specific applications: maximum membership, mean of maximum, largest of maximum, smallest of maximum, weighted average and so on. Generally speaking, these defuzzification methods are computation intensive Theoretical results are available [4, 5]. Even if Fuzzy Inference Systems are largely used, these lack in compositionality, in the sense that the rules interactions as well as how a rule interferes with the others is not completely clear. Moreover, it is difficult to compare different implementations of a system as well as verifying the properties satisfied by a given specification.

Process algebras are a set of mathematically rigorous languages with well defined semantics that permit modelling behaviour of concurrent and communicating systems. Verification of concurrent systems within the process algebraic approach can be performed by checking that processes enjoy properties described by some temporal logic's formulae.

In this paper we propose a process algebra approach to specification and analysis of fuzzy behaviours. Indeed, we introduce a Fuzzy variant of CCS (*Calculus of Communicating Processes*), that permits compositionally describing fuzzy behaviours. Operational semantics of Fuzzy CCS (FCCS) is described by means of *Fuzzy Labelled Transition Systems* [6] (FLTS). These are an extension of Labelled Transition Systems where fuzziness is used for modeling imprecision in concurrent systems.

Moreover, we also show how standard process algebra formal tools can be used for supporting fuzzy reasoning. Indeed, we define a fuzzy behavioural equivalence and a fuzzy modal logic that can be used for specifying and verifying properties of fuzzy systems.

The rest of the paper is organised as follows. In Section 2 we introduce Process Algebras conceptually. In Section 3 we recall the basic notions related to L-Fuzzy Sets while in Section 4 we present the Fuzzy CCS. In Section 5 we show how Fuzzy Hennessy-Milner Logic can be used for specifying and verifying properties of Fuzzy Systems. Finally, Section 6 concludes the paper.

2 Process Algebras

Process algebras are a set of mathematically rigorous languages with well defined semantics that permits describing and verifying properties of concurrent communicating systems. They can be seen as mathematical models of processes, regarded as agents that act and interact continuously with other similar agents and with their common environment. The agents may be real-word objects (even people), or they may be artefacts, embodied perhaps in computer hardware or software systems.

Process algebras provide a number of constructors for system description and are equipped with an operational semantics that describes systems evolution. Moreover, they often

ISBN: 978-989-95079-6-8

come equipped with observational mechanisms that permit identifying (through behavioural equivalences) those systems that cannot be taken apart by external observations. In some cases, process algebras have also complete axiomatizations, that capture the relevant identifications.

There has been a huge amount of research work on process algebras carried out during the last 25 years that started with the introduction of CCS [7], CSP [8] and ACP [9].

The main ingredients of a specific process algebra are:

- a minimal set of well thought operators capturing the relevant aspect of systems behaviour and the way systems are composed;
- a transition system associated with the algebra via structural *operational semantics* to describe the evolution of all systems that can be built from the operators;
- an equivalence notion that permits abstracting from irrelevant details of systems descriptions.

Verification of concurrent systems within the process algebraic approach is performed either by resorting to behavioural equivalences or by checking that processes enjoy properties described by some temporal logic's formulae. Equivalences are used for proving conformance of a process to specifications, expressed within the same process algebra notation. Verification with logical formulae is implemented by "model checking", an automatic method to prove properties verification.

In the former case two descriptions of a given system, one very detailed and close to the actual concurrent implementation, the other more abstract describing the abstract tree of relevant actions the system has to perform, are provided and tested for equivalence.

In the latter case, concurrent systems are specified as terms of a process description language while properties are specified as temporal logic formulae. Labelled Transition Systems are associated with terms via a set of structural operational semantics rules and model checking is used to determine whether the transition system associated with those terms enjoys the property specified by the given formulae.

Process algebras and modal logics have been largely used as tools for specifying and verifying properties of concurrent systems. This, also thanks to model checking algorithms, permits verifying whether a given specification satisfies the expected properties.

3 L-Fuzzy Sets

Human perception of real world abounds with concepts without strictly defined constraints (*fat*, *very*, *more*, *slowly*, *old*, etc...). Such concepts can be described by means of Fuzzy Sets: classes of objects in which transit from membership to not membership gradually takes place. A fuzzy set is a simple and intuitive generalization of the classical *crisp*[1] one.

Fuzzy sets are denoted by means of a generalised *membership function* that gives the *membership degree* of each element of the *universe*. This degree usually takes values in $[0, 1]$, the interval of real numbers from 0 to 1 inclusive. Although above range is the one most commonly used for representing membership degrees, any arbitrary set with some natural full or partial ordering can be used. Elements of this set are not required to be numbers as long as the ordering among them can be interpreted as representing various strengths of membership degree [10].

Let U be a universal set and L be a *complete lattice*, a *L-Fuzzy Set* [2, 11] A is denoted by a membership function $\mu_A : U \to L$.

Standard operations on sets and lattices L, like complement, intersection and union, can be generalised to L-Fuzzy Sets. These operations rely on the use of three function $c(\cdot)$, $i(\cdot,\cdot)$ and $u(\cdot,\cdot)$ that, respectively, give the measure of complement, intersection and union of fuzzy degrees.

The complement of a L-Fuzzy Set A, denoted by $\overline{A}$, is specified by a function $c : L \to L$ which assigns a value $\mu_{\overline{A}}(x) = c(\mu_A(x))$ to each membership degree $\mu_A(x)$. This assigned value is interpreted as the membership degree of the element x in the L-Fuzzy Set representing the negation of the concept represented by A.

Intersection and union of two L-Fuzzy Sets A and B are defined using functions $i : L \times L \to L$ and $u : L \times L \to L$. For each element a in the universal set U, these functions take as argument the pair consisting of the membership degrees of a in A and in B, respectively. Function i, also named *t-norm*, yields the membership degree of a in $A \cap B$, while u, also named *t-conorm*, returns the membership degree of a in $A \cup B$. Thus, $\mu_{A\cap B}(a) = i(\mu_A(a), \mu_B(a))$ while $\mu_{A\cup B}(a) = u(\mu_A(a), \mu_B(a))$.

Function c, i and u operating on L-Fuzzy Sets must be *continuous* on L and satisfy all axioms in Table 1, where 0 and 1 denote respectively the least and the greatest element in L while $\leq$ denotes the partial ordering on L. In the rest of this paper we will use $\mathcal{L}$ to denote a tuple $\langle L, c, i, u\rangle$ containing a complete lattice L together with its complement, intersection and union functions used for defining a family of L-Fuzzy Sets.

Table 1: Axioms of fuzzy operations

Axioms for $c(\cdot)$

$$c(1) = 0 \qquad c(0) = 1 \qquad \frac{x \leq y}{c(y) \leq c(x)} \qquad c(c(x)) = x$$

Axioms for $i(\cdot,\cdot)$

$$i(1,1) = 1 \qquad i(0,x) = 0 \qquad i(1,x) = x \qquad i(x,y) = i(y,x)$$

$$\frac{x_1 \leq x_2 \quad y_1 \leq y_2}{i(x_1,y_1) \leq i(x_2,y_2)} \qquad i(i(x,y),z) = i(x,i(y,z))$$

Axioms for $u(\cdot,\cdot)$

$$u(0,0) = 0 \qquad u(0,x) = x \qquad u(1,x) = 1 \qquad u(x,y) = u(y,x)$$

$$\frac{x_1 \leq x_2 \quad y_1 \leq y_2}{u(x_1,y_1) \leq u(x_2,y_2)} \qquad u(u(x,y),z) = u(x,u(y,z))$$

Fuzzy set theory was initially formulated by considering the

[1] Crisp set is defined to split individuals belonging to a certain universe into two groups: members (who surely belong to the set) and not members (who surely do not belong).

ISBN: 978-989-95079-6-8

complete lattice $[0,1]$, denoting the interval of real numbers from 0 to 1 (inclusive), and the following complement, intersection and union functions:

$$c(x) = 1 - x \quad i(x,y) = \min[x,y] \quad u(x,y) = \max[x,y]$$

It is easy to prove that these functions are continuous on $[0,1]$ and satisfy the axioms of Table 1.

4 A Fuzzy Process Algebra

Fuzzy systems are used for handling behaviours that depend on data without precise values that are described by means of variables taking values between 0 and 1. Variables, like temperature or speed, are represented by means of fuzzy sets each of which identifies a different range of values (e.g. *cold*, *warm*, *hot*,...).

As an example, we can consider a *Temperature Control System* (TCS). This is composed of an air conditioner (AC) system and a temperature sensor installed in the room. The TCS regulates the AC power, according to the values read from the sensor, to guarantee a suitable room temperature.

Following the fuzzy approach, room temperature can be modeled by considering different states identifying different range of values. For instance, *cold*, *warm* and *hot* like in figure 1.

Behaviour of TCS is described by a set of rules like: *if the temperature is warm then slightly increase the AC power*, *if the temperature is hot then increase the AC power*.

These systems are classically represented by means of Fuzzy Inference Systems (FIS).

In this section we present a fuzzy variant of CCS (*Calculus of Communicating Processes* [12]), named *Fuzzy CCS* (FCCS). In the new calculus the standard CCS actions are enriched with a fuzzy value modeling the enabling-degree. We aim at defining a formal language that permits compositionally describing fuzzy systems and that can be used for supporting fuzzy reasoning. This, also thanks to the use of standard formal tools like modal logics.

Like other process algebrae, FCCS provides a set of operators that permit describing the complete system starting from the specification of its subcomponents. Following the CCS approach, components interact with each other by means of *actions*, atomic and not interruptible steps, which represent input/output operations on communication ports or internal computations of the system. Let Λ be an infinite numerable set of labels or ports, not containing τ. A FCCS action can be: $a \in \Lambda$, the action to receive a signal on port a; $\bar{a}$ with $a \in \Lambda$, the action to deliver a signal on port a; τ, an internal computation step. We assume $\bar{\bar{\alpha}} \triangleq \alpha$, where $\alpha \in \Lambda \cup \{\bar{a} \mid a \in \Lambda\} \cup \{\tau\}$. Actions $\bar{\alpha}$ and α are said complementary, they represent input and output actions on the same channel.

The fundamental difference from CCS is the introduction of an attribute "x_i", that we define *action execution degree*. It is a fuzzy value able to represent, on a quality level, action behaviour. By this way CCS actions become fuzzy and therefore more representative of real world.

We define FCCS sintax by means of the following grammar:

$$\begin{aligned} Q &::= nil \mid X \mid \sum_{i \in I} (act_i,\, x_i).Q_i \\ P &::= Q \mid P_1 \mid P_2 \mid P\backslash A \mid P\,[f] \\ act &::= \bar{a} \mid a \mid \tau \end{aligned}$$

Now a brief description of operators:

- nil is the *inactive* process.
- X is the *process constant*, if $X \triangleq P$ then X denotes the invocation of process P. It is useful in defining recursive processes.
- $\sum_{i \in I}(act_i,\, x_i).Q_i$ is the *choice or sum* operator and denotes a choice among i possible behaviours that evolve with action act_i and degree x_i.
- $P_1 \mid P_2$ is the *parallel composition* operator and represents the concurrent execution of processes P_1 e P_2. If during the composition two complementary actions match, the resulting composed action is the internal one τ.
- $P\backslash A$ is the *restriction* operator. $A \in \Lambda$ and $P\backslash A$ behaves like P exception made for the impossibility to interact using actions in A.
- $P\,[f]$ is the *relabelling* operator. $f : \Lambda \to \Lambda$ allows relabelling process actions, to ease the description of complex processes.

4.1 Fuzzy Operational Semantics

Operational semantics of FCCS processes is defined in term of Fuzzy Labelled Transition Systems [6]. These generalize Labelled Transition Systems by defining the transition relation in term of a L-Fuzzy Set. This approach permits modeling situations like: "the transition takes place *rarely*" or "the transition occurs *frequently*" which may be distinguished and treated as a consequence.

Definition 4.1 ($\mathcal{L}$-FLTS) *Let $\mathcal{L} = \langle L, c, i, u\rangle$, a Fuzzy Labelled Transition System $\mathcal{F}$ for $\mathcal{L}$ ($\mathcal{L}$-FLTS) is a tuple $\langle Q, A, \chi_{\to}\rangle$ where:*

- *Q is a set whose elements are called states*
- *A is a finite set whose elements represent actions*
- *$\chi_{\to} : (Q \times A \times Q) \to L$ is the total membership function.*

We will write $q_0 \xrightarrow{\alpha}_{\varepsilon} q_1$ to denote that a transition from state q_0 to state q_1 by action α has a membership degree ε to the automaton.

In FLTS next states are selected nondeterministically. The membership degree associated to each transition is used to give a measure to computations. This measure is not *exact* as the probability one induced by PLTS [13], but can be used as the base for *approximate* reasoning in the spirit of Fuzzy Logic. The membership degree associated to a transition can be thought of as a value describing how much this transition is *enabled* in the FLTS.

Semantics for FCCS processes is defined by considering function $\mathcal{N}$ associating to each process P and transition α the fuzzy set $\mathcal{P}$ of processes reachable from P with α; we identify $\mathcal{P}$ with the set

$$\{Q : \varepsilon \mid Q \text{ is reachable from } P \text{ with action } \alpha \text{ and degree } \varepsilon\}$$

ISBN: 978-989-95079-6-8

Function $\mathcal{N}$ is formally defined in Table 2 where we use $\mathcal{P}|Q$ (resp. $Q|\mathcal{P}$) denotes the fuzzy set obtained by composing each element of $\mathcal{P}$ in parallel with Q. Similarly, $\mathcal{P}|\mathcal{Q}$ is the fuzzy set containing the parallel composition of each P in $\mathcal{P}$ with each Q in $\mathcal{Q}$, where the membership degree of $P|Q$ in $\mathcal{P}|\mathcal{Q}$ is defined as $i\left(\mu_{\mathcal{P}}(P), \mu_{\mathcal{Q}}(Q)\right)$.

Table 2: $\mathcal{N}$ext Function

$$\mathcal{N}\left(\langle\alpha,\varepsilon\rangle.P,\beta\right) = \begin{cases} \{P:\varepsilon\} & if\ \alpha=\beta \\ \emptyset & else \end{cases}$$

$$\mathcal{N}\left(P+Q,\alpha\right) = \mathcal{N}\left(P,\alpha\right) \cup \mathcal{N}\left(Q,\alpha\right)$$

$$\mathcal{N}\left(P\backslash L,\alpha\right) = \begin{cases} \mathcal{N}\left(P,\alpha\right)\backslash L & if \alpha \notin L \\ \emptyset & else \end{cases}$$

$$\mathcal{N}\left(P\left[f\right],\alpha\right) = \bigcup_{\beta:f(\beta)=\alpha} \mathcal{N}\left(P,\beta\right)\left[f\right]$$

$$\mathcal{N}\left(P|Q,\alpha\right) = \begin{cases} \left[\mathcal{N}\left(P,\alpha\right)|Q\right] \cup \left[P|\mathcal{N}\left(Q,\alpha\right)\right] & if \alpha \neq \tau \\ \left[\mathcal{N}\left(P,\tau\right)|Q\right] \cup \left[P|\mathcal{N}\left(Q,\tau\right)\right] \cup \left[\bigcup_{\alpha\in\Lambda}\left(\mathcal{N}\left(P,\alpha\right)|\mathcal{N}\left(Q,\bar{\alpha}\right)\right)\right] \cup \left[\bigcup_{\alpha\in\Lambda}\left(\mathcal{N}\left(P,\bar{\alpha}\right)|\mathcal{N}\left(Q,\alpha\right)\right)\right] & if \alpha = \tau \end{cases}$$

$$\mathcal{N}\left(A,\alpha\right) = \mathcal{N}\left(P,\alpha\right) \qquad if\ A \triangleq P$$

We say that P can evolve to Q (written $P \succ Q$) if and only if there exists an action α such that if $\mathcal{P} = \mathcal{N}(P,\alpha)$, $\mathcal{P}(Q) \neq 0$. We use $\succ^*$ to denote the reflexive and transitive closure of $\succ$.

Let P be a FCCS process, we denote with $FLTS(P) = \langle S, \Lambda, \chi_{\rightarrow}\rangle$ the FLTS such that:

- $S = \{Q \mid P \succ^* Q\}$;
- $\chi_{\rightarrow}(P,\alpha,Q) = \mathcal{P}(Q)$, where $\mathcal{P} = \mathcal{N}(P,\alpha)$.

Standard behavioural equivalences, like for instance bisimulation, can be easily generalized in order to take into account fuzziness. This kind of equivalences are useful when one aims at comparing different specifications. Definition of Fuzzy Bisimulation is straightforward and it is somehow reminiscent of Stochastic Bisimulation [14].

Definition 4.2 (Fuzzy Bisimulation) *Let $\langle S, A, \chi_{\rightarrow}\rangle$ a Fuzzy LTS. An equivalence relation $R \subseteq S \times S$ is a* fuzzy bisimulation *if and only if for each p and q in S, for each equivalent class C of R in S, and for each transition label α:*

$$\chi_{\rightarrow}(p,\alpha,C) = \chi_{\rightarrow}(q,\alpha,C)$$

where:

$$\chi_{\rightarrow}(p,\alpha,C) = \bigvee_{p'\in C} \chi_{\rightarrow}(p,\alpha,p')$$

Definition 4.3 (Fuzzy Bisimilarity) *Let $\langle S, A, \chi_{\rightarrow}\rangle$ be a Fuzzy LTS. We say that $p, q \in Q$ are bisimilar ($p \sim_F q$), if there exists a fuzzy bisimulation R such that pRq.*

Relation $\sim_F$ is a *fuzzy bisimilarity* and can be defined as the largest fuzzy bisimulation, namely:

$$\sim_F \triangleq \bigcup \{R \mid R \quad \text{is a fuzzy bisimulation}\}$$

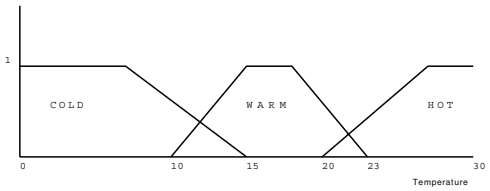


Figure 1: Temperature Fuzzy sets: COLD, NORMAL, HOT

Modeling Fuzzy Systems with FCCS We can now use FCCS for modeling the TCS described in section 4, where we consider three fuzzy sets for describing possible values of the room temperature. These sets are represented in Figure 1. Systems TCS is modeled in FCCS by splitting it into four subcomponents representing behaviours of the system. They "mime" TCS by interacting among each of them. Interactions are the result of synchronization between an input and an output action with the same name. The fuzzy value of the resulting action is calculated following semantics in Tab. 2. Process SYS_{st} is defined as follows:

$$SYS_{st} \triangleq (SENS_{st}|H|N|C)\backslash\{hot, warm, cold, inc, dec, noop\}$$

where st is the starting temperature while process $SENS_t$, which models the behaviour of temperature sensor, is defined as follows:

$$\begin{aligned} SENS_t \triangleq\ & \left\langle \overline{hot}, \mu_{hot}(t)\right\rangle.AC_t \\ & + \left\langle \overline{warm}, \mu_{warm}(t)\right\rangle.AC_t \\ & + \left\langle \overline{cold}, \mu_{cold}(t)\right\rangle.AC_t \\ & + \left\langle temp_t, 1\right\rangle.SENS_t \end{aligned}$$

$$\begin{aligned} AC_t \triangleq\ & \left\langle inc, 1\right\rangle.SENS_{t+1} \\ & + \left\langle dec, 1\right\rangle.SENS_{t-1} \\ & + \left\langle noop, 1\right\rangle.SENS_t \end{aligned}$$

Notice that in the processes above, non-determism is used for modeling the unpredictable changes in room temperature.

Controller behaviour is rendered by means of processes H, N and C that modify the AC power. These processes are defined as follows:

$$\begin{aligned} H &\triangleq \left\langle hot, 1\right\rangle.\left\langle \overline{inc}, 1\right\rangle.H \\ N &\triangleq \left\langle warm, 1\right\rangle.\left\langle \overline{noop}, 1\right\rangle.N \\ C &\triangleq \left\langle cold, 1\right\rangle.\left\langle \overline{dec}, 1\right\rangle.C \end{aligned}$$

For this system one could be interested on verifying that for each fixed starting temperature, the system always is able to reach a state where the room temperature is in a given range. In the next section, we will introduce a modal logic that will permit specifying and verifying properties of fuzzy systems.

An alternative description of this system could be obtained by considering process SYS^2_{st} defined as follows:

$$SYS^2_{st} \triangleq (SENS_{st}|CONT)\backslash\{hot, warm, cold\ inc, dec, noop\}$$

ISBN: 978-989-95079-6-8

where, differently from the previous implementation, the controller is obtained as a single process $CONT$ that nondeterministically can behave like H, N and C defined above:

$$CONT \triangleq (H + N + C).CONT$$

It easy to prove that SYS_{st} and SYS^2_{st} provide the same behaviour. Namely, that $SYS_{st} \sim_F SYS^2_{st}$.

5 Fuzzy Hennessy-Milner Logic

Fuzzy Hennessy-Milner Logic (FHML) [6] is an extension of HML, which aims at specifying properties of concurrent systems whose behavior is detailed by means of FLTS. Let $\mathcal{L} = \langle L, c, i, u\rangle$, $\Phi_{\mathcal{L}}$ be the set of formulas φ defined by the following syntax:

$$\varphi ::= tt \mid \neg\varphi \mid \varphi \bowtie \varepsilon \mid \varphi_1 \wedge \varphi_2 \mid \langle\alpha\rangle\varphi \mid X \mid \nu X.\varphi$$

where $\varepsilon \in L$.

Table 3: Formulae semantics

$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![tt]\!]\delta(p) = 1$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\neg\varphi]\!]\delta(p) = \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(p)$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_1 \wedge \varphi_2]\!]\delta(p) = i\left(\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_1]\!]\delta(p), \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_2]\!]\delta(p)\right)$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi \bowtie \varepsilon]\!]\delta(p) = \begin{cases} 1 & if\ \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(p) \bowtie \varepsilon \\ 0 & else \end{cases}$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\langle\alpha\rangle\varphi]\!]\delta(p) = u_{q\in Q}\left(i\left(\chi_{\rightarrow}(p,\alpha,q), \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(q)\right)\right)$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![X]\!]\delta = \delta(X)$$
$$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\nu X.\varphi]\!]\delta = \cup\left\{\chi \mid \chi \leq \mathfrak{F}^{\delta,\varphi}_X(\chi)\right\}$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![tt]\!]\delta(p) = 0$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\neg\varphi]\!]\delta(p) = \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(p)$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi_1 \wedge \varphi_2]\!]\delta(p) = u\left(\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi_1]\!]\delta(p), \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi_2]\!]\delta(p)\right)$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi \bowtie \varepsilon]\!]\delta(p) = \begin{cases} 1 & if\ \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(p) \not\bowtie \varepsilon \\ 0 & else \end{cases}$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\langle\alpha\rangle\varphi]\!]\delta(p) = $$
$$= i_{q\in Q}\left(u\left(i\left(\chi_{\rightarrow}(p,\alpha,q), \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(q)\right), c\left(\chi_{\rightarrow}(p,\alpha,q)\right)\right)\right)$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![X]\!]\delta = \delta(X)$$
$$\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\nu X.\varphi]\!]\delta = \cap\left\{\chi \mid \chi \geq^{\frown} \mathfrak{F}^{\delta,\varphi}_X(\chi)\right\}$$

where

$$\mathfrak{F}^{\delta,\varphi}_X = \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta[\chi/X] \qquad {}^{\frown}\mathfrak{F}^{\delta,\varphi}_X = \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta[\chi/X]$$

Operators are usual logical tt (true), $\neg$ (not), $\wedge$ (and). Moreover $\nu X.\varphi$ is Tarski's fixed point and $\langle\alpha\rangle\varphi$ the modal operator representing the property of evolving with action α to a state that behaves like φ. FHML extends HML by considering new operator $\varphi \bowtie \varepsilon$ ($\bowtie \in \{<, >\}$) that states about the satisfaction degree of a given formula in a given state. Such operator makes possible to describe the satisfaction degree of a formula in terms of an upper or a lower bound[2].

[2]This is somehow reminiscent of operator $[\varphi]_p$ proposed by Parma and Segala [15]

Other operators can be defined as macros in FHML. In the rest of this paper we let: $ff = \neg tt$, $\varphi_1 \vee \varphi_2 = \neg(\neg\varphi_1 \wedge \neg\varphi_2)$, $[\alpha]\varphi = \neg\langle\alpha\rangle\neg\varphi$ and $\mu X.\varphi = \neg\nu X\neg\varphi[\neg X/X]$.

Semantics of FHML is defined in term of functions $\mathcal{M}_{\mathcal{L},\mathcal{F}}$ and $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}$ that for each formula φ yield the L-Fuzzy Set that gives the measure, respectively, of satisfaction and unsatisfaction of φ. This approach permits defining semantics of FHML in a general way without considering any special constraint on the underlying L-Fuzzy Sets. Indeed, in general, standard properties of sets do not hold in the case of L-Fuzzy Sets. For instance, the intersection between a L-fuzzy set and its complement could be not empty. Notice that, in general, $\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\neg\varphi]\!] \neq c\left(\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\right)$.

To give a semantics to recursive formulae, interpretation functions $\mathcal{M}_{\mathcal{L},\mathcal{F}}$ and $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}$ take also, as a parameter, a function $\delta : Q \rightarrow L$, that associates to each logical variable X a L-fuzzy set.

We assume the usual definitions on positive and negative variables in [6] for guaranteeing well-definedness of interpretation formulae.

Definition 5.1 (Well formed formula) *A formula is well formed if in each subformula of the form $\nu X.\varphi$, variable X occurs positive in φ.*

Interpretation functions are parameterized with respect to $\mathcal{L}$, used for defining the underlying L-Fuzzy Set and the relative operations, and with respect to the $\mathcal{L}$-FLTS $\mathcal{F}$ used for interpreting formulae. Note that, interpretation of FHML coincides with the standard interpretation of HML when considering standard Boolean lattices.

Definition 5.2 (Formulae Semantics) *Let $\mathcal{L} = \langle L, c, i, u\rangle$ and $\mathcal{F}_{\mathcal{L}} = \langle Q, A, \chi_{\rightarrow}\rangle$. Functions $\mathcal{M}_{\mathcal{L},\mathcal{F}} : \Phi_{\mathcal{L}} \rightarrow Q \rightarrow L$ and $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}} : \Phi_{\mathcal{L}} \rightarrow Q \rightarrow L$ are inductively defined in Table 3.*

$\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!]\delta(q)$ denotes the satisfaction degree of formula φ by q. Formulae tt and ff are satisfied by every state with degree 1 and 0 respectively. A state p satisfies $\neg\varphi$ with degree ε if and only if p *does not satisfy* φ with degree ε. Fuzzy set $\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_1 \wedge \varphi_2]\!]$ is defined as the intersection between $\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_1]\!]$ and $\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi_2]\!]$. If a state p satisfies φ with a degree that is $<$ (resp. $>$) of ε then p satisfies $\varphi < \varepsilon$ (resp. $\varphi > \varepsilon$) with degree 1 (resp. 0). $\mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\langle\alpha\rangle\varphi]\!]\delta(p)$, which gives a measure of how p can reach with a transition labelled α a state satisfying φ, is defined as the disjunction, for each q, of $i(\chi_{\rightarrow}(p,\alpha,q), \mathcal{M}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!](q))$. Where $\chi_{\rightarrow}(p,\alpha,q)$ returns the membership degree of the transition from p to q with label α to the behaviour of the system. Finally, interpretation of $\nu X.\varphi$ is defined as greates fixed point of the interpretation of φ ($\mathfrak{F}^{\delta,\varphi}_X$), which is monotone in the complete lattice of Q L-Fuzzy Sets. This, thanks to the Tarski's fixed point theorem [16], guarantees the well-definedness of functions $\mathcal{M}_{\mathcal{L},\mathcal{F}}$ and $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}$.

The definition of function $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}$ is similar and straightforward. However, more attention has to be paid for $\mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\langle\alpha\rangle\varphi]\!]$. Each state q, contributes to the unsatisfaction of $\langle\alpha\rangle\varphi$ by p as a factor that depends on the degree of the transition α from p to q and on the unsatisfaction degree of formula φ by q. This value is obtained as a disjunction of: $c\left(\chi_{\rightarrow}(p,\alpha,q)\right)$ and $i\left(\chi_{\rightarrow}(p,\alpha,q), \mathcal{M}^{\frown}_{\mathcal{L},\mathcal{F}}[\![\varphi]\!](q)\right)$. The former indicates how much the transition from p to q with label

ISBN: 978-989-95079-6-8

α *does not belong* to the behaviour of the system. The latter, gives the measure of the unsatisfaction of φ when the action is executed.

Example 5.1 *FHML can be used for specifying that system process SYS_{st} can always reach a configuration where the room temperature is between* 18 *and* 20 *degrees. The following formula states that a configuration where temperature is between* 18 *and* 20 *degrees is eventually reached:*

$$\varphi = \mu X.\langle\overline{temp}_{18}\rangle tt \vee \langle\overline{temp}_{19}\rangle tt \vee \langle\overline{temp}_{20}\rangle tt \vee \langle\tau\rangle X$$

While formula:

$$\nu Y.\varphi \wedge [\tau]Y$$

states that φ is always satisfied.

Logical characterization of Fuzzy Bisimulation Formulae satisfaction induces an equivalence on the interpretation model and two states in a Fuzzy LTS are equivalent if (and only if) they satisfy the same set of formulae. A classical result relating modal logic and behavioural equivalence is the one in [12] showing that the equivalence induced by HML coincides with the bisimulation equivalence. A similar result can be proved when one considers FHML and Fuzzy Bisimulation. However, to prove this correspondence, one has to guarantee that the considered Fuzzy LTS is *finite-branching*.

Definition 5.3 (Finite-branching) *A Fuzzy LTS $\mathcal{F} = \langle S, A, \chi_{\rightarrow}\rangle$ is finite-branching if and only if $\forall p \in S$, $\forall \alpha \in A$ and $\forall \varepsilon \in \chi_{\rightarrow}$, the set $\left\{q' \in Q | q \xrightarrow{\alpha}_{\varepsilon} q'\right\}$ is finite.*

Theorem 1 *Let $\mathcal{F} = \langle S, A, \chi_{\rightarrow}\rangle$ be finite-branching. For each $p, q \in S$,*

$$p \sim_F q \quad \Leftrightarrow \quad \forall \varphi.\, \mathcal{M}[\![\varphi]\!](p) = \varepsilon \Leftrightarrow \mathcal{M}[\![\varphi]\!](q) = \varepsilon \text{ and } \mathcal{M}^{\smile}[\![\varphi]\!](p) = \varepsilon \Leftrightarrow \mathcal{M}^{\smile}[\![\varphi]\!](p) = \varepsilon$$

Due to lack of space, the proof is not reported.

6 Conclusions and Future Works

In this paper we have presented FCCS, a Fuzzy variant of CCS (*Calculus of Communicating Processes*), that aims at compositionally describing fuzzy behaviours. Operational semantics of FCCS has been defined by means of *Fuzzy Labelled Transition Systems* [6] (FLTS). These are an extension of Labelled Transition Systems where fuzziness is used for modeling uncertainty in concurrent systems. In the paper we have also shown how standard process algebras formal tools, like modal logics and behavioural equivalences, can be used for supporting fuzzy reasoning.

The idea to combine process algebrae and fuzzy sets is not new. For instance, in [17] is introduced a model based on interval values, to solve the problem of nondeterministic choices arising in concurrency and communication systems. However, differently from previous approaches, the present work introduces both a behavioural equivalence and a modal logic for reasoning about fuzzy specifications.

As a future work we plan to use the proposed approach for modeling the examples proposed in literature [18, 19, 20] where Fuzzy Theory is used for modeling the behaviour of "systems", like for instance those involving human interactions, where imprecision is a central feature.

References

[1] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] G.J. Klir and T.A. Folger. *Fuzzy sets, uncertainty, and information*. Prentice-Hall International, Inc, 1988.

[3] Jyh-Shing Roger Jang and Chuen-Tsai Sun. Neuro-fuzzy modeling and control. *Proceedings of the IEEE*, 83(3), 1995.

[4] Qiang Song and Robert P. Leland. Adaptive learning defuzzification techniques and applications. *Fuzzy Sets and Systems*, 81(3):321 – 329, 1996.

[5] Werner Van Leekwijck and Etienne E. Kerre. Defuzzification: criteria and classification. *Fuzzy Sets and Systems*, 108(2):159 – 178, 1999.

[6] Liliana D'Errico and Michele Loreti. Modeling fuzzy behaviours in concurrent systems. In L. Laura G. Italiano, E. Moggi, editor, *Proceedings of the 10th Italian Conference on Theoretical Computer Science ICTCS'07*, pages 94–105, 2007.

[7] R. Milner. *Communication and concurrency*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989.

[8] S. D. Brookes, C. A. R. Hoare, and A. W. Roscoe. A theory of communicating sequential processes. *J. ACM*, 31(3):560–599, 1984.

[9] J.A. Bergstra and J.W. Klop. Algebra of communicating processes with abstraction. *Theoretical Computer Science*, (37):77–121, 1985.

[10] L. A. Zadeh. Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, 4(2):103–111, 1996.

[11] J. A. Goguen. L-fuzzy sets. *J. Math. Anal. Appl.*, 18:145–174, 1967.

[12] Matthew Hennessy and Robin Milner. Algebraic laws for nondeterminism and concurrency. *J. ACM*, 32(1):137–161, 1985.

[13] Roberto Segala and Nancy Lynch. Probabilistic simulations for probabilistic processes. *Nordic J. of Computing*, 2(2):250–273, 1995.

[14] Ed Brinksma and Holger Hermanns. Process algebra and markov chains. In Ed Brinksma, Holger Hermanns, and Joost-Pieter Katoen, editors, *Euro Summer School on Trends in Computer Science*, volume 2090 of *Lecture Notes in Computer Science*, pages 183–231. Springer, 2001.

[15] A. Parma and R. Segala. Logical characterizations of bisimulations for discrete probabilistic systems. In Springer-Verlag, editor, *FOSSACS 2007, LNCS*, volume 4423, pages 287–301, 2007.

[16] A. Tarski. A lattice-teoretichal fixpoint theorem and its applications. *Pacific J. Math.*, 5:285–309, 1955.

[17] Yixiang Chen and Jie Zhou. Fuzzy interval-valued processes algebra. In S.Barry Cooper, Benedict Lowe, and Leen Torenvliet, editors, *CiE 2005 Abstracts*, ILLC Publications, pages 44–57, 2005.

[18] L. A. Zadeh. Commonsense reasoning based on fuzzy logic. In *WSC '86: Proceedings of the 18th conference on Winter simulation*, pages 445–447, New York, NY, USA, 1986. ACM Press.

[19] Eviatar Tron and Michael Margaliot. Mathematical modeling of observed natural behavior: a fuzzy logic approach. *Fuzzy Sets and Systems*, 2004.

[20] Bernd Möller, Wolfgang Graf, and Song Ha Nguyen. Modeling the life cycle of a structure using fuzzy processes. *Computer-Aided Civil and Infrastructure Engineering*, 2004.

ISBN: 978-989-95079-6-8

A Generalized Numerical Solution for Fuzzy Relation Equations

Iman Mohammadi Ardehali[1] Milad Avazbeigi[2]

1 Faculty of Management, University of Tehran
Tehran, Iran
2 Department of Industrial Engineering, Amirkabir University of Technology
Tehran, Iran
Email: iman@hamgerayesh.com, miladavazbeigi@aut.ac.ir

Abstract— *In this paper, line search based on Sequential Quadratic Programming is implemented in order to find a solution to Fuzzy Relation Equations. Sequential Quadratic Programming is a gradient-based method that uses a quadratic estimation of the objective function in each iteration's neighborhood. Unlike analytical approaches, the method can handle equations with any combinations of t-norms and t-conorms and at any dimensions. It is assumed that the FRE problem has at least one solution.*

Keywords— Fuzzy Relation Equation (FRE), Fuzzy Triangular Norms, Line Search, Numerical Solution, Sequential Quadratic Programming (SQP).

1 Introduction

Fuzzy Relation Equations were first introduced by Sanchez [1] in 1976. Most of the works done on fuzzy relation equations are focused on specific non-parametric t-norms (min and product) and t-conorms (max and probabilistic sum). Markovskii [2], Shieh [3-5], Molai and Khorram [6], Hana et al. [7], and Perfilieva and Nosková [8] have presented some algorithms for FREs with such compositions. In special cases, an exact solution can be obtained through direct analytical methods. However, for parametric operators like *Schweizer-Sklar*, *Frank*, *Yager*, *Sugeno-Weber*, *Dombi*, and *Dubois-Prade* there is no direct solution available. This is mostly because of the complex non-linear nature of these operators.

An FRE can be formed as:

$$P \circ Q = R \tag{1}$$

where $\circ$ denotes the S-T composition. The problem involves finding P from Q and R.

Stamou and Tzafestas [9] developed a criterion for the existence of minimal solutions, and interpreted a *Fuzzy Inference System* (FIS) as an FRE, indicating that FREs are appropriate for FIS implementation. Shieh [3, 4] recently extended the discussion to a general condition using continuous t-norms. In such cases, when the solution set for (1) is not empty, it can be completely determined by a unique maximum solution and a finite number of minimal solutions.

However, the current models mostly discuss analytical approaches to solve the FRE problem in case-specific conditions—such as max-min or max-product compositions and sizes of $1\times m$ and $1\times n$ for binary fuzzy relations P and R.

In this paper, we propose a new numerical method based on *Sequential Quadratic Programming* (SQP) line search. The proposed model's advantages are solving FREs for any set of parametric operators and capability of handling sizes of $p\times m$ and $p\times n$ for P and R.

It should be noted that the fundamental assumption of this effort is that there exists at least one solution to the FRE problem.

The rest of this paper is organized as follows. In section 2, basic concepts of Fuzzy Relation Equations and Sequential Quadratic Programming are presented. Section 3 describes the method in detail. In section 4, some examples are provided for demonstration of the model's performance. Finally, section 5 gives the conclusions and suggests ideas for future research.

2 Basic Concepts

2.1 Parametric Fuzzy Operators

The most familiar fuzzy operators are max and algebraic sum (as t-conorms or s-norms) and min and product (as t-norms). But they are not limited to these special cases, and can be parametrically constructed.

The advantages of using parametric operators are:

- Most of the parametric operators have the capability to reduce to the aforementioned non-parametric operators, with specific values for the parameters. For example, the Frank t-norm reduces to min when $p = 0$.
- The parameters can be tuned according to the specific requirements of the system in question.

On the other hand, the disadvantage is that the increasing complexity leads to longer running times and more sensitivity to the tuning mechanism.

2.2 Fuzzy Relation Equations

An FRE is as presented in (1), in which $\circ$ acts as the S-T composition, meaning:

$$\bigvee_{j=1}^{m}\left[T\left(p_{ij}, q_{jk}\right)\right] = r_{ik} \tag{2}$$

where P, Q, and R are fuzzy binary relations and p, q, and r are their elements respectively [10]. The Yager series of t-norms and t-conorms are well-known fuzzy operators:

$$T_w(a,b) = 1 - \min\left\{1, \left((1-a)^w + (1-b)^w\right)^{1/w}\right\} \tag{3}$$

$$S_w(a,b) = \min\left\{1, \left(a^w + b^w\right)^{1/w}\right\} \tag{4}$$

ISBN: 978-989-95079-6-8

Equation (3) represents the t-norm and (4) is the t-conorm of Yager operators.

2.3 Sequential Quadratic Programming

SQP methods represent the state of the art in nonlinear programming methods. Schittkowski [11], for example, has implemented and tested a version that outperforms every other tested method in terms of efficiency, accuracy, and percentage of successful solutions, over a large number of test problems.

The method closely mimics Newton's method for constrained optimization just as is done for unconstrained optimization. At each major iteration, an approximation is made of the Hessian of the Lagrangian function using a quasi-Newton updating method. This is then used to generate a Quadratic Programming sub-problem whose solution is used to form a search direction for a line search procedure.

Consider the following general non-linear minimization problem (*P*):

$$Minimize\ f(x)$$

$$subject\ to \begin{cases} h_i(x) = 0; & i = 1,...,l \\ h_j(x) \le 0; & j = l+1,...,m \end{cases}$$

where $x \in R^n, f: R^n \to R, h: R^n \to R^m$.

At *k*th iteration, the SQP algorithm generates a search direction (*d*) solving the definite quadratic sub-problem (*Q*) below:

$$Minimize\ \nabla f_k^T d + \tfrac{1}{2} d^T B_k d$$

$$subject\ to \begin{cases} h_i(x_k) + \nabla h_i^T(x_k) d = 0; & i = 1,...,l \\ h_i(x_k) + \nabla h_i^T(x_k) d \le 0; & j = l+1,...,m \end{cases}$$

where B_k is a positive definite approximation to the Hessian matrix of the Lagrangian function of the original problem (*P*):

$$L(x,\mu) = f(x) + \sum_{i=1}^{m} \mu_i h_i(x) \tag{5}$$

At each major iteration a positive definite quasi-Newton approximation of the Hessian of the Lagrangian function, *L*, is calculated using Broyden–Fletcher–Goldfarb–Shanno (BFGS) method [12].

A detailed description of the SQP method for nonlinear optimization can be found in Fletcher [13], Gill et al. [14], Powell [15], Hock and Schittkowski [16], Nocedal and Wright [17], and Bartholomew-Biggs [18].

3 The Method

SQP, as any other line search method, requires an initial solution to start from. A fairly large random search of the solution space can provide us with this initial solution. Random solutions are generated as matrices of *p*×*m*—dimensions of *P* in (1)—with elements uniformly distributed between 0 and 1.

There are various line search methods for solving nonlinear problems, the most prominent of which are gradient descent, the Newton method, quasi-Newton methods, and Sequential Quadratic Programming. These four were thoroughly tested for our problem and SQP turned out to be the most superior in terms of running time and error measure. So SQP is the line search algorithm of choice for our method.

Since our goal is to minimize the final error, the objective function would rationally be *RMSE*, that is, the root-mean-square of the errors obtained by the difference between the actual *R* and the one calculated from the composition below.

$$R' = P' \circ Q \tag{6}$$

where *P′* is the solution in search of *P*. Thus, the objective function will be as follows.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{p}\sum_{k=1}^{n}(r_{ik} - r'_{ik})^2}{p \times n}} \tag{7}$$

Using (2), we arrive at this formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{p}\sum_{k=1}^{n}\left\{\bigcirc\!\!\!\!\!\!\sum_{j=1}^{m}\left[T(p_{ij}, q_{jk})\right] - \bigcirc\!\!\!\!\!\!\sum_{j=1}^{m}\left[T(p'_{ij}, q_{jk})\right]\right\}^2}{p \times n}} \tag{8}$$

Constraints for *P′* elements are obviously the lower band of 0 and the upper band of 1.

This approach guarantees that regardless of the operators chosen, the difference between *R* and *R′* will be minimized. Moreover, continuity and differentiability are necessary conditions for the line search. Specifically, operators containing such terms as $\max\{c, f(a,b)\}$ render SQP unusable. In situations like this, the following substitutions will be used:

$$\max\{c, f(a,b)\} = \frac{c + f(a,b)}{2} + \frac{|c - f(a,b)|}{2} \tag{9}$$

$$\min\{c, f(a,b)\} = \frac{c + f(a,b)}{2} - \frac{|c - f(a,b)|}{2} \tag{10}$$

It's worth mentioning that unlike the "minimum" and "maximum" functions, the "absolute value" function is differentiable using symbolic math—across all its domain except where $u(x) = 0$.

$$\frac{d}{dx}|u(x)| = \frac{|u(x)|}{u(x)} u'_x(x) \tag{11}$$

4 Sample Results

The MATLAB 2008a environment was chosen to implement the proposed method, because of its efficiency in matrix operations and also the built-in functions and methods. The sample M code for Yager's operators can be found in the appendix.

For each run, *P* and *Q* are generated randomly and *R* is obtained through $P \circ Q$ with predetermined t-norm and t-conorm parameters to guarantee the existence of at least one solution for *P*. This initial value of *P* is then discarded so as not to interfere with the search process.

P, *Q*, and *R* are all 5×5 in the samples, since dimensions do not cause any significant change in the method's behavior. The initial random search produces 2000 solutions and chooses the one with the least *RMSE* for the SQP initialization. The stopping criterion for SQP is the number of *RMSE* evaluations, which was determined to be 1500.

ISBN: 978-989-95079-6-8

Operators of the same type (e.g. Dombi) were chosen in each run. Sample results are summarized in Table 1 for the following triangular norms and conorms:

- Frank
- Dombi
- Schweizer & Sklar 1
- Schweizer & Sklar 2
- Schweizer & Sklar 3
- Schweizer & Sklar 4
- Yager

More information on parametric fuzzy operators can be found in [10].

The method was repeated 10 times for each operator, with the best, worst, and average results for *RMSE* reported in the Table 1. Also included is the average running time.

Table 1 shows the performance of the proposed method with various parametric operators. The results imply that the Frank operators provide the most suitable search space for this method.

Table 1: test results for various fuzzy operators

Fuzzy Operator	Worst RMSE	Best RMSE	Average RMSE	Average Running Time
Frank	0.0008	0.00005	0.000313	222.48
Dombi	0.064	0.0009	0.026438	88.11
Schweizer & Sklar 1	0.014	0.0003	0.003795	300.98
Schweizer & Sklar 2	0.064	0.0050	0.018533	188.55
Schweizer & Sklar 3	0.046	0.0006	0.016644	166.08
Schweizer & Sklar 4	0.036	0.0003	0.008164	363.53
Yager	0.072	0.0001	0.020254	1006.42

For the sake of clarity, a numerical example will be fully featured here .Given the following values for Q and R, we want to estimate P such that

$$P \circ Q = R$$

using the Frank norm.

$$Q = \begin{bmatrix} 0.3500 & 0.3500 & 0.2900 & 0.0800 & 0.1300 \\ 0.2000 & 0.8300 & 0.7600 & 0.0500 & 0.5700 \\ 0.2500 & 0.5900 & 0.7500 & 0.5300 & 0.4700 \\ 0.6200 & 0.5500 & 0.3800 & 0.7800 & 0.0100 \\ 0.4700 & 0.9200 & 0.5700 & 0.9300 & 0.3400 \end{bmatrix}$$

$$R = \begin{bmatrix} 0.8551 & 0.9785 & 0.9356 & 0.9769 & 0.6889 \\ 0.6664 & 0.9419 & 0.8942 & 0.6672 & 0.6712 \\ 0.6423 & 0.9227 & 0.8260 & 0.8592 & 0.5874 \\ 0.4230 & 0.8127 & 0.8054 & 0.5878 & 0.5892 \\ 0.6713 & 0.9630 & 0.9023 & 0.9533 & 0.6681 \end{bmatrix}$$

After 57 iterations (1500 evaluations), P' is the estimated value for P.

$$P' = \begin{bmatrix} 0.2921 & 0.4785 & 0.7671 & 0.9538 & 0.8433 \\ 0.6804 & 0.9593 & 0.2627 & 0.5497 & 0.2491 \\ 0.2617 & 0.5455 & 0.3835 & 0.4141 & 0.7563 \\ 0.1811 & 0.5821 & 0.7026 & 0.1342 & 0.2487 \\ 0.0139 & 0.2564 & 0.8796 & 0.3088 & 0.9254 \end{bmatrix}$$

Then, R' is calculated according to (6):

$$R' = P' \circ Q$$

$$R' = \begin{bmatrix} 0.8549 & 0.9785 & 0.9360 & 0.9772 & 0.6888 \\ 0.6664 & 0.9418 & 0.8942 & 0.6672 & 0.6713 \\ 0.6417 & 0.9236 & 0.8279 & 0.8590 & 0.5859 \\ 0.4231 & 0.8125 & 0.8052 & 0.5878 & 0.5894 \\ 0.6715 & 0.9629 & 0.9021 & 0.9536 & 0.6682 \end{bmatrix}$$

And here's the difference between the actual R and the one obtained by the estimation of P (P'), which will lead to our error measure:

$$R' - R = \begin{bmatrix} -0.0002 & 0.0000 & 0.0005 & 0.0003 & -0.0001 \\ 0.0000 & -0.0001 & 0.0000 & 0.0000 & 0.0001 \\ -0.0005 & 0.0009 & 0.0019 & -0.0002 & -0.0015 \\ 0.0001 & -0.0002 & -0.0002 & 0.0000 & 0.0002 \\ 0.0002 & -0.0001 & -0.0002 & 0.0003 & 0.0000 \end{bmatrix}$$

Finally, by applying (7), we arrive at the value of *RMSE* as a universal error measure.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{5}\sum_{k=1}^{5}(r_{ik} - r'_{ik})^2}{5 \times 5}} = 5.5246\text{E-}04$$

5 Conclusions and Future Research

In this paper, we have proposed a new numerical method based on Sequential Quadratic Programming for solving fuzzy relation equations. This work is distinguished from previous methods in that:

- it views the FRE problem in the most general sense, being capable of handling sizes of $p \times m$ and $p \times n$ for P and R; and
- unlike analytical approaches, it can handle equations with any combinations of t-norms and t-conorms and at any dimensions.

However, the proposed method is not capable of tuning the parameters of t-norms and t-conorms. Also, it can be extended to be able to handle systems of Fuzzy Relation Equations. Such improvements can be considered for future research.

Appendix

The sample M file only for Yager's fuzzy operators is available through the following URL:
http://h1.ripway.com/commonlove1985/SQP.txt

References

[1] E. Sanchez, Resolution of Composite Fuzzy Relation Equations, *Information and Control*, 30: 38-48, 1976.

[2] A. V. Markovskii, On the Relation between Equations with Max-Product Composition and the Covering Problem, *Fuzzy Sets and Systems*, 153: 261-273, 2005.

[3] B. Shieh, Solutions of Fuzzy Relation Equations Based on Continuous t-norms, *Information Sciences*, 177: 4208-4215, 2007.

ISBN: 978-989-95079-6-8

[4] B. Shieh, Infinite Fuzzy Relation Equations with Continuous t-norms, *Information Sciences*, 178: 1961-1967, 2008.

[5] B. Shieh, Deriving Minimal Solutions for Fuzzy Relation Equations with Max-Product Composition, *Information Sciences*, 178: 3766-3774, 2008.

[6] A. A. Molai and E. Khorram, An Algorithm for Solving Fuzzy Relation Equations with Max-T Composition Operator, *Information Sciences*, 178: 1293-1308, 2008

[7] S. Hana, H. Lia and J. Wanga, Resolution of Finite Fuzzy Relation Equations Based on Strong Pseudo-t-norms, *Applied Mathematics Letters,* 19: 752-757, 2006.

[8] I. Perfilieva and L. Nosková, System of Fuzzy Relation Equations with inf-⟶ Composition: Complete Set of Solutions, *Fuzzy Sets and Systems*, 159: 2256-2271, 2008

[9] G. B. Stamou and S. G. Tzafestas, Resolution of Composite Fuzzy Relation Equations Based on Archimedean Triangular Norms, *Fuzzy Sets and Systems*, 120: 395-407, 2001.

[10] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall PTR, 1995.

[11] K. Schittkowski, *NLQPL: A FORTRAN-Subroutine Solving Constrained Nonlinear Programming Problems*, Annuals of Operations Research, 5: 485-500, 1985.

[12] J. Nocedal, S. J. Wright, *Numerical Optimization* (2nd ed.), Berlin, New York: Springer-Verlag, 2006

[13] R. Fletcher, *Practical Methods of Optimization*, John Wiley and Sons, 1987.

[14] P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization*, London. Academic Press, 1981.

[15] M. J. D. Powell, *Variable Metric Methods for Constrained Optimization. Mathematical Programming: The State of the Art*, Springer Verlag, 288-311, 1983.

[16] W. Hock, K. Schittkowski, *A Comparative Performance Evaluation of 27 Nonlinear Programming Codes*, Computing, 30: 335, 1983

[17] J. Nocedal and S. J. Wright, *Numerical Optimization (Second Edition): Springer Series in Operations Research and Financial Engineering*, Springer Science+Business Media, LLC., 2006.

[18] M. Bartholomew-Biggs, *Nonlinear Optimization with Engineering Applications: Springer Optimization and Its Applications*, Springer Science+Business Media, LLC., 2008.

ISBN: 978-989-95079-6-8

A Multi-objective Evolutionary Algorithm for Tuning Fuzzy Rule-Based Systems with Measures for Preserving Interpretability.

M.J. Gacto[1] R. Alcalá, F. Herrera[2]

1 Dept. Computer Science, University of Jaen
Jaen, Spain
2 Dept. Computer Science and Artificial Intelligence, University of Granada
Granada, Spain
Email: mjgacto@ugr.es, {alcala,herrera}@decsai.ugr.es

Abstract— *In this contribution we propose a multi-objective evolutionary algorithm for Tuning Fuzzy Rule-Based Systems by considering two objectives, accuracy and interpretability. To this aim we define a new objective that allows preserving the interpretability of the system. This new objective is an interpretability index which is the union of three metrics to preserve the original shapes of the membership functions as much as possible while a tuning of the membership function parameters is performed. The proposed method has been compared to a single objective accuracy-guided algorithm in two real problems showing that many solutions in the Pareto front dominate to those obtained by the single objective-based one.*

Keywords— Fuzzy Rule-Based Systems, Tuning, Interpretability, Multi-Objective Evolutionary Algorithms.

1 Introduction

Fuzzy modeling usually tries to improve the accuracy of the system without inclusion of any interpretability measure, an essential aspect of Fuzzy Rule-Based Systems (FRBSs). However, the problem of finding the right trade-off between accuracy and interpretability has achieved a growing interest [1].

One of the most widely-used approaches to enhance the performance of FRBSs is the *tuning* of the Membership Functions (MFs). It consists of refining a previous definition of the Data Base (DB) once the Rule Base (RB) has been obtained. Generally, tuning is a variation in the shape of the MFs that improves their global interaction. Classically, the tuning methods refine the parameters that identify the MFs associated to the labels comprising the DB [2].

For this model, to take into account interpretability issues it is necessary to use a measure to quantify the interpretability of the tuned DB. This kind of measure could be used as an additional objective to maintain the interpretability of the fuzzy partitions when the tuning is carried out.

In the literature, many authors improve the difficult trade-off between accuracy and interpretability of FRBSs, obtaining linguistic models not only accurate but also interpretable. We can distinguish two kinds of approaches for managing the interpretability:

1. The complexity of the model [3, 4, 5, 6] (usually measured as number of rules, variables, labels per rule, etc.)
2. Measuring the interpretability of the fuzzy partitions [4, 7, 8] by means of a semantic interpretability measure.

A way for optimizing both objectives (accuracy and interpretability) is the use of the Multi-Objective Evolutionary Algorithms (MOEAs) [9, 10]. Since this problem presents a multi-objective nature the use of MOEAs to obtain a set of solutions with different degrees of accuracy and interpretability is an interesting way to work [3, 5, 6, 8].

In this work, we focus our attention in measuring the interpretability of the fuzzy partitions. We propose a semantics interpretability index for maintaining the interpretability of the system by means of the aggregation of several metrics, with the aim of preserving the original form of the MFs while a tuning is performed. These metrics try to minimize the displacement of the central point of the MFs, besides maintaining the symmetry and the area of the original MFs associated to the linguistic labels. To this end, we apply a specific MOEA to obtain accurate and interpretable linguistic fuzzy models by performing a tuning of the MFs parameters with two main objectives:

- The system error
- The proposed interpretability index.

This algorithm is based on the well known $SPEA2$ [11] and is called SPEA2$_{SI}$ ($SPEA2$ for Semantic Interpretability). In order to improve its search ability, SPEA2$_{SI}$ implements some concepts as incest prevention and restarting [12]. Moreover, it is focused on the most accurate solutions to finally form a wide Pareto front. Thus, SPEA2$_{SI}$ is aimed at generating a complete set of Pareto-optimum solutions with different trade-offs between accuracy and interpretability. We have not considered the well-known NSGA-II [13] algorithm since in [3], approaches based on SPEA2 were shown to be more effective when performing a tuning of the MFs.

To show the good performance of the proposed method it is compared with a single objective accuracy-guided tuning algorithm [14] by applying both of them to initial linguistics models obtained from automatic learning methods. Two real-world problems with different complexities have been considered showing that the solutions of the accuracy based algorithm are dominated by those obtained by SPEA2$_{SI}$. We can see as both objectives required are certainly contradictory as the obtained Pareto fronts moves clearly from the most accurate solutions to the most interpretable ones.

Next section describes the classic tuning used in this work. Section 3 presents the proposed index to control the interpretability of the MFs. Section 4 presents the SPEA2$_{SI}$ algorithm describing its main characteristics and the genetic operators considered. Section 5 shows the experimental study

ISBN: 978-989-95079-6-8

and the results obtained. Finally, in section 6 we point out some conclusions.

2 Preliminaries: Tuning of MFs

This approach, usually called DB tuning, involves refining the MF shapes from a previous definition once the remaining FRBS components have been obtained [15, 16, 17, 18, 19, 20]. The classic way to refine the MFs is to change their definition parameters. For example, if the following triangular-shape MF is considered:

$$\mu(x) = \begin{cases} \frac{x-a}{b-a}, & \text{if } a \leq x < b \\ \frac{c-x}{c-b}, & \text{if } b \leq x \leq c \\ 0, & \text{otherwise} \end{cases}$$

changing the basic parameters — a, b, and c — will vary the shape of the fuzzy set associated to the MF, thus influencing the FRBS performance (See Figure 1). The same yields for other shapes of MFs (trapezoidal, gaussian, sigmoid, etc.).

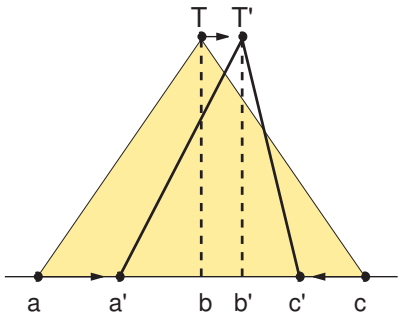


Figure 1: Tuning by changing the basic MF parameters

Tuning involves fitting the characterization of the MFs associated to the primary linguistic terms considered in the system. Thus, the meaning of the linguistic terms is changed from a previous definition (an initial DB). In order to ensure the semantics integrity through the MFs optimization process [1, 21, 22], some researchers have proposed several properties. Considering one or more of these properties several constraints can be applied in the design process in order to obtain a DB maintaining the linguistic model comprehensibility to the higher possible level [15, 23, 24, 25].

In this work we consider strong fuzzy partitions to define the initial MFs. In order to maintain semantics integrity we consider also these constraints by defining the variation intervals for each MF parameter. For each $MF_j = (a_j, b_j, c_j)$ where j=(1,...m) and m is the number of MFs in a given DB, the variation intervals are calculated in the following way:

$$[I^l_{a_j}, I^r_{a_j}] = [a_j - (b_j - a_j)/2, a_j + (b_j - a_j)/2] \ ,$$
$$[I^l_{b_j}, I^r_{b_j}] = [b_j - (b_j - a_j)/2, b_j + (c_j - b_j)/2] \ ,$$
$$[I^l_{c_j}, I^r_{c_j}] = [c_j - (c_j - b_j)/2, c_j + (c_j - b_j)/2] \ .$$

Thanks to these restrictions it is possible to maintain the comprehensibility of MFs to a reasonable level. In any case, it would be very interesting to have a measure for the quality of the tuned MFs. We propose three metrics trying to preserve the original form of the MFs, so improving if possible the trade-off between accuracy and interpretability.

3 An Interpretability Index based on Three Interpretability Metrics

In this section, we propose several metrics to measure the interpretability when a tuning is performed on the DB. At this point, we should remark that these metrics are based on the existence of the variation intervals (integrity constraints) defined in the previous section and, therefore, on the assumption that the initial DB is comprised of triangular uniformly distributed MFs (strong fuzzy partitions) with an associated linguistic meaning. Based on these assumptions, significant changes in the DB can influence negatively in the interpretability. In this way, each metric is proposed to control how good are different desirable aspects of the tuned MFs with respect to the original ones. The metrics proposed are:

- MFs displacements (δ): This metric measures the proximity of the central points of the MFs to the original corresponding ones. It should be higher as closer to the original point.
- MFs symmetry (γ): This metric measures the symmetry of the MFs. It should be higher when the two parts of the support of the MFs (left and right) present more similar lengths.
- MFs area similarity (ρ): This metric measures the similarity of the area of the tuned MFs to the original one. It should be higher when the tuned area and the original area are closer.

In the following subsections the three proposed metrics will be explained in depth.

3.1 MFs displacements (δ)

This metric can control the displacements in the central point of the MFs. It is based on computing the normalized distance between the central point of the tuned MF and the central point of the original MF, and is calculated through keeping the maximum displacement obtained on all the MFs. For each MF_j: $\delta_j = |b_j - b'_j|/I$, where $I = (I^r_{b_j} - I^l_{b_j})/2$ represents the maximum variation for each central parameter. Thus δ^* is defined as: $\delta^* = max_j\{\delta_j\}$.

The δ^* metric takes values between 0 and 1, thereby value near to 1 represent that the MFs present a great displacement. The following transformation is made so this metric represents proximity (maximization):

$$\delta = 1 - \delta^* \ ,$$

3.2 MFs symmetry (γ)

This metric can be used to control the symmetry of the MF shapes. It is based on relating the two parts of the support of the tuned MFs. Let us define that $leftS'_j = |a'_j - b'_j|$ is the length of the left part of the MF support and that $rightS'_j = |b'_j - c'_j|$ is the right part. γ_j is calculated using the following equation:

$$\gamma_j = \frac{min\{leftS'_j, rightS'_j\}}{max\{leftS'_j, rightS'_j\}} \ .$$

Values near to 1 mean that the two parts of the support of the MFs are more similar (higher symmetry). Finally γ is calculated by keeping the minimum value obtained.

$$\gamma = min_j\{\gamma_j\} \ .$$

ISBN: 978-989-95079-6-8

3.3 MFs area similarity (ρ)

This metric can be used to control the area of the MF shapes. It is based on relating the areas of the original and the tuned MFs. Let us define that A_j is the area of the triangle representing the original MF_j, and A'_j is the new area. ρ_j is calculated using the following equation for each MF:

$$\rho_j = \frac{min\{A_j, A'_j\}}{max\{A_j, A'_j\}} \ .$$

Values near to 1 mean that the original area and the tuned area of the MFs are more similar (less changes). The ρ metric is calculated through keeping the minimum value obtained:

$$\rho = min_j\{\rho_j\} \ .$$

3.4 A global Semantics Interpretability index based on the aggregation of the three metrics

We propose an aggregation of the metrics in a global index based on the geometric mean, that is denoted as GM3M index. It is clear that if only one of the proposed metrics has very low values a problem in the interpretability will arise. The aggregation operator should consider this fact:

$$GM3M = \sqrt[3]{\delta * \gamma * \rho}$$

The value of $GM3M$ ranges between 0 (the lowest level of interpretability) and 1 (the highest level of interpretability).

4 Proposed Multi-objective Evolutionary Algorithm

The proposed algorithm performs a parametric tuning to improve the system accuracy as a first objective and uses the $GM3M$ index to try to preserve the interpretability. It is called $SPEA2$ for Semantic Interpretability ($SPEA2_{SI}$) and is based on the well-known $SPEA2$[11] algorithm. In the next subsections the main components of this algorithm are described and the specific characteristics are presented.

4.1 Coding Scheme and Initial Gene Pool

We consider a real coding scheme, being m^i the number of labels of each of the n variables comprising the DB,

$$\begin{array}{l} C_i = (a_1^i, b_1^i, c_1^i, \ldots, a_{m^i}^i, b_{m^i}^i, c_{m^i}^i), \\ \ i = 1, \ldots, n \ , \\ C = C_1 C_2 \ldots C_n \ . \end{array}$$

The initial DB is included as first individual of the initial population. The remaining individuals are generated at random within the corresponding variation intervals defined in the previous section.

4.2 Objectives

In this algorithm we use two objectives. They are:

- The interpretability index ($GM3M$) that is a maximization objective representing the geometric mean of the three proposed metrics (interpretability).
- The Mean Squared Error (MSE) that is a minimization objective (accuracy):

$$\text{MSE} = \frac{1}{2 \cdot |E|} \sum_{l=1}^{|E|} (F(x^l) - y^l)^2,$$

with $|E|$ being the data set size, $F(x^l)$ being the output obtained from the FRBS decoded from such chromosome when the l-th example is considered and y^l being the known desired output. The fuzzy inference system considered to obtain $F(x^l)$ is the *center of gravity weighted by the matching* strategy as defuzzification operator and the *minimum t-norm* as implication and conjunctive operators.

4.3 Main Characteristics of $SPEA2_{SI}$

The proposed algorithm makes use of the $SPEA2$ selection mechanism. However in order to improve the search ability of the algorithm the following changes are considered:

- The proposed algorithm includes a mechanism for incest prevention based on the concepts of CHC [12], for maintaining population diversity. This mechanism avoids premature convergence. Only those parents whose hamming distance divided by 4 is higher than a threshold are crossed. The well-known BLX-0.5 [26] crossover is applied to obtain the offspring.

  Since we consider a real coding scheme, we have to transform each gene considering a Gray Code with a fixed number of bits per gene ($BGene$) determined by the system expert. In this way, the threshold value is initialized as:

  $$L = (\#C * BGene)/4,$$

  where $\#C$ is the number of genes of the chromosome.

  At each generation of the algorithm, the threshold value is decremented by one which allows crossing closer solutions. Following the concept of the CHC algorithm that does not used mutation, in our algorithm we do not include mutation operator.

- A restarting operator is applied by maintaining the most accurate individual, and the most interpretable individual as a part of the new population (external population is forced to be empty) and obtaining the remaining individuals with the tuning parameters generated at random within the corresponding variation intervals. This way preserves both the most accurate and the most interpretable solutions obtained.

  Restarting should be applied when we detect that all crossovers are allowed. However in order to avoid premature convergence we apply the first restart if 50 percent of crossovers are detected at any generation. This condition is updated each time restarting is performed as $\%_{Required} = (100 + \%_{Required})/2$. Moreover, the most accurate solution should be improved before each restarting. To preserve a well formed Pareto front at the end, the restarting is not applied in the last evaluations. The number of evaluations without restart can be estimated as the number of evaluations needed to apply the first restart multiplied by 10. Additionally, restart is disabled if it was never applied before reaching the mid of the total number of evaluations.

ISBN: 978-989-95079-6-8

- In each stage of the algorithm (between restarting points), the number of solutions in the external population ($\overline{P}_{t+1}$) considered to form the mating pool is progressively reduced, by focusing only on those with the best accuracy. To do that, the solutions are sorted from the best to the worst (considering accuracy as sorting criterion) and the number of solutions considered for selection is reduced progressively from 100% at the beginning to 50% at the end of each stage by taking into account the value of L.

  In the last evaluations when restart is disabled. This mechanism, whose main objective is focusing on the most accurate solutions, is also disabled in order to obtain a wide well formed Pareto front.

5 Experiments

To evaluate the goodness of the proposed approach, two real-world problems with different complexities (different number of variables and available data) are considered to be solved (these data sets are available at, *http://www.keel.es/*) [27]:

- Estimating the maintenance costs of medium voltage lines in a town (ELE).
- Predicting the Abalone Age (ABA).

In both cases, the well-known *ad hoc* data-driven learning algorithm of Wang and Mendel [28] is applied to obtain an initial set of candidate linguistic rules. To do so, we will consider strong fuzzy partitions of triangular-shaped MFs. Once the initial RB is generated, the proposed post-processing algorithm can be applied.

Methods considered for the experiments are:

- T method performs a classic MFs parameter tuning by only considering the accuracy of the model as the sole objective [14].
- $SPEA2_{SI}$ is the proposed MOEA method for the classic tuning considering two objectives, precision and the semantics interpretability index (GM3M).

5.1 Experimental Set-up

We consider a *5-fold cross-validation model*, i.e., 5 random partitions of data each with 20%, and the combination of 4 of them (80%) as training and the remaining one as test. For each one of the 5 data partitions, the tuning methods have been run 6 times, showing for each problem the averaged results of a total of 30 runs.

In the case of $SPEA2_{SI}$ the averaged values are calculated considering the most accurate solution from each Pareto obtained. In this way, $SPEA2_{SI}$ can be compared with the single objective method T.

The values of the input parameters considered by T are: population size of 61, 100000 evaluations, 0.6 as crossover probability and 0.2 as mutation probability per chromosome. The values of the input parameters considered by $SPEA2_{SI}$ are: population size of 200, external population size of 61, 100000 evaluations and 30 bits per gene for the Gray codification.

5.2 Results and Analysis

Table 1 shows the results obtained with WM, where $\#R$ stands for the number of rules, $\text{MSE}_{tra/tst}$ for the averaged error obtained over the training/test data, σ for their respective standard deviations and $GM3M$ for the interpretability index. This method obtains the initial knowledge bases that will be tuned by T and $SPEA2_{SI}$. For the WM method the interpretability index takes value 1 that is the highest level of interpretability.

Table 1: Results obtained with WM method

Dataset	#R	MSE_{tra}	σ_{tra}	MSE_{tst}	σ_{tst}	GM3M
ELE	65	56136	1498	56359	4685	1
ABA	68	8.407	0.443	8.422	0.545	1

The results obtained by both post-processing methods are shown in Table 2. In addition we also show δ, γ and ρ that represent the individual values of the metrics comprising, and *t* represents the results of applying a *test t-student* (with 95 percent confidence) in order to ascertain whether differences in the performance of the best results are significant when compared with that of the other algorithm in the table. The interpretation of the *t* column is:

- $\star$ represents the best averaged result.
- $+$ means that the best result has better performance than that of the related row.

Analysing the results showed in Table 2, we can highlight the following facts:

- The proposed method obtains the best results in training and test with respect to T in both problems. In the case of the electrical problem, SPEA2_{SI} improves among 7% and 8.5% in training and test respectively and in the abalone problem it obtains an improvement of 1.5%.
- The most accurate solutions from SPEA2_{SI} improve the accuracy and obtain more interpretable models, with 29%(ELE) and 52%(ABA) of improvement in the interpretability index with respect to the T method.

Figure 2 shows the Pareto front obtained with SPEA2_{SI} method, the solution obtained by T and the initial knowledge base obtained by WM in the same data partition and seed of ELE. We can observe that the obtained Pareto front is quite wide. In fact, the number of non dominated solutions is always equal to the external population size. Moreover, the WM solution coincides with the last point of the Pareto front and the solution obtained with T is dominated by several solutions from SPEA2_{SI}. Furthermore there is not overfitting in the results obtained with the proposed method.

The Pareto front obtained allows selecting solutions with different degrees of accuracy and interpretability. Figure 2 presents that an improvement in the interpretability produces lack of precision and an improvement in the precision produces lack of interpretability. This figure clearly shows that

ISBN: 978-989-95079-6-8

Table 2: Results obtained in both problems

Dataset	Method	MSE_{tra}	σ_{tra}	t-test	MSE_{tst}	σ_{tst}	t-test	GM3M	σ_{GM3M}	t-test	δ	γ	ρ
ELE	T	17020	1893	+	21027	4225	=	0.225	0.046	+	0.058	0.337	0.694
	$SPEA2_{SI}$	**15884**	1191	*	**19257**	2893	*	**0.319**	0.170	*	0.225	0.398	0.648
ABA	T	2.688	0.063	+	2.770	0.242	=	0.144	0.051	+	0.032	0.182	0.636
	$SPEA2_{SI}$	**2.648**	0.051	*	**2.744**	0.276	*	**0.298**	0.153	*	0.177	0.392	0.625

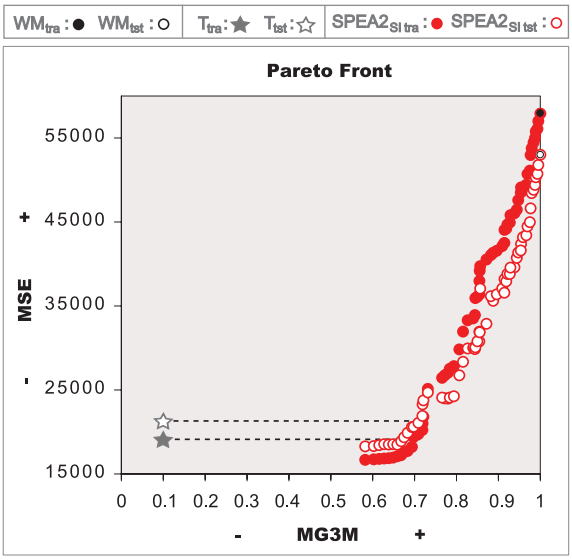


Figure 2: Pareto Front obtained in ELE

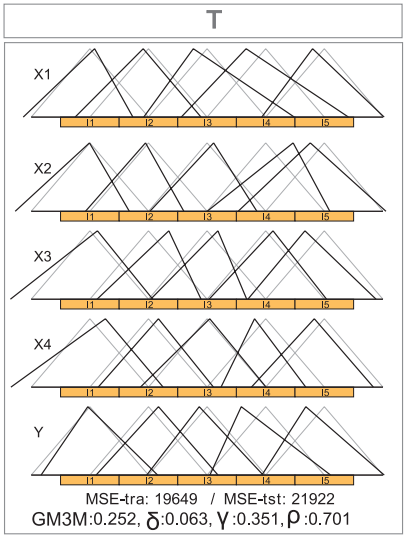


Figure 3: A DB obtained with T method in ELE

both objectives are actually contradictory. In the extremes of the Pareto front an improvement in one objective represents small lost in the other objective. On the contrary in the mid part of the Pareto front improvements in one objective deteriorates the other objective.

Figure 3 presents an illustrative DB obtained with T. Some DBs obtained with $SPEA2_{SI}$, in the same data partition and seed than this shown for T, are shown in Figure 4. This figure includes three DBs, one with the most accurate solution, other with a solution not only accurate but also interpretable and another highly interpretable DB, that obtains 35% of improvement with respect to the WM method with a value of the interpretability index near to 1.

6 Conclusions

The proposed index is useful to preserve the original shape of the MFs, in order to maintain the interpretability and it is a measure of the quality of the DB.

The proposed method obtains wide well formed Pareto fronts that provide a large variety of solutions to select from more accurate solutions to more interpretable ones.

Finally, we can stand out that $SPEA2_{SI}$ algorithm is very competitive and efficient since it is able to maintain the DB interpretability at a better level while accuracy is greatly improved.

Acknowledgment

Supported by the Spanish Ministry of Education and Science under grant no. TIN2008-06681-C06-01.

References

[1] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, editors. *Interpretability issues in fuzzy modeling*, volume 128 of *Studies in Fuzziness and Soft Computing*. Springer, 2003.

[2] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena. *Genetic Fuzzy Systems. Evolutionary tuning and learning of fuzzy knowledge bases*, volume 19 of *Advances in Fuzzy Systems - Applications and Theory*. World Scientific, 2001.

[3] R. Alcalá, M. J. Gacto, F. Herrera, and J. Alcalá-Fdez. A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(5):539–557, 2007.

[4] José M. Alonso, Luis Magdalena, and Serge Guillaume. Hilk: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *International Journal of Intelligent Systems*, 23(7):761–794, 2008.

[5] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index. *Soft Computing*, 13(5):437–449, 2008.

[6] H. Ishibuchi and Y. Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy

ISBN: 978-989-95079-6-8

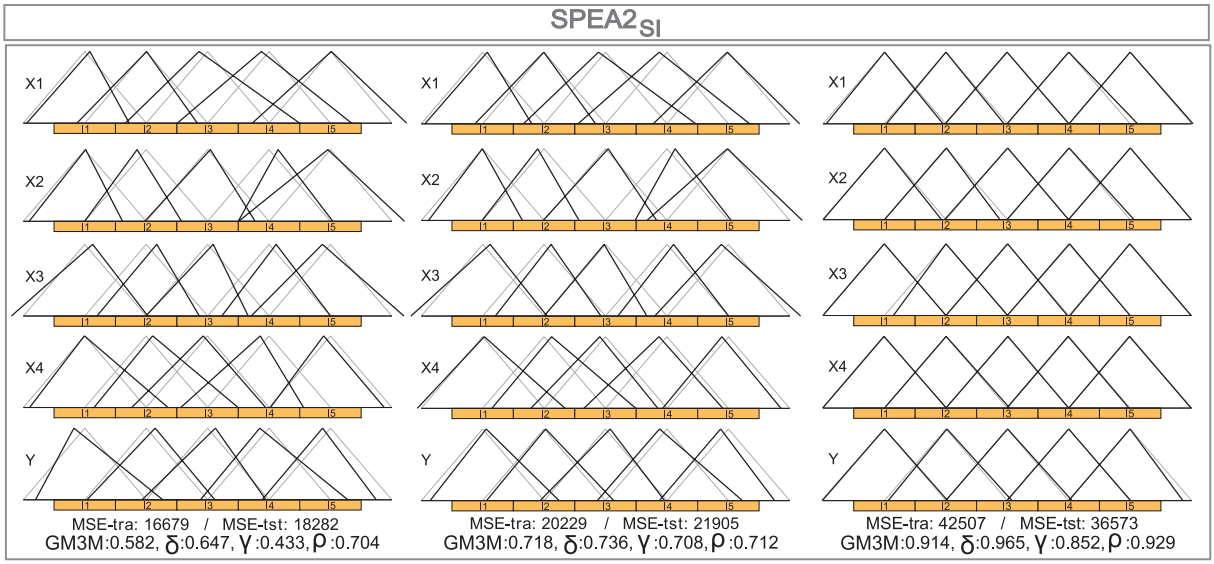


Figure 4: Three DBs obtained in a run of SPEA2$_{SI}$ in ELE

genetics-based machine learning. *International Journal of Approximate Reasoning*, 44(1):4–31, 2007.

[7] J.M. Alonso, O. Cordón, S. Guillaume, and L. Magdalena. Highly interpretable linguistic knowledge bases optimization: Genetic tuning versus solis-wetts. looking for a good interpretability-accuracy trade-off. In *Proceedings of the 2007 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2007)*, pages 901–906, 2007.

[8] P. Fazendeiro and J. V. de Oliveira. A working hypothesis on the semantics/accuracy synergy. In *EUSFLAT-LFA 2005*, pages 266–271, 2005.

[9] C. A. Coello, D. A. Van Veldhuizen, and G. B. Lamont, editors. *Evolutionary algorithms for solving multi-objective problems*. Kluwer Academic Publishers, 2002.

[10] K. Deb. *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, NY, USA, 2001.

[11] E. Zitzler, M. Laumanns, and L. Thiele. Spea2: Improving the strength pareto evolutionary algorithm for multiobjective optimization. In *Proc. Evolutionary Methods for Design, Optimization and Control with App. to Industrial Problems*, pages 95–100, Barcelona, Spain, 2001.

[12] L. J. Eshelman. The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. In G.J.E. Rawlin, editor, *Foundations of genetic Algorithms*, volume 1, pages 265–283. Morgan Kaufman, 1991.

[13] K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolutionary Computation*, 6(2):182–197, 2002.

[14] J. Casillas, O. Cordón, M. J. Del Jesus, and F. Herrera. Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Trans. Fuzzy Syst.*, 13(1):13–29, 2005.

[15] O. Cordón and F. Herrera. A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples. *International Journal of Approximate Reasoning*, 17(4):369–407, 1997.

[16] H. B. Gürocak. A genetic-algorithm-based method for tuning fuzzy logic controllers. *Fuzzy Sets and Systems*, 108(1):39–47, 1999.

[17] F. Herrera, M. Lozano, and J. L. Verdegay. Tuning fuzzy logic controllers by genetic algorithms. *International J. of Approximate Reasoning*, 12:299–315, 1995.

[18] Y. Jin, W. von Seelen, and B. Sendhoff. On generating fc^3 fuzzy rule systems from data using evolution strategies. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(6):829–845, 1999.

[19] C.L. Karr. Genetic algorithms for fuzzy controllers. *AI Expert*, 6(2):26–33, 1991.

[20] D. Nauck and R. Kruse. Neuro-fuzzy systems for function approximaton. *Fuzzy Sets and Systems*, 101(2):261–271, 1999.

[21] J. Valente de Oliveira. Semantic constraints for membership function optimization. *IEEE Trans. Syst., Man, Cybern. - Part A: Systems and Humans*, 29(1):128–138, 1999.

[22] J. Valente de Oliveira. Towards neuro-linguistic modeling: constraints for optimization of membership functions. *Fuzzy Sets and Systems*, 106(3):357–380, 1999.

[23] U. Bodenhofer and P. Bauer. A formal model of interpretability of linguistic variables. In Casillas et al. [1], pages 524–545.

[24] F. Cheong and R. Lai. Constraining the optimization of a fuzzy logic controller using an enhanced genetic algorithm. *IEEE Trans. Syst., Man, Cybern. - Part B: Cybernetics*, 30(1):31–46, 2000.

[25] J. Espinosa and J. Vandewalle. Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm. *IEEE Trans. Fuzzy Syst.*, 8(5):591–600, 2000.

[26] L. J. Eshelman and J. D. Schaffer. Real-coded genetic algorithms and interval-schemata. *Foundations of Genetic Algorithms*, 2:187–202, 1993.

[27] J. Alcalá-Fdez, L. Sánchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, and F. Herrera. Keel: A software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, 13(3):307–318, 2009.

[28] L. X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Trans. Syst., Man, Cybern.*, 22(6):1414–1427, 1992.

ISBN: 978-989-95079-6-8

On Tolerant Fuzzy c-Means Clustering with L_1-Regularization

HAMASUNA Yukihiro[1] ENDO Yasunori[2] MIYAMOTO Sadaaki[2]

1. Doctoral Program in Risk Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
2. Faculty of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
Email: yhama@soft.risk.tsukuba.ac.jp, {endo, miyamoto}@risk.tsukuba.ac.jp

Abstract— *We have proposed tolerant fuzzy c-means clustering (TFCM) from the viewpoint of handling data more flexibly. This paper presents a new type of tolerant fuzzy c-means clustering with L_1-regularization. L_1-regularization is well-known as the most successful techniques to induce sparseness. The proposed algorithm is different from the viewpoint of the sparseness for tolerance vector. In the original concept of tolerance, a tolerance vector attributes to each data. This paper develops the concept to handle data flexibly, that is, a tolerance vector attributes not only to each data but also each cluster. First, the new concept of tolerance is introduced into optimization problems. These optimization problems are based on conventional fuzzy c-means clustering (FCM). Second, the optimization problems with tolerance are solved by using Karush-Kuhn-Tucker conditions and an optimization method for L_1-regularization. Third, new clustering algorithms are constructed based on the explicit optimal solutions. Finally, the effectiveness of the proposed algorithm is verified through some numerical examples.*

Keywords— fuzzy c-means clustering, L_1-regularization, optimization, tolerance, uncertainty

1 Introduction

The aim of data mining is to discover important knowledge from a large quantity of data. From this viewpoint, clustering techniques have been actively studied. Clustering is one of the well-known unsupervised classification methods. For example, hard c-means clustering (HCM) is the most basic method. Fuzzy c-means clustering (FCM) is one of the well-known and useful clustering methods. For example, standard fuzzy c-means clustering (sFCM) [1] and entropy regularized fuzzy c-means clustering (eFCM) [2] are representatives. The entropy regularized fuzzy c-means clustering is constructed by regularization with maximum entropy function.

By the way, there are some difficulties of handling a set of data by clustering methods. Some clustering algorithms have been proposed to solve significant problems, for example, data with uncertainty, cluster size, noise or isolated data and so on. When we handle a set of data, data contains inherent uncertainty. For example, errors, ranges or some missing value of attributes are much caused. In these cases, each data is represented by an interval or a set instead of a point. In case of handling data with uncertainty, some significant methods have been proposed [3, 4]. These methods can not only handle data with uncertainty but also obtain high quality results by considering data with uncertainty. Thus, handling data with uncertainty is a very important problem in the field of data mining.

Therefore, some of the authors have proposed the original concept of tolerance [5, 6] which handle data with the above-mentioned uncertainty by using tolerance vector, and constructed some clustering algorithms [7, 8]. In these algorithms, tolerance is defined as hypersphere [5, 6] or hyper-rectangle [7, 8]. In case of hyper-rectangle, the missing value of attributes are handled successfully.

On the other hand, it is difficult to obtain clusters with different size or shape by a single-objective function, e.g., conventional FCM. In general, multi-objective optimization is considered to solve such problems. However, there are some problems by optimization, for example, how to select objective function. We have thought to be solved the above-mentioned problems by introducing a kind of "flexibility" for a pattern space and we have proposed the method to handle such "flexibility" by tolerance vector [9].

It is considered that the constraint for tolerance vector much affects classification results. It means that unsuited parameter makes trivial solutions. Therefore, regularization technique, e.g., Tikhonov's regularization [10] method have been used to solve these ill-posed problems. The quadratic and maximum entropy functions are typical regularization techniques. Moreover, L_1-regularization is the most efficient technique to induce sparseness. In the field of regression models or machine learning, some methods with L_1-regularization have been proposed and given sparse classifiers [12, 13, 14]. In algorithms applied to L_1-regularization, a lot of vari-

ISBN: 978-989-95079-6-8

ables are calculated zero. Non-zero variables are essential to understand classification results.

First, we consider optimization problems of tolerant fuzzy c-means clustering with L_1-regularization. Second, the optimal solutions are derived by Karush-Kuhn-Tucker (KKT) conditions and an optimization method for L_1-regularization . Third, we construct a new clustering algorithm by above-mentioned processes. Finally, the effectiveness of the proposed algorithm is verified through some numerical examples.

2 Preparation

Let data set, cluster and its cluster center be $X = \{x_k \mid x_k = (x_k^1, \ldots, x_k^p)^T \in \Re^p, k = 1 \ldots n\}$, C_i $(i = 1, \ldots, c)$ and $v_i = (v_i^1, \ldots, v_i^p)^T \in \Re^p, (i = 1, \ldots, c)$, respectively. Moreover, u_{ki} is the membership grade of x_k belonging to C_i and we denote the partition matrix $U = (u_{ki})_{k=1\sim n,\ i=1\sim c}$.

2.1 Fuzzy c-means clustering

Fuzzy c-means clustering is based on optimizing an objective function under constraint for membership grade.

We consider following two types of objective functions J_s and J_e.

$$J_s(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^m \|x_k - v_i\|^2,$$

$$J_e(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}\|x_k - v_i\|^2 + \lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}\log u_{ki}.$$

J_s is a well-known objective function of standard fuzzy c-means clustering (sFCM) proposed by Bezdek [1] and J_e is an entropy regularized fuzzy c-means clustering (eFCM) [2].

The constraint for u_{ki} is as follows :

$$\sum_{i=1}^{c} u_{ki} = 1\ ,\ u_{ki} \in [0,1]\ ,\ {}^{\forall}k.$$

2.2 Regularization

As above-mentioned, regularization is efficient techniques in the field of data mining. Tikhonov's regularization [10] method has been used to solve ill-posed problems. Many data mining algorithms which are applied to regularization have been actively studied to determine a variety of solution [12, 13, 14].

In the field of clustering, many membership regularization techniques have been proposed to obtain a variety of membership functions.

$$\min_U J(U,V) + \lambda\Omega(U),$$

where λ is a regularization parameter and $\Omega(U)$ is a regularization term.

The choice of a regularization term is important to determine a shape of a membership function. The quadratic function and maximum entropy function are typical examples.

$$\Omega(U) = \frac{1}{2}\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^2,$$

$$\Omega(U) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}\log u_{ki}.$$

3 The concept of tolerance

In the new concept of tolerance, the vector which attributes not only each data but also cluster center is defined as tolerance vector ε_{ki}. The tolerance κ_{ki} is defined as upper bound of tolerance vector.

From these formulation, the proposed methods can handle data more flexible than conventional methods.

We define κ_{ki}^j as the upper bound of each attribute of tolerance vector $\kappa_{ki} = (\kappa_{11}^1, \ldots, \kappa_{nc}^p) \geq 0$ and tolerance vector $E = \{\varepsilon_{ki} \mid \varepsilon_{ki} = (\varepsilon_{ki}^1, \ldots, \varepsilon_{ki}^p)^T \in \Re^p\ \}$ which mean the admissible range of each data, and the vector within the range of tolerance, respectively. The constraint for ε_{ki}^j is as follows :

$$|\varepsilon_{ki}^j|^2 \leq (\kappa_{ki}^j)^2\ (\kappa_{ki}^j \geq 0), {}^{\forall}k,\ i,\ j.$$

Figure 1 is an illustrative example about the new concept of tolerance defined as hyper-rectangle in $\Re^2$.

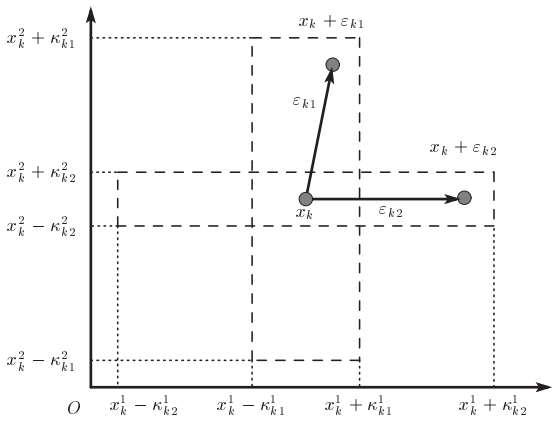


Figure 1: The new concept of tolerance defined as hyper-rectangle in $\Re^2$.

4 Tolerant fuzzy c-means clustering

In this section, we discuss optimization problems for clustering. We formulate the tolerant fuzzy c-means clustering (TFCM) by introducing the notion of tolerance into optimization problems and consider the way

ISBN: 978-989-95079-6-8

to optimize this objective function under the constraints for u_{ki} and ε_{ki}^j. The squared Euclidean-norm is used as dissimilarity, that is,

$$d_{ki} = \|x_k + \varepsilon_{ki} - v_i\|_2^2 = \sum_{j=1}^{p}(x_k^j + \varepsilon_{ki}^j - v_i^j)^2.$$

4.1 Tolerant standard fuzzy c-means clustering

The optimization problem is as follows :

$$J_{ts}(U,E,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^m d_{ki}, \tag{1}$$

under the following constraints,

$$\sum_{i=1}^{c} u_{ki} = 1\ ,\ u_{ki} \in [0,1]\ ,\ {}^\forall k, \tag{2}$$

$$|\varepsilon_{ki}^j|^2 \le (\kappa_{ki}^j)^2\ (\kappa_{ki}^j \ge 0), {}^\forall k,\ i,\ j. \tag{3}$$

The goal is to find the solutions which minimize the objective function (1) under the constraints (2) and (3).

From the convexity of (1), we introduce the following Lagrangian function to solve this optimization problem.

The Lagrangian function L_{tsr} is as follows :

$$\begin{aligned} L_{tsr} =& J_{ts}(U,E,V) + \sum_{k=1}^{n}\gamma_k(\sum_{i=1}^{c}u_{ki} - 1) \\ &+ \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p}\delta_{ki}^j(|\varepsilon_{ki}^j|^2 - (\kappa_{ki}^j)^2). \end{aligned}$$

Karush-Kuhn-Tucker conditions (KKT conditions) are as follows:

$$\left\{\begin{array}{l} \frac{\partial L_{tsr}}{\partial v_i^j} = 0,\ \frac{\partial L_{tsr}}{\partial \varepsilon_{ki}^j} = 0,\ \frac{\partial L_{tsr}}{\partial u_{ki}} = 0,\ \frac{\partial L_{tsr}}{\partial \gamma_k} = 0, \\ \frac{\partial L_{tsr}}{\partial \delta_{ki}^j} \le 0,\ \delta_{ki}^j\frac{\partial L_{tsr}}{\partial \delta_{ki}^j} = 0,\ \delta_{ki}^j \ge 0. \end{array}\right. \tag{4}$$

First we consider u_{ki}, from $\frac{\partial L_{tsr}}{\partial u_{ki}} = 0$, we have,

$$u_{ki} = \left(\frac{\gamma_k}{m d_{ki}}\right)^{\frac{1}{m-1}}. \tag{5}$$

In addition, from the constraint (2),

$$\sum_{l=1}^{c}\left(\frac{\gamma_k}{m d_{kl}}\right)^{\frac{1}{m-1}} = 1. \tag{6}$$

From (5) and (6), we have,

$$u_{ki} = \frac{\frac{1}{(d_{ki})^{\frac{1}{m-1}}}}{\sum_{l=1}^{c}\frac{1}{(d_{kl})^{\frac{1}{m-1}}}}. \tag{7}$$

If some $x_k + \varepsilon_{ki} = v_i$, we set $u_{ki} = 1/|C'|$. Here, $|C'|$ is number of cluster centers which satisfies $x_k + \varepsilon_{ki} = v_i$.

For v_i^j, from $\frac{\partial L_{tsr}}{\partial v_i^j} = 0$,

$$v_i^j = \frac{\sum_{k=1}^{n}(u_{ki})^m(x_k^j + \varepsilon_{ki}^j)}{\sum_{k=1}^{n}(u_{ki})^m}. \tag{8}$$

For ε_{ki}^j from $\frac{\partial L_{tsr}}{\partial \varepsilon_{ki}^j} = 0$, we can get

$$\varepsilon_{ki}^j = -\frac{(u_{ki})^m(x_k^j - v_i^j)}{(u_{ki})^m + \delta_{ki}^j}. \tag{9}$$

From $\delta_{ki}^j\frac{\partial L_{tsr}}{\partial \delta_{ki}^j} = 0$,

$$\delta_{ki}^j(|\varepsilon_{ki}^j|^2 - (\kappa_{ki}^j)^2) = 0. \tag{10}$$

From (10), we should consider two cases, i.e., $\delta_{ki}^j = 0$ and $|\varepsilon_{ki}^j|^2 = (\kappa_{ki}^j)^2$. First, we consider the case of $\delta_{ki}^j = 0$. In this case, the constraint (3) is not considered. From (9), we can get,

$$\varepsilon_{ki}^j = -(x_k^j - v_i^j).$$

On the other hand, in case that $|\varepsilon_{ki}^j|^2 = (\kappa_{ki}^j)^2$,

$$|\varepsilon_{ki}^j|^2 = |-\frac{(u_{ki})^m(x_k^j - v_i^j)}{(u_{ki})^m + \delta_{ki}^j}|^2 = (\kappa_{ki}^j)^2.$$

From $(u_{ki})^m + \delta_{ki}^j > 0$,

$$\frac{(u_{ki})^m}{(u_{ki})^m + \delta_{ki}^j} = \frac{\kappa_{ki}^j}{|x_k^j - v_i^j|}. \tag{11}$$

From (9), (11),

$$\varepsilon_{ki}^j = \frac{-\kappa_{ki}^j(x_k^j - v_i^j)}{|x_k^j - v_i^j|}.$$

From the above, we can get an optimal solution for ε_{ki}^j as follows :

$$\begin{aligned} \varepsilon_{ki}^j =& -\alpha_{ki}^j(x_k^j - v_i^j), \\ \alpha_{ki}^j =& \min\left\{\frac{\kappa_{ki}^j}{|x_k^j - v_i^j|}, 1\right\}. \end{aligned} \tag{12}$$

4.2 Tolerant entropy regularized fuzzy c-means clustering

The optimization problem is as follows :

$$\begin{aligned} J_{te}(U,E,V) =& \sum_{k=1}^{n}\sum_{i=1}^{c}u_{ki}d_{ki} \\ &+ \lambda^{-1}\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ki}\log u_{ki}. \end{aligned}$$

ISBN: 978-989-95079-6-8

Constraints are same as (2) and (3).
The Lagrangian function L_{er} is as follows :

$$L_{ter} = J_{te}(U,E,V) + \sum_{k=1}^{n}\gamma_k(\sum_{i=1}^{c}u_{ki}-1) + \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p}\delta_{ki}^{j}(|\varepsilon_{ki}^{j}|^2-(\kappa_{ki}^{j})^2).$$

KKT conditions are same as (4).
The optimal solutions are derived as follows :

$$u_{ki} = \frac{\exp(-\lambda d_{ki})}{\sum_{l=1}^{c}\exp(-\lambda d_{kl})}, \quad (13)$$

$$v_i^j = \frac{\sum_{k=1}^{n}u_{ki}(x_k^j+\varepsilon_{ki}^j)}{\sum_{k=1}^{n}u_{ki}}, \quad (14)$$

$$\varepsilon_{ki}^j = -\alpha_{ki}^j(x_k^j-v_i^j),$$

$$\alpha_{ki}^j = \min\left\{\frac{\kappa_{ki}^j}{|x_k^j-v_i^j|},1\right\}.$$

5 Tolerant fuzzy c-means clustering with L_1-regularization

In this section, we will consider tolerant fuzzy c-means clustering with L_1-regularization. In this method, constraint for ε_{ki}^j is not considered.

5.1 L_1-regularization

It is well-known that L_1-regularization can induce the strong sparseness of the variables [12, 13, 14]. A lot of ε_{ki}^j become zero, by L_1-regularization described as follows :

$$\Omega(E) = \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p}|\varepsilon_{ki}^j|.$$

Here, we will describe objective function with L_1-regularization as J_{tl}. The partial derivatives of J_{tl} which respect to ε_{ki}^j will be uniformly zero, as follows :

$$\left|\frac{\partial J_{tl}}{\partial\varepsilon_{ki}^j}\right| = \nu \qquad \text{if } |\varepsilon_{ki}^j|>0,$$

$$\left|\frac{\partial J_{tl}}{\partial\varepsilon_{ki}^j}\right| < \nu \qquad \text{if } |\varepsilon_{ki}^j|=0.$$

Here, $\nu > 0$ is a regularization parameter. This denotes that if the partial derivatives of J below ν, ε_{ki}^j will be set exactly zero.

5.2 Tolerant standard fuzzy c-means clustering with L_1-regularization

We will consider the following objective function with L_1-regularization.

$$J_{tsl}(U,E,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ki})^m d_{ki} + \nu\sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p}|\varepsilon_{ki}^j|.$$

In the tolerant fuzzy c-means clustering, each tolerance vector can be solved separately. So, we consider the following semi-objective function J_{tsl}^{kij} :

$$J_{tsl}^{kij}(E) = (u_{ki})^m(x_k^j+\varepsilon_{ki}^j-v_i^j)^2 + \nu|\varepsilon_{ki}^j|.$$

To obtain partial derivatives respect to ε_{ki}^j, we will decompose $\varepsilon_{ki}^j = \xi^+ - \xi^-$, where all element of ξ^+ and ξ^- are nonnegative. Thus, the semi-objective function can be written as follows :

$$J_{tsl}^{kij}(E) = (u_{ki})^m(x_k^j+\varepsilon_{ki}^j-v_i^j)^2 + \nu(\xi^+ + \xi^-).$$

The constraints are as follows :

$$\varepsilon_{ki}^j \le \xi^+,$$
$$\varepsilon_{ki}^j \ge -\xi^-,$$
$$\xi^+, \xi^- \ge 0.$$

Introducing the Lagrange multiplier β^+, β^-, δ^+ and $\delta^- \ge 0$, Lagrangian L_{tsl} is as follows :

$$L_{tsl} = (u_{ki})^m(x_k^j+\varepsilon_{ki}^j-v_i^j)^2 + \nu(\xi^+ + \xi^-) + \beta^+(\varepsilon_{ki}^j-\xi^+) + \beta^-(-\varepsilon_{ki}^j-\xi^-) - \delta^+\xi^+ - \delta^-\xi^-.$$

Here,

$$\frac{\partial L_{tsl}}{\partial\xi^+} = \nu-\beta^+-\delta^+, \quad (15)$$

$$\frac{\partial L_{tsl}}{\partial\xi^-} = \nu-\beta^--\delta^-. \quad (16)$$

Since $\delta^+,\ \delta^- \ge 0$, conditions $0 \le \beta^+ \le \nu$ and $0 \le \beta^- \le \nu$ are obtained from (15) and (16), respectively. By using (15) and (16), the Lagrangian L_{tsl} is simplified as follows :

$$L_{tsl} = (u_{ki})^m(x_k^j+\varepsilon_{ki}^j-v_i^j)^2 + \beta\varepsilon_{ki}^j. \quad (17)$$

Here, $\beta = \beta^+ - \beta^-$ and satisfies condition $-\nu \le \beta \le \nu$.
From $\frac{\partial L_{tsl}}{\partial\varepsilon_{ki}^j} = 0$,

$$\frac{\partial L_{tsl}}{\partial\varepsilon_{ki}^j} = 2(u_{ki})^m(x_k^j+\varepsilon_{ki}^j-v_i^j)+\beta = 0.$$

From above,

$$\varepsilon_{ki}^j = -(x_k^j-v_i^j) - \frac{\beta}{2(u_{ki})^m}. \quad (18)$$

Introducing (18) to (17), the Lagrangian dual problem is written as follows :

$$L_{tsld} = -\frac{\beta^2}{4(u_{ki})^m} - \beta(x_k^j-v_i^j).$$

ISBN: 978-989-95079-6-8

This objective function is a quadratic function respect to β. From $\frac{\partial L_{tsld}}{\partial \beta} = 0$, this dual problem is solved as,

$$\beta = -2(u_{ki})^m (x_k^j - v_i^j). \tag{19}$$

From considering (18) and (19), the optimal solution of primal problem is derived. First, if $\beta \leq -\nu$, the optimal solution is $\beta = -\nu$. Second, if $-\nu \leq \beta \leq \nu$, the optimal solution is $\beta = -2(u_{ki})^m (x_k^j - v_i^j)$. Third, if $\nu \leq \beta$, the optimal solution is $\beta = \nu$. Finally, the optimal solution for ε_{ki}^j of L_1-regularized objective function is derived as follows :

$$\varepsilon_{ki}^j = \begin{cases} -(x_k^j - v_i^j) + \frac{\nu}{2(u_{ki})^m} & (\beta \leq -\nu), \\ 0 & (-\nu \leq \beta \leq \nu), \\ -(x_k^j - v_i^j) - \frac{\nu}{2(u_{ki})^m} & (\nu \leq \beta). \end{cases} \tag{20}$$

5.3 Tolerant entropy regularized fuzzy c-means clustering with L_1-regularization

From the same procedure, the optimal solution for ε_{ki}^j of tolerant entropy regularized fuzzy c-means clustering with L_1-regularization is derived as follows :

$$\varepsilon_{ki}^j = \begin{cases} -(x_k^j - v_i^j) + \frac{\nu}{2u_{ki}} & (\beta \leq -\nu), \\ 0 & (-\nu \leq \beta \leq \nu), \\ -(x_k^j - v_i^j) - \frac{\nu}{2u_{ki}} & (\nu \leq \beta). \end{cases} \tag{21}$$

6 Algorithms

Algorithms of TFCM derived in the above section are called as follows. In case of hyper-rectangle, we call these methods TsFCM(R) and TeFCM(R). In case of L_1-regularization, we call these methods TsFCM-L_1R, and TeFCM-L_1R.

Each algorithm of TFCM is calculated according to the following procedure. Eqs. **A**, **B** and **C** used in each algorithm follow **Table 1** .

Algorithm 1

TFCM1 *Set the initial values and parameters.*

TFCM2 *Calculate $u_{ki} \in U$ by Eq.* **A**.

TFCM3 *Calculate $v_i \in V$ by Eq.* **B**.

TFCM4 *Calculate $\varepsilon_{ki} \in E$ by Eq.* **C**. *If convergence criterion is satisfied, stop. Otherwise, go back to* **TFCM2**.

In these algorithms, the convergence criterion is convergence of each variable, value of objective function or number of repetition.

Table 1: The optimal solutions of each algorithm.

Algorithm	Eq.**A**	Eq.**B**	Eq.**C**
TsFCM(R)	(7)	(8)	(12)
TeFCM(R)	(13)	(14)	(12)
TsFCM-L_1R	(7)	(8)	(20)
TeFCM-L_1R	(13)	(14)	(21)

7 Numerical examples

In this section, some numerical examples of classification are shown. In these examples, '$\circ$', '$\triangle$', '$\square$' and '$*$' mean each cluster and cluster centers, respectively. Moreover, tolerance vectors are expressed by arrowed line. The value of each data are normalized between 0 and 10. In addition, $m = 2.0$ in TsFCM and $\lambda = 1.0$ in TeFCM.

A polaris data set is mapped into two dimensional pattern space and consists of 51 points. This data set should be classified into three clusters [1]. Fig. 2 and 3 are classification results of TFCM with L_1-regularization. In Fig. 2, $\nu = 2.0$ and the number of non-zero tolerance vector is 30. In Fig. 3, $\nu = 4.0$ and the number of non-zero tolerance vector is 15. Moreover, Fig. 4 shows the relation between regularization parameter ν and the average of zero parameter ratio out of 1000 trials.

From these results, it is verified that proposed algorithm with L_1-regularization can induce the strong sparseness and its sparseness is controlled by regularization parameter ν.

8 Conclusions

In this paper, we have formulated the optimization problems based on concept of tolerance and derived the optimal solutions of tolerant fuzzy c-means clustering with L_1-regularization. From these results, we have constructed new clustering algorithm. Moreover, we have verified the effectiveness of proposed algorithm through some numerical examples.

The proposed technique is essentially different from the past one from the viewpoint of handling data more "flexible" than conventional methods.

In future works, we will calculate with real data which includes data with uncertainty, e.g., the missing value of attributes. Next, we will consider another type of regularization term or constraint for tolerance vector to induce sparseness.

Acknowledgment

This study is partly supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists and the Grant-in-Aid for Scientific Research (C) and (B) (Project No.21500212 and

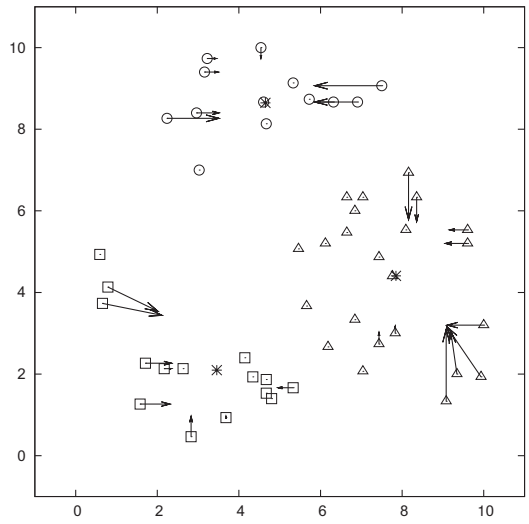


Figure 2: Result of TsFCM with L_1-regularization.

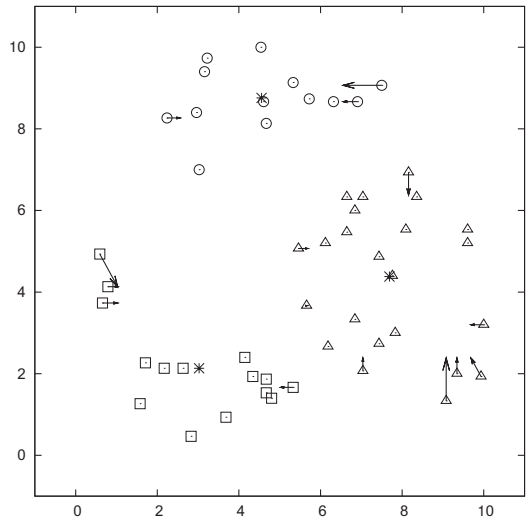


Figure 3: Result of TeFCM with L_1-regularization.

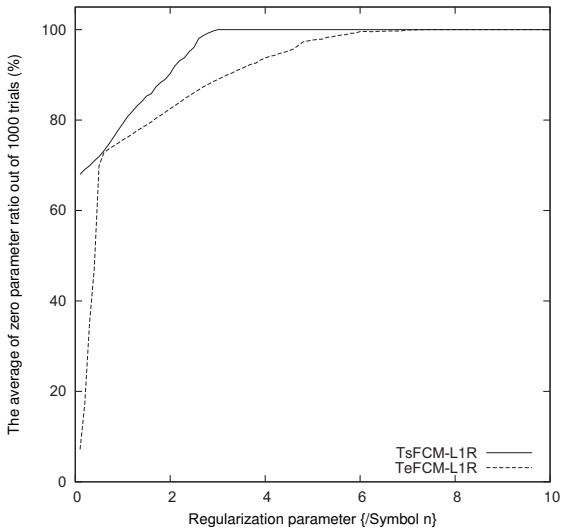


Figure 4: The relation between regularization parameter ν and the average of zero parameter ratio out of 1000 trials

No.19300074) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] J. C. Bezdek, 'Pattern Recognition with Fuzzy Objective Function Algorithms', Plenum Press,New York, 1981.

[2] S. Miyamoto and M. Mukaidono, "Fuzzy c-means as a regularization and maximum entropy approach", Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97), June 25-30, 1997, Prague, Czech, Vol. 2, pp. 86–92, 1997.

[3] W. K. Ngai, B. Kao, C. K. Chui, R. Cheng, M. Chau and K. Y. Yip, "Efficient Clustering of Uncertain Data", Proc. of the Sixth International Conference on Data Mining (ICDM'06), December 18-22, 2006, Hong Kong, China, pp. 436–445, 2006.

[4] O. Takata, and S. Miyamoto, "Fuzzy clustering of Data with Interval Uncertainties", Journal of Japan Society for Fuzzy Theory and Systems, Vol.12, No.5, pp. 686–695, 2000 (in Japanese).

[5] Y. Endo, R. Murata, H. Haruyama and S. Miyamoto, "Fuzzy c-Means for Data with Tolerance", Proc. of International Symposium on Nonlinear Theory and Its Applications, pp. 345–348, 2005.

[6] R. Murata, Y. Endo, H. Haruyama and S. Miyamoto, "On Fuzzy c-Means for data with Tolerance", Journal of Advanced Computational Intelligence and Intelligent Informatics 10(5), pp. 673–681, 2006.

[7] Y. Hasegawa, Y. Endo, Y. Hamasuna and S. Miyamoto, "Fuzzy c-Means for Data with Tolerance Defined as Hyper-rectangles", Modeling Decisions for Artificial Intelligence (MDAI2007), pp. 237-248. Springer, Heidelberg 2007.

[8] Y. Hamasuna, Y. Endo, S. Miyamoto and Y. Hasegawa, "On Hard Clustering for Data with Tolerance", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol.20, No.3, pp. 388–398, 2008 (in Japanese).

[9] Hamasuna Y., Endo Y. and Miyamoto S., "On Tolerant Fuzzy c-Means", Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), (to be published).

[10] Tikhonov, A. N. and V. Y. Arsenin, 'Solutions of Ill-Posed Problems', V. H. Winston, 1977.

[11] P. M. Williams, "Bayesian Regularization and Pruning Using a Laplace Prior", Neural Computation, Vol. 7, No.1, pp. 117–143, 1995.

[12] K. Tsuda and T. Kudo, "Clustering Graphs by Weighted Substructure Mining", Proc. of the 23rd international conference on Machine learning, pp. 953–960, 2006.

[13] R. Tibshirani, "Regression shrinkage and selection via the Lasso", Journal of the Royal Statistical Society, Series B, Vol. 58, No.1, pp. 267-288, 1996.

[14] J. Kazama and J. Tsuji, "Evaluation and extension of maximum entropy models with in equality constraints", In Proc. of EMNLP, pp. 137–144, 2003.

ISBN: 978-989-95079-6-8

A parametric approach to solve quadratic programming problems with fuzzy environment in the set of constraints

Carlos Cruz[1] Ricardo C. Silva[2] José L. Verdegay[1] Akebo Yamakami[2]

1.Department of Computer Science and Artificial Intelligence,
University of Granada, E-18071, Granada, Spain
Email: {carloscruz,verdegay}@decsai.ugr.es
2.Department of Telematics, School of Electrical and Computer Engineering,
University of Campinas, P.O.Box 6101, 13083-970, Campinas-SP, Brazil.
Email: {rcoelhos,akebo}@dt.fee.unicamp.br

Abstract— *Quadratic programming can be seen both as a general approach to linear programming and a special class of nonlinear programming. Moreover, Quadratic Programming problems are of utmost importance in a variety of relevant practical fields, such as, portfolio selection. This work presents and develops a novel fuzzy-sets-based method that solves a class of quadratic programming problems with vagueness in the set of constraints. As vagueness is natural and ever-present in real-life situations requiring solutions, it makes perfect sense to attempt to address them using fuzzy quadratic programming. This kind of problem modeling is being applied in an increasing variety of practical fields especially those with logistics problems. Some illustrative numerical examples illustrating the solution approach are solved and analyzed to show the efficiency of this proposed method.*

Keywords— Fuzzy sets, decision making, fuzzy mathematical programming, quadratic optimization.

1 Introduction

In the early sixties, based on the fact that classical logic does not reflect, to the extent that it should, the omnipresent imprecision in the real world, L. A. Zadeh proposed the Theory of Fuzzy Sets and Fuzzy Logic. Nowadays Fuzzy Logic, or rather Soft Computing, is employed with great success in the conception, design, construction and utilization of a wide range of products and systems whose functioning is directly based on how human beings reason. This is specifically patent in the case of optimization problems, and particularly in so called Mathematical Programming problems. Mathematical programming is an area which solves problems that involve minimization (or maximization) of the objective function in a function domain that can be constrained or not. In this area, Quadratic Programming represents a special class where the objective function is a quadratic function and the constraints are linear. This set of problems can be formalized in the following form:

$$\begin{aligned} \min \quad & \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q x} \\ \text{s.t.} \quad & \mathbf{A x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{1}$$

where $\mathbf{c}$ is an n vector, $\mathbf{b}$ is an m vector, $\mathbf{A}$ is an $m \times n$ matrix, and $\mathbf{Q}$ is an $n \times n$ symmetric matrix.

Thus, on the one hand, it is clear that quadratic programming encompasses all linear problems, including applications in scheduling, planning and flow computations, and they may be used to solve some interesting combinatorial optimization problems. On the other hand, quadratic programming is a particular kind of nonlinear programming. There are several classes of problems that are naturally expressed as quadratic problems. Examples of such problems can be found in game theory, engineering modeling, design and control, problems involving economies of scale, facility allocation and location problems, problems in microeconomics amongst others. Several applications and test problems for quadratic programming can be found in [5, 7, 8, 16, 17]. Some traditional methods are available in the literature [2, 24] for solving such problems. An interesting web page about quadratic programming is [6].

Among the several applications, we will present a portfolio selection problem which is an important research field in modern finance. This problem was first introduced by Markowitz [14, 15], and provided a risk investment analysis. Vagueness, approximate values and lack of precision in this problem are very frequent in that context, and quadratic programming problems have shown to be extremely useful in solving a variety of portfolio models. In any case, it is important to point out that the aim of this work is not to solve portfolio models. They are only considered here for the sake of illustrating the fuzzy quadratic programming problems solution approach presented, which in fact is the goal and main aim of this contribution. Some works about portfolio selection problem by using fuzzy approaches can be found in [9, 11, 19, 21, 22].

Moreover, there are some cases where the parameters of the real-world problems are seldom known exactly and have to be estimated by the decision maker. Therefore, the application of Soft Computing, and Fuzzy Logic in particular, has shown, in recent years, great potential for modeling systems which are non-linear, complex, ill-defined and not well understood. Fuzzy Logic is a way to describe this vagueness mathematically and it has found numerous and different applications due to its easy implementation, flexibility, tolerant nature to imprecise data, low cost implementations and ability to model non-linear behavior of arbitrary complexity because of its basis in terms of natural language. The uncertainties can be found in the relation, constants, decision variables or in all parameters of the problem. Some authors have applied Soft Computing methodologies to quadratic programming as can be seen in [1, 3, 4, 12, 13, 18, 20, 23, 25, 26, 27].

With this in mind, the goal of this paper is to present a novel approach that transforms a quadratic programming problem with uncertainties in the coefficients and order relation of the

ISBN: 978-989-95079-6-8

set of constraints into a parametric problem and to obtain a set of optimal solutions of this new problem that belong to the fuzzy solution.

Thus, Problem (1), which is a classic quadratic programming problem, can be rewritten in the following way:

$$\begin{aligned} \min \quad & \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q x} \\ \text{s.t.} \quad & \tilde{\mathbf{A}}\mathbf{x} \leq^f \tilde{\mathbf{b}} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{2}$$

where $\mathbf{x}$ and $\mathbf{c}$ are $n \times 1$ vectores with real numbers and $\tilde{\mathbf{b}}$ is an $m \times 1$ vector with fuzzy numbers, $\mathbf{Q}$ is an $n \times n$ matrix with real numbers and $\tilde{\mathbf{A}}$ is an $m \times n$ matrix with fuzzy components, and the symbol "$\leq^f$" shall mean just that the decision-maker is willing to permit some violations in the accomplishment of the constraints. Both fuzzy numbers and these violations are measured by membership functions $\mu_i : \mathbb{R} \to [0,1],\ i = 1,\ldots,m$ that for each solution provide the decision maker's degree of satisfaction with the accomplishment of each constraint.

The paper is organized as follows: Section 2 demonstrates that a fuzzy quadratic programming problem can be transformed into a parametric quadratic problem and that the solution obtained by traditional techniques is an acceptable fuzzy solution. To illustrate the approach, Section 3 presents a general portfolio selection problem formulated as a fuzzy quadratic programming; Section 4 presents numerical simulations for the proposed problems and an analysis of the results obtained. Finally, in Section 5 some conclusions are pointed out.

2 A general model for fuzzy quadratic programming

In this section, a novel approach is presented to solve quadratic programming problems with uncertainties in the order relation and coefficients in the set of constraints. This approach transforms this fuzzy quadratic problems into a parametric quadratic programming problem. A quadratic programming problem with this fuzzy environment has been formulated in (2) and the following sub-section will show how the uncertainties can be dealt with.

2.1 Parametric ideas for solving a fuzzy quadratic programming problem

The fuzzy coefficients of the quadratic problem (2) are defined with fuzzy nature, that is, some violations in the accomplishment of such restriction functions are permitted. Therefore these fuzzy parameters can be determined by the decision maker.

It is clear that each membership function will give the degree of membership (satisfaction) such that any $\mathbf{x} \in \mathbb{R}^n$ accomplishes the corresponding fuzzy objective function and constraint upon which it is defined. These membership functions can be formulated as follows

$$\mu_i : \mathbb{R} \to (0,1], \quad i \in I$$

where μ is an linear membership function (and formally can also be a non linear one), and I is the set that contains all fuzzy parameters.

In order to solve this problem in a two-phase method, first let us define for each fuzzy constraint, $i \in I$

$$X_i = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \tilde{A}_i \mathbf{x} \leq^f \tilde{b}_i, \mathbf{x} \geq 0 \right\}.$$

If $\mathbf{X} = \bigcap_{i \in I} X_i$ then the former fuzzy quadratic problem can be addressed in a compact form as

$$\min \{ f(\mathbf{x}) \mid \mathbf{x} \in \mathbf{X} \}$$

It is clear that $\forall \alpha, \gamma \in (0,1]$, an (α,γ)-cut of the fuzzy constraint set will be the classical set

$$X(\alpha,\gamma) = \{ \mathbf{x} \in \mathbb{R}^n \mid \mu_X(\mathbf{x}) \geq \min\{\alpha,\gamma\} \}$$

where $\forall \mathbf{x} \in \mathbb{R}^n$,

$$\mu_X(\mathbf{x}) = \inf \mu_i(\mathbf{x}), i \in I$$

Hence an (α,γ)-cut of the i-th constraint will be denoted by $X_i(\alpha,\gamma)$. Therefore, if $\forall \alpha, \gamma \in (0,1]$,

$$S(\alpha,\gamma) = \{ \mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = \min\ f(\mathbf{y}),\ \mathbf{y} \in X(\alpha,\gamma) \}$$

the fuzzy solution to the problem will therefore be the fuzzy set defined by the following membership function

$$\mu_{S(\alpha,\gamma)}(\mathbf{x}) = \begin{cases} \sup\{\alpha,\gamma : \mathbf{x} \in S(\alpha,\gamma)\} & \mathbf{x} \in \bigcup_\alpha S(\alpha,\gamma) \\ 0 & \text{otherwise.} \end{cases}$$

Provided that $\forall \alpha \in (0,1]$,

$$X(\alpha,\gamma) = \bigcap_{i \in I} \{ \mathbf{x} \in \mathbb{R}^n \mid [(\tilde{A})_\alpha x]_i \leq r_i(\alpha,\gamma), \mathbf{x} \geq 0, \mathbf{x} \in \mathbb{R}^n \}$$

with $r_i(\alpha,\gamma) = [(\tilde{b})_\alpha]_i + [(\tilde{d})_\alpha]_i (1-\gamma)$.

Thus, the operative solution to the former problem can be found, (α,γ)-cut by (α,γ)-cut, by means of the following auxiliary parametric quadratic programming model,

$$\begin{aligned} \min \quad & \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q x} \\ \text{s.t.} \quad & \tilde{\mathbf{A}}_\alpha \mathbf{x} \leq \tilde{\mathbf{b}}_\alpha + \tilde{\mathbf{d}}_\alpha (1-\gamma) \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{3}$$

Therefore, the fuzzy quadratic programming problem was parameterized at the end of the first phase. In the second phase, the parametric quadratic programming problem is presented as a multiobjective approach which is solved for each of the different α and γ values using any multiobjective optimization technique. We must find efficient solutions to parametric problem for each α and γ that satisfy Karush-Kuhn-Tucker's necessary efficient optimality conditions.

2.1.1 Methods to solve quadratic programming problems with fuzzy relations

By using the parametric idea described above a general method to solve quadratic programming problems with fuzzy relations is presented here, as described in [18]. This method is an extension of the method that was developed to solve fuzzy linear programming problems;

$$\begin{aligned} \min \quad & \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q x} \\ \text{s.t.} \quad & (Ax)_i \leq b_i + d_i(1-\alpha), i \in I \\ & x_j \geq 0, j \in J, \alpha \in (0,1]. \end{aligned} \tag{4}$$

where d_i is the violation permitted for each constraint i, $\mathbf{Q}$ and $\mathbf{A}$ are an $n \times n$ matrix and an $m \times n$ matrix with real numbers, and $\mathbf{c}$ and $\mathbf{b}$ are an n vector and m vector with real numbers.

ISBN: 978-989-95079-6-8

2.1.2 Methods to solve quadratic programming problems with fuzzy coefficients in the set of constraints

This multiobjective approach was developed in [10] to solve fuzzy nonlinear programming problems. This approach will be used to solve Problem (2) with fuzzy coefficients in the constraints set. The fuzzy solution is obtained by transforming a fuzzy nonlinear programming problem into a parametrical multiobjective nonlinear programming problem in which the parameters $\alpha, \gamma, \beta_{ij} \in [0,1],\ i = 1,\ldots,m, j = 1,\ldots,n+1$ can be treated as new decision variable.

The goal of this parametrical multiobjective problem is to minimize $f(\mathbf{x})$, maximize γ, and maximize and minimize $\alpha, \gamma,\ \beta_{ij},\ i = 1,\ldots,m,\ j = 1,\ldots,n+1$, simultaneously. Therefore, Problem (3) is transformed into a multiobjective quadratic programming problem that is stated as follows:

$$\begin{array}{ll} \min & \left[\mathbf{c^t x} + \frac{1}{2}\mathbf{x^t Q x}, \beta_{11}, 1-\beta_{11}, \ldots, \beta_{m,n+1}, 1-\beta_{m,n+1}\right] \\ \text{s.t.} & \sum_{j=1}^{n}\left((a_\alpha^L)_{ij} + \beta_{ij}\left((a_\alpha^U)_{ij} - (a_\alpha^L)_{ij}\right)\right)x_j \leq (b_\alpha^L + d_\alpha^L)_i + \\ & \quad + \beta_{i,n+1}\left((b_\alpha^U)_i - (b_\alpha^L)_i + \left((d_\alpha^U)_i - (d_\alpha^L)_i\right)(1-\gamma)\right) \\ & \mathbf{x} \geq \mathbf{0}, \alpha, \beta_{ij} \in [0,1],\ i = 1,\ldots,m,\ j = 1,\ldots,n+1 \end{array} \tag{5}$$

where it considers $m(n+3)$ new decision variables α, γ and $\beta_{ij},\ i = 1,\ldots,m,\ j = 1,\ldots,n+1$, to transform the intervals $I_{ij}(\alpha) = [h_{ij}^{-1}(1-\alpha), g_{ij}^{-1}(1-\alpha)]$ into functions of the form $z_{ij}(\alpha, \beta_{ij}) = h_{ij}^{-1}(1-\alpha) + \beta_{ij}(g_{ij}^{-1}(1-\alpha) - h_{ij}^{-1}(1-\alpha))$.

The obtained results for each α and γ value generate a set of solutions $S(\alpha, \gamma)$ and then the Representation Theorem can be used to integrate all these specific alpha-solutions and ending the second phase. Therefore the outlined solution to the parametric method is a valid solution to the fuzzy quadratic problem.

3 Portfolio selection problem

As said previously, in order to illustrate the above described two phases method for solving fuzzy quadratic programming problems, we now focus on general Portfolio Problems. It is important to point out that up to now we have not tried to improve other solution methods for this kind of important problems, but only to show how our solution approach performs. A description of a classical portfolio selection problem that was formulated by Markowitz as a quadratic programming problem is given in [15]. Assume that there are n securities denoted by S_j $(j = 1,\ldots,n)$, then this quadratic problem can be written in the following form:

$$\begin{array}{ll} \min & \mathbf{x^t \Sigma x} \\ \text{s.t.} & \mathbf{x^t E(R)} \geq \rho \\ & \mathbf{1x} = 1 \\ & \mathbf{x} \geq \mathbf{0} \end{array} \tag{6}$$

where $\mathbf{x}$ is an n vector that represents the percentage of money invested in assets, i.e., the proportion of total investment funds devoted to each security; $\mathbf{E(R)}$ is the average vector of returns over m periods because $\mathbf{R} = [r_{ij}]$ is an $m \times n$ matrix that represents the random variables of the returns of assets varying in m discrete times; ρ is a parameter representing the minimal rate of return required by an investor; and $\Sigma = [\sigma_{ij}^2]$ is a covariance $n \times n$ matrix between returns of asset which can be written as:

$$\sigma_{ij}^2 = \sum_{k=1}^{m} \frac{\left(r_{ki} - E(r_i)\right)\left(r_{kj} - E(r_j)\right)}{m-1}. \tag{7}$$

Therefore, the objective of Problem (6) is minimizing the risk variance and the investment diversification subject to a given average return ρ.

The expected return rate, ρ, is a decision maker's value that represents an expert's knowledge, then a fuzzy approach can be used in the constraint of the portfolio selection problem. Problem (6) with the use of fuzzy sets can be formulated in the following form:

$$\begin{array}{ll} \min & \mathbf{x^t \Sigma x} \\ \text{s.t.} & \mathbf{x^t E(R)} \geq^f \rho \\ & \mathbf{1x} = 1 \\ & \mathbf{x} \geq \mathbf{0}. \end{array} \tag{8}$$

Thus, Problem (8) can be rewritten as

$$\begin{array}{ll} \min & \mathbf{x^t \Sigma x} \\ \text{s.t.} & \mathbf{x^t E(R)} \geq \rho - d(1-\alpha) \\ & \mathbf{1x} = 1 \\ & \mathbf{x} \geq \mathbf{0}. \end{array} \tag{9}$$

4 Numerical experiments

The problems we use to evaluate this method are one hypothetic mathematical formulation and one fuzzy portfolio selection problem with the fuzzy approach described in Section 3. Nevertheless, they are efficient in validating the realized study. First, in subsection 4.1 we will show the formulation of the problems. Then in subsection 4.2 the computational results and a comparative analysis of the classic methods and the iterative methods responses will be presented.

The tests were all performed on a PC with two 2.26GHZ Intel® Core(™) 2 Duo processors, 4GB RAM running Ubuntu 8.10 operational system. All the problems presented in this work were resolved using **fminimax** function to solve constraint programming problems of *ToolBox Optimization* of MATLAB® 7.4.0 program.

4.1 Formulation of the numerical examples

We present some real-world and theoretical problems found in the literature with a view to validating the proposed algorithm. We simulate two quadratic programming problems both with uncertainties in the order relation and coefficients in the set of contraints.

Table 1 provides one theoretical quadratic problem that is described in [8]. The optimal solution to the problems without uncertainties is presented in the columns $\bar{\mathbf{x}}'$ and $f(\mathbf{x})$ of table 1.

In order to show the performance of our method, we use the set of historical data shown in Table 2 introduced by Markowitz. The columns 2-10 represent American Tobacco, A.T.&T., United States Steel, General Motors, Atcheson&Topeka&Santa Fe, Coca-Cola, Borden, Firestone and Sharon Steel securities data, respectively. The returns on the nine securities, during the years 1937-54, are presented in Table 2.

This example will consider performances of portfolios with respect to "return" thus defined. This assumes that a dollar of realized or unrealized capital gains is exactly equivalent to a

ISBN: 978-989-95079-6-8

dollar of dividends, no better and no worse. This assumption is appropriate for certain investors, for example, some types of tax-free institutions. Other ways of handling capital gains and dividends, which are appropriate for other investors, can be viewed in [15].By computing the average value of all the years of each column of random variables of Table 2, we obtained the expected values of each return of the securities that are described in (10). Then, using Equation (7), we computed the covariance matrix that is presented in (11).

4.2 Results and Analysis

Here we show the results obtained for the problem by the fuzzy quadratic programming method introduced in Section 2. The problem described in this work was solved by using linear membership functions as presented by Problem (3). Table 3 shows the solution of the theoretical problem described by Table 1, while the solution of the real-world portfolio selection problem, described by Table 2, is presented in Table 4.

Table 3 presents the results of the hypothetical problem with fuzzy coefficients and relation in the constraints set. Each row describes parametrical solutions to different α-cut levels of the fuzzy coefficients and each column represents parametrical solutions to different violations of the fuzzy relation in the set of constraints. Now, by applying the Representation Theorem in these parametrical solutions, we can define a fuzzy solution that describes a satisfactory solution of the fuzzy quadratic programming problem shown in Table (1).

Table 4 presents the results of the fuzzy portfolio selection problem to different α-cut levels of the fuzzy coefficients and violations of the fuzzy relation in the set of constraints. Thus, a fuzzy solution can be defined by applying the Representation Theorem in these parametrical solutions. This fuzzy solution describes a satisfactory solution of the fuzzy portfolio selection problem.

5 Conclusions

Fuzzy quadratic programming problems are of utmost importance in an increasing variety of practical fields because real-world applications inevitably involve some degree of uncertainty or imprecision, for example in logistics management. One of these problems is the portfolio selection problem, where the imperfect knowledge of the returns on the assets and the uncertainty involved in the behaviour of financial markets may be introduced by means of fuzzy quantities and/or fuzzy constraints. In this context this paper has presented an operative and novel method for solving Fuzzy Quadratic Programming problems which is carried out by performing two phases that finally provide the user with a fuzzy solution. The method has been validated by solving a number of practical problems. The solutions obtained aid the authors to follow along this line of research to try to solve real problems in practice, in such a way that oriented Decision Support Systems involving Fuzzy Quadratic Programming problems can be built.

A parametric approach that solve a fuzzy quadratic problem was proposed. The set of optimal solutions obtained by one parametric approach to each $\alpha, \beta, \gamma \in [0, 1]$ constructs the fuzzy solution by using the Representation Theorem. The authors aim is to extend the line of investigation involving Fuzzy Quadratic Programming problems in order to try to solve other practical problems.

Acknowledgments

The authors want to thank the support provided by the Brazilian agency CAPES and the Spanish projects TIN2005-08404-C04-01 (70% of which are FEDER funds) and MINAS (TIC-00129-JA).

References

[1] AMMAR, E., KHALIFA, H.A.: Fuzzy portfolio optimization a quadratic programming approach. Chaos, Solitons and Fractals **18**, 1045–1054 (2003)

[2] BEALE, E.: On quadratic programming. Naval Research Logistics Quarterly **6**, 227–244 (1959)

[3] BECTOR, C.R., CHANDRA, S.: Fuzzy mathematical programming and fuzzy matrix games, *Studies in Fuzziness and Soft Computing*, vol. 169. Springer, Berlin (2005)

[4] BELLMAN, R.E., ZADEH, L.A.: Decision-marking in a fuzzy environment. Management Science **17**(4), B141–B164 (1970)

[5] FLOUDAS, C.A., PARDALOS, P., ADJIMAN, C., ESPOSITO, W.R., GÜMÜS, Z.H., HARDING, S.T., KLEPEIS, J.L., MEYER, C.A., SSHWEIGER, C.A.: Handbook of test problems in local and global optimization, *Nonconvex Optimization and its Applications*, vol. 33. Kluwer Academic Publishers, Dordrecht (1999)

[6] GOULD, N., TOINT, P.: A quadratic programming page. www.numerical.rl.ac.uk/qp/qp.html

[7] GOULD, N.I.M., TOINT, P.L.: A quadratic programming bibliography. RAL Numerical Analysis Group Internal Report 2000-1, Department of Mathematics, University of Namur, Bruxelles, Belgium (2000)

[8] HOCK, W., SCHITTKOWSKI, K.: Test examples for nonlinear programming codes, *Lecture Notes in Economics and Mathematical Systems*, vol. 187. Spring-Verlag (1981)

[9] INUIGUCHI, M., RAMIK, J.: Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem. Fuzzy Sets and Systems **111**(1), 3–28 (2000)

[10] JIMÉNEZ, F., CADENAS, J., SÁNCHEZ, G., GÓMEZ-SKARMETA, A., VERDEGAY, J.: Multi-objective evolutionary computation and fuzzy optimization. International Journal of Approximate Reasoning **43**(1), 59–75 (2006)

[11] LEON, T., LIERN, V., VERCHER, E.: Viability of infeasible portfolio selection problems: A fuzzy approach. European Journal of Operational Research **139**, 178–189 (2002)

[12] LIU, S.T.: Quadratic programming with fuzzy parameters: A membership function approach (2007). Chaos, Solitons and Fractals. Article in press. doi:10.1016/j.chaos.2007.07.054

[13] LIU, S.T.: Solving quadratic programming with fuzzy parameters based on extension principle. In: FUZZ-IEEE 2007, pp. 1–5. IEEE International Conference Fuzzy Systems, London, UK (2007)

[14] MARKOWITZ, H.: Portfolio selection. The Journal of Finance **7**(1), 77–91 (1952)

[15] MARKOWITZ, H.M.: Portfolio Selection: Efficient Diversification of Investments, second edn. Blackwell Publisher, Massachusetts (1991)

ISBN: 978-989-95079-6-8

Table 1: Fuzzy quadratic programming problem

$f(\mathbf{x})$	$\mathbf{x}_{initial}$	Constraints	Violation	Classic solution	
				$\overline{\mathbf{x}}^T$	$f(\mathbf{x})$
$9-8x_1-6x_3-4x_3+$ $+2x_1^2+2x_2^2+x_3^2+2x_1x_2+2x_1x_3$	$[0.5;0.5;0.5]^T$	$x_1+x_2-2x_3-3 \leq^f 0$ $x_i \geq^f 0, i=1,2,3$	$d_1=\widetilde{0.3}$	$[4/3;7/9;4/9]^T$	1/9

Table 2: Fuzzy portfolio selection problem

	#1	#2	#3	#4	#5	#6	#7	#8	#9
Year	Am.T	A.T&T.	U.S.S.	G.M.	A.T.&S.	C.C.	Bdm.	Frstn.	S.S.
1937	-0.305	-0.173	-0.318	-0.477	-0.457	-0.065	-0.319	-0.4	-0.435
1938	0.513	0.098	0.285	0.714	0.107	0.238	0.076	0.336	0.238
1939	0.055	0.2	-0.047	0.165	-0.424	-0.078	0.381	-0.093	-0.295
1940	-0.126	0.03	0.104	-0.043	-0.189	-0.077	-0.051	-0.09	-0.036
1941	-0.28	-0.183	-0.171	-0.277	0.637	-0.187	0.087	-0.194	-0.24
1942	-0.003	0.067	-0.039	0.476	0.865	0.156	0.262	1.113	0.126
1943	0.428	0.300	0.149	0.255	0.313	0.351	0.341	0.580	0.639
1944	0.192	0.103	0.260	0.290	0.637	0.233	0.227	0.473	0.282
1945	0.446	0.216	0.419	0.216	0.373	0.349	0.352	0.229	0.578
1946	-0.088	-0.046	-0.078	-0.272	-0.037	-0.209	0.153	-0.126	0.289
1947	-0.127	-0.071	0.169	0.144	0.026	0.355	-0.099	0.009	0.184
1948	-0.015	0.056	-0.035	0.107	0.153	-0.231	0.038	0	0.114
1949	0.305	0.030	0.133	0.321	0.067	0.246	0.273	0.223	-0.222
1950	-0.096	0.089	0.732	0.305	0.579	-0.248	0.091	0.650	0.327
1951	0.016	0.090	0.021	0.195	0.040	-0.064	0.054	-0.131	0.333
1952	0.128	0.083	0.131	0.390	0.434	0.079	0.109	0.175	0.062
1953	-0.010	0.035	0.006	-0.072	-0.027	0.067	0.21	-0.084	-0.048
1954	0.154	0.176	0.908	0.715	0.469	0.077	0.112	0.756	0.185

$$\mathbf{E}(\mathbf{R}) = \begin{bmatrix} 0.0659 & 0.0616 & 0.1461 & 0.1734 & 0.1981 & 0.0551 & 0.1276 & 0.1348 & 0.1156 \end{bmatrix} \tag{10}$$

$$\Sigma = \begin{pmatrix} 0.0565 & 0.0228 & 0.0303 & 0.0518 & 0.0172 & 0.0341 & 0.0257 & 0.0464 & 0.0383 \\ 0.0228 & 0.0155 & 0.0199 & 0.0259 & 0.0085 & 0.0106 & 0.0153 & 0.0265 & 0.0221 \\ 0.0303 & 0.0199 & 0.0905 & 0.0663 & 0.0470 & 0.0141 & 0.0111 & 0.0836 & 0.0445 \\ 0.0518 & 0.0259 & 0.0663 & 0.1011 & 0.0546 & 0.0307 & 0.0220 & 0.0775 & 0.0388 \\ 0.0172 & 0.0085 & 0.0470 & 0.0546 & 0.1354 & 0.0136 & 0.0221 & 0.0683 & 0.0476 \\ 0.0341 & 0.0106 & 0.0141 & 0.0307 & 0.0136 & 0.0437 & 0.0119 & 0.0254 & 0.0229 \\ 0.0257 & 0.0153 & 0.0111 & 0.0220 & 0.0221 & 0.0119 & 0.0305 & 0.0229 & 0.0184 \\ 0.0464 & 0.0265 & 0.0836 & 0.0775 & 0.0683 & 0.0254 & 0.0229 & 0.1024 & 0.0553 \\ 0.0383 & 0.0221 & 0.0445 & 0.0388 & 0.0476 & 0.0229 & 0.0184 & 0.0553 & 0.0839 \end{pmatrix} \tag{11}$$

ISBN: 978-989-95079-6-8

Table 3: Results of the first phase of the hypothetical problem.

γ/α	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
1.0	0.1111	0.1045	0.0982	0.0920	0.0860	0.0803	0.0747	0.0693	0.0642	0.0592	0.0544
0.9	0.1037	0.0973	0.0911	0.0851	0.0794	0.0738	0.0684	0.0632	0.0583	0.0535	0.0489
0.8	0.0958	0.0896	0.0836	0.0779	0.0723	0.0669	0.0618	0.0568	0.0521	0.0475	0.0432
0.7	0.0874	0.0814	0.0757	0.0701	0.0648	0.0597	0.0548	0.0501	0.0456	0.0413	0.0372
0.6	0.0784	0.0727	0.0672	0.0620	0.0569	0.0521	0.0474	0.0430	0.0388	0.0348	0.0311
0.5	0.0689	0.0635	0.0583	0.0534	0.0486	0.0441	0.0398	0.0357	0.0319	0.0282	0.0248
0.4	0.0588	0.0538	0.0490	0.0444	0.0400	0.0359	0.0320	0.0283	0.0248	0.0216	0.0186
0.3	0.0483	0.0437	0.0393	0.0351	0.0312	0.0275	0.0240	0.0208	0.0178	0.0150	0.0125
0.2	0.0374	0.0333	0.0294	0.0258	0.0224	0.0192	0.0163	0.0136	0.0112	0.0090	0.0070
0.1	0.0265	0.0230	0.0197	0.0167	0.0139	0.0114	0.0092	0.0072	0.0054	0.0039	0.0026
0.0	0.0160	0.0132	0.0108	0.0085	0.0066	0.0048	0.0034	0.0022	0.0013	0.0006	0.0002

Table 4: Results of the first phase of the portfolio selection problem.

γ/α	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
1.0	0.0343	0.0332	0.0322	0.0313	0.0305	0.0297	0.0290	0.0284	0.0278	0.0273	0.0268
0.9	0.0305	0.0298	0.0291	0.0286	0.0281	0.0275	0.0270	0.0266	0.0261	0.0257	0.0253
0.8	0.0282	0.0277	0.0272	0.0268	0.0263	0.0259	0.0255	0.0251	0.0247	0.0243	0.0239
0.7	0.0265	0.0261	0.0257	0.0253	0.0249	0.0245	0.0241	0.0238	0.0234	0.0230	0.0227
0.6	0.0266	0.0247	0.0243	0.0240	0.0240	0.0236	0.0233	0.0226	0.0223	0.0220	0.0216
0.5	0.0252	0.0249	0.0245	0.0245	0.0229	0.0234	0.0223	0.0216	0.0213	0.0210	0.0207
0.4	0.0240	0.0237	0.0234	0.0230	0.0227	0.0225	0.0222	0.0218	0.0215	0.0202	0.0209
0.3	0.0232	0.0228	0.0225	0.0222	0.0218	0.0215	0.0212	0.0209	0.0206	0.0203	0.0200
0.2	0.0220	0.0217	0.0215	0.0212	0.0209	0.0206	0.0204	0.0201	0.0198	0.0196	0.0193
0.1	0.0212	0.0209	0.0207	0.0204	0.0201	0.0199	0.0196	0.0194	0.0192	0.0189	0.0187
0.0	0.0204	0.0202	0.0199	0.0197	0.0195	0.0192	0.0190	0.0188	0.0186	0.0184	0.0182

[16] MAROS, I., MÉSZÁROS, C.: A repository of convex quadratic programming problems. Department Technical Report DOC 97/6, Department of Computing, Imperial College, London, UK (1997)

[17] SCHITTKOWSKI, K.: More test examples for nonlinear programming codes, *Lecture Notes in Economics and Mathematical Systems*, vol. 282. Spring-Verlag (1987)

[18] SILVA, R.C., VERDEGAY, J.L., YAMAKAMI, A.: Two-phase method to solve fuzzy quadratic programming problems. In: FUZZ-IEEE 2007, pp. 1–6. IEEE International Conference Fuzzy Systems, London, UK (2007)

[19] TANAKA, H., GUO, P., TÜRKSEN, B.: Portfolio selection based on fuzzy probabilities and possibity distributions. Fuzzy Sets and Systems **111**, 387–397 (2000)

[20] TANG, J., WANG, D.: An interactive approach based on a genetic algorithm for a type of quadratic programming problems with fuzzy objective and resources. Computers and Operations Research **24**(5), 413–422 (1997)

[21] VERCHER, E., BERMÚDEZ, J.D., SEGURA, J.V.: Fuzzy portfolio optimization under downside risk measures. Fuzzy Sets Syst. **158**(7), 769–782 (2007)

[22] WATADA, J.: Fuzzy portfolio selection and its applications to decision making. Tatra Mountains Mathematics Publication **13**, 219–248 (1997)

[23] WERNERS, B.: Interative multiple objective programming subject to flexible constraints. European Journal of Operational Research **31**, 342–349 (1987)

[24] WOLFE, P.: The simplex method for quadratic programming. Econometrica **27**, 382–398 (1959)

[25] XU, Z., DA, Q., CHEN, Q.: Priority approach based on quadratic programming model to fuzzy preference relation. Journal of Southeast University (English Edition) **21**(1), 108–110 (2005)

[26] ZIMMERMANN, H.J.: Fuzzy programming and linear programming with several objective functions. Fuzzy Sets and Systems **1**, 45–55 (1978)

[27] ZIMMERMANN, H.J.: Fuzzy mathematical programming. Computers & Operations Research **10**(4), 291–298 (1983)

ISBN: 978-989-95079-6-8

Interval-Valued Fuzzy System for Segmentation of Prostate Ultrasound Images

Miguel Pagola[1] Edurne Barrenechea[1] Aranzazu Jurio[1] Mikel Galar[1] Pedro Couto[2]
Farhang Sahba[3] Hamid R. Tizhoosh[4]
1.Automática y Computación, Universidad Pública de Navarra, Pamplona (Spain)
Email: miguel.pagola@unavarra.es
2.CITAB, Universidade Tras-os-Montes e Alto Douro, Vilareal, (Portugal)
Email: pcouto@utad.pt
3.University of Toronto, Toronto (Canada)
Email: farhang@ieee.org
4.Systems Design Engineering, University of Waterloo, Ontario (Canada)
Email: tizhoosh@uwaterloo.ca

Abstract— *In this paper we introduce an application of interval-valued systems to the segmentation of prostate ultrasound images. The system classifies each pixel as prostate or background. The input variables are the values of each pixel in different processed images as proximity, edginess and enhanced image. The system has 20 rules and is trained with ideal images segmented by an expert. Interval-valued fuzzy systems have been used due to their potential to capture uncertainty in a more robust way compared to ordinary fuzzy systems.*

Keywords— Interval-valued fuzzy sets, ultrasound image, image segmentation

1 Introduction

Ultrasound images are used in clinical settings to evaluate anomalies, tissues and organs. Ultrasound imaging is a quite common modality due to its low cost, portability and harmlessness to human body. Transrectal ultrasound images of the prostate in male patients are frequently used for both diagnosis and treatment purposes. Detecting the outline of the prostate – prostate segmentation – is therefore the first step regardless the following step is of diagnostic nature or treatment planning as for brachytherapy of prostate cancer.

Ultrasound image segmentation is strongly affected by the lower quality of these images. Speckle noise, shadows and prostate inhomogeneity make the segmentation a complicated task. There exist several methods to segment ultrasound images, a complete review can be found in [12]. Some methods use machine learning techniques to extract the objects/lesions. In those methods ideally segmented images created by an expert radiologist are used to train the system. In [18] reinforcement learning is used to train an agent devoted to segment prostate ultrasound images. Zhang et al. [20] optimize the weights of a weighed wavelet to detect microcalcifications in mamographic images. Neural networks were used in [13] to identify possible injuries in liver images and in [5] a genetic algorithm optimizes the weights of a self-organized neural network (Kohonen net).

One of the problems of the Neural Networks is the black box reasoning. In medical applications it is very useful to know how a certain task has been done the process of segmenting the image. Hence, we propose to use a fuzzy logic system to segment ultrasound images since they provide a convenient way of interpreting the tasks/results.

Due to the complexity of the ultrasound images we are going to use interval-valued fuzzy sets enabling us to represent the uncertainty that is within these images. In [9] Mendel proposes an adaptation of the fuzzy rule learning algorithm to interval type 2 fuzzy rules. Note what he called interval type 2 fuzzy sets are the same as interval-valued fuzzy sets in some cases (when $a = 1$, please see section 2). The objective of this paper is to develop an interval-valued fuzzy logic system (from now IT2FLS following the notation given in [9]) to segment prostates in transrectal ultrasound images.

Why to use interval-valued fuzzy systems instead of classical fuzzy systems? When we design a fuzzy system that is going to be trained using typical machine learning algorithms, we must choose the input variables, the output of our system and the number of rules. If we train the system with the past data, we can understand this process as function fitting problem, in which the fuzzy rule base system is a parametrized function and the training process is the modification of said parameters. This means that the output of the system fits the training data. If the system uses interval-valued fuzzy sets instead of classical fuzzy sets, the number of parameters to define an interval fuzzy rule is larger than the one needed for a classical fuzzy rule. So, in the training process, with the same data, the interval fuzzy system has more parameters, more degrees of freedom, which means it can be adjusted in a better way to the data. Some researchers suggest that this is an advantage compared to ordinary fuzzy systems. Therefore it means that for the same linguistic complexity (number of rules and number of variables) interval-valued fuzzy systems can, at least in theory, achieve better accuracy. We have made a comparative study to verify this hypotheses.

This work is organized as follows: first we present an introduction of interval-valued fuzzy logic systems. In section 3 we review Mendel's algorithm to generate an IT2FLS from training data. Later, in section 4 we present the model that we propose to segment ultrasound images. Finally we show some experimental results, conclusions and future research.

2 Interval-valued fuzzy systems

An interval-valued fuzzy set constitutes that the membership degree of every element to the set is given by a closed subinterval of interval [0,1]. The concept of type 2 fuzzy sets was introduced by Zadeh [19] as a generalization of an ordinary

ISBN: 978-989-95079-6-8

fuzzy set. The membership degree of an element to a type 2 fuzzy set is a fuzzy set in [0,1].

An interval type 2 fuzzy set $\overline{\overline{A}}$ in U is defined as

$$\overline{\overline{A}} = \{(u, A(u), \mu_u(x)) | u \in U, A(u) \in L([0,1])\},$$

where $A(u) = [\underline{A}(u), \overline{A}(u)]$ is a membership function; i.e., a closed subinterval is $[0,1]$, and function $\mu_u(x)$ represents the fuzzy set associated with the element $u \in U$ obtained when x is within $[0,1]$; $\mu_u(x)$ is given in the following way:

$$\mu_u(x) = \begin{cases} a \text{ if } \underline{A}(u) \leq x \leq \overline{A}(u) \\ 0 \text{ otherwise} \end{cases},$$

where $0 \leq a \leq 1$.

In [8]–[10], it is proved that an interval type 2 fuzzy set is the same as an interval-valued fuzzy set if $a = 1$.

Example: We can represent an interval-valued fuzzy set by means of an upper bound membership function and a lower bound membership function. In this work, all of the membership functions are going to be represented by a Gaussian function:

$$\mu_k^l(x_k) = exp\left[-\frac{1}{2}\left(\frac{x_k - m_k^l}{\sigma_k^l}\right)\right] \ \sigma_k^l \in [\sigma_{k1}^l, \sigma_{k2}^l] \quad (1)$$

In fig. 1 we show the membership function with parameter $m_k^l = 0.4$. We take $\sigma_{k1}^l = 0.05$ to represent the lower bound and $\sigma_{k2}^l = 0.1$ for the upper bound.

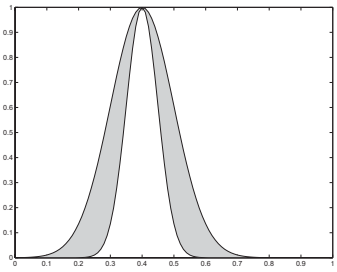


Figure 1: Interval-valued membership function with a Gaussian primary membership function and $m = 0.4$, $\sigma_{k1}^l = 0.05$ y $\sigma_{k2}^l = 0.1$.

An IT2FLS is a rule-based system in which the sets that represent the antecedents and consequents are interval-valued fuzzy sets (or interval type 2 fuzzy sets). In fig. 2 we depict the most important modules of an IT2FLS (see [9]).

The input of the system is a set of values. The fuzzfier transforms the inputs into interval-valued fuzzy sets. Then, the main part of the system, using the rules carries out the inference to generate conclusions, represented by interval-valued fuzzy sets. To use these conclusions (decisions) in the real world, the defuzzifier transforms these sets into crisp values. Commonly, before defuzzification, we can reduce the sets from interval-valued fuzzy sets to classical fuzzy sets, such a way typical defuzzifying techniques could be used to obtain crisp values.

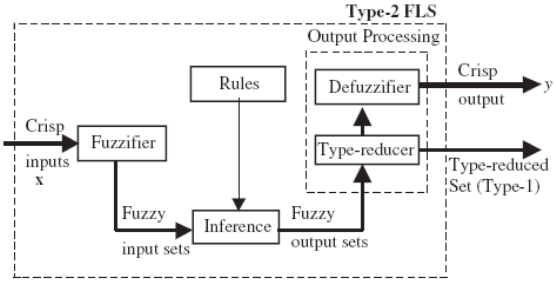


Figure 2: Interval Type 2 Fuzzy Logic System

Considering an IT2FLS with p inputs $x_1 \in X_1, \cdots, x_p \in X_p$ and one output $y \in Y$. Then we assume that the system has M rules in the following way:

$$R^l: \text{ IF } x_1 \text{ is } F_1^l \text{ AND } \ldots \text{ AND } x_p \text{ is } F_p^l, \text{ THEN } y \text{ is } G^l \text{ con } l = 1, \ldots, M \quad (2)$$

where $F_1^l, \ldots, F_p^l, G^l$ are interval-valued fuzzy sets. Each rule is interpreted as an implication:

$$R^l : F_1^l \times F_2^l, \cdots \times F_p^l \rightarrow G^l = A^l \rightarrow G^l \text{ with } l = 1, \ldots, M \quad (3)$$

R^l is described by the membership function $\mu_{R^l}(\mathbf{x}, y) = \mu(x_1, \ldots, x_p, y)$ where,

$$\mu_{R^l}(\mathbf{x}, y) = \mu_{A^l \rightarrow G^l}(\mathbf{x}, y) = \left[\bigcap_{i=1}^{p} \mu_{F_i^l}(x_i)\right] \bigcap \mu_{G^l}(y) \quad (4)$$

Using the extension of the Zadeh's compositional rule to interval-valued fuzzy sets, the consequent is calculated as follows:

$$\mu_{B^l}(y) = \mu_{A^l \circ R^l}(y) = \bigcup_{x \in X} \left[\mu_{A_x}(\mathbf{x}) \bigcap \mu_{R^l}(\mathbf{x}, y)\right] \text{ with } y \in Y, l = 1, \ldots, M \quad (5)$$

There exist several works regarding the inference with interval-valued fuzzy rules [2, 6], but in this work we are going to use the method proposed by Mendel in [9] (for a more detailed study see [10]).

3 Design of Interval-valued fuzzy systems from data

In this section we show how we can design an IT2FLS from training data.

Given a collection of N pairs of input-output data $(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}), \cdots (\mathbf{x}^{(N)}, y^{(N)})$ where $\mathbf{x}^{(t)}$ is the input vector and $y^{(t)}$ is the output value of the t training pair, we define the error of the IT2FLS for the t-th input as:

$$e^{(t)} = f_{s2}(\mathbf{x}^{(t)}) - y^{(t)} \quad t = 1, \ldots, N \quad (6)$$

where f_{s2} is the output of the IT2FLS. Such a value depends on the parameters that define the interval-valued fuzzy sets of the antecedents and consequents of the corresponding rules (m_k^l, σ_{k1}^l, σ_{k2}^l, y_l^j and y_r^j, with $k = 1, \ldots, p$, for p inputs, $j = 1, \ldots, M$, for M rules). Also y_l^j and y_r^j represent the bounds of the membership function of the consequent after

ISBN: 978-989-95079-6-8

the type reduction.
We define the quadratic error of the IT2FLS in t-th input as:

$$E^{(t)} = \frac{1}{2}[e^{(t)}]^2 \quad t = 1, \ldots, N \tag{7}$$

and the performance of the system as

$$E = \sum_{i=1}^{N} E^{(t)} \quad t = 1, \ldots, N \tag{8}$$

The training of an IT2FLS aims at finding the optimal values of the parameters that define the system such that equations (7) or (8) are minimized. If we minimize equation (7) (such a method is explained in [9]), the learning process is done in an incremental way, because the parameters are adjusted iteratively after processing every training pair. If we minimize equation (8), as we have done in this work, the learning process is done as a batch processing, because the parameters are adjusted after processing all of the training pairs. When we process all of the N training pairs and modify the parameters it is called a training epoch. In this work, in every epoch the parameters are adjusted proportional to the error gradient $\frac{\partial E}{\partial m_k^l}$, $\frac{\partial E}{\partial \sigma_{k1}^l}$, $\frac{\partial E}{\partial \sigma_{k2}^l}$, $\frac{\partial E}{\partial y_l^j}$ and $\frac{\partial E}{\partial y_r^j}$.

This quantity is calculated using the *resilient backpropagation* algorithm (RPROP, [14]). This algorithm automatically adjusts its own parameters during the training process, and is very easy to implement and can achieve a high convergence speed.

4 Proposed method

The objective of our method is to classify each pixel of the image. That is, decide if each pixel belongs to the important area (prostate) to segment or if it belongs to the background. Since the size and the position of the prostate are subject to change, the user must select the central point of the region (also called the seed point) to avoid false detections (and to focus on the segmentation itself). Therefore, the method is semi-supervised due to user suppling some relevant information. However, automation of this step is quite possible and has been repeatedly reported in literature.

The IT2FLS that we propose has 5 inputs, one output and consists of 20 rules. The inputs are:

1. pos_i: The distance, in pixels, of the pixel considered from the central point in the horizontal axes.
2. pos_j: The distance, in pixels, of the pixel considered from the central point in the vertical axes.
3. $dist$: The proximity of the pixel w.r.t. to the central point, calculated via a flooding algorithm.
4. edg: Edginess of each pixel.
5. mgr: Average gray level of the pixels neighborhood (e.g. 5×5 neighborhoods) in the enhanced image.

In the following subsections we present the techniques used to obtain the values of enhancement, proximity and edginess.

4.1 Enhanced image

The algorithm used to enhance ultrasound images is the one proposed by Sahba et al. [15, 16, 17], in which fuzzy rules such as the following have been used:

IF the pixel does not belong to the prostate,
THEN leave it unchanged
IF the pixel belongs to the prostate AND is dark,
THEN make it darker
IF the pixel belongs to the prostate AND is gray,
THEN make it dark
IF the pixel belongs to the prostate AND is bright,
THEN make it brighter

We use a simplified version of these rules in form of

IF the pixel *belongs* to the object AND is *dark*,
THEN make it *darker*,

or in a even more simple formulation and to save time we can use rules such as:

IF the pixel *belongs* to the object,
THEN make it *darker*,

where the degree of "belonging" of each pixel to the object is a function of its distance to the central point of the object or the inside of an initial/coares segment as proposed by Sahba et al. The main idea of enhancement is to eliminate the noise in the images and enhance the gray levels of selected area (regional contrast enhancement). First the noise is eliminated using a median filter (7×7 or 9×9). Then each pixel is fuzzyfied depending on its intensity with a membership function that is constructed taking into account the mean level of gray of the surroundings and the position of the selected point.

4.2 Proximity image

The proximity image represents the proximity of every pixel to the central point (similar to [7, 15, 16]), but taking into account the edges that separate the different regions of the original image. First we calculate the edges of the enhanced image using the Canny algorithm [4]. Then, starting from the central pixel selected by the user the algorithm labels the pixels with their distance to the central pixel step by step. In the first step the neighbors are labeled with distance 1 and so on (Fig. 3). The pixels marked as edge by the Canny edge detector are used as walls and cannot be labeled, so the proximity values generated are related with areas of the image.

4.3 Edginess

To create the edginess image, we calculate what is commonly called *false edges*. In [3] we presented a method to obtain false edges by means of t-norms and t-conorms. For every pixel a neighborhood matrix is constructed (3×3, 5×5, etc.). Applying t-norms to the elements of the matrix we obtain a lower bound of an interval. Applying t-conorms we obtain the upper bound of that interval. We called the length of the interval, that is the difference between the upper bound and the lower bound of the interval, false edge. The most known case is obtained when using minimum as the t-norm and maximum as the t-conorm.

ISBN: 978-989-95079-6-8

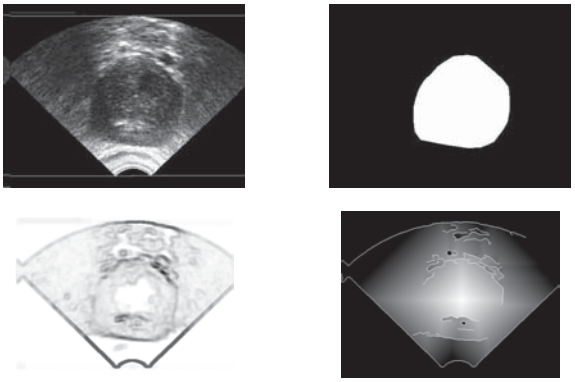


Figure 3: Top left to bottom right: Original image, ideally segmented image, edginess image, proximity image

5 Experimental results

To evaluate the performance of the IT2FLS we have a set of ten prostate ultrasound images with their corresponding ideal segmentation created by an expert. Each image has a central point of the prostate provided by the expert (for many prostate images automated detection of central point is relatively easy [1]). We use two of these images for the training and the other six for the validation. From the two training images we select randomly 200 pixels which are the training data pairs. This training data set is used to adjust the parameters of the system as described in section 3. The data is also split into two groups, the training set (80% of the data pairs) and the validation set (20%). To evaluate the performance of the IT2FLS we use an overlap measure S_A between the areas of the IT2FLS result and the ideally segmented image.

$$S_A = \frac{|Ideal \cap IBF|}{|Ideal \cup IBF|}, \tag{9}$$

where $Ideal$ is the binary image segmented by the expert, IBF is the binary image obtained using our method, $\cap$ and $\cup$ are the intersection and the union between crisp sets, respectively. Due to the initial values of the parameters of the system are selected randomly before training, we can obtain different possible solutions. In Table 1 we show the best ones after 20 trials. In Fig. 4 and 5 we show the binary images obtained by IT2FLS.

Table 1: Area overlap of segmented images with ideally segmented prostates.

Image	1	2	3	4	5	6	7	8	9	10
IVFS S_A(%)	84	61	74	75	77	75	62	64	71	72
Fuzzy S_A	86	67	77	78	77	80	71	72	70	73

Table 2: Percentage of convergence.

System	Percentage of convergence
Fuzzy	90 %
IVFS	30 %

In the experimental results we show that interval-valued fuzzy rule systems perform similar to classical fuzzy systems. The point is that the convergence of IVFS systems is really poor and also the mean performance achieved by these systems is a bit worse than ordinary fuzzy sets. It means that the extra adjustable parameters, if we dont use a specific learning algorithm with the IVFS system, are not worth in this case.

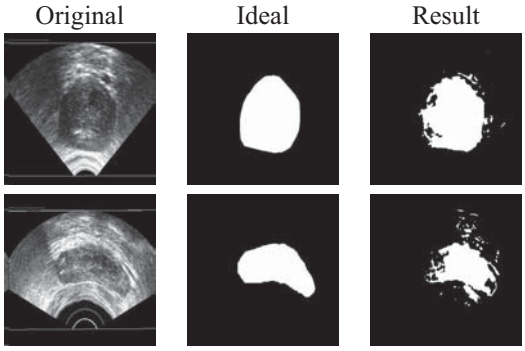


Figure 4: Comparison between ideally segmented images and IT2FLS segmented images used for training.

6 Conclusions and future research

We have proposed a new method to segment ultrasound images using an interval-valued fuzzy system. The system has 5 inputs, 20 rules and one output and can be trained using images previously segmented by an expert. An average overlap of 72% between the segmented area and the ideal segmentation has been reached. Nevertheless the results of the classical fuzzy systems are a bit better than the IVFSs system, mainly due to poor convergence of the learning algorithm.Taking into account the size of the proposed system on one hand and the challenging nature of prostate ultrasound segmentation on the other hand, the results can be regarded as promising.

The results show poor spatial consistency, which could be improved by adding some constraints or rules regarding shape and boundaries of the areas to the system. Also, we want to obtain interval input values to capture the uncertainty existing in the ultrasound images and deal with it via the IT2FLS in order to obtain better results.

Acknowledgment - This paper has been partially supported by the National Science Foundation of Spain, Reference TIN2007-65981

References

[1] J. Awad, Prostate Segmentation and Regions of Interest Detection in Transrectal Ultrasound Images. PhD Thesis, University of Waterloo, Canada, 2007.

[2] H. Bustince, “Indicator of inclusion grade for interval-valued fuzzy sets. Application to approximate reasoning based on interval-valued fuzzy sets”, Int. Journal of Approximate Reasoning, vol. 23, pp. 137–209, 2000.

[3] H. Bustince, E. Barrenechea, M. Pagola, R. Orduna, Construction of interval type 2 fuzzy images to represent images in grayscale. False edges. FUZZY IEEE 2007.

[4] J. Canny, A Computational approach to edge detection, IEEE Transactions on Pattern Recognition and Machine Intelligence, 8(6), 679-698 (1986).

ISBN: 978-989-95079-6-8

[5] Z. Dokur and T. Olmez, Tissue segmentation in ultrasound images by using genetic algorithms, *Expert Systems with Applications*, 34(4), 2739-2746 (2008)

[6] M.B. Gorzalczany, "A method of inference in approximate reasoning based on interval-valued fuzzy sets", *Fuzzy Sets and Systems*, vol. 21, pp. 1–17, 1987.

[7] M. Mancas, B. Gosselin, B. Macq, Segmentation Using a Region Growing Thresholding. In "Image Processing: Algorithms and Systems IV". Edited by Dougherty, Edward R.; Astola, Jaakko T.; Egiazarian, Karen O. Proceedings of the SPIE, Volume 5672, pp. 388-398, 2005.

[8] J.M. Mendel, R. I. Bob John. Type-2 fuzzy sets made simple, *IEEE Transactions on Fuzzy Systems*, **10**(2), pages 117–127, 2002.

[9] J.M. Mendel. Uncertain Rule-Based Fuzzy Logic Systems, *Prentice-Hall, Upper Saddle River*, NJ, 2001.

[10] J.M. Mendel. Advances in type-2 fuzzy sets and systems, *Information Sciences*, Volume 177(1), pp. 84–110, 2007.

[11] S. Mitra, Y. Hayashi, Neuro-fuzzy rule generation: survey in soft computing framework, IEEE Transactions on Neural Networks, 11(3), 748 - 768 May (2000).

[12] J.A. Noble and D. Boukerroui, Ultrasound image segmentation: A survey, IEEE Transactions on Medical Imaging 25 (2006), pp. 987–1010.

[13] K. Ogawa, M. Fukushima, K. Kubota and N. Hisa, Computer-aided diagnostic system for diffuse liver-diseases with ultrasonography by neural networks, IEEE Trans. on Nuclear Science 45(6) 3069-3074, 1998.

[14] M. Riedmiller, H. Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco,1993

[15] F. Sahba, H.R. Tizhoosh, M.M. Salama, A coarse-to-fine approach to prostate boundary segmentation in ultrasound images, BioMedical Eng. OnLine 2005, 4:58

[16] F. Sahba, H.R. Tizhoosh, M.M.A. Salama, Segmentation of Prostate Boundaries Using Regional Contrast Enhancement. The IEEE International Conference on Image Processing (ICIP), Volume 2, pp. 1266-9, 2005.

[17] H.R. Tizhoosh, Observer-Dependent Image Enhancement, Proceeding of FUZZ-IEEE 2001.

[18] H.R. Tizhoosh. Image thresholding using type II fuzzy sets, *Pattern recognition*, 38(12), pp. 2363-2372, 2005.

[19] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning – I, *Information Science*, **8**, 199–249, 1975.

[20] W. Zhang, H. Yoshida, R.M. Nishikawa, K. Doi, Optimally weighted wavelet transform based on supervised training for detection of microcalcs in digital mammograms, Medical Physics 25(6), 1998, pp. 949–956.

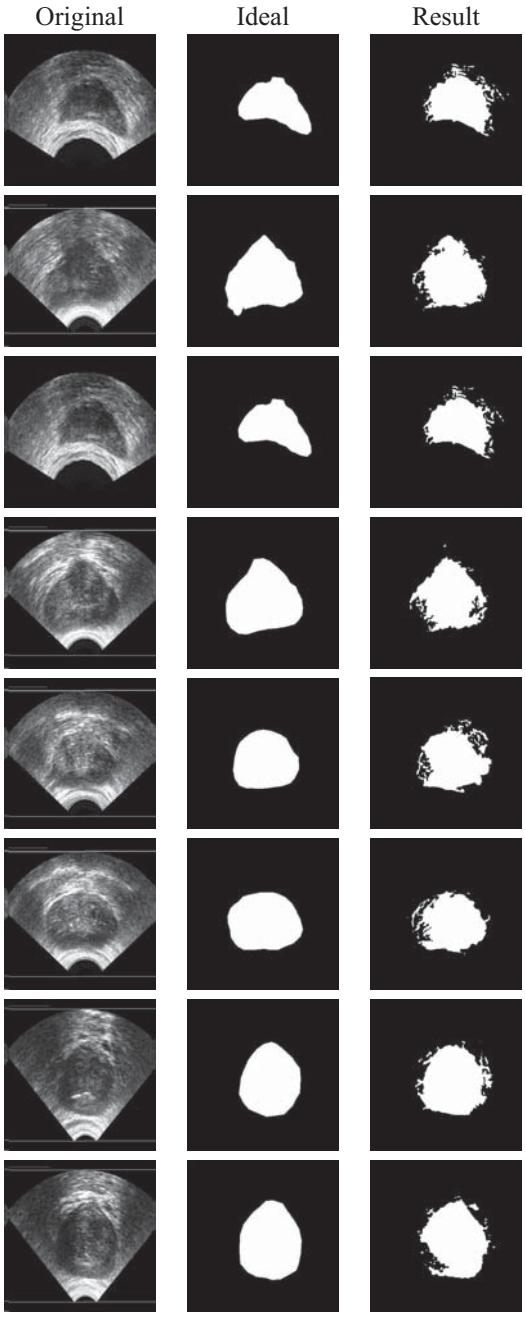


Figure 5: Comparison between ideally segmented images and IT2FLS segmented images. These are test images.

ISBN: 978-989-95079-6-8

A toward framework for generic uncertainty management

Ernesto Damiani Paolo Ceravolo Marcello Leida

Università degli Studi di Milano, Dipartimento di Tecnologie dell'Informazione
via Bramante 65, 26013 Crema (CR), Italy
Email: {ernesto.damiani, paolo.ceravolo, marcello.leida}@unimi.it

Abstract*— *The need for an automatic inference process able to deal with information coming from unreliable sources is becoming a relevant issue both on corporate networks and on the open Web. Mathematical theories to reason with uncertain information have been successfully applied in several situations, but each one of these models is tailored to deal with a specific semantics of uncertainty. In this paper, we put forward the idea of using explicit representations of the different types of uncertainty for partitioning the inference process into parts. By coordinating multiple independent reasoning processes, we are sometimes able to apply a specific model to each type of uncertain information, and recombine the final results via a suitable reconciliation process. We validated our approach applying it to the classic* ***schema matching problem*, and using the* ***Ontology Alignment Evaluation Initiative****, (OAEI) tests to assess the results.*

Keywords— Uncertainty, Ontology Matching, Reasoning, Rules.

1 Introduction

The problem of automatic inference is one of the most challenging problem in computer science [1], becoming even harder when knowledge is uncertain, due to lack of reliability of the source of information, approximation, dependencies and other factors. While many mathematical models for reasoning on uncertain information have been proposed, the general problem of handling and interpretation of uncertain knowledge is still to be solved. In this paper, we put forward the idea of using explicit representations of the different uncertainties present in the knowledge base according to different uncertainty models, coordinating multiple independent reasoning processes. By splitting the inference process into parts, we are able to apply a specific model to each type of uncertain information, recombining the final results via a suitable reconciliation process. Although the interoperability among multiple inference models been studied [2, 3], in the literature we are not aware of any hybrid reasoning processes which can handle the flexible integration of different models. As a proof of concept of this approach we present a semantics-aware matching strategy, that we apply to the well-known problem of ontology alignment [4]. The paper is structured as follows: Section 2 introduces the problem of uncertain information in knowledge management, briefly presenting the different types of uncertainty and the mathematical models used for the inference process. The section also introduces the need for an explicit representation of the various types of uncertainty, referring to the Ontology of Uncertainty [5] proposed by W3C's UR3W-XG incubator group[1]. Section 3 presents a case study, applying our technique to the classic *schema matching problem*, testing it via the Ontology Alignment Evaluation Initiative (OAEI) tests and comparing the results to the participants to the OAEI 2007 contest [4]. Conclusions and future work on our framework are presented in Section 5.

[1] http://www.w3.org/2005/Incubator/urw3/

2 Uncertain Information Representation and Reasoning

Experience has shown that the open Web and other platforms for hosting user-generated content can provide little quality control at content production time. As a result, most publicly available information can be considered *uncertain* to some degree. In order to clarify the notion of uncertainty, it is important to distinguish between **degrees of truth** and **degrees of uncertainty** in the information [6]. A degree of truth can be defined as the degree of compatibility between a statement and a knowledge base, which is limited to what the system knows about reality: a statement S is true if this assumption agrees with the set of statements in the knowledge base. Instead, Uncertainty of a statement arises when the knowledge base does not provide sufficient information to decide if a statement is true or false. Therefore uncertainty falls at a meta-level with respect to truth [7]. In case of truth values, we briefly mention two major theories [7]: *Classical two-valued logic* and *Fuzzy Logic*. In the first case, a statement's truth value can only assume one of two values [8], namely `0` or `1`. In the second case [9], truth values belong to the entire interval $[0, 1]$. For uncertainty representation, we distinguish between *Probability* and *Possibility theory*. The degree of probability associated to a statement is a typical example of gradual uncertainty. In sentences like: *"The player tossing a coin* wins *with 50% probability"*, "The player *wins*" is a true statement: it cannot happen that a player "half wins". The "50%" at the end of the sentence is not the statement truth value, but its level of uncertainty. *Possibility theory* is an alternative to probability theory, which separates the uncertainty of statements in possibilities and necessities [7]. Uncertainty can be classified as **Epistemic**, if it comes from the limited knowledge of the agent that generates the assertion or **Aleatory** if it is intrinsic in the observed world. Depending on the features of the agent that generates uncertain statements, is possible to identify two different types of uncertainty: **Objective** if the uncertainty derives from a repeatable observation and **Subjective** if the uncertainty in the information is derived from an informal evaluation. Furthermore, uncertainty can depend on the type of statement it is associated to: **Ambiguous**, **Inconsistent**,**Vague**, **Incomplete** and **Empiric**.

For our purposes, uncertainty can be represented as an annotation about a statement, expressing the level of certainty about it. We shall call "uncertain information" the triple

ISBN: 978-989-95079-6-8

(S, t, l) formed by a statement S, its truth value t and its corresponding uncertainty level l. It is important to remark that in many practical scenarios, we may encounter statements whose uncertainty levels have diverse semantics, especially when uncertain information is generated by multiple unsupervised processes. For example, Web-based weather forecasting services provide uncertain information in different forms; the uncertainty can be possibilistic or probabilistic (*Cloudy:"50%"*, *Rain:"10%"*) while truth value ranges can be classical (*Rainy*, *Fair*) or fuzzy (*Partially Covered*, *Heavy Rain*). Traditional approaches to uncertain reasoning support extensions of logic models dealing with inference on statements and their truth values, including mathematical theories able to deal with the uncertainty levels. In many cases, however, handling uncertainty hs an impact on the complexity and even the decidability of the inference problem. Scientific areas very active in the integration of classical logic with mathematical theories dealing with uncertainty is Artificial Intelligence and Knowledge Representation. Recently the effort was concentrated on the languages for the Web, such as in particular Semantic Web standards. For example, several probabilistic extensions of *Description Logics* (DL) [10] and *First Order Logic* (FOL) are available. Here, we shall focus the discussion on Description Logics, which are the logical model underpinning the OWL-Lite and OWL-DL ontology languages [11] used on the Semantic Web [12]. OWL-Lite and OWL-DL correspond respectively to $\mathcal{SHIF}(\mathbf{D})$ and $\mathcal{SHOIN}(\mathbf{D})$ Description Logics respectively, which are known to be tractable. A sound and complete fragment of $\mathcal{SROIQ}(\mathbf{D})$ is $\mathcal{SHOQ}(\mathbf{D})$ [13]. This fragment can be extended to handle uncertainty; in [14] Lukasiewicz defines a Probabilistic Description Logic $\mathcal{P} - \mathcal{SHOQ}(\mathbf{D})$. Based on Lukasiewicz's Probabilistic Description Logics, Klinov in [15] has implemented the probabilistic reasoner Pronto, that can reason with ontologies where a probability interval is assigned to statements, specifying the probability that a certain statement is true. As far as truth-values are concerned, the work of Straccia and Bobillo [16] extends the classical two-valued Description Logics to fuzzy sets. Namely, the authors present an extension of $\mathcal{SHIF}(\mathbf{D})$ Description Logics to the fuzzy case, dealing with different definitions for the logic operators (Zadeh logic, Lukasiewicz logic and Classical Logic); this approach, moreover, provides support to backward reasoning in case of Classical Logic semantics.

At first sight, one might hope that uncertainty representations and truth value ranges can be freely mixed according to the characteristics of the problem at hand. On the Web, vague information is usually modelled using fuzzy truth-values, knowledge uncertainty due to incomplete or defective observations is represented by Probability theory, and uncertainty arising from common sense knowledge and guessing can be handled with Possibility theory [17].

Unfortunately, the problem of dealing simultaneously with probability-based uncertainty and fuzzy truth values has been widely treated in literature but, as stated in [7] probability and possibility theories are not fully compositional with respect to all the logical connectives, without a relevant loss of expressiveness. This consideration leads to the consequence that uncertain calculi and degrees of truth are not fully compositional either. Nevertheless, some work in this direction has been proposed, by imposing restrictions to the expressiveness of the logics. The most relevant studies are: [18, 19] where the authors define *Probabilistic Description Logics Programs* (PDLP) by combining stratified fuzzy Description Logics programs with respect to degrees of probabilities in a unified framework. In [20] a definition of possibilistic fuzzy Description Logics has been proposed by associating weights, representing degrees of uncertainty, to the fuzzy Description Logic formulas. An extension of the fuzzy Description Logics in the field of Possibility theory has been presented also in [21] by annotating logic axioms with possibilities and necessity measures; by extending the approach presented in [20].

It is also important to underline that different approaches can be used to tackle the same type of uncertainty (i.e. in case of incomplete information is possible to use Possibilistic or Probabilistic theory) the choice of the best theory depends on the context. For instance, in [22] the authors use Dempster-Shafer's beliefs theory to resolve inconsistencies.

2.1 Ontology of Uncertainty

As mentioned in Section 2 uncertainty is generated from different situations and represented under different semantics. it is possible to create a classification of assertions based on several criteria: *nature*, *derivation*, *temporal validity* and *type*. Nature of uncertainty can be divided in *epistemic* and *aleatory*; *objective* and *subjective*; based on the temporal validity of a statement that can be valid for a period of time or can be valid always. A statement is *contingent* if refers to a particular situation or instant; in the second case a statement refers to situations that summarize trends (e.g. laws of physics, common sense knowledge, statistical knowledge) and is classified as *generic* statement. The nature of uncertainty, moreover, also depends on the statement it is attached to. Statements can be *ambiguous* in case the statement can be represented in different worlds with more than one interpretation, *inconsistent* if there is no possible world where the statement can hold, *vague*, and *incomplete* in case the knowledge about the observed world do not provide enough information to take a decision. Finally, a statement is *empiric* when is satisfied at least in one world. A first effort toward capturing all aspects of uncertainty is the *Ontology of Uncertainty*, published by the UR3W-XG incubator group[5]. This ontology tries to capture the nature, type and source of uncertainty that are specific of an assertion and allows moreover to relate the assertion to the correct computational model of inference. Statements, in the Ontology of Uncertainty, are represented by the concept `Sentence`, that provides information about the source (`Agent`) of the assertion, the subject (`World`) of the assertion and the semantics information about the `Uncertainty` model related to the assertions. There is not much to say about the two concepts `Agent` and `World`, which respectively represent the producer and the subject of an assertion. More interesting is the case of concept `Uncertainty`, the central concept of the ontology. This concept is related to all the various elements used to classify a `Sentence` under precise semantics. Other concepts related to `Uncertainty` are then used to describe type, derivation, validity and nature of the statement[2]. The ontology provides a generic meta-model rep-

[2]Currently, the Ontology of Uncertainty defined by the URW3-XG does not include a concept of `Validity` related to the temporal

ISBN: 978-989-95079-6-8

resented in OWL-DL [11] for representing uncertainty associated to various assertions and provides some use case scenarios, where, according to the semantics of the uncertainty, the correct inference model is selected. Unfortunately, the document [5] produced by the URW3-XG incubator group does not specify how to deal with situations where more than one model is involved in the inference process.

2.2 Using the Ontology of Uncertainty to support Reasoning

The Ontology of Uncertainty provides information on which mathematical model of uncertainty can be employed for managing a specific set of statements. When more than one model is involved in the reasoning process, the problem of integrating the results of multiple inference processes arises : if the subsets of statements handled by each model are disjoint, i.e. inferences are independent form each other, however, there are no particular problems in re-conciliating the results of the various reasoning processes;s. Some work in this direction has been carried out by the *Rule Interchange Format* (RIF[3]). In [2] the authors propose a framework for sharing information between three different models of uncertainty, where the fuzzy linguistic truth values are propagated through the three models in a non-monotonic way, by exploiting the extension principle [23] and aggregation of linguistic values. This approach is promising but is grounded to fixed fuzzy values (linguistic truth) that are used by all the different models and then aggregated according to non-monotonic rules.

In our approach, instead, we make use of the Ontology of Uncertainty as a way to model different types of uncertainty in a unified framework. The inference process involves three different steps: the first step is to partition the knowledge base in subsets according to the specific model, the second step is to carry out independent inferences, and the third one aggregates the results of the independent inference processes, following the *First Inference Then Aggregation* (FITA) approach [24], which also supports parallel reasoning. In our strategy the various reasoning processes are independent; we use the Ontology of Uncertainty classification to divide the various matching relations according to the uncertainty model to be used for the reasoning process. This way, the reasoners can be modelled as parallel processes. Directives on how to divide the heterogeneous knowledge base and how to recombine the results of the different reasoning processes are explicitly specified as DL-Safe Horn rules [25].

The knowledge base containing the information for our matching strategy is composed by a set of statements, generated independently by different sources, seen as instances of the concept `Sentence` in the Ontology of Uncertainty. To each statement S, information about uncertainty is associated by instantiating the ontology concept `Uncertainty` that defines the correct semantics. DL-Safe rules and SPARQL queries are largely involved in this process. A first set of DL-Safe rules is used to associate the statements to the correct type of uncertainty, and a set of SPARQL queries is used to divide the knowledge base according to the reasoning model associated to the statement. Finally a third set of rules is used to aggregate the results of the various reasoners. Currently, our approach requires the manual definition of rule sets. However, while the first set of rules is largely application dependent, the set of SPARQL queries and the third set of rules can be reused in many applications[4]. A detailed application of our approach is presented in Section 3.

3 A Semantics-aware Matching Strategy

Schema Matching is the time-honored problem of identifying the relations between the entities of two data source schemata. In case these schemata are represented as ontologies, this problem is also known as Ontology Alignment. In the literature different matching operators for a wide range of situations are available: the most exhaustive survey is [26].

Recent proposals tackle the schema matching problem by considering more than one matching operator at once and combining the final results through a matching strategy [27]. A **Matching Strategy** can be defined as *the process of transformation from a set of Matching Relations* Mr *to a new set* Mr', while a **Matching Operator** can be defined as *a function that takes as input two schema elements and creates as output a Matching Relation between the two elements*.

So far, even if some logic-based approaches are available [28, 29], most strategies proposed in literature neither consider explicitly the semantics of the various matching operators nor the different meanings of the relations that they generate. Instead, our matching strategy explicitly models the semantics of different matching operators as a Description Logic. In our approach, matching relations are stored as instances of a domain ontology, describing our application scenario. This domain ontology is extended with the *Ontology of Uncertainty* [5], which is used to associate the respective uncertainty to the different types of assertions. The *classification* of the various uncertainty types is performed by applying SWRL rules [30, 25] to the knowledge base. Once the classification is performed, the knowledge base is divided (by a *splitting* process) according to the specific uncertainty model. Each one of this inference models is used to perform a classification of the various relations in order to discover the most reliable ones. Finally a *reconciliation* process aggregates the best results[5].

3.1 The Matching Ontology

A matching relation mr_i is a $1:1$ relation associating two elements (concepts or attributes) of the two ontologies to align by a relation r from a collection of set theory operators ($\equiv$,$\subset$,$\supset$,$\cap$,$\neq$) and a degree of matching δ represented as:

$$mr_i = \langle e_h, e_k, r, \delta \rangle \tag{1}$$

The relation r between the elements depends on the particular feature that is analysed by the matching operator: as an example, in case of *JaroWinkler* matching operator, the features are composed by the label of the elements e_h and e_k to

validity of an assertion, but its addition is straightforward.

[3]The mission of the Rule Interchange Format (RIF) Working Group is to produce W3C Recommendations for rules interchange. `http://www.w3.org/2005/rules/wiki/RIF_Working_Group`

[4]Although they may require some fine tuning. This especially true of the third set, which is closely related to the aggregation procedure.

[5]The *classification*, *splitting* and *reconciliation* process in our system are rule driven, but this is not a strict requirement: this phases of the strategy can also be inferred from a Description Logics reasoning process used to classify various instances.

ISBN: 978-989-95079-6-8

match, while in case of an *Instance-based* matching operator the features are composed by the instances related to the elements to match.

We generally consider a matching operator as a process that generates a matching relation mr_i represented by the relation r between the elements e_h, e_k in input, associated with the strength δ of the relation. A matching operator is defined by a function f used to extract a particular feature that is analysed by the operator, a function θ which generates the relation r and a function ϕ that is in charge of the creation of the value δ, which represents the *strength* of the relation:

$$mo = \langle f, \theta, \phi \rangle \qquad (2)$$
$$mo(e_h, e_k) \rightarrow \langle e_h, e_k, r, \delta \rangle$$

Obviously, different matching operators carry different semantics that needs to be considered in the matching generation process. Also, the same relation can be generated by different matching operators; but in this case, the syntax of the relations is identical while the semantics is different. Consequently, different theories can be adopted, in order to infer the most suitable matching relations: the relations generated by the data type based matching operator can be considered as *possibilities*, while the relations generated by the instance based and string based matching operator can be modelled as *necessities* [6]. As we have seen, the Matching Ontology models the information created by the various matching operators, as well as information describing each specific matching operator. The attributes, relations and concepts from the ontologies to align are stored in the Matching Ontology as instances of the concept `Element`, represented by a unique identifier (e.g. URI) which is used by the system to retrieve the element in the original ontologies. The concept `MatchingOperator`, in the Matching Ontology, represents a generic matching operator; the specific matching operators (e.g. JaroWinkler, WordNet, Instances) are modelled as subclasses of this concept. The various matching relations are defined as instances of the concept `MatchingRelation`, which is composed by the concept `RelatedElements` that has a `subject` and an `object` that represent the two elements that have been related by one or more matching operators; and by a `confidence` value representing the strength of the matching relation. Subclasses of the concept `Matching Relation` are created in order to specialize the matching relation according to the specific relation (e.g. `Equivalent`, `Subset`, `Superset`, `Disjoint`, `Intersection`).

Instances of the Matching Ontology are generated during the matching process. The various matching operators generate a set of matching relations in the form $\langle e_h, e_k, r, \delta \rangle$; these relations are stored in the Matching Ontology with a reference to the instance of the matching operator that has generated the relation. When the Matching Ontology is generated in this way, some information is certain (e.g. the information provided by the original ontologies), and some other information in uncertain under different semantics (e.g. the matching relations that can disagree or that may have a probability degree to consider). This consideration motivates the use of additional annotations able to model the semantics of uncertainty of the various matching relations.

3.2 Managing uncertainty in the matching strategy

In a scenario like the one described in Section 3.1, we use the Ontology of Uncertainty to identify situations where is important to explicitly describe the type of uncertainty related to an assertion. First of all, the two ontologies (Matching Ontology and Ontology of Uncertainty) have to be linked somehow. Assertions in the Ontology of Uncertainty are represented by the concept `Sentence`. A sentence is `saidBy` an `Agent`, which we identify with the matching operator; moreover a sentence have also an object of the assertion (`saidAbout`) which is represented by the concept `World`. The two concepts `Agent` and `World` are then the linking point between the two ontologies; the link is defined by declaring `MatchingRelation` and `MatchingOperator` as sub-concepts of `World` and `Agent` respectively. Here, the Ontology of Uncertainty is used basically to drive the reasoning process: each type of uncertainty is processed by its specific reasoner and a final process, based on SWRL rules, integrates the results of the various reasoners. The application flow of the Matching Strategy operates as follows: a process takes as input the uncertain knowledge base generated by the matching operators, afterwards it divides the assertions according to their uncertainty and each sub part of the ontology is processed by its specific reasoner: in the system we consider a Probabilistic Description Logic reasoner, such as *Pronto* [15], a fuzzy Description Logic reasoner such as *FuzzyDL* [16] and a Defeasible Logic Reasoner such as *DR-Prolog* [3], but other models can be easily added. In our scenario the sources of information to be analysed, according to different uncertainty models are independent and no intersections among them has to be managed. This particular case allows a straightforward use of the Ontology of Uncertainty to drive the reasoning process, although in general the assumption of independence among the source of information is a lucky case.

The first part of the matching strategy is to assign to the various assertions (`Sentence`), the correct information about their uncertainty semantics. This information is classified according to a set of pre-defined SWRL rules that assigns the correct semantics in relation to several factors. The assignment is based on: (i) the `Agent` that has generated the relation; as an example: some agents can generate objective or subjective assertions: we can identify objective statements as necessities and subjective as possibilities; (ii) the presence of a degree of probability: as an example, sentences with a degree of probability can be handled with probabilistic theory models or with possibility theories. In this case is important to identify which statements need to be modelled with probability theory and which ones need to be modelled with possibility theory. (iii) the level of inconsistency among matching relations: as an example, if a sentence asserts that two elements are equivalent and another matching relation asserts that they are disjoint; (iv) the trustiness level of the matching operator: some operators are more reliable than others; (v) the level of detail of the assertion: the assertion created by a Data Type matching operator is more vague then an assertion created by a Regular Expression matching operator; (vi) the Data Type

[6] Note that this does not have a direct implication on the mappings to be created by the systems; for instance, it is clear that a necessary matching of strings does not imply a necessary mapping.

ISBN: 978-989-95079-6-8

of the elements to match can be used to establish priorities between operators.

$$\begin{aligned} uncertainty : Sentence(?sentence) \wedge \\ matching : MatchingRelation(?matrel) \wedge \\ uncertainty : saidBy(?sentence, ?operator) \wedge \\ matching : MatchingOperator(?operator) \wedge \\ uncertainty : saidAbout(?sentence, ?matrel) \wedge \\ uncertainty : Objective(?derivation) \wedge \\ uncertainty : Uncertainty(?uncer) \wedge \\ uncertainty : hasUncertainty(?sentence, ?uncer) \rightarrow \\ uncertainty : derivationType(?uncer, ?derivation) \end{aligned} \tag{3}$$

There are one or more rules for each specific uncertainty type, nature, model, derivation and temporal validity. An example of a rule that we use in this step of the matching strategy is reported in (3). In our strategy the rules are applied to the matching ontology with the use of a rule engine such as Jess[7]; the rules we use have to be restricted to the DL-Safe [25] subset to ensure tractable complexity of the reasoning processes. By applying the rules, all the matching relations are associated to their respective uncertainty.

At this point, a set of sub-knowledge bases is created by dividing the various instances of the concept `sentence` according to their uncertainty model. A SPARQL [32] query is used in this case to select the instances of the concept `sentence` that respect the desired restrictions. As an example the following SPARQL query returns the instances of `Sentence` that are associated to a probability theory model.

```
SELECT ?sentence ?type ?derivation ?nature ?model
WHERE {?sentence uncertainty:hasUncertainty ?uncertainty.
?uncertainty uncertainty:nature ?nature.
?uncertainty uncertainty:derivationType ?derivation.
?uncertainty uncertainty:uncertaintyType ?type.
?uncertainty uncertainty:uncertaintyModel ?model.
?probabilistic rdf:type uncertainty:Probability.
FILTER (?model = ?probabilistic)}
```

Once the set of assertions has been partitioned, the parallel reasoning processes can be launched. The reasoning processes are performed locally, exploiting the information provided by each assertion and the information provided by the ontologies to align. According to the information that has been provided to each reasoner, the process has to return back to the matching strategy the set of assertions that they believe to be the most reliable ones. Each reasoning process returns the results as instances of its representative concept (sub-concepts of the concept `Sentence`).

When the parallel reasoning processes come to an end, results are propagated back to the matching ontology by a reconciliation process. This process can be another reasoning process; in the case of our matching strategy we make use of SWRL rules to aggregate the results. Basically the reconciliation process follows the priority between the various reasoners that need to be made explicit. Some models are more reliable than others and this preference is defined in our system by SWRL rules. In our matching strategy we have to deal with inconsistencies because different relations on the same pair of elements can be classified as reliable relations by different reasoning processes. In case the preferences between inference models can not solve this situation we handle inconsistencies by assigning preferences between relations and operators, according to contextual factors: (e.g., analysing the data types of the elements to match. *Equivalence* relations have the highest priority in case of Strings, because Instance-based equivalence between integers is less reliable than the one between Strings). Rules are also used to propagate the best matching relations to other elements of the ontologies exploiting structural information from the original ontologies to match. This case can again be managed using a defeasible rules system such as DR-Prolog [3], which provides different precedences between rules, to help the decision process in inconsistent situations.

[7] in the Java prototype we developed, we have used Jess Rule engine (`http://herzberg.ca.sandia.gov/`) with the support of JessTab [31] and Protégé (`http://protege.stanford.edu/`) to translate SWRL rules and assertions from OWL to Jess and backwards.

4 Experimental Evaluation

Let us now describe the use of OAEI as a validating test for our data integration system. The comparison has been carried out with the results of the 2007 contest [4]. The majority of the participants of the context are based on a linear combination of several matching operators. In some cases this aggregating function is adaptive (Asmov [33], Prior+ [34]), while in some other cases it is fixed and defined by several experimentation (RiMOM [35], Sambo [36], Soda [37], Ola [38], TaxoMap [39], X-Som [40]). Some approaches are based on Possibility theory such as DSSim [28] and OWL-CM [41]. The test is performed using a palette of five semantically different matching operators (*JaroWrinkler Matching Operator, WordNet Label Matching Operator, Description Matching Operator, Type-based Matching Operator* and *Instance-based Matching Operator*). The results are stored in a matching ontology and the most reliable relations are extracted by our matching strategy. Each relation is associated to the respective matching operator. The reasoners that we used are a Fuzzy Description Logic reasoner [16] for the matching relations classified as probabilities (e.g. JaroWinkler matching operator) and a classic Description Logic reasoner [42] to process certain matching relations that the rules classify as necessities (e.g. String matching operator). The matching strategy has been developed as a Java prototype that we used to run the tests.

4.1 Results

The testbed can be divided in five subcategories:

101-104 This test set is the easiest set of tests. The first task is to match the reference ontology 101 with itself, the second test requires to generate a matching from the reference ontology to an ontology totally irrelevant (102 is a wine ontology). The ontology 103 represents a language generalization (unavailable constraints are replaced by their generalization): this ontology is an OWL-Lite generalization of the ontology 101. Finally the ontology 104 represents a language restriction with respect to the reference ontology 101. The constraints that are not available in OWL-Lite are simply removed. In the case of this first set of tests, our algorithm does not generate perfect alignments. In the case of 103 the propagation mechanism of the strategy does not perform well because of the generalization of the ontology. The various matching operators generate several reliable relations that the strategy can not recombine correctly in relation to their priority.

201-210 In this set of test cases, the structure of ontology

ISBN: 978-989-95079-6-8

Test #	Name	Prec.	Rec.	fMeas.
101	Reference Alignment	1	1	1
102	Irrelevant Ontology	NaN	NaN	NaN
103	Language Generalization	0,81	0,56	0,66
104	Language Restriction	0,94	0,94	0,94

is preserved. Syntactical changes are introduced: labels and identifiers are replaced by random names, misspellings, synonyms, language translation and moreover the comments in some cases have been suppressed. Our matching strategy obtains good results in this set of tests, because of the variety of the palette of matching operators that are considered in the strategy. In some cases (202, 209, 210) the Recall value is low, this because in case a matching operator does not create a possible relation this can not be automatically generated by the strategy. SWRL rules are used to propagate the best results to related elements, but if the matching relation is not created by any matching operator the propagation mechanism is not effective.

Test #	Name	Prec.	Rec.	fMeas.
201	No Names	0,92	0,90	0,91
202	No Names, No Comments	0,88	0,15	0,26
203	No Comments	0,97	0,97	0,97
204	Naming Conventions	0,92	0,92	0,92
205	Synonyms	0.91	0,90	0,90
206	Translation	0,96	0,70	0,95
207		0,98	0,70	0,97
208		0,96	0,72	0,82
209		0,87	0,34	0,49
210		0,92	0,12	0,22

221-247 In this case the set of tests can be divided into two subgroups: **221-231** and **232-247**. The first subgroup contains several structural modifications, such as the hierarchy that is flattened or expanded, and individuals, restrictions and data types that are suppressed. Each one of the documents in this subgroup has been modified by a structural change. Because of the fact that the labels and comments are preserved, the modifications have little influence on our system. The use of *Description Matching Operator* and *Name Matching Operator* allows the strategy to find most of the correct alignments using just the labels and comments information. In the second subgroup (232-247), the modifications made by combinations of the single modifications used in 221-231. Our system obtains good results for 232-247 as well.

Test #	Name	Prec.	Rec.	fMeas.
221	No Specialization	0,92	0,92	0,92
222	Flattened Hierarchy	0,92	0,92	0,92
223	Expanded Hierarchy	0,93	0,93	0,93
224	No Instance	0,93	0,93	0,93
225	No Restrictions	0,91	0,91	0,91
228	No Properties	0,97	0,97	0,97
230	Flattened Classes	0,95	0,96	0,95
231	Expanded Classes	0,93	0,93	0,93
232		0,95	0,95	0,95
233		1	1	1
236		1	1	1
237		0,92	0,92	0,92
238		0,91	0,91	0,91
239		0,93	0,97	0,95
240		0,85	0,88	0,87
241		1	1	1
246		0,97	1	0,98
247		0,91	0,94	0,93

Note that the observations about the Recall value we made about the previous tests, here are not valid since the ontologies provide sufficient information to the matching operators that can create reliable relations. Also in case of 222 where there is no hierarchy the results are satisfactory because of the information provided by the ontology.

248-266 This set of documents represents the most challenging case. This set combines structural and syntactical suppressions. The single challenges represented by the two previous set of documents are mixed in this set. All labels and identifiers are replaced by random names, and the comments are also suppressed. In this case our system does not perform well because no hierarchical representation is preserved so the strategy can not propagate the few correct matching found. This results in a low level of Recall. However, not enough information is provided in the ontologies, and the matching strategy can only find few alignments. The tests from 254 to 262, are the most difficult since almost all literal (labels and comments) and structural information are removed. In this case the propagation of the results can not take place because the relations that the strategy discovers can not be associated to other elements because of the structural information that is missing. When some structural information is preserved, the strategy can exploit this information in order to create possible matches, starting from the relations discovered by the single matching operators.

Test #	Name	Prec.	Rec.	fMeas.
248		0,88	0,14	0,25
249		0,74	0,14	0,24
250		0,92	0,33	0,49
251		0,86	0,13	0,22
252		0,65	0,11	0,19
253		0,88	0,14	0,25
254		0,90	0,27	0,42
257		0,92	0,33	0,49
258		0,86	0,13	0,22
259		0,65	0,11	0,19
260		0,82	0,31	0,45
261		0,69	0,27	0,39
262		0,90	0,27	0,42
265		0,82	0,31	0,45
266		0,69	0,27	0,39

301-304 This test set is composed by real ontologies contextually related to the reference ontology 101. The ontologies in this test set represent bibliographical information and they have been defined independently each other. This test represents a real world case of ontology alignment. Our strategy performs in the average with respect to the other systems evaluated. In the case of ontologies 301 our approach finds most of the correct alignments, but it also returns some wrong results. The alignment results for 302 and 303 are far from satisfactory. The reason is that these ontologies do not provide individuals and with shallow class hierarchy, where classes and properties are not related. In this case as well, the recall value is low: the matching operators are very sensible to the noise in data; moreover without a useful hierarchy the few matching relations identified by the matching operators can not be propagated or enforced with the support of hierarchical information. The ontology 304 has similar structure and vocabularies to the reference ontology 101 and in this case, the results are slightly better than the previous alignments.

Test #	Name	Prec.	Rec.	fMeas.
301	Real: BibTeX/MIT	0,9	0,6	0,71
302	Real: BibTeX/UMBC	0,79	0,46	0,58
303	Real: Karlsruhe	0,83	0,49	0,62
304	Real: INRIA	0,81	0,62	0,70

ISBN: 978-989-95079-6-8

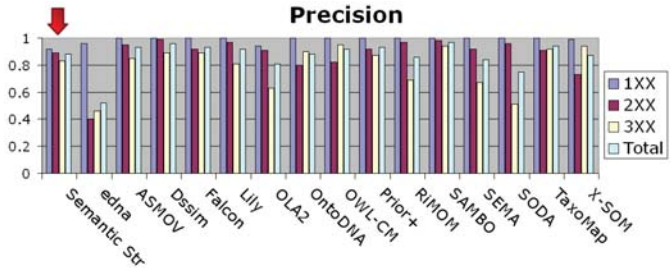


Figure 1: Graph of the Precision values of the different matching algorithms.

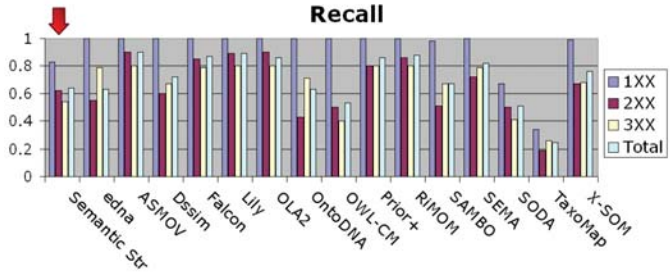


Figure 2: Graph of the Recall values of the different matching algorithms.

As one can see from the graphs in Figure 1 and 2 , our algorithm has acceptable results but under the best scores. These results are anyway reasonably encouraging, because our strategy is neither entailed to a particular reasoning model nor to a set of specific matching operator. More important our strategy does not imply a training phase that is typically required by the other systems. Our semantics-aware strategy is a new approach to ontology matching problem that provide a approach generally valid and not dependent on the domain. At this stage the results are very sensible to factors independent to the strategy; such as the quality of the matching operators and the available reasoners.

For this test we used a Probabilistic Description Logics reasoner and a traditional Description Logic reasoner in case of matching relations without a confidence degree. Moreover the inconsistencies, in the knowledge base, are not treated by a reasoning process but with the use of SWRL rules. As soon as new reasoners or new models appears we can exploit their use with our strategy. Moreover or strategy is highly customizable: if a new matching operator is plugged in the matching strategy is just necessary to create the related concept in the matching ontology and the SWRL rules reacting to the semantics of the new matching operator.

In case of the first series of test the strategy do not obtain the best results but, instead provides results under the average. Even the most simple matching operators, such as Edna, performs better than our strategy. The reason of this behaviour has been identified in the confusion of the strategy that considers the relations generated by different operator all reliable at the same level. We had run the same test only with JaroWinker matching operator and we obtained an average value of Precision 1 and Recall 1 for all the first series; this result confirm our previous assumption. In the case of the second and the third series the results are in the average of the approaches presented to the contest. The graph in Figure 1 shows the comparison of the Precision of our strategy with respect to the other algorithms. As is possible to see the values are very closed to 1, except in some isolated cases (Edna). These results show that the various strategies generate correct relations. The graph in Figure 2 shows the comparison of the Recall of our strategy with respect to the other algorithms. As is possible to see the values this time instead are far from the limit. Only few strategies (Asmonv, Falcon, Lily, OLA2) obtain results around 0.8, while the others still remain over the $0,5$. This generally low value of Recall means that the algorithm does not propagate well the good matching relations by exploiting the structure of the ontologies.

5 Conclusions

In this paper we presented preliminary work on a framework for managing different types of uncertainty and a possible application to the Schema Matching problem. The Ontology of Uncertainty, proposed by the W3C's UR3W-XG incubator group, provides a vocabulary to annotate different sources of information with different types of uncertainty. We argue that such annotations should be clearly mapped to corresponding reasoning and representation strategies. This mapping allows the system to analyse the information on the basis of its uncertainty model, running the inference process according to the respective uncertainty. In this way we can also deal with complex situations that do not tailor to the traditional strategies: as an example, matching relations provided by a user with her specific level of expertise. In our scenario the sources of information analysed, according to different uncertainty models are independent and no intersection among them has to be managed. This particular case allows a straight-forward use of the Ontology of Uncertainty to drive the reasoning process, although in general the assumption of independence among the source of information is a lucky case. The need of additional work on the Ontology of Uncertainty is necessary in order to support reasoning processes when combinations of uncertainty models are applied to a single source of information. This outlook is promising in order to provide more expressive frameworks for reasoning under different types of uncertainties, but definitely the need of more research in this direction is evident.

Acknowledgements

This work was partly funded by the Italian Ministry of Research under FIRB contract n. RBNE05FKZ2_004, TEKNE.

References

[1] Maurizio Lenzerini. Data integration is harder than you thought. In *CooplS '01: Proceedings of the 9th International Conference on Cooperative Information Systems*, pages 22–26, London, UK, 2001. Springer-Verlag.

[2] X. Luo, C. Zhang, and N. R. Jennings. A hybrid model for sharing information between fuzzy, uncertain and default reasoning models in multi-agent systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(4):401–450, 2002.

[3] Grigoris Antoniou and Antonis Bikakis. Dr-prolog: A system for defeasible reasoning with rules and ontologies on the semantic web. *IEEE Trans. Knowl. Data Eng.*, 19(2):233–245, 2007.

ISBN: 978-989-95079-6-8

[4] Jérôme Euzenat, Antoine Isaac, Christian Meilicke, Pavel Shvaiko, Heiner Stuckenschmidt, Ondrej Svab, Vojtech Svatek, Willem Robert van Hage, and Mikalai Yatskevich. Results of the ontology alignment evaluation initiative 2007. In *Proc. 2nd ISWC 2007 international workshop on ontology matching (OM), Busan (KR)*, pages 96–132, 2007.

[5] I Kenneth J. Laskey, Kathryn B. Laskey, Paulo C. G. Costa, Mieczyslaw M. Kokar, Trevor Martin, and Thomas Lukasiewicz. Uncertainty reasoning for the world wide web. *W3C Incubator Group Report*, 2008.

[6] B. de Finetti. La logique de la probabilité. In *Actes du Congres International de Philosophie Scientifique a Paris 1935, Tome IV*, pages 1–9. 1936.

[7] Didier Dubois and Henri Prade. Can we enforce full compositionality in uncertainty calculi. In *In Proc. of the 11th Nat. Conf. on Artificial Intelligence (AAAI-94*, pages 149–154. AAAI Press / MIT Press, 1994.

[8] George Boole. The calculus of logic, 1848.

[9] L.A. Zadeh, K.S Fu, K. Tanaka, and M. Shimura. *Fuzzy sets and their applications to Cognitive and Decision Processes*. 1975.

[10] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*, 2003.

[11] Deborah L. McGuinness and Frank van Harmelen. Owl web ontology language overview. *W3C Recommendation*, 2004.

[12] Shiyong Lu, Ming Dong, and Farshad Fotouhi. The semantic web: Opportunities and challenges for next-generation web applications. *International Journal of Information Research*, 7:2002, 2002.

[13] I. Horrocks and U. Sattler. Ontology reasoning in the SHOQ(D) description logic. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, 2001.

[14] Thomas Lukasiewicz. Expressive probabilistic description logics. *Artif. Intell.*, 172(6-7):852–883, 2008.

[15] Pavel Klinov. Pronto: Probabilistic dl reasoning with pellet, 2008.

[16] Fernando Bobillo and Umberto Straccia. fuzzydl: An expressive fuzzy description logic reasoner. In *2008 International Conference on Fuzzy Systems (FUZZ-08)*, pages –. IEEE Computer Society, 2008.

[17] Janusz Kacprzyk and Slawomir Zadrozny. Computing with words in intelligent database querying: standalone and internet-based applications. *Inf. Sci.*, 134(1-4):71–109, 2001.

[18] Thomas Lukasiewicz and Umberto Straccia. Description logic programs under probabilistic uncertainty and fuzzy vagueness. In *ECSQARU*, pages 187–198, 2007.

[19] Thomas Lukasiewicz and Umberto Straccia. Uncertainty and vagueness in description logic programs for the semantic web, 2007.

[20] Didier Dubois, Jérôme Mengin, and Henri Prade. Possibilistic uncertainty and fuzzy features in description logic. A preliminary discussion. In E. Sanchez, editor, *Fuzzy logic and the semantic web*, pages 101–113. http://www.elsevier.com/, 2006. DMenP001.

[21] Fernando Bobillo, Miguel Delgado, and Juan Gómez-Romero. Extending fuzzy description logics with a possibilistic layer. In *URSW*, volume 327 of *CEUR Workshop Proceedings*, 2007.

[22] Andriy Nikolov, Victoria S. Uren, Enrico Motta, and Anne N. De Roeck. Using the dempster-shafer theory of evidence to resolve abox inconsistencies. In *URSW*, volume 327, 2007.

[23] L.A. Zadeh. Fuzzy sets. *Information Control*, 8:338–353, 1965.

[24] Hiroshi Maeda and Yuji Nobusada. A study on the parallel computability of multi-fold approximate reasoning. *Fuzzy Sets Syst.*, 97(2):129–144, 1998.

[25] Boris Motik, Ulrike Sattler, and Rudi Studer. Query answering for owl-dl with rules. In *Journal of Web Semantics*, pages 549–563, 2004.

[26] Erhard Rahm and Philip A. Bernstein. A survey of approaches to automatic schema matching. *VLDB Journal: Very Large Data Bases*, 10(4):334–350, 2001.

[27] Pavel Shvaiko and Jérôme Euzenat. *Ontology matching*. Springer-Verlag, Heidelberg (DE), 2007.

[28] Miklos Nagy, Maria Vargas-Vera, and Enrico Motta. Dssim-ontology mapping with uncertainty. In *Ontology Matching*, 2006.

[29] Andrea Calì, Thomas Lukasiewicz, Livia Predoiu, and Heiner Stuckenschmidt. A framework for representing ontology mappings under probabilities and inconsistency. In *URSW*, 2007.

[30] Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, and Mike Dean. Swrl: A semantic web rule language combining owl and ruleml. *W3C Member Submission*, 2004.

[31] Henrik Eriksson. Using jesstab to integrate protégé and jess. *IEEE Intelligent Systems*, 18(2):43–50, 2003.

[32] Eric Prud'hommeaux and Andy Seaborne. Sparql query language for rdf. *W3C Member Submission*, 2008.

[33] Yves Jean-Mary and Mansur Kabuka. Asmov results for oaei 2007. In *Proc. 2nd Ontology matching workshop, Busan (KR)*, pages 141–150, 2007.

[34] Ming Mao and Yefei Peng. The prior+: Results for oaei campaign 2007. In *2nd International Workshop on Ontology Matching (OM) at 6th International Semantic Web Conference (ISWC), Busan, Korea, 2007*, 2007.

[35] Yi Li, Qian Zhong, Juanzi Li, and Jie Tang. Result of ontology alignment with rimom at oaei'07. In *The Second International Workshop on Ontology Matching*, 2007.

[36] Patrick Lambrix and He Tan. Sambo-a system for aligning and merging biomedical ontologies. *Web Semant.*, 4(3):196–206, 2006.

[37] Sami ZGHAL, S. Ben Yahia, Engelbert MEPHU NGUIFO, and Y. Slimani. Soda: an owl-dl based ontology matching system. In *OAEI'07 - OM'07 - ISWC'07*, Busan (Korea), nov 2007.

[38] Jérôme Euzenat, David Loup, Mohamed Touzani, and Petko Valtchev. Ontology alignment with ola. pages 333–337, 2004.

[39] Haifa Zargayouna, Brigitte Safar, and Chantal Reynaud. TaxoMap in the OAEI 2007 alignment contest . In *Ontology Alignment Evaluation Inititiative (OAEI) 2007 Campaign - Workshop on Ontology Matching*, 2007.

[40] Carlo Curino, Giorgio Orsi, and Letizia Tanca. X-som: A flexible ontology mapper. In *DEXA '07: Proceedings of the 18th International Conference on Database and Expert Systems Applications*, pages 424–428, Washington, DC, USA, 2007. IEEE Computer Society.

[41] Boutheina Ben Yaghlane and Najoua Laamari. Owl-cm: Owl combining matcher based on belief functions theory.

[42] Evren Sirin, Bijan Parsia, Bernardo C. Grau, Aditya Kalyanpur, and Yarden Katz. Pellet: A practical owl-dl reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):51–53, June 2007.

ISBN: 978-989-95079-6-8

Fuzzy Descriptions to Identify Temporal Substructure Changes of Cooccurrence Graphs

Matthias Steinbrecher[1] Rudolf Kruse[1]

1. Department of Knowledge Processing and Language Engineering
Otto-von-Guericke University of Magdeburg
Universitätsplatz 2, 39106 Magdeburg, Germany
E-mail: {msteinbr, kruse}@ovgu.de

Abstract— *Cooccurrence graphs easily grow very dense when applied to represent binary association patterns of large amounts of data. Therefore, postprocessing is needed to extract valuable information from them. We propose an approach to identify subgraphs of cooccurrence graphs that show a certain temporal behavior. This behavior is described with linguistic variables and fuzzy connectives defined over the change rate domains of certain graph measures. These measures assess graph properties whose change over time the user is interested in. To justify our proposed method, we are going to present evidence from a real-world dataset.*

Keywords— Cooccurrence graphs, Temporal change, Fuzzy description

1 Introduction

Frequent pattern mining has become a prominent method for identifying patterns in large volumes of data and led to algorithms for postprocessing these patterns [1] or transfer the underlying ideas to other data structures such as graphs [2, 3, 4]. Since the search for frequent patterns necessarily has to deal with subsets of input data, one easily runs into the problem of combinatorial explosion which is reflected by the common problem of finding more patterns than there are input data. We have addressed this issue in previous work [5], arguing that the user requires tools that allow to filter the results in order to identify only those results that meet his criteria (such as interestingness, novelty, etc.). In addition to that we also provided arguments and empirical evidence [6] that patterns usually do not arise all of a sudden but evolve or disappear rather slowly as time passes. In consequence we proposed a method that allows the user to specify linguistically (in terms of fuzzy variables) the temporal behavior of the values of association rules' evaluation measures that he is interested in. The presented algorithm thinned out the entire rule set retaining just those rules that matched the users' concepts (to some degree).

In this paper we develop this idea further while keeping the way of describing temporal behavior (explained in the background section) but transferring it to a different area of application.

This area of application comprises the identification of interesting substructures in cooccurrence graphs. These graphs arise quite naturally wherever, theoretically speaking, one is interested in the fact that two entities share some property with respect to a so-called location (which not necessarily has to be a spatial artifact but often is). Two authors being cited by the same paper [7], two persons having visited the same website [8, 9, 10], or two crimes being committed at the same location are just three examples of cooccurrences. We will elaborate the possible applications later in the future work section. All these examples can be represented by an undirected graph with the node set comprising all possible locations and the weighted edges representing the cooccurrences. A typical graph of a real-world application is depicted in Fig. 1. It represents 100 locations in a 3D gaming environment. Whenever two locations have been visited by the same players, an edge is inserted (which also gets assigned a weight drawn as the width of the edge representing the number of common visitors, but this is omitted here). It is pretty obvious that a user needs some assistance tools that allow him to identify interesting substructures (edge combinations). This becomes even more important if we have multiple such graphs representing the cooccurrences of different time frames (e. g. weeks or months). The focus of this paper is to present a straightforward yet powerful approach how to identify common substructures in a collection of cooccurrence graphs by means of linguistics expressions that address the temporal change in these patterns.

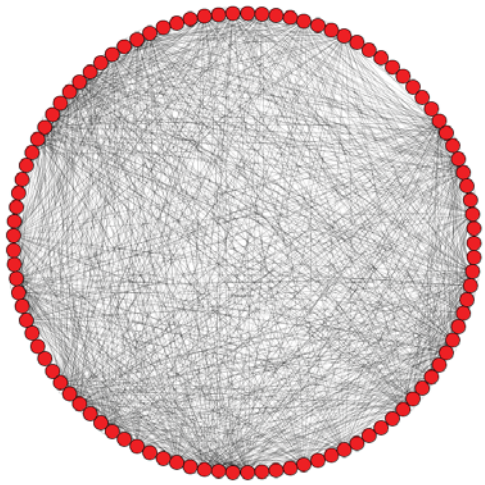


Figure 1: A typical cooccurrence graph with 100 nodes as it arose in a small-sized application (online gaming players that having visited the same locations in a 3D world). It is clear that this representation calls for some means of postprocessing in order to extract usable information from it.

The remainder of this paper is organized as follows: Section 2 introduces the notation used throughout the paper and revisits the linguistic filtering introduced in [6]. Section 3 mo-

ISBN: 978-989-95079-6-8

tivates and presents the method of extracting substructures of cooccurrence graphs that exhibit a certain user-specified temporal behavior. To underpin our proposal, we use a real-world dataset to illustrate the different stages of analysis in Section 4. Since this dataset reveals substructures that could not have been better generated manually, we refrain from handcrafting an artificial dataset and use selected subsets instead. We conclude our paper in Section 5 and propose other applications as well as possible promising extensions.

2 Background and Nomenclature

2.1 Graph Notations

In this paper we are going to deal exclusively with undirected graphs which we model as a tuple $G = (V, E)$ with vertices V and edge set E with

$$E \subseteq V \times V \setminus \{(v, v) \mid v \in V\},$$

and the constraint

$$(u, v) \in E \;\Rightarrow\; (v, u) \in E$$

to emphasize the undirected character. We will interpret the graphs as cooccurrence graphs where edges determine the number of cooccurrences (of whatever kind). This is taken into account with an edge weight function for every edge $e = (u, v)$:

$$w : E \to \mathbb{N}_0 \;\text{ with }\; w(e) = w(u, v) = w(v, u).$$

In the figures, this weight is represented as the edge width, thus we use the notion *width* and *weight* interchangeably. Given a subset $W \subseteq V$, we can induce a subgraph $G_W = (W, E_W)$ with

$$E \supseteq E_W = \{(u, v) \mid u, v \in W \wedge (u, v) \in E\}.$$

In the remainder we will sometimes use such a subset W in the context of a graph; it is G_W that we then refer to. A threshold θ defines the subgraph $G_\theta = (V, E_\theta)$ with

$$E_\theta = \{(u, v) \mid (u, v) \in E \wedge w(u, v) \geq \theta\},$$

i. e., as the graph containing only edges with a weight greater or equal to θ. Both operations can of course be combined, i. e., $G_{W,\theta}$ represents the subgraph of G induced by the node set W after having removed all edges with weight less than θ.

Since we will deal with sequences of graphs, we denote the temporal index as a superscript. All graphs share the same node set V and differ only in their edge sets or edge weights or both. Given a sequence $G^{(1)}, \ldots, G^{(n)}$ of graphs, we define the sum of these graphs as follows: $G_\Sigma = (V, E_\Sigma)$ with

$$E_\Sigma = \bigcup_{i=1}^{n} E^{(i)} \;\text{ and }\; w_\Sigma(u, v) = \sum_{i=1}^{n} w^{(i)}(u, v).$$

2.2 Linguistic Filtering Revisited

As described in [6] it is not only important to find patterns that meet some predefined constraints (such as minimum support or confidence) but also interpret these patterns in terms of temporal change. If a pattern describes a problem in some domain or an interesting customer behavior, it might be of interest to find such evolving patterns early. The underlying idea is as follows: given a pattern, we devise a set of evaluation measures that characterize the particular pattern (in [6] we used association rule measures such as support, lift, confidence, etc.). Next, the time series of a user-selected subset of these measures is calculated for every pattern. Each time series in turn is aggregated to a single value representing the overall trend, if any. The domain of this aggregate (i. e., the change rate domain) is equipped with an adequate fuzzy partition. Given a fuzzy rule antecedent (representing the user's intention of what temporal behavior of which measure(s) he is interested in), for every rule a membership degree to this concept is computed and an ordered list of rules according to these degrees is returned.

3 Spatio-temporal Filtering

3.1 Motivation

The objective of our approach is to answer questions of the following type (given a sequence of cooccurrence graphs):

> "First, what are interesting candidates for subgraphs that it would be worth looking at over time?"

and

> "Second, given a (still intractable large) set of subgraphs, which graphs become more sparse and less balanced over time?"

Before we turn to the algorithmic part of our approach, we need to negotiate which types of substructures within the graphs are most interesting to users. We will exploit the edge weights for this purpose. Several measures are needed to quantify for every subgraph aspects such as size, completeness, edge balance, etc.

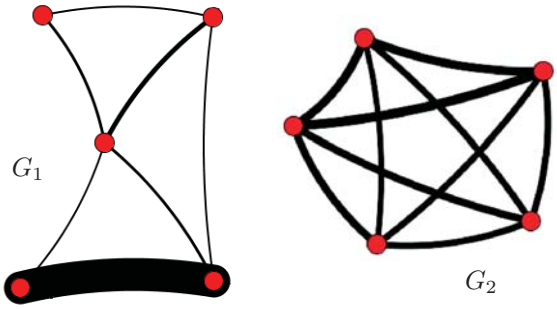
If the cooccurrence graphs represent visits of different web pages within the same online shop portal, then it might be desirable to know whether customers are able to use the web portal as intended by the owners. Are there dead ends where users are stuck? What are the "hot spot" sites, i. e., the pages that attract the most users and are visitors able to find the recently introduced shortcut to related pages? How do the accesses to the support area of the site change after renewing the navigational aids, etc.

We explicitly stress that subgraphs that are heavily interconnected with large edge weights only provide us with a *hint* that there may be an interesting visiting pattern. However, we can never conclude transitivity just from the cooccurrence graph! This is due to the fact that it only represents *binary* cooccurrences. Even a fully connected graph does not tell us anything about individual events. The sets of cooccurring events whose cardinality is represented by the edge weights even might be mutually disjoint. However, these subgraphs are found to be valuable hints that are worth being investigated.

3.2 Graph Measures

Focussing on the before-mentioned type of aspects one can identify highly connected subgraphs with large edge weights to be one type of substructures that are most interesting to users. Another type may be single edges just connecting two nodes or substructures that are highly interconnected but with

ISBN: 978-989-95079-6-8



Measure		G_1	G_2
size	Size	5.0	5.0
comp	Completeness	0.7	1.0
wght	Total Weight	5082.0	7906.0
avg	Average Weight	726.0	790.6
med	Median Weight	257.0	770.0
dev	Std. Dev.	1191.1	129.3

Figure 2: Two graphs with the same number of nodes. G_1 lacks three edges to be complete, therefore the completeness is just 0.7, whereas the clique G_2 yields 1.0. The rather large difference between average weight and median weight for G_1 (in contrast to G_2) indicate an imbalanced edge widths distribution which is strengthened by the large standard deviation value. The two graphs obviously justify this finding. The layout only acts as a visual cue and thus does not have any influence on the graph measures.

a large imbalance in the edge weights. The latter might represent two active sets of websites (large edge weights) between which users are able to navigate back and forth (numerous edges in between but with small weights since not every user is likely to use the offered navigational freedom).

The last arguments call for measures that on the one hand capture the mentioned properties of subgraphs and on the other hand allow to build a fuzzy partition on their domains since we are not going to ask for subgraphs *with 9 nodes and edges with weights greater than 100* but for *large* subgraphs with *moderately sized* edges. We use the following set of (sub)graph measures to quantify different aspects:

$$\begin{aligned}
&\text{Size} && \operatorname{size}(G_W) = |W| \\
&\text{Completeness} && \operatorname{comp}(G_W) = \frac{2\,|E_W|}{|W|^2 - |W|} \\
&\text{Edge Weight} && \operatorname{wght}(G_W) = \sum_{e \in E_W} w(e)
\end{aligned}$$

The size simply represents the number of nodes of the subgraph, whereas completeness refers to the relative number of edges compared to the maximal number. Zero represents an isolated graph (no edges) while a value of 1 designates a clique. Finally, the edge weight simply returns the sum of all edge weights without giving any clue about the distribution of these weights among the edges. Therefore, three additional measures are used: $\operatorname{avg}(G_W)$ calculates the arithmetic mean of all edge weights, $\operatorname{med}(G_W)$ returns the median of the weights and $\operatorname{dev}(G_W)$ represents the standard deviation of the weights. Fig. 2 illustrates these intentions with two graphs of the same size. Note, that these are subgraphs from real-world data and no artificial graphs that were crafted to meet the requirements.

3.3 *Candidate Graph Generation*

As we are now equipped with measures to assess certain aspects of subgraphs that we would like to track over time, the remaining question is how to determine such candidate graphs? It is clear that a brute-force approach (testing all subsets of nodes as potential subgraph node sets) fails immediately due to runtime problems, even for small node sets. We therefore promote the following heuristic: The graphs of all time frames are added as shown in Section 2 to arrive at the sum graph G_Σ (or simply the cooccurrence graph if we ignore the time frames). Next, a threshold θ is chosen and the graph's components (disconnected subgraphs) $\mathcal{C}_{G_\Sigma} = \{C_1, \ldots, C_j, \ldots, C_m\}$ of $G_{\Sigma,\theta}$ are taken as the candidate subgraphs. The choice of θ can be entirely left to the user (e. g. by offering a graphical preview tool that shows the components instantly whenever the user selects a new threshold via a slider) or θ may be determined in such a way to limit either the number of components or the (average) size of the components.

3.4 *Matching Against Linguistic Concepts*

Whatever way of determining the granularity of components is chosen, we are left with a set of mutual disjoint node sets $\mathcal{C}_{G_\Sigma}$ that are used to create a sequence of subgraphs $\left\langle G^{(i)}_{C_j} \right\rangle$, $i = 1, \ldots, n$, $j = 1, \ldots, m$ (one sequence for every subgraph induced by the node set) that are evaluated against the user-specified temporal behavior description. A time series is generated for every measure referenced in this user description. The temporal change within this time series is computed and the degree of membership to the user description is calculated. We will employ a simple regression approach, i. e., we fit a regression line into the time series and interpret its slope as an indicator of decrease, stability and increase.

The example concept from the motivation of this section is repeated here:

"Completeness is decreasing and std. deviation is increasing"

Translated into a linguistic concept, the user may specify

$$\langle\, \Delta_{\text{comp}} \texttt{ is decr} \,\wedge\, \Delta_{\text{dev}} \texttt{ is incr} \,\rangle\,,$$

which is evaluated to

$$\top\left(\mu^{(\text{decr})}_{\Delta_{\text{comp}}}(C_j), \mu^{(\text{incr})}_{\Delta_{\text{dev}}}(C_j)\right),$$

where $\top$ represents a t-norm modelling the fuzzy conjunction (we use $\top_{\min}(a,b) = \min\{a,b\}$ in this paper). Fig. 3 depicts an example subgraph consisting of 9 nodes. Five time frames are shown with the respective edge weights. The graph is obviously becoming less dense with time, i. e., the completeness is decreasing. The chart for this measure is depicted in Fig. 4. In analogy to this, Fig. 5 shows the increasing deviation of the edge weights which is attributed to the emergence of the strong cooccurrence (the sudden appearance in this case can be explained with time frames that were too large to appropriately cover the short period during which this strong cooccurrence emerged). If we equipped the change rate domains

ISBN: 978-989-95079-6-8

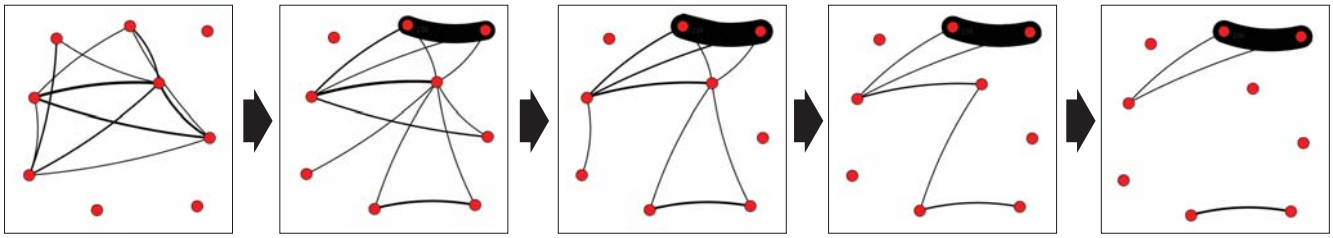


Figure 3: The temporal evolution of the graphs induced by a set of nine nodes. The number of edges is decreasing with time resulting in an almost isolated graph. Simultaneously, the edges that are remaining grow more and more unbalanced, i. e., the deviation of the edge weights is increasing. Both time series of the corresponding measures comp and dev are shown in Fig. 4 and Fig. 5, respectively.

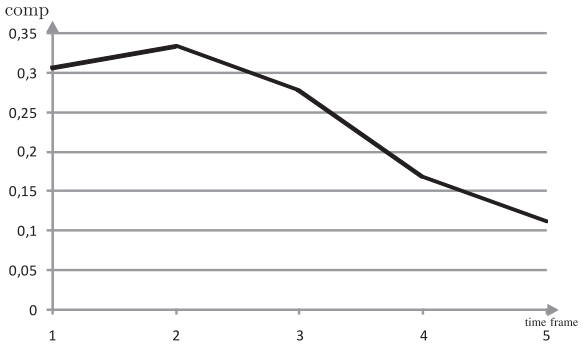


Figure 4: Time series with decreasing trend for the completeness of the edge weights for the five graphs of the time frames depicted in Fig. 3.

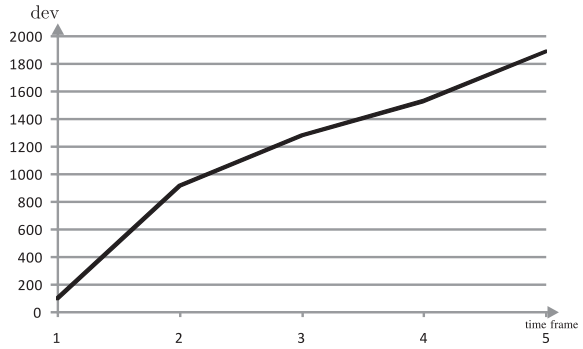


Figure 5: Time series with increasing trend for the standard deviation of the edge weights for the five graphs of the time frames depicted in Fig. 3.

(i. e., domains of the slopes of the regressions lines of the two time series) with appropriate fuzzy partitions (as we will do it in the experiments section) we could calculate the membership degree of this node set to the above-mentioned linguistic concept.

Summarizing, we state the following procedure:

1. Given a sequence $G^{(0)}, \ldots, G^{(n)}$ of cooccurrence graphs with their sum graph being G_Σ.
2. Based on an appropriate value θ, we calculate the candidate graph node sets $\{C_1, \ldots, C_m\}$ which are the vertices of the components of $G_{\Sigma,\theta}$.
3. The user provides a set of linguistic descriptions (fuzzy rule antecedents) that refer to the temporal change of the graph measures.
4. Provide fuzzy partitions for every domain of the change rate of the measures used in the descriptions of step 3.
5. Evaluate for every graph G_{C_j} the degree of membership to the linguistic concepts of step 3.
6. For every linguistic concept sort the graphs in descending order with respect to their membership degrees.

4 Experiments

Now that we have a tool at hand that allows us to determine the structural changes of subgraphs over time, we demonstrate the applicability with a real-world dataset that was already used to illustrate the examples above.

This dataset contains player contacts at certain locations within a 3D environment over a time period of six months. We carefully selected a subset of 100 such locations and discretized month-wise in order to result in a dataset large enough to justify the need for filtering while simultaneously being able to extract subgraphs that exhibit a structure and behavior that could not have been crafted more exemplary by hand. Therefore, we refrain from creating an artificial dataset with just the same structures.

Fig. 6 shows the sum graph of the dataset, i. e., the cooccurrences of six months among 100 locations. The edge weights indicate some subgraphs that are worth looking closer at. The threshold θ has been chosen to be 1000 to induce the candidate node sets.

We will match two linguistic concepts against these graph candidates: first, we are interested in decay, i. e., in graphs that show kind of a dissolving behavior, translating into a decreasing completeness and decreasing total weight. In the sample data this might indicate locations whose attractiveness is diminishing. A second concept that we would like to assess is that of an establishing pattern. An increasing average edge weight and deviation (of edge weight) might point out a phase of initial apparent random visiting of multiple locations which accumulates into a strong favored visiting pattern.

4.1 Concept 1: Decreasing Completeness and Weight

In order to evaluate the membership degrees to the linguistic concept

$$\langle\, \Delta_{\mathrm{comp}} \texttt{ is decr} \wedge \Delta_{\mathrm{wght}} \texttt{ is decr} \,\rangle,$$

ISBN: 978-989-95079-6-8

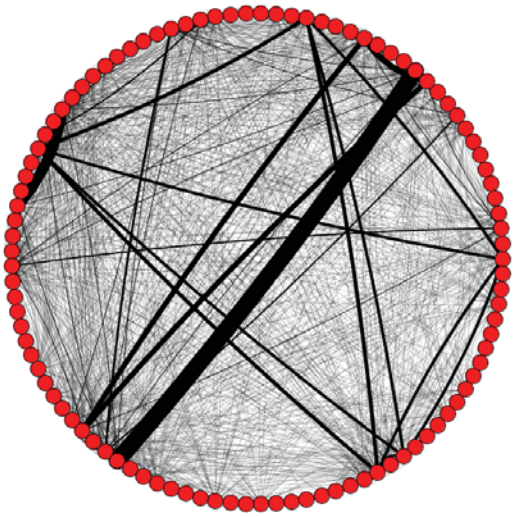


Figure 6: The sum graph of six months of visiting history of players in a 3D environment. We will match the major components (extracted via a user-specified threshold) against two linguistic concepts.

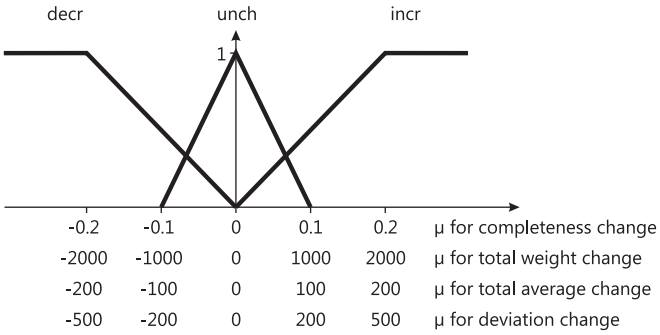


Figure 7: Fuzzy partitions for the four graph measures used in the two experiments.

we need to declare a fuzzy partition on the change rate domains of comp and wght. Fig. 7 displays all used fuzzy partitions. We apply three fuzzy sets. Note the asymmetric slopes of the borders. This setup has proven to be useful in this context since "unchanged" has a more strict semantic to users than the adjectives "decreasing" and "increasing". The respective values that determine the particular fuzzy partition can be read from the four different horizontal scales. These values have been selected with respect to the dataset since the quantity that renders a slope to be highly decreasing or increasing differs, of course, from dataset to dataset.

If we apply the linguistic concept to the candidate graphs and select the one with the highest membership (ignoring the remaining ones here for brevity), the subgraph whose history is depicted in the upper row in Fig. 8 scores 0.71. Most of the high degree can be attributed to the rapid loss of visits in the last two months. The membership degree was evaluated via

$$\min\{\mu^{(\text{decr})}_{\Delta_{\text{comp}}}(C_1), \mu^{(\text{decr})}_{\Delta_{\text{wght}}}(C_1)\} = \min\{0.71, 0.84\} = 0.71,$$

with C_1 being the set containing the five nodes. Data inspection revealed a newly set up structure which was heavily frequented shortly after opening followed by abating excitement.

4.2 Concept 2: Increasing Average and Deviation

We follow the same procedure to find the subgraph that scores best on the concept

$$\langle\, \Delta_{\text{avg}} \texttt{ is incr} \,\wedge\, \Delta_{\text{dev}} \texttt{ is incr} \,\rangle\,.$$

The lower part of Fig. 8 shows the resulting graph with a score of

$$\min\{\mu^{(\text{incr})}_{\Delta_{\text{avg}}}(C_2), \mu^{(\text{incr})}_{\Delta_{\text{dev}}}(C_2)\} = \min\{0.83, 0.89\} = 0.83,$$

The graph shows an establishing link between two nodes in parallel with a weakening in the remaining edges thus rendering the graph history becoming more unbalanced.

5 Conclusion and Future Work

In this paper we discussed the need to postprocess cooccurrence graphs if they grow very dense in practical use. We put forward a heuristic that allows to restrict the candidate node sets, and transferred the fuzzy concept matching approach from [6] to the graphical setting. We provided empirical evidence of the applicability by analyzing a real-world dataset containing six months of game player visits to 100 locations in a 3D environment.

As indicated in the introductory section, there are many other scenarios for which such an analysis might be interesting. We are most interested in applying our method to a web click stream analysis.

Since the used real-world dataset contained time stamps for every event, it is possible to generate a directed graph that also indicates which nodes were the source and target nodes for different users. This requires, of course, some heuristic with which to decide what should be the maximal period in which a user has to commit two visits to different nodes in order to assume that it was a transition and not just two independent visits.

A shortcoming of the presented approach is the heuristic to generate the candidate node sets. It worked well on the underlying datasets but one can imagine that it will return fewer and more interconnected components the more the threshold θ is reduced. Such subgraphs are also referred to as *giant connected components* [11]. To address this problem, we intend to phrase the problem in the setting of the emerging area of graph mining, more specific finding common substructures in a *single* graph [2, 12]. However, we will need a more specific subgraph definition since we have to account for the edge weights and not only the edge presence. This in turn automatically leads to another area of investigation: devising more sophisticated graph measures such as the edge betweenness centrality [13].

References

[1] Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami. Mining Association Rules between Sets of Items in Large Databases. In Peter Buneman and Sushil Jajodia, editors, *Proc. ACM SIGMOD Int. Conf. on Management of Data, Washington, D.C., May 26-28, 1993*, pages 207–216. ACM Press, 1993.

[2] Mathias Fiedler and Christian Borgelt. Subgraph support in a single graph. In *Proc. IEEE Int. Workshop on Mining Graphs and Complex Data (MGCS 2007 at ICDM 2007, Omaha, NE)*, pages 399–404. IEEE Press, Piscataway, NJ, USA, 2007.

ISBN: 978-989-95079-6-8

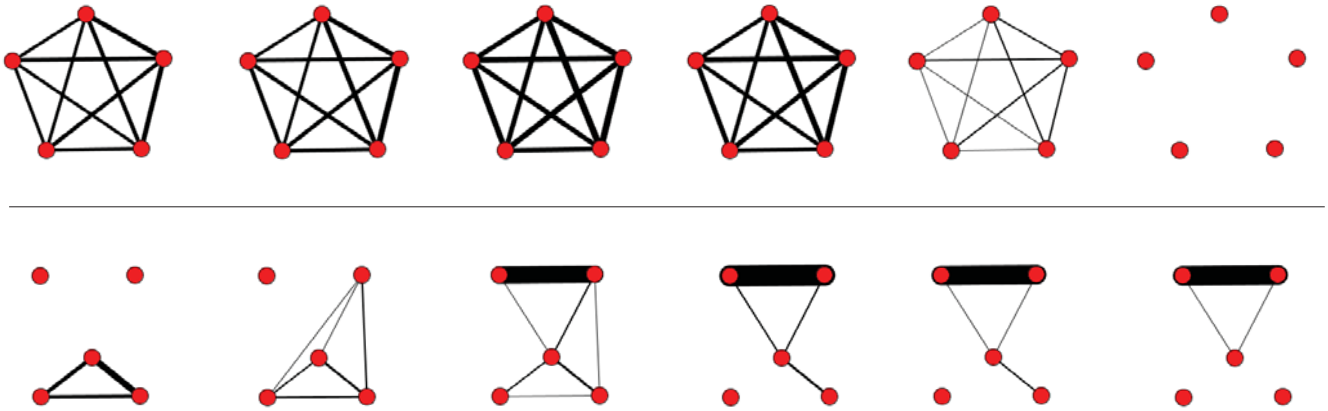


Figure 8: The histories of the two subgraphs that scored highest in the two experiments. The upper row shows a six-month development of five locations that were heavily visited but declined rather rapidly towards the end. It yielded a membership degree of 71% to the concept "completeness is decreasing and total weight is decreasing". The lower row depicts the best-scoring subgraph of the second experiment resulting in a membership degree of 83% to the concept "average is increasing and deviation is increasing."

[3] Christian Borgelt. On canonical forms for frequent graph mining. In *Workshop on Mining Graphs, Trees, and Sequences (MGTS'05 at PKDD'05, Porto, Portugal)*, pages 1–12, Porto, Portugal, 2005. ECML/PKDD'05 Organization Committee.

[4] Tobias Werth, Alexander Dreweke, Marc Wörlein, Ingrid Fischer, and Michael Philippsen. Dagma: Mining directed acyclic graphs. In IADIS, editor, *Proceedings of the ECDM 2008 (IADIS European Conference on Data Mining (ECDM))*, pages 11–17. IADIS PRESS, 2008.

[5] Rudolf Kruse, Christian Borgelt, Detlef D. Nauck, Nees Jan van Eck, and Matthias Steinbrecher. The role of soft computing in intelligent data analysis. In *Proc. 16th IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE'07, London, UK)*, pages 9–17. IEEE Press, Piscataway, NJ, USA, 2007.

[6] Matthias Steinbrecher and Rudolf Kruse. Identifying temporal trajectories of association rules with fuzzy descriptions. In *Proc. Conf. North American Fuzzy Information Processing Society (NAFIPS 2008)*, pages 1–6, New York, USA, May 2008.

[7] Levent Bolelli, Seyda Ertekin, and C. Lee Giles. *Knowledge Discovery in Databases: PKDD 2006*, volume 4213/2006, chapter Clustering Scientific Literature Using Sparse Citation Graph Analysis, pages 30–41. Springer Berlin / Heidelberg, 2006.

[8] Prasanna Desikan and Jaideep Srivastava. *Advances in Web Mining and Web Usage Analysis*, volume 3932/2006 of *Lecture Notes in Computer Science*, chapter Mining Temporally Changing Web Usage Graphs, pages 1–17. Springer Berlin / Heidelberg, 2006.

[9] M. E. J. Newman. Modularity and community structure in networks. *PROC.NATL.ACAD.SCI.USA*, 103:8577, 2006.

[10] Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical Review E*, 70:066111, 2004.

[11] Béla Bollobás. *Random Graphs*, chapter The Evolution of Random Graphs — the Giant Component, pages 130–159. Cambridge Univ Press, 2nd edition, 2001.

[12] Björn Bringmann and Siegfried Nijssen. *Advances in Knowledge Discovery and Data Mining*, volume 5012/2008 of *Lecture Notes in Computer Science*, chapter What Is Frequent in a Single Graph?, pages 858–863. Springer Berlin / Heidelberg, 2008.

[13] U. Brandes and C. Pich. Centrality estimation in large networks. *International Journal of Bifurcation and Chaos 17*, 7:2303–2318, 2007.

ISBN: 978-989-95079-6-8

Online Measurement of Oily-water Using Rule-based Fuzzy Concentration Map and Halogen Fluorescence

Gyeongdong Baek[1], Moon-Ho Choi[2], Tae-Ryong Jeon[1], Jaewoo Cho[1], and Sungshin Kim[1]

1 School of Electrical Engineering, Pusan National University, Busan, Korea
Email: {gdbaek, jtr, jwjw777, sskim}@pusan.ac.kr
2 MARSEN CO. Ltd., Busan, Korea
Email: master@marsen.co.kr

Abstract—*The aim of this study is to develop oil content measurement system using transmitted and scattered intensity. Two fluorescence analyses would be suggested with methods that it can use rapidly, continuously, conveniently and accurately in long-distance cruise vessels. The present paper consists of the structure of fluorescence measurement hardware and an interpretative analysis of experimental data. Proposed rule-based fuzzy concentration map offers the possibility of the conversion of fluorescence into oil content. This measurement method would be applied oil discharge monitoring control equipment in cruise vessels.*

Keywords—oil content meter, oil discharge monitoring system, fluorescence analysis, rule-based fuzzy concentration map.

1 Introduction

Maritime pollution controls are effected through the working of the MARPOL Convention and its associated Annexes, as developed by the IMO(international maritime organization) Maritime Environmental Protection Committee, MEPC. The vessel have an OFE(oil filtering equipment) on board. From 6 July 1998, all ships of 400 gross tonnages and above, regardless of age, must be fitted with 15ppm OFE. Some ships were able to use 100ppm OFE, but this is no longer acceptable. According to regulation 10(3)(b) of Annex I processed bilge water from machinery spaces is only allowed to be discharged into the sea through OFE system in a Special Area of Annex I to MARPOL 73/78 Convention [1].

The problem of detecting oil in water has practical significance, especially in the environment [2-3]. Optical measurement techniques are analytical methods commonly used in instrumentation for online measurement of oily water [4-5]. Online optical sensors have been used to monitor for example the coagulation and filtration processes in the water industry [6]. Online nephelometers have been used to measure oily water distributions [7]. In a study of smart nephelometer, artificial neural networks have been used to interpret online nephelometer data [8].

This paper, after working the sensing system with the halogen fluorescence, will attempt to analyze characteristic of the oil content. By comparison with the transmitted and scattered intensity of 25*mg/l* and 1000*mg/l* samples, proposed rule- based fuzzy concentration map is classified into different kinds of oil content. The following section provides a experimental system. Section 3 presents proposed rule-based fuzzy concentration map, section 4 and 5 show experimental result and contains concluding remarks.

2 Experimental System

A fluorescence measurement system is shown in Fig. 1. And system prototype is shown in Fig. 2. The experimental system consists of light emitting part, light receiving part, signal measurement part, and analyzing part.

2.1 Light Emitting Part

The component of the hydrocarbon of oil had strong fluorescence absorbance at the wavelength of about between 600*nm* and 900*nm* [9-10]. The halogen lamp provides the necessary spectral response for emission at between 200*nm* and 900*nm* [11]. So halogen light exhibits high absorption in

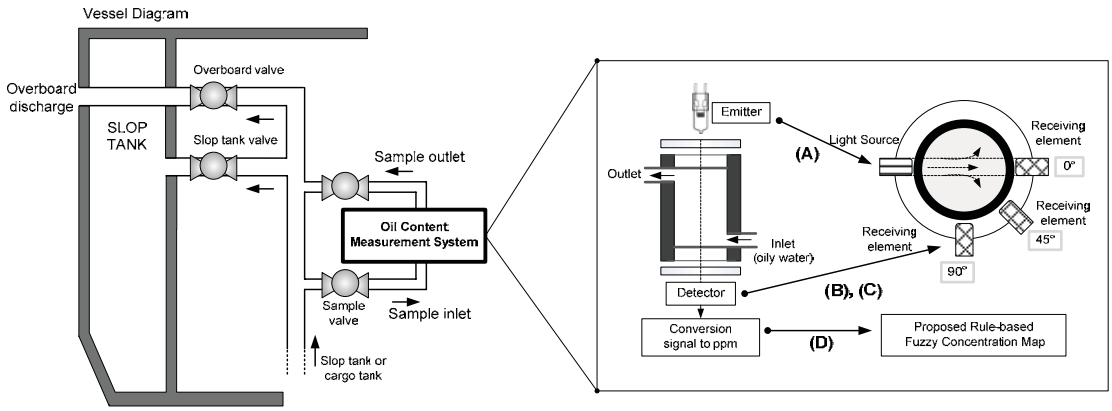


Figure 1: Schematic of fluorescence measurement system,
(A) light emitting part, (B-C) light receiving and signal measurement parts, (D) analyzing part.

ISBN: 978-989-95079-6-8

oil, system prototype was to use halogen lamp as the source of light.

2.2 *Light Receiving Part*

Since the required photo-detector of oily water should detect wavelengths about between 600*nm* and 1000*nm*, the most suitable sensor was found to be the photodiode that spectral response range is between 320*nm* and 1100*nm*. As in the case of the system prototype, voltages are being measured. Since photodiodes generate low current, it is connected to a resistor to transform the measured current to voltage as shown in Fig. 4. And resistor should have a large value in order to attain high voltages the data acquisition board, because there is no amplifier in this system. The transmitted and scattered fluorescence intensities are measured. Each sensor is positioned at 0˚, 45˚, 90˚ as shown in Fig. 1.

2.3 *Signal Measurement Part*

The signal sources of photodiodes are floating or nonreferenced signal source [12]. A floating signal source is not connected to the building ground system, but has an isolated ground-reference point. For examples of floating signal sources are outputs of thermocouples, battery-powered devices, optical isolators, and isolation amplifiers. Since the voltage level of the floating source can move, data acquisition board then saturates, causing erroneous readings. So resistors are used as illustrated in Fig. 3. These resistors, called bias resistors, provide a DC path. V_m is the measurement voltage. In this experiment, light oil samples were prepared with two types, 25*ppm* and 1000*ppm*. The data acquisition were being measured 30 seconds, and sampling rate of 2*kHz*.

2.4 *Analyzing Part*

The experimental results are shown in Fig. 4-5. There is a significant disparity between transmitted and scattered. It shows that the more oil content increases, the more transmitted intensity is decreased. And scattered intensities are increased. Given experimental results, the component of the oil contamination may be seen as absorbing and scattering wavelength based on the hydrocarbon. Although there is not a stark contrast between samples with marked oil content, data-sets have fluorescence patterns to classify. So attention was directed to conversion from fluorescence intensity to oil content using proposed rule-based fuzzy concentration map.

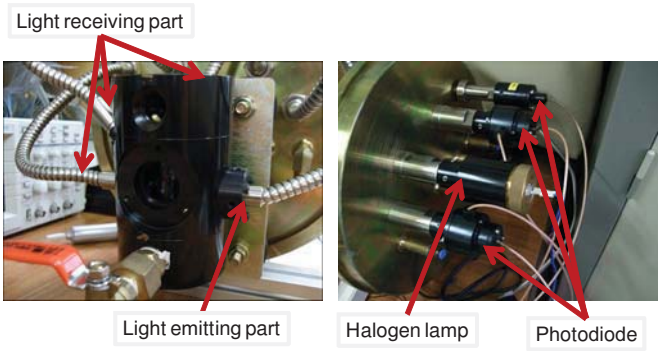


Figure 2: System prototype; Left: front of system, optical fibre to the sensors; Right: back of system, halogen lamp and photodiodes.

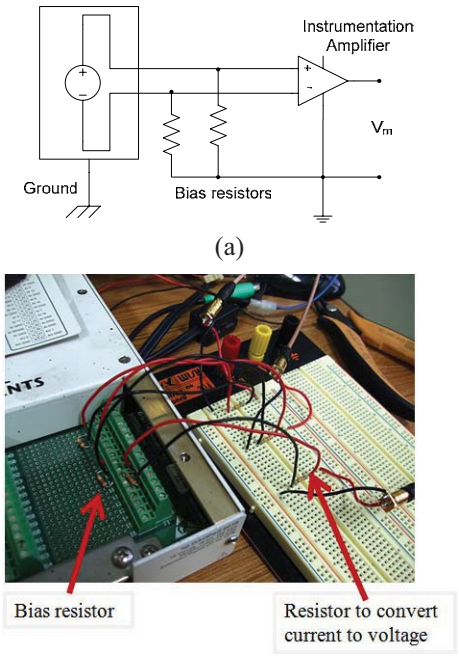


Figure 3: (a) Schematic of connecting a floating signal source, (b) Practical application.

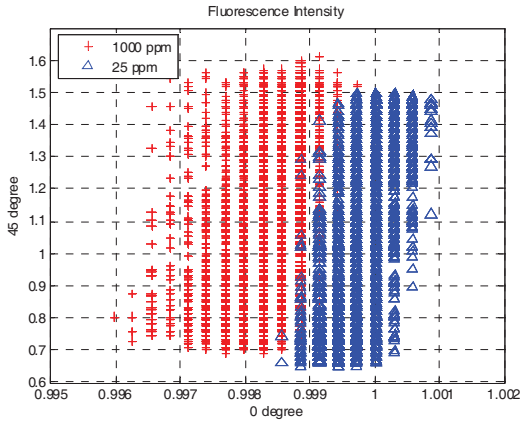


Figure 4: Intensity variations of transmitted (0°) and scattered (45°) in light oil.

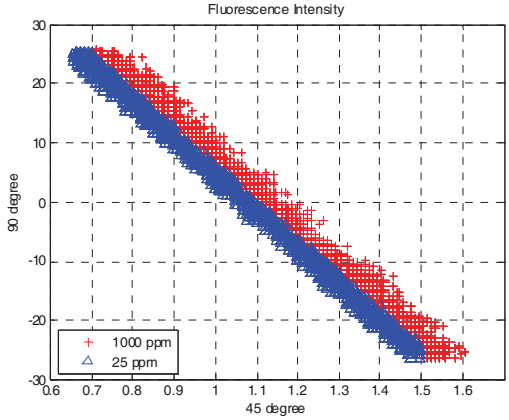


Figure 5: Intensity variations of scattered (45° and 90°) in light oil.

ISBN: 978-989-95079-6-8

3 Rule-based Fuzzy Concentration Map

Proposed algorithm consist of two variables of fuzzy inference system`s output and one fusion parameter as shown in Fig. 6. ρ_t, ρ_c, ρ_f are predicted oil content, λ is fusion effect ratio, and s_0, s_{45}, s_{90} are optical sensor signals that is positioned 0˚, 45˚, 90˚. The hypothesized relationship between transmitted intensity and scattered intensity is written:

$$\rho_f = (1-\lambda)\rho_t + \lambda\rho_c + \varepsilon \tag{1}$$

where ρ_t, ρ_c denote the independent variables of sub predicted oil content of ANFIS output, ρ_f denotes the final estimated oil content, λ is the fusion effect ratio, and ε is the noise term reflecting other factors that influence content.

The following sections provide proposed structure, fuzzy concentration map and multiple regression analysis.

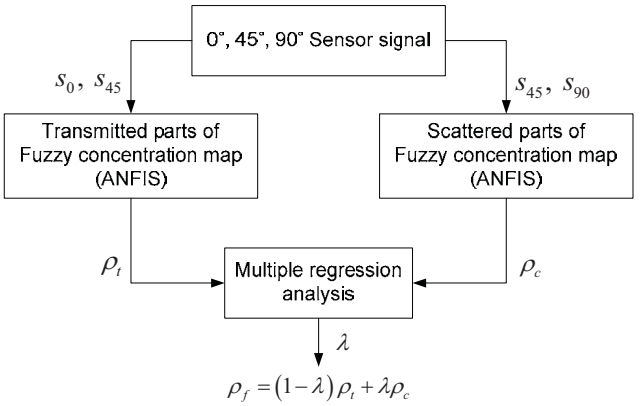


Figure 6: Proposed rule-based fuzzy concentration map.

3.1 Fuzzy Concentration Map

Proposed rule-based fuzzy concentration map presents a modeling method for oil content measuring in the form of ANFIS [13,14]. ANFIS(adaptive neuro-fuzzy inference system) represent TSK(Takagi-Sugeno-Kang) fuzzy models as shown in Fig. 7(a). ANFIS are adaptive networks that are functionally equivalent to TSK fuzzy inference system as shown in Fig. 7(b). *Prod* is the product of all incoming signal or *T*-norm operator that *AND* can be used. *Norm* is the normalized firing weights. The fuzzy concentration map shown in Fig. 7(c) classifies the input space of sensor signal into *n* regions which are assigned a labelled class of oil content. Every *n* region is an adaptive node with a node equation (2).

$$O_i = \bar{w}_i f_i = \bar{w}_i \left(p_i x + q_i y + r_i\right) \tag{2}$$

where $\bar{w}_i$ denotes the normalized firing weight from *i*, and p_i, q_i, r_i are the parameter set of *i* region. These are referred to as consequent parameters.

The fuzzy concentration map predicts oil content of transmitted and scattered part, which computes the overall output as the summation of all *n* regions in (3).

$$\rho_{ppm} = \sum_i O_i = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{3}$$

The ANFIS can be trained by hybrid learning algorithm with combining back propagation and least squares method. In the forward pass the algorithm uses least squares method to identify the consequent parameters. In the backward pass the errors are propagated backward and the premise parameters are updated by gradient descent.

In this paper, proposed method consists of two ANFIS. The two ANFIS have training dataset that is (0˚ sensor, 45˚ sensor, target *ppm*) and (45˚ sensor, 90˚ sensor, target *ppm*). Training was made using MATLAB software.

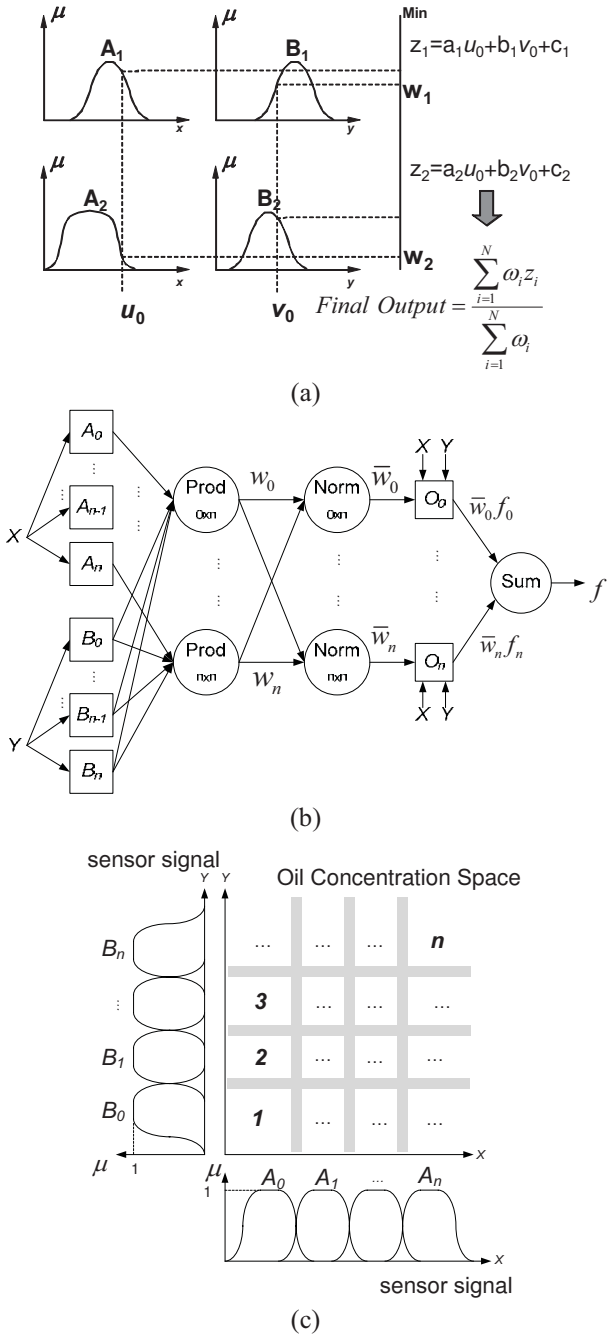
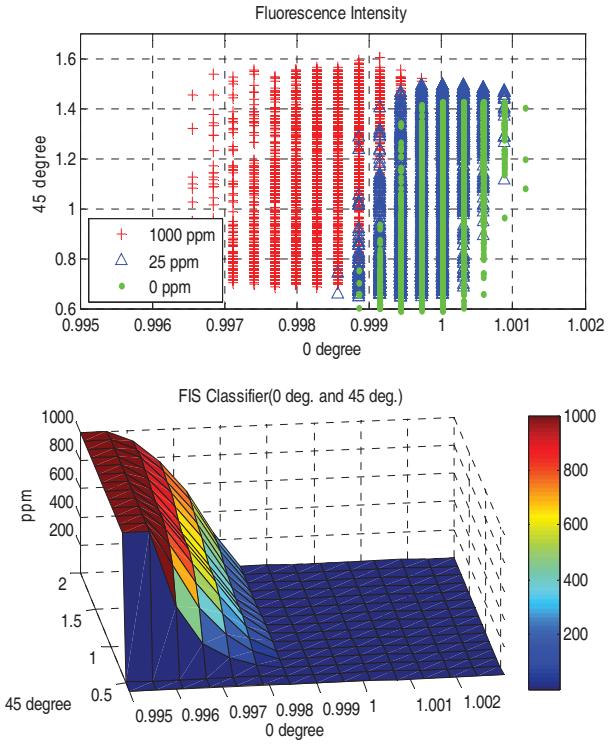
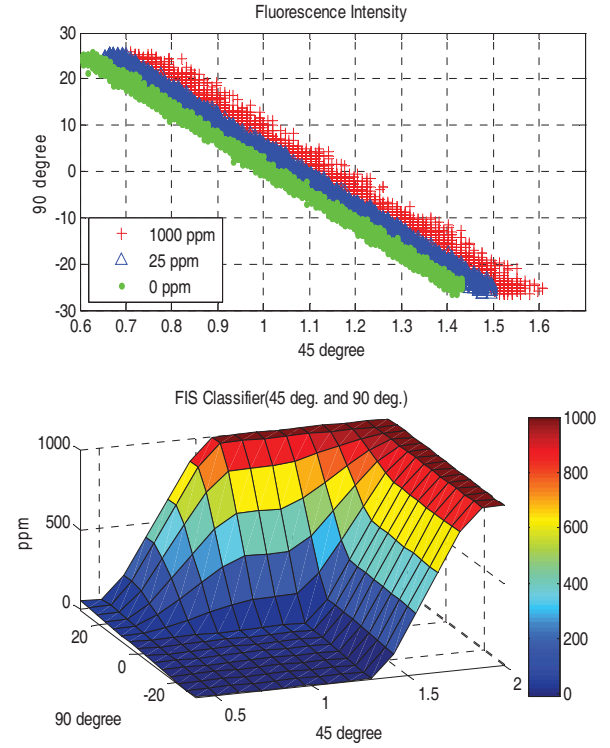


Figure 7: (a) Representation of TSK fuzzy reasoning, (b) Equivalent ANFIS Architecture, (c) Feature space partitioning of oil content.

ISBN: 978-989-95079-6-8



(a)



(b)

Figure 8: Intensity variations and FIS output surfaces; (a) Experimental results and classifier of 0° and 45°, (b) Experimental results and classifier of 45° and 90°.

3.2 Multiple Regression Analysis

Regression analysis is a statistical method for the investigation of relationships between variables [15]. Proposed rule-based fuzzy concentration map is similar to multiple linear regression analysis. Regression analysis is a statistical method for the investigation of relationships between variables. A dependent variable is called response variable or measurement, one or more independent variables are called explanatory variables or predictors in regression analysis. For example, a line in a two dimensional or two-variable space is written:

$$Y = a + bX \tag{4}$$

where a denote term of a constant or intercept, b is a slope as the regression coefficient or B coefficient.

The multiple regression procedures will be estimated in (5).

$$Y = a + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n \tag{5}$$

The purpose of multiple regressions is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable [16]. The computational problem that needs to be solved in multiple regression analysis is to fit a straight line to number of points. When there is more than one independent variable, the regression line cannot be visualized in the two dimensional space, but can be computed. With n explanatory variables, multiple regression analysis will estimate the equation of a hyper-plane in n-space such that the sum of squared errors has been minimized.

Given experimental data, it shows that the more oil content increases, the more transmitted intensities are decreased, and scattered intensities are increased. Multiple regression analysis provides an update on relation of intensity in (6).

$$\begin{aligned} \rho_f &= b_1 \rho_t + b_2 \rho_c \\ &= (1-\lambda)\rho_t + \lambda \rho_c, \qquad \lambda = 1 - e^{-\eta \bar{\rho}} \end{aligned} \tag{6}$$

where ρ_t, ρ_c denote the independent variables of sub predicted oil content of ANFIS output, ρ_f denotes the final estimated, λ is the fusion effect ratio, η is the gradient scale, and $\bar{\rho}$ is the mean of two predicted oil content.

4 ANFIS Results and Discussion

This paper has attempted to make online measurement system of oil concentration with insufficient samples. It was used samples of 25 *ppm* and 1000 *ppm*, and analyzed contrast between samples with marked oil concentration. Proposed rule-based fuzzy concentration map method was used to convert sensing signal to *ppm*. The generated FIS output surface is shown in Fig. 8. When the measurement voltages are compared to the inference of oil content, this analysis method seems reasonable. After analysis method was established, insufficient content ranges of oil sample were calcu-

ISBN: 978-989-95079-6-8

lated as shown in Table 1. While The output of ANFIS(0° and 45°) was no change in good condition, FIS(45° and 90°) was sensitive. And the more *ppm* increases, the more ANFIS(45° and 90°) increases than ANFIS(0° and 45°). Considered in this relation, λ and η of multiple regression analysis parameters should be updated.

Table 1: Predicted *ppm*(*mg/l*) based on ANFIS.

$[A, B]_x = A \le x \le B$, condition λ=0.5

0 deg.	45 deg.	90 deg.	ANFIS(*ppm*) (0˚&45˚)	ANFIS(*ppm*) (45˚&90˚)	***ppm***
[0.998, 1.001]	[0.6, 1.4]	-20	[0.951, 91.184]	[4.492, 87.651]	[2.721, 89.417]
		-10	[0.951, 91.184]	[17.571, 99.364]	[9.261, 95.274]
		0	[0.951, 91.184]	[28.592, 109.550]	[14.771, 100.367]
		10	[0.951, 91.184]	[52.605, 314.557]	[26.778, 202.871]
		20	[0.951, 91.184]	[106.429, 806.196]	[53.690, 448.690]

5 Conclusion

In this paper, we have developed online content measurement system using transmitted and scattered intensity. We attempt to outline an alternative analysis of measurement of oil content using ANFIS. Two part of fluorescence intensities can be used oil content information. Given experimental data-set, it shows that the more oil content increases, the more fluorescence absorbance and scattering of hydrocarbon (-CH) are increased. So transmitted intensity is decreased, and scattered intensities are increased. Proposed rule-based fuzzy concentration map offers the possibility of conversion sensor signal into *ppm*. Because proposed method uses hybrid learning algorithm, the proposed method can provide an input-output mapping. Proposed analysis method and sensing system provide a stepping stone for developing basic technique for designing oil content meter.

Acknowledgment

This work was supported by the Grant of the Korean Ministry of Education, Science and Technology (The Regional Core Research Program/Institute of Logistics Information Technology).

References

[1] MARPOL, How to do it, Manual of the Practical Implication of Ratifying and Implementing MARPOL 73/78, *International Maritime Organization*, London, 1993.

[2] Clarke. R.B, Marine pollution, *Oxford Science 2nd edition*, 1989.

[3] Mason, C.F., Biology of fresh water pollution, *Longman Harlow 2nd edition*, 1991.

[4] Zhu B., Yuan H., Classification of oils an water using LIF spectrum parameters independent of the thickness of oil film, *SICE '95. Proceedings of the 34th SICE Annual Conference*, pp. 1539-1541, Jul. 1995.

[5] Smith, P.R., Green, D.A., Naimimohasses, R., Some investigations on neural processing of scattered light in water quality assessment, *IEE Proceedings Vision, Image and Signal Processing*, vol. 141, no. 4, pp. 261-266, Aug. 1994.

[6] Eisenlauer, J., Horn, D., Fibre-optic online flocculant dose control in water treatment operations, *Colloids and Surfaces 25*, pp. 111-129, 1987.

[7] Snel, D., Pitt, G.D., Oil content monitoring (practical considerations), *International conference proceedings on Optical techniques and process control*, pp.27-41, 1983.

[8] Smith, P.R., Green, D.A., Naimimohasses, R., Thomason, H., Aritificial neural network processing of scattered light measurements in the detection of immiscible water pollutants, *3rd international conference on Artificial neural networks*, pp. 282-285, 1993.

[9] Haavard A., Trond S., Reidar B. S., An in-line NIR/endoscope technique for observations in real hydrocarbon multiphase systems, *Flow Measurement and Instrumentation*, vol. 16, pp. 289-293, 2005.

[10] Xin B., Liankui D., Partial least squares with outlier detection in spectral analysis: A tool to predict gasoline properties, *Fuel*, vol. 88, pp. 1216-1222, 2009.

[11] Cristina Yubero, M. Carmen Garcia, M. Dolores Calzada, Using a halogen lamp to calibrate an optical system for UV-VIS radiation detection, *Optica Applicata*, vol. 38, no. 2, pp. 353-363, 2008.

[12] National Instruments, Field wiring and noise consideration for analog signals, *NI developer zone*, Feb. 2006.

[13] Jyh-Shing R.J., ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE Trans. on Systems, Man and Cybernetics*, vol. 23, no. 3, 1993.

[14] Kwang H. Lee, First Course on Fuzzy Theory and Applications, *Springer*, ISBN-10: 3540229884, 2004.

[15] Rawlings, J.O., Pantula, S.G., Dickey, D., Applied Regression Analysis, *Springer Verlag*, ISBN-10: 0387984542, 1998.

[16] Timothy Z. Keith, Multiple Regression and Beyond, *Prentice Hall*, ISBN-10: 0205326447, 2006.

Features stereo matching based on fuzzy logic

Grzegorz Nieradka, Bohdan Butkiewicz
Warsaw University of Technology
00-665 Warsaw, Nowowiejska 15/19 str, Poland
Email: G.Nieradka@ise.pw.edu.pl, B.Butkiewicz@ise.pw.edu.pl

Abstract— *This paper presents an entire scheme for the estimation of a sparse disparity map from stereo pair images. In contrary to a dense disparity map, for which disparity values are calculated for each pixel of image, the sparse disparity map is determined only for some distinguished set of pixels from an image. The pixels belong to this set are called features. Therefore, in the first stage the algorithm for determining sparse disparity map has to be able to detect and specify some pixels as the feature pixels. In this article a method for specifying features is introduced. The core of the presented method relies on a fuzzy edges detector. The algorithm of fuzzy edges detection, as well as a manner enabling determination of the features' set, are introduced. The disparity is calculated at the fuzzy domain based on a similarity measure which is the correlation of fuzzy sets. The results of the obtained disparity maps for some benchmark stereo pairs, and the comparison with the well known Marr-Poggio-Grimson algorithm designed for sparse disparity map estimation are presented.*

Keywords— feature points, stereo matching, disparity map, edges detection, fuzzy logic

1 Introduction

Stereo vision is a vivid researched technique of machine vision area which allows to obtain a three dimensional image of a scene observed. Its importance arises from its main characteristics, hence, it is a passive method and the equipment which allows to use it is rather cheap. The estimation of disparity map is one of the central problems of stereo vision application.

In spite of thirty years of research there is no good enough method which always gives a satisfactory solution. The stereo vision and especially the stereo matching research could be divided into two main streams [1, 2]. One of them are algorithms which try to match all pixels from images. As a result of this kind of algorithms the so called dense disparity map is obtained. [3]. The "dense" means that the disparity values are estimated for each pixel of the images. These algorithms, however suffer from the calculation being time consuming and the matching in uniform regions of images being almost impossible. But doubtless the advantage of this group of methods is better model reconstruction possibility. This paper does not consider the methods which produce the dense disparity map.

An alternative approach towards finding a solution to the stereo vision problem is a feature matching strategy. Although, the obtained disparity map has established values in much less amount of pixels, in contrary to the former group of methods, the result can be used in many real tasks as navigation of autonomous robots, manipulation of industrial robots or view morphing. Moreover, the disparity values could be interpolated in the areas where the matching is not made. The disadvantage of this approach is an insufficient amount of disparity values for visually good reconstruction of observed scene. On the other hand, the methods based on features reduce the searching space from full image pixels set to the pixels which were recognized and assigned as the features. It attracts reduction of calculation time and in a natural way, it causes eliminating the ambiguity that may occur inside the uniform regions of images, because in these regions there are usually no feature points.

Each clearly distinct pixel from image can be selected as the features considered for matching task. However, the most popular features taken under consideration are corners [4, 5] and edges [6]. The edges matching techniques are more popular for the reason that the disparity map estimated with dependence only on corners is too sparse. Matching the edges of the images gives a lot of more dense disparity map therefore matching edges is considered in this paper.

In another group of feature based approaches image segmentation is done as the first step. Later, they try to match the segmented regions [7]. This group of methods is not considered here, and the comparison of results would not give meaningful results because the ability of estimation disparity map depends mainly on the segmentation step, which is a totally different issue in machine vision research.

In order to improve results of features stereo matching there is a big group of methods which aim at matching piecewise-linear edge segments [8, 9]. The matching of edge segments has advantages that the error of isolated pixels in edges has very little influence on the position and orientation of the edge segments. There are different modifications in the descriptions of edge segments [10] and the edges could be matched as edge chains [11]. This group of methods is out of our interest because they do not depend on an edge detector and have a different philosophy behind them than matching edges directly.

The well known algorithm for matching features is the Marr-Poggio-Grimson algorithm, MPG in short [12, 13]. This algorithm is an implementation of the human stereo vision theory that was developed by Marr and Poggio [14]. Because the MPG algorithm works mainly with reliance on the edge detection it has been used here as the reference algorithm. All the details of this algorithm are given in the Sec. 5.1.

The paper proceeds in three stages: the first gives a formal statement of the stereo matching problem; the second part contains an algorithm based on fuzzy logic for edges detection and a manner for establishing feature points as well as a method for the calculation of a disparity map; the third part of the paper presents evaluation results of the introduced method and the comparison with the MPG algorithm.

ISBN: 978-989-95079-6-8

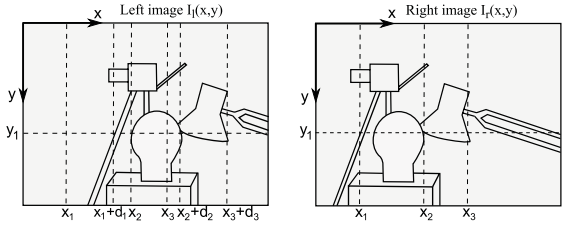


Figure 1: Depiction of the stereo matching problem

2 Problem statement

The stereo vision problem relies on establishing the correspondence between physically the same points in two different images representing the same scene. However, these two images have to be taken under some geometric restrictions in order to ensure they constitute a proper stereo pair [15, 16]. If two images constitute a proper stereo pair then there is a horizontal shift between these two images called *disparity*. Finding this shift is the main aim of almost every stereo matching algorithm. This task is illustrated in Fig. 1 which shows, that the same physical points (x_1, x_2, x_3 in the Fig. 1) in the left and the right images have different abscissa in the image coordinate system. The matching task is usually posed as the problem of finding points in the right image I_r corresponding to the selected points in the left image I_l. The value of abscissa shift of the determined point, called *disparity* is here refereed to as d. The task is to find the proper values of disparity d which correspond to the same physical points.

Usually, the epipolar constraint [15] restricts the search to a horizontal line of images. A further simplification is attained by limiting the maximum possible value of disparity d, for which a good estimate can exist. In this paper we restrict our considerations to a parallel rectified pair of cameras. This means that the searching is done for each line of images and the lines of the left and right images correspond to each other.

3 Features detector

Edge detection is a well known problem for the machine vision community. There are many more or less popular algorithms in the literature considering on image processing tasks [17, 18]. In this paper we used some new kind of an edge detector. This edge detector is based on fuzzy relation and with its characteristic feature being rather fat edges obtained as the detection results [19]. The thickness of the detected edges could be a disadvantage in some applications but in the matching problem has become an advantage, because the edges have more information about themselves and about their neighborhood.

3.1 Edges detection

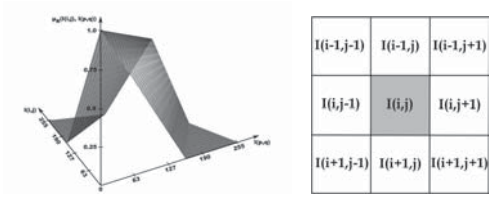
The principle behind the presented algorithm is the fact that for the pixel which belongs to an edge, the neighboring pixels have different values. Therefore if the edge is strong then in the determined small neighborhood of the pixel under consideration there are numerous pixels with the big difference of value.

The distinction between these pixels is made by using the fuzzy relation [20]. As it is pointed in [19] a lot different shapes of fuzzy relation can be used. In this paper we used the fuzzy relation in the triangular form (1). The shape of the triangular similarity relation is shown in Fig. 2(a) and its strength depends mainly on α value.

Formally the triangular fuzzy relation between the pixels $I(i,j)$ and $I(k,l))$ is expressed:

$$\mu_{RT}\left(I\left(i,j\right),I\left(k,l\right),\alpha\right)=\begin{cases}1-\frac{|I(i,j)-I(k,l)|}{\alpha} & if \quad |I\left(i,j\right)-I\left(k,l\right)|<\alpha\\ 0 & if \quad |I\left(i,j\right)-I\left(k,l\right)|\geq\alpha\end{cases} \tag{1}$$

In the edge detection stage the distinctions between one distinguished pixel $I(i,j)$ and its neighbors are calculated. If the differences are big, there are likely big changes in image, and there could be an edge present. The algorithm for edge detection is described as follows:



(a) (b)

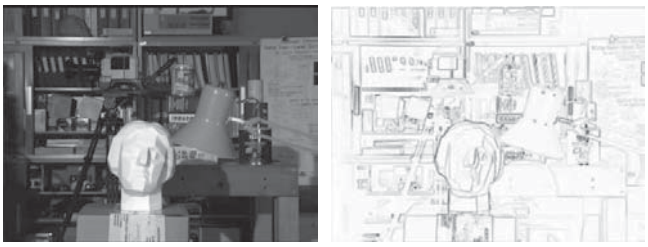
Figure 2: The shape of triangular fuzzy relation used for the edge detection (a) and the neighborhood (b) of distinguished pixel.

1. The determination of the differences between the distinguished pixel and its small established neighborhood. This situation of the distinguished pixel and its neighboring pixels is shown in Fig. 2(b). In this case we use the window size eight but many different sizes of this neighborhood can be used [19].
   The vector of homogenity description h_v is defined. The length of homogenity vector h_v is equal to the size of the neighborhood. For each pixel of the image the values of the vector h_v are calculated by (2):
   $$h_v=\bigcup_{k,l=-\rho}^{\rho}\mu_{RT}\left[I\left(i,j\right),I\left(i+k,j+l\right)\right] \tag{2}$$
2. The final response of the edge detector is defined as the sum of the homogenity vector values:
   $$h_e\left(i,j\right)=\sum h_v \tag{3}$$

If the region is uniform, the response of the edge detector would be big because all of the pixels have big similarity close to one. If the region is diverse, and if a strong edge is present in the region, the response of the detector is smaller. The results of the detector working for two images are shown in Fig. 3 and 4. We reverse the colors of the edge images in order to get the better presentation in black in white background. Originally the colors for the edges are brighter, the bigger response of the edge detector while the background is black.

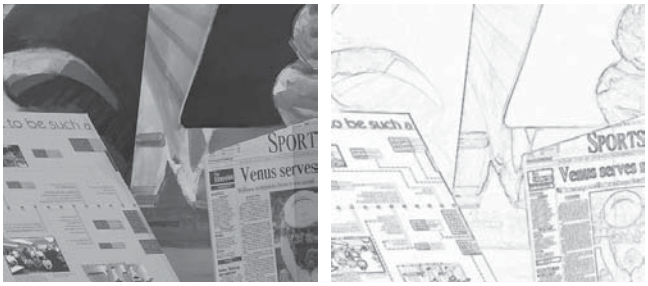
Moreover, the values of the edge detector response called "edginess" could be considered as values of some membership function. The bigger the response, the bigger the pixel's membership function value.

ISBN: 978-989-95079-6-8



(a) (b)

Figure 3: The left of stereo pair image called "Tsukuba" (a) and the result of detected edges (b).



(a) (b)

Figure 4: The left image of an another stereo pair called "Venus" (a) and the result of edges detection (b).

3.2 Edges thinning

The edges detected by the presented edges detection procedure are rather fat therefore we need to further reduce the number of points which are subject to match-making. In order to obtain the reduction of the edge points, a thinning algorithm is executed at the edge images. The thinning is done in three independent steps.

The full thinning algorithm proceeds in the following way:

1. *Thresholding:* The first step of the edge thinning is attenuation of small values by using the global threshold P_{th} with the value given by:

$$P_{th} = a_{th} \times E_{mean} \tag{4}$$

where a_{th} is an arbitrary chosen empirical threshold value. In this work it was set at the value equals 1.25. The term E_{mean} determines the average value of the edges image:

$$E_{mean} = \frac{1}{M \cdot N} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} e\left(i,j\right) \tag{5}$$

where $e\left(i,j\right)$ is the pixel value of the edge image.

2. *Horizontal non-maximum suppression:* This step has similar idea to the non-maximum suppression algorithm. But in this step, it is made only in the horizontal direction of the image, as carried out:

$$p_e\left(i,j\right) = \begin{cases} p_e\left(i,j\right) & \textbf{if} \quad p_e\left(i,j\right) > p\left(i-1,j\right) \\ & \textbf{and if} \quad p_e\left(i,j\right) > p\left(i+1,j\right) \\ 0 & \textbf{if} \quad else \end{cases} \tag{6}$$

3. *Vertical non-maximum suppression:* The last step is practically the same as the previous step, but it is done in the vertical direction of the image:

$$p_e\left(i,j\right) = \begin{cases} p_e\left(i,j\right) & \textbf{if} \quad p_e\left(i,j\right) > p\left(i,j-1\right) \\ & \textbf{and if} \quad p_e\left(i,j\right) > p\left(i,j+1\right) \\ 0 & \textbf{if} \quad else \end{cases} \tag{7}$$

As the result the images with the edges of one pixel thickness are obtained. In these images the positions of the remaining pixels determine the locations of the features which are to be matched.

The examples of the results obtained by the edge thinning algorithm and the locations of the feature points for the images considered earlier (see Fig. 3 and Fig. 4) are shown in Fig. 5.

(a) (b)

Figure 5: The feature points after edge thinning applied for the images in Fig. 3 (a) and Fig. 4 (b) respectively.

4 Feature points matching

A typical algorithm for finding a stereo matching solution works for an arbitrary range of disparity values from d_{min} to the d_{max}. There are two windows established, one of which, called a reference window is set at a stable position in one (usually left) of the images constituting a stereo pair and the other called a search window moves along the horizontal line in the other image (usually right) within the established disparity range. For each position of the search window some kind of matching measure is calculated. The value of d for which position of the search window the value of measure attains its supreme is taken as the proper value.

The presented algorithm works in the mentioned scheme, that is reference and search windows, but during the matching stage the feature image and the edges images are used. The feature image determines the position of the feature points received after using the edge thinning. It is used for establishing the position of the reference window in the reference edge image exactly at location of the feature point. After establishing the position of the reference window the matching is carried out by moving the search window in the other edge image.

The measure of matching is calculated based on the edge images. For each point which have determined as the feature point in the feature images, the calculation of matching score in the edge images is done. As it was mentioned earlier the values of the edge detector can be interpreted as the values of the membership function which gives the degree of pixel belonging to the "edginess" variable. For this reason the calculation of matching score is done in a fuzzy fashion.

ISBN: 978-989-95079-6-8

The matching measure used in this algorithm is the correlation between fuzzy sets [21]. Let us assume that there is a sequence of paired data $((\mu_A(x_1), \mu_B(x_1)), \ldots, (\mu_A(x_n), \mu_B(x_n)))$ which corresponds to the grades of the membership functions of fuzzy sets A and B defined on X. The correlation coefficient, $f_{A,B}$, between the fuzzy sets A and B is calculated as:

$$r_{A,B} = \frac{\sum_{i=1}^{n} (\mu_A(x_i) - \bar{\mu}_A)(\mu_B(x_i) - \bar{\mu}_B) / (n-1)}{S_A \cdot S_B} \quad (8)$$

where $\bar{\mu}_A$ and $\bar{\mu}_B$ denote the average membership grades of the fuzzy sets A and B according to the dependence:

$$\bar{\mu}_A = \frac{\sum_{i=1}^{n} \mu_A(x_i)}{n} \quad (9)$$

and S_A and S_B are the standard deviations of fuzzy sets A and B calculated as: $S_A = \sqrt{S_A^2}$, where S_A^2 represents the degree of variation of the membership function determined by:

$$S_A^2 = \frac{\sum_{i=1}^{n} (\mu_A(x_i) - \bar{\mu}_A)^2}{n-1} \quad (10)$$

The measure of correlation (8) is simply adapted to the measure of correlation between "edginess" in the reference window and "edginess" in the search window. As in a typical matching algorithm the score is calculated for each shift from the chosen range and the value d_k for which the correlation attains its maximal value is taken as the proper disparity value.

5 Results evaluation and comparison

For the testing purposes we used some benchmark stereo pairs made public accessible [3] from the Middlebury Stereo Vision Page [22]. The big advantage of these stereo pairs is that they have the true disparity map available. This allows to make a reasonable comparison of different algorithm in reference to the true result which should be obtained.

5.1 Reference feature matching algorithm

As the reference algorithm the Marr-Poggio-Grimson (MPG) has been selected. This algorithm is very popular, easy to implementation and gives good results. The implementation which is used here was based on the scheme described in [12, 13].

The MPG algorithm works in the coarse to fine scheme. In the description of this algorithm the word "channel" is used and it refers to image filtering with a specific filter window with established size w. The coarse channel means the filtering with the bigger size of filter window and the finest channel means that the image was filtered by using filter with smaller size.

The typical scheme of the MPG algorithm consists of six steps. The steps are outlined below:

1. *LOG filtering:* Started from the coarsest channel the left and right image of stereo pair is convolved with LOG filter. The LOG is a smoothed second derivative of the image signal and assumes the following form:

$$\nabla^2 (G(x,y)) = \left(\frac{x^2+y^2}{\sigma^2}\right) \exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right) \quad (11)$$

where ∇^2 is the Laplacian $\nabla^2 = \left(\frac{\partial^2}{\partial x^2}\right) + \left(\frac{\partial^2}{\partial y^2}\right)$ and $G(x,y))$ is the Gaussian function, which acts as a low-pass-filter to the image:

$$G(x,y) = \sigma^2 \exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right) \quad (12)$$

where the width of the channel "w" is related to σ as follows: $w = \sqrt{2}\sigma$

2. *Zero crossing extraction:* In the filtered image the detection of zero crossing is done by scanning the image horizontally for adjacent elements of opposite sign or for three horizontally adjacent elements, where the middle one is zero. Also the sign of zero crossing is searched. The located position of zero crossing with their signs are remembered.

3. *Matching:* For each scan line in the left image is centered reference window at the founded zero crossing point as the possible candidate to matching. In the right image the search window is moved in established disparity range. If in the right image is a zero crossing point on the search window and has the same sign as the left point, then this zero-crossing point produces a match. The disparity, difference between locations of zero crossing points in the left and right image, is stored in a dynamic buffer.

   Based on the matching process, the left zero crossing is marked as:

   - *unique match*, if only one right image zero crossing is matched with the left one
   - *multiple matches*, if more than one match is found
   - *no match*, if no match is found

4. *Disambiguation:* In this step the disparity map is checked for possible double matches. This kind of ambiguity is resolved by checking the disparities within the same region of the representation at the previous channel. If there is no coarser channel, or there is no a disparity value within this region in the coarser channel, or the disparity is not consistent with coarser level disparity, then the disparity is discarded. The disparity can be approximated by taking the average of the multiple match.

5. *Loop:* When the final map disparity for the current channel has been completed, the process return to the convolution to the next finer representation.

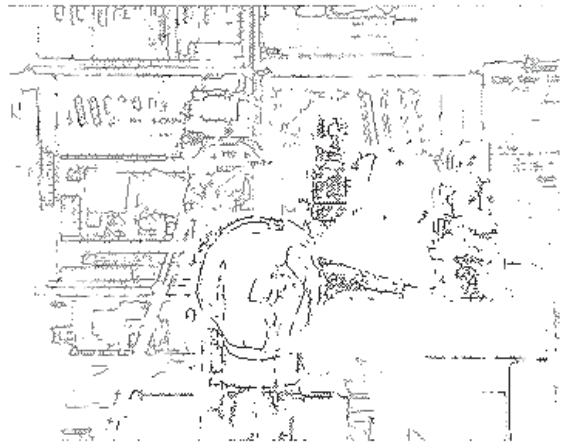
6. *Consistency:* After calculation of disparity in each channel has been completed, there is one final test done. Each disparity value at the finest channel is tested for consistency by checking its value with values at the coarser channel. If the values are inconsistent the matching is eliminated.

In this paper we used, as in the Grimson [12] implementation, four channels with size of 5, 9, 13 and 17 pixels respectively.

ISBN: 978-989-95079-6-8

5.2 *Results evaluation*

The algorithm proposed here and the MPG algorithm were applied for two different stereo pairs. The Fig. 6 presents the estimated disparity maps for "Tsukuba" stereo pair. Similar as in the edge images the colors were reversed for better visual presentation. If the color at the disparity map is darker the original disparity value has bigger value, and the element has bigger disparity value. The second image Fig. 7 presents the disparity maps obtained for the "Venus" stereo pair images.



(a) The result of the proposed algorithm.

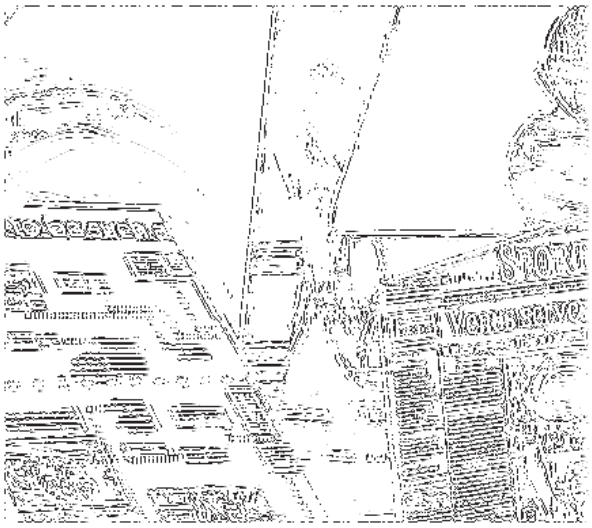
(b) The result of the MPG algorithm.

Figure 6: Disparity maps for the stereo pair "Tsukuba" (a) estimated by the proposed fuzzy algorithm and the MPG (b).

Visually these results are not so different. In the "Tsukuba" stereo pair can be seen that MPG algorithm gives some different matching but the comparison is very hard to done. Just comparison of numerical statistic given in the Tab.1 gives better view at obtained results. It can be seen that the amount of points classified as the features points is similar in both algorithms but the proposed algorithm based on fuzzy edge detection gives much more true matches. During generation of the results enclosed in Tab. 1 a point was classified as the proper matching point if the difference between its disparity and the true disparity taken from the true disparity map was equal zero or one. In this case the small mistake equal one was allowed. If the difference between estimated disparity value and the true value was bigger than one the matching point was classified as wrongly matched point.

In order to check the characteristics of these algorithms in some more deep manner the second experiment was done. In this case their behavior in the presence of noise in images was investigated. For this reason some amount of Gaussian noise was added to the both images consist of stereo pair. The amount of added noise is determined in the logarithmic scale and was established at $30dB$ value. The value denotes the amount of noise and is calculated by using expression:

(a) The result of the proposed algorithm



(b) The result of the MPG algorithm

Figure 7: Disparity maps for the stereo pair "Venus" (a) estimated by the proposed fuzzy algorithm and the MPG (b).

$$SNR = 10 \cdot \log_{10}\left(\frac{P_{image}}{P_{noise}}\right) \tag{13}$$

For discrete image I with size $M \times N$ and intensities $I(i,j)$ of pixels the power P_{image} is expressed as following:

$$P_{image} = \frac{1}{M \cdot N}\sum_{i,j=1}^{M,N} I^2(i,j) \tag{14}$$

For the purpose to produce the image with assigned SNR to the image the Gaussian noise was added iteratively and the power of noise was calculated for the noise image:

$$I_{noise} = I_{distorted} - I_{original} \tag{15}$$

ISBN: 978-989-95079-6-8

Table 1: The quantitative results of comparison for matching algorithms.

Image	Points	Right	Bad	Right %	Bad %
Results obtained by the fuzzy matching					
Tsukuba	11579	10392	1187	89.8	10.2
Venus	18588	16479	2109	88.7	11.3
In the presence of noise in image; $SNR = 30dB$					
Tsukuba	11898	2629	9269	22.1	77.9
Venus	19269	3313	15956	17.2	82.8

Results obtained by the Marr-Poggio-Grimson algorithm					
Tsukuba	11579	9549	2030	82.4	17.6
Venus	19143	12449	6694	65.1	34.9
In the presence of noise in image; $SNR = 30dB$					
Tsukuba	12345	582	11763	4.7	95.3
Venus	22705	564	22141	2.5	97.5

where the power $I_{original}$ had been calculated (13) before adding noise and the P_{noise} was calculated for the noised image. The amount of noise had been increasing to the moment for which the SNR attained the value equal to $30dB$.

As it is shown in Tab.1 both algorithms have very small resistance at the noise. The number of features does not decrease but the the amount of the proper matching is decreasing very significantly. But the proposed fuzzy approach still give better results.

6 Conclusion

In this paper the full scheme for features stereo matching based on fuzzy edge detection has been presented. The application of fuzzy logic allowed to obtain results better than the good and stable MPG algorithm.

However, both of these algorithms are very sensitive to the noise present in the images under matching consideration. It seems that the most sensitive stage is the edge detection. This aspect of these algorithms should be researched more deeply because the noise is very important in the real applications of stereo vision. In the real world numerous kind of distortions should be considered and this topic should be examined more carefully.

The obtained results are probably also image depended. In the case where the image is more complicated and there are many more edges in an image, as in "Tsukuba" stereo pair, the number of properly matched points is always bigger. In the "Venus" stereo pair, in which long straight edges are present, both algorithms found much more points classified as the feature points but the matching in both cases was worse. In that kind of images probably the algorithms based on segments edge matching would give better results. But when the image encloses a big number of short edges algorithms presented here should have better properties. This is probably a clue for further research but it is out of scope of this paper at this moment.

References

[1] M.Z. Brown, D. Burschka, and G.D. Hager. Advances in computational stereo. *IEEE Trans. on Pattern Analys. and Machine Intell.*, 25(8):993–1008, 2003.

[2] U.R. Dhond. Structure from stereo - a review. *IEEE Transc. on Systems, Man and Cybernetics*, 19(6):1489–1510, 1989.

[3] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Computer Vision*, 47(1/2/3):7–42, 2002.

[4] F. Tiesheng, N. Bing, W. Qingsong, W. Tao, and Q. Dapeng. Novel stereo matching method on multi-scale harris corner points. *IEEE Proc. of the Int. Symposium on Comp. Science and Computat. Techn.*, pages 167–170, 2008.

[5] D. Zhao, D. Liu, and Y. Yang. A new stereo matching method based on sub-pixel corner detection. In *IEEE Proc. of the Int. Conf. on Comp. Science and Soft. Eng.*, pages 899–902, 2008.

[6] X. Su and T.M. Khoshgoftaar. A progressive edge-based stereo correspodence method. In *Proc. of the Int. Symposioum on Visual Comput. (ISVC)*, volume 1, pages 248–257, 2007.

[7] K. Kyung-Yoo, K. Sung-Yeol, and H. Yo-Sung. Stereo matching using global and local segmentation. In *Proc. of the Int. Technical Conf. on Circuts/Systems, Comp. and Communicat. (ITC-CSCC'07)*, pages 1225–1226, 2007.

[8] H. Bay, V. Ferrari, and L. Van Gool. Wide-baseline stereo matching with line segments. In *Proc. of the IEEE Comp. Society Conf. on Comp. Vision and Pattern Recogn. (CVPR'05)*, 2005.

[9] A. Luo, W. Tao, and H. Burkhardt. A new multilevel line-base stereo vision algorithm based on fuzzy techniques. In *Proc. of the Int. Conf. on. Pattern Recogn. (ICPR'96)*, volume 1, page 383, 1996.

[10] X. Shenghua, L. Jiping, P. Jianping, and W. Yong. Straight line extraction for stereoimage feature matching. In *IEEE Proc. of the Congress on Image and Signal Processing*, 2008.

[11] B. Tang, D. Ait-Boudaoud, J. Matuszewski, and L. Shark. An efficient feature based matching for stereo images. In *Proc. of the Geometric Modeling and Imaging - New Trends (GMAI'06)*. IEEE, 2006.

[12] W. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transc. on Pattern Analysis and Machine Intell.*, 7(1):17–34, 1985.

[13] D.J. Braunegg. Stereo feature matching in disparity space. *IEEE*, pages 796–803, 1990.

[14] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proc. of the Royal Society of London*, B(204):301–328, 1979.

[15] O.D. Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, Cambridge, 1993.

[16] E. Trucco E. and A. Verri. *Introductory techniques for 3-D computer vision*. Prentice Hall, New Jersey, 1998.

[17] W.K. Pratt. *Digital image processing*. John Wiley & Sons, Inc., third edition edition, 2001.

[18] R. Nevatia. *Machine perception*. Prentice-Hall, Inc., 1982.

[19] G. Nieradka. Method for edge detection in images using fuzzy relation. In *Proceedings of the SPIE*, volume 6937, pages 69373E–69373E–8, 2008.

[20] G.J. Klir and B. Yuan. *Fuzzy sets and fuzzy logic. Theory and applications.* Prentice Hall PTR, 1995. Upper Saddle River, New Jersey.

[21] D.-A. Chiang and N.P. Lin. Correlation of fuzzy sets. *Fuzzy sets and systems*, 102(2):221–226, 1999.

[22] Middlebury Stereo Vision Page. Available from: `http://vision.middlebury.edu/stereo`. [cited: 20.01.2009].

ISBN: 978-989-95079-6-8

Non-additive robust ordinal regression with Choquet integral, bipolar and level dependent Choquet integrals

Silvia Angilella Salvatore Greco Benedetto Matarazzo

Department of Economics and Quantitative Methods,
Faculty of Economics, University of Catania,
Corso Italia 55, I-95129 Catania, Italy.
Email: angisil@unict.it, salgreco@unict.it, matarazz@unict.it

Abstract— *Choquet integral has proved to be an effective aggregation model in multiple criteria decision analysis when interactions between criteria have to be taken into consideration. Recently, some generalizations of Choquet integral have been proposed to take into account more complex forms of interaction. This is the case of the bipolar Choquet integral and of the level dependent Choquet integral. To apply Choquet integral and its generalizations in decision problems it is necessary to determine one capacity permitting to represent the preferences of the Decision Maker (DM). In general the capacities are determined on the basis of some exemplary decisions supplied by the DM. It has been observed that effectively there is not only one capacity compatible with the DM's preferences, but rather a whole set of capacities. The determination of the whole set of compatible capacities and the consequent definition of proper preference relations is the domain of the non-additive robust ordinal regression. The authors have already proposed a methodology for non-additive robust ordinal regression when dealing with classical Choquet integral in ranking or choice decision problems. In this presentation, we want to give the basis of a general methodology for non-additive robust ordinal regression for Choquet integral and its generalizations (therefore also the bipolar Choquet integral and the level dependent Choquet integral) in the whole spectrum of decision problems (i.e. not only ranking and choice, but also multicriteria classification).*

Keywords— Choquet integral; Bi-Capacity; Bipolar Choquet integral; Level dependent Choquet integral; Non-additive robust ordinal regression.

1 Introduction

In the field of Multiple Criteria Decision Aid (MCDA), the main purpose of the Multi-Attribute Utility Theory (MAUT) [1] is to represent the preferences of the Decision Maker (DM) on a set of alternatives, taking in account different and conflicting points of view, called *criteria*, to get an overall utility function of every alternative.

The principal assumption underlying MAUT is the independence of criteria, not well suited to many real decision problems in which some interactions between criteria should be considered. In this last direction, Choquet integral [2] has proved to be an effective aggregation model in multiple criteria decision analysis when interactions between criteria have to be taken into consideration. Recently, according to some studies in psychology [3], many real decision problems are often based on *affect*. It seems natural to consider the criteria evaluations on a scale going from negative (bad) to positive (good) values, with a central neutral value. Consequently, the criteria are evaluated on a *bipolar* scale, *i.e.* some complex interactions among criteria arise depending on their good or positive values. To handle bipolar scales of criteria, the notion of capacity has been extended to that of bi-capacity by Grabisch and Labreuche in [4] and independently, to the bipolar capacity by Greco, Matarazzo and Słowiński in [5]. As a result, the Choquet integral has been generalized with the bipolar Choquet integral [6].

In this work we also consider a recent generalization of the Choquet integral: the level dependent Choquet integral [7]. The level dependent Choquet integral handles either unipolar scales or bipolar scales and takes into account the fact that importance of criteria depends also on the level of their evaluation.

In this paper, we propose a general framework for *non-additive robust ordinal regression* (see [8]) permitting to determine the capacities of the Choquet integral and of its above generalizations being compatible with the DM's preferences. On the basis of these sets of compatible capacities, we can define appropriate preference relations permitting to give a recommendation in the whole spectrum of decision problems (*i.e.* not only ranking and choice, but also multicriteria classification). Moreover, we propose to help the DM by identifying one capacity being the most representative among the many compatible capacities for the decision problem at hand.

The paper is organized as follows. In Section 2, we present the basic concepts relative to the Choquet integral and its generalizations, the bi-polar Choquet integral and the level dependent Choquet integral. In Section 3, the non-additive robust ordinal regression is extended to such generalizations. Section 4 contains some conclusions.

2 Choquet integral and its generalizations

2.1 Preliminary notation

In a multiple criteria decision analysis, let $X = X_1 \times X_2 \times \cdots X_n$ with $X_1, \cdots, X_i, \cdots, X_n \subseteq \mathbb{R}^n$ be the possible values taken by n criteria describing a finite set X of m alternatives. We denote every alternative $\mathbf{x} \in X$ by the evaluation vector $\mathbf{x} = (x_1, \cdots, x_i, \cdots, x_n) \in X$ and the index set of criteria by $N = \{1, \cdots, i, \cdots n\}$.

2.2 Choquet integral

A *fuzzy measure* (called also capacity) on N is a set function

$$\mu : 2^N \rightarrow [0, 1]$$

with $\mu(\emptyset) = 0$, $\mu(N) = 1$ (*boundary conditions*) and $\forall\ A \subseteq B \subseteq N,\ \mu(A) \leq \mu(B)$ (*monotonicity condition*).

ISBN: 978-989-95079-6-8

In the framework of multicriteria decision problems, the value $\mu(A)$ on the set of criteria A can be interpreted as the importance weight given by the DM to the set of criteria A.

A fuzzy measure is *additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$, for any $A, B \subseteq N$ such that $A \cap B = \emptyset$.

In case of additive fuzzy measures, $\mu(A)$ is simply obtained by $\mu(A) = \sum_{i \in A} \mu(\{i\}), \forall\ A \subseteq N$. In the other cases, we have to define a value $\mu(A)$ for every subset A of N, obtaining 2^n coefficients values.

Given $\mathbf{x} \in X \subseteq \mathbb{R}^n_+$ and μ being a fuzzy measure on N, then the *Choquet integral* [2] is defined by:

$$\begin{array}{l} \mathcal{C}_\mu(\mathbf{x}) = \sum_{i=1}^n \left[x_{(i)} - x_{(i-1)}\right] \mu\left(A_i\right) = \\ = \sum_{i=1}^n x_{(i)} \Big(\mu\left(A_i\right) - \mu\left(A_{i+1}\right) \Big) \end{array}$$

where $_{(\cdot)}$ stands for a permutation of the indices of criteria such that:

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq ... \leq x_{(n)}, \qquad (1)$$

with $A_i = \{(i), ..., (n)\}$ where $A_{n+1} = \{\emptyset\}$ $(i = 1, .., n)$ and $x_{(0)} = 0$.

2.3 Bipolar Choquet integral

Let $\mathcal{S}(N) = \{(C, D)\ :\ C \subseteq N,\ D \subseteq N,\ C \cap D = \emptyset\}$ be the set of pairs of subsets of N.

A *bi-polar capacity*, defined in [5], is a function

$$\breve{\mu}\ :\ \mathcal{S}(N) \rightarrow [0, 1] \times [0, 1]$$

such that,

1. $\breve{\mu}(A, \emptyset) = (a, 0)$ and $\breve{\mu}(\emptyset, B) = (0, b)$, with $A, B \in \mathcal{S}(N)$ and $a, b \in [0, 1]$;
2. $\breve{\mu}(N, \emptyset) = (1, 0)$ and $\breve{\mu}(\emptyset, N) = (0, 1)$;
3. For each $(C, D), (E, F) \in \mathcal{S}(N)$, such that $C \supseteq E$ and $D \subseteq F$, we have $\breve{\mu}(C, D) = (c, d)$ and $\breve{\mu}(E, F) = (e, f)$, $c, d, e, f \in [0, 1]$, with $c \geq e$ and $d \leq f$.

The properties 1) and 2) are the *boundary conditions*, while the property 3) is the *monotonicity condition*.

Given $(C, D) \in \mathcal{S}(N)$ with $\breve{\mu}(C, D) = (c, d)$, we use the following notation, $\breve{\mu}^+(C, D) = c$ and $\breve{\mu}^-(C, D) = d$.

A *bi-capacity*, defined in [4], is a function, $\hat{\mu}\ :\ \mathcal{S}(N) \rightarrow [-1, 1]$ such that,

1. $\hat{\mu}(\emptyset, \emptyset) = 0$ and $\hat{\mu}(N, \emptyset) = 1$ and $\hat{\mu}(\emptyset, N) = -1$ *(boundary conditions)*;
2. If $C \supseteq E$ and $D \subseteq F$, then $\hat{\mu}(C, D) \geq \hat{\mu}(E, F)$ *(monotonicity conditions)*.

From each bi-polar capacity, a corresponding bi-capacity is obtained by

$$\hat{\mu}(C, D) = \breve{\mu}^+(C, D) - \breve{\mu}^-(C, D),\ \ \forall\ (C, D) \in \mathcal{S}(N).$$

Let be $_{(\cdot)}$ a permutation of the elements of N such that,

$$|x_{(1)}| \leq |x_{(2)}| \leq \ldots \leq |x_{(i)}| \leq \ldots \leq |x_{(n)}|,$$

and $|x_{(0)}| = 0$.

Let be the following two subsets of N, $A_i^+ = \{j \in N\ :\ x_j \geq |x_{(i)}|\}$ and $A_i^- = \{j \in N\ :\ x_j < 0,\ -x_j \geq |x_{(i)}|\}$.

The bi-polar Choquet integral of the positive part is defined as,

$$\mathcal{C}^+(\mathbf{x}, \breve{\mu}) = \sum_{i \in N} \Big(|x_{(i)}|\ - |x_{(i-1)}| \Big) \breve{\mu}^+(A_i^+, A_i^-).$$

Analogously, *the bi-polar Choquet integral* of the negative part is defined as,

$$\mathcal{C}^-(\mathbf{x}, \breve{\mu}) = \sum_{i \in N} \Big(|x_{(i)}| - |x_{(i-1)}| \Big) \breve{\mu}^-(A_i^+, A_i^-).$$

The bi-polar Choquet integral is defined (see [5]) as:

$$\mathcal{BC}(\mathbf{x}, \hat{\mu}) = \mathcal{C}^+(\mathbf{x}, \breve{\mu}) - \mathcal{C}^-(\mathbf{x}, \breve{\mu}).$$

$\mathcal{BC}(\mathbf{x}, \hat{\mu})$ can be also formulated in terms of bi-capacities (see [6]), as follows,

$$\mathcal{BC}(\mathbf{x}, \hat{\mu}) = \sum_{i \in N} \Big(|x_{(i)}| - |x_{(i-1)}| \Big) \hat{\mu}(A_i^+, A_i^-).$$

2.4 Some particular submodels

In the bipolar decision making setting, one of the main drawbacks is the huge number of parameters to be elicitated from the DM in order to define the bi-capacity or the bipolar capacity. In fact, since $\mathcal{S}(N)$ is isomorphic to the set of functions from $\mathbb{N}$ to $\{-1, 0, 1\}$, then $|\mathcal{S}(N)| = 3^n$. Let us remark that in this contest the non-additive ordinal regression is particularly useful, because it does not require the elicitation of all the parameters, but it is mainly based on some holistic preferences on a reference set of alternatives from which the whole set of bi-capacities or bipolar capacities compatible with the DM's preferences. Nevertheless, dealing with a model with smaller number of parameters is always useful and for this reason, in the following sections we recall some well-known bipolar submodels, that are interesting from the point of view of applications and a decomposition of the bi-polar capacities, introduced in [9], more meaningful from a DM's point of view.

2.4.1 Decomposable bi-polar measures

We define a $2-$order decomposable bi-polar measure (see [9]) such that

$$\begin{array}{rcl} \mu^+(C, D) & = & \sum_{i \in C} a^+(\{i\}, \emptyset) + \sum_{\{i,j\} \subseteq C} a^+(\{i, j\}, \emptyset) + \\ & & + \sum_{i \in C,\ j \in D} a^+(\{i\}, \{j\}) \end{array}$$

$$\begin{array}{rcl} \mu^-(C, D) & = & \sum_{j \in D} a^-(\emptyset, \{j\}) + \sum_{\{i,j\} \subseteq D} a^-(\emptyset, \{i, j\}) + \\ & & \sum_{i \in C,\ j \in D} a^-(\{i\}, \{j\}) \end{array}$$

ISBN: 978-989-95079-6-8

The above decomposition of the bi-polar capacity has a more manageable and meaningful interpretation according to the DM's preferences.

In fact, $a^{\pm}(\cdot)$ can be interpreted in the following way:

- $a^{+}(\{i\},\emptyset)$, represents the power of the criterion i by itself; this value is always positive.
- $a^{+}(\{i,j\},\emptyset)$, represents the interaction between i and j, when their values are both positive; when its value is zero there is no interaction; on the contrary, when the value is positive there is a synergy effect when putting together i and j; a negative value means that the two criteria are redundant.
- $a^{+}(\{i\},\{j\})$, represents the power of the criterion j against the criterion i, when the criterion i has a positive value and j has a negative value; this provokes always a reduction or no effect on the value of μ^{+} since this value is always non-positive.

Analogous interpretation can be applied to the value of $a^{-}(\emptyset,\{j\})$, $a^{-}(\emptyset,\{i,j\})$, and $a^{-}(\{i\},\{j\})$.

In what follows, for the sake of simplicity, we will use a_i^{+}, a_{ij}^{+}, $a_{i|j}^{+}$, instead of $a^{+}(\{i\},\emptyset)$, $a^{+}(\{i,j\},\emptyset)$, and, $a^{+}(\{i\},\{j\})$, respectively; and a_j^{-}, a_{ij}^{-}, $a_{i|j}^{-}$, instead of $a^{-}(\emptyset,\{j\})$, $a^{-}(\emptyset,\{i,j\})$, and $a^{-}(\{i\},\{j\})$, respectively.

2.5 The level dependent Choquet integral

Let us also recall a recent further generalization of the bi-capacity (see [7]): the generalized capacity. Such measures take in account the fact that importance of criteria depends also on the level of their evaluation.

In particular, we define a *generalized capacity* a function $\mu^{\mathcal{G}}: 2^N \times \mathbb{R}_+ \to [0,1]$ such that

1. for all $t \in \mathbb{R}_+$ and $A, B \subseteq N$, $\mu^{\mathcal{G}}(A,t) \leq \mu^{\mathcal{G}}(B,t)$;
2. for all $t \in \mathbb{R}_+$, $\mu^{\mathcal{G}}(\emptyset,t) = 0$ and $\mu^{\mathcal{G}}(N,t) = 1$;

We define the generalized Choquet integral of $\mathbf{x} = [x_1,\cdots,x_n] \in \mathbb{R}_+^n$, with respect to the generalized capacity $\mu^{\mathcal{G}}$ as follows:

$$\mathcal{GC}(\mathbf{x},\mu^{\mathcal{G}}) = \int_0^{+\infty} \mu^{\mathcal{G}}(A(\mathbf{x},t),t)\,dt$$

where $A(\mathbf{x},t) = \{i \in N\colon x_i \geq t\}$.

Let us remark that the generalized Choquet integral can always be written as:

$$\mathcal{GC}(\mathbf{x},\mu^{\mathcal{G}}) = \sum_{i=1}^{n} \int_{x_{(i-1)}}^{x_{(i)}} \mu^{\mathcal{G}}(A(\mathbf{x},t),t)\,dt.$$

The level dependent Choquet integral can be defined also with respect to a generalized bi-capacity.

In particular, we define a *generalized bi-capacity* a function $\mu^{\mathcal{G}}: \mathcal{S}(N) \times \mathbb{R} \to \mathbb{R}$ such that

1. for all $t \in \mathbb{R}$ and $(A,B),(C,D) \in \mathcal{S}(N)$, $A \subseteq C$, $B \supseteq D$, $\mu^{\mathcal{G}}(A,B,t) \leq \mu^{\mathcal{G}}(C,D,t)$;
2. for all $t \in \mathbb{R}$, $\mu^{\mathcal{G}}(\emptyset,N,t) = 0$ and $\mu^{\mathcal{G}}(N,\emptyset,t) = 1$;

We define the generalized bipolar Choquet integral of $\mathbf{x} = [x_1,\cdots,x_n] \in \mathbb{R}^n$, with respect to the generalized bi-capacity $\mu^{\mathcal{G}}$ as follows:

$$\mathcal{GC}(\mathbf{x},\mu^{\mathcal{G}}) = \int_0^{+\infty} \mu^{\mathcal{G}}(A^{+}(\mathbf{x},t),A^{-}(\mathbf{x}t),t)\,dt$$

where $A^{+}(\mathbf{x},t) = \{i \in N\colon x_i \geq t\}$ and $A^{-}(\mathbf{x},t) = \{i \in N\colon -x_i \geq t\}$.

2.5.1 Interval level dependent capacity

The generalized Choquet integral is quite difficult to calculate and consequently, to be applied since a capacity $\mu^{\mathcal{G}}(A,t)$ is needed for each level $t \in \mathbb{R}$. For this reason in [7], a manageable class of generalized capacities are proposed, namely the *interval level dependent capacities*. For the sake of simplicity, but without loss of the generality, in the following we consider $\mathbf{x} \in [0,1]^n$.

A generalized capacity $\mu^{\mathcal{G}}$ is defined *interval level dependent capacity* if there exist

1. $a_0, a_1, a_2, , a_{m-1}, a_m \in [0,1]$, with m a positive integer such that
$$0 = a_0 < a_1 < a_2 < \cdots < a_{m-1} < a_m = 1$$
2. m capacities $\mu_1,\cdots,\mu_m$ on N.

such that, for all $A \subseteq N$ and all $t \in [0,1]$, $\mu^{\mathcal{G}}(A,t) = \mu_j(A)$ if $t \in]a_{j-1},a_j[$, $j = 1,\cdots,m$. In this case, $\mu^{\mathcal{G}}$ is an interval level dependent capacity relative to the breakpoints $a_0, a_1, a_2, , a_{m-1}, a_m \in]0,1[$ and to the capacities $\mu_1,\cdots,\mu_m$.

Finally, we recall an useful theorem, proposed in [7], which permits to split $\mathcal{GC}(\mathbf{x},\mu^{\mathcal{G}})$ in the sum of a finite number of classical Choquet integrals, more precisely one Choquet integral for each interval $]a_{j-1},a_j[$, $j = 1,\cdots,m$.

Theorem 1 *If $\mu^{\mathcal{G}}$ is the interval level dependent capacity relative to the breakpoints $a_0, a_1, a_2, , a_{m-1}, a_m \in]0,1[$ and to the capacities $\mu_1,\cdots,\mu_m$ then for each $\mathbf{x} \in [0,1]^n$*

$$\mathcal{GC}(\mathbf{x},\mu^{\mathcal{G}}) = \sum_{j=1}^{n} \mathcal{C}_{\mu}(\mathbf{x}^j,\mu_j)$$

where $\mathbf{x}^j \in [0,1]^n$ is the vector having its elements x_i^j , $i = 1,\cdots,n$, defined as follows:

$$x_i^j = \begin{cases} 0 & \text{if} \quad x_i < a_{j-1} \\ x_i - a_{j-1} & \text{if} \quad a_{j-1} \leq x_i \leq a_j \\ a_j - a_{j-1} & \text{if} \quad x_i \geq a_j. \end{cases}$$

3 Non-additive robust ordinal regression

3.1 Sorting problems

Within the multicriteria aggregation-disaggregation framework, *ordinal regression* aims at inducing the parameters of a decision model, for example those of a utility function, which have to represent some holistic preference comparisons of the DM. Usually, among the many utility functions representing the DM's preference information, only one utility function is

ISBN: 978-989-95079-6-8

selected (for example, in the UTA method [10]). In this context we also remember some UTA like-methods within the Choquet integral framework, proposed in [11]. Since such a choice is arbitrary to some extent, recently *additive robust ordinal regression* has been proposed with the purpose of taking into account all the sets of parameters *compatible* with the DM's preference information (for more details, see the multicriteria methodologies UTA$^{\text{GMS}}$ and GRIP proposed, respectively, in [12] and [13]). Let us remark that the principles of the *additive robust ordinal regression* have been applied also in sorting problems (see the UTADIS$^{\text{GMS}}$ method [14]).

Until now, robust ordinal regression has been implemented to additive utility functions under the assumption of criteria independence. In [15], the authors have proposed a *non-additive robust ordinal regression* on a set of alternatives X, whose utility is evaluated in terms of the Choquet integral which permits to represent the interaction among criteria, modeled by the fuzzy measures, parameterizing their approach.

In [15], besides holistic pairwise preference comparisons of alternatives from a subset of reference alternatives $X' \subseteq X$, the DM is also requested to express the intensity of preference on pairs of alternatives from X' and to supply pairwise comparisons on the importance of criteria, and the sign and intensity of interaction among pairs of criteria.

In the following, we recall the binary preference relations on the set of reference alternatives defined in [15].

Let us suppose that the preference of the DM is given by a partial pre-order $\succsim$ on $X' \subseteq X$.

The preference relation $\succsim$ can be decomposed into its symmetric part $\sim$ and into its asymmetric part $\succ$, whose semantics are, respectively:

$$\mathbf{x} \sim \mathbf{y} \;\Leftrightarrow\; \mathbf{x} \text{ is indifferent to } \mathbf{y},$$
$$\mathbf{x} \succ \mathbf{y} \;\Leftrightarrow\; \mathbf{x} \text{ is preferred to } \mathbf{y}, \text{ with } \mathbf{x}, \mathbf{y} \in X'.$$

The relation on the intensity of preference on pairs alternatives is represented by a partial pre-order $\succsim^*$ on $X' \times X'$, whose semantics is: for $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t} \in X'$

$$(\mathbf{x}, \mathbf{y}) \succsim^* (\mathbf{z}, \mathbf{t}) \;\Leftrightarrow\; \mathbf{x} \text{ is preferred to } \mathbf{y}$$
$$\text{at least as much as } \mathbf{z} \text{ is preferred to } \mathbf{t}.$$

The following system of linear constraints synthesizes the DM's preference information expressed in the approach proposed in [15].

$$\begin{cases} \mathbf{x} \succsim \mathbf{y} \;\Leftrightarrow\; \mathcal{C}_\mu(\mathbf{x}) \geq \mathcal{C}_\mu(\mathbf{y}), \text{ with } \mathbf{x}, \mathbf{y} \in X', \\ (\mathbf{x}, \mathbf{y}) \succsim^* (\mathbf{z}, \mathbf{t}) \;\Leftrightarrow\; \mathcal{C}_\mu(\mathbf{x}) - \mathcal{C}_\mu(\mathbf{y}) \geq \mathcal{C}_\mu(\mathbf{z}) - \mathcal{C}(\mathbf{t}), \\ \text{with } \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t} \in X', \\ \vdots \\ \text{Constraints on the importance and interaction of criteria} \\ \vdots \\ \text{Boundary, monotonicity conditions} \end{cases}$$

The output of the approach defines a set of fuzzy measures (capacities) μ defined *compatible* with the DM's preference information if the Choquet integral, calculated with respect to it, restores the DM's ranking on X', *i.e.*

$$\mathbf{x} \succsim \mathbf{y} \;\Leftrightarrow\; \mathcal{C}_\mu(\mathbf{x}) \geq \mathcal{C}_\mu(\mathbf{y}) \;\; \forall \mathbf{x}, \mathbf{y} \in X'.$$

Moreover, using linear programming, our decision model establishes two preference relations:

- for any $\mathbf{r}, \mathbf{s} \in X$, the *necessary* weak preference relation $\succsim^N$, if for all *compatible* fuzzy measures the utility of $\mathbf{r}$ is not smaller than the utility of $\mathbf{s}$, *i.e.* $\mathbf{r} \succsim^N \mathbf{s} \;\Leftrightarrow\; \mathcal{C}_\mu(\mathbf{r}) \geq \mathcal{C}_\mu(\mathbf{s})$,
- for any $\mathbf{r}, \mathbf{s} \in X$, the *possible* weak preference relation $\succsim^P$, if for at least one *compatible* fuzzy measure the utility of $\mathbf{r}$ is not smaller than the utility of $\mathbf{s}$, *i.e.* $\mathbf{r} \succsim^P \mathbf{s} \;\Leftrightarrow\; \mathcal{C}_\mu(\mathbf{r}) \geq \mathcal{C}_\mu(\mathbf{s})$.

Since as it is shown in literature, bi-capacities are the right tools to represent many decision bipolar behaviors, in this work we suggest to extend the *non-additive robust ordinal regression*, described in [15], to the bipolar decision setting and to the level dependent Choquet integral.

However, the greater flexibility of decision strategies with the criteria on a bipolar scale and level dependent Choquet integral is offset by the huge number of parameters to be elicitated by the DM and the not easy interpretation of the bi-capacities and level dependent capacity for the DM.

Let us remark that in our approach the DM is not compelled to give preference information on all criteria.

Moreover, not all criteria are compulsory on a bipolar scale; some criteria could be on the usual unipolar scale (see for a result in this topic the concept of partially symmetric bi-capacities, introduced in [16]).

The set of constraints relative to the bi-polar Choquet integral is as follows:

2) $$\begin{cases} \mathbf{x} \succsim \mathbf{y} \;\Leftrightarrow\; \mathcal{BC}(\mathbf{x}, \hat{\mu}) \geq \mathcal{BC}(\mathbf{y}, \hat{\mu}), \text{ with } \mathbf{x}, \mathbf{y} \in X', \\ (\mathbf{x}, \mathbf{y}) \succsim^* (\mathbf{z}, \mathbf{t}) \;\Leftrightarrow \\ \mathcal{BC}(\mathbf{x}, \hat{\mu}) - \mathcal{BC}(\mathbf{y}, \hat{\mu}) \geq \mathcal{BC}(\mathbf{z}, \hat{\mu}) - \mathcal{BC}(\mathbf{t}, \mu), \\ \text{with } \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t} \in X', \\ \vdots \\ \text{Constraints on the bipolar measures,} \\ \vdots \\ \text{Boundary, monotonicity conditions.} \end{cases}$$

The first set constraints are interpreted in the same way of the first set of constraints of system **1)**, with the only difference that every alternative is evaluated by the bi-polar Choquet integral.

Concerning the constraints on the bipolar measures, we suggest to adopt the decomposition explained in subsection 2.4.1. Such decomposition is more easy for the DM to be interpreted and to be elicitated.

The set of constraints relative to the level dependent Choquet integral is as follows:

3) $$\begin{cases} \mathbf{x} \succsim \mathbf{y} \;\Leftrightarrow \\ \mathcal{GC}(\mathbf{x}, \mu^{\mathcal{G}}) \geq \mathcal{GC}(\mathbf{y}, \mu^{\mathcal{G}}), \text{ with } \mathbf{x}, \mathbf{y} \in X', \\ (\mathbf{x}, \mathbf{y}) \succsim^* (\mathbf{z}, \mathbf{t}) \;\Leftrightarrow \\ \mathcal{GC}(\mathbf{x}, \mu^{\mathcal{G}}) - \mathcal{GC}(\mathbf{y}, \mu^{\mathcal{G}}) \geq \mathcal{GC}(\mathbf{z}, \mu^{\mathcal{G}}) - \mathcal{GC}(\mathbf{t}, \mu^{\mathcal{G}}), \\ \text{with } \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t} \in X', \\ \vdots \\ \text{Constraints on the level dependent capacity.} \end{cases}$$

ISBN: 978-989-95079-6-8

In this case, we adopt the interval level dependent capacity defined in Section 2.5.1, since it is more easy for the DM to express some ordinal constraints on such capacities. For example, if $X \subseteq [0,1]^n$ and we split the interval $]0,1[$ in two equal subintervals $I_1 =]0, \frac{1}{2}[$, $I_2 =]\frac{1}{2}, 1[$, DM's preference statement could be that criterion i is more important if the level of evaluation of every alternative is in I_2 than if it is in I_1.

We denote the system of constraints (types **1)**, **2)** and **3)**) on the reference alternatives X' by $E_\varepsilon(X')$.

Since linear programming is not able to handle strict inequalities in $E_\varepsilon(X')$, we put the constraints in the form of weak inequalities, by adding a small arbitrary positive value ε (see [17] for a result on this topic).

$$\textbf{4)} \begin{cases} U(\mathbf{x}) \geq U(\mathbf{y}) + \varepsilon, \text{ with } \mathbf{x}, \mathbf{y} \in X', \\ U(\mathbf{x}) - U(\mathbf{y}) \geq U(\mathbf{z}) - U(\mathbf{t}) + \varepsilon, \\ \text{with } \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t} \in X', \\ \vdots \\ \text{Constraints on the bipolar measures,} \\ \vdots \\ \text{Boundary, monotonicity conditions,} \end{cases}$$

where U stands for the aggregation operator than can be the bipolar Choquet integral or the level dependent Choquet integral. A similar approach can be used with the bipolar level dependent Choquet integral.

Once the decision making setting on the set of reference alternatives X' is chosen by the DM (types **1)**, **2)** and **3)**), two different optimization problems arise to establish a *necessary* and *possible* preference relation for any $\mathbf{r}, \mathbf{s} \in X$:

$$\begin{array}{ll} \max & \varepsilon \\ \text{s.t.} & E_\varepsilon^{X'} \quad \text{plus the constraint } U(\mathbf{s}) \geq U(\mathbf{r}) + \varepsilon \end{array} \tag{2}$$

and

$$\begin{array}{ll} \max & \varepsilon \\ \text{s.t.} & E_\varepsilon^{X'} \quad \text{plus the constraint } U(\mathbf{r}) \geq U(\mathbf{s}). \end{array} \tag{3}$$

If the problem (2) finds a solution with $\varepsilon \leq 0$, then $U(\mathbf{r}) \geq U(\mathbf{s})$ for all compatible sets of fuzzy measures, that implies $\mathbf{r} \succsim^N \mathbf{s}$ with $\mathbf{r}, \mathbf{s} \in X$.

On the contrary, if a positive ε solves the linear program indicated in (3), then there exists at least one *compatible* fuzzy measure such that $U(\mathbf{r}) \geq U(\mathbf{s})$, that implies $\mathbf{r} \succsim^P \mathbf{s}$ with $\mathbf{r}, \mathbf{s} \in X$.

3.2 The most representative value function

To consider the whole set of fuzzy measures compatible with the preference expressed by the DM reduces arbitrariness in the decision process. However to take into account one specific fuzzy measure can help the DM in understanding the decision process. As already proposed for robust ordinal regression with additive value functions [18], the most representative fuzzy measure is that one which better represents the necessary ranking maximizing the difference of evaluations between alternatives for which there is a preference in the necessary ranking. As secondary objective, one can consider minimizing the difference of evaluations between actions for which there is not a preference in the necessary ranking.

This comprehensive "most representative" value function can be determined through the following procedure:

1. Determine the necessary and the possible rankings in the considered set of actions.
2. For all pairs of alternatives $(\mathbf{x}, \mathbf{y})$, such that $\mathbf{x}$ is necessarily preferred to $\mathbf{x}$, add the following constraints to the linear programming constraints of types **1)**, **2)** or **3)**: $U(\mathbf{x}) \geq U(\mathbf{y}) + \varepsilon$.
3. Maximize the objective function ε.
4. Add the constraint $\varepsilon = \varepsilon^*$, with $\varepsilon^* = \max \varepsilon$ in the previous point, to the linear programming constraints of robust ordinal regression of types **1)**, **2)** and **3)**.
5. For all pairs of actions $(\mathbf{x}, \mathbf{y})$, such that neither $\mathbf{x}$ is necessarily preferred to $\mathbf{y}$ nor $\mathbf{y}$ is necessarily preferred to $\mathbf{x}$, add the following constraints to the linear programming constraints of of types **1)**, **2)** and **3)** and to the constraints considered in above point 4): $U(\mathbf{x}) - U(\mathbf{y}) \leq \delta$ and $U(\mathbf{y}) - U(\mathbf{x}) \leq \delta$.
6. Minimize the objective function δ.

3.3 Sorting problems

In this section, we briefly illustrate the principles of nonadditive robust ordinal regression to sorting problems in case of nonadditive value function represented by Choquet integral or some of its generalizations.

In MCDA, the sorting problem consists in the assignment of m alternatives of a finite set X into $h = 1, \cdots, k$ homogeneous classes $C_1, \cdots C_h, \cdots C_k$, which are increasingly ordered with respect to preference, i.e. all the elements in the class C_h have a better evaluation than the elements in class C_{h-1}. Let $b_1, b_2, \cdots, b_h$ be a set of thresholds relative to the h classes. We have that $\mathbf{x}$ belongs to class h if $U(\mathbf{x}) \geq b_{h-1}$ and $U(\mathbf{x}) < b_h$, where U stands for the aggregation operator than can be the Choquet integral or one of its generalizations. In this case, the DM's preference information will consist in the assignment of some reference alternatives to some classes. Taking into account the Choquet integral, the proposed approach determines the sets of pairs $(\mu, \mathbf{b})$, where μ is a capacity and $\mathbf{b} = [b_1, b_2, \cdots, b_h]$, compatible with the DM's assignment of the reference alternatives [19]. An alternative $\mathbf{x} \in X$ is said that *possibly belongs to class* C_h if there is at least one pair $(\mu, \mathbf{b})$ for which $\mathcal{C}_\mu(\mathbf{x}) \geq b_{h-1}$ and $\mathcal{C}_\mu(\mathbf{x}) < b_h$. One can deal analogously with the above mentioned extensions of Choquet integral.

As proposed in [20] with respect to robust ordinal regression for sorting problems based on additive value function through the UTADIS$^{\text{GMS}}$ method [14], also in this case we can help the DM with the concept of the "most representative" value function.

The idea is to select among compatible fuzzy measure that one which better highlights the possible sorting which is considered the most stable part of the robust sorting obtained by UTADIS$^{\text{GMS}}$. Thus it is selected the fuzzy measure that maximizes the difference of evaluations between alternatives for

ISBN: 978-989-95079-6-8

which the intervals of possible sorting are disjoints. As secondary objective, one can consider maximize the minimal difference between values of actions **x** and **y**, such that for any compatible fuzzy measure, **x** is assigned to a class not worse than the class of **y**, and for at least one compatible value function, **x** is assigned to a class which is better than the class of **y**. In case there is still more than one such fuzzy measure, the most representative fuzzy measure minimizes the maximal difference between values of alternatives being in the same class for all compatible fuzzy measures, and between values of alternatives for which the order of classes is not univocal.

4 Conclusions

Nonadditive value functions and bipolar decision making are one of the underpinning directions of research within MCDA, as many studies have shown that in a multicriteria decision problem the criteria under consideration could be on a bipolar scale. In such context, we have proposed a multicriteria methodology taking inspiration from some recent approaches based on the principle of the *robust ordinal regression*: the UTA$^{\text{GMS}}$ [12] and GRIP [13] in the context of choice and ranking problems, the UTADIS$^{\text{GMS}}$ method relative to sorting problems [14]. On this basis, with the aim of representing interactions between criteria, we proposed the *non-additive robust ordinal regression* which consists in the extension of the idea of robust ordinal regression to non-additive decision. More precisely, the Choquet integral and its generalizations have been adopted as utility function in different decision problems such as ranking, choice and sorting (for the specific case of Choquet integral applied to ranking and choice problems see [15]). The non-additive robust ordinal regression seems very useful because it permits to take into account the whole set of capacities compatible with the DM's preferences which are expressed through very simple questions such as:

- "is the representative alternative a better than the representative alternative b"?
- "to what class the representative alternative a belongs"?

References

[1] R.L. Keeney and H. Raiffa. *Decision with Multiple Objectives - Preferences and value Tradeoffs*. Wiley, New York, 1976.

[2] G. Choquet. *Theory of capacities*. Annales de l'Institut Fourier, 1953.

[3] P. Slovic, E. Peters M. Finucane, and D. G. MacGregor. *Heuristics and biases: the psychology on intuitive judgement*, chapter The affect heuristic, pages 397–420. Cambridge University Press, 2002.

[4] M. Grabisch and C. Labreuche. Bi-capacities–I: definition, Möbius transform and interaction. *Fuzzy Sets and Systems*, 151:211–236, 2005.

[5] S. Greco, R. Słowiński, and B. Matarazzo. Bipolar Sugeno and Choquet integrals. pages 191–196, Varenna, Italy, September 2002. EUROFUSE Workshop on Informations Systems.

[6] M. Grabisch and C. Labreuche. Bi-capacities–II: the Choquet integral. *Fuzzy Sets and Systems*, 151:237–259, 2005.

[7] S. Greco, S. Giove, and B. Matarazzo. The Choquet integral with respect to a level dependent capacity. *Fuzzy Sets and Systems*, submitted, 2008.

[8] S.Greco, R. Słowiński, J. Figueira, and V. Mousseau. *New Trends in Multiple Criteria Decision Analysis*, chapter Robust ordinal regression. Springer Science + Business Media, Inc, in press.

[9] J. Figueira and S. Greco. Dealing with interactivity beween bi-polar multiple criteria preferences in outranking methods. Manuscript, 2003.

[10] E. Jacquet-Lagrèze and Y. Siskos. Assessing a set of additive utility functions for multicriteria decision-making, the uta method. *European Journal of Operational Research*, 10:151–164, 1982.

[11] S. Angilella, F. Lamantia., S. Greco, and B. Matarazzo. Assessing non-additive utility for multicriteria decision aid. *European Journal of Operational Research*, 158:734–744, 2004.

[12] S. Greco, V. Mousseau, and R. Słowiński. Ordinal regression revisited: Multiple criteria ranking with a set of additive value functions. *European Journal of Operational Research*, 191:416–436, 2008.

[13] J. Figueira, S. Greco, and R. Słowiński. Building a set of additive value functions representing a reference preorder and intensities of preference: Grip method. *European Journal of Operational Research*, 195 (2):460–486, 2008.

[14] S. Greco, V. Mousseau, and R. Słowiński. Multiple criteria sorting with a set of additive value functions. *European Journal of Operational Research*, submitted.

[15] S. Angilella, S. Greco, and B. Matarazzo. Non-additive robust ordinal regression: a multiple criteria decision model based on the Choquet integral. *European Journal of Operational Research*, doi:10.1016/j.ejor.2009.02.023, 2009.

[16] M. Grabisch and C. Labreuche. Partially unipolar bi-capacities in MCDM. In: SCIS-ISIS, Yokohama, Japan, 21-24 September 2004.

[17] J.L. Marichal and M. Roubens. Determination of weights of interacting criteria from a reference set. *European Journal of Operational Research*, 124:641–650, 2000.

[18] J. Figueira, S. Greco, and R. Słowiński. Identifying the "most representative " value function among all compatible value functions in the grip method. 68$^{\text{th}}$ Meeting of the European Working Group on Multiple Criteria Decision Aiding, Chania, October 2-3, 2008.

[19] S. Angilella, S. Greco, and B. Matarazzo. Sorting decisions with interacting criteria. A.M.A.S.E.S. conference, Trento, Italy, September 1-4, 2008.

[20] S. Greco, M. Kadziński, and R. Słowiński. The most representative value functionin robust multiple criteria sorting. 69$^{\text{th}}$ Meeting of the European Working Group on Multiple Criteria Decision Aiding, Bruxelles, April 2-3, 2009 and submitted to Computers & Operations Research.

ISBN: 978-989-95079-6-8

Towards a unified logical framework of fuzzy implications to compare fuzzy sets

Hoel Le Capitaine Carl Frélicot

MIA Laboratory, University of La Rochelle
La Rochelle, France
Email: {hoel.le_capitaine,carl.frelicot}@univ-lr.fr

Abstract*— *In fuzzy set theory, comparison of fuzzy sets plays an important role. Among the several ways to compare fuzzy sets, we address the logical theoretic approach using fuzzy implications. We propose a general framework allowing to generate many measures of comparison: inclusion, similarity and distance, and study their properties. Since the literature on the use fuzzy implications for defining such measures is abundant, we also attempt to relate this work to existing research.

***Keywords*—** aggregation operators, distance, fuzzy implication, inclusion, similarity

1 Introduction

In fuzzy set theory, inclusion (or subsethood) and similarity measures between fuzzy sets are basic concepts. Among the different approaches to compare fuzzy sets, one finds three categories. The first one is based on measuring a distance between two real functions and do not refer to a specific interpretation, e.g. the Minkowski r-distances. The second category involves set-theoretic operations for fuzzy sets (fuzzy intersection, union, cardinality) [1, 2]. This paper addresses the third category which relies on logical theory, mainly approached by using fuzzy implications. Following the early paper by Bandler and Kohout [3] introducing the use of implications to measure inclusion of fuzzy sets, many authors have proposed new measures satisfying some specific axioms [4, 5, 6].

This paper will be organized as follows. We first recall some basic definitions and theorems on aggregation operators, triangular norms and fuzzy implications that will be used in the sequel in section 2. Next, in section 3 we propose a general framework of comparison measures of fuzzy sets (i.e. inclusion, similarity and distance) based on logical considerations, i.e. using fuzzy implications. A general description of the overall aggregation method is also provided, characterizing each measure belonging to this category. In section 4, we present the existing approaches that could be linked to this general formulation. We finally conclude and mention some application domains we have in mind in section 5.

2 Preliminaries and notations

Aggregating numbers plays an important role in decision-making systems. Values to be aggregated are generally defined on a finite real interval or on ordinal scales. In this paper, we assume with no loss of generality that they come from the unit interval. If not, a simple transformation can be found to make this assumption true. Given n numbers, an aggregation operator is a mapping $\mathcal{A} : [0,1]^n \to [0,1]$, satisfying boundary conditions

$$\mathcal{A}(0,\cdots,0) = 0 \quad \text{and} \quad \mathcal{A}(1,\cdots,1) = 1. \tag{1}$$

and monotonocity

$$\begin{aligned} \forall n \in \mathbb{N}, x_1 \leq y_1, \cdots, x_n \leq y_n \quad &\text{implies} \\ \mathcal{A}(x_1,\cdots,x_n) \leq \mathcal{A}(y_1,\cdots,y_n). \end{aligned} \tag{2}$$

Adding properties like idempotency, continuity, associativity lead to others definitions. In the literature, one finds many aggregation operators, e.g.: triangular norms (t-norms for short), OWA (*Ordered Weighted Averaging*) operators, γ-operators, or fuzzy integrals. They belong to several categories, depending on the way the values are aggregated: conjunctives, disjunctives, compensatory, and weighted operators. Briefly, an aggregation operator is said to be *conjunctive* if its output value is lower than the minimum of the input values, *compensatory* if the output value lies between the minimum and the maximum input value, and *disjunctive* if its output is greater than the maximum value of the inputs, refer to [7] for a large survey on aggregation operators.

Theorem 1. *If $\mathcal{A}$ is a strictly monotonic compensatory aggregation operator, then*

$$\mathcal{A}(x_1,\ldots,x_n) = 1 \Leftrightarrow x_1 = \cdots = x_n = 1 \tag{3}$$

and

$$\mathcal{A}(x_1,\ldots,x_n) = 0 \Leftrightarrow x_1 = \cdots = x_n = 0 \tag{4}$$

Proof.

(3) ($\Leftarrow$) obvious with Eq. (1).
($\Rightarrow$) by contraposition. We show that $\exists i$ such that $x_i < 1$ implies $\mathcal{A}(x_1,\cdots,x_n) < 1$. If $x_i < 1$ for some i, then, since $x_j \leq 1$ for all j, by strict monotonicity of $\mathcal{A}$, we get $\mathcal{A}(x_1,\cdots,x_n) < \mathcal{A}(1,\cdots,1) = 1$, which ends the contraposition.

(4) ($\Leftarrow$) obvious with Eq. (1).
($\Rightarrow$) by contraposition. We show that $\exists i$ such that $x_i > 0$ implies $\mathcal{A}(x_1,\cdots,x_n) > 0$. If $x_i > 0$ for some i, then, since $x_j \geq 0$ for all j, by strict monotonicity of $\mathcal{A}$, we get $\mathcal{A}(x_1,\cdots,x_n) > \mathcal{A}(0,\cdots,0) = 0$, which ends the contraposition.

□

Example 1. *The family of mean operators are strictly monotonic compensatory aggregation operators.*

ISBN: 978-989-95079-6-8

Remark 1. *From Theorem 1, we immediately see that a disjunctive aggregation operator cannot be strictly monotonic.* $\forall j\ x_j \leq 1$, *and* $\exists i,\ x_i < 1$, $\exists k,\ x_k = 1$. *If* $\mathcal{A}$ *is strictly monotonic,* $\mathcal{A}(x_1, \cdots, x_n) < \mathcal{A}(1, \cdots, 1) = 1$. *Since* $\mathcal{A}$ *is disjunctive, i.e.* $\mathcal{A}(x_1, \cdots, x_n) \geq \max(x_1, \cdots, x_n)$, *we have* $1 = \max(x_1, \cdots, x_n) \leq \mathcal{A}(x_1, \cdots, x_n) < 1$, *which is a contradiction.*

A t-norm is an increasing, associative and commutative mapping $\top : [0,1]^2 \to [0,1]$ satisfying the boundary condition $\top(x,1) = x$ for all $x \in [0,1]$. The most popular continuous t-norms are:

- Standard: $\top_S(x,y) = \min(x,y)$
- Algebraic (Product): $\top_A(x,y) = x\,y$
- Łukasiewicz: $\top_L(x,y) = \max(x+y-1, 0)$

Alternatively, the dual operators with respect to a strict negation are called triangular conorms (t-conorms for short). A t-conorm is an increasing, associative and commutative mapping $\perp : [0,1]^2 \to [0,1]$ satisfying the boundary condition $\perp(x,0) = x$ for all $x \in [0,1]$. The most popular continuous t-conorms are:

- Standard: $\perp_S(x,y) = \max(x,y)$
- Algebraic (Product): $\perp_A(x,y) = x + y - x\,y$
- Łukasiewicz: $\perp_L(x,y) = \min(x+y, 1)$

Various parametrical families have been introduced, e.g. the Hamacher family defined by: given $\gamma \in [0, +\infty[$,

- $x \top_H y = \frac{x\,y}{\gamma + (1-\gamma)(x+y-x\,y)}$
- $x \perp_H y = \frac{x+y-x\,y-(1-\gamma)\,x\,y}{1-(1-\gamma)\,x\,y}$.

Such families result in an infinite number of t-norm couples, including non parametrical ones e.g. the Algebraic couple for $\gamma = 1$.

A general problem in fuzzy logic is to handle conditional statements *if* x, *then* y where x and y are fuzzy predicates. A widely used method consists in managing them by mappings $I : [0,1] \times [0,1] \to [0,1]$ such that the truth value of I depends on the initial propositions x and y. We generally speak about an *implication function* if I is non-increasing in the first variable, non-decreasing in the second variable and $I(0,0) = I(1,1) = 1$, and $I(1,0) = 0$, see [8] for a survey on fuzzy implication functions and [9] for a large overview on the use of parametrical implications in fuzzy inference systems. The four most usual implications are:

1. *S-implications*, defined by:
$$I_\perp(x,y) = \perp(x^c, y) \tag{5}$$
where $(.)^c$ is the usual complementation $x^c = 1 - x$. This implication is an immediate generalization of the usual boolean implication $x \to y \equiv x^c \vee y$.

2. *R-(for Residual) implications*, defined by:
$$I_\top(x,y) = \sup_t \{t \in [0,1] | \top(x,t) \leq y\}. \tag{6}$$
Note that if $\top$ is a left-continuous t-norm, the *supremum* operation can be substituted by the *maximum* one.

3. *QL-(for Quantum mechanic Logic) implications*, defined by:
$$I_{QL}(x,y) = \perp(x^c, \top(x,y)). \tag{7}$$

4. *D-implications*, defined by:
$$I_D(x,y) = \perp(\top(x^c, y^c), y) \tag{8}$$
which are the contraposition of QL-implications with respect to the complementation.

In the sequel, we will denote by:

- $X = \{x_1, \cdots, x_n\}$ the (supposed finite) universe of discourse,
- $\mathcal{C}(X)$ and $\mathcal{F}(X)$ the sets of all crisps and fuzzy sets in X, respectively,
- $f_A(x)$, $\forall x \in X$, the membership function of a fuzzy set A over X,
- $[\frac{1}{2}]$ the constant fuzzy set defined by $[\frac{1}{2}](x) = \frac{1}{2}$ for any $x \in X$.

3 General framework

There are several ways to compare fuzzy values or fuzzy quantities. The first one is based on a broad class of measures of equality based on a distance measure which is specified for membership functions of fuzzy sets. This approach takes its roots from studies on how to measure the distance between two real functions and do not refer to any specific interpretation. The general form of a Minkowski r-metric is usually taken and leads to well known distance functions (Hamming, Euclidean, Chebyshev):

$$d_r(A,B) = \left(\sum_{x \in X} |f_A(x) - f_B(x)|^r \right)^{1/r} \tag{9}$$

A second way to compare fuzzy values comes from some basic set-theoretic considerations where union, intersection and complementation are defined for fuzzy sets. Cardinal and possibility based measures belong to this category. In this paper, we will focus our attention on the third way: the logical framework. We present the design a unified logical framework to compare fuzzy sets, which includes usual measures depending the aggregation operators we use. The use of parametrical t-norms and t-conorms to define the implications will enable to obtain parametrical measures of comparison, where the parameters can be set according to user needs.

3.1 Inclusion measures

An inclusion measure is a relation between two fuzzy sets which indicates to which extent one fuzzy set is contained in another one. Since its original definition by $A \subset B$ iff $f_A(x) \leq f_B(x)$, for all x in X, which was crisp assessment, Bandler and Kohout enlarged this point of view by giving a degree of subsetness [3], more in the spirit of the fuzzy theory. Inclusion measures are generally defined by a set of axioms [4] and by using fuzzy implication operators [3, 10].

Definition 1. *A mapping* $\mathcal{I} : \mathcal{F}(X) \times \mathcal{F}(X) \to [0,1]$ *is called an inclusion (or subsethood) measure if it satisfies*

ISBN: 978-989-95079-6-8

(P1) $\mathcal{I}(A,B) = 1$ *iff* $A \subseteq B$, $\quad \forall A, B \in \mathcal{F}(X)$.

(P2) *if* $[\frac{1}{2}] \subseteq A$, *then* $\mathcal{I}(A, A^c) = 0$ *iff* $A = X$.

(P3) $\forall A, B, C \in \mathcal{F}(X)$, *if* $A \subseteq B \subseteq C$, *then* $\mathcal{I}(C,A) \leq \mathcal{I}(B,A)$, *and if* $A \subseteq B$, $\mathcal{I}(C,A) \leq \mathcal{I}(C,B)$.

Note that if a measure $\mathcal{I}$ satisfies only the two last properties, it is called a weak inclusion measure.

Theorem 2. *Let $I_\top$ be a residual implication function. Given X and arbitrary $A, B \in \mathcal{F}(X)$, let*

$$\mathcal{I}(A,B) = \mathop{\mathcal{A}}_{i=1}^{n} I_\top \left(f_A(x_i), f_B(x_i)\right) \tag{10}$$

for all x_i in X, where $\mathcal{A}$ is a conjunctive or a strictly monotonic compensatory aggregation operator satisfying Eqs.(1) and (2). Then $\mathcal{I}$ is an inclusion measure.

Proof.

(P1) ($\Leftarrow$) if $A \subseteq B$, then $f_A(x_i) \leq f_B(x_i)$ for all $x_i \in X$. Since if $x \leq y$ then $I_\top(x,y) = 1$, we have $I_\top(f_A(x_i), f_B(x_i)) = 1$ for all $x_i \in X$, which gives $\mathcal{I}(A,B) = 1$ by boundary condition on $\mathcal{A}$.
($\Rightarrow$) two cases are considered
- if $\mathcal{A}$ is disjunctive, then we obviously have $f_A(x_i) \leq f_B(x_i)$ for all $x_i \in X$, since the minimum value of all implications on X is 1
- if $\mathcal{A}$ is a strictly monotonic compensatory aggregation operator, then by using Theorem 1, we have $I_\top\left(f_A(x_i), f_B(x_i)\right) = 1$ for all $x_i \in X$, giving $A \subseteq B$ by the confinement principle: $x \leq y$ *iff* $I(x,y) = 1$.

(P2) ($\Leftarrow$) if $A = X$, then $I_\top(f_X(x_i), \emptyset(x_i)) = 0$ for all $x_i \in X$. By boundary condition on $\mathcal{A}$, it follows that $\mathcal{I}(A, A^c) = 0$.
($\Rightarrow$) if $\mathcal{I}(A, A^c) = 0$, then $I_\top(f_A(x_i), 1 - f_A(x_i)) = 0$ for all $x_i \in X$ by Theorem 1. Assuming that $A \neq X$, then $\exists i$, x_i such that
$\frac{1}{2} \leq f_A(x_i) < 1$, i.e. $0 < 1 - f_A(x_i) \leq \frac{1}{2}$.
By non-increasingness in the first variable of $I_\top$, we have $I_\top(f_A(x_i), 1 - f_A(x_i)) \geq I_\top(1, 1 - f_A(x_i)) = 1 - f_A(x_i) \neq 0$, since $I_\top$ satisfies the border principle: $I(1,x) = x, \forall x \in [0,1]$. This is a contradiction with $I_\top(f_A(x_i), 1 - f_A(x_i)) = 0$, so that $A = X$.

(P3) if $A \subseteq B \subseteq C$, $f_A(x_i) \leq f_B(x_i) \leq f_C(x_i)$, for all $x_i \in X$. By non-increasingness in the first variable and non-decreasingness in the second variable, it follows that $I_\top(f_C(x_i), f_A(x_i) \leq I_\top(f_B(x_i), f_A(x_i))$ and $I_\top(f_C(x_i), f_A(x_i)) \leq I_\top(f_C(x_i), f_B(x_i))$ for all i. By monotonicity of $\mathcal{A}$, we have $\mathcal{I}(C,A) \leq \mathcal{I}(B,A)$ and $\mathcal{I}(C,A) \leq \mathcal{I}(C,B)$, which concludes the proof.

□

A necessary condition to define a strong inclusion measure $\mathcal{I}$ is that the implication I holds the confinement principle and the border principle. It is easy to show that the four usual implications $I_\top$, $I_\perp$, I_{QL} and I_D satisfy the latter, but only $I_\top$ satisfies the former. Therefore S, QL and $D-$implications define weak inclusion measures while $R-$implications define strong inclusion measures provided the operator $\mathcal{A}$ is not disjunctive, see Remark 1.

3.2 Similarity and distance measures

Definition 2. *A mapping $\mathcal{S} : \mathcal{F}(X) \times \mathcal{F}(X) \to [0,1]$ is called a similarity measure if it satisfies*

(P1) $\mathcal{S}(A,B) = \mathcal{S}(B,A), \quad \forall A, B \in \mathcal{F}(X)$.

(P2) $\mathcal{S}(A,A) = 1$, $\forall A \in \mathcal{F}(X)$.

(P3) $\mathcal{S}(D, D^c) = 0, \quad \forall D \in \mathcal{C}(X)$.

(P4) $\forall A, B, C \in \mathcal{F}(X)$, *if* $A \subseteq B \subseteq C$, *then* $\mathcal{S}(A,C) \leq \mathcal{S}(A,B) \wedge \mathcal{S}(B,C)$
or, equivalently
$\forall A, B, C, D \in \mathcal{F}(X)$, *if* $A \subseteq B \subseteq C \subseteq D$, *then* $\mathcal{S}(A,D) \leq \mathcal{S}(B,C)$

However, the symmetry property (P1) is still subject to experimental investigation: if $\mathcal{S}(x,y)$ is the answer to the question *how is x similar to y?*, then, when making comparisons, subjects focus more on the feature x than on y. This corresponds to the notion of saliency [1] of x and y: if y is more salient than x, then x is more similar to y than vice versa, which is experimentally confirmed. Accordingly to [11], for a certain element of the universe of discourse X, a degree of equality of two fuzzy elements x and y can be defined by using implications as follows:

$$(x \equiv y) = \frac{1}{2}\big((x \to y) \wedge (y \to x) + (x^c \to y^c) \wedge (y^c \to x^c)\big) \tag{11}$$

where $\wedge$ stands for minimum, $\to$ is an R-implication.
Then applying Eq. (6) with $x \geq y$, gives

$$(x \equiv y) = \frac{1}{2}\big((x \to y) + (y^c \to x^c)\big) \tag{12}$$

since 1 is the neutral element of t-norms.
Furthermore, due to the opposition law: $I(x,y) = I(y^c, x^c)$, we obtain

$$(x \equiv y) = (x \to y) \tag{13}$$

A convenient way to define a similarity measure is to quantify to which extent two fuzzy membership degrees are similar, so that it is closely related to the problem of matching fuzzy quantities, or fuzzy sets similarity. So we propose to use fuzzy implication functions as similarity measures by the following theorem:

Theorem 3. *Let $I_\top$ be a residual implication function. For arbitrary $A, B \in \mathcal{F}(X)$, let*

$$\mathcal{S}(A,B) = \mathop{\mathcal{A}}_{i=1}^{n} I_\top \left(f_{(1)}(x_i), f_{(2)}(x_i)\right) \tag{14}$$

for all x_i in X, where $f_{(.)}$ is a permutation of f_A and f_B such that $f_{(1)}(x_i) = (f_A \cup f_B)(x_i)$, $f_{(2)}(x_i) = (f_A \cap f_B)(x_i)$, and $\mathcal{A}$ an aggregation operator satisfying Eqs. (1) and (2). Then $\mathcal{S}$ is a similarity measure.

Proof.

(P1) we have $I_\top(x,x) = 1$, for any $x \in [0,1]$. By boundary conditions on $\mathcal{A}$, see Eq. (1), we obtain $\mathcal{S}(A,A) = 1$.

ISBN: 978-989-95079-6-8

(P2) by commutativity of union and intersection of fuzzy sets, we have

$$\begin{aligned}\mathcal{S}(A,B) &= \mathop{\mathcal{A}}_{i=1}^{n} I_\top\left((f_A \cup f_B)(x_i), (f_A \cap f_B)(x_i)\right)\\ &= \mathop{\mathcal{A}}_{i=1}^{n} I_\top\left((f_B \cup f_A)(x_i), (f_B \cap f_A)(x_i)\right)\\ &= \mathcal{S}(B,A)\end{aligned}$$

(P3) by definition, $I(1,0)=0$. By boundary conditions on $\mathcal{A}$, see Eq. (1), we obtain $\mathcal{S}(D,D^c)=0$.

(P4) since $A \subseteq B \subseteq C \subseteq D$, we have for all $x_i \in X$

$$f_D(x_i) \geq f_C(x_i) \tag{15}$$
$$f_B(x_i) \geq f_A(x_i) \tag{16}$$

By non-increasingness in the first variable and non-decreasingness in the second variable of $I_\top$, we obtain for all $x_i \in X$

$I_\top(f_D(x_i), f_A(x_i)) \leq I_\top(f_C(x_i), f_A(x_i))$ by Eq. (15)

$I_\top(f_C(x_i), f_A(x_i)) \leq I_\top(f_C(x_i), f_B(x_i))$ by Eq. (16)

Last, monotonicity of $\mathcal{A}$, see Eq. (2), gives $\mathcal{S}(A,D) \leq \mathcal{S}(B,C)$ which concludes the proof.

□

In contrast to inclusion measures, no restriction is imposed to the $\mathcal{A}$ operator, it can be freely chosen provided Eqs.(1) and (2) are satisfied. On another hand, here again, the confinement principle is necessary to obtain property (P1) of similarity measures. Since $\mathcal{S}(A,B)$ is reflexive and symmetrical, i.e. $\mathcal{S}(A,A)=1$ and $\mathcal{S}(A,B)=\mathcal{S}(B,A)$ hold for any $A,B \in \mathcal{F}(X)$, $\mathcal{S}$ is a proximity relation on $\mathcal{F}(X)$.

Definition 3. *A mapping $\mathcal{D}: \mathcal{F}(X)\times\mathcal{F}(X) \to [0,1]$ is called a distance measure if $\mathcal{D}$ has the following properties*

(P1) $\mathcal{D}(A,B) = \mathcal{D}(B,A), \quad \forall A,B \in \mathcal{F}(X)$.

(P2) $\mathcal{D}(A,A) = 0, \forall A \in \mathcal{F}(X)$.

(P3) $\mathcal{D}(D,D^c) = 1, \quad \forall D \in \mathcal{C}(X)$.

(P4) $\forall A,B,C \in \mathcal{F}(X)$, *if* $A \subseteq B \subseteq C$, *then* $\mathcal{D}(A,B) \leq \mathcal{D}(A,C)$ *and* $\mathcal{D}(B,C) \leq \mathcal{D}(A,C)$

Proposition 1. *For arbitrary $A,B \in \mathcal{F}(X)$, and $\mathcal{S}(A,B)$ a similarity measure defined by Eq. (14), then the mapping $\mathcal{D}: \mathcal{F}(X)\times\mathcal{F}(X) \to [0,1]$ defined by $\mathcal{D}(A,B) = \mathcal{S}^c(A,B)$ is a distance measure between A and B.*

3.3 Examples

Some examples of inclusion and similarity measures of the literature, and the new Hamacher inclusion and similarity measures, all obtained through the proposed general logical framework are given in Table 1. Moreover, this table shows how it is easy to check wether an inclusion measure is strong (non disjunctive $\mathcal{A}$ and $I_\top$) or weak (non disjunctive $\mathcal{A}$), see Definition 1, as well as the proper definition of a similarity measure (any $\mathcal{A}$ and $I_\top$), see Theorem 3.

For illustration purpose, Fig. 1 shows some examples of fuzzy similarity measures as well as the influence of the parameter γ for Hamacher residual implications, which are defined by:

$$I_{\top_{H_\gamma}}(x,y) = \begin{cases} 1 & \text{if } y \geq x \\ \frac{y(\gamma+x-\gamma x)}{y(\gamma+x-\gamma x)+x-y} & \text{if } y \leq x \end{cases} \tag{17}$$

where $\gamma \in [0,+\infty[$. The different plots show the similarity value of a given fuzzy set $A = \{0.4/x_1, 0.3/x_2\}$ to all the possible two-dimensional fuzzy sets B for various $I_\top$. As one could expect, the closer to x_1 or x_2 the higher the similarity. One can also note that different $I_\top$ lead to different shapes for isosurfaces of the similarity.

4 Related works

Let us relate, as far as possible, the proposed framework to some works issued from the abundant literature on the use of fuzzy implications for defining inclusion measures. In their seminal paper [3], Bandler and Kohout propose to use implication functions in order to quantify the inclusion of each element in another, and then aggregate these individual measures by a conjunctive operator: the minimum. Some years later, Hirota and Pedrycz propose to use implications for matching fuzzy quantities [11]. They aggregate the different implication truth values over X by a Choquet integral, computing the fuzzy measure with the help of a family of fuzzy sets taken as fuzzy prototypes. They also propose an entropy measure based on this matching measure, which gives an *impression concerning the uncertainty of matching*. In [14], Kosko criticizes the original definition of fuzzy set containment: B contains A if and only if $f_A(x) \leq f_B(x)$ for all x in X by Zadeh, pointing out that *if this inequality holds for all but just a few x, we can still consider A to be a subset of B to some degree*. So he proposes a second definition based on fuzzy intersection of two fuzzy sets which can be seen as the conditional probability $P(B|A)$ under certain circumstances. Furthermore, he defines the fuzziness of a fuzzy set A as the inclusion measure of $A \cup A^c$ in $A \cap A^c$, which satisfies the axioms of fuzzy entropy of De Luca and Termini [15]. Inclusion and similarity measures from a general set-theoretic point of view, coming from the proposition of Tversky [1], are described by Bouchon-Meunier *et al.* in [2]. The authors introduce a general framework for similitude, satisfiability and inclusion measures, also study the aggregation of measures of comparison, but take only two examples for the aggregation operator $\mathcal{A}$: a t-norm and the OWA operator. In this paper, we propose a study of the properties of $\mathcal{A}$ in its more general meaning. The same remark applies to the work by Young [4] proposing an axiomatization of inclusion measures and their connection to implication operators since she restricts herself to both the minimum and the arithmetical mean for $\mathcal{A}$. Furthermore, this work do not give necessary conditions on implications for the definition of strong or weak inclusion measure whereas our's does (Definition 1). Wang [13] presents two similarity measures which can be viewed as particular cases of the framework we propose. He enlarges his definition to the similarity of fuzzy elements belonging to various fuzzy sets. Again, the framework we propose allows to obtain a similarity between fuzzy elements since we compute implications for each element. Starting from Kosko and Young observations, Botana [16] presents a set of new measures derived

ISBN: 978-989-95079-6-8

Table 1: Inclusion and similarity measures of the literature, and the new Hamacher inclusion and similarity measures, all obtained through the proposed general logical framework.

Inclusion Measure $\mathcal{I}$	**Aggregation Operator** $\mathcal{A}$	**Implication** I
$\mathcal{I}(A,B) = \min_{x\in X}(\min(1, 1 - f_A(x) + f_B(x)))$ as defined in [12]	min	$I_{\top_L}$
$\mathcal{I}(A,B) = \frac{1}{n}\sum_{x\in X}\min(1, 1 - f_A(x) + f_B(x))$	arithmetical mean	$I_{\top_L}$
$\mathcal{I}(A,B) = \frac{1}{n}\sum_{x\in X}\max(1 - f_A(x), f_B(x))$	arithmetical mean	$I_{\perp_S}$
$\mathcal{I}(A,B) = \frac{1}{n}\sum_{x\in X} 1 - f_A(x) + f_A(x)\, f_B(x)$ as defined in [4]	arithmetical mean	$I_{\perp_A}$
$\mathcal{I}(A,B) = \frac{1}{n}\sum_{x\in X}\frac{f_B(x)(\gamma + f_A(x) - \gamma f_A(x))}{f_B(x)(\gamma + f_A(x) - \gamma f_A(x)) + f_A(x) - f_B(x)}$	arithmetical mean	$I_{\top_{H_\gamma}}$
Similarity Measure $\mathcal{S}$	**Aggregation Operator** $\mathcal{A}$	**Implication** I
$\mathcal{S}(A,B) = \frac{1}{n}\sum_{x\in X}\frac{\min(f_A(x), f_B(x))}{\max(f_A(x), f_B(x))}$ as defined in [13]	arithmetical mean	$I_{\top_A}$
$\mathcal{S}(A,B) = \frac{1}{n}\sum_{x\in X} 1 - \lvert f_A(x) - f_B(x)\rvert$ as defined in [13]	arithmetical mean	$I_{\top_L}$
$\mathcal{S}(A,B) = \max_{x\in X}\min(f_A(x), f_B(x))$	max	$I_{\top_S}$
$\mathcal{S}(A,B) = 1 - \max_{x\in X}\lvert f_A(x) - f_B(x)\rvert$	min	$I_{\top_L}$
$\mathcal{S}(A,B) = \frac{1}{n}\sum_{x\in X}\frac{f_{(2)}(x)(\gamma + f_{(1)}(x) - \gamma f_{(1)}(x))}{f_{(2)}(x)(\gamma + f_{(1)}(x) - \gamma f_{(1)}(x)) + f_{(1)}(x) - f_{(2)}(x)}$	arithmetical mean	$I_{\top_{H_\gamma}}$

from fuzzy implications. He studies whether the introduced inclusion measures satisfy Young's axioms [4] and/or those in [12] when using Wu, Goguen, modified Goguen, Gödel and Standard strict implications. He also gives the formulation of the corresponding entropy in the sense of [15, 14]. Another approach to aggregation operators is proposed in [17] consisting in combining implication truth values through respectively disjunctive and conjunctive functions for an optimistic and pessimistic aggregation. Fan et al. discuss the links between inclusion, entropy and fuzzy implications, and propose some new axioms for these measures [5]. Burillo et al. present a family of implication operators derived from the Łukasiewicz implication in order to define a family of inclusion grade operators [6] using the minimum operator for $\mathcal{A}$. In [18], Kehagias and Konstantinidou introduce L-fuzzy valued inclusion, similarity and distance measures, i.e. mappings $\mathcal{I}, \mathcal{S}, \mathcal{D} : \mathcal{F}(X) \times \mathcal{F}(X) \to [0,1]^n$, but restrict the output to crisp values. As pointed out by the authors, the vector output can lead to a difficult interpretation of the result. Furthermore, as vectors are partially ordered they are even harder to compare, they require a new measure to compare outputs. The framework we propose can provide a vector output since an implication on each element of X is computed. By contrast to [18], it would be a *fuzzy L-fuzzy valued* measure. More recently [19, 20], a distance between fuzzy operators, fuzzy implication functions in particular, is proposed. It leads to normalized tensor-norms which allow to define a similarity of fuzzy sets, and present an heuristic to choose the most suitable implication function to a fuzzy inference system. Zhang and Zhang propose an hybrid inclusion measure in [10] and use it to define similarity and distance measures of fuzzy sets. They restrict themselves to the weighted mean for $\mathcal{A}$, so it can be viewed as a special case of the work we propose, but contrarily to Young the involved implications satisfy the confinement and the border principle. Let us finally mention the work by Fono et al. [21] where as many difference operations as many implications are used to define measures of comparison of fuzzy sets in the set-theoretic framework, but we remind that such measures are out of the scope of this paper.

5 Conclusion and perspectives

In this paper, we propose a unified logical framework to compare fuzzy sets as well as fuzzy elements. Within this framework, new measures of inclusion, similarity and distance can be easily derived. These measures depend on any fuzzy implication I, provided it satisfies the necessary conditions we give (the border and the confinement principles) and any aggregation operator $\mathcal{A}$ (for similarity and distance) or any conjunctive and strictly monotonic compensatory aggregation operator $\mathcal{A}$ (for inclusion). Choosing specific $(\mathcal{A}, I)$ enables to retrieve most of the measures of the literature.
We hope that results of this work would be of great help to set comparison functions in many fields: fuzzy mathematical morphology [6], cluster validity [5], as well as other domains e.g. image retrieval or feature selection.

References

[1] A. Tversky. Features of similarity. *Psychological review*, 84(4):327–352, 1977.

[2] B. Bouchon-Meunier, M. Rifqi, and S. Bothorel. Towards general measures of comparison of objects. *Fuzzy Sets and Systems*, 84(2):143–153, 1996.

[3] W. Bandler and L. Kohout. Fuzzy power sets and fuzzy implication operators. *Fuzzy Sets and Systems*, 4:13–30, 1980.

[4] V. R. Young. Fuzzy subsethood. *Fuzzy Sets and Systems*, 77(3):371–384, 1996.

ISBN: 978-989-95079-6-8

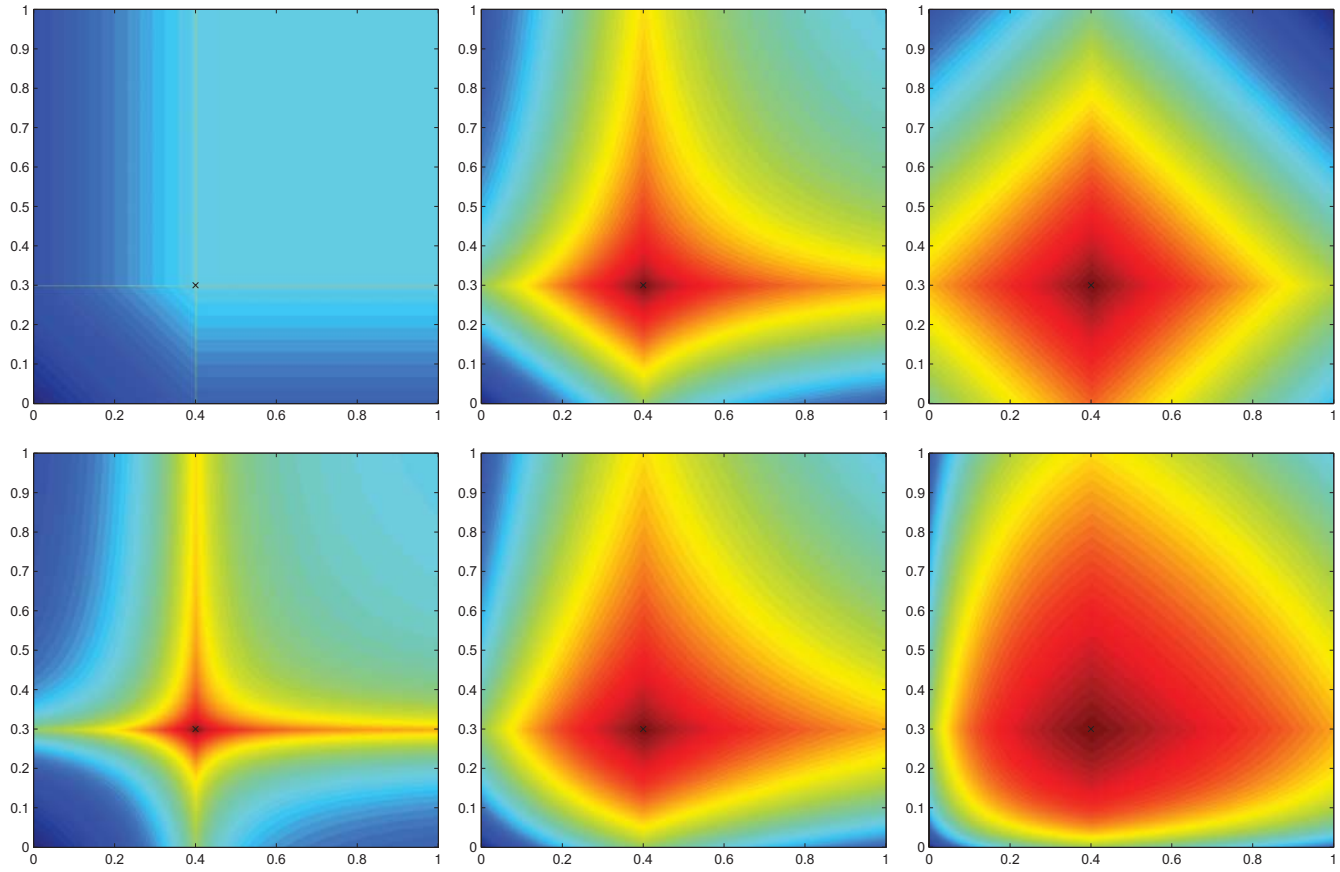


Figure 1: Examples of similarity measures of $A = \{0.4/x_1, 0.3/x_2\}$ and all fuzzy sets $\mathcal{F}(X)$, $n = 2$ where high and low values correspond to red and blue colors respectively. First row: $\mathcal{A}$ is the arithmetical mean, and we use $I_{\top_S}$, $I_{\top_A}$ and $I_{\top_L}$, respectively from left to right. Second row: $\mathcal{A}$ is the arithmetical mean and we used the Hamacher implication $I_{\top_{H_\gamma}}$, where $\gamma = 0, 2, 5$ from left to right.

[5] J. Fan, W. Xie, and J. Pei. Subsethood measure: new definitions. *Fuzzy Sets and Systems*, 106(2):201–209, 1999.

[6] P. Burillo, N. Frago, and R. Fuentes. Inclusion grade and fuzzy implication operators. *Fuzzy Sets and Systems*, 114(3):143–153, 2000.

[7] T. Calvo, A. Kolesarova, M. Komornikova, and R. Mesiar. *Aggregation Operators: New Trends and Applications*, volume 97 of *Studies in Fuzziness and Soft Computing*, chapter Aggregation Operators: Properties, Classes and Construction Methods, pages 1–104. Physica Verlag, 2002.

[8] M. Mas, M. Monserrat, J. Torrens, and E. Trillas. A survey on fuzzy implication functions. *IEEE Transactions on Fuzzy Systems*, 15(6):1107–1121, 2007.

[9] T. Whalen. Parameterized r-implications. *Fuzzy Sets and Systems*, 134(2):231–281, 2003.

[10] H-Y. Zhang and W-X. Zhang. Hybrid monotonic inclusion measure and its use in measuring similarity and distance between fuzzy sets. *Fuzzy Sets and Systems*, 160(1):107–118, 2009.

[11] K. Hirota and W. Pedrycz. Matching fuzzy quantities. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(6):1580–1586, 1991.

[12] D. Sinha and E. R. Dougherty. Fuzzification of set inclusion : theory and applications. *Fuzzy Sets and Systems*, 55(1):15–42, 1993.

[13] W-J. Wang. New similarity measures on fuzzy sets and on elements. *Fuzzy Sets and Systems*, 85:305–309, 1997.

[14] B. Kosko. *Neural networks and fuzzy systems*. Prentice Hall, Englewood Cliffs, NJ, 1992.

[15] A De Luca and S Termini. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information Control*, 20:301–312, 1972.

[16] F. Botana. Deriving fuzzy subsethood measures from violations of the implication between elements. In *11th Int. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, volume LNCS 1415, pages 234–243. Springer, 1998.

[17] P. Fonck, J. Fodor, and M. Roubens. An application of aggregation procedures to the definition of measures of similarity between fuzzy sets. *Fuzzy Sets and Systems*, 97(1):67–74, 1998.

[18] A. Kehagias and M. Konstantinidou. L-fuzzy valued inclusion measure, l-fuzzy similarity and l-fuzzy distance. *Fuzzy Sets and Systems*, 136(3):313–332, 2003.

[19] B.K. Papadopoulos, G. Trasanides, and A.G. Hatzimichailidis. Optimization method for the selection of the appropriate fuzzy implication. *Journal of Optimization Theory and Applications*, 134(1):135–141, 2007.

[20] V. Balopoulos, A. Hatzimichailidis, and B. Papadopoulos. Distance and similarity measures for fuzzy operators. *Information Sciences*, 177:2336–2348, 2007.

[21] L. Fono, H. Gwet, and B. Bouchon-Meunier. Fuzzy implication operators for difference operations for fuzzy sets and cardinality-based measures of comparison. *European Journal of Operational Research*, 183(1):314–326, 2007.

ISBN: 978-989-95079-6-8

Preservation of properties of interval-valued fuzzy relations

B. Pękala
Institute of Mathematics, University of Rzeszów
Rejtana 16a, 35-310 Rzeszów, Poland
Email: bpekala@univ.rzeszow.pl

Abstract— *The goal of this paper is to consider properties of the composition of interval-valued fuzzy relations which were introduced by L.A. Zadeh in 1975. Fuzzy set theory turned out to be a useful tool to describe situations in which the data are imprecise or vague. Interval-valued fuzzy set theory is a generalization of fuzzy set theory which was introduced also by Zadeh in 1965. This paper generalizes some properties of interval matrices considered by Pękala (2007) on these of interval-valued fuzzy relations.*

Keywords— Fuzzy relations, interval-valued fuzzy relations, properties of interval-valued fuzzy relations

1 Introduction

The idea of a fuzzy relation was defined in [28]. An extension of fuzzy set theory is interval-valued fuzzy set theory. Any interval-valued fuzzy set is defined by an interval-valued membership function: a mapping from the given universe to the set of all closed subintervals of [0,1] (it means that information is incomplete). In this paper we study properties of the composition of interval-valued fuzzy relations. Consideration of diverse properties of the composition is interesting not only from a theoretical point of view but also for the applications, since the composition of interval-valued fuzzy relations has proved to be useful in several fields, see for example, [20] (performance evaluation), [27] (genetic algorithm), [19] (approximate reasoning) or in other (see [1, 16]). In Section 2, we recall elementary properties of the composition of interval-valued fuzzy relations. Next, we consider preservation of properties of interval-valued fuzzy relations by the composition and lattice operations.

We give the following definition. Let X, Y, Z be crisp finite non-empty sets.

Definition 1 (cf. [29, 26]). Let $Int([0,1])$ be the set of all closed subintervals of $[0,1]$. An interval-valued fuzzy relation R in a universe X, Y is a mapping $R : X \times Y \to Int([0,1])$ such that $R(x,y) = [\underline{R}(x,y), \overline{R}(x,y)] \in Int([0,1])$, for all pairs $(x,y) \in (X \times Y)$. The class of interval-valued fuzzy relations in a universe $X \times Y$ will be denoted by $IVFR(X \times Y)$ or $IVFR(X)$ for $X = Y$.

Interval-valued fuzzy relations reflect the idea that membership grades are often not precise and the intervals represent such uncertainty.

The boundary elements in $IVFR(X \times Y)$ are $\mathbf{1} = [1,1]$ and $\mathbf{0} = [0,0]$. The relation R^{-1} between Y and X is defined by $R^{-1}(y,x) = R(x,y)$ for all $(x,y) \in (X \times Y)$ which we will call the inverse relation of R.

Let us look at immediate properties for interval-valued fuzzy relations: Let $P, R \in IVFR(X \times Y)$.

Then for every $(x,y) \in (X \times Y)$ we can define

$$P(x,y) \leq R(x,y) \Leftrightarrow \underline{P}(x,y) \leq \underline{R}(x,y), \overline{P}(x,y) \leq \overline{R}(x,y),$$

$$(P \vee R)(x,y) = [\max(\underline{P}(x,y), \underline{R}(x,y)), \max(\overline{P}(x,y), \overline{R}(x,y))],$$

$$(P \wedge R)(x,y) = [\min(\underline{P}(x,y), \underline{R}(x,y)), \min(\overline{P}(x,y), \overline{R}(x,y))],$$

where operations $\vee$ and $\wedge$ are the supremum and the infimum in $IVFR(X \times Y)$, respectively. For arbitrary set $T \neq \emptyset$ similarly we use

$$(\bigvee_{t \in T} R_t)(x,y) = [\bigvee_{t \in T} \underline{R_t}(x,y), \bigvee_{t \in T} \overline{R_t}(x,y)],$$

$$(\bigwedge_{t \in T} R_t)(x,y) = [\bigwedge_{t \in T} \underline{R_t}(x,y), \bigwedge_{t \in T} \overline{R_t}(x,y)].$$

We know that $([0,1], \max, \min)$ is a complete, distributive lattice, and therefore $(IVFR(X \times Y), \vee, \wedge)$ is also a complete, distributive lattice. So it is a particular case of lattices considered by Goguen in [18].

Interval-valued fuzzy relations (sets) are equivalent to some other extensions of fuzzy relations (sets) (see [12]). Among others, interval-valued fuzzy relations are isomorphic to Atanassov's fuzzy relations. This fact was noticed by several authors [5, 11, 12]. An Atanassov's fuzzy relation is a pair of fuzzy relations, namely a membership and a nonmembership functions, which represent positive and negative aspects of the given information. This objects introduced by Atanassov and originally called intuitionistic fuzzy relations were recently suggested to be called Atanassov's intuitionistic fuzzy relations or just bipolar fuzzy relations [15]. An Atanassov's fuzzy set theory is also widely applied in solving real-life problems. An example of such application is the optimization in Atanassov's intuitionistic fuzzy environment (an extension of fuzzy optimization and an application of bipolar fuzzy sets) where by applying this concept it is possible to reformulate the optimization problem by using degrees of rejection of constraints and values of the objective which are non-admissible. This concept allows one to define a degree of rejection which cannot be simply a complement of the degree of acceptance [2]. The idea of a positive and negative information was confirmed by psychological investigations [9]. Moreover, multiattribute decision making using Atanassov's intuitionistic fuzzy sets is possible (see [22, 23]).

Definition 2 (cf. [3]). Let $X \neq \emptyset$, $\mathfrak{R},\ \mathfrak{R}^d : X \times Y \to [0,1]$ be fuzzy relations fulfilling the condition

$$\mathfrak{R}(x,y) + \mathfrak{R}^d(x,y) \leqslant 1, \quad (x,y) \in (X \times Y).$$

A pair $\rho = (\mathfrak{R}, \mathfrak{R}^d)$ is called an Atanassov's intuitionistic fuzzy relation. The family of all Atanassov's intuitionistic fuzzy relations described in a given sets X, Y is denoted by $AIFR(X \times Y)$.

ISBN: 978-989-95079-6-8

Basic operations on Atanassov's intuitionistic fuzzy relations $\rho = (\mathfrak{R}, \mathfrak{R}^d)$, $\sigma = (\mathfrak{S}, \mathfrak{S}^d)$ are defined in the following way

$$\rho \cup \sigma = (\max(\mathfrak{R}, \mathfrak{S}), \min(\mathfrak{R}^d, \mathfrak{S}^d)),$$

$$\rho \cap \sigma = (\min(\mathfrak{R}, \mathfrak{S}), \max(\mathfrak{R}^d, \mathfrak{S}^d)),$$

$$\rho \leqslant \sigma \ \Leftrightarrow \ (\mathfrak{R} \leq \mathfrak{S}, \ \mathfrak{S}^d \leq \mathfrak{R}^d).$$

The isomorphism which proves the equivalence between Atanassov's intuitionistic fuzzy relations and interval-valued fuzzy relations is the following

Theorem 1 (cf. [11]). *The mapping $\psi : IVFR(X \times Y) \to AIFR(X \times Y)$, such that $R \to \rho$ is an isomorphism between the lattices $(IVFR(X \times Y), \vee, \wedge)$ and $(AIFR(X \times Y), \cup, \cap)$, where $R(x,y) = [\underline{R}(x,y), \overline{R}(x,y)]$, $R \in IVFR(X \times Y)$ and $\rho(x,y) = (\underline{R}(x,y), 1 - \overline{R}(x,y))$, $(x,y) \in (X \times Y)$.*

2 Composition of interval-valued fuzzy relations

Now, we consider the composition of interval-valued fuzzy relations.

Definition 3 (cf. [8]). Let $* : [0,1]^2 \to [0,1]$,
$P \in IVFR(X \times Y)$, $R \in IVFR(Y \times Z)$.
By the $\sup -*$ composition of relations P and R we call the relation $P \circ R \in IVFR(X \times Z)$,

$$(P \circ R)(x,z) = [(\underline{P} \circ \underline{R})(x,z), (\overline{P} \circ \overline{R})(x,z)], \quad (1)$$

where

$$(\underline{P} \circ \underline{R})(x,z) = \bigvee_{y \in Y} (\underline{P}(x,y) * \underline{R}(y,z)),$$

$$(\overline{P} \circ \overline{R})(x,z) = \bigvee_{y \in Y} (\overline{P}(x,y) * \overline{R}(y,z))$$

and $(\underline{P} \circ \underline{R})(x,z) \leq (\overline{P} \circ \overline{R})(x,z)$.

Other types of compositions may be also considered. For example, the composition of interval-valued fuzzy relations with supremum and $*$ being a triangular norm or a triangular conorm is examined in [8].

Definition 4 (cf. [24]). A triangular norm T (conorm S) is an increasing, commutative, associative operation
$T(S) : [0,1]^2 \to [0,1]$ with a neutral element 1 (0).

In [8] it was proved, for finite sets X, Y, Z, that compositions mentioned there are associative if and only if the first operation in composition is $\vee$ or $\wedge$ and the second is an arbitrary triangular norm or conorm. This is why the choice of the first operation in the composition (1) is reasonable.

For our further considerations we need the following properties

Definition 5 (cf. [18]). Let $* : [0,1]^2 \to [0,1]$. Operation $*$ is infinitely sup-distributive, if

$$\bigvee_{t \in T} (x_t * y) = (\bigvee_{t \in T} x_t) * y, \quad \bigvee_{t \in T} (y * x_t) = y * (\bigvee_{t \in T} x_t).$$

Definition 6 (cf. [6]). Operation $* : [0,1]^2 \to [0,1]$ is isotone if it fulfils the condition:

$$\underset{x,y,z \in [0,1]}{\forall} x \leq y \Rightarrow x * z \leq y * z, z * x \leq z * y.$$

By generalization of the results of the papers [14, 25] and using [13] we obtain

Lemma 1. *If $*$ is isotonic, then $\sup -*$ composition is also isotonic.*

Proof. Let $*$ be left side isotonic.
For $P, R \in IVFR(X \times Y), Q \in IVFR(Y \times Z)$

$$P \leq R \Leftrightarrow [\underline{P}, \overline{P}] \leq [\underline{R}, \overline{R}] \Leftrightarrow$$

$$\underset{x,y \in X \times Y}{\forall} \underline{P}(x,y) \leq \underline{R}(x,y), \overline{P}(x,y) \leq \overline{R}(x,y)$$

by isotonicity of the $*$ and supremum we have for
$x \in X, y \in Y, z \in Z$

$$\bigvee_{y \in X} (\underline{P}(x,y) * \underline{Q}(y,z)) \leq \bigvee_{y \in X} (\underline{R}(x,y) * \underline{Q}(y,z)),$$

$$\bigvee_{y \in X} (\overline{P}(x,y) * \overline{Q}(y,z) \leq \bigvee_{y \in X} (\overline{R}(x,y) * \overline{Q}(y,z)) \Leftrightarrow$$

$$\underline{P} \circ \underline{Q} \leq \underline{R} \circ \underline{Q}, \overline{P} \circ \overline{Q} \leq \overline{R} \circ \overline{Q} \Leftrightarrow P \circ Q \leq R \circ Q.$$

The right side isotonicity of $\circ$ may be proved similarly. □

Lemma 2. *Let the operation $*$ have a zero element z. If $*$ has a neutral element e, then $\sup -*$ composition has a neutral element $\mathbf{S}_e = [S_e, S_e]$,*

$$S_e(x,y) = \begin{cases} e & if \ x = y \\ z & if \ x \neq y \end{cases} \quad for \ x \in X, y \in Y.$$

Proof. Let $*$ operation have a zero element z and a neutral element e.

$$(\mathbf{S}_e \circ R)(x,y) =$$

$$[\bigvee_{z \in X} (S_e(x,z) * \underline{R}(z,y)), \bigvee_{z \in X} (S_e(x,z) * \overline{R}(z,y))] =$$

$$[(e * \underline{R}(x,y)) \vee \bigvee_{x \neq z} (z * \underline{R}(z,y)), (e * \overline{R}(x,y)) \vee \bigvee_{x \neq z} (z * \overline{R}(z,y))] =$$

$$[e * \underline{R}(x,y), e * \overline{R}(x,y)] = [\underline{R}(x,y), \overline{R}(x,y)] = R(x,y).$$

The proof for $R \circ \mathbf{S}_e = R$ is similar. □

If $*$ is an isotonic operation, then we may prove

$$\bigvee_{t \in T} (P_t \circ R) \leq (\bigvee_{t \in T} P_t) \circ R, \ \bigwedge_{t \in T} (P_t \circ R) \geq (\bigwedge_{t \in T} P_t) \circ R. \quad (2)$$

Now we examine the problem of sup-distributivity. Authors [8] in Theorem 6 present condition for sup-distributivity by $*$ equal to any t-norm or t-conorm and by finite non-empty X, Y. By generalization we obtain

Lemma 3. *If an operation $*$ is infinitely sup-distributive, then $\sup -*$ composition is also infinitely sup-distributive i.e., for $P_t \in IVFR(X \times Y), R \in IVFR(Y \times Z)$*

$$\bigvee_{t \in T} (P_t \circ R) = (\bigvee_{t \in T} P_t) \circ R. \quad (3)$$

ISBN: 978-989-95079-6-8

Proof. Let $*$ be sup-distributive, $x \in X, y \in Y, z \in Z$. Then

$$((\bigvee_{t\in T} P_t) \circ R)(x,z) =$$

$$[\bigvee_{y\in Y}((\bigvee_{t\in T}\underline{P_t})(x,y)*\underline{R}(y,z)), \bigvee_{y\in Y}((\bigvee_{t\in T}\overline{P_t})(x,y)*\overline{R}(y,z))] =$$

$$[\bigvee_{y\in Y}(\bigvee_{t\in T}(\underline{P_t}(x,y)*\underline{R}(y,z))), \bigvee_{y\in Y}(\bigvee_{t\in T}(\overline{P_t}(x,y)*\overline{R}(y,z)))] =$$

$$[\bigvee_{t\in T}(\bigvee_{y\in Y}(\underline{P_t}(x,y)*\underline{R}(y,z))), \bigvee_{t\in T}(\bigvee_{y\in Y}(\overline{P_t}(x,y)*\overline{R}(y,z)))] =$$

$$[\bigvee_{t\in T}(\underline{P_t}\circ\underline{R})(x,z), \bigvee_{t\in T}(\overline{P_t}\circ\overline{R})(x,z)] = \bigvee_{t\in T}(P_t\circ R)(x,z).$$

□

Here we discuss the most important property of binary operations, i.e. the associativity. The associativity of $*$ is not sufficient for the associativity of the $\sup -*$ composition. Some results of this problem we see in ([8], Theorem 9) for t-norms or t-conorms on finite sets. As a result the question about the associativity of the composition of interval-valued fuzzy relations we get for arbitrary sets X, Y, Z, U the following lemmas.

Lemma 4. *If an operation $*$ is associative and infinitely sup-distributive, then* $\sup -*$ *composition is associative.*

Proof. Let operation $*$ be associative and infinitely sup-distributive, $P \in IVFR(X\times Y), R \in IVFR(Y\times Z)$, $Q \in IVFR(Z\times U)$ and $x \in X, y \in Y, z \in Z, u \in U$.

$$((P\circ R)\circ Q)(x,u) = [((\underline{P}\circ\underline{R})\circ\underline{Q})(x,u), ((\overline{P}\circ\overline{R})\circ\overline{Q})(x,u)] =$$

$$[\bigvee_z(\underline{P}\circ\underline{R})(x,z)*\underline{Q}(z,u), \bigvee_z(\overline{P}\circ\overline{R})(x,z)*\overline{Q}(z,u)] =$$

$$[\bigvee_z(\bigvee_y(\underline{P}(x,y)*\underline{R}(y,z))*\underline{Q}(z,u)),$$

$$\bigvee_z(\bigvee_y(\overline{P}(x,y)*\overline{R}(y,z))*\overline{Q}(z,u))] =$$

$$[\bigvee_z\bigvee_y(\underline{P}(x,y)*(\underline{R}(y,z)*\underline{Q}(z,u))),$$

$$\bigvee_z\bigvee_y(\overline{P}(x,y)*(\overline{R}(y,z)*\overline{Q}(z,u)))] =$$

$$[\bigvee_y(\underline{P}(x,y)*\bigvee_z(\underline{R}(y,z)*\underline{Q}(z,u))),$$

$$\bigvee_y(\overline{P}(x,y)*\bigvee_z(\overline{R}(y,z)*\overline{Q}(z,u)))] =$$

$$[(\underline{P}\circ(\underline{R}\circ\underline{Q}))(x,u), (\overline{P}\circ(\overline{R}\circ\overline{Q}))(x,u)] = (P\circ(R\circ Q))(x,u).$$

□

As a direct consequence of the above lemmas we observed that set of all interval-valued fuzzy relations with the composition (1) create a semigroup.

Proposition 1. *If $*$ is associative, infinitely sup-distributive operation with a zero element z=0 and a neutral element e=1, then $(IVFR(X),\circ)$ is an ordered semigroup with the identity $\boldsymbol{I} = [I, I]$.*

In the sequel we denote by D the set of all binary operations $* : [0,1]^2 \to [0,1]$ which are associative and infinitely sup-distributive (these conditions imply that $*$ is isotonic). As a result a special case of $*$ may be a left-continous triangular norm or conorm. If $* \in D$, then in a semigroup $(IVFR(X),\circ)$ we can consider the powers of its elements, i.e. relations R^n for $R \in IVFR(X)$, $n \in \mathbb{N}$. By analogy to [21] we define

Definition 7. By the powers of a relation $R \in IVFR(X)$ we call interval-valued fuzzy relations

$$R^1 = R,\ R^{m+1} = R^m \circ R, \text{ where } m = 1, 2, \ldots .$$

By the upper closure $R^\vee$ and the lower closure $R^\wedge$ of the relation R we call, respectively

$$R^\vee = \bigvee_{k=1}^{\infty} R^k,\ \ R^\wedge = \bigwedge_{k=1}^{\infty} R^k, \text{ where } R^k = [\underline{R}^k, \overline{R}^k]. \quad (4)$$

Proposition 2. *If $* \in D$ and $P, R \in IVFR(X)$, then*

$$\bigvee_{n\in\mathbb{N}}(P\vee R)^n \geq P^n \vee R^n,\ \ (P\wedge R)^n \leq P^n \wedge R^n, \quad (5)$$

$$(P\vee R)^\vee \geq P^\vee \vee R^\vee,\ \ (P\wedge R)^\vee \leq P^\vee \wedge R^\vee, \quad (6)$$

$$(P\vee R)^\wedge \geq P^\wedge \vee R^\wedge,\ \ (P\wedge R)^\wedge \leq P^\wedge \wedge R^\wedge. \quad (7)$$

Proof. Let $n \in \mathbb{N}$. From the isotonicity of $*$ and the Lemma 1 we know that the $\sup -*$ composition is also isotonic. As a result we obtain the isotonicity for powers. Then by

$$P\vee R \geq P,\ P\vee R \geq R$$

$$(\underline{P}\vee\underline{R} \geq \underline{P},\ \underline{P}\vee\underline{R} \geq \underline{R},\ \overline{P}\vee\overline{R} \geq \overline{P},\ \overline{P}\vee\overline{R} \geq \overline{R})$$

we have

$$(\underline{P}\vee\underline{R})^n \geq \underline{P}^n,\ (\underline{P}\vee\underline{R})^n \geq \underline{R}^n \text{ and}$$

$$(\overline{P}\vee\overline{R})^n \geq \overline{P}^n,\ (\overline{P}\vee\overline{R})^n \geq \overline{R}^n,$$

so we obtain

$$(P\vee R)^n \geq P^n,\ (P\vee R)^n \geq R^n \Rightarrow (P\vee R)^n \geq P^n \vee R^n.$$

Similarly, we can prove $(P\wedge R)^n \leq P^n \wedge R^n$. Moreover, supremum and infimum are isotonic, so from the condition (5) also closures have this property
$P^\vee \leq (P\vee R)^\vee$, $R^\vee \leq (P\vee R)^\vee \Rightarrow P^\vee \vee R^\vee \leq (P\vee R)^\vee$,
$(P\wedge R)^\vee \leq P^\vee$, $(P\wedge R)^\vee \leq R^\vee \Rightarrow (P\wedge R)^\vee \leq P^\vee \wedge R^\vee$.
Similarly, we may prove the inequalities in (7). □

Proposition 3. *Let $* \in D$ and $P, R \in IVFR(X)$.*
If $P\circ R = R\circ P$, then

$$\bigvee_{n\in\mathbb{N}}(P\circ R)^n = P^n \circ R^n, \quad (8)$$

$$(P\circ R)^\vee \leq P^\vee \circ R^\vee,\ \ (P\circ R)^\wedge \geq P^\wedge \circ R^\wedge. \quad (9)$$

ISBN: 978-989-95079-6-8

Proof. The given equality in (8) may be proved by the mathematical induction and associativity of $*$ and commutativity of powers, which is implied by commutativity of P and R. Then

$$(P \circ R)^{\vee} = \bigvee_{k=1}^{\infty} (P \circ R)^k = [\bigvee_{k=1}^{\infty} (\underline{P}^k \circ \underline{R}^k), \bigvee_{k=1}^{\infty} (\overline{P}^k \circ \overline{R}^k)],$$

by (3), (2) and isotonicity of the operation $\sup - *$ we have

$$[\bigvee_{k=1}^{\infty} (\underline{P}^k \circ \underline{R}^k), \bigvee_{k=1}^{\infty} (\overline{P}^k \circ \overline{R}^k)] \leq [\bigvee_{k=1}^{\infty} (\underline{P}^{\vee} \circ \underline{R}^k), \bigvee_{k=1}^{\infty} (\overline{P}^{\vee} \circ \overline{R}^k)]$$

$$\leq [\underline{P}^{\vee} \circ \bigvee_{k=1}^{\infty} \underline{R}^k, \overline{P}^{\vee} \circ \bigvee_{k=1}^{\infty} \overline{R}^k] = [\underline{P}^{\vee} \circ \underline{R}^{\vee}, \overline{P}^{\vee} \circ \overline{R}^{\vee}] = P^{\vee} \circ R^{\vee},$$

which proves the first condition in (9). The second condition in (9) one may be justified in a similar way. □

3 Properties of interval-valued fuzzy relations

Now, we will examine whether the given properties are preserved by the composition of interval-valued fuzzy relations. We see that many properties of the $*$ operation are transposed to the operation of $\sup - *$ composition but not all of them. Namely, if $* = \min$, then $\sup - \min$ composition is not commutative. Now, we examine very interesting properties, namely subidempotency and superidempotency. These properties are of the large interest for example in economy where they are applied in valuation of supply and demand. Similarly to definitions of properties of fuzzy relations in [21] we have

Definition 8. Let $R \in IVFR(X)$. The relation R is called idempotent, subidempotent (transitive) or superidempotent if $R^2 = R,\ R^2 \leq R,\ R^2 \geq R$, respectively.

For the Boolean matrices of dimension $n \times n$ the following computations hold true

Table 1: The subidempotent relations in the family of all relations.

n	all	subidempotent	%
2	16	13	81,25000
3	512	171	33,39844
4	65536	3994	6,09436
5	33554432	154303	0,45986

A similar situation takes place in any distributive and bounded lattice, it means that the percentage of the subidempotent matrices is rapidly decreasing with the growth of n. This is why the consideration and determination of the upper closure $R^{\vee}$ is very important.

Theorem 2. *Let $* \in D$ and $R \in IVFR(X)$. $R^{\vee}$ is the least subidempotent relation greater than or equal to R. Moreover, the relation R is subidempotent if and only if $R = R^{\vee}$.*

Proof. Let $R, S, Q \in IVFR(X)$. If R is subidempotent, then

$$\bigwedge_{n \in \mathbb{N}} R^n \leq R \text{ and } R \leq R^{\vee} = \bigvee_{n \in \mathbb{N}} R^n \leq R,$$

so we obtain $R^{\vee} = R$.
If $R^{\vee} = R$, then for $S = R^{\vee}$ we have

$$S^2 = \bigvee_{k=2}^{\infty} R^k \leq R^{\vee} = S$$

and $R^{\vee}$ is subidempotent.
We show that if there exists $Q = [\underline{Q}, \overline{Q}]$ such that

$$[\underline{R}, \overline{R}] \leq [\underline{Q}, \overline{Q}] \text{ and } [\underline{Q}^2, \overline{Q}^2] \leq [\underline{Q}, \overline{Q}],$$

then by isotonicity the $\sup - *$ composition (Lemma 1) we obtain

$$[\underline{R}^2, \overline{R}^2] \leq [\underline{R} \circ \underline{Q}, \overline{R} \circ \overline{Q}] \leq [\underline{Q}^2, \overline{Q}^2]$$

thus

$$[\underline{R}^k, \overline{R}^k] \leq [\underline{Q}^k, \overline{Q}^k] \leq [\underline{Q}, \overline{Q}] \text{ for } k \in \mathbb{N}.$$

So by isotonicity of the supremum $R^{\vee} \leq Q$. □

The closures and powers of interval-valued fuzzy relations also preserve some properties of such relations.

Theorem 3. *Let an operation $* \in D$ be commutative and $R \in IVFR(X)$. If R is subidempotent (superidempotent), then R^n, $R^{\wedge}$ ($R^{\vee}$) are subidempotent (superidempotent).*

Proof. Let $R^2 \leq R$. Then R^n are also subidempotent, $n \in \mathbb{N}$. By (2) and subidempotency of R we have

$$(R^{\wedge})^2 \leq (R^2)^{\wedge} \leq R^{\wedge},$$

so $R^{\wedge}$ is subidempotent. The property of superidempotency may be proved analogously. □

Theorem 4. *Let $T \neq \emptyset$ and $* \in D$ and $R_t \in IVFR(X)$, $t \in T$. If $(R_t)_{t \in T}$ is a family of subidempotent relations, then the relation $R = \bigwedge_{t \in T} R_t$ is subidempotent.*

Proof. Let $R_t^2 \leq R_t$, $t \in T$, then by (2)

$$R^2 = (\bigwedge_{t \in T} R_t)^2 = (\bigwedge_{s \in T} R_s) \circ (\bigwedge_{t \in T} R_t) \leq$$

$$\bigwedge_{t \in T} ((\bigwedge_{s \in T} R_s) \circ R_t) \leq \bigwedge_{s,t \in T} (R_s \circ R_t) \leq \bigwedge_{t \in T} R_t = R$$

i.e., the relation R is subidempotent. □

Theorem 5. *Let $* \in D$ and $P, R \in IVFR(X)$. If P, R are subidempotent, then $P \vee R$ is subidempotent if and only if $P \circ R \vee R \circ P \leq P \vee R$.*

Proof. Since $P^2 \leq P$, $R^2 \leq R$ so by (3) we obtain

$$(P \vee R)^2 = [(\underline{P} \vee \underline{R}) \circ (\underline{P} \vee \underline{R}), (\overline{P} \vee \overline{R}) \circ (\overline{P} \vee \overline{R})] =$$

$$[(\underline{P} \circ \underline{P}) \vee (\underline{R} \circ \underline{P}) \vee (\underline{P} \circ \underline{R}) \vee (\underline{R} \circ \underline{R}),$$
$$(\overline{P} \circ \overline{P}) \vee (\overline{R} \circ \overline{P}) \vee (\overline{P} \circ \overline{R}) \vee (\overline{R} \circ \overline{R})] \leq$$

$$[\underline{P} \vee (\underline{R} \circ \underline{P}) \vee (\underline{P} \circ \underline{R}) \vee \underline{R}, \overline{P} \vee (\overline{R} \circ \overline{P}) \vee (\overline{P} \circ \overline{R}) \vee \overline{R}] =$$

$$[\underline{P} \vee \underline{R} \vee (\underline{R} \circ \underline{P}) \vee (\underline{P} \circ \underline{R}), \overline{P} \vee \overline{R} \vee (\overline{R} \circ \overline{P}) \vee (\overline{P} \circ \overline{R})],$$

as a result

$$(P \vee R)^2 \leq P \vee R \Leftrightarrow P \circ R \vee R \circ P \leq P \vee R,$$

because

$$P \leq R \Leftrightarrow P \vee R = R.$$

□

ISBN: 978-989-95079-6-8

Theorem 6. *Let $* \in D$ and $P, R \in IVFR(X)$. If P, R are subidempotent (superidempotent) and $P \circ R = R \circ P$, then $P \circ R$ is subidempotent (superidempotent).*

Proof. If $\underline{P} \circ \underline{R} = \underline{R} \circ \underline{P}$ and $\overline{P} \circ \overline{R} = \overline{R} \circ \overline{P}$, $P^2 \leq P$, $R^2 \leq R$, then by the associativity and monotonicity of the $\sup -*$ composition we have

$$(P \circ R)^2 = [(\underline{P} \circ \underline{R})^2, (\overline{P} \circ \overline{R})^2] =$$

$$[\underline{P} \circ (\underline{R} \circ \underline{P}) \circ \underline{R}, \overline{P} \circ (\overline{R} \circ \overline{P}) \circ \overline{R}] =$$

$$[\underline{P} \circ \underline{P} \circ \underline{R} \circ \underline{R}, \overline{P} \circ \overline{P} \circ \overline{R} \circ \overline{R}] = [\underline{P}^2 \circ \underline{R}^2, \overline{P}^2 \circ \overline{R}^2] \leq$$

$$[\underline{P} \circ \underline{R}, \overline{P} \circ \overline{R}] = P \circ R.$$

As a result $P \circ R$ is subidempotent. The proof for superidempotency is similar. □

4 Conclusion

In this work we present only some problems connected with the preservation of interval-valued fuzzy relation properties by the $\sup -*$ composition and related to it operation. We can also consider preservation of other properties (for example, symmetry, asymmetry, antisymmetry, reflexivity, irreflexivity, connectedness) by the composition of interval-valued fuzzy relations and also by its powers and lattice operations.
We may consider dual composition to the one defined in (1). This is the $\inf -*'$ composition with the dual binary operation $*'$, where $x *' y = 1 - (1 - x) * (1 - y)$ for $x, y \in [0, 1]$. The properties of this composition may be deduced from the $\sup -*$ composition.
Moreover, we may study interval-valued t-norms and t-conorms as operations on *Int*([0,1]) which are important functions because they are useful in approximate reasoning, in medical diagnosis and information retrieval. For example, the authors of [7, 10] examine the construction of t-norms and t-conorms in the lattice $(Int[0, 1], \vee, \wedge)$ and analyze some properties of them.

References

[1] E. Agustench, H. Bustince and V. Mohedano. Method of least squares applied to the GMP with international-valued fuzzy sets. *Mathware & Soft Computing*, 6(2-3):267–276, 1999.

[2] P.P. Angelov, Optimization in an intuitionistic fuzzy environment, *Fuzzy Sets and Systems*, 86:299–306, 1997.

[3] K.T. Atanassov, Intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 20:87–96, 1986.

[4] K.T. Atanassov, *Intuitionistic fuzzy sets*, Physica-Verlag, New York, 1999.

[5] K.T. Atanassov and G. Gargov, Interval valued intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 31:343–349, 1989.

[6] G. Birkhoff, *Lattice theory*, AMS Coll. Publ. 25, Providence 1967.

[7] H. Bustince, E. Barrenechea and M. Pagola. Generation of interval-valued fuzzy and Atannasov's intuitionistic fuzzy connectives from fuzzy connectives and from K_α operators: laws for conjunction and disjunction, amplitude. *International Journal of Intelligent Systems*, 23:680-714, 2008.

[8] H. Bustince and P. Burillo. Mathematical analysis of interval-valued fuzzy relations: Application to approximate reasoning. *Fuzzy Sets and Systems*, 113:205-219, 2000.

[9] J.T. Cacioppo, W.L. Gardner and G.G. Berntson. Beyond bipolar conceptualization and measures: the case of attitudes and evaluative spaces. *Personality and Social Psychology Review*, 1:3–25, 1997.

[10] G. Deschrijver. A representation of t-norms in interval-valued L-fuzzy set theory. *Fuzzy Sets and Systems*, 159:1597-1618, 2008.

[11] G. Deschrijver and E.E. Kerre, On the relationship between some extensions of fuzzy set thory, *Fuzzy Sets and Systems*, 133(2):227–235, 2003.

[12] G. Deschrijver and E.E. Kerre, On the position of intuitionistic fuzzy set theory in the framework of theories modelling imprecision, *Information Sciences*, 177:1860–1866, 2007.

[13] J. Drewniak and K. Kula, Generalized compositions of fuzzy relations, *International Journal of Uncertainty, Fuzziness Knowledge-Based Systems*, 10:149-163, Suppl., 2002.

[14] J. Drewniak and B. Pękala. Properties of fuzzy relations powers. *Kybernetika*, 43(2):133-142, 2007.

[15] D. Dubois, S. Gottwald, P. Hajek, J. Kacprzyk and H. Prade, Terminological difficulties in fuzzy set theory - The case of "Intuitionistic Fuzzy Sets", *Fuzzy Sets and Systems*, 156:485–491, 2005.

[16] D. Dubois, H. Prade and S. Sessa. Recent literature. *Fuzzy Sets and Systems*, 159:857–865, 2008.

[17] D. Dubois and H. Prade. An introduction to bipolar representations of information and preference, *International Journal of Intelligent Systems*, 23(8):866–877, 2008.

[18] A. Goguen. L-fuzzy sets. *Journal of Mathematical Analysis and Applications*, 18:145–174, 1967.

[19] M.B. Gorzałczany. A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, 21(1):1-17, 1987.

[20] Y.-Y. Guh, M.-S. Yang, R.-W. Po and E.S. Lee. Interval-valued fuzzy relation-based clustering with its application to performance evaluetion. *Computers and Mathematics with Applications*, 57:841–849, 2009.

[21] A. Kaufmann. *Introduction to the theory of fuzzy subsets*, Acad. Press, New York, 1975.

[22] D.-F. Li. Multiattribute decision making models and method using intuitionistic fuzzy sets, *Journal of Computer and System Sciences*, 70:73–85, 2005.

[23] L. Lin, X-H. Yuan and Z-Q. Xia. Multicriteria fuzzy decision-making methods based on intuitionistic fuzzy sets, *Journal of Computer and System Sciences*, 73:84–88, 2007.

[24] E.P. Klement, R. Mesiar and E. Pap. *Triangular norms*. Kluwer Acad. Publ., Dordrecht, 2000.

[25] B. Pękala. Operations on interval matrices, in: M. Kryszkiewicz et al. (eds.), *Rough Sets and Intelligent Systems Paradigms*, Springer-Verlag, Berlin, 613–620, 2007.

[26] R. Sambuc. Fonctions ϕ-floues. Application á l'aide au diagnostic en pathologie thyroidienne. Ph.D. Thesis, Université de Marseille, France, 1975.

[27] L. Sánchez and I. Couso. Advocating the use of imprecisely observed data in genetic fuzzy systems. *IEEE Transaction on Fuzzy Systems*, 15(4):551–562, 2007.

[28] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[29] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-I, *Information Sciences*, 8:199–249, 1975.

ISBN: 978-989-95079-6-8

A fuzzy approach for the model of sliding window. An application to behaviour patterns mining

María Ros, Miguel Delgado, Amparo M. Vila[1]

1.Department of Computer Science and A.I. University of Granada, España
Email: {marosiz, mdelgado, vila}@decsai.ugr.es

Abstract— *Many traditional environment applications base their operations on using sensor. For instance, the Tagged World uses the information from sensors to identify user behaviour and to provide services. We present a method to limit an user behaviour through external user knowledge: a Temporal Window. However, an user behaviour has random imprecision by definition. Thus, we have adapted the method to manage the imprecision, Fuzzy Temporal Window, which includes Fuzzy Logic concepts. As an application, we present a method to extract behaviour patterns on time defined by a Fuzzy Temporal Window.*

Keywords— Behaviour, Fuzzy Windows, Quantified Sentences, Tagged World, Frequent Itemsets

1 Introduction

In the last years, Smart Computers have suffered a great evolution that is transforming the world. *Ubiquitous Computing*[1] uses them to make user daily life easier. Computers are included on daily life without disturbing normal user activities. Different devices are added in traditional environment from common PDA or Mobile Phone to sensors. Nowadays, the last trend in *Ubiquitous Computing* is to provide the environment with different mechanisms to get information.

Sensors are devices that are able to know what is happening in the environment. There are many kind of sensors from temperature sensor to mechanical one. We also find RFID sensors [2] that identify every object in the environment with an unique number by means of radio frequency. Placing RFID tags in an environment we get to construct an intelligent space in which everyone can enjoy services. An intelligent space obtains user position, user behaviour, and environment information around a user and so on by a sensor network [3]. Other example about these spaces can be found in [4]. We find an intelligent space that provides services in a variety of scenes, but hardly ever, without reasoning or inference. The development of these projects is the Tagged World concept.

In [5, 3], the Tagged World project is presented. This project is developed in the University of Ritsumeikan, Japan and consists on providing appropriate personalized services for each user, to make their life easier and safer by recognizing and reasoning the human behaviour. They use a wearable computer as a Pocket Assistant that compares an access log with patterns to recognize human activities. The system is based on a Bayesian network and obtains results with a probability value for every extracted behaviour.

Other alternative is proposed by Philipose and et al. [6]. They proposed a system to infer Activities of Daily Living in a Tagged World. They present a new paradigm for ADL inferencing leverage radio-frequency-identification technology, data mining and a probabilistic inference engine to recognize ADLs based on objects that people use. As sensor, they use RFID technology with other sensor streams to fill in the gaps. The system represents activities as linear sequences of activity stages, and annotates each stage with the involved objects and the probability of their involvement.

In [7][8] is proposed a system to identify correct behaviour using Data Mining Techniques. The system is divided into two main parts: inductive learning mechanism, which produces a behaviour database and a reasoning system for the recognition of sequences that uses this database. The first stage uses Frequent Itemsets, while the second one does Regular Grammar.

This paper is organized in six main sections. In Section 2 we present the formal problem to solve. In Section 3, Fuzzy Temporal Window concept is explained. Section 4 presents the method that manage problem uncertainty. In section 5, we show some empirical results, comparing crisp and fuzzy models. Finally, the conclusions and future works are reported in section 6.

2 Formal Problem

This section defines the formal problem: to obtain sequence patterns to identify user behaviour that are defined on a specific domain and context. Thus, we have to define some basic concepts which are the objectives of the system.

DEFINITION **2.1 (Action)** *An action, a, is an activity that happens using a specific object.*

However, it results very interesting to know when an action happens on time.

DEFINITION **2.2 (Action on time)** *An action on time, a, is an activity that happens over a specific object in a known time and it is denoted as a pattern $a = (h, l)$, where h is a fact and l is a temporal label that defines the time of the action.*

These definitions give us a basic element to work, but our aim is to find different actions that make a behaviour up.

DEFINITION **2.3 (Behaviour)** *Let $A = \{a_1, ..., a_n\}$ be set of possible user actions in some situations or domains. An user behaviour is a finite set of actions:*

$$\beta = \{\alpha_1, \alpha_2, , ...\alpha_{p(\beta)}\}$$

with $\alpha_j \in A \forall j$, and where α_j is performed before α_k iif $j \leq k$.

ISBN: 978-989-95079-6-8

In this definition, we do not pay attention on the time, although, in general, a behaviour happens in a specific moment on time.

DEFINITION **2.4 (Behaviour on time)** *Let $A = \{a_1, ..., a_n\}$ be set of possible user actions in some situations or domains. Let τ be the temporal line, then A user behaviour on time is a finite set of pair such as:*

$$b_I = \{(\alpha_1, t_1), ...(\alpha_{n(b)}, t_{n(b)})\}$$

where $\alpha_j \in A \forall j$, where α_j is performed before α_k iif $j \leq k$, $t_j \in [0, \tau]$ $\forall t_i < t_j$ if $i \geq j$, with $I = [t_1, t_{n(b)}]$

The System is based on a database about user actions. This database has been obtained from user observation, i.e., from actions that user has done. In rest of this document, we name this database as Observation Data Base (ODB).

The representation of ODB is a transactional database T, where every row is an observation over an user. Normally, we study his activity in a whole day; and every column is a possible action from A set, $A = \{a_1, ..., a_n\}$.

3 Fuzzy Temporal Windows. Extract behaviour patterns tool.

In previous section, we have presented behaviour sequence pattern problem. In [7][8] we study behaviours in a crisp way, however the human behaviour is not crisp. Human activity has random imprecision by definition.

3.1 Motivations

In [7][8] we present a process to obtain correct sequence patterns of actions from user behaviour. This method extracts from a particular ODB common actions. A general ODB contains actions that a user has realized for a whole day. For this reason, we do not know where a behaviour starts or ends. The proposed method needs an expert to indicate when a behaviour is performed. This proposal is not applicable in a real system, since every person has his/her specific activities and habits.

Therefore, we have to design a mechanism to identify the interval of the ODB that we have to study. In this point, we use a knowledge about the behaviour :generally, there exist some behaviours that are realized by user at the same time every day.

EXAMPLE **3.1** *Luis leaves home at 8:30 o'clock.*

We use this knowledge to situate an interval on the temporal line. The result is a subset from particular ODB to the studied behaviour.

In spite of this fact, nobody usually does an action at the exactly moment every day. They usually do actions roughly at the same moment. This raises a new problem: how we fix the interval's ranges to detect a specific activity. Ranges should consider random imprecision of the situations.

EXAMPLE **3.2** *Let us suppose that* Luis often leaves home at 8:30 o'clock*, then we could control actions that happen about 8:30 (from 8:20 to 8:40, for instance).*

We name the interval defined over temporal line as *Temporal Window*. It permits to get a subset from each ODB tuple (See figure 1). In this subset, we know that actions have not the same degree of importance, since actions are about the specific time are more important than actions far from the interval centre. So, we could assign a degree of importance at each action, in function of its situation in the interval. Then, we can define the interval as a fuzzy set and name it as *Fuzzy Temporal Window* (See figure 2).

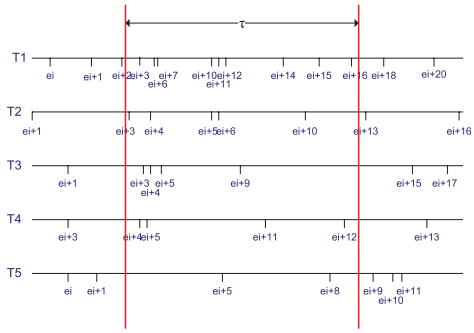


Figure 1: Temporal Window over a ODB

3.2 Model Formulation

In this section we present a formal representation to *Temporal Window* and *Fuzzy Temporal Window*.

DEFINITION **3.1 (Temporal Window, W)** *Let I be a interval from the temporal line τ, ODB the Observation Data Base and $t \in ODB$ a tuple from ODB. Let i_j be an action $i_j = (h_j, l_j)$, then a Temporal Window, W, for a specific behaviour, which happens in interval I of τ, is defined as a subset of t where*

$$\forall i_j \in W(t) \text{ then } i_j \in t \text{ and } l_j \in I$$

DEFINITION **3.2 (Fuzzy Temporal Window, FW)** *Let I be a interval from the temporal line τ, ODB the Observation Data Base and $t \in ODB$ a tuple from ODB. Let i_j be an action $i_j = (h_j, l_j)$ and a Temporal Window, W, to a specific behaviour. Let fs be a fuzzy set over τ, then it defines a* Fuzzy Temporal Window*, FW, as a Temporal Window where*

$$\forall i_j \in W(t)\ \mu_{FW}(i_j) = \mu_{fs}(l_j)$$

If we apply a Fuzzy Temporal Window over the ODB, we obtain an image where each item has its membership degree. As we have represented an ODB as a Transactional Database, we obtain a Transactional Database image too.

DEFINITION **3.3 (Fuzzy ODB applying W)** *Let ODB be an Observation DataBase and W a Fuzzy Temporal Window, it defines $\tilde{ODB}$ as a Fuzzy Observation Data Base constructs as W(ODB) such as*

$$\forall t \in ODB, W(t) \in \tilde{ODB}$$

DEFINITION **3.4 (Fuzzy Transactional Data Base)** *Let T be a Transactional Data Base and W a Fuzzy Temporal Window, it defines $\tilde{T}$ as a Fuzzy Transactional Data Base constructs as W(t) $\forall t \in T$.*

We represent the Fuzzy Transactional DataBase as a table where for every row and column we have membership degree corresponding to the Fuzzy Temporal Window.

ISBN: 978-989-95079-6-8

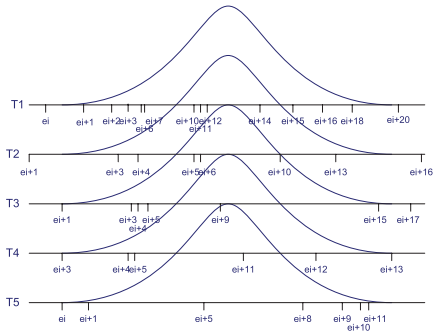


Figure 2: Fuzzy Temporal Window over a ODB

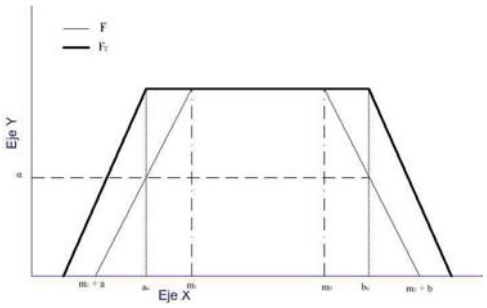


Figure 3: Transformations over a Trapezoidal Fuzzy Temporal Windows

3.3 Extending the Fuzzy Temporal Windows.

Until this point, we establish a Fuzzy Temporal Window with previous behaviour knowledge. We suppose this knowledge is completely true, although this statement always is not correct.

EXAMPLE **3.3** *Let us suppose that we want to check the behaviour* `Luis leaves home at 8:30 o'clock`. *However, we know that Luis sometimes arrives late at work. So, we can affirm that* `Luis leaves home at time almost always`.

In the example 3.3, we give a quantified adjective to actions of a behaviour to express certain. We employ a semantic approach based on evaluation of quantified sentences [9]. A quantified sentence is an expression of the form "Q of F are G", where F and G are two fuzzy subsets of a finite set X, and Q is a relative fuzzy quantifier. Some examples could be (The most times, Luis leaves home, 8:30), (Almost never, Luis leaves home, on time), (Almost all times, Luis leaves home, 8:30).

This knowledge is used to expand the Fuzzy Temporal Window size. We use the method presented in [10]. This method transforms the used function to define a specific behaviour according with the knowledge expressed by a quantified sentence. The proposed process obtains a new window F' in two steps:

1. Firstly, they truncate a fuzzy number, our Fuzzy Temporal Window, using the certainty degree α associated to the fuzzy value A. After this operation, we obtain a non normalized fuzzy set A^{α}.
2. Secondly, they normalize the fuzzy set. The authors assume that uncertainty is being translated into imprecision under the condition of the amount of information provided by the fuzzy number remains equal before and after normalization process.

The transformation function is defined as:

DEFINITION **3.5** *Let $A \in \tau$ be a fuzzy number such that $A = \{(m_1, m_2, a, b), \alpha_A\}$, where m_1, m_2, a, b are the values that defines a trapezoidal fuzzy number and α_A is the height of A. Let $\alpha \in (0, 1]$. We will denote $\triangle(\alpha_A, \alpha) = \triangle$ and define*

$$T_{\alpha}(A) = \left\{ \left(m_1, m_2, a + \frac{\triangle}{k}, b + \frac{\triangle}{k} \right), \alpha \right\} \quad (1)$$

for those α in which the transformation makes sense.

To use the definition 3.5, we need to define two basic parameters:

- k.- scale parameter
- α.- value for doing the transformation

In figure 3 we represent the followed process over a trapezoidal window.

In this point, we introduced the quantified sentences over a Fuzzy Temporal Windows, using their evaluation to know α value. When we have this value, we transform A set in its α version (A^{α}), which is the basis of the final transformation A^{T}. Example 3.4 shows the followed process. We use a trapezoidal Fuzzy Temporal Window for the sake of clarity.

EXAMPLE **3.4** *Let us suppose that we want to check the behaviour* `Luis hardly ever leaves home at 8:30 o'clock`. *Thus, our Fuzzy Temporal Windows, represented as a lr-number as $F = (m_1, m_2, a, b)$ where m_1, m_2 is the higher points of a trapezoidal fuzzy number, and a, b are the time to establish the lower points.*

$$F = (8:25, 8:35, 0:05, 0:05) \quad (2)$$

Firstly, we have to evaluate a quantified sentence to find out the α-value. There exists a lot of ways to evaluate it. Here we have selected the basic way, but not the best: Zadeh. The expression of the Zadeh Cardinal is $\alpha_A = \frac{P(A)}{|X|}$ where $P(A) = sum_{x \in X} A(x)$ and $|X|$ is the X set cardinal.

So, $\alpha_A = \frac{15}{20} = 0.75$. The evaluation of the quantified sentence is $Z_Q(A) = Q(\alpha_A) = 0.53$.

Next, we do the transformation of our windows using α value.

$$F = \{(8:25, 8:35, 0:05, 0:05), 1\} \rightarrow \quad (3)$$

$$F^{\alpha} = \left\{ \begin{pmatrix} 8:25 - 0:05(1-\alpha), 8:35 - \\ 0:05(1-\alpha), 0:05\alpha, 0:05\alpha \end{pmatrix}, \alpha \right\} \rightarrow \quad (4)$$

$$F^{T} = \left\{ \begin{pmatrix} 8:25 - 0:05(1-\alpha), 8:35 - 0:05 \\ (1-\alpha), 0:05\alpha - \frac{1-\alpha}{k\alpha}, 0:05\alpha - \frac{1-\alpha}{k\alpha} \end{pmatrix}, \alpha \right\} \quad (5)$$

Replacing α value with $Z_Q(A)$, i.e., $\alpha = 0.53$ we expand our Fuzzy Temporal Windows to

$$F = (8:23, 8:37, 0:08, 0:08) \quad (6)$$

ISBN: 978-989-95079-6-8

4 Fuzzy method to extract behaviour patterns. Obtaining sequence patterns by alpha-cuts

Once we have information, we should get common actions and valid sequences for every behaviour. However, we can not apply the method explained in [7] [8] directly, because we have to manage membership degree corresponded to the Fuzzy Temporal Window.

Here, we cannot present our Crisp model but we have to clarify some basic concepts of it that are need to can understand the process.

4.1 Previous concepts

In this section, we want to specify some important concepts: common actions and valid sequences.

1. Common actions: We define the common actions as a sequence of events that occurs more often in the observed knowledge. We can identify the common actions with the concept of Frequent Itemset: every frequent itemset corresponds to a particular common behaviour [8]. However, an itemset is a set and, so, it has not order relationship among its elements. However, a basic condition for a behaviour is the existence of this relationship. Since, we have to control which sequences of these itemsets are valid.

2. Valid/Correct sequences We define the valid sequences as the permutations of the set whose order relationship among their elements appear in the knowledge of the *ODB*. We use the permutation concept and an designed algorithm to determine if a sequence is valid or not, since not all permutation will be accepted as a pattern sequence.

4.2 Changes to manage the random imprecision

In this subsection, we present the different changes to manage the random imprecision. Thus, the objective is to develop a system that obtains stimulation from user who realizes some actions sequentially. The problem consists on extracting the sequence patterns to specific behaviour when we have a Fuzzy ODB $O\tilde{D}B$, represented as $\tilde{T}$, and a Fuzzy Temporal Window W defined from user knowledge.

Firstly, we have to transform a fuzzy problem to a crisp problem to apply the method explained in [7] [8]. We could use α-cut set [11]. By the α-cut, we would have a new crisp image of $\tilde{T}$, T^{α} where every value is in $\{0, 1\}$. Then, we apply classical method to T^{α}, obtaining frequent itemsets and sequence patterns to specific α value (I^{α} and P^{α}, respectively). After extracting all sequence patterns, we create a fuzzy set **P** by sequence patterns which have obtained to every α-value applying the Representation Theorem.

As we use frequent itemsets, we need to represent extracted frequent itemsets as an unique fuzzy set. Thus, we have to ensure *consistent restriction* between every α frequent itemset I^{α_i}. For sequence patterns representation is the same. The proof of these statements can be studied in [8].

4.3 An illustrative example

In this example we want to show the operation of the method. We start from ODB which is collected from the touched object by the user in the daily activity[1].

Let W a Fuzzy Temporal Window defines by the following fuzzy set over temporal line τ:

$$W = \left\{ \begin{array}{c} \frac{0}{8:20}, \frac{0,2}{8:21}, \frac{0,4}{8:22}, \frac{0,6}{8:23}, \frac{0,8}{8:24}, \frac{1}{8:25}, \frac{1}{8:26} \\ \frac{1}{8:27}, \frac{1}{8:28}, \frac{1}{8:29}, \frac{1}{8:30}, \frac{1}{8:31}, \frac{1}{8:32}, \frac{1}{8:33}, \frac{1}{8:34} \\ \frac{1}{8:35}, \frac{0,8}{8:36}, \frac{0,6}{8:37}, \frac{0,4}{8:38}, \frac{0,2}{8:39}, \frac{0}{8:40} \end{array} \right\} \quad (7)$$

Then, we apply W over T and obtain a Fuzzy T $\tilde{T}$ where every time of actions is replaced with the membership degree value in W. We apply the α-cut for α values α_1 = 0.4, α_2 = 0.6, α_3 = 0.8, α_4 = 1.0. We obtain for each α values the T^{α}, the result of apply the α-cut in $\tilde{T}$. After these operations, we have transformed the fuzzy problem to a crisp problem. Now, we extract frequent itemsets and sequence patterns with the model explained in [7] and [8].

We have executed the Apriori Algorithm with two support values: *minsup = 0.8 and minsup = 0.9*. The results are showed in tables 1 and table 2.

Since we have the common actions to specific behaviour, the next stage consists of obtaining the valid sequences patterns and the final representation as a fuzzy set applying the Identity Principle. The valid patterns are showed in table 3 and table 4. And the final pattern representation in equation 8 to minsup = 0.8 and equation 9 to minsup = 0.9.

Table 1: Frequent itemset to minsup = 0.8

I^{α_1}	{Shoes, Bag, Keys, ODoor, MobilePhone} {Keys, Keys2, ODoor, ODoor2, MobilePhone}
I^{α_2}	{Shoes, Bag, Keys, ODoor, MobilePhone} {Keys, Keys2, ODoor, ODoor2, MobilePhone}
I^{α_3}	{Shoes, Bag, Keys, ODoor, MobilePhone}
I^{α_4}	{Shoes, Bag, Keys, ODoor, MobilePhone}

Table 2: Frequent itemset to minsup = 0.9

I^{α_1}	{Keys, ODoor, MobilePhone}
I^{α_2}	{Keys, ODoor, MobilePhone}
I^{α_3}	{Keys, ODoor}
I^{α_4}	{ODoor}

Table 3: Valid patterns to minsup = 0.8

P^{α_1}	{Shoes, Bag, Keys, MobilePhone, ODoor}	p_1
	{Shoes, Keys, MobilePhone, Bag, ODoor}	p_2
	{Shoes, MobilePhone, Bag, Keys, ODoor}	p_3
	{Keys, MobilePhone, ODoor, Keys2, ODoor2}	p_4
	{MobilePhone, Keys, ODoor, Keys2, ODoor2}	p_5
P^{α_2}	IDEM	
P^{α_3}	{Shoes, Bag, Keys, MobilePhone, ODoor}	p_1
	{Shoes, Keys, MobilePhone, Bag, ODoor}	p_2
	{Shoes, MobilePhone, Bag, Keys, ODoor}	p_3
P^{α_4}	IDEM	

$$\tilde{P} = \left\{ \frac{p_1}{\alpha_4}, \frac{p_2}{\alpha_4}, \frac{p_3}{\alpha_4}, \frac{p_4}{\alpha_2}, \frac{p_5}{\alpha_2} \right\} = \left\{ \frac{p_1}{1}, \frac{p_2}{0.4}, \frac{p_3}{1}, \frac{p_4}{1}, \frac{p_5}{0.6} \right\} \quad (8)$$

[1] In this paper, we do not show the Data Base due to the lack of space.

ISBN: 978-989-95079-6-8

Table 4: Valid patterns to minsup = 0.9

P^{α_1}	{Keys, MobilePhone, ODoor}	p_1
	{MobilePhone, Keys, ODoor}	p_2
P^{α_2}	IDEM	
P^{α_3}	{Keys, ODoor}	p_3
P^{α_4}	{ODoor}	p_4

$$\tilde{P} = \left\{ \frac{p_1}{\alpha_2}, \frac{p_2}{\alpha_2}, \frac{p_3}{\alpha_3}, \frac{p_4}{\alpha_4} \right\} = \left\{ \frac{p_1}{0.6}, \frac{p_2}{0.6}, \frac{p_3}{0.8}, \frac{p_4}{1} \right\} \quad (9)$$

5 Performance Evaluation

In this section, our objective is to evaluate fuzzy model versus crisp model presented in [7][8]. In other hand, we present advantages of modifying the windows using quantified sentences. To do this proof, we used a database that represents daily activities of the user. To be precise, we focus study on the analysis of Leave Home behaviour, to simplify the process. We study two databases. These databases try to force the temporal model, thus they imply different temporal relations between their items. In the first one, the interval between two consecutive actions is always the same. In the second database, we break this temporal relation.

With these experiments, we want to control the fuzzy window influence over the database.

As well as, we need a goodness measure to analyze obtained outcomes. We design a measure to evaluate the method that works with a sequence of words. So, we have to design a way to transform these sequences in a numerical way. We have to design this measure because it tries to pay attention all important aspects: correction, rubbish, alpha, minsup, time, etc.

Therefore, we have to define a goodness measure that makes the result evaluation easier:

$$c = \alpha_1 correction + \alpha_2 alpha + \alpha_3 \frac{1}{time} \quad (10)$$

where,

- correction.- similarity measure of every pattern to the correct pattern. We define a function that associates a numeric value with sequences that a higher value indicates greater similarity.
- alpha.- mean of the alpha-values used for the α-cut for every pattern.
- time.- amount of millisecond that the method inverts in obtaining results.

Goodness measure is inversely proportion with time raises to the power of minus one and directly proportion with the correction and confidence measure. So, if the absolute value of the correction gets bigger, the absolute value of goodness measure gets bigger too.

For the experiments, we use a specific values for α. We have chosen this values because we believe they are rougher to the reality. We use: α_1 and α_2 equals 0.3 and α_3 equals 0.4.

5.1 Crisp model vs Fuzzy model

We were evaluating our system with three different fuzzy distribution: a trapezoidal distribution, a the Gauss bell shape

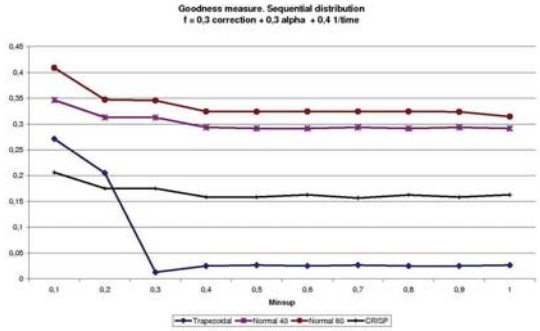


Figure 4: Crisp vs Fuzzy. Database a

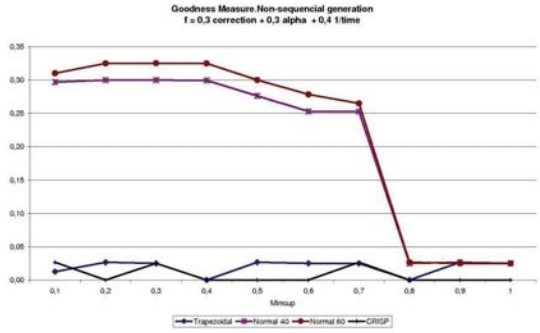


Figure 5: Crisp vs Fuzzy. Database b

with mean 40 minutes and a the Gauss bell shape with mean 60 minutes. We have used some distributions more, however we believe these distributions show the most interesting results. In figure 4 and 5, two graphics show goodness measure for every window. In addition, we show results of crisp model to compare the methods. If we observe results, we extract some conclusions which are convenient to study:

- In general, we conclude that the Fuzzy model obtains better results than the Crisp one. The Fuzzy method includes a way to provide importance to some actions that are key but are irregular. On the other hand, the Crisp method consider all actions at the same importance, and can lose this key actions. In addition, the Crisp method has to study all the data, while the Fuzzy method can reduce the size of transactions and can do the elements become frequent, minimizing the computational cost. This reduction could be very interesting when we have a forgetful user, because we would have a database with many mistakes that we would have to avoid. Moreover, this size reduction of the database makes it easier to find if the order of the sequences are valid.
- Among the Fuzzy Temporal Windows, the best results are obtained with the Gauss bell shape. It relaxes the degree of membership of their elements to the fuzzy window. When we perform the cut with a specific α, we include more items to study. Instead, in the trapezoidal shape the α-cuts are more abrupt, then more elements are lost.
- For the second database, the Crisp method provide good results: if we study all possible data, you always get

ISBN: 978-989-95079-6-8

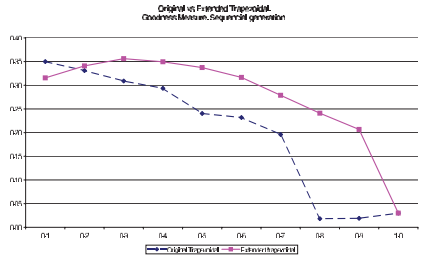


Figure 6: Original vs Extended Trapezoidal. Database a

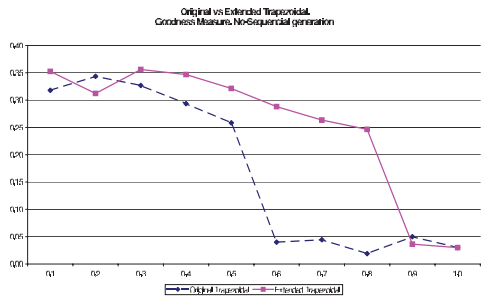


Figure 7: Original vs Extended Trapezoidal. Database b

key actions. However, with any Fuzzy Temporal Window with a Gauss bell shape we can extract almost the same results than the Crisp model minimizing the computational cost. Other alternative is to extend the variance of the fuzzy window, as we will see in the next subsection.

- Thus, the outcomes obtained with the fuzzy model are better than using crisp model.

5.2 *Analysis of process of extending the Fuzzy Temporal Window*

Once we have checked our fuzzy method, we analysed the process of extending the Fuzzy Temporal Window from a knowledge expressed as a quantified sentence. We are going to continue with example 3.4 that we proposed on section 3.3. We applied original Trapezoidal Temporal Window over two databases used in the previous section, and the new window obtained after expanding the window. Goodness measure is represented on figures 6 and 7: From these graphics, we observe that sometimes we extract results with the second window where the first window do not. This event happens because with the second window we have expanded the amount of data we studied to obtain the results. Although, always, we studied less data than the Crisp Method. Therefore, we adjusted the window to get better results than original window, without studying the whole interval.

6 Conclusions

In this paper, we have presented a method to detect user behaviour. As we have indicated, this problem is random imprecision by definition, because we do not know the interval of the time in which the action happen. Imprecision appears due of a behaviour is not static on a context or domain and is specific for every user. Thus, we have designed a method which uses fuzzy logic to limit events in a specific behaviour using a Fuzzy Temporal Window. With a Window, we can assign a membership degree for every event for the studied behaviour. In addition, we extend the Fuzzy Temporal Window concept, to adjust the interval using the knowledge from user we have. This adjustment is made with quantified sentences. With this adjustment we solve the possibility of a person changes his habits.

System's output is the fuzzy pattern sequences which define the behaviour. To manage non-random imprecision, we use the α-cuts and the Decomposition Theorem to obtain a fuzzy result applying the crisp method.

We have demonstrated that the fuzzy method is better than crisp method and that the fuzzy method obtains better outcomes than crisp one. As well as we conclude that results depend on the type of window we use. In addition, we have studied that outcomes are independent from the database.

In future work, we want to distinguish between three problems of our system:

- Non-random imprecision
- Uncertainty, because it could exists an error in the sensor reading. There are many reasons to provoke mistakes in sensors reading: the sensors or the reader could fail, the environment affects to the sensors, sensors damage, etc.
- Uncertainty in the Fuzzy Temporal Window, because the behaviour does no always happen exactly equal.

Our objective consists on extracting patterns that could control these kinds of problems, obtaining an adjusted pattern.

References

[1] M. Weiser. The computer for the twenty-first century. *Scientific American*, 1:94, 1994.

[2] Roy Want. Rfid: A key to automating everything. *Scientific American Magazine*, -:10, Enero 2004.

[3] Hiroyuki Yamahara et al. An individual behavioral pattern to provide ubiquitous service in intelligent space. volume 6, pages 562–569, 2007.

[4] T. Mori et al. Sensing room: Distributed sensor environment from measurement of human daily behaviour. *Intelligent Robots and Systems, 2005. (IROS 2005).IEEE/RSJ International Conference*, 2:1703 – 1709, 2-6 Aug. 2005.

[5] Kyohei Koyamaet al. Embedded action detector to enhance freedom from care. In *Proc. of 11th International Conference on Human-Computer Interaction*, page 8, Las Vegas, July 2005.

[6] Matthai Philipose et al. Inferring activities from interactions with objects. *IEEE Pervasive Computing*, 3(4):50–57, 2004.

[7] María Ros et al. Un sistema para identificacin de comportamientos correctos en un mundo marcado. Primeras Jornadas Cientficas sobre RFID, 21-23 Noviembre 2007.

[8] María Ros et al. Correct behavior identification system in a tagged world. *Expert systems with applications*, 36:9899–9906, 2009.

[9] L A Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computing and Mathematics with Applications*, (9):149–184, 1983.

[10] A. Gonzalez et al. Dealing with uncertainty and imprecision by means of fuzzy numbers. *International Journal of Approximate Reasoning*, 21:233–256(24), August 1999.

[11] Hung T. Nguyen and Elbert A. Walker. *A first Course in Fuzzy Logic*. Chapman & Hall/CRC Taoylr & Frances Group, 2007.

ISBN: 978-989-95079-6-8

Organizational Risk Assessment using Adaptive Neuro-Fuzzy Inference System

J.Jassbi [1] S.khanmohammadi [2]

1 Islamic Azad University, Science & Research Branch,
Poonak sq. Ashrafi Esfahani Blv., Hesarak, Tehran, Iran
2 Faculty of electrical Engineering, University of
Tabriz, Tabriz, Iran
Email: Jassbi@srbiau.ac.ir, khan@tabrizu.ac.ir

Abstract— *In this paper a fuzzy model based on Adaptive Neuro-Fuzzy Inference System (ANFIS) is introduced for calculating the level of risk in managerial problems. In this model, affecting factors on the level of risk are considered as inputs and the level of risk as the output. Using the fuzzy model, the risky condition changes smoothly in a fuzzy environment as it is the case in the real world; while in classic models, we may have some stepwise changes in the state of the system caused by an infinity small deviations in input parameters. The main advantage of the introduced model is that for continuous values of input factors, the counters of risk surface represent a more realistic behavior for different systems. The model is designed for a system of three inputs (probability, impact and ability to react) and one output (risky situation). The ability of using historical data as well as experts' knowledge and flexibility of adaptation to unusual risky situations are some benefits of the introduced model. This model which is originally used in strategic management system to analyze the external environment and the level of threats can also be used in contingency management for incidents (CMI) or as a tool for Comprehensive Emergency Management Program (CEMP). The strategic Risk of Roche is considered as the bench mark for comparing the difference between fuzzy and classic systems.*

Keywords— Risk assessment, Adaptive Neuro-Fuzzy Inference System, Risk Counters, Impact, Ability to react,

1 Introduction

Risk management has many applications in different fields such as financial management, industrial engineering, military and manufacturing systems as a Meta-disciplinary field. Companies and organizations have to monitor, scan and investigate the environmental factors and to update their strategic plans based on the level of recognized threats to select appropriate policies and strategies for reducing the risks. There are several steps for implementation of global risk management, starting with risk identification and calculation of its level for taking an appropriate decision. One of the most important sub-processes for this purpose is the methodology of calculation of risk and comparison of risky factors to determine the priorities. Considerable quantitative models have been introduced for this purpose in literature, where it is tried to calculate the level of the risk, which is simply defined as the rate of threat or future deficit of any system imposed by controllable or uncontrollable variables (Chavas, 2004; Doherty, 2000). Several factors such as probability of occurrence, impact, severity and ability to react are introduced as effecting factors on the risk. Then it is tried to find the mathematical relation between affecting factors and the value (level) of the risk (McNeil and Frey and Embrechts 2005; Li and Liao, 2007). The concept of risk is considerably wide. It can contain strategic, financial, operational or any other type of the risk. Based on literature review, the models which are mostly used in different fields of risk analysis can be classified to three types:

a) Data oriented models
b) Analytical models
c) Models based on judgment

Although it should be mentioned that this classification is not only related to Risk analysis, but in general it can be applied to system identification tasks.

a) Data oriented models:

In these models the structure is not needed to be known and the only important issue is the system's behavior. Some of the features of this type of models are: dependence to historical data, being behavioral, and storing experience and knowledge, based on the recognition of the input and output patterns of the system. Artificial intelligent models such as ANFIS and Neural Networks as well as statistical methods by concept of math average approach can be classified as this type of models. The more the number of the historical data is, the more reliable the models and their results would be. Figure 1 presents this type of model which is considered as a black box.

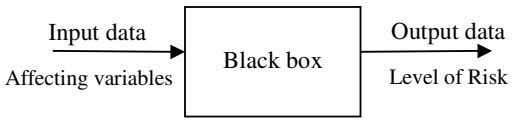


Figure. 1. Data-based System as a black box

This black box shows that there is no information on the analytical model of the system. It means that the relation among the variables as well as the way they affect each other is still unknown for user.

b) Analytical models.

In these systems it is possible to design a homomorphous model based on the system analysis process and structure of the problem. Being structure-based, analytical and independent from the data, are some of the features of this class of models. Many of the existing mathematical models,

ISBN: 978-989-95079-6-8

System Dynamics and a variety system analysis models are included in this class (Marrison2002). In this type of systems the structures must be recognized. It means that components and their relations must be identified to develop the analytical model. Figure 2 represents the concept of a structure based system. As it is mentioned the component and their relations must be recognized for developing an analytical model of the system.

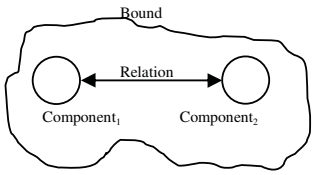


Figure. 2:. Structure-Based Systems (analysis)

For example if component$_1$ is Risk of investment when constructing a building, and component$_2$ represent the probability of earthquake in the region, the relation between them can be presented as :
Risk of investment=f (probability of earthquake)
Although it is a dream for researches to find an analytical model for such problems, however the major weakness of this kind of models is that they need precise data for modeling and the process of verification is considerably hard.

c) Models based on judgment

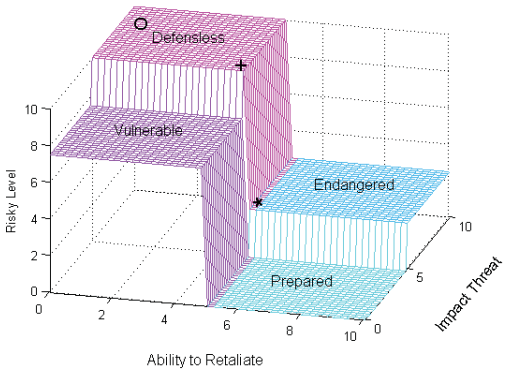
As it is mentioned, sometimes the first and second methods can not be used in real world problems. For example when the risk is based on the human behavior as a complex system, it is impossible for researches to find a pattern or mathematical model for it with available tools. So the judgment and knowledge of experts can help us to determine the level of risk and to find suitable solutions for risk management.

In this type of models, it is tried to overcome the major weakness of the first type consisting their uselessness in the case of no preexisting patterns. The growth of environmental dangers and the rapid increase of their variations as well as the increase of demand for such models by insurance companies have caused rapid and accelerated developments of such models (Roberets, 2005; Michael, 2004; Bernadell and CArdon and Coche and Diebold and Manganelli, 2004). This classification will help us to use the appropriate model in different situations.

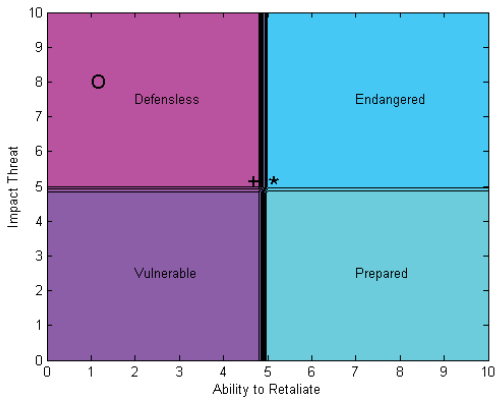
In this paper, the classic models of second type, commonly used in risk assessment of threats in strategic planning, are analyzed and criticized. Afterward a new model based on fuzzy inference is introduced for calculation of risk levels.

2. Problem Statement

Figure 3 shows a classic model of risk analysis. It consists of two factors: Impact threat and the ability to retaliate. In this model the risk value is classified in four groups. Each group represents a risky condition for the organization. During the implementation, first the opinions of the experts on the impact threat and the ability to retaliate are processed by means of any appropriate method such as group decision making, Delphi, …, and then the risky situation of organization is recognized. The distributions of the points with the same risk levels (contours of different levels) are also presented in Fig. 3(b), Points O and + represent the risky situation for two organizations with ability to retaliate and impact threats of (1, 8) and (4.9, 5.1) respectively.



(a)



(b)

Figure 3: (a) Risky situations classified in 4 levels, (b) counters

This model is very simple, but it has some structural drawbacks. For example the organization + which is in Endangered situation will change to completely opposite condition (Vulnerable) point(*) with infinity small deviations (A-da) in ability to retaliate. Also because of its geometrical structure, this model suffers from the lack of considering additional parameters such as ability to react.

Another method which has gained more attraction in the risk analysis literature is the model based on the linear combination of ability to retaliate and impact threat as:

$$Risk = (\text{ability to retaliate}) \times (\text{impact threat}) \quad (1)$$

Figure 4 represents a practical continuous increasing surface (levels), instead of stepwise levels for risk values. Two particular levels are shown by the cutting planes K1 and K2. Positions O, + and * are also presented in this figure. Figure 4(b) shows some contours of risky surface. As it is seen, in this model any small change in the values of probability and severity will cause a very small deviation in its risky level of organization. Also sometimes the risky level remains unchanged. This model is more realistic than the one

ISBN: 978-989-95079-6-8

presented by figure 3. However, it also has its limitations for real world applications because it simplifies the complicated relation between different factors that affect the risky conditions of organizations to a simple multiplication of two factors (ability to retaliate and impact threat).

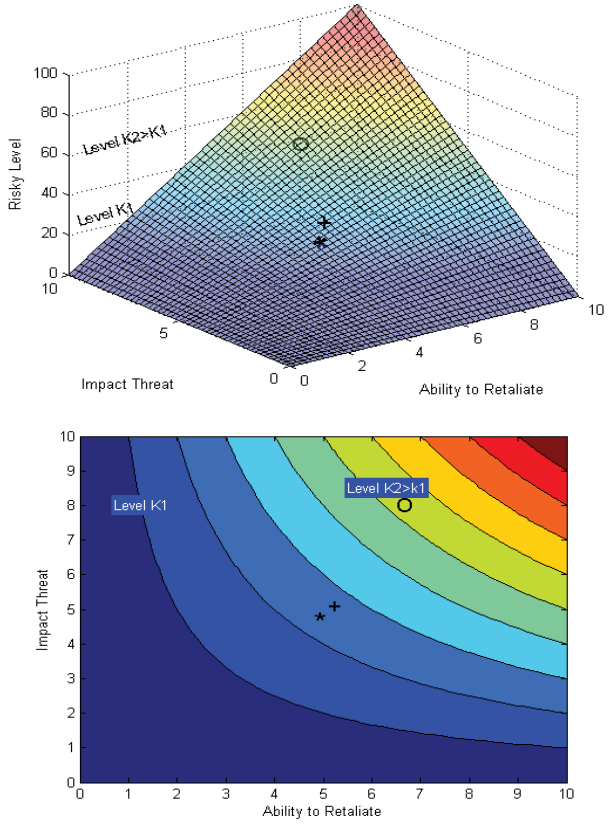


Figure. 4: Continuous surface for risk levels and counters

These kinds of models suffer from some weaknesses such as considering the relation between variables as a linear relation. Also when the number of inputs increases, the complexity of the system will increase dramatically while the accuracy of the model will decrease. Moreover due to inflexibility of the model, it can not be improved considering our experiences based on real world situations or the growth of experts' knowledge.

Hence we can formulate our problem as an input output system by:

$$R=F(X) \qquad (2)$$

Where X is the set of input variables which affect the level of the risk, R is the level of the risk and F(.) is a nonlinear function (Kreinovich and Nguyen, 2000).

The problem here is to find an appropriate solution by which the level of risk of the system (Organization) can be determined in complex situations when there is no access to all data, or the historical data is useless.

In this paper we try to develop a rule based model by using fuzzy inference system.

3. Fuzzy model

Fuzzy inference systems (FIS) are rule-based systems with concepts and operations associated with fuzzy set theory (Zadeh, 1965) and fuzzy logic (Ross, 2004; Mendel, 2001). These systems are mappings from an input space to an output state; therefore, they allow constructing structures that can be used to generate responses (outputs) to certain stimulations (inputs), based on stored knowledge on how the responses and stimulations are related. Sometimes this knowledge is obtained by eliciting information from specialists, in which case these systems are known as fuzzy expert systems (Takács, 2004). A fuzzy system also can be created to mach any set of input-output data. Adaptive Neuro-fuzzy Inference System (ANFIS) is one of the well known methods for creating Input-Output based models. (Krus , Gebhart and Palm, 1994). ANFIS only support Sugeno systems (Sugeno, 1985) subject to the structure of the system such as unit weight for each rule. Since its introduction, ANFIS has successfully been proved in many engineering applications (Jang, 1993). Another common denomination for FIS is fuzzy control systems (see for example (Mendel, 2001)).

FIS are usually divided in two categories (Mendel, 2001; Takagi and Sugeno, 1985): multiple input, multiple output (MIMO) systems, where the system returns several outputs based on the inputs it receives; and multiple input, single output (MISO) systems, where only one output is returned from multiple inputs. Since MIMO systems can be decomposed into a set of MISO systems working in parallel, all that follows will be exposed from a MISO point of view (Mamdani and Assilian, 1999).

FIS suffers of adjusting the linguistic knowledge of the expert with available data; so in this paper a fuzzy model based on ANFIS is introduced for calculating the risky situations of organizations by considering different factors such as probability, Impact threat and ability to retaliate (Nguene and Finger, 2007; Hyo and Hyun, and Yoon, 2002). This model is developed in three phases. To clarify the process of modeling in different phases a benchmark case study adopted from "strategic management by Rowe, et all" (Row et all, 1999) is used as a test example.

Phase 1: Data generation

The expert s' judgments may be extracted by means of different group decision making methods. Our experience in this work shows that the Delphi method as well as Nominal group is more practical. In case there are hard data as a result of historical behavior, a similar methodology can be used. By the way combination of hard data and information extracted from expertise can be used to develop first table which is necessary for next step.

Table 1 shows the set of generated data for test example.

Table.1

Probability	Impact	Ability to React	Risk (Vulnerability)
1	10	4	7.5
0	0	10	0
.5	5	5	1
1	5	5	3
.2	5	2	0
.8	2	8	5
.4	7	3	1

ISBN: 978-989-95079-6-8

1	0	5	0
.7	8	2	4
.8	8	7	2
.2	9	5	.5
.2	3	5	0
.7	10	3	4.5
.5	10	2	2.5
.5	10	10	0
1	8	0	10
1	2	8	.5
.6	10	4	3
.1	2	6	0
.3	6	8	0

Phase 2: Rule making

In this step the Adaptive Neuro-Fuzzy Inference System (ANFIS) is used for generating the rules.

Figure 5 Shows the surfaces of the rule base system adapted for the data of table 1 with:

numMFs = 25

mfType = gaussmf

epoch_n = 20

After constructing rule-base surface the model is improved by investigating the critical states (which are not approved by expert or when there is illogical behavior). Then by adding more rules or by imposing some minor deviations to data set the final rules may be obtained as shown in figure.5. By this way any specific behaviors cause by unusual relation between some particular inputs and output of the system can be supported which is significant advantage benefit of introduced model compare to other model such as Neural Network.

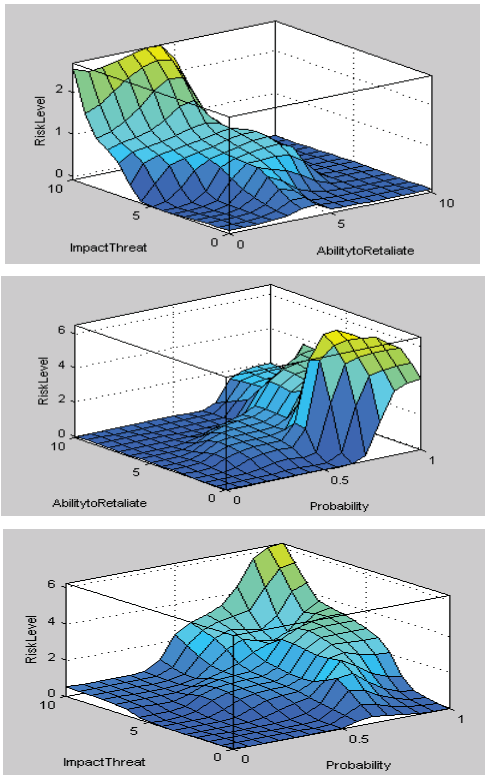
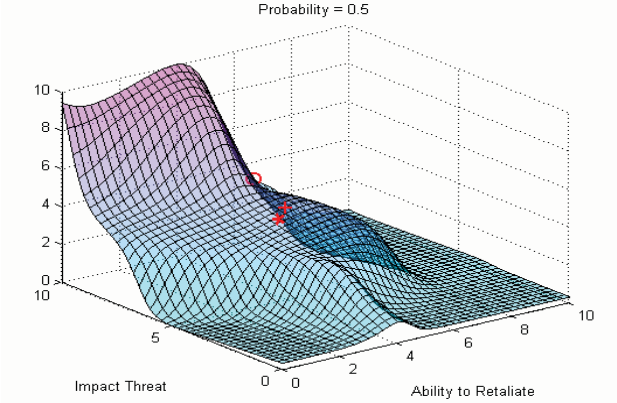
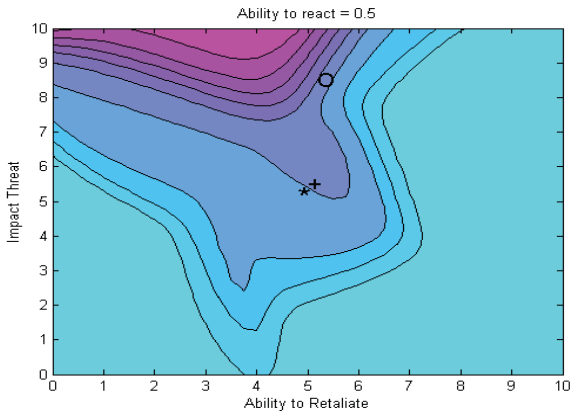


Figure.5: Surfaces of generated rules.

Phase 3. Implementation

This model is implemented to the simple example of section one to have an idea on the main difference between this method and the classic model. Figure 6 shows the surface and counters of risky levels of organizations +, and O. The results are shown for 50% of probability of impact to visualize the obtained results.





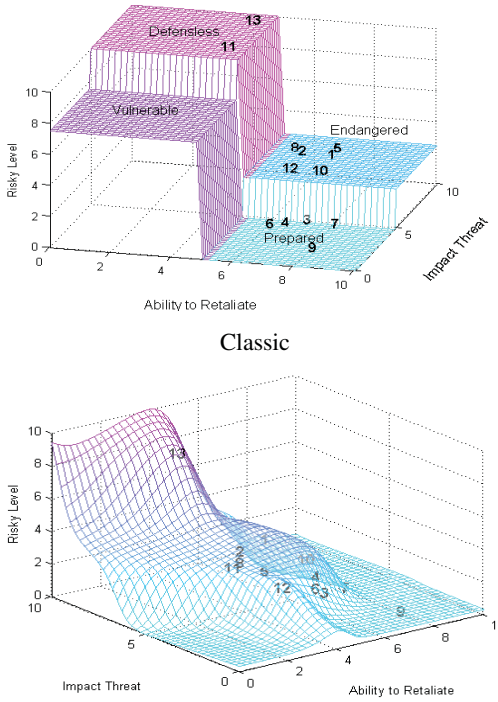
(b)

Figure. 6: (a) Risky surfaces, (b) counters for simple example

As it is seen, organization + which is appropriately Endangered situation will change to appropriately defenseless, point(*), with infinity small deviations (A-da) and (T-dt) in ability to retaliate and impact threat which is more realistic than the classic one.

The vulnerability analysis for Roche Company is also considered as a benchmark for implementing and comparing the obtained results with the classic methods. Figures 7 and 8 compare the results obtained by implementing the introduced model. Note that in the rest of paper only the surfaces and counters for probability of 50% are shown.

ISBN: 978-989-95079-6-8



Classic

Fuzzy

Figure.7: application of classic and fuzzy models for vulnerability analysis of Roche Company

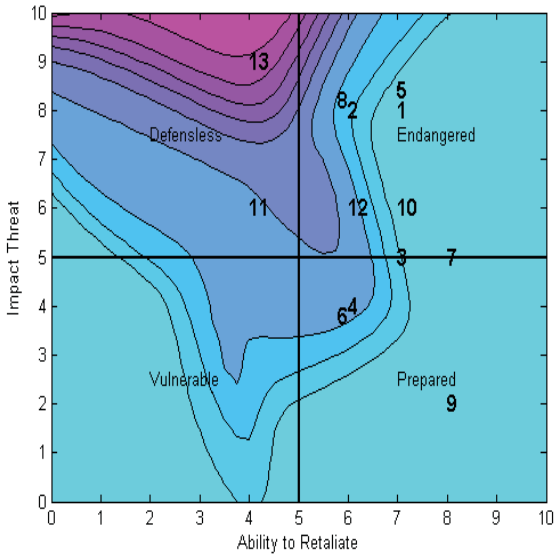


Figure 8: Risky levels of Roche by using classic and fuzzy models

Table 2 compares the results of this method by the classic one.

Table 2

Assumptions	Impact 0-10	Probability 0-1	Capability 0-10	Vulnerability Assessment Strategic management Approach	Vulnerability Assessment Classic	Vulnerability Assessment Fuzzy
1. Needs and wants served by products	8	0.7	7	Endangered	2.5	1.9960
2. Resources and Assets	8	0.6	6	Endangered	2.5	1.6119
3. Cost position relative to competition	5	0.4	7	Prepared	0	0.1490
4. Customer base	4	0.5	6	Prepared	0	1.9093
5. Technologies	8	0.4	7	Endangered	2.5	0.0446
6. Special skills	4	0.4	6	Prepared	0	1.1455
7. Corporate identity	5	0.5	8	Prepared	0	0.0298
8. Institutional barriers to competition	8	0.4	6	Endangered	2.5	0.8793
9. Social values	2	0.2	8	Prepared	0	0.0015
10. Sanctions, supports, and incentives	6	0.7	7	Endangered	2.5	1.7019
11. Customer goodwill	6	0.5	4	Defenseless	10	2.1834
12. Complementary products or services	6	0.3	6	Vulnerable	7.5	0.2943
13. Regulatory agencies	9	0.7	4	Defenseless	10	7.8129

It is seen that using the fuzzy method appropriate continues vulnerability levels are obtained to analyze the risky levels.

As it can be seen in Table.2 the Result of Vulnerability assessment by using Strategic management approach (as shown in fig.3) is stepwise and limited to 4 categories; hence the real value of risk is unclear. The second model (Classic Vulnerability Assessment) suffers from rigidity and can not be customized in different situations. So its efficiency can not be improved by considering available data and knowledge of experts. It is why in real world applications the decision Makers are not generally satisfied (trust the model) by the results obtained by this method.

ISBN: 978-989-95079-6-8

The introduced fuzzy based method overcomes the mentioned drawbacks. For example considering rows 1,2,5,8,10 in table.2, by using classical method as well as strategic management approach the same levels of risk are obtained for different assumptions where by using introduced method different level of risk is determined.

4. Conclusion

In this paper the fuzzy risk analysis procedure is compared with classic methods. It is demonstrated that the fuzzy analysis seems to be more reasonable and applicable because of its smoothness in calculating the risky situations. By fuzzy analysis the dynamic (time dependent) behavior of risk conditions can be defined and used for convenient and reliable decision makings. The model presented in this paper has some features as follows: a) Relations between variables in real life are nonlinear. Abstracting the situation and simplifying the problem to a linear model will cause the missing of some vital data where by utilizing the introduced model the relation between Risk and variables can be considered as a nonlinear function. b) Using FIS brings the advantage of developing and improving the model based on historical data. It is noted that the original model is constructed only based on the experts' knowledge due to lack of historical data. c) The model can be extended to be used for any number of inputs, where expanding the classic models to more inputs is not an easy task. d) This model considers probability as an input where other models usually solve the problem in a probabilistic environment. e) Any Specific behavior caused by unusual relation between some particular inputs and output of the system can be supported by imposing new rules to the model.

References

[1] Bernadell, C., CArdon, P., Coche, J., Diebold, F, X., Manganelli, S., *Risk management for central bank foreign reserves*. European Central Bank. 2004

[2] Carr, V., Tah, J, H, M., *A Fuzzy approach to construction project risk assessment and analysis*: construction project risk management system. Advances in Engineering Software, 32 (10 – 11), 847-857. 2001,

[3] Chavas, J, P., *Risk Analysis in Theory and Practice*. English, ISBN – 10: 0121706214. 2004,

[4] Doherty, N., *Integrated Risk Management*: Techniques and Strategies for Managing Corporate Risk, McGraw-Hill. 2000,

[5] Hall, L. O., Kandel, A., *The evolution from expert systems to fuzzy expert systems.* In Fuzzy Expert Systems, edited by A. Kandel, CRC press. 1991,

[6] Hyo, N, C., Hyun, H, C., Yoon, B, K., *A risk assessment methodology for incorporating uncertainties using fuzzy concepts.* Reliability Engineering & System Safety, 78, 173-183. 2002,

[7] Jang, J.S.R., *Adaptive-Network-based Fuzzy Inference System, IEEE Transactions on systems, Man, and Cybernetics* ,23(3) ,665-685,1993

[8] Kreinovich, V., Nguyen, H. T., *Fuzzy Systems are Universal Approximators for a Smooth Function and its Derivatives.* International Journal of Intelligent Systems, 15, 565-574. 2000,

[9] Kruse R. and Gebhardt J. and Palm R., Fuzzy Systems in Computer Science, Verlag Vieweg, P.O. Box 5829, D-65048 Wiesbaden, 1994"

[10] Li, Y., Liao, X,. *Decision support for risk analysis on dynamic alliance.* Decision Support Systems, 42, 2043-2059. 2007,

[11] Mamdani, E. H., Assilian, S., *An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller.* In. International Journal of Human-Computer Studies, 51 (2) pp. 135-147. 1999,

[12] Marrison C. , The Fundamentals of Risk Measurement, McGraw-Hill, 2002

[13] McNeil, A. J., Frey, R., Embrechts, P., *Quantitative Risk Management:* Concepts, Techniques, and Tools, Princeton University Press. 2005,

[14] Mendel, J., *Uncertain rule-based fuzzy inference systems:* Introduction and new directions. Prentice-Hall. 2001,

[15] Michael, P., *Systems modeling*; theory and practice, John Wiley & Sons, Ltd, England. 2004,

[16] Nguene, G, N., Finger, M., *A fuzzy – based approach for strategic choices in electric energy supply. The case of a swiss power provider on the eve of electricity market opening.* Engineering applications of Artificial Intelligence, 20, 37-48. 2007,

[17] Roberets, B. B., *The benefits of integrated quantitative risk management.* 12th annual international symposium council on system engineering, Melbourne. 2005,

[18] Ross, T. J., *Fuzzy logic with engineering applications.* John Wiley & Sons, Ltd. 2004,

[19] Rowe, Alan J., Mason, Richard O., Dickel, Karl E., Mann, Richard B., Mockler, Robert J. *Strategic Management: A Methodological Approach.* Addison N.Y. Page 204. 1999.

[20] Sugeno M., Industrial Applications of Fuzzy Control, Amsterdam, The Netherlands: North Holland, 1985

[21] Takács, M., *Critical Analysis of Various Known Methods for Approximate Reasoning in Fuzzy Logic Control.* 5th International Symposium of Hungarian Researchers on Computational Intelligence, Budapest, Hungary. 2004,

[22] Takagi, T., Sugeno, M., *Fuzzy identification of systems and its applications to modelling and control,* IEEE Trans. on Systems, Man and Cybernetics, 15:116-132,85. 1985,

[23] Zadeh L.A., Fuzzy Sets, Information and Control, vol.8, pp. 338--353, 1965"

ISBN: 978-989-95079-6-8

On the use of min-based revision under uncertain evidence for possibilistic classifiers

Salem Benferhat[1] Karim Tabia[1]

1.CRIL CNRS UMR8188
Artois University, France
Email: benferhat@cril.univ-artois.fr, tabia@cril.univ-artois.fr

Abstract— *Possibilitic networks, which are compact representations of possibility distributions, are powerful tools for representing and reasoning with uncertain and incomplete knowledge. According to the operator conditioning is based on, there are two possibilistic settings: quantitative and qualitative. This paper deals with qualitative possibilistic network classifiers under uncertain inputs. More precisely, we first present and analyze Jeffrey's rule for revising possibility distributions by uncertain observations in the qualitative setting. Then, we propose an efficient algorithm for revising possibility distributions encoded by naive possibilistic networks for classification purposes. This algorithm consists in a series of efficient and equivalent transformations of initial naive possibilistic classifiers.*

Keywords— Min-based possibilistic networks, classification under uncertain inputs

1 Introduction

Graphical models such as Bayesian networks [1][2], influence diagrams and possibilistic networks [3][4] are well-known formalisms widely used for representing and reasoning with uncertain and incomplete knowledge. Possibilistic networks are like Bayesian ones but lie on possibility theory [5][6] to handle imprecise and incomplete knowledge. They allow to factor a global joint possibility distribution into a set of local possibility distributions. There are two types of possibilistic networks: quantitative and qualitative. In the former, conditioning is based on the $product$ operator while it is based on the min-operator in the latter.

Classification is an important task in many real world applications consisting in predicting the class instance corresponding to an observation. This task is a special kind of inference: given an observed instance of each observable variable A_i, it is required to determine the class instance c_k for the instance to classify among a predefined set of class labels. It is important to note that possibilistic classifiers have not been sufficiently studied in spite of the fact that they are very appropriate for problems where knowledge is imprecise or missing as in real-time classification problems or problems where some inputs are missing or uncertain. These problems require classifiers under uncertain inputs. However, only few works used possibilistic classifiers [7][8][9][10] and to the best of our knowledge, there is only one preliminary work [11] addressing possibilistic network classifiers under uncertain inputs in the quantitative setting. More precisely, this work proposes an algorithm suitable of classification under uncertain inputs using product-based possibilistic networks.

This paper also addresses possibilistic-based classification with uncertain observations but in the qualitative setting. While quantitative possibilistic networks are quite similar to the Bayesian ones, the qualitative ones show significant differences. For instance, normalizing conditional possibility distributions in product-based possibilistic networks is the same as in the probabilistic setting while it is significantly different in min-based networks. We first recall the qualitative possibilistic counterpart of Jeffrey's rule [12] for revising possibility distributions by uncertain observations and show that this rule cannot be directly applied for revising possibilistic knowledge encoded by a possibilistic network. Indeed, classification by directly using the possibilistic counterpart of Jeffrey's rule is exponential in the number of attributes and attribute domains. Then, we proposed an efficient method for revising naive min-based possibilistic networks suitable for classification with uncertain inputs. This algorithm is based on a series of equivalent and polynomial transformations of initial possibilistic networks taking into account uncertain inputs.

The rest of this paper is organized as follows: Section 2 briefly presents basic background about possibility theory and possibilistic networks. In section 3, we address possibilistic belief revision based on the possibilistic counterpart of Jeffrey's rule. Section 4 proposes a new efficient algorithm for naive possibilistic network classification with uncertain inputs. Finally, section 5 concludes this paper.

2 Basic background on possibility theory and possibilistic networks

Let us first fix our notations: V=$\{A_1, A_2, .., A_n\}$ denotes the set of variables. D_A=$\{a_1, a_2, .., a_m\}$ denotes the finite domain of variable A. a_i denotes an instance (value) of variable A_i. A, X,.. denote subsets of variables from V. D_X=$\times_{A_i \in X} D_{A_i}$ represents the cartesian product relative to variables A_i involved in subset X. Ω=$\times_{A_i \in V} D_{A_i}$ denotes the universe of discourse and consists in the cartesian product of all variable domains involved in V. A tuple w=$(a_1, a_2, .., a_n)$ or w=$a_1 a_2 .. a_n$ which is an instance of Ω represents a possible state of the world. ϕ, φ denote subsets of Ω called events while $\overline{\phi}$ denotes the complementary of ϕ in Ω ($\overline{\phi}$=Ω-ϕ).

2.1 Possibility theory

Possibility theory was introduced by Zadeh [5] and developed by Dubois and Prade [6]. It is an uncertainty theory based on a pair of dual measures in order to evaluate knowledge/ignorance relative to event in hand. The concept of possibility distribution π is one of the important building blocks of possibility theory: It is a mapping from the universe of discourse Ω to the unit scale $[0, 1]$ which can be either quantitative or qualitative. In both these settings, a possibility degree $\pi(w_i)$ expresses to what extent it is consistent that w_i can be

ISBN: 978-989-95079-6-8

the actual state of the world. In particular, $\pi(w_i)$=1 means that w_i is totally possible and $\pi(w_i)$=0 denotes an impossible event. The relation $\pi(w_i)>\pi(w_j)$ means that w_i is more possible than w_j. A possibility distribution π is said to be normalized if $max_{w_i \in \Omega}(\pi(w_i))$=1. It is said to be sub-normalized otherwise.

The second important concept in possibility theory is the one of possibility measure denoted $\Pi(\phi)$ and computing the possibility degree relative to an event $\phi \subseteq \Omega$. It evaluates to what extent ϕ is consistent with the current knowledge encoded by possibility distribution π on Ω. It is defined as follows:

$$\Pi(\phi) = max_{w_i \in \phi}(\pi(w_i)). \quad (1)$$

The term $\Pi(\phi)$ denotes the possibility degree relative to having one of the events involved in ϕ as the actual state of the world.

The necessity measure is the dual of possibility measure and evaluates the certainty implied by the current knowledge of the world. Namely, $N(\phi)$=1-$\Pi(\overline{\phi})$ where $\overline{\phi}$ denotes the complementary of ϕ.

Given a possibility distribution π on Ω, marginal distributions π_X relative to subset of variables X ($X \subseteq V$) are computed using the max operator as follows:

$$\pi_X(x) = max_{w_i \in \Omega}(\pi(w_i) : w_i[X] = x), \quad (2)$$

where term $w_i[X] = x$ denotes the fact that x is the instantiation of X in w_i.

According to the interpretation underlying the possibilistic scale [0,1], there are two variants of possibility theory:

- **Qualitative possibility theory**: In this case, the possibility distribution is a mapping from the universe of discourse Ω to an "ordinal" scale where only the "ordering" of values is important.
- **Quantitative possibility theory**: In this case, the possibilistic scale [0,1] is numerical and possibility degrees are like numeric values that can be manipulated by arithmetic operators.

In this paper, we only focus on qualitative setting. Conditioning is a fundamental notion concerned with updating the current knowledge (encoded by a possibility distribution π) when an evidence (a sure event) is observed.

In the qualitative setting, conditional possibility degree of w_i given an event ϕ is computed as follows (we assume that $\Pi(\phi) \neq 0$) [13]:

$$\pi_m(w_i|\phi) \begin{cases} 1 & \text{if } \pi(w_i)=\Pi(\phi) \text{ and } w_i \in \phi; \\ \pi(w_i) & \text{if } \pi(w_i)< \Pi(\phi) \text{ and } w_i \in \phi; \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

2.2 Brief description of possibilistic networks

Like Bayesian networks, possibilistic ones [4] involves two components:

1. A graphical component consisting in a DAG[1] which encodes direct influence relationships existing between domain variables.

2. A numerical component which is a "numerical" component composed of a set of local a priori and conditional possibility distributions. The latter measure the influence endured by each domain variable A_i in the context of its parents U_{A_i}. Local possibility distributions should satisfy the normalization condition denoted as follows:

$$max_{a_{ij} \in D_{A_i}}(\pi(a_{ij}|U_{A_i})) = 1 \quad (4)$$

The joint possibility distribution encoded by the network is computed using the min-based chain rule. Namely,

$$\Pi(A_1, A_2, .., A_n) = min_{i=1..n}(\pi(A_i|U_{A_i})) \quad (5)$$

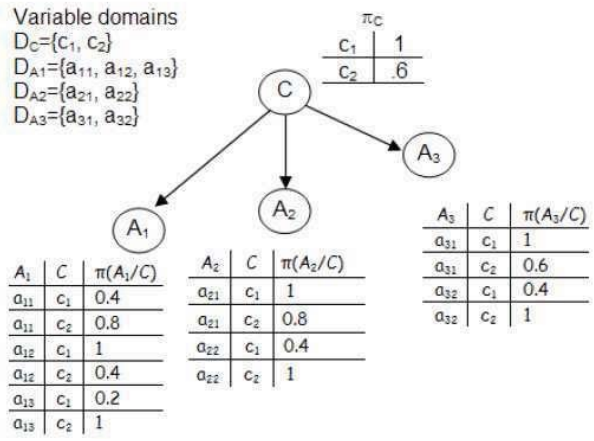
In classification problems, there exists one node associated with the class variable C which is not observable (it is the target variable) while the remaining nodes represent the attributes A_1, A_2,..,A_n that may be observable. Classification is ensured by computing the most plausible class instance given the instance to classify. Namely, given an observation denoted A=(a_1,a_2,..,a_n) of $\{A1, A_2, .., A_n\}$, the predicted class c is determined as follows:

$$c = argmax_{c_k \in D_C}(\Pi(c_k|A)) \quad (6)$$

Note that term $\Pi(c_k|A)$ denotes the possibility degree of having c_k the actual class instance given the observation A=(a_1,a_2,..,a_n). It is important to note that a class instance c_k is candidate for a given observation $a_1..a_n$ implies that $\Pi(c_k|a_1..a_n)$=1. Note also that several class instances may be totally possible for the observation to classify.

2.3 Min-based naive possibilistic network classifiers

A naive network classifier is the simplest form of possibilistic network classifiers. It lies on the strong independence assumption of attributes in the context of the parent node: attributes are assumed independent in the context of the class node. As it is shown in Figure 1, the only dependencies allowed in naive classifiers are from the class node C to each attribute A_i. As for the quantitative component of a naive possibilistic network, it involves the prior possibility distribution relative to the class node and the conditional possibility distributions relative to attributes given the class node. Figure 2 gives the joint possibility distribution encoded by network of Figure 1.



C	$\pi(C)$
c_1	1
c_2	.6

A_1	C	$\pi(A_1/C)$
a_{11}	c_1	0.4
a_{11}	c_2	0.8
a_{12}	c_1	1
a_{12}	c_2	0.4
a_{13}	c_1	0.2
a_{13}	c_2	1

A_2	C	$\pi(A_2/C)$
a_{21}	c_1	1
a_{21}	c_2	0.8
a_{22}	c_1	0.4
a_{22}	c_2	1

A_3	C	$\pi(A_3/C)$
a_{31}	c_1	1
a_{31}	c_2	0.6
a_{32}	c_1	0.4
a_{32}	c_2	1

Figure 1: Naive possibilistic network structure

Classification is ensured using min-based conditioning (see Equation 3) by computing the a posteriori possibility degree for each class instance c_k given the observation $A_1..A_n$ (namely $\Pi(c_k|A_1..A_n)$). Namely,

$$\Pi(c_k|A_1..A_n) = min(\pi(c_k, A_1..A_n), \Pi(A_1..A_n)) \quad (7)$$

[1] Direct Acyclic Graph

ISBN: 978-989-95079-6-8

A_1	A_2	A_3	C	$\pi(CA_1A_2A_3)$
a_{11}	a_{21}	a_{31}	c_1	0.4
a_{11}	a_{21}	a_{31}	c_2	0.6
a_{11}	a_{21}	a_{32}	c_1	0.4
a_{11}	a_{21}	a_{32}	c_2	0.6
a_{11}	a_{22}	a_{31}	c_1	0.4
a_{11}	a_{22}	a_{31}	c_2	0.6
a_{11}	a_{22}	a_{32}	c_1	0.4
a_{11}	a_{22}	a_{32}	c_2	0.6
a_{12}	a_{21}	a_{31}	c_1	1
a_{12}	a_{21}	a_{31}	c_2	0.4
a_{12}	a_{21}	a_{32}	c_1	0.4
a_{12}	a_{21}	a_{32}	c_2	0.4

A_1	A_2	A_3	C	$\pi(CA_1A_2A_3)$
a_{12}	a_{22}	a_{31}	c_1	0.4
a_{12}	a_{22}	a_{31}	c_2	0.4
a_{12}	a_{22}	a_{32}	c_1	0.4
a_{12}	a_{22}	a_{32}	c_2	0.4
a_{13}	a_{21}	a_{31}	c_1	0.2
a_{13}	a_{21}	a_{31}	c_2	0.6
a_{13}	a_{21}	a_{32}	c_1	0.2
a_{13}	a_{21}	a_{32}	c_2	0.6
a_{13}	a_{22}	a_{31}	c_1	0.2
a_{13}	a_{22}	a_{31}	c_2	0.6
a_{13}	a_{22}	a_{32}	c_1	0.2
a_{13}	a_{22}	a_{32}	c_2	0.6

Figure 2: Joint possibility distribution encoded by naive classifier of Figure 1

Note that according to min-based conditioning, Equation 7 is only valid when term $\pi(c_k, A_1..A_n) < \Pi(A_1..A_n)$ (See min-based conditioning of Equation 3). To the best of our knowledge, there is no work that addresses the problem of classification under uncertain inputs using min-based possibilistic networks.

3 Min-based revision for classification under uncertain inputs

In our context, uncertainty relative to uncertain/missing attribute A_i is represented by a possibility distribution π'_{A_i} given for example by the expert. Moreover, uncertainty can bear on any attribute subset or on the whole attribute set.
Jeffrey proposed in [12] a method based on probability kinematics for revising a probability distribution p into p' given uncertainty bearing on a set of mutually exclusive and exhaustive events λ_i. In this method, uncertainty is of the form (λ_i, α_i) with α_i=$p'(\lambda_i)$. Jeffrey's rule states that although there is uncertainty about events λ_i, conditional probability of any event $\phi \subseteq \Omega$ given any uncertain event λ_i. The possibilistic counterpart of this rule has been investigated in [14]. In the possibilistic framework, revised possibility distribution π' must comply with the principle stating that uncertainty about events λ_i must not alter the conditional possibility degree of any event $\phi \subseteq \Omega$ given any event λ_i. Namely,

$$\forall \lambda_i \in \Omega, \quad \forall \phi \subseteq \Omega, \Pi'(\phi|\lambda_i) = \Pi(\phi|\lambda_i) \tag{8}$$

Note that contrary to the probabilistic and quantitative possibilistic settings, Equation 8 may have several solutions. In this case, the least specific distribution is selected according to the principle stating that if an event is not explicitly discarded, then it must remain possible. Min-based conditioning allows to obtain from Equation 8 the following one:

$$\forall \phi, \Pi'(\phi) = max_{\lambda_i}(min(\pi'(\lambda_i, \phi), \pi(\lambda_i, \phi), \Pi(\phi)) \tag{9}$$

In classification with uncertain inputs problems, the possibilistic counterpart of Jeffrey's rule cannot directly be applied to revise the possibility distribution encoded by the classifier because of two major problems:
1- The first problem concerns the fact that Jeffrey's rule can be applied only if uncertainty bears on a set of exhaustive and mutually exclusive events while in classification with uncertain inputs problems, uncertainty is bearing on a set of attributes A_i which are not mutually exclusive. Indeed, if inputs A_1,..,A_n are uncertain, then this uncertainty is encoded by $\pi'(A_1)$,..,$\pi'(A_n)$ respectively. In order to apply Jeffrey's rule, we must compute a joint possibility distribution $\pi'(A_1..A_n)$ relative to $A_1..A_n$ where uncertain events $a_1..a_n$ are exhaustive and mutually exclusive. Obviously, the term $\pi'(A_1..A_n)$ is function of $\pi'(A_1)$,..,$\pi'(A_n)$ and will be henceforth denoted by $\pi'(A_1..A_n)$=$f(\pi'(A_1)$,..,$\pi'(A_n))$. Uncertain inputs $\pi'(A_1)$,..,$\pi'(A_n)$ can be combined using a combination operator according to the problem constraints and objectives. It is important to note that the algorithm we propose in next section works regardless the combination function f provided that this latter satisfies the following natural propriety (called henceforth unanimity):
i) If $\forall$i=1..n, $\pi'(a_i)$=1 then $\pi'(a_1..a_n)$=1.
ii) If $\exists a_i \in D_{A_i}$ such that $\pi'(a_i)$=α<1, then $\pi'(a_1..a_i..a_n)$<1 whatever are the possibility degrees of the other variable values.
One can easily check that the $product$ and min (and more generally t-$norm$ operators), which are the most used combination operators in the possibilistic framework satisfy this natural propriety. However, max and more generally, t-$conorm$ operators do not satisfy it. In this paper, we will use the min operator in order to induce the joint possibility distribution $\pi'(A_1..A_n)$ from uncertain inputs $\pi'(A_1)$,..,$\pi'(A_n)$. Namely,

$$\pi'(a_1..a_n) = min_{k=1..n}(\pi'(a_k)) \tag{10}$$

2- The second problem concerns the computational complexity of performing classification under uncertain inputs based on revising the possibility distribution encoded by the classifier by uncertain inputs using the possibilistic counterpart of Jeffrey's rule. Namely, given the initial possibility degrees of class instances c_k, we want to revise this possibility distribution given uncertain attributes A_1,..,A_n using Jeffrey's rule. More precisely, we need to compute $\pi'(c_k)$=$\pi(c_k|\pi'(A_1, .., A_n))$ defined as follows:

$$\pi'(c_k) = max_{A_1..A_n}(min(\Pi(c_k|A_1..A_n), \pi'(A_1..A_n))) \tag{11}$$

It is clear that Equation 11 cannot be used to compute $\pi'(c_k)$ because this computation is exponential in the number of attributes and attribute domains.

Example:
Let us illustrate possibilistic classification with uncertain inputs using Jeffrey's rule on the naive possibilistic classifier of Figure 1. The uncertain inputs to classify are provided by Figure 3. Classification is ensured by determining the most

π'_{A1}	
a_{11}	1
a_{12}	.4
a_{13}	1

π'_{A2}	
a_{21}	1
a_{22}	1

π'_{A3}	
a_{31}	1
a_{32}	.4

Figure 3: Uncertain inputs to classify

plausible class instance(s) given these uncertain inputs. Figure 4 gives the results of revising the initial joint possibility distribution (see Figure 2) using the possibilistic counterpart of Jeffrey's rule (see Equation 11). One can easily deduce from joint possibility distribution π' of Figure 4 that only class instance c_2 is totally plausible ($\pi'(c_2)$=1) given the uncertain observations of Figure 3. However, in order to obtain such a result, we computed the revised joint possibility distribution. We would like to ensure the same classification without computing the joint possibility distribution. Next section proposes

ISBN: 978-989-95079-6-8

A_1	A_2	A_3	C	$\pi'(CA_1A_2A_3)$
a_{11}	a_{21}	a_{31}	c_1	0.4
a_{11}	a_{21}	a_{31}	c_2	**1**
a_{11}	a_{21}	a_{32}	c_1	0.4
a_{11}	a_{21}	a_{32}	c_2	0.4
a_{11}	a_{22}	a_{31}	c_1	0.4
a_{11}	a_{22}	a_{31}	c_2	**1**
a_{11}	a_{22}	a_{32}	c_1	0.4
a_{11}	a_{22}	a_{32}	c_2	0.4
a_{12}	a_{21}	a_{31}	c_1	0.4
a_{12}	a_{21}	a_{31}	c_2	0.4
a_{12}	a_{21}	a_{32}	c_1	0.4
a_{12}	a_{21}	a_{32}	c_2	0.4

A_1	A_2	A_3	C	$\pi'(CA_1A_2A_3)$
a_{12}	a_{22}	a_{31}	c_1	0.4
a_{12}	a_{22}	a_{31}	c_2	0.4
a_{12}	a_{22}	a_{32}	c_1	0.4
a_{12}	a_{22}	a_{32}	c_2	0.4
a_{13}	a_{21}	a_{31}	c_1	0.2
a_{13}	a_{21}	a_{31}	c_2	**1**
a_{13}	a_{21}	a_{32}	c_1	0.2
a_{13}	a_{21}	a_{32}	c_2	0.4
a_{13}	a_{22}	a_{31}	c_1	0.2
a_{13}	a_{22}	a_{31}	c_2	**1**
a_{13}	a_{22}	a_{32}	c_1	0.2
a_{13}	a_{22}	a_{32}	c_2	0.4

Figure 4: Revised joint possibility distribution π'

an efficient algorithm suitable for naive min-based possibilistic network classification with uncertain inputs.

4 A polynomial algorithm for naive possibilistic classifier under uncertain inputs

Classification based on possibilistic networks is ensured by determining if each class instance $c_k \in D_C$ is totally possible given the inputs ($\Pi(c_k|A_1..A_n)$=1). This is the basic idea of our algorithm: only search for class instances having a posteriori possibility degrees equal to 1 on the basis of uncertain inputs. Our algorithm ensures this search through a series of equivalent transformations on the initial possibilistic network taking into account the inputs to classify. The five steps of our algorithm are detailed in the following.

4.1 Step 1: Eliminating not totally possible observations

In our context, classification is ensured by revising the possibility distribution encoded by the classifier according to the possibilistic counterpart of Jeffrey's rule (using Equation 11). Since we are only interested in determining if a given class label c_k has an a posteriori possibility degree equal to 1, it is possible to discard all instances of $A_1A_2..A_n$ where $\pi'(A_1A_2..A_n)$<1 because such instances force the value of $\pi'(c_k)$ to be less than 1. According to the *unanimity* propriety, each input $a_i \in D_{A_i}$ which is not totally possible ($\pi'(a_i)$<1) will prevent every input configuration $a_1..a_i..a_n$ from being totally possible. It is clear that such configurations are useless for the classification task and can be discarded. This leads to eliminating from each attribute domain $D_{A_i}^{\Pi G}$ (relative to attribute A_i in initial network ΠG) values whose possibility degrees in π'_{A_i} are less than 1. In this step, attribute domain $D_{A_i}^{\Pi G}$ is changed to $D_{A_i}^{\Pi G_{S_1}}$ which denotes A_i's domain in network ΠG^{S_1} obtained from ΠG after Step 1. Namely, $D_{A_i}^{\Pi G_{S_1}} = D_{A_i}^{\Pi G} - \{a_i, if\ \ \pi'_{A_i}(a_i)<1\}$. All the remaining attribute instances are totally possible after this step. Then we have the following proposition:

Proposition 1 Let ΠG be the naive possibilistic classifier encoding the initial knowledge and π'_{A_1}, π'_{A_2},..,π'_{A_n} be the possibility distributions encoding uncertainty relative to attributes A_1, A_2,.., A_n respectively. Let ΠG^{S_1} be the naive possibilistic network obtained by eliminating not totally possible instances. Then,
$\pi'^{\Pi G}(c_k) = max_{A_1..A_n}(\Pi^{\Pi G}(c_k|a_1..a_n) * \pi'(a_1..a_n))$=1
if and only if
$\pi^{\Pi G_{S_1}}(c_k) = max_{A_1..A_n}(\Pi^{\Pi G_{S_1}}(c_k|a_1..a_n)) = 1$

This proposition states that if there is a class instance c_k which is totally possible in the initial network ΠG given the uncertain inputs then it is also totally possible in network ΠG^{S_1} obtained after eliminating not totaly possible inputs.

Example (continued): Network of Figure 5 is obtained by respectively substituting $D_{A_1}^{\Pi G}$=$\{a_{11}, a_{12}, a_{13}\}$, $D_{A_2}^{\Pi G}$=$\{a_{21}, a_{22}\}$ and $D_{A_3}^{\Pi G}$=$\{a_{31}, a_{32}\}$ by $D_{A_1}^{\Pi G_{S_1}}$=$\{a_{11}, a_{13}\}$, $D_{A_2}^{\Pi G^{S_1}}$= $\{a_{21}, a_{22}\}$ and $D_{A_3}^{\Pi G^{S_1}}$=$\{a_{31}\}$.

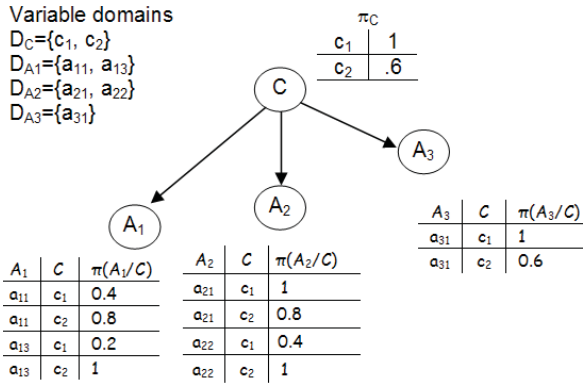


Figure 5: ΠG^{S_1}: Possibilistic network after Step 1

It is important to note that the possibility distribution relative to node C in ΠG^{S_1} is exactly the same as in ΠG.

4.2 Step 2: Eliminating unary variables

After the instance elimination step, some attributes may have their domains only containing one instance. In the example of Figure 5, attribute A_3's domain $D_{A_3}^{\Pi G_{S_1}}$ exactly contains one element (which is a_{31}). Such unary variables can be eliminated provided that class node distribution is adapted to keep the joint possibility distribution unchanged. The possibility degree of any instance $c_k a_1 a_2..a_n$ in network ΠG^{S_1} is computed using the min-based chain rule as follows:

$$\Pi^{\Pi G_{S_1}}(c_k a_1..a_n) = min(\pi^{\Pi G_{S_1}}(c_k), \pi^{\Pi G_{S_1}}(a_1|c_k), .., \pi^{\Pi G_{S_1}}(a_n|c_k)) \quad (12)$$

Unary attributes can be eliminated giving a new network ΠG^{S_2} where class node distribution has to be adjusted in order to guarantee that $\pi^{\Pi G_{S_1}}(c_k a_1..a_n)$ and $\pi^{\Pi G_{S_2}}(c_k a_1..a_n)$ remain equal. Hence, for an unary attribute A_i whose domain $D_{A_i}^{\Pi G_{S_1}}$ contains only one value a_i, it is possible to achieve this transformation by substituting each $\pi^{\Pi G_{S_1}}(c_k)$ with $\pi^{\Pi G_{S_2}}(c_k)$=min($\pi^{\Pi G_{S_1}}(c_k), \pi^{\Pi G_{S_1}}(a_i|c_k)$). This finding is formalized by the following proposition:

Proposition 2 Let ΠG^{S_1} be the naive possibilistic classifier whose nodes involve class node C and attribute nodes A_1, A_2,..,A_n. Assume that A_1 is a unary attribute whose domain only contains the instance a_1 and let ΠG^{S_2} be the naive possibilistic network involving C, A_2,..,A_n such that: $\pi^{\Pi G_{S_2}}(a_i)$= $\pi^{\Pi G_{S_1}}(a_i)$ for i=2,..,n and $\pi^{\Pi G_{S_2}}(c_k)$=min($\pi^{\Pi G_{S_1}}(a_1|c_k), \pi^{\Pi G_{S_1}}(c_k)$). Then,
$\forall c_k \in D_C, \forall a_i \in D_{A_i}^{\Pi G_{S_1}}$ for i=1..n,
$\pi^{\Pi G_{S_1}}(c_k a_1 a_2..a_n)$=$\pi^{\Pi G_{S_2}}(c_k a_2..a_n)$.

Example (continued): Let us continue our example. In network of Figure 5, attribute A_3 is unary and will be removed. After Step 2, we obtain network of Figure 6.

ISBN: 978-989-95079-6-8

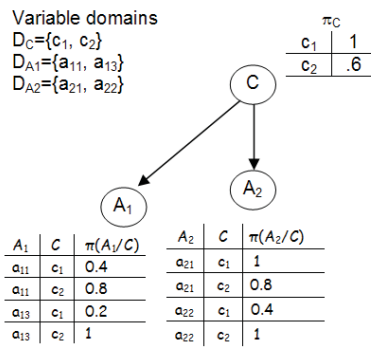


Figure 6: ΠG^{S_2}: Possibilistic network after Step 2

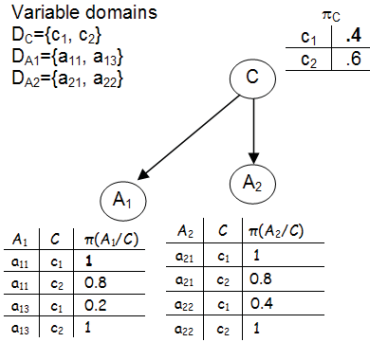


Figure 7: ΠG^{S_3}: Possibilistic network example after Step 3

4.3 Step 3: Re-normalizing local possibility distributions

As a consequence of the instance elimination step (Steps 1), local possibility distributions of some attributes might not be normalized. Indeed, it may exist a variable A_i and a class label c_k such that $max_{A_i}(\pi^{\Pi G_{S_2}}(a_i|c_k))=\alpha$ ($\alpha<1$). Step 3 deals with this problem considering two cases:

- If $\alpha=0$, then $\forall a_i \in D_{A_i}^{\Pi G_{S_2}}$, $\pi^{\Pi G_{S_2}}(c_k a_1..a_i..a_n)=0$ meaning that whatever is the value of A_i, this forces $\pi^{\Pi G_{S_2}}(c_k a_1 a_2..a_n)=0$. Hence, the class c_k cannot be among plausible ones, and can be removed from $D_C^{\Pi G_{S_2}}$.
- If $0<\alpha<1$, re-normalizing the conditional possibility distribution relative to attribute A_i can be done by building a new network ΠG^{S_3} by substituting $\pi^{\Pi G_{S_2}}(a_i|c_k)$ with

$$\begin{cases} 1 & \text{if } \pi(a_i|c_k)=max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k)); \\ \pi(a_i|c_k) & \text{otherwise.} \end{cases}$$

and substituting $\pi^{\Pi G_{S_2}}(c_k)$ with $min(\pi^{\Pi G_{S_2}}(c_k), max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k)))$. After this transformation, $\pi^{\Pi G_{S_3}}(c_k a_1..a_n)$ is conserved while sub-normalized local possibility distribution becomes normalized.

Proposition 3 Let ΠG^{S_2} be the naive possibilistic network involving nodes C,A_1,A_2,..,A_n obtained from Step 2. Assume that conditional possibility distribution relative to node A_1 is not normalized ($\exists c_k$ such that $max_{A_1}(\pi^{\Pi G_{S_2}}(a_1|c_k))=\alpha$ and $0<\alpha<1$). Let ΠG^{S_3} be the naive possibilistic network having same structure as ΠG^{S_2} where $\pi^{\Pi G_{S_3}}(a_1|c_k)=1$ if $\pi(wa_i|c_k)=max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k))$, $\pi(a_i|c_k)$ otherwise and $\pi^{\Pi G_{S_3}}(a_i|c_k)=\pi^{\Pi G_{S_2}}(a_i|c_k)$ for i=2,..,n and $\pi^{\Pi G_{S_3}}(c_k)$=min$(\pi^{\Pi G_{S_2}}(c_k), max_{a_j \in D_{A_i}}(\pi^{\Pi G_{S_2}}(a_j|c_k)))$. Then, $\forall c_k \in D_C^{\Pi G_{S_3}}$, $\forall a_i \in D_{A_i}^{\Pi G_{S_3}}$, $\pi^{\Pi G_{S_3}}(c_k a_1 a_2..a_n)=\pi^{\Pi G_{S_2}}(c_k a_1 a_2..a_n)$

Note that contrary to product-based possibilistic networks, normalizing conditional possibility distribution of min-based networks is significantly different from normalizing local probability distributions of Bayesian networks.

Example (continued): In network ΠG^{S_2} of Figure 6, local possibility distribution of attribute A_1 is not normalized ($max_{a_i \in D_{A_1}}(\pi^{\Pi G_{S_2}}(a_i|c_1))$=.4). After re-normalizing ΠG^{S_2}, we obtain network of Figure 7. Possibility distribution relative to A_1 has been re-normalized and possibility distribution relative to class node C has been adjusted accordingly. One can easily check that joint possibility distributions encoded by networks of Figure 6 and Figure 7 are equal.

4.4 Step 4: Prior totally possible class lookup

The aim of Steps 1, 2 and 3 is simplifying and re-normalizing the initial network. Step 4 allows to search for class instances which are totally possible in network NP^{S_3}. Once the network is simplified and re-normalized, it is immediate that each class instance c_k having the utmost prior possibility degree in network NP^{S_3} (c_k=$argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$, is totally possible given the uncertain inputs to classify. In particular, if $\pi^{NP_{S_3}}(c_k)$=1, then we can assert that there exists an attribute configuration $a_1..a_n$ allowing c_k to be totally plausible. This result is formalized in the following proposition:

Proposition 4 Let NP^{S_3} be the naive possibilistic network obtained after Steps 1, 2 and 3. Then, $\forall c_k \in D_C^{\Pi G_{S_3}}$ such that $\pi^{\Pi G_{S_3}}(c_k)$=1 or c_k=$argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$, then $max_{A_1..A_n}(\pi^{\Pi G_{S_3}}(c_k|a_1..a_n))=1$

The attribute configuration guaranteeing that c_k is totally possible is the one where $\pi^{\Pi G_{S_3}}(a_i|c_k)$=1 for i=1..n.

Example (continued): Proposition 4 allows to assert that in network ΠG^{S_3} of Figure 7, class instance c_2 is totally possible since c_2=$argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$. The attribute configuration allowing this result is $a_{13}a_{22}$. One can easily check that $\Pi^{\Pi G_{S_3}}(c_2|a_{13}a_{22})$=1.

4.5 Step 5: Conditionally totally possible class lookup

We assume in this step that the class variable C is a binary variable (namely, $D_C^{\Pi G_{S_3}}$={c_1, c_2}) and class label c_2 is totally possible ($\pi^{\Pi G_{S_3}}(c_2)>\pi^{\Pi G_{S_3}}(c_1)$).

In Step 4, only a subset of plausible class instances are found. Indeed, other class instances c_k having $\pi^{\Pi G_{S_3}}(c_k)=\alpha$ ($0<\alpha<1$) and $c_k \neq argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$ can be totally possible given the uncertain inputs to classify. Namely, it may exist an attribute configuration $a_1 a_2..a_n$ such that $\Pi^{\Pi G_{S_3}}(c_k|a_1 a_2..a_n)$=1 even if $\pi^{\Pi G_{S_3}}(c_k)<1$ and $c_k \neq argmax_{c_j \in D_C^{NP_{S_3}}}(\pi^{NP_{S_3}}(c_j))$. Hence, there is a need to check if every class instance c_k which is not found totaly possible in Step 4 can be conditionally totally possible without exploring all configurations of $\Pi^{\Pi G_{S_3}}(c_k|a_1 a_2..a_n)$. If class instance c_1 is totally possible, then this implies

ISBN: 978-989-95079-6-8

that there exists an attribute configuration $a_1..a_n$ where $\Pi^{\Pi G_{S_3}}(c_1|a_1a_2..a_n)$=1. In order to find such a configuration, we can use the following decomposition:
$\Pi^{\Pi G_{S_3}}(c_1|a_1..a_n)$=1 if and only if
$\pi^{\Pi G_{S_3}}(c_1a_1..a_n)$=$\Pi^{\Pi G_{S_3}}(a_1..a_n)$. Recall that
$\Pi^{\Pi G_{S_3}}(a_1..a_n)$=$max(\pi^{\Pi G_{S_3}}(c_1a_1..a_n), \pi^{\Pi G_{S_3}}(c_2a_1..a_n))$
Hence, $\pi^{\Pi G_{S_3}}(c_1a_1..a_n)\geq\pi^{\Pi G_{S_3}}(c_2a_1..a_n)$. Then
$\frac{min(\pi(c_1),\pi(a_1|c_1),..,\pi(a_n|c_1)}{min(\pi(c_2),\pi(a_1|c_2),..,\pi(a_n|c_2)}\geq 1$.
Let us define $\pi_{c_1}(a_i/c_1)$=1 if
$\pi(a_i|c_1)\geq\pi(a_i|c_2)$ and $\pi_{c_1}(a_i/c_1)$=$\pi(a_i|c_2)$ otherwise. Then
$\frac{min(\pi(c_1),\pi(a_1|c_1),..,\pi(a_n|c_1)}{min(\pi(c_2),\pi(a_1|c_2),..,\pi(a_n|c_2)}\geq 1$ implies that

$$\pi(c_1) \leq min(\pi_{c_1}(a_1|c_1),..,\pi_{c_1}(a_n|c_1)). \qquad (13)$$

Decomposition of Equation 13 can be used to build a new network ΠG^{S_5} by transforming conditional possibility distributions relative to each attribute A_i in ΠG^{S_3} by substituting every term $\pi^{\Pi G_{S_3}}(a_i|c_1)$ by 1 if $\pi^{\pi G_{S_5}}(a_i|c_1)\geq\pi^{\Pi G_{S_3}}(a_i|c_2))$ and by $\pi^{\Pi G_{S_5}}(a_i|c_1)$ otherwise and discarding c_2 since it is known to be totally possible. After this transformation, some local conditional possibility distributions in network ΠG may not be normalized. Repeating Step 2 (Re-normalization) on sub-normalized local distributions in network ΠG^{S_5}) allows to re-normalize them. Once re-normalization accomplished, the new possibility distribution $\pi_C^{\Pi G_{S_5}}$ relative to class node shows whether class instance c_1 is totally possible. Then we have the following proposition:

Proposition 5 Let ΠG^{S_3} be the naive possibilistic network obtained after Steps 1, 2 and 3, and let $D_C^{\Pi G_{S_3}}$=$\{c_1, c_2\}$, $\pi^{\Pi G_{S_3}}(c_1)$<$\pi^{\Pi G_{S_3}}(c_2)$. Let ΠG be the possibilistic network having same structure as ΠG^{S_3} where $D_C^{\Pi G^{S_5}}$=$\{c_1\}$. Let for i=1..n $\pi^{\Pi G^{S_5}}(a_i|c_2)$=1 if $\pi^{\Pi G_{S_3}}(a_i|c_1)\geq\pi^{\Pi G_{S_3}}(a_i|c_2))$ and $\pi^{\Pi G^{S_5}}(a_i|c_2)$= $\pi^{\Pi G^{S_3}}(a_i|c_2)$ otherwise. Then, $\pi^{\Pi G^{S_5}}(c_1)$=1 if and only if there exists an attribute configuration $a_1..a_n$ such that $\Pi^{\Pi G_{S_3}}(c_1|a_1..a_n)$=1.

It is important to note that in comparison with the lookup for conditionally totally possible class instances in product-based networks (see [11]), this step shows significant differences in min-based networks.

Example (continued): Transformation of Step 5 on network ΠG^{S_3} and its re-normalization gives network ΠG^{S_5} of Figure 8 where network of the left side gives the network obtained after Step 5 before re-normalization while network of right side represents ΠG^{S_5} after re-normalization according to Step 3. After this transformation, we can assert that class instance c_1 is not totally possible ($\pi^{\Pi G^{S_5}}(c_1) < 1$). This result confirms the one obtained by directly applying the possibilistic counterpart of Jeffrey's rule (see Figure 4).

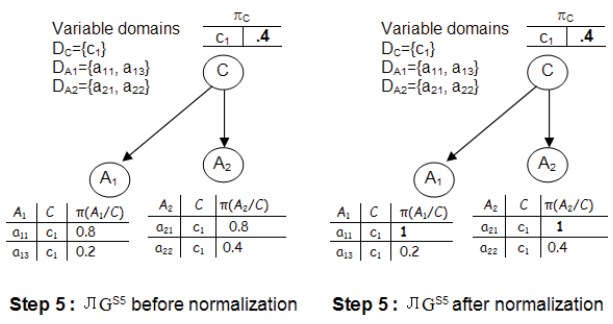


Figure 8: ΠG^{S_5}: Possibilistic network after Step 5

5 Conclusion

This paper dealt with min-based possibilistic network classifiers under uncertain inputs. It first addressed the min-based possibilistic counterpart of Jeffrey's rule for revising possibilistic knowledge encoded by a qualitative naive possibilistic network. After showing that classification based on the min-based revision of a possibility distribution encoded by a qualitative possibilistic classifier is exponential in the number of uncertain inputs, we proposed a polynomial algorithm for revising a naive min-based possibilistic network given uncertain inputs. This algorithm applies a series of equivalent and polynomial transformations on the initial network taking into account the uncertain evidence to classify. In future works, we will address general possibilistic classifiers under uncertain inputs.

References

[1] F. V. Jensen. *An Introduction to Bayesian Networks*. UCL press, London, 1996.

[2] A. Mittal. *Bayesian Network Technologies: Applications and Graphical Models*. IGI Publishing, USA, 2007.

[3] S. Benferhat, D. Dubois, L Garcia, and H. Prade. Possibilistic logic bases and possibilistic graphs. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence UAI 99*, pages 57–64, Stockholm, 1999.

[4] C. Borgelt and R. Kruse. *Graphical Models: Methods for Data Analysis and Mining*. John Wiley & Sons, Inc., USA, 2002.

[5] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(3), 1978.

[6] D. Dubois and H. Prade. *Possibility theory*. Plenum Press, New-York, 1988.

[7] B.a Haouari, Z. Elouedi N. Ben-Amor, and K. Mellouli. Naive possibilistic network classifier. In *Proceedings of The Thirteenth Congress of International Association for Fuzzy-Set Management and Economy, SIGEF 2006*, 2006.

[8] T. Denoeux and L. M. Zouhal. Handling possibilistic labels in pattern classification using evidential reasoning. *Fuzzy Sets and Systems*, 122(3):409 – 424, 2001.

[9] H. Eyke. Possibilistic induction in decision-tree learning. In *ECML '02: Proceedings of the 13th European Conference on Machine Learning*, pages 173–184, London, UK, 2002.

[10] I. Jenhani, N. Benamor, and Z. Elouedi. Decision trees as possibilistic classifiers. *Int. J. Approx. Reasoning*, 48(3):784–807, 2008.

[11] S. Benferhat and K. Tabia. An efficient algorithm for naive possibilistic classifiers with uncertain inputs. In *The Second International Conference on Scalable Uncertainty Management(SUM'08)*, Naples, Italy, oct 2008. Springer LNAI.

[12] R. C. Jeffrey. *The Logic of Decision*. McGraw Hill, NY, 1965.

[13] E. Hisdal. Conditional possibilities independence and non interaction. *Fuzzy Sets and Systems*, pages 283–297, 1978.

[14] D. Dubois and H. Prade. A synthetic view of belief revision with uncertain inputs in the framework of possibility theory. *Int. J. of Approximate Reasoning*, 17(2–3):295–324, aot 1997.

ISBN: 978-989-95079-6-8

Parallel Coordinates-based Fuzzy Color Distance and its Applications

Constantin Vertan

Image Analysis and Processing Laboratory, Politehnica University of Bucharest
Bucharest, Romania
Email: cvertan@alpha.imag.pub.ro

Abstract— *Although simple and based on physical realizable color primaries, the RGB color space cannot allow the direct definition of a topology-preserving, perceptually-compatible inter-color distance. The classical normalized inter-color distances are rather complex to implement and tune and do not account for the natural incertitude and confusion regarding both the numerical color values and the color perception. We propose the use of a planar color representation plot, derived from multivariate data visualization, which allows the introduction of a luminance-invariant, geometry-based fuzzy inter-color distance, used with good results in the implementation of distance-based color filters and edge detectors.*

Keywords— fuzzy distance, fuzzy color, fuzzy image filtering.

1 Introduction

The classical inter-color distances are based on standard colorimetric representations, being metrics (or almost metrics) in the color gamut space (a subset of $\mathbf{R}^3$). We propose to investigate a new approach, inspired by the reduced ordering principle of Barnett [1] and also used in multivariate data visualization [2]: namely, the multivariate data (colors) are mapped to some familiar, two-dimensional objects, that can be grouped, compared, and plotted with more ease and are more suited for human perception. Examples of such mappings are the Chernoff faces [3], the Andrews curves [4] and their possible extensions [5], the basic and modified parallel coordinates [6] and the star glyphs [2]. Previous work showed that we could extend the geometric distance between simplified, plane color representations (such as the star glyphs) to color distances [7].

This work proposes the embedding of the uncertainty regarding the exact values of the color components describing the color into a fuzzy color distance, inspired by the two-dimensional geometrical representation of the color by means of the parallel coordinates.

The remainder of the paper is organized as follows: section 2 describes the basic parallel coordinates representation of multivariate data and its application to the representation of colors, section 3 introduces the distance between colors based on their parallel coordinates representation, section 4 introduces the fuzzy extension of the inter-color distance and, finally, section 5 presents some applications of the proposed fuzzy distance in non-linear color image filtering and color edge extraction.

2 Parallel coordinates: the basics

The parallel coordinates representation is a visualization technique that basically allows plotting n-dimensional points and patterns into the bi-dimensional plane. Thus, the parallel coordinates representation transforms multidimensional problems into two-dimensional patterns, without loss of information.

Let us consider a n-dimensional data point $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, originally represented in a Cartesian coordinate system. If each of the n axes of the coordinate system are lined up in parallel with all others into the plane, separated by a fixed distance Δ, we obtain the parallel coordinates system (as presented in Fig. 1) [6]. Obviously, the approach is scalable with respect to the data dimension, n. Visualization is facilitated by viewing the two-dimensional representation of the n-dimensional data points as lines, crossing the n parallel axes, each of them representing one dimension of the original feature space. A n-dimensional point $\mathbf{x}$ is represented in the plane by the series of $n-1$ connected line segments defined by the end points $[(kD, x_1), ((k+1)D, x_2)]$, $[((k+1)D, x_2), ((k+2)D, x_3)]$, , $[((k+n-2)D, x_{n-1}), ((k+n-1)D, x_n)]$ (as shown in Fig. 1).

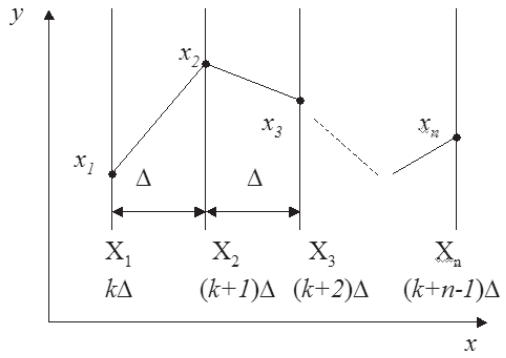


Figure 1: Basic parallel coordinates representation of a n-dimensional data point.

Several interesting mathematical results have been proven for the parallel coordinates representation [6], [8] and several applications were proposed, such as clustering [9], fuzzy rule representation [10], air traffic control [8], and color image filtering [11] based on the independent processing of the points along the individual coordinates axes.

We may notice the strong resemblance of the parallel coordinate representation with the Andrews curves [4], [2] representation method. The Andrews curve representation maps each n-dimensional point into a continuous, analytical curve, obtained as a n-term series expansion having the point coordinates as coefficients, according to a fixed functional basis. The original method uses the Fourier expansion (sine and cosines functions), but modifications were also proposed (e.g. the use of Haar wavelets for color image filtering [5]). For a discrete representation of the Andrews curves, the continuous basis functions are sampled at regular intervals, similar to the

ISBN: 978-989-95079-6-8

use of some parallel axes.

In the following of the paper, we will focus on the representation of colors - three-dimensional data points. Thus, each data point is a triple, $\mathbf{x} = (x_1, x_2, x_3)$. This representation is straightforward for the representation of RGB color data, due to the similar nature of the RGB components; thus, the data points will be $\mathbf{x} = (R, G, B)$, as proposed in [11]. Still, we propose to use a slight modification of the classical model, by using a four-dimensional data vector for the representation of the color, namely $\mathbf{x} = (R, G, B, R)$ as shown in Fig. 2. A color will be thus represented by the polygonal line or by the subgraph (lower limited by the horizontal axis) of that polygonal line, which is itself a polygon and will be subsequently named color polygon.

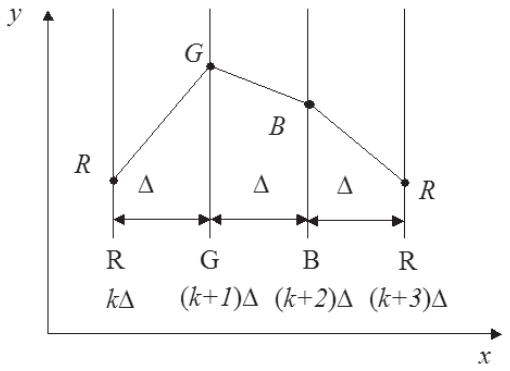


Figure 2: Proposed parallel coordinates representation of a RGB color, viewed as a 4-dimensional data vector.

3 Computing color distances

Color is represented by a three-component vector; the basic representation is the RGB model, for which the vector components are the relative amounts of normalized red, green and blue that additively mix in order to produce that color. Although simple and based on physical realizable color primaries, the RGB color space cannot allow the direct definition of a topology-preserving, perceptually-compatible inter-color distance. The same observation holds for the spectral color representations as well. As normalized by CIE (Commission Internationale de l'Eclairage), perceptual-compatible inter-color distances are obtainable from the Lab color representation. For any two colors $C_1 = (L_1, a_1, b_1)$ and $C_2 = (L_2, a_2, b_2)$, the simplest inter-color distance is the Euclidean distance in the Lab color gamut [12].

Advanced color difference formulas (such as the CMC, BFD or CIE94 [12]) have been introduced, since the basic Lab Euclidean metric do not accurately quantify small-to-medium-sized color differences. Such correctly-measured (as correlated to the subjective estimation) color difference are very complex and untractable for general-use in color image processing. Furthermore, they are not invariant to the change or luminance.

Obviously, the parallel coordinates polygon's shape is related to the nature of the color (properties like hue, saturation, colorfulness, etc.) and the overall area of the polygon relates to the luminance or brightness of the color. It can be accepted that a similarity measure between colors is given by the area of intersection of the two associated parallel coordinates polygons. In order to construct a symmetrical and normalized measure, we will follow an approach similar to the definition of the Canberra distance: the normalization is performed with respect to the average area of the two color polygons. Thus, the distance between colors C_1 and C_2, represented by their corresponding parallel coordinates polygons P_1 and P_2 can be introduced as:

$$d(C_1, C_2) = 2\frac{Area_{P_1 \nabla P_2}}{Area_{P_2} + Area_{P_2}} \tag{1}$$

From simple plane geometry considerations one can easily show that the distance in (1) can be approximately put in the form of:

$$d(C_1, C_2) = \left(D_{RG} + D_{GB} + D_{BR}\right)/3 \tag{2}$$

where D_{RG} is given in (3) and D_{GB} and D_{BR} have similar forms.

$$D_{RG} = 2\begin{cases} \frac{|R_1-R_2|+|G_1-G_2|}{R_1+R_2+G_1+G_2} & if \Delta R \Delta G \geq 0 \\ \frac{(R_1-R_2)^2+(G_1-G_2)^2}{(R_1+R_2)^2-(G_1+G_2)^2} & if \Delta R \Delta G < 0 \end{cases} \tag{3}$$

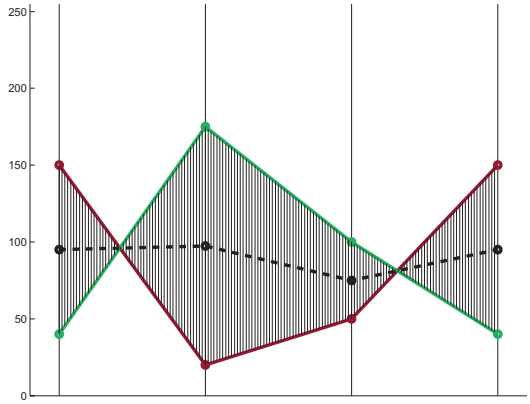


Figure 3: Proposed parallel coordinates distance for two colors (red with $RGB = (150, 25, 50)$ and green with $RGB = (40, 170, 110)$): the hashed area is the uncommon area of the parallel coordinates polygons associated to the colors; the dashed central line is the parallel coordinates polygon having the average area.

It can be easily shown that for unsaturated colors (grays), the proposed inter-color distance is the Canberra distance. The distance in (2) is normalized in the $[0; 1]$ range. It can be easily shown that the proposed distance is invariant to luminance changes, that is $d(C_1, C_2) = d(\alpha C_1, \alpha C_2), \forall \alpha \in \mathbf{R}$.

4 The fuzzy color distance

All image processing algorithms must deal with the imprecision and vagueness that naturally arise in the digital representation of visual information. Noise, quantization and sampling errors, the tolerance of the human visual system are the cause of this imprecision. This strongly suggests that fuzzy models may be used for taking them into account.

In the case of color images, color attributes and color differences play a particularly important role in the perception of

ISBN: 978-989-95079-6-8

object boundaries. The process of measuring color differences must be designed to maintain a balance between the computed and the perceived difference. Still, the simple use of any given color representation does not account for the similarity perception and the visual confusion of colors. We propose to deal with this factors in the framework of a fuzzy representation of the basic RGB color components and obtain a fuzzy distance between colors based on the parallel coordinate representation of the fuzzy colors.

The assumed model means that we will consider each of the R, G, and B color components as fuzzy numbers, centered at their ideal (correct, crisp) value. Thus, the polygonal line representing the given color in the parallel coordinates system will evolve into some fuzzy polygonal line, composed by fuzzy line segments.

We will define the fuzzy distance $\widetilde{D}$ between two fuzzy colors by means of its α-cuts. The corresponding α-cuts of any fuzzy color component are closed intervals along the corresponding parallel coordinate axes. Under these circumstances we can easily determine (from basic geometrical considerations) the minimal and maximal common areas of the corresponding color parallel coordinates polygon and thus to determine the lower and upper bounds of the corresponding α-cut of the fuzzy distance $\widetilde{D}$.

In the particular case of modelling the fuzzy RGB components by fuzzy triangular numbers, the limits of the α-cut intervals of the fuzzy distance are particulary simple to obtain. We will assume that the fuzzy color components $\widetilde{V}_i$ (where V_i is any of the RGB components) are defined by triangular fuzzy membership functions around their crisp value V_{i0}:

$$\mu_{\widetilde{V}_i}(V) = max(0, 1 - \frac{|V - V_{i0}|}{2 * \delta}) \tag{4}$$

In equation (4) above, δ is the support of the membership function, measuring its width around the crisp (correct) value and being thus the direct measure of the incertitude regarding the correct component value V_{i0}.

The α-cuts of any of the terms that sum to the color distance defined in (2) are bounded intervals, like $[D^-_{XY}(\alpha); D^-_{XY}(\alpha)]$. These upper and lower bounds of the α-cuts intervals of color distance terms are given by the expressions in (5) and (6) (for the D_{RG} term given in (3)), according to the cases when the line segments within two successive parallel coordinate axes intersect (corresponding to $\Delta R \Delta G < 0$) or not ($\Delta R \Delta G \geq 0$).

$$D^{\pm}_{RG}(\alpha) = 2\frac{|R_1 - R_2| + |G_1 - G_2| \pm 4(1-\alpha)\delta}{R_1 + R_2 + G_1 + G_2} \tag{5}$$

$$\begin{aligned} P_1 &= (R_1 + R_2 + G_1 + G_2 \mp 4(1-\alpha)\delta) \\ P_2 &= (|R_1 - R_2| + |G_1 - G_2| \mp 4(1-\alpha)\delta) \\ U_1 &= (R_1 - R_2 \pm 2(1-\alpha)\delta)^2 \\ U_2 &= (G_1 - G_2 \pm 2(1-\alpha)\delta)^2 \\ D^{\pm}_{RG}(\alpha) &= 2\frac{U_1 + U_2}{P_1 P_2} \end{aligned} \tag{6}$$

The fuzzy color distance is obtained by its α-cuts, by summing the α-cuts of the corresponding D_{XY} terms in (2). The choice of the width δ of the triangular membership function that models the RGB color components and the discrete nature of the component values imply that the number of distinct α-cuts of the color distance is limited to the rounded value of $\delta + 1$. For instance, the fuzzy distance between the two colors represented in parallel coordinates in figure 3 (red with $RGB = (150, 25, 50)$ and green with $RGB = (40, 170, 110)$) computed according to $\delta = 2.2$ is defined by three α-cuts: 2.53 for $\alpha = 1$, $[2.49; 2.58]$ for $\alpha \in [0.55; 1)$ and $[2.44; 2.62]$ for $\alpha \in [0.09; 0.55)$.

5 Applications

A significant part of nonlinear filters for color images are distance-based, i.e. they rely on the computation of inter-color distances between the colors selected by the filtering window. We shall prove the use of the proposed fuzzy inter-color distance in both smoothing filters (such as the vector median filter) and edge extraction filters, such as standard derivative-based filters.

5.1 Vector median filtering using the fuzzy inter-color distance

The classical VMF (Vector Median Filter) [13] defines the vector (or color) median as the vector characterized by a minimal aggregated distance with respect to all other vectors in the filtering neighborhood; this ordering of the color vectors is an instance of the reduced ordering principle [1]. Usually, the distance between color is computed as a L^k norm of the RGB color vectors. We will show here that the use of the proposed fuzzy inter-color distance provides good filtering results.

The fuzzy aggregate distance associate to each color vector within the filtering window is computed based on the α-cuts of the individual inter-color distances. Finally, the aggregated distances (which are also fuzzy numbers defined by their α-cuts) are ranked, and the minimal fuzzy aggregated distance corresponds to the median color.

The ranking of the fuzzy numbers (the fuzzy aggregated distances) is performed according to the classical total integral value. The total integral value of a fuzzy number $\widetilde{D}$ was introduced in [14] as:

$$I_T(\widetilde{D}) = \int_0^1 \mu^{-1}_{\widetilde{D}}(y)dy \tag{7}$$

According to the ranking proposed in [14], the smallest fuzzy number has the smallest total integral value (we may also notice that this ranking is an instance of Barnett's reduced ordering principle [1]).

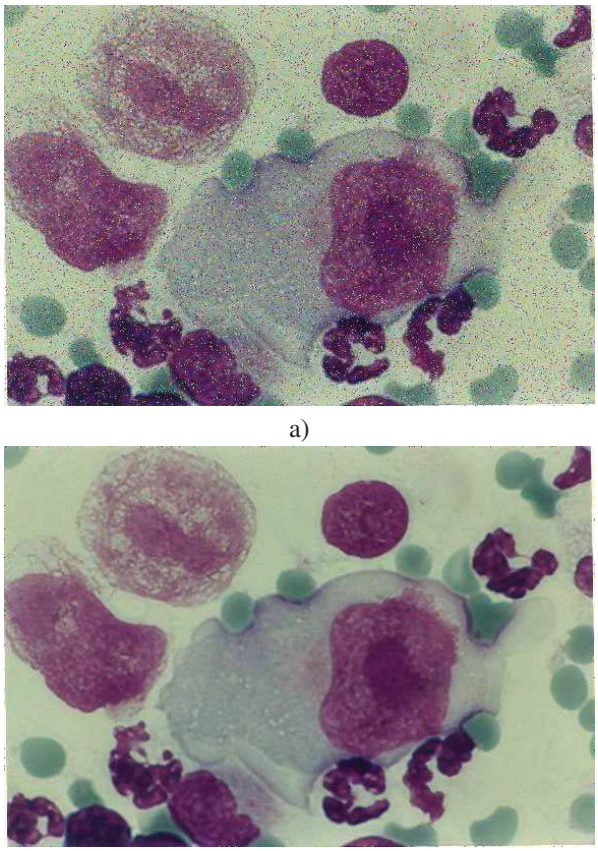
Figure 4 presents the result of a fuzzy-VMF filter applied for impulsive noise reduction in a color image.

5.2 Fuzzy color edge detection

Basically, all edge detection operators rely on the computation of an edge intensity map, which is further thresholded in order to obtain a binary edge map. The edge intensity map exhibits important values for the pixels that are on the boundaries of uniform regions, characterized by a discontinuity (or variation) of their colors.

One of the simplest contour extraction methods is the image Laplacian. We shall use the luminance-invariant fuzzy inter-color distance introduced in (2) for the implementation

ISBN: 978-989-95079-6-8



a) b)

Figure 4: a) 5% Impulsive noise degraded image and b) fuzzy-VMF filtered image within a 3×3 window, with a fuzzy color model characterized by $\delta = 2.2$.

of a fuzzy color Laplacian operator L. The proposed operator is a modification of the classical (derivative-type) V4-neighborhood Laplacian operator; the color Laplacian is the average inter-color distance within the color at the current processed location (i, j) and its immediate neighboring colors from the color image f. Mathematically we can express the proposed fuzzy Laplacian at location (i, j) as:

$$\widetilde{L}(i,j) = \frac{1}{4} \sum_{(k,l) \in V_4} \widetilde{D}(f(i+k, j+l), f(i,j)) \qquad (8)$$

The fuzzy Laplacian will be defined by the use of fuzzy color distances in (8). The summation of the fuzzy color distance is performed via the corresponding summation of the limits of their α-cut intervals, yielding an α-cut definition of the fuzzy color Laplacian at each image location.

The extraction of the binary edge map image implies the thresholding of the fuzzy Laplacian edge intensity map. According to the desired level of detail that we want to extract form the image, we may use as color edge intensity map either the lower or the upper limit of a specified α-cut interval, as shown in figures 5 and 6. The choice of a particular value of α acts as trimming parameter for the extraction of a more accurate and detailed or a more rough binary edge map, while keeping a same fixed general threshold. This is not possible when using a non-fuzzy color distance, either in RGB or Lab or other color representations.

6 Conclusions

In this contribution we presented a new fuzzy inter-color distance measure, derived from geometrical considerations related to a plane, two-dimensional, reversible color representation. This color representation by parallel coordinates (representing a color as a an open polygonal line) was primarily used in multivariate data representation. The proposed fuzzy inter-color distance can be used with good results in the median (non-linear) filtering of color images and fuzzy edge detection.

Acknowledgment

This work was supported in part by the Romanian National Agency for Scientific Research under the PNCDI2 51-021.2/2007 and 12-136/P2/2008 research grants.

References

[1] V. Barnett. The ordering of multivariate data. *J. of. Royal Stat. Soc. A*, 139(3):318–354, 1976.

[2] W. J. Krzanowski. *Principles of Multivariate Analysis: A User's Perspective*. Clarendon Press, Oxford, 1993.

[3] H. Chernoff. The use of faces to represent points in k- dimensional space graphically. *J. of the American Statistical Association*, 68:125–136, 1973.

[4] D. F. Andrews. Plots of high-dimensional data. *Biometrics*, 28:125–136, 1972.

[5] C. Vertan, M. Ciuc, and V. Buzuloiu. Multichannel median filtering by wavelet series expansion. In *Proc. of the 6th International Workshop on Systems, Signals and Image Processing IWSSIP 1999*, pages 77–80, Bratislava, Slovakia, 2-4 Jun. 1999.

[6] Inselberg A. and Dimsdale B. Multidimensional lines I: representation. *SIAM J. Appl. Math.*, 54(2):559–577, 1994.

[7] C. Vertan, M. Zamfir, E. Zaharescu, V. Buzuloiu, and C. Fernandez-Maloigne. Non-linear color image filtering by color to planare shape mapping. In *Proc. of ICIP 2003*, volume 1, pages 808–811, Barcelona, Spain, 10-13 Sept. 2003.

[8] Inselberg A. and Dimsdale B. Multidimensional lines II: proximity and applications. *SIAM J. Appl. Math.*, 54(2):578–596, 1994.

[9] S. Y. Chou, S. W. Lin, and C. S. Yeh. Cluster identification with parallel coordinates. *Pattern Recognition Letters*, 20:562–572, 1999.

[10] L. O. Hall and M. R. Berthold. Fuzzy parallel coordinates. In *Proc. of NAFIPS 2000*, pages 74–78, Atlanta GA, USA, 13-15 Jul. 2000.

[11] C. Vertan, M. Ciuc, and V. Buzuloiu. Parallel coordinates for color image processing. In *Proc. of International Conference on Signals, Circuits and Systems SCS 2001*, pages 469–472, Iaşi, România, 10-11 Jul. 2001.

[12] M. R. Luo. Colour science. In S. J. Sangwine and R. E. N. Horne, editors, *The Colour image processing handbook*, pages 26–66, London, UK, 1998. Chapman and Hall.

[13] J. Astola, P. Haavisto, and Y. Neuvo. Vector median filters. *Proc. IEEE*, 78(4):678–689, Apr. 1990.

[14] T. S. Lion and M. J. Wang. Ranking fuzzy numbers with integral value. *Fuzzy Sets and Systems*, 50:247–255, 1992.

ISBN: 978-989-95079-6-8

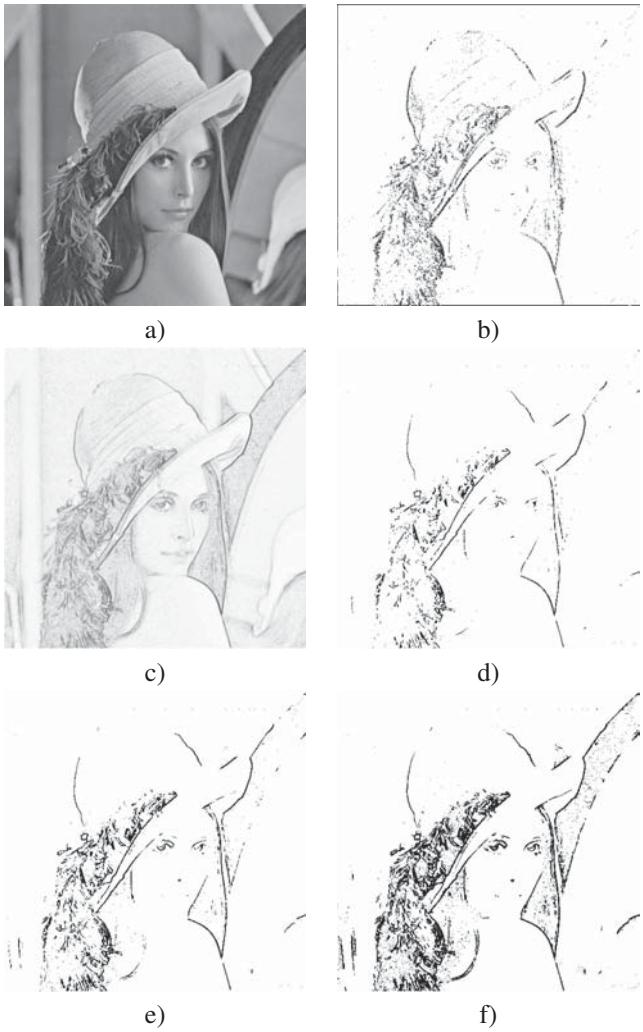


Figure 5: Fuzzy Laplacian edge extraction according to the proposed fuzzy color distance: a) Original color image; b) Classical Laplacian binary edge map; c) Fuzzy color Laplacian edge strength map computed according to (8) and (2) for $\delta = 3.2$ corresponding to an α-cut at α=1; d) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 0.66$ and thresholding of the lower bound of the α-cuts interval; e) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 1$; f) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 0.66$ and thresholding of the upper bound of the α-cut interval. The fuzzy color model is defined by $\delta = 3.2$.

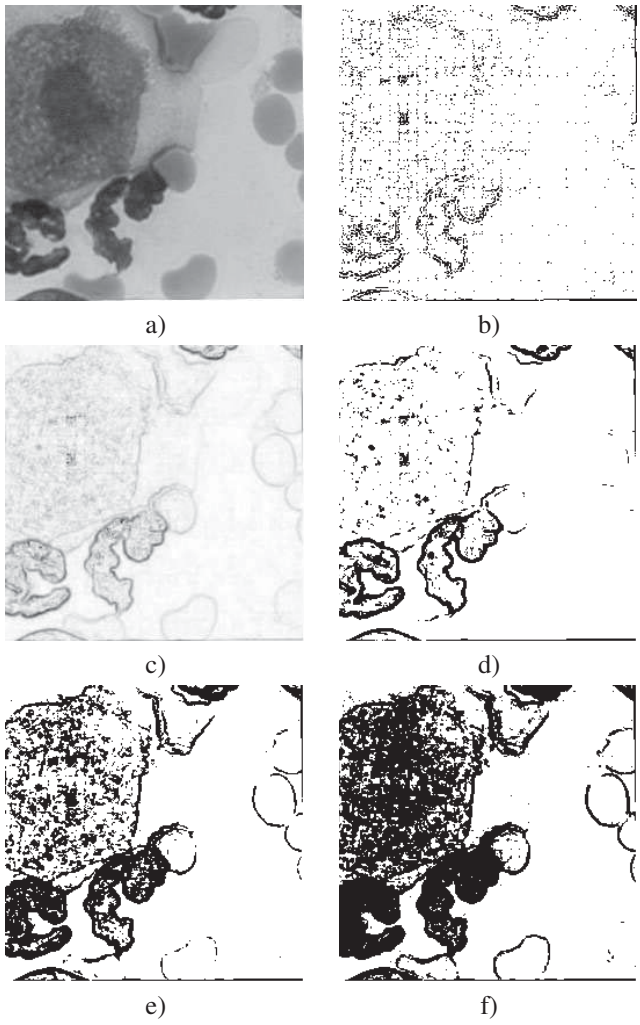


Figure 6: Fuzzy Laplacian edge extraction according to the proposed fuzzy color distance: a) Original color image; b) Classical Laplacian binary edge map; c) Fuzzy color Laplacian edge strength map computed according to (8) and (2) for $\delta = 3.2$ corresponding to an α-cut at α=1; d) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 0.66$ and thresholding of the lower bound of the α-cuts interval; e) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 1$; f) Fuzzy color Laplacian binary edge map computed from c) according to an α-cut at $\alpha = 0.66$ and thresholding of the upper bound of the α-cut interval. The fuzzy color model is defined by $\delta = 3.2$.

ISBN: 978-989-95079-6-8

Definition of fuzzy Pareto-optimality by using possibility theory

Ricardo C. Silva[1] Akebo Yamakami[1]

1. Department of Telematics, School of Electrical and Computer Engineering,
University of Campinas, P.O.Box 6101, 13083-970, Campinas-SP, Brazil.
Email: {rcoelhos,akebo}@dt.fee.unicamp.br

Abstract— *Pareto-optimality conditions are crucial when dealing with classic multi-objective optimization problems because we need to find out a set of optimal solutions rather than only one optimal solution to optimization problem with a single objective. Extensions of these conditions to the fuzzy domain have been discussed and addressed in recent literature. This work presents a novel approach based on the use of possibility theory as a comparison index to define a fuzzily ordered set with a view to generating the necessary conditions for the Pareto-optimality of candidate solutions in the fuzzy domain. Making use of the conditions generated, one can characterize fuzzy efficient solutions by means of carefully chosen single-objective problems. The uncertainties are inserted into the formulation of the studied fuzzy multi-objective optimization problem by means of fuzzy coefficients in the objective function. Some numerical examples are analytically solved to illustrate the efficiency of the proposed approach.*

Keywords— Possibility theory, multi-objective optimization, fuzzy Pareto-optimality conditions, fuzzy mathematical programming.

1 Introduction

One of the most significant characteristics of human beings is the decision making of day-by-day problems. This characteristics is used to solve several practical problems including economic, environmental, social and technical. These problems are multidimensional and have multiple objectives that are often non-commensurable and conflict with each other. Thus, they are inserted in the set of problems that are solved by using the theory of multi-objective optimization which is a generalization of traditional single objective optimization. Although multi-objective optimization problems differ from single objective problems only in the plurality of objective functions, it is important to realize that the notion of optimality condition change because now the decision maker must find solutions that satisfy or create a compromise among the multiple objectives. These solutions are called Pareto optimal or efficient or non-dominated solutions.

Optimization is a procedure of finding and comparing feasible solutions until no better solution can be found. These solutions are defined good or bad in terms of one or several objectives when is used any optimization model. The optimization models often use classical mathematical programming, which attempts to develop an exact model to the optimization problem of interest. Such modeling may overlook ambiguities that all too frequently exist in actual optimization operations. In recent years, Fuzzy Logic [16] has showed great potential for modeling systems which are non-linear, complex, ill-defined and not well understood. Fuzzy Logic has found numerous applications due to its ease of implementation, flexibility, tolerant nature to imprecise data, and ability to model non-linear behavior of arbitrary complexity because of its basis in terms of natural language. In [18] is discussed the use of fuzzy logic that is a precise logic of imprecision and approximate reasoning. We refer to [6, 10, 19], for some applications in the fields of pattern recognition, data analysis, optimal control, economics and operational research, among others.

The representation and arithmetic manipulation of uncertain numerical quantities can be defined by means of fuzzy sets. Unfortunately, the comparison among two or more fuzzy numbers, intervals and/or sets is not easy. Some approaches to compare them (see some examples in [2, 6, 9, 10, 14]) were developed, each one being based upon a different point of view. The possibility theory, which is analogous to the probability theory, was proposed by Zadeh [17] to aggregate the concept of a possibility distribution to the theory of fuzzy sets. Comparison indexes to rank fuzzy numbers and intervals employing possibility theory were proposed in [7]. Their importance stems from the fact that much of the information on which human decisions rely upon have a possibilistic, rather than a probabilistic nature. On the other hand, some works describe a fuzzy optimization problem in a classical problem and they use the classical theory to find the Pareto optimal set. [8] transform a fuzzy single objective problem into a classical multiple objective one where the number of objectives is defined by fuzzy coefficients from the fuzzy problem.

Possibility theory emerged from the notion of fuzzy sets and his concept tries to take account of the fact that an object may more or less correspond to a certain category in which one attempts to place it. In the calculus of degree of possibility emphasizes the double relationship between possibility theory and set theory and the concept of measure, respectively. One merit of possibility theory is to represent imprecision and to quantify uncertainty at the same time.

This work is organized as follows. Section 2 presents an overview about the formulation of classical multi-objective programming problems and classical concepts to obtain the set of Pareto optimal solutions. Also, it is shown an extension of these concepts to Pareto optimal solutions of fuzzy multi-objective programming problems. Section 3 introduces a novel approach to fuzzy Pareto-optimality. A fuzzily ordered set is defined by using a possibility distribution function as a comparison measure. This section also presents the characterization of the efficient solutions through the use of well defined scalar problems. This relation between efficient solutions and scalar problems can be determined by certain theorems. To clarify the above developments, two numerical

ISBN: 978-989-95079-6-8

examples are analyzed in section 4. Finally, conclusions are presented in Section 5.

2 Multiobjective programming problem formulation and concepts

Choosing the goal to be optimized is a critical step in the process of modeling real-world problems. The local or global optimal solution depends totally upon this choice. In the vast majority of real-world problems, various objective functions could be defined. Many times these functions are conflicting and/or non-measurable. Multi-objective optimization is the branch of mathematical optimization theory devoted to developing methods to solve problems with various objective functions. A classical multi-objective problem can be formulated as follows:

$$\begin{array}{ll} \min & F(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in \Omega \end{array} \tag{1}$$

where $F = (f_1, f_2, \ldots, f_m),\ (m \geq 2)$ is a vector of objectives and $\Omega \subset \mathbb{R}^n$ is the set of feasible solutions.

Due to the issue of conflicting objectives, the classical concept of optimality that are used to obtain the optimal solution to optimization problems with an objective function does not fit into the multi-objective framework. Hence, in such a framework, one settles for the so-called efficient, non-dominated or Pareto-optimal solution.

The classical works [12] and [13], by Vilfredo Pareto, introduced the concept of Pareto-optimality and started the field of multi-objective optimization. A solution $x^* \in \Omega$ is said to be *Pareto-optimal* or *non-dominated* if there exists no alternative solution in Ω that improves some of the objective functions without degrading at least another objective function. Then, we can define mathematically a non-dominated solution as

Definition 1 (Pareto optimal solution) $\mathbf{x}^* \in \Omega$ *is said to be a* non-dominated solution *of Problem (1) if there exists no other feasible* $\mathbf{x} \in \Omega$ *such that* $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*),\ \forall\ i = 1, \ldots, m$ *with strict inequality for at least one* i.

However, when non-linear programming problems with single objective are solved by any non-linear programming method, only local optimal solutions are guaranteed in practical. Then, the concept of local non-dominated solution can be defined in the following way:

Definition 2 (local Pareto optimal solution) $\mathbf{x}^* \in \Omega$ *is said to be a* local non-dominated solution *of Problem (1) if and only if there exists a real number* $\delta > 0$ *such that* $\mathbf{x}^*$ *is non-dominated in* $\Omega \bigcap \mathcal{N}(\mathbf{x}^*, \delta)$*, i.e., there does not exist another feasible* $\mathbf{x} \in \Omega \bigcap \mathcal{N}(\mathbf{x}^*, \delta)$ *such that* $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*),\ \forall\ i = 1, \ldots, m$ *with strict inequality for at least one* i.

where $\mathcal{N}(\mathbf{x}^*, \delta)$ denotes the δ neighbourhood of $\mathbf{x}^*$ defined by $\{\mathbf{x} \in \mathbb{R} \| \|\mathbf{x} - \mathbf{x}^*\| < \delta\}$. Thus, it is possible to see by these definitions that the solution of a multi-objective programming problem consists of an infinite number of points.

Many methods to solve multi-objective programming problems were proposed and some specific methods can be found in [4]. These methods are classified according to the instant the decision maker applies their criteria. Three methods are proposed: (i) *a-Priori method*, where the decision maker assigns weights to the objective functions a-priori, thus obtaining a single mono-objective criterion; (ii) *a-Posteriori method*, where the decision maker strives to create some type of efficient solutions to chose from a-posteriori; and (iii) *interactive methods*, where the decision maker informs their preferences during the search process of an efficient solution.

Although the mathematical formulation of optimization problems with multiple objectives be well defined, some values of the real-world problems have vagueness, imprecision and uncertainty. These values which have been estimated by decision maker are parameters in the set of constraints and in one or several objective functions. These uncertainties can be formulated by logic fuzzy which is a way to describe this vagueness mathematically and it has found numerous and different applications due to its easy implementation, flexibility, tolerant nature to imprecise data, low cost implementations and ability to model non-linear behavior of arbitrary complexity because of its basis in terms of natural language. The concept of fuzzy decision to obtain a solution to fuzzy programming problems was introduced by Bellman and Zadeh[1] which proved that the fuzzy programming problems can be reduced to a conventional programming problem under some assumptions determined by decision maker.

Based on this work, many researches developed methods that solve optimization problems under fuzzy environment and in [15] is used α-cut sets to define a non-dominated solution to multi-objective programming problems with fuzzy parameters which is called α-Pareto optimal solution. In this case, the fuzzy parameters can be inserted in Problem (1) and it is transformed in the following way:

$$\begin{array}{ll} \min & F(\tilde{\mathbf{a}}; \mathbf{x}) = (f_1(\tilde{\mathbf{a}}_1; \mathbf{x}), \ldots, f_m(\tilde{\mathbf{a}}_m; \mathbf{x})) \\ \text{s.t.} & \mathbf{x} \in X(\tilde{\mathbf{b}}) \triangleq \{x \in \mathbb{R}^n | g_i(\tilde{\mathbf{b}}_i; \mathbf{x}) \leq 0, i = 1, \ldots, p\} \end{array} \tag{2}$$

where $\tilde{\mathbf{a}}_i$ and $\tilde{\mathbf{b}}_i$ represent a vector of fuzzy parameters involved in the objective functions and in the functions that form the set of constraints, respectively. These fuzzy parameters which reflect the expert's ambiguous understanding of the nature of the parameters in the problem formulation process, are assumed to be characterized as fuzzy numbers.

Definition 3 (α-level set) *The* α*-level set of the fuzzy numbers* $\tilde{c}$ *and* $\tilde{d}$ *is defined as the ordinary set* $(\tilde{c}, \tilde{d})_\alpha$ *for which the degree of their membership functions exceeds the level* α*:*

$$(\tilde{c}, \tilde{d})_\alpha = \{(c, d) | \mu_{\tilde{c}}(c) \geq \alpha \text{ and } \mu_{\tilde{d}}(d) \geq \alpha\}.$$

Now, suppose that the decision maker considers that the degree of all of the membership functions of the fuzzy numbers involved in Problem (2) should be greater than or equal to a certain value of α. Then, for such a degree α, Problem (2) can be interpreted as a conventional multi-objective programming problems in the following way:

$$\begin{array}{ll} \min & F(\mathbf{a}; \mathbf{x}) = (f_1(\mathbf{a}_1; \mathbf{x}), \ldots, f_m(\mathbf{a}_m; \mathbf{x})) \\ \text{s.t.} & \mathbf{x} \in X(\mathbf{b}) \triangleq \{x \in \mathbb{R}^n | g_i(x, b_i) \leq 0, i = 1, \ldots, p\} \\ & (\mathbf{a}, \mathbf{b}) \in (\tilde{\mathbf{a}}, \tilde{\mathbf{b}})_\alpha \end{array} \tag{3}$$

where the coefficient vector $(\mathbf{a}, \mathbf{b}) \in (\tilde{\mathbf{a}}, \tilde{\mathbf{b}})_\alpha$ and $(\mathbf{a}, \mathbf{b})$ are arbitrary for any value in $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})_\alpha$. This fuzzy subset is formatted for all the vectors whose degree of each membership function exceeds the level α. Here, it is possible to observe that the parameters $(\mathbf{a}, \mathbf{b})$ are treated as decision variables of Problem (3) rather than constants of Problem (2). Through the

ISBN: 978-989-95079-6-8

description of Problem (3), the concepts of (local) α-Pareto optimality is defined as follows.

Definition 4 ((Local) α-Pareto optimal solution) *$x^* \in X(b)$ is said to be a (local) α-Pareto-optimal solution to the α-MONLP(Multi-Objective Non-Linear programming) if and only if there exists no other $x \in X(b)(\bigcap \mathcal{N}(x^*, r))$ and $(a, b) \in (A, B)_\alpha(\bigcap \mathcal{N}(a^*, b^*, r'))$ such that $f_i(x, a_i) \leq f_i(x^*, a_i^*)$, $i = 1, 2, \ldots, k$, with strict inequality holding for at least one i, where the corresponding values of parameters a^* and b^* are called α-level optimal parameters (and $\mathcal{N}(x^*, r) \triangleq \{x \in \mathbb{R}^n | \|x - x^*\| < r\}$ denote the r neighborhood of x^*).*

The (local) α-Pareto-optimal solutions can be obtained through a direct application of the scalarization methods, which transform a multi-objective programming problem into a single-objective programming problem, and the optimal solution obtained is formed by point $(\mathbf{x}, \mathbf{a}, \mathbf{b})$. These set of solutions, however, generally comprises an infinite number of points and the decision should select a single (local) solution based on a subjective criterion.

3 The use of possibility theory in multi-objective optimization under fuzzy environment

Normally, when a multi-objective optimization problems is formulated, many parameters need to be assigned by the decision maker and they may be described by possible values. In most practical situations, it is natural to consider that the possible values of these parameters are often only vaguely known and it is appropriated to interpret them by the decision maker's understanding. This parameters can be represented by fuzzy numbers which intent to describe the possible values that are inserted in the real-world problems. Then, the resulting multi-objective programming problem with fuzzy parameters would be viewed as the more realistic version of the conventional one. In addition, it is necessary to define a model for the quantification of the imprecise data that interpret the decision maker's judgment. There are many comparison approaches among fuzzy numbers and one of them is the possibility theory which was chosen in this work.

3.1 Fuzzy basic concepts

Mathematical programming problems need a precise definition of both the constraints and the objective function to be optimized. Fuzzy sets help handle uncertainties when multi-objective programming problems are formalized in the following form:

$$\begin{array}{ll} \widetilde{\min} & F(\tilde{\mathbf{a}}; \mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in \tilde{\Omega} \end{array} \tag{4}$$

where $F = (f_1, f_2, \ldots, f_m)(m \geq 2)$ is a vector of objectives, $\tilde{\mathbf{a}} \in \mathbb{F}(\mathbb{R}^{n \times m})$ represent the fuzzy parameters in the objective functions and $\tilde{\Omega} \subset \mathbb{F}(\mathbb{R}^n)$ is a subset of feasible solutions. $\mathbb{F}(\mathbb{R})$ defines the set of fuzzy numbers, $\mathbb{F}(\mathbb{R}^n)$ defines the set of n-dimensional vector with fuzzy parameters and $\mathbb{F}(\mathbb{R}^{n \times m})$ defines the set of $n \times m$-dimensional matrix with fuzzy parameters. However, we shall only address the uncertainties of the parameters in the objective functions, in this work.

In order to be able to sort fuzzy numbers in an increasing (decreasing) order, one has to opt for a comparison measure, therefore a possibility measure is used which is defined as:

Definition 5 (Possibility measure) *Let A be a fuzzy subset of U and let $\prod_X$ be a possibility distribution associated with a variable X which takes values in U. The* possibility measure*, $\pi(A)$, of A is defined by*

$$Poss\{X \text{ is } A\} \triangleq \pi(A) \triangleq \sup_{u \in U} \min(\mu_A(u), \pi_X(u)) \tag{5}$$

where μ_A is the membership function of A and π_X is the possibility distribution function of X. It can be interpreted as the possibility that the value X belongs to the set A *and it is defined to be numerically equal to the membership function of X.*

Then, it is possible define a way to compare two fuzzy numbers and this index can be formulated as follows

$$Poss\{\tilde{a}_1 \leq \tilde{a}_2\} = \sup_{u, v \in U; u \leq v} \min(\mu_{\tilde{a}_1}(u), \mu_{\tilde{a}_2}(v))$$

where $\mu_{\tilde{a}_1}$ and $\mu_{\tilde{a}_2}$ are membership functions of $\tilde{a}_1$ and $\tilde{a}_2$. Possibility degree $Poss\{\tilde{a}_1 \leq \tilde{a}_2\}$ shows to what extent $\tilde{a}_1$ is possibly less than or equal to $\tilde{a}_2$, as described in [7, 11].

The definition above enables one to define a fuzzily ordered set $\mathbb{F}(\mathbb{R})$ that is an extension of the classical ordered set. A set is said to be *completely ordered* if it satisfies the following conditions:

Definition 6 (Ordered fuzzily set) *A fuzzy subset $A \subset \mathbb{F}(\mathbb{R})$ is fuzzily ordered with respect to the possibility measure if each element in A satisfies the following basic properties:*

1. $Poss[\tilde{\mathbf{a}}_1 \leq \tilde{\mathbf{a}}_1] = 1$;
2. $Poss[\tilde{\mathbf{a}}_1 \leq \tilde{\mathbf{a}}_2] \geq \alpha_1$ *and* $Poss[\tilde{\mathbf{a}}_2 \leq \tilde{\mathbf{a}}_3] \geq \alpha_2 \Rightarrow$ $Poss[\tilde{\mathbf{a}}_1 \leq \tilde{\mathbf{a}}_3] \geq \min\{\alpha_1, \alpha_2\}$;
3. $Poss[\tilde{\mathbf{a}}_1 \leq \tilde{\mathbf{a}}_2] \geq \alpha_1$ *and* $Poss[\tilde{\mathbf{a}}_2 \leq \tilde{\mathbf{a}}_1] \geq \alpha_2 \Rightarrow$ $Poss[\tilde{\mathbf{a}}_1 = \tilde{\mathbf{a}}_2] \geq \min\{\alpha_1, \alpha_2\}$;

$\forall\, \tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \tilde{\mathbf{a}}_3 \in A$ *and* $\forall\, \alpha_1, \alpha_2 \in [0, 1]$.

According to the expressions above, a fuzzy subset $A \subset \mathbb{F}(\mathbb{R})$ is completely ordered. However, a fuzzy subset of $\mathbb{F}(\mathbb{R}^m)$ is only partially ordered. Therefore, the concept of optimal solution for single objective problems, which was defined in [3, 15], does no fit into the multi-objective formulation, unless the problem admits the so-called *ideal solution*, i.e. a single solution that simultaneously minimizes all objective functions as below:

Definition 7 (Ideal solution) *The ideal solution $\tilde{\mathbf{y}}$ of the multi-objective problem is defined as*

$$\tilde{\mathbf{y}}_i = f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^i), \;\; i = 1, \ldots, m$$

where $\mathbf{x}^i = \arg\min_{\mathbf{x} \in \mathbb{R}^n} f_i(\tilde{\mathbf{a}}_i; \mathbf{x})$.

The problem is said to admit an ideal solution whenever the set of arguments $\{\mathbf{x}^i,\ i = 1, \ldots, m\}$, possesses a single element. Because the multi-objective framework is most often

ISBN: 978-989-95079-6-8

employed in problems with conflicting objectives, a typical multi-objective problem is unlikely to admit such a solution. However, due the existence of an ideal solution is very rare, such a possibility will not be considered in the present analysis.

Conceptually, an efficient solution is one which is not dominated by any other feasible solution. Hence, the domination concept of a fuzzy multi-objective problem should reflect the decision maker's preferences. In this work, a fuzzy dominance concept is proposed which can be adjusted to the decision maker's preferences. This renders the proposed approach flexible and customizable, and possibly applicable to a wide range of problems. For any point $\mathbf{x}^0 \in \mathbb{R}^n$, consider the following subsets

$$\Omega_<(\mathbf{x}^0;\alpha) \triangleq \{\mathbf{x} \in \mathbb{R}^n : Poss[F(\tilde{\mathbf{a}};\mathbf{x}) \leq F(\tilde{\mathbf{a}};\mathbf{x}^0)] \geq \alpha \text{ and } Poss[F(\tilde{\mathbf{a}};\mathbf{x}) = F(\tilde{\mathbf{a}};\mathbf{x}^0)] < 1\}$$

$$\Omega_\geq(\mathbf{x}^0;\alpha) \triangleq \{\mathbf{x} \in \mathbb{R}^n : Poss[F(\tilde{\mathbf{a}};\mathbf{x}) \geq F(\tilde{\mathbf{a}};\mathbf{x}^0)] \geq \alpha\}$$

$$\Omega_\sim(\mathbf{x}^0;\alpha) \triangleq \{\mathbf{x} \in \mathbb{R}^n : \max\{Poss[F(\tilde{\mathbf{a}};\mathbf{x}) \leq F(\tilde{\mathbf{a}};\mathbf{x}^0)], Poss[F(\tilde{\mathbf{a}};\mathbf{x}) \geq F(\tilde{\mathbf{a}};\mathbf{x}^0)]\} \leq \alpha\}$$

The subset $\Omega_<(\mathbf{x}^0;\alpha)$ comprises the points in $\mathbb{R}^n$ that dominate $\mathbf{x}^0$, whereas $\Omega_\geq(\mathbf{x}^0;\alpha)$ encompasses the points in $\mathbb{R}^n$ that are dominated by $\mathbf{x}^0$. The set of points that neither dominate nor are dominated by $\mathbf{x}^0$ is denoted by $\Omega_\sim(\mathbf{x}^0;\alpha)$. The parameter α is a vector where each one of the terms, α_i with $i = 1, 2, \ldots, m$, belong to the interval $[0, 1]$. Those sets being defined, one can denote the set of fuzzy Pareto-optimal solutions as below:

Definition 8 (Fuzzy Pareto-optimal solution) *$\mathbf{x}^* \in \Omega$ is said be a fuzzy Pareto-optimal solution if there exists no other $x \in \Omega$ such that $Poss[f_i(\tilde{\mathbf{a}}_i;\mathbf{x}) \leq f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*)] \geq \alpha_i$, $\forall i$ and $Poss[f_j(\tilde{\mathbf{a}}_j;\mathbf{x}) = f_j(\tilde{\mathbf{a}}_j;\mathbf{x}^*)] < 1$ for at least one j, where $\alpha_i \in [0, 1]$, $\forall i$.*

For difficult optimization problems it is often the case that a local optimal solution is acceptable. A local efficient solution for the proposed problem is defined below:

Definition 9 (Fuzzy local Pareto-optimal solution) *$\mathbf{x}^* \in \Omega$ is said to be a fuzzy local Pareto-optimal solution if there is a real number $\delta \geq 0$ such that there exists no other $x \in \Omega \cap \mathcal{N}(x^*, \delta)$ such that $Poss[f_i(\tilde{\mathbf{a}}_i;\mathbf{x}) \leq f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*)] \geq \alpha_i$, $\forall i$ and $Poss[f_j(\tilde{\mathbf{a}}_j;\mathbf{x}) = f_j(\tilde{\mathbf{a}}_j;\mathbf{x}^*)] < 1$ in at least one j, where $\alpha_i \in [0, 1]$, $\forall i$.*

Note that the definition above implies that a candidate solution to the proposed fuzzy problem is (locally) non-dominated or efficient, if one cannot find (in a certain vicinity) another solution that simultaneously improves all the objective functions. This interpretation matches the classical counterpart of multi-objective optimization.

The convexity hyphotesis determine that the neighbourhood of each one local solution involves whole the feasible region.

Theorem 1 *Let $f_i : \Omega \subset \mathcal{X} \to \mathbb{F}(\mathcal{Y})$, $i = 1, \ldots, m$ a convex fuzzy functions about a convex subset Ω of a linear space $\mathcal{X}$. Then whole locally efficient solution is globaly efficient solution.*

Proof: Let $\mathbf{x}^* \in \Omega$ is a locally efficient solution. By definition of convex subset, we obtain $\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x} \in \Omega$, $\forall \mathbf{x} \in \Omega - \mathcal{N}(\mathbf{x}^*, \epsilon)$, with $\epsilon > 0$ e $\lambda \in [0, 1]$. Suppose $\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x} \in \Omega \cap \mathcal{N}(\mathbf{x}^*, \epsilon)$ then by the fuzzy Pareto-optimal solution definition, we obtain $Poss[f_i(\tilde{a}_i;\mathbf{x}^*) \leq f_i(\tilde{a}_i;\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x})] \geq \alpha_i^1$, $i \in \mathcal{I} = \{1, 2, \ldots, m\}$, and $Poss[f_j(\tilde{a}_j;\mathbf{x}^*) = f_j(\tilde{a}_j;\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x})] < 1$ for at least one $j \in \mathcal{I}$, where $\alpha_i^1 \in (0, 1]$, $\forall i \in \mathcal{I}$. By the convex fuzzy function definition, we obtain $f_i(\tilde{a}_i;\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x}) \lesssim \lambda f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*) + (1-\lambda) f_i(\tilde{\mathbf{a}}_i;\mathbf{x})$, $\forall i \in \mathcal{I}$, which it can be rewritten by using Possibility Theory as $Poss[f_i(\tilde{a}_i;\lambda\mathbf{x}^* + (1-\lambda)\mathbf{x}) \leq \lambda f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*) + (1-\lambda) f_i(\tilde{\mathbf{a}}_i;\mathbf{x})\}] \geq \alpha_i^2$, where $\alpha_i^2 \in (0, 1]$, $\forall i \in \mathcal{I}$. Thus, by using the ordered fuzzily subset definition, we obtain $Poss[f_i(\tilde{a}_i;\mathbf{x}^*) \leq \lambda f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*) + (1-\lambda) f_i(\tilde{\mathbf{a}}_i;\mathbf{x})\}] \geq \min\{\alpha_i^1, \alpha_i^2\} \Rightarrow Poss[f_i(\tilde{a}_i;\mathbf{x}^*) \leq f_i(\tilde{a}_i;\mathbf{x})] \geq \min\{\alpha_i^1, \alpha_i^2\}$, $\forall i \in \mathcal{I}$. By selecting a determined objective function $k \in \mathcal{I}$ and $k \neq i$, by the Theorem 2, we guarantee that $Poss[f_k(\tilde{a}_k;\mathbf{x}^*) = f_k(\tilde{a}_k;\mathbf{x})] < 1$ for at least one $k \in \mathcal{I}$. ■

3.2 Characterization of fuzzy efficient solutions

The characterization of efficient solutions, $efi(\Omega)$, by means of well defined scalar problems is a recurrent approach in fuzzy multi-objective problems. The following theorem relates efficient solutions and scalar problems.

Theorem 2 *$\mathbf{x}^* \in efi(\Omega)$ if and only if $\mathbf{x}^*$ solves the m scalar problems*

$$P_k: \quad \min_{\mathbf{x}\in\Omega} \quad f_k(\tilde{\mathbf{a}}_k;\mathbf{x}) \qquad s.t. \quad f_l(\tilde{\mathbf{a}}_l;\mathbf{x}) \lesssim f_l(\tilde{\mathbf{a}}_l;\mathbf{x}^*), \quad l = 1, 2, \ldots, m, \quad \forall\, l \neq k. \tag{6}$$

Proof: ($\Rightarrow$) If $\mathbf{x}^* \in efi(\Omega)$, then there exist no other $x \in \Omega$ such that $Poss[f_i(\tilde{\mathbf{a}}_i;\mathbf{x}) \leq f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*)] \geq \alpha_i$, $i = 1, 2, \ldots, m$, and $Poss[f_j(\tilde{\mathbf{a}}_j;\mathbf{x}) = f_j(\tilde{\mathbf{a}}_j;\mathbf{x}^*)] < 1$, for any j. In this case $\mathbf{x}^*$ solves (6) for all k.
($\Leftarrow$) Suppose $\mathbf{x}^*$ solves (6), but $\mathbf{x}^* \notin efi(\Omega)$, then there exists another $\mathbf{x} \in \Omega$ such that $Poss[f_i(\tilde{\mathbf{a}}_i;\mathbf{x}) \leq f_i(\tilde{\mathbf{a}}_i;\mathbf{x}^*)] \geq \alpha_i$, $\forall i$, and for some j, $Poss[f_j(\tilde{\mathbf{a}}_j;\mathbf{x}) \leq f_j(\tilde{\mathbf{a}}_j;\mathbf{x}^*)] < 1$. Therefore, $\mathbf{x}^*$ does not solve Problem (6). This contradiction concludes the proof. ■

The development of analytical conditions to efficient solutions, based on the characterization of non-dominated solutions to problems P_k, $k = 1, 2, \ldots, m$, is an important tool in the theoretical analysis. However, such an analysis yields only m non-dominated solutions, one to each scalar problem and is therefore, unable to generate the whole Pareto-optimal set.

Employing a similar analysis to the one presented above, we now establish the relationship between non-dominated solutions of a fuzzy multi-objective problem and solutions to the weighting problem. An alternative characterization based on the linear combination of the objectives can be expressed as

Theorem 3 *Let $\mathbf{x}^* \in \Omega$ solve the problem*

$$P_w: \quad \min_{\mathbf{x}\in\Omega} \quad \langle \mathbf{w}, F(\tilde{\mathbf{a}};\mathbf{x})\rangle = \textstyle\sum_{i=1}^m \omega_i f_i(\tilde{\mathbf{a}}_i;\mathbf{x}) \tag{7}$$

for some $\mathbf{w} \in \mathbb{R}^m$, $\mathbf{w} \geq \mathbf{0}$ and $\sum_{i=1}^m w_i = 1$. Then $\mathbf{x}^* \in efi(\Omega)$ if

ISBN: 978-989-95079-6-8

(i) $\mathbf{x}^*$ *is the unique solution(7), or*

(ii) $w_i > 0,\ i = 1, \ldots, m.$

Proof: *(i)* If $\mathbf{x}^* \in \Omega$ is a unique solution of (7), then $\forall \mathbf{x} \in \Omega$ and by definition, we obtain $Poss\left[\sum_{i=1}^{m} w_i \left(f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^*) - f_i(\tilde{\mathbf{a}}_i; \mathbf{x})\right) < 0\right] \geq \min_i\{\alpha_i\}$. Suppose $\mathbf{x}^* \notin efi(\Omega)$, i.e., there exists at least one $\mathbf{x}^0 \in \Omega$ such that $Poss[f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^0) \leq f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^*)] \geq \alpha_i,\ i = 1, 2, \ldots, m$ and $Poss[f_j(\tilde{\mathbf{a}}_j; \mathbf{x}^0) = f_j(\tilde{\mathbf{a}}_j; \mathbf{x}^*)] < 1$, for some j. This contradicts the uniqueness hypothesis, because $\mathbf{w} \geq \mathbf{0}$. Thus, $\mathbf{x}^* \in efi(\Omega)$.
(ii) Suppose $\mathbf{x}^* \notin efi(\Omega)$, but $\mathbf{x}^*$ is a solution of (7). Then there exists a $\mathbf{x}^0 \in \Omega$ such that $Poss[f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^0) \leq f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^*)] \geq \alpha_i,\ i = 1, 2, \ldots, m$ and $Poss[f_j(\tilde{\mathbf{a}}_j; \mathbf{x}^0) = f_j(\tilde{\mathbf{a}}_j; \mathbf{x}^*)] < 1$, for any j. Hence,

$$Poss\left[\sum_{i=1}^{m} w_i \left(f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^*) \leq f_i(\tilde{\mathbf{a}}_i; \mathbf{x}^0)\right) > 0\right] \geq \min_i\{\alpha_i\},\ \forall i.$$

A contradiction and therefore $\mathbf{x}^* \in efi(\Omega)$. ■

4 Results and analysis

The problems we use to evaluate this method are two hypothetic mathematical formulations, which are described in [5], with the fuzzy approach described in Section 3. Nevertheless, they are efficiency in validating the realized study. They were resolved using an modified implementation of **NSGA-II** that solves multi-objective programming problems with constraints or not. This modification was made in the comparison of the objective functions with fuzzy parameters between two feasible solutions which uses the concepts of fuzzy Pareto optimal solutions described in this work.

The vagueness was inserted into the costs of the objective function and fuzzy numbers are interpreted in the form $(a, \underline{a}, \overline{a})_{LR}$ where a is the modal value, $\underline{a}$ is the scattering left and $\overline{a}$ is the scattering right of each fuzzy number.

Example 1 (Schaffer's problem)

$$\begin{aligned} \min \quad & f_1(\tilde{\mathbf{a}}_1; \mathbf{x}) = (x + \tilde{a}_1)^2 \\ \min \quad & f_2(\tilde{\mathbf{a}}_2; \mathbf{x}) = (x - \tilde{a}_2)^2 \\ s.t \quad & -5 \leq \mathbf{x} \leq 5 \end{aligned} \tag{8}$$

where $\tilde{a}_1 = (0, 0, 2)_{LR}$ *and* $\tilde{a}_2 = (2, 1, 1)_{LR}$.

The figures below present the fuzzy solution and fuzzy front.

In Figure 2, each star represents one solution of this multi-objective problem by using the fuzzy Pareto optimal concept defined by Sakawa to $\alpha = 0.8$*, while each square represents one solution by using the fuzzy Pareto optimal concept described in this work to* $\alpha = 0.8$*, too. It can be observed that a range of possible Pareto optimal solutions is formed when the squares are merged. We can also see that many stars are inside some squares, i.e., this solutions have a degree of possibility great or equal to 0.8 and belong to the range of Pareto optimal solutions obtained by the definition that uses possibility theory.*

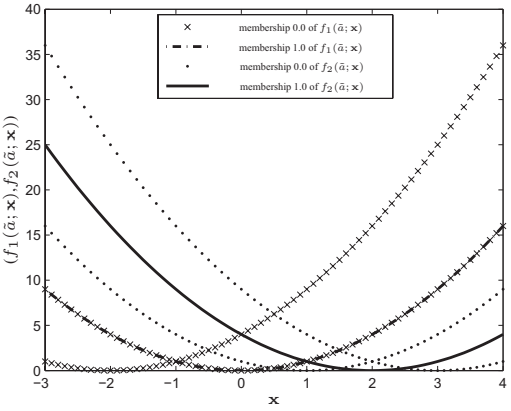


Figure 1: Objective functions

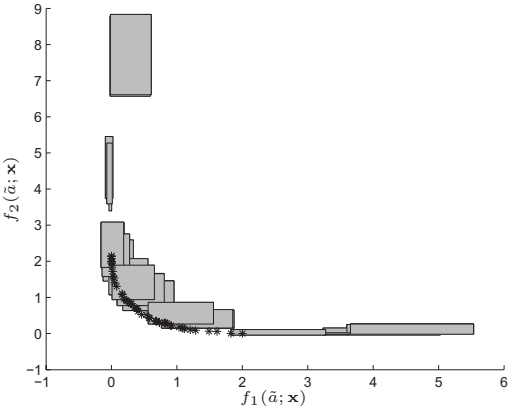


Figure 2: Pareto front

Example 2 (Binh and Kern's problem)

$$\begin{aligned} \min \quad & f_1(\tilde{\mathbf{a}}_1; \mathbf{x}) = \tilde{a}_{11} x_1^2 + \tilde{a}_{12} x_2^2 \\ \min \quad & f_2(\tilde{\mathbf{a}}_2; \mathbf{x}) = (x_1 - \tilde{a}_{21})^2 + (x_2 - \tilde{a}_{22})^2 \\ s.t. \quad & (x_1 - 5)^2 + x_2^2 \leq 25 \\ & (x_1 - 8)^2 + (x_2 + 3)^2 \geq 7.7 \\ & 0 \leq x_1 \leq 5,\ \ 0 \leq x_2 \leq 3 \end{aligned} \tag{9}$$

where $\tilde{a}_{11} = \tilde{a}_{12} = (4, 1, 1)_{LR}$, $\tilde{a}_{21} = \tilde{a}_{22} = (5, 1, 1)_{LR}$.

The figures below present the fuzzy solution and fuzzy front.

In Figure 3 are shown the function objectives of Problem (2) where the superior drawing represents the function $f_1(\tilde{\mathbf{a}}_1; \mathbf{x})$ *and the inferior one represents the function* $f_2(\tilde{\mathbf{a}}_2; \mathbf{x})$*. The drawings with solid line represent the value of the objective functions with* $\alpha = 1$ *while the ones with dotted line represent the value of the objective functions with* $\alpha = 0$*.*

Again, each star represents one solution by using the concept defined by Sakawa to $\alpha = 0.8$*, while each square represents one solution by using the concept described here to* $\alpha = 0.8$*, too. In this case, the stars are in the imaginary bound of the Pareto front range formed by possible Pareto solutions, i.e., the solutions obtained by Sakawa's definition have a satisfaction level closed in* $\alpha = 0.8$*.*

ISBN: 978-989-95079-6-8

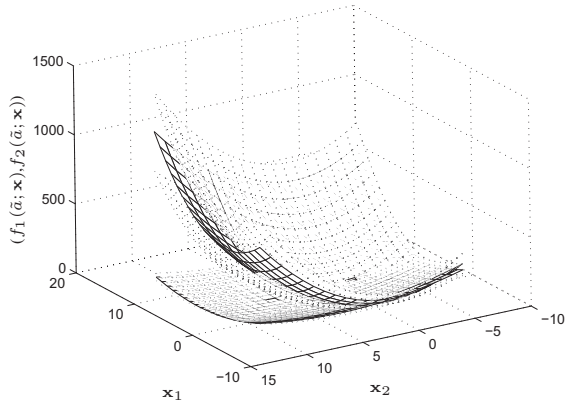


Figure 3: Objective functions

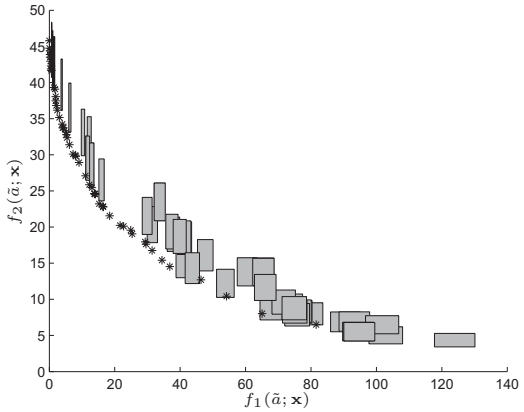


Figure 4: Pareto front

5 Conclusion

The use of defuzzification methods to solve fuzzy mathematical programming problems is very common. These methods transform a fuzzy number into a classical one, but some data are always lost in each case. The Pareto-optimality presented in this work makes use of different defuzzification methods in various stages of the definitions.

Multi-objective Programming problems are very important in a variety of both theoretical and practical areas. As ambiguity and vagueness are natural and ever-present in real-life situations that require precise solutions, it makes perfect sense to attempt to address these problems using Fuzzy Multi-objective Programming problems. In this context, this paper presented a novel theory to determine Pareto-optimality conditions which provide the user with a fuzzy solution. This theory is an expansion of the classical Pareto-optimality theory and demonstrates the necessary conditions for fuzzy Pareto-optimality. Some numerical examples are presented to validate the theory outlined.

The authors aim firstly to extend the line of investigation regarding Fuzzy Quadratic Programming problems in order to try to solve practical real-life problems by facilitating the building of Decision Support Systems. This requires the involvement of fuzzy costs as well as fuzzy coefficients, as a must.

Acknowledgments

The authors want to thank the support provided by the Brazilian agency CAPES.

References

[1] R. E. BELLMAN and L. A. ZADEH, Decision-marking in a fuzzy environment. *Management Science* 17(4):B141–B164, 1970

[2] L. CAMPOS and J. L. VERDEGAY. Linear programming problems and ranking of fuzzy numbers. *Fuzzy Sets and Systems* 32:1–11, 1989.

[3] L. A. P. CANTÃO and A. YAMAKAMI, Nonlinear programming with fuzzy parameters: Theory and applications In: International Conference on Computational Intelligence for Modelling, Control and Automation, CIMCA 2003, Vienna, Austria, 2003.

[4] V. CHANKONG and Y. Y. HAIMES. *Multiobjective decision making: Theory and Methodology*, volume 8 of *North Hollando series in system science and engineering*. North Holland, New York, USA, 1983.

[5] K. DEB. *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, Chichester, UK, 2001.

[6] D. DUBOIS and H. PRADE. *Fuzzy sets and systems: Theory and Application*. Academic Press, San Diego, USA, 1980.

[7] D. DUBOIS and H. PRADE. Ranking fuzzy numbers in the setting of possibility theory. *Information Sciences*, 30(3):183–224, 1983.

[8] F. JIMÉNEZ and J. M. CADENAS and G. SÁNCHEZ and A. F. GÓMEZ-SKARMETA and J. L. VERDEGAY. Multi-objective evolutionary computation and fuzzy optimization. *International Journal of Approximate Reasoning* 43:59–75, 2006.

[9] A. KAUFMANN and M. M. GUPTA. *Introduction to fuzzy arithmetic: theory and applications*. Van Nostrand Reinhold, New York, USA, 1984.

[10] G. J. KLIR and B. YUAN. *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall, New Jersey, USA, 1995.

[11] M. INUIGUSHI: On possibilistic/fuzzy optimization. In: 12th IFSA World Congress, Cancun, Mexico, June 2007, pag. 351-360

[12] V. PARETO. *Cours d'economique politique*, volume I. Macmillan, Paris, FR, 1987.

[13] V. PARETO. *Le cours d'economique politique*, volume II. Macmillan, London, UK, 1987.

[14] W. PEDRYCS and F. GOMIDE. *An introduction of fuzzy sets: analisys and design*. MIT press, Cambridge, UK, 1998.

[15] M. SAKAWA. *Fuzzy sets and interactive multiobjective optimization*. Plenum Press, New York, USA, 1993.

[16] L. A. ZADEH. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[17] L. A. ZADEH. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1):3–28, 1978.

[18] L. A. ZADEH. Is there a need for fuzzy logic?. *Information Sciences*. 178:2751–2779, 2008.

[19] H.-J. ZIMMERMANN. *Fuzzy set theory and its applications*. Kluwer Academic Publishers, Massachusetts, USA, third edition, 1996.

ISBN: 978-989-95079-6-8

Terrain Morphology Classification over Fuzzy Digital Elevation Models

Jorge M. F. Santos[1]
Departamento de Matemática da Universidade de Coimbra and INESC Coimbra, Portugal 1
Email: jmfs@mat.uc.pt

Abstract— *One way to incorporate uncertainty associated to the data and to the interpolation method in a digital elevation model (DEM) is to use Fuzzy Sets Theory. A fuzzy DEM can be generated from fuzzy data, where elevations are represented by fuzzy numbers. To use that type of terrain models in further applications, a generalization of known methods is needed. An example about the generalization of terrain morphology classification methods is proposed in this work. Given a fuzzy DEM, a procedure to find ridges and other morphological features will be analyzed. Two examples of this approach will be presented to illustrate the proposed methods.*

Keywords— Fuzzy DEM, fuzzy morphological feature, signal of a fuzzy number.

1 Introduction

Automated identification of objects representing terrain features (such as peaks, channels and ridges) from a digital elevation model is a well known procedure in Geomorphology. The development of those methods has important implications in terms of spatial analysis and for the generalization of spatial databases, for example. Another important application is the study of water flow over the terrain surface that represents a fundamental geomorphological process and it is closely related to its shape: it constraints the water route over the surface and it is mainly shaped by this flow.

The topographic surface is usually defined through a gridded DEM constructed by interpolation from a finite set of samples and it is a well-known problem in geographic modeling. The uncertainty associated to this process generates uncertain elevation values which can be modeled using fuzzy numbers [1, 2, 3]. Therefore, methods to identify morphological terrain features have to be generalized to fuzzy DEMs in order to express the impact of elevation uncertainty in terrain classification. In this work we will study one approach to this problems, generalizing known methods over fuzzy DEMs.

2 Morphology classification

To identify morphological objects over the terrain surface it is adopted here the strategy developed by Wood [4]. This problem can be seen as an image processing problem to classify each *pixel* (grid cell) in one class of *geomorphological feature* (usually, *peak*, *ridge*, *pass*, *plane*, *channel* or *pit*). The feature classification is based on table 1, where the signals [1] of the four morphological parameters: *slope* (S), *cross sectional curvature* ($crosc$), *maximum profile convexity* ($maxic$) and *minimum profile convexity* ($minic$), are given by

$$sign(S) = sign(\arctan\sqrt{d^2+e^2}), \quad (1)$$

$$sign(crosc) = sign(cde - bd^2 - ae^2), \quad (2)$$

$$sign(maxic) = -sign(a+b-\sqrt{(a-b)^2-c^2}), \quad (3)$$

$$sign(minic) = -sign(a+b+\sqrt{(a-b)^2-c^2}). \quad (4)$$

The values of a, b, c, d and e are evaluated from the first and second order derivatives of the quadric surface

$$z = ax^2 + by^2 + cxy + dx + ey + f, \quad (5)$$

usually fitted to a 3x3 data window [4].

Table 1: Feature classification criteria [4].

Feature	$slope$	$crosc$	$maxic$	$minic$
Peak	0	all	+	+
Ridge	0	all	+	0
	+	+	all	all
Pass	0	all	+	-
Plane	0	all	0	0
	+	0	all	all
Channel	0	all	0	-
	+	-	all	all
Pit	0	all	-	-

3 Fuzzy morphology classification

The above procedures are based on a regular elevation grid. Considering that the elevations have some degree of uncertainty which is expressed by fuzzy numbers, we will study the generalization of those procedures to a fuzzy DEM. The first step is to evaluate the first and second order fuzzy derivatives. This is done using interval arithmetic over the fuzzy numbers $\alpha - levels$. Those fuzzy derivatives will be used in equations (1 to 4) to obtain fuzzy morphological parameters using again interval arithmetic [5].

3.1 Signal of a fuzzy number

To use the fuzzy version of the above *slope*, *cross sectional curvature*, *maximum profile curvature* and *minimum profile curvature* it is need now a generalization of the $signal$ function. It is proposed here to adapt the concept of *overtaking between fuzzy numbers* in [3] to decide how much a fuzzy value is negative, positive or zero. This concept is based on the notion of *overtaking between intervals* defined by

[1] We assume that $sign$ is a three-valued function in $\{-1, 0, 1\}$.

ISBN: 978-989-95079-6-8

$$\sigma(X,Y) = \begin{cases} 0, & X_u \leq Y_l \\ \frac{X_u - Y_l}{X_u - X_l}, & X_u > Y_l \wedge X_l \leq Y_l \\ 1, & X_l < Y_l \end{cases}, \quad (6)$$

where X and Y are intervals defined by $X = [X_l, X_u]$ and $Y = [Y_l, Y_u]$.

Following, the *overtaking between fuzzy numbers* is

$$\sigma(\tilde{X}, \tilde{Y}) = \int_0^1 \sigma([\tilde{X}]_\alpha, [\tilde{Y}]_\alpha) w(\alpha) d\alpha, \quad (7)$$

where $[\tilde{X}]_\alpha$ and $[\tilde{Y}]_\alpha$ are the $\alpha - levels$ of fuzzy numbers $\tilde{X}$ and $\tilde{Y}$ and $w(\alpha)$ is a weight function $w : [0,1] \rightarrow \Re$ such that

$$\begin{aligned} & w(\alpha) \geq 0 \\ & \alpha < \alpha' \Rightarrow w(\alpha) < w(\alpha') \\ & \int_0^1 w(\alpha) d\alpha = 1. \end{aligned} \quad (8)$$

.

To apply this strategy, it is needed to define a fuzzy zero value. When fuzzy numbers like the fuzzy morphological parameters have very different support widths, ranging from 10 to 0.01, approximately, scaling problems may arise if comparison is needed. To prevent those problems, the zero value will depend on the fuzzy value to be evaluated. The fuzzy zero will have the same membership of the respective fuzzy value but there will be a shift over the real line such that the $\alpha - level$ of maximum membership will be centered in the zero of real line. For triangular fuzzy numbers it will be $\tilde{0} = (X_l - X_m/0/X_u - X_m)$ for fuzzy number $\tilde{X} = (X_l/X_m/X_u)$, see figure 1. This type of fuzzy zero value will be called here *twin zero* of $\tilde{X}$.

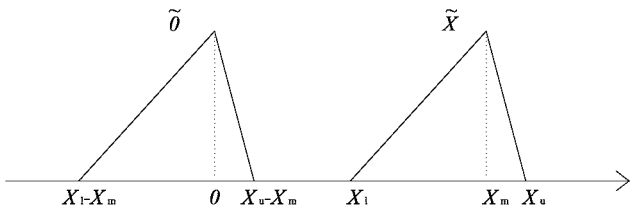


Figure 1: Twin zero of a triangular fuzzy number $\tilde{X}$.

First, to evaluate how much an interval X is positive, it is used a named $positive$ function defined by

$$pos(X) = \begin{cases} 0, & X_l \leq 0_l \\ \frac{X_u - 0_u}{X_u - X_l}, & 0_l < X_l < 0_u \\ 1, & X_l \geq 0_u \end{cases}, \quad (9)$$

where $[0_l, 0_u]$ is the twin zero of $X = [X_l, X_u]$, that is, it has the same width and it is centered in the real zero value. In the same way, $negative$ and $zero$ functions for intervals can be defined by

$$neg(X) = \begin{cases} 0, & X_u \geq 0_u \\ \frac{0_l - X_l}{X_u - X_l}, & 0_l < X_u < 0_u \\ 1, & X_u \leq 0_l \end{cases}, \quad (10)$$

and

$$zer(X) = \begin{cases} 0, & X_l \geq 0_u \\ \frac{0_u - X_l}{X_u - X_l}, & 0_l < X_l < 0_u \\ 1, & X_l = 0_l \\ \frac{X_u - 0_l}{X_u - X_l}, & 0_l < X_u < 0_u \\ 0, & X_u \leq 0_l \end{cases}. \quad (11)$$

Since the width of the zero interval and of X are the same, it is easy to see that[2]: $pos(X) + neg(X) + zer(X) = 1$ and $zer(X) = 1 - pos(X) \dot{\vee} zer(X) = 1 - neg(X)$. To extend those functions to fuzzy numbers it is used the same approach of (7):

$$pos(\tilde{X}) = \int_0^1 pos([\tilde{X}]_\alpha) w(\alpha) d\alpha, \quad (12)$$

where $w(\alpha)$ is the weight function in (8). The same can be done for $negative$ and $zero$ functions.

3.2 Fuzzy morphological features

From table 1, we can evaluate how much a grid cell belongs to a feature class using the above functions to evaluate the signals of the fuzzy morphological parameters and the standard fuzzy conjunction and disjunction defined by min and max operators [6]. This procedure will generate the fuzzy morphological classification, providing a method to quantify the membership of every terrain grid cell to the *fuzzy peak*, *fuzzy ridge*, *fuzzy pass*, *fuzzy plane*, *fuzzy channel* or *fuzzy pit* classes.

4 Examples

Two examples will be presented in this section to illustrate the methods discussed above. In the first case it will be used a digital elevation model of Coimbra region, crossed by Mondego river with narrow valleys and steep slopes (see figure 2). The second example is Kilimanjaro region, where a big volcanic cone lies down on the middle of a plain (see figure 5). In the following figures, all distance units are in meters and angles in degrees.

4.1 Coimbra region

The study region is a 2.0 by 2.1 kilometers rectangle, with elevations ranging from 20 do almost 300 meters. The elevation uncertainty model here is just the root mean square error (RMSE) provided by the map producer. It was introduced a small random gaussian variation added to that RMSE to define the support of symmetrical triangular fuzzy number to express elevation uncertainty. The modal values are taken from the elevations of the crisp DEM. Some results can be seen in figures 3 and 4.

4.2 Kilimanjaro mountain

It was used a rectangle with about 85 by 80 kilometers to enclose Kilimanjaro mountain, with elevations going from 800 to almost 6000 meters. The DEM for this region was taken from SRTM mission, version 4, with original gaps filled by interpolation and with other elevation data available available for that region. To quantify the distribution of DEM elevation uncertainty, about 100 GPS control points were measured around the mountain and in the top. The difference between the elevations of those points and the DEM were calculated

[2]The symbol $\dot{\vee}$ represents xor operator

ISBN: 978-989-95079-6-8

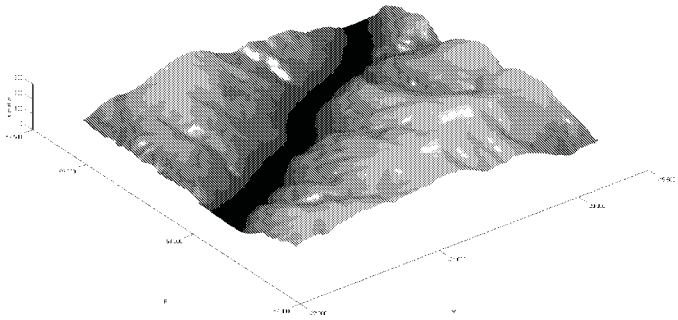


Figure 2: Coimbra digital elevation model.

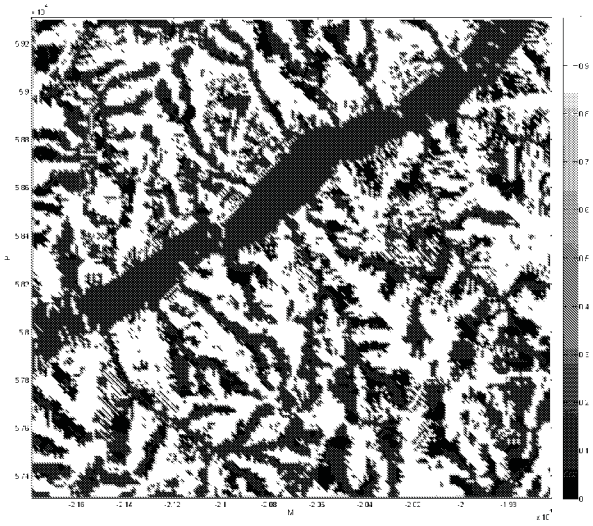


Figure 3: Fuzzy ridge membership for Coimbra study area.

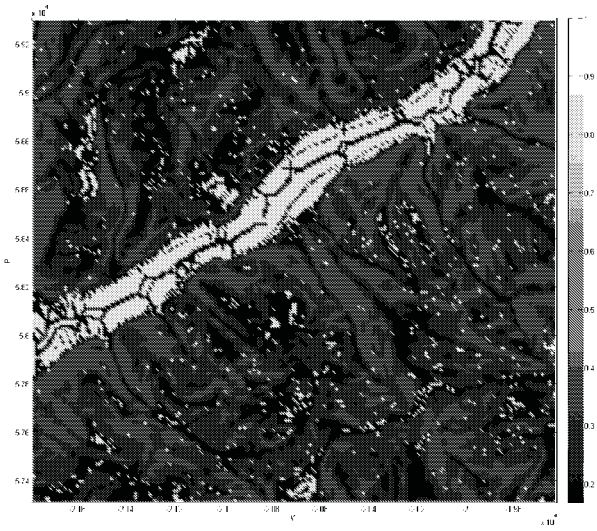


Figure 4: Fuzzy plane membership for Coimbra.

and an error surface were interpolated from those error samples. The error surface in figure 6 was used to estimate the support of symmetric triangular fuzzy numbers that will express the fuzzy elevations. The modal values of those fuzzy numbers are given by the crisp DEM. Some results can be seen in figures 7 and 8. Instead of crisp 0 or 1 membership values, now there are partial membership values to the fuzzy morphological objects, reflecting the uncertainty of the terrain model. Comparing figures 7 and 8, it is possible to detect visually some correlation between the error surface and the fuzzy membership values.

5 Conclusions

This work gave an example that methods applicable to crisp digital elevation models can be generalized to be also applicable to fuzzy DEMs or to similar gridded data expressed by fuzzy numbers. This approach has the advantage to allow the inclusion of uncertainty modeled by Fuzzy Set Theory in geographic analysis. The examples applications presented here are just a first step to develop more robust and advanced terrain classification methods to apply over fuzzy DEMs. Fuzzy values of morphological parameters, like slope and curvatures, have sometimes support intervals too wide. Therefore, those numerical operations are very sensitive to the input values and maybe more robust algorithms are needed here. These methods can also be combined with multi-scale classification methods used in [7, 8, 9, 10] or others approaches like in [11], for example, in order to improve the automated identification of morphological objects.

ISBN: 978-989-95079-6-8

Appendix

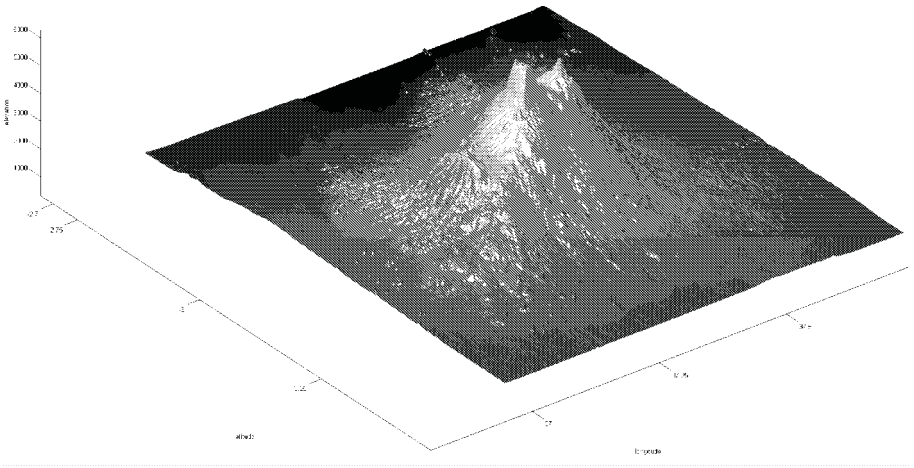


Figure 5: Kilimanjaro digital elevation model.

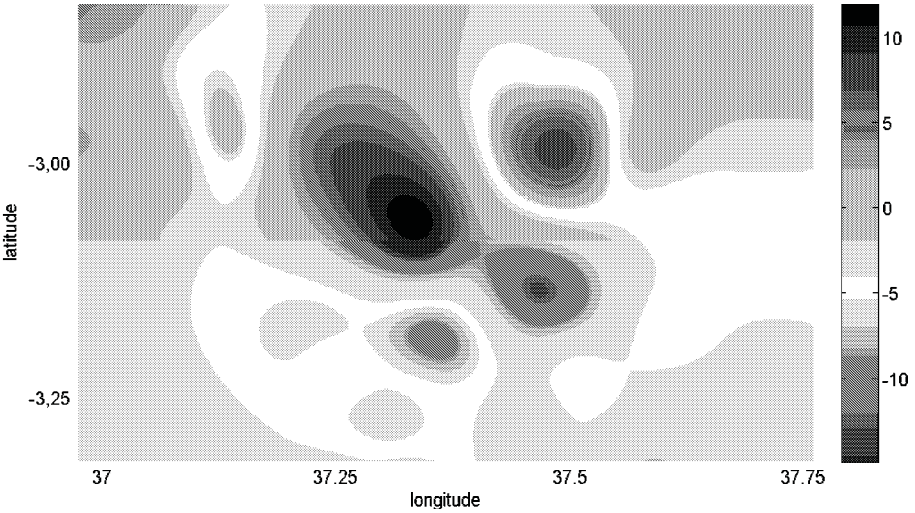


Figure 6: Kilimanjaro uncertainty spatial distribution.

ISBN: 978-989-95079-6-8

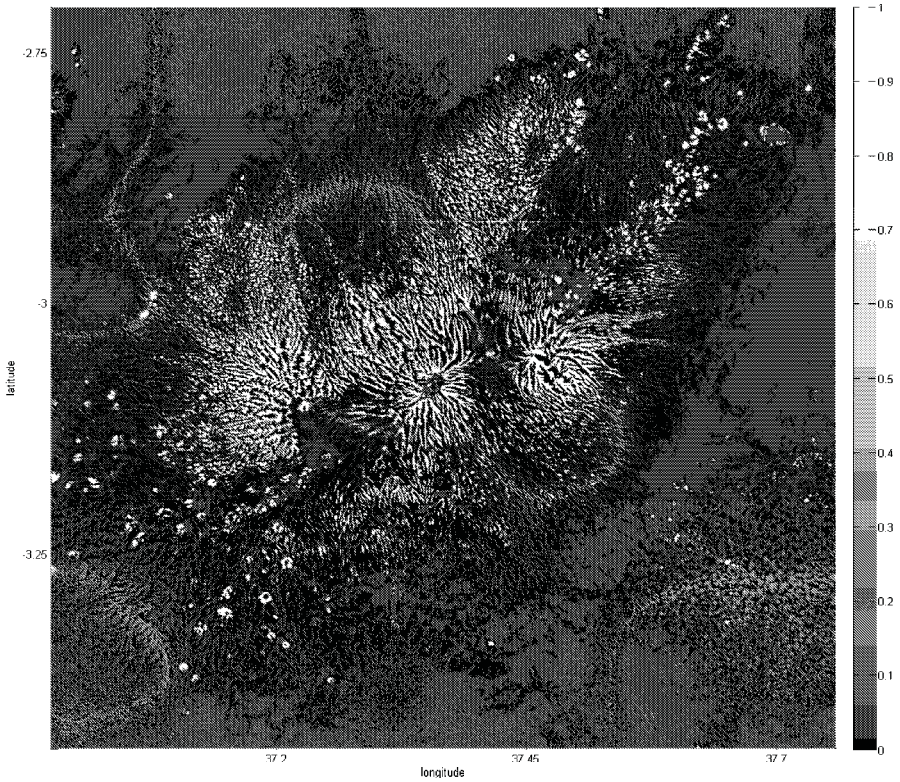


Figure 7: Fuzzy ridge membership for Kilimanjaro region.

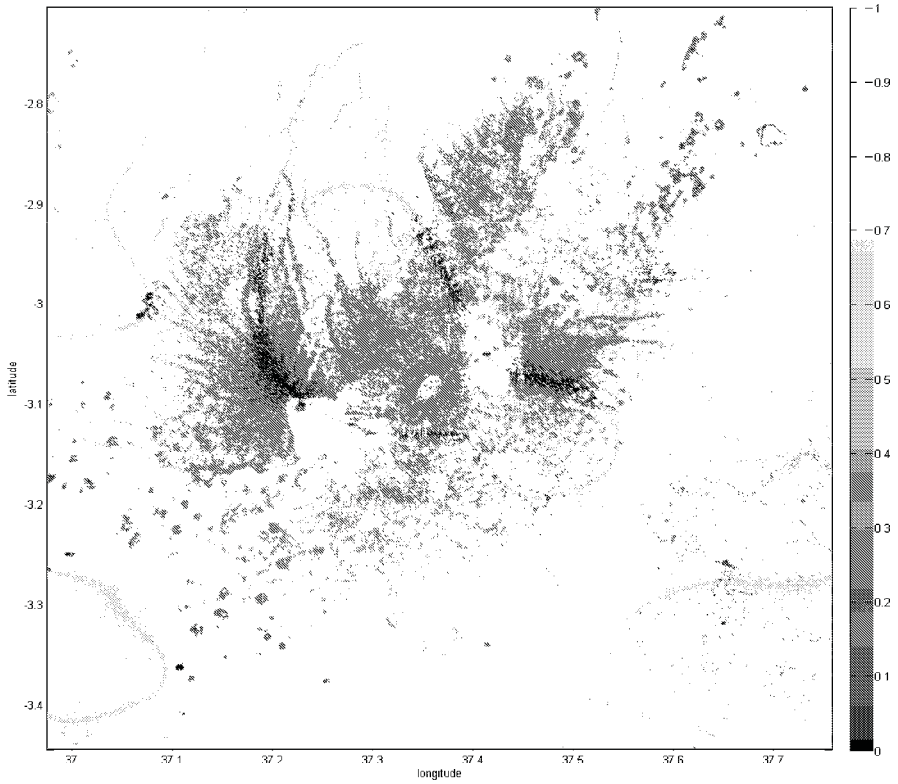


Figure 8: Fuzzy plane membership for Kilimanjaro region.

ISBN: 978-989-95079-6-8

Acknowledgments

Author thanks to the rest of *Kili 2008 Expedition* team (http://kili2008-expedition.blogspot.com/2008/09/ complete-list-of-participants.html) the hard work to provide GPS control points for DEM quality assessment of Kilimanjaro mountain and to Dora Santos, who helped to compile cartographic information of Coimbra.

References

[1] A. M. Anile, B. Falcidieno, G. Gallo, M. Spagnuolo, and S. Spinello. Modeling uncertain data with fuzzy b-splines. *Fuzzy Sets and Systems*, 113(3):397–410, 2000.

[2] W. Lodwick and J. Santos. Constructing consistent fuzzy surfaces from fuzzy data. *Fuzzy Sets and Systems*, 135(2):259–277, 2003.

[3] W. Lodwick, editor. *Fuzzy Surfaces in GIS and Geographical Analysis: Theory, Analytical Methods, Algorithms and Applications*. CRC Press, New York, 2008.

[4] J. Wood. *The geomorphological characterisation of digital elevation models*. PhD thesis, University of Leicester, UK, 1996.

[5] J. Santos. *Interpolação Difusa na Modelação de Fenómenos Geográficos*. PhD thesis, University of Coimbra, Portugal, 2004.

[6] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic, Theory and Applications*. Prentice Hall Inc., New Jersey, 1995.

[7] P. Fisher, J. Wood, and T. Cheng. Where is helvellyn? fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers*, 29:106–128, 2004.

[8] P. Fisher, J. Wood, and T. Cheng. *Fuzzy Modelling with Spatial Information for Geographic Problems*, chapter Fuzziness and ambiguity in multi-scale analysis of landscape morphometry, pages 209–232. Springer, New York, 2005.

[9] P. Fisher, J. Wood, and T. Cheng. Higher order vagueness in geographical information: Empirical geographical population of type n fuzzy sets. *Geoinformatica*, 11:311–330, 2007.

[10] Y. Deng and J. P. Wilson. Multi-scale and multi-criteria mapping of mountain peaks as fuzzy entities. *International Journal of Geographical Information Science*, 22:205–218, 2007.

[11] P. A. Burrough, P. F. M. Van Gaans, and R. Hootsmans. High-resolution landform classification using fuzzy k-means. *Fuzzy Sets and Systems*, 113:37–52, 2000.

ISBN: 978-989-95079-6-8

Modeling of Polyester Dyeing Using an Evolutionary Fuzzy System

Maryam Nasiri[1] Stefan Berlik[2]

1 Institute of Knowledge Based Systems and Knowledge Management, Siegen University
Siegen,Germany
2 Software Engineering Institute, Siegen University
Siegen, Germany

Email: maryam.nasiri@uni-siegen.de, berlik@informatik.uni-siegen.de

Abstract*—The aim of this study is to apply and compare statistical regression and an evolutionary fuzzy system to model color yield in the polyester high temperature (HT) dyeing as a function of disperse dyes concentration, temperature and time. The predictive power of the obtained models was evaluated by means of MSE value. It seems that for modeling cases such as the one considered in this study, the evolutionary fuzzy system with a minimum MSE showed a better predictive capability than the model based on statistical regression.*

Keywords— Evolutionary Algorithm, Fuzzy Inference System, Polyester Dyeing, Modelling.

1 Introduction

After introducing and developing soft computing methods many attempts have been made to use and apply them in textile research. In this work, we describe two modeling methodologies, namely; statistical regression and evolutionary-fuzzy systems and compare their capability for modeling of color yield in polyester dyeing. So, a summary of the evolutionary fuzzy approach and a brief note on polyester dyeing are described followed by literature review in use of soft computing in textile engineering.

1.1 Evolutionary Fuzzy System

Many real world problems are so complex that classical computing methods fail to deal with them efficiently. An alternative approach to deal with these problems is soft-computing. The Term soft computing refers to a family of computing techniques comprising four different partners: fuzzy logic, evolutionary computation, neural network and probabilistic reasoning. The term soft computing distinguishes these techniques from hard computing that is considered less flexible and computationally demanding. From this set of techniques, Fuzzy Inference System (FIS) are a powerful tool for modeling real control systems [1,2].
After the initial fuzzy inference system has been set up, the parameters of its fuzzy sets have been optimized in a subsequent step using the covariance matrix adaptation evolution strategy (CMA-ES) [3,4]. CMA-ES is an advanced evolution strategy that relies on an enhanced update mechanism for the mutation distribution's covariance matrix. A survey about genetic-fuzzy systems and its application can be found in [2].

1.2 Polyester Dyeing

Polyester is the most important man made fiber, which is produced by melt spinning process [5]. The dyeing of polyester is limited to only disperse dyes [6] and requires special conditions such as high temperature (≈ 130°C), dry heat (190-220°C), or using carrier in the dye bath [7]. The chemical structure of disperse dyes contains polar groups but there are no ionic groups present which leads to their very low solubility in water. The three main chemical structures of disperse dyes are azo, anthraquinone, and nitro diphenylamine [8]. Temperature, time and disperse dye concentration are the primary factors affecting the color yield in dyeing polyester. The relative importance of these factors can be seen in models representing the color yield as a function of them. These models may also have application in processing and cost minimization. The color yield is shown by K/S. K/S shows the ratio of the absorbed light by an opaque substrate relative to the scattered light from it. This ratio is calculated by Kubelka-Munk theory as [1,9]:

$$(K/S)_\lambda = \frac{(1-R_\lambda)^2}{2R_\lambda} \quad (1)$$

1.3 Literature Review

The research related to the subject of this work can be summarized as follows:
Kim et al. studied Fuzzy modeling, control and optimization of textile processes [10]. Soft computing methods in textile sciences has been reviewed by Sztandera and Pastore [11]. Zarandi et al. presented a fuzzy expert system for textile manufacturing system using fuzzy cluster analysis [12]. FIS has been applied by Hung and to control continuous dyeing [13]. Rautenberg et al. used fuzzy sets for color recipe specification in the textile print shop [14]. Jahmeerbacusa et al. studied fuzzy dye bath pH control in exhaust dyeing [15]. Fuzzy- based simulation model for declorization of industrial waste water has been presented by Abdou et al. [16] Marjoniemi and Mantysalo applied Adaptive Neuro Fuzzy Inference Systems (ANFIS) in modeling dye solution and concentration [17,18]. Active Tension Control of High Speed Splitting Machines using Fuzzy PID has been studied by Chung et al. [19]. Tavanai et al. used fuzzy regression method to model the color yield in dyeing [20]. Smith et al.

ISBN: 978-989-95079-6-8

studied improving computer control of batch dyeing operations [21]. Kim et al. proposed an emotion-based textile indexing system using colors, texture and patterns. Their system utilizes both fuzzy rules and neural networks [22]. A genetic-fuzzy approach has been applied by Nasiri et al. to model polyester dyeing [1]. Callhof and Wulfhort applied ANFIS [23], Veit et al. employed Neural Networks [24], and Nasiri used fuzzy regression in texturing [25]. Guifen et al. predicted the warp breakage rate in weaving using neural network techniques [26]. Predicting the properties of melt spun fibers using neural network has been studied by Chung-Feng [27]. Jeng-Jong applied an evolutionary algorithm to obtain the best combination of weaving parameters for woven fabric designs [28]. Sette et al. used soft computing techniques to Fiber-to-Yarn production process [29]. Peeva et al. examined fuzzy relational calculus theory with applications in various engineering subjects such as textile [30].Blaga studied the application of evolutionary algorithms in knitting technology [31]. An automatic textile sales forecast using fuzzy treatment of explanatory variables was used by Thomassey et al. [32]. Wong et al. studied genetic optimization of JIT operation schedules for fabric-cutting process in apparel manufacture [33]. Siddaiah applied automation in cotton ginning [34]. Aggregation as similarity in a morphological framework for the processing of textile images has been studied by Soria-Frisch [35].

2 Modeling of Polyester Dyeing

The aim of this study is to model variations of color yield of C.I. Disperse Blue 266 versus time, temperature, and disperse dye concentration in the high temperature (HT) polyester dyeing process using an evolutionary fuzzy system and the statistical regression method. A total number of 120 polyester samples were dyed according to the conditions in Table 1. The models based on statistical regression and the evolutionary fuzzy system was developed and compared [1].

Table 1: Dyeing conditions [1]

Dye Concentration (%owf)	0.75 1.5 3 4.5 6
Temperature (°C)	100 110 115 120 125 130
Time (min)	12 24 36 48

2.1 Modeling by Statistical Regression

Modeling of a dependent variable as a function of one or more independent variable(s) can be carried out by means of regression. A general multiple regression for modeling the color yield can be considered to take the following form:

$$Y = A_0 + A_1 x_1 + A_2 x_2 + A_3 x_3 \qquad (2)$$

After using the least squares method to obtain the coefficients in the above equation, the statistical regression model is obtained as fallows:

K/S=-79.4+1.54Conc.+0.11Time+0.71Temp.

To verify the necessary statistical regression conditions, the following four conditions can be used [1,36]:

1) Linear form for the normal plot of the residuals.
2) I chart of residuals should lie between the upper and lower control limits without any specific pattern.
3) The residuals histogram should be of an approximately normal form.
4) Residuals versus fitted values should show no specific pattern.

Fig. 1 shows the information related to the statistical regression model obtained for C.I. Disperse Blue 266 [1]. Regarding the above mentioned four conditions for the validity of the models, it can be said that the linear statistical model cannot be accepted [1,36].

Obviously, the model discussed above with its just four degrees of freedom is of limited capability and there are more powerful e.g. nonlinear regression models. However, these tend to be quite complicated and that is why in the following an evolutionary fuzzy approach will be considered.

2.2 Modeling by Evolutionary Fuzzy System

Variation of K/S function versus two variables of temperature, time and concentration for dyed samples in C.I. Disperse Blue 266 as curves and surfs in some figures like Fig. 2 are shown to study the effect and behaviour of each variable. In view of this figure, for example, it is clearly observed that temperature has a greater effect on K/S value than time has.

In the same way, the FIS model has been investigated along the following lines to model the colour yield of C.I. Disperse Blue 266 in polyester high temperature dyeing [1,37,38].

First, membership functions for input and output variables according to Table 2 and Table 3 have been determined. The Gaussian membership function, and Mamdani max-min Inference was used for all input and output variables [1].

Table 2: Parameters of fuzzy set for input variables [1].

Fuzzy set	Concentration		Time		Temperature	
	Mean	Std	Mean	Std	Mean	Std
Low	1.28	0.848	13.1	14.2	100	9.61
Medium	3.38	0.891	-	-	119	1.64
High	5.87	1.33	44.4	11.7	129	3.97

Table 3: Parameters of fuzzy set for output variable [1].

Fuzzy set	Color yield (K/S)	
	Mean	Std
Very low	3.29	0.62
Low	4.11	1.42
Medium	12.8	1.69
High	19.3	1.86
Very high	29.8	2.23

Secondly, the following nine rules were defined according to the physical and chemical structure of polyester fiber, HT dyeing of polyester, and the behavior of 120 samples dyed in C.I. Disperse Blue 266 [1].

1) If (temperature is low) and (time is low) and (concentration is low), then (K/S is very low).
2) If (temperature is medium) and (concentration is high), then (K/S is high).

ISBN: 978-989-95079-6-8

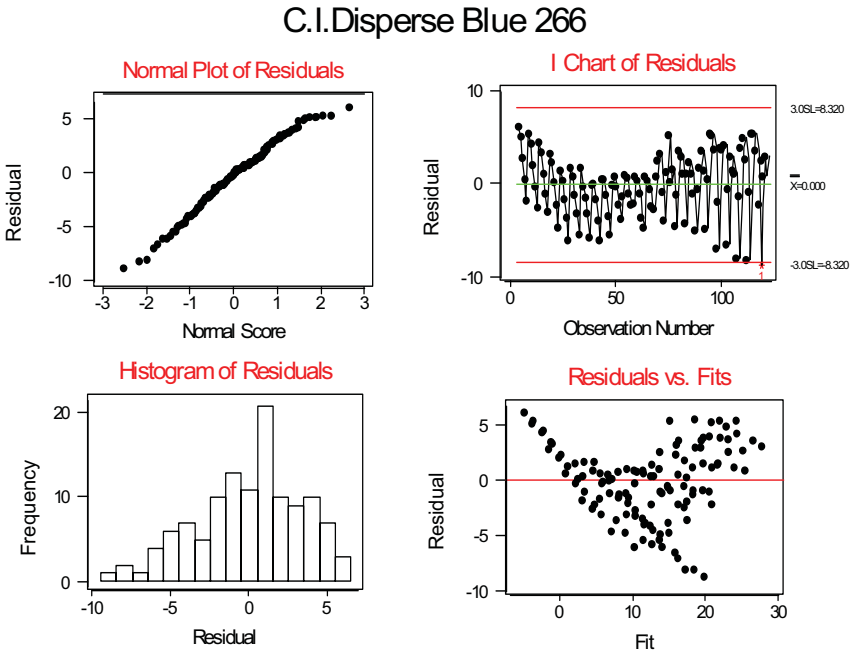


Figure 1: Plot of Residuals for K/S of C.I. Disperse Blue 266 [1].

3) If (temperature is high) and (concentration is low), then (K/S is medium).
4) If (temperature is high) and (concentration is medium), then (K/S is high).
5) If (temperature is low) and (time is high) and (concentration is low), then (K/S is very low).
6) If (temperature is high) and (concentration is high), then (K/S is very high)
7) If (temperature is medium) and (time is low) and (concentration is high), then (K/S is medium).
8) If (temperature is medium) and (time is high) and (concentration is high), then (K/S is high).
9) If (temperature is low) and (time is low) and (concentration is high), then K/S is low.

Then, after a centroid defuzzification of predicted K/S by FIS a Mean Square Errors (MSE) of 3.307 for the FIS model has been obtained [1].

2.3 Optimization of the Initial Fuzzy System

In a final step the fuzzy sets were tuned using the evolutionary algorithm. Hence, input of the CMA-ES algorithm is a vector of parameters of the predefined fuzzy inference system, in this setting the 16 values as given in Table 2. The CMA-ES was used without any modification, i.e. according to the 16 dimensions of the problem, 6 parents breeding 12 offspring were used as recommended. Also the build-in recombination and mutation operators have been applied. The initial step size of the optimization algorithm, also known as mutation strength, was set to 0.1. During the evolutionary loop the fuzzy sets' parameters are successively modified by the CMA-ES and in turn the squared error of the resulting fuzzy inference system is computed for the sample set. The optimization ends after 1000 generations.

3 Results and Interpretation

The FIS model has been improved along the following lines to model the colour yield of C.I. Disperse Blue 266 in polyester high temperature dyeing using an evolutionary algorithm. Table 4 shows parameters of improved fuzzy sets for input variables. As it can be seen in Table 4, concentration has two fuzzy sets in this new model. In the same way, the effect of rule 4 has been tested. Because of its low influence on the results, this rule can be neglected.

Table 4: Parameters of fuzzy sets for input variables.

Fuzzy set	Concentration		Time		Temperature	
	Mean	Std	Mean	Std	Mean	Std
Low	0.53	0.60	12.3	14.21	104.5	9.76
Medium	-	-	-	-	119.2	1.69
High	5.05	1.64	44.1	12.1	129.5	6.10

Regarding the remaining eight rules and changed input parameters according to Table 4, a Mean Square Error (MSE) of 2.333 for evolutionary fuzzy system has been obtained.

Fig. 3 shows the results of the FIS system applied to dyed samples in C.I. Disperse Blue 266.

ISBN: 978-989-95079-6-8

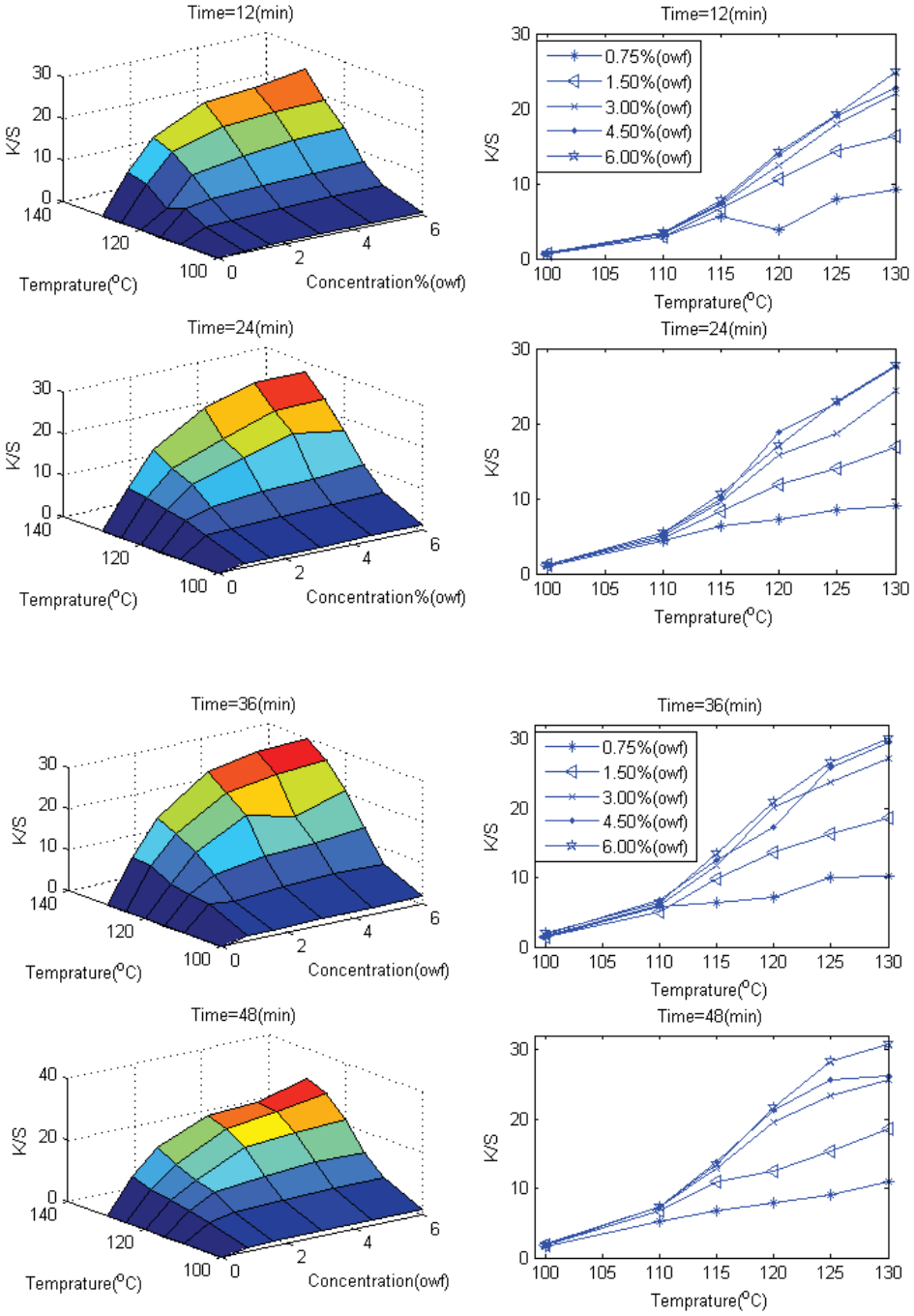


Figure 2: K/S in terms of temperature and concentration [1].

ISBN: 978-989-95079-6-8

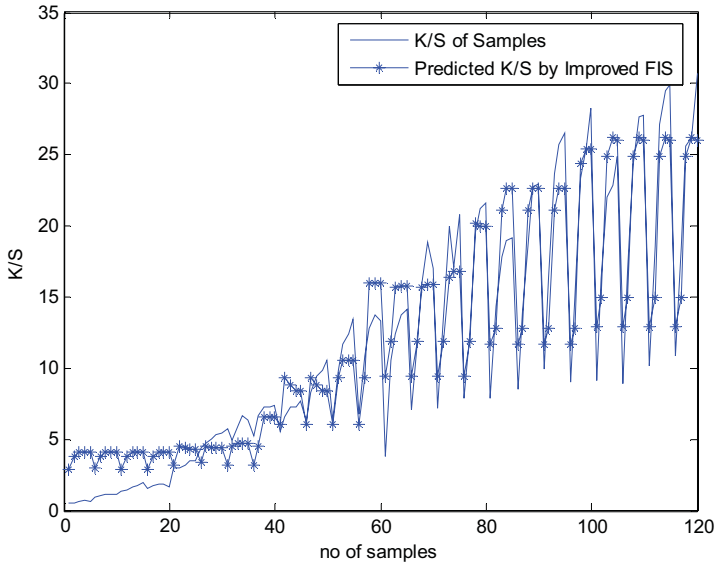


Figure 3: K/S of dyed polyester and predicted K/S by improved FIS.

3.1 Comparison of Statistical Regression and Evolutionary Fuzzy System

Table 5 shows a comparison of the predictive powers by statistical regression, the fuzzy inference system and the evolutionary fuzzy system using MSE. According to this table, the evolutionary fuzzy with minimal MSE and simpler structure showed the best predictive capability in comparison to statistical regression and the fuzzy inference system.

Table 5: Parameters of fuzzy set for output variables.

Method	MSE
Statistical Regression	3.382
Fuzzy Inference System	3.307
Evolutionary Fuzzy System	2.333

On the basis of the above considerations, it can be said that the evolutionary fuzzy system provides an appropriate method to predict the color yield.

4 Conclusions

This research employed statistical regression and an evolutionary fuzzy system to model the (HT) polyester dyeing process. Color yield has been predicted in terms of time, temperature, and disperse dye concentration. We improved the fuzzy sets and rules of the FIS model using an evolutionary algorithm. The results show that the prediction performance is best for the evolutionary fuzzy system.

Acknowledgment

The authors thanks Prof. Hossein Tavanai for his supervision during the first author's work on her MS dissertation when he lent his full support and knowledge to obtain the data used here and to study the behavior of the dyes and also Dr. S. Mahmoud Taheri for his helpful suggestions and assistance.

References

[1] M. Nasiri, S.M. Taheri and H. Tarkesh, Applying Genetic-Fuzzy Approach to Model Polyester Dyeing, Analysis *and Design of Intelligent Systems Using Soft Computing Techniques*, 41: 608-617, 2007.

[2] O. Cordon, F. Herrera, F. Hoffmann and L. Magdalena, Genetic-Fuzzy Systems, Evolutionary *Tuning and Learning of Fuzzy Knowledge Bases*, Advances in Fuzzy Systems - Application and Theory, 19, 2001.

[3] N. Hansen and A. Ostermeier, Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adaptation. IEEE *International Conference on Evolutionary Computation*, 312–317, 1996.

[4] N. Hansen, The CMA Evolution Strategy: A tutorial, 2005. *URL www.bionik.tu-berlin.de/user/niko/cmatutorial.pdf*.

[5] J. Goorden Cook, HandBook *of Textile Fibers, Man Made Fiber*, Merro Technical Library, 1984.

[6] E. R. Trotman, Dyeing *and Chemical Technology of Textile Fibers*; Charles Griffin, 1970.

[7] R.W. Moncrieff , Man-*Made Fibers*, Newnes-Butterworths, 1971.

[8] A. Johnon, The *Theory of Coloration of Textiles*, Second Ed., Bardford: S.D.C. Pub., 1989.

[9] E. Allen, Optical Radiation *Measurements, 2 Color Measurement*, Academic Press, New York, 1980.

[10] S. Kim, A. Kumar, J.L. Dorrity, and G. Vachtsevanos, Fuzzy modeling, control and optimization of textile processes, *Proc 1 Int Jt Conf NAFIPS IFIS NASA*, 32-38,1994.

[11] L.M. Sztandera and C. Pastore, Soft *Computing in Textile Sciences*, Springer, Berlin, 2003.

[12] Mohammad H. Faze1 Zarandi and M. Esmaeilian, A Systematic Fuzzy Modeling for Scheduling of Textile Manufacturing System, *Fuzzy Information Processing Society, NAFIPS 2003. 22nd International Conference of the North American*, 359-364, 2003.

[13] C.C. Hung and W.H. Yu, Control of Dye Concentration, PH, and Temperature in Dyeing Processes, *Textile Research Journal* 69: 914-918, 1999.

ISBN: 978-989-95079-6-8

[14] S. Rautenberg, D. Fanderuff, R. Pacheco, A. Martins, Jose L. Todesco, Richardo M. Barcia, The Use of Fuzzy Sets for Color Recipe Specification in the Textile Print Shop, *Fuzzy Systems Proceedings, IEEE World Congress on Computational Intelligence.,* 885-890, 1998.

[15] M.I. Jahmeerbacusa, N. Kistamahb and R.B. Ramgulamb, Fuzzy Control of Dyebath PH in Exhaust Dyeing, *Coloration Technology*, 120:51-55, 2004.

[16] Abuda etal., Fuzzy- Based Simulation Model for Declorization of Industrial Waste Water, *Journal of Applied Sciences Research*, 4(2): 178-187, 2008

[17] M. Marjoniemi, E. Mantysalo, Neuro-Fuzzy Modeling of Spectroscopic Data. Part A- Modelling of Dye Solutions, *J.S.D.C 113*, 13-17, 1997.

[18] M. Marjoniemi, E. Mantysalo, Neuro-Fuzzy Modeling of Spectroscopic Data. Part B– Dye Concentration Prediction, *J.S.D.C 113*, 64-67, 1997.

[19] Byeong M. Chung, Sung G. Lee, and Che- Seung Cho, Active Tension Control of High Speed Splitting Machines using Fuzzy PID, *IEEE International Conference on Mechatronics*, Taipei, Taiwan, 2005.

[20] H. Tavanai, S.M. Taheri, and M. Nasiri, Modeling of Colour Yield in Polyethylene Terphethalate Dyeing with Statistical and Fuzzy Regression, *Iranian Polymer Journal,* 14: 954-968, 2005.

[21] B. Smith, J. Lu, Improving Computer Control of Batch Dyeing Operations, *American Dyestuff Reporter*, 82(9): 11, 1993.

[22] Soo J. Kim, Eun Y. Kim, K. Jeong and Jee I. Kim, Emotion-Based Textile Indexing Using Colors, Texture and Patterns, *Advances in Visual Computing,* 4292:9-18, 2006.

[23] C. Callhof, B. Wulfhort: Neuro-fuzzy Networks- A Tool For the Development of New Yarn Contacting Exemplary on Friction Discs, *Man Made Fiber Yarn Book* , 95, 1999.

[24] D. Veit, P. Batista-de –Souza, B. Wulfhorst, Application of a Neural Network in the False Twist Texturing Process, *Chemical Fibers International,* 48: 155-156, 1998.

[25] M. Nasiri, Fuzzy Regression Modeling of False-Twist Texturing Yarn, *IFSA 2005 Conference, Beijing, China*, 142-146, 2005.

[26] Y. Guifen, G. Jiansheng, Z. Yongyuan, Predicting the Warp Breakage Rate in Weaving by Neural Network Technicqes, *Textile Research Journal*, 75: 274-278, 2005.

[27] K.J. Chung-Feng, H. Kun-Iuan, W. Yi-Shiuan, Using Neural Network Theory to Predict the Properties of Melt Spun Fibers, *Textile Research Journal*, 74: 840-843, 2004.

[28] L. Jeng-Jong, A Genetic Algorithm for Searching Weaving Parameters for Woven Fabrics, *Textile research journal,* 73: 105-112, 2003.

[29] S. Sette, L. Boullart, Van Langenhof, L.: Building a Rule Set for the Fiber-to-Yarn Production Process by Means of Soft Computing Techniques, Textile Research Journal 70 (2000 Treatment of Explanatory Variables), *European Journal of Operational Research,* 16 : 275-284, 2005.

[30] K. Peeva K., Y. Kyosev, Fuzzy Relational Calculus, *Advanced in Fuzzy Systems- Application and Theory,* 22: 375-386, 2005

[31] M. Blaga, M. Draghici, Application of Genetic Algorithms in Knitting Technology, *The Journal of the Textile Institute*, 97: 175-178, 2004.

[32] S. Thomassey, M. Happiette, J.M. Castelain, An Automatic Textile Sales Forecast Using Fuzzy Treatment of Explanatory Variables, *European Journal of Operational Research,* 16: 275-284, 2005.

[33] W. K. Wong, C.K. Kwong, P.Y. Mok and W.H. Ip, Genetic Optimization of JIT Operation Schedules for Fabric-Cutting Process in Apparel Manufacture., *Journal of Intelligent Manufacturing*, 17(3): 341, 2006.

[34] M. Siddaiah, Michael A. Lieberman, Nadipuram R. Prasad, S.E. Hughs, Automationin Cotton Ginning, *International Journal of Intelligent Systems*, 19 (1-2): 111-129, 2004.

[35] A. Soria-Frisch, M. Koppen, B. Nickolay, Aggregation as Similarity in a Morphological Framework for the Processing of Textile Images *IEEE International Conference on Fuzzy Systems*, 3: 1739-1744, 2004.

[36] D.C. Montgomery, C.A. Peek, Introduction *to Linear Regression Analysis*, Second Ed., J. Wiley, 1991.

[37] H.J. Zimmerman, Fuzzy *Set Theory and Its Applications*, Third Ed. Kluwer Academic Publishers, 1998.

[38] W. Pedrycz, F. Gomidee, An *Introduction to Fuzzy Sets, Analysis and Design*, Prentice-Hall, 1998.

ISBN: 978-989-95079-6-8

Fuzzy concept lattice is made by proto-fuzzy concepts.

Ondrej Krídlo, Stanislav Krajči

Insitute of Computer Science, University of P.J.Šafárik in Košice
Email: ondrej.kridlo@upjs.sk, stanislav.krajci@upjs.sk

Abstract *— An L-fuzzy context is a triple consisting of a set of objects, a set of attributes and an L-fuzzy binary relation between them. An l-cut is a classical context over the same sets with relation as a set of all object attribute pairs, which fuzzy relation assigns truth degree geater or equal than l. Proto-fuzzy concept is a triple made of a set of objects and a set of attributes, which form a concept in some cut of L-fuzzy context and a supremum of all degrees in which cuts this concept exists. Aim of the paper is to show the connection of the structure of proto-fuzzy concepts and fuzzy concept lattice constructed in the way of [1][5]. This connection can help to generate of all fuzzy concepts.*

Keywords — formal concept analysis, fuzzy concept lattice, fuzzy Galois connection

1 Preliminaries

Basic notions of Formal Concept Analysis(FCA) are *formal context* and *formal concept*.

Definition 1 *A formal context $\langle B, A, R\rangle$ consists of a set of objects B, a set of attributes A and a relation R between B and A.*

Definition 2 *Define the mappings $\uparrow\colon {}^B2 \to {}^A2$ and $\downarrow\colon {}^A2 \to {}^B2$. The first assigns to the set $X \subseteq B$ the set of all attributes common to all objects of the set X*

$$\uparrow(X) = \big\{a \in A : \big(\forall o \in X\big)(o,a) \in R\big\}$$

and the second assigns to the set $Y \subseteq A$ the set of all objects common to all attributes of the set Y

$$\downarrow(Y) = \big\{o \in B : \big(\forall a \in Y\big)(o,a) \in R\big\}.$$

Definition 3 *A formal concept of the context $\langle B, A, R\rangle$ is a pair $\langle X, Y\rangle$ such that $X \subseteq B$, $Y \subseteq A$, $\uparrow(X) = Y$ and $\downarrow(Y) = X$.*

Ganter and Wille in [3] showed that the pair of mappings $\big(\uparrow, \downarrow\big)$ is a Galois connection and the composite mappings $\uparrow\downarrow\colon {}^B2 \to {}^B2$ and $\downarrow\uparrow\colon {}^A2 \to {}^A2$ are closure operators. Authors prooved an important theorem in FCA well known as The Basic Theorem On Concept Lattices.

Theorem 1 *(The Basic Theorem on Concept Lattices)The Concept Lattice (lattice of concepts with ordering $\langle X_1, Y_1\rangle \leq \langle X_2, Y_2\rangle$ iff $X_1 \subseteq X_2$ iff $Y_1 \supseteq Y_2$) is a complete lattice in which infimum and supremum are given by*

$$\bigwedge_{i\in I}\langle X_i, Y_i\rangle = \Big\langle \bigcap_{i\in I} X_i, \uparrow\downarrow\Big(\bigcup_{i\in I} Y_i\Big)\Big\rangle$$

$$\bigvee_{i\in I}\langle X_i, Y_i\rangle = \Big\langle \downarrow\uparrow\Big(\bigcup_{i\in I} X_i\Big), \bigcap_{i\in I} Y_i\Big\rangle.$$

A complete lattice V is isomorphic to the concept lattice of some context $\langle B, A, R\rangle$ if and only if there are mappings $\beta : B \to V$ and $\alpha : A \to V$, such that $\beta(B)$ is supremum-dense in V and $\alpha(A)$ is infimum-dense in V and $(o, a) \in R$ is equivalent to $\beta(o) \leq \alpha(a)$ for all $o \in B$ and $a \in A$. In particular V is isomorphic to the concept lattice of context $\langle V, V, \leq\rangle$.

Bělohlávek and Krajči in [1, 2, 5, 6] showed that above mentioned basic notions may be generalized by applying the fuzzy logic.

Everybody knows that reality provides situations where many of attributes are rather fuzzy than crisp. Answer of question "Does the object has the attribute?" is rather somewhere in the middle of false (0) and true (1).

Definition 4 An L-fuzzy formal context *is a triple $\langle B, A, r\rangle$ consists of a set of objects B, a set of attributes A and an L-fuzzy binary relation r, i.e. the L-fuzzy subset of $B \times A$ or mapping from $B \times A$ to L, where L is a complete residuated lattice.*

The class of all L-fuzzy sets in X will be denoted by ${}^XL$. If L is complete then the relation $\subseteq$ (defined by $f \subseteq g$ iff $f(x) \leq g(x)$ for all $x \in X$) makes ${}^XL$ into a complete lattice.

Definition 5 *A* complete residuated lattice *is an algebra $\boldsymbol{L} = \langle L, \wedge, \vee, \otimes, \to, 0, 1\rangle$ where*

(a) *$\langle L, \wedge, \vee, 0, 1\rangle$ is a complete lattice with the least element 0 and the greatest element 1,*

(b) *$\langle L, \otimes, 1\rangle$ is a commutative monoid,*

(c) *$\otimes$ and $\to$ satisfy adjointness, i.e.*

$$a \otimes b \leq c \iff a \leq b \to c$$

for each $a, b, c \in L$ ($\leq$ is the lattice ordering).

Definition 6 (Bělohlávek) *A triple $\langle B, A, r\rangle$ is an L-fuzzy context where $r : B \times A \to L$ and L is a complete residuated lattice. Define mappings $\uparrow\colon {}^BL \to {}^AL$ and $\downarrow\colon {}^AL \to {}^BL$ such that for every $f \in {}^BL$ and $g \in {}^AL$*

$$\uparrow(f)(a) = \bigwedge_{o\in B}\big(f(o) \to r(o,a)\big)$$

$$\downarrow(g)(o) = \bigwedge_{a\in A}\big(g(a) \to r(o,a)\big).$$

ISBN: 978-989-95079-6-8

Table 1: Example of L-fuzzy formal context.

	a_1	a_2	a_3	a_4
o_1	0,2	0,8	0,8	1
o_2	1	1	0,8	1
o_3	0,2	0,6	0,4	1
o_4	0,8	0,4	0,2	0
o_5	0,2	0,4	0,4	0,2

Table 2: 1-cut, 0,6-cut, 0,2-cut

	a_1	a_2	a_3	a_4
o_1				•
o_2	•	•		•
o_3				•
o_4				
o_5				

	a_1	a_2	a_3	a_4
o_1		•	•	•
o_2	•	•	•	•
o_3		•		•
o_4	•			
o_5				

	a_1	a_2	a_3	a_4
o_1	•	•	•	•
o_2	•	•	•	•
o_3	•	•	•	•
o_4	•	•	•	
o_5	•	•	•	•

The aim of this paper is to give a new equivalent definition of Belohlavek's mappings, but using so-called protto-fuzzy concepts as basic building units for constructing the fuzzy concepts. In the next section an l-cuts of L-fuzzy contexts will be defined for any truth degree $l \in L$, main properties of their concepts and a relationship of concepts of different cuts will be showed. Finally Proto-fuzzy concepts will be defined. Then an equality of new mappings with Bělohlávek ones will be showed.

2 Cuts of L-fuzzy context

Lets have an example of L-fuzzy context (Table 1). In our example is the lattice of truth degrees is $\langle\{1;0,8;0,6;0,4;0,2;0,\};\leq\rangle$.

Definition 7 *Let $l \in L$ be an arbitrary truth degree. An l-*cut *of some L-fuzzy set $f \in {}^X L$ is a classical set denoted by*

$$f_l = \{x \in X : f(x) \geq l\}.$$

Definition 8 *An l-*cut *of the L-fuzzy formal context $\langle B, A, r\rangle$ for $l \in L$ is the classical context $\langle B, A, r_l\rangle$ where $r_l = \{(o,a) \in B \times A : r(o,a) \geq l\}$.*

Some of cuts of our example are in the tables.

Definition 9 *For every truth value $l \in L$ lets define mappings $\uparrow_l$: ${}^B2 \to {}^A2$ and $\downarrow_l$: ${}^A2 \to {}^B2$. For every object or attribute subset $X \subseteq B$ and $Y \subseteq A$ put*

$$\uparrow_l (X) = \{a \in A : (\forall o \in X) r(o,a) \geq l\}$$

$$\downarrow_l (Y) = \{o \in B : (\forall a \in Y) r(o,a) \geq l\}.$$

Lemma 1 *Let $K \subseteq L$ be an arbitrary subset of truth degrees. Then for every set of objects $X \subseteq B$ and attributes $Y \subseteq A$ holds that*

$$\uparrow_{(\bigvee K)} (X) = \bigcap_{l \in K} \uparrow_l (X)$$

$$\downarrow_{(\bigvee K)} (Y) = \bigcap_{l \in K} \downarrow_l (Y).$$

Proof:

$\boxed{\subseteq}$ Cuts of the L-fuzzy context were defined such that for every $l_1, l_2 \in L$ if $l_1 \leq l_2$ then $r_{l_1} \supseteq r_{l_2}$. Hence for every subset of objects X or subset of attributes Y, $\uparrow_{l_1} (X) \supseteq \uparrow_{l_2} (X)$ and $\downarrow_{l_1} (Y) \supseteq \downarrow_{l_2} (Y)$, which for every $l \in K$ implies $\uparrow_{\bigvee K} (X) \subseteq \uparrow_l (X)$ and $\downarrow_{\bigvee K} (Y) \subseteq \downarrow_l (Y)$. And from above we have $\uparrow_{(\bigvee K)} (X) \subseteq \bigcap_{l \in K} \uparrow_l (X)$ and $\downarrow_{(\bigvee K)} (Y) \subseteq \bigcap_{l \in K} \downarrow_l (Y)$.

$\boxed{\supseteq}$ Let a be an arbitrary attribute from $\bigcap_{l \in K} \uparrow_l (X)$. For all $l \in K$ and for every object $o \in X$, $r(o,a) \geq l$. From the properties of supremum is $r(o,a) \geq \bigvee K$ for all objects $o \in X$. It means that $a \in \uparrow_{(\bigvee K)} (X)$. Hence $\bigcap_{l \in K} \uparrow_l (X) \subseteq \uparrow_{(\bigvee K)} (X)$.

The second part can be prooved dually. □

Lemma 2 *For all $l \in L$ the pair $(\uparrow_l, \downarrow_l)$ forms a Galois connection between the power-set lattices ${}^B2$ and ${}^A2$.*

Now lets define the concept on the l-cut for some truth degree $l \in L$.

Definition 10 *Let $\langle B, A, r\rangle$ be the L-fuzzy context. A pair $\langle X, Y\rangle$ is called an l-*concept *iff*

$$\uparrow_l (X) = Y, \text{ and } \downarrow_l (Y) = X,$$

hence the pair is a concept in a classical context $\langle B, A, r_l\rangle$. The set of all l-concepts will be assigned $\mathrm{C}_l(B, A, r)$, shortly C_l.

2.1 *Relationship of concepts in different cuts*

Lemma 3 *Let $l_1, l_2 \in L$ be an arbitrary truth values, such that $l_2 \leq l_1$. Let $\langle X, Y\rangle \in \mathrm{C}_{l_1}$. Then there exists an interval $\mathcal{I}$ in concept lattice C_{l_2}, such that for every l_2-concept $\langle Z, W\rangle \in \mathcal{I}$ holds that $X \subseteq Z$ and $Y \subseteq W$.*

Proof: From $\langle X, Y\rangle \in \mathrm{C}_{l_1}$ we know that $\uparrow_{l_1} (X) = Y$ and $\downarrow_{l_1} (Y) = X$. So as the greatest element of wanted interval we can use the $\langle \downarrow_{l_2}\uparrow_{l_2} (X), \uparrow_{l_2} (X)\rangle$ and the least one $\langle \downarrow_{l_2} (Y), \uparrow_{l_2}\downarrow_{l_2} (Y)\rangle$. From the fact $r_{l_1} \subseteq r_{l_2}$ we have inclusions

$$Y = \uparrow_{l_1} (X) \subseteq \uparrow_{l_2} (X)$$

$$X = \downarrow_{l_1} (Y) \subseteq \downarrow_{l_2} (Y).$$

From closure property of conclusion of mappings we have

$$\downarrow_{l_2}\uparrow_{l_2} (X) \supseteq X \text{ and } \uparrow_{l_2}\downarrow_{l_2} (Y) \supseteq Y.$$

From $Y = \uparrow_{l_1} (X) \subseteq \uparrow_{l_2} (X)$ we have

$$\downarrow_{l_2} (Y) \supseteq \downarrow_{l_2}\uparrow_{l_2} (X)$$

and from properties of concepts from [3] we know that it is equivalent to

$$\uparrow_{l_2}\downarrow_{l_2} (Y) \subseteq \uparrow_{l_2} (X).$$

□

ISBN: 978-989-95079-6-8

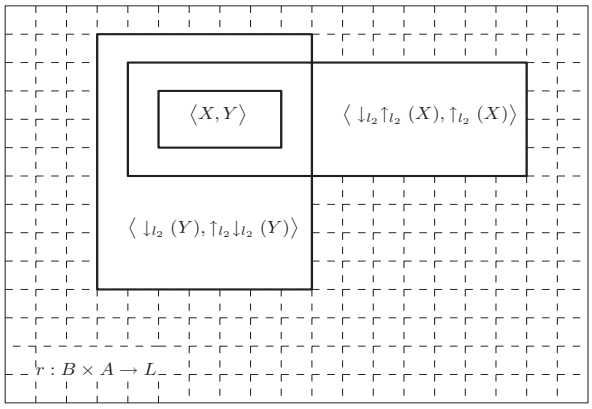


Figure 1: l_1-concept and its l_2-superconcepts

Definition 11 *Define an ordering "$\preceq$" on the set of all l-concepts of all cuts of L-fuzzy context. Lets have two concepts of two different cuts $\langle X, Y\rangle \in \mathrm{C}_{l_1}$ and $\langle Z, W\rangle \in \mathrm{C}_{l_2}$ (because of these pairs are concepts in two different cuts so we can assign them $\langle X, Y\rangle_{l_1}$ and $\langle Z, W\rangle_{l_2}$) then $\langle X, Y\rangle_{l_1} \preceq \langle Z, W\rangle_{l_2}$ iff $X \subseteq Z$ and $Y \subseteq W$ and $l_2 \leq l_1$.*

The lattice in the picture 3 isn't the concept lattice. It is the lattice of all l-concepts for all $l \in L$ and lines is assigning new ordering from definition.

Lemma 4 *Let $l_1, l_2 \in L$ are arbitrary truth values such that $l_2 \leq l_1$. Let $\langle X_1, Y_1\rangle, \langle X_2, Y_2\rangle \in \mathrm{C}_{l_1}$, such that $\langle X_1, Y_1\rangle \preceq \langle X_2, Y_2\rangle$. Than the geatest and least elements of corresponding intervals of l_2-concepts are ordered same as corresponding l_1-concepts.*

Proof: If $X_1 \subseteq X_2$ from properties of closure operator we have $\downarrow_{l_2}\uparrow_{l_2}(X_1) \subseteq \downarrow_{l_2}\uparrow_{l_2}(X_2)$ And from $Y_2 \subseteq Y_1$ we have $\downarrow_{l_2}(Y_1) \subseteq \downarrow_{l_2}(Y_2)$. □

3 Proto-fuzzy concepts

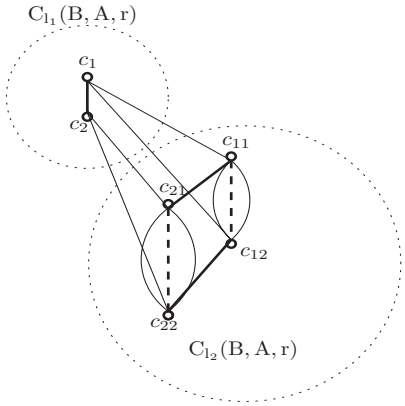


Figure 2: The relationship of concepts of different cuts

As we can see in the figure 3, some of l-concepts are equal, but in different cuts. If we fix some l-concept and look on the set of truth degrees in which cuts the conceot exists, we can see two important properties described in next theorems.

Theorem 2 is sayng that for every concept of some cut, a set of all truth degrees in which cut the concept exists is closed under its supremum.

Theorem 2 *Let $K \subseteq L$ be an arbitrary set of truth degrees, $\langle X, Y\rangle \in \mathrm{C}_l$ for all $l \in K$. Then $\langle X, Y\rangle \in \mathrm{C}_{\bigvee K}$.*

Proof: The lemma 1 implies

$$\uparrow_{\bigvee K}(X) = \bigcap_{l\in K} \uparrow_l (X) = \bigcap_{l\in K} Y = Y,$$

$$\downarrow_{\bigvee K}(Y) = \bigcap_{l\in K} \downarrow_l (Y) = \bigcap_{l\in K} X = X.$$

Hence $\langle X, Y\rangle \in \mathrm{C}_{\bigvee K}$. □

Next theorem 3 is saying that if some concept exist in two different cuts, then exists in every cut between them.

Theorem 3 (Convexity) *Let $l_1, l_2 \in L$ be an arbitrary truth degrees, and let $\langle X, Y\rangle \in \mathrm{C}_{l_1} \cap \mathrm{C}_{l_2}$. Then for all $l \in L$, such that $l_1 \leq l \leq l_2$, $\langle X, Y\rangle \in \mathrm{C}_l$.*

Proof: The lemma 1 implies, that for every set of object X and any two arbitrary truth degrees $k, m \in L$, such that $k \leq m$ holds

$$\uparrow_k (X) \subseteq \uparrow_k (X) \cap \uparrow_m (X) = \uparrow_{\bigvee\{k,m\}} (X) = \uparrow_m (X)$$

and

$$\downarrow_k (Y) \subseteq \downarrow_k (Y) \cap \downarrow_m (Y) = \downarrow_{\bigvee\{k,m\}} (Y) = \downarrow_m (Y).$$

So

$$Y = \uparrow_{l_1} (X) \supseteq \uparrow_l (X) \supseteq \uparrow_{l_2} (X) = Y,$$

$$X = \downarrow_{l_1} (Y) \supseteq \downarrow_l (Y) \supseteq \downarrow_{l_2} (Y) = X.$$

Hence $\uparrow_l (X) = Y$ and $\downarrow_l (Y) = X$, which implies $\langle X, Y\rangle \in \mathrm{C}_l$. □

Definition 12 *Let $\langle X, Y\rangle \in \bigcup_{k\in L} \mathrm{C}_k(B, A, r)$ be the concept of some cut of the fuzzy context $\langle B, A, r\rangle$. Triple $\langle X, Y, l\rangle$ such that $l = \bigvee\{k \in L : \langle X, Y\rangle \in \mathrm{C}_k(B, A, R)\}$ will be called a* proto-fuzzy concept. *The set of all proto-fuzzy concepts will be denoted by* $\mathrm{PFC}(B, A, r)$.

Definition 13 *Define a mapping* $\mathrm{pd} : {}^B2 \times {}^A2 \to L$ *such that for every set of objects $Z \subseteq B$ and set of attributes $W \subseteq A$ is*

$$\mathrm{pd}(Z, W) = \bigvee\{l \in L : (\exists X \subseteq B)(\exists Y \subseteq A)$$

$$\langle X, Y, l\rangle \in \mathrm{PFC}(B, A, r) Z \subseteq XW \subseteq Y\}.$$

This mapping assigns to every pair of sets of objects and attributes the truth degree of highest proto-fuzzy concept which owns them. Notation pd means proto-degree of input sets.

4 Alternative definition of Bělohlávek's mappings

Lets go back to introduction.

Definition 14 *A triple $\langle B, A, r\rangle$ is L-fuzzy context where $r : B \times A \to L$ and L is the complete residuated lattice. Define mappings $\uparrow: {}^BL \to {}^AL$ and $\downarrow: {}^AL \to {}^BL$ such that for every $f \in {}^BL$ and $g \in {}^AL$*

$$\uparrow (f)(a) = \bigwedge_{o\in B} \big(f(o) \to r(o, a)\big)$$

$$\downarrow (g)(o) = \bigwedge_{a\in A} \big(g(a) \to r(o, a)\big).$$

ISBN: 978-989-95079-6-8

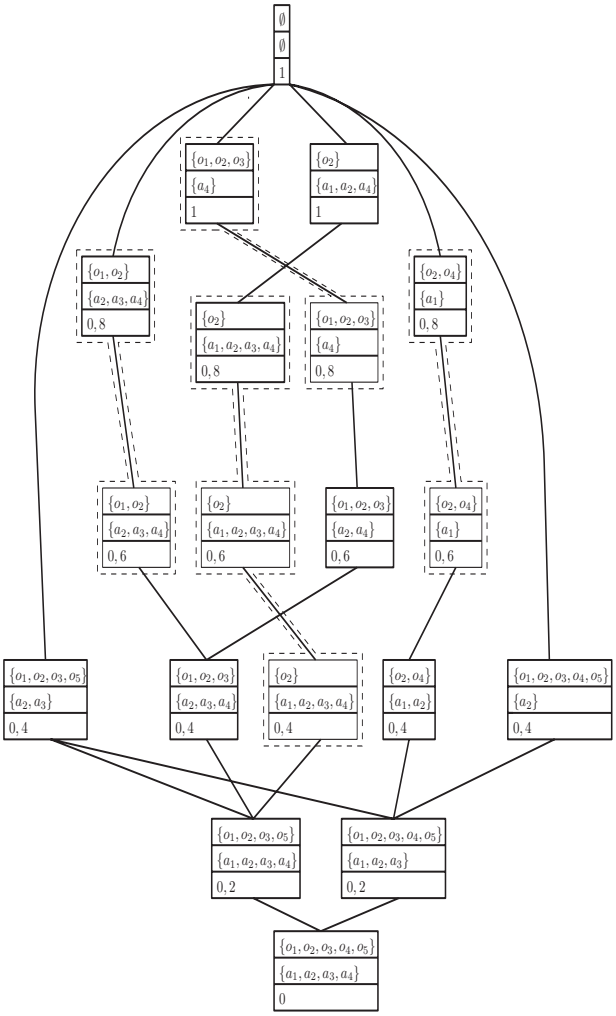


Figure 3: Lattice of all l-concepts of example for all $l \in L$, $\left\langle \cup_{l\in L} \mathbf{C}_l(B, A, r), \preceq \right\rangle$

Lets define new mappings and show that they are equivalent to the mappings above.

Definition 15 *A triple $\langle B, A, r\rangle$ is L-fuzzy context where $r : B \times A \to L$ and L is the complete residuated lattice. Define mappings $\Uparrow$: ${}^B L \to {}^A L$ and $\Downarrow$: ${}^A L \to {}^B L$ such that for every $f \in {}^B L$ and $g \in {}^A L$*

$$\Uparrow(f)(a) = \bigwedge_{l\in rng(f)} \big(l \to \mathrm{pd}(f_l, \{a\})\big)$$

$$\Downarrow(g)(o) = \bigwedge_{l\in rng(g)} \big(l \to \mathrm{pd}(\{o\}, g_l)\big).$$

Theorem 4 *For every set Z of objects, an every set W of attributes,*

$$\bigwedge_{(o,a)\in Z\times W} r(o, a) = \mathrm{pd}(Z, W).$$

Proof:

$\boxed{\leq}$ Let $l = \bigwedge_{(o,a)\in Z\times W} r(o, a)$. If $a \in W$ then $(\forall o \in Z) r(o, a) \geq l$, i.e. $a \in{\uparrow_l}(Z)$, so $W \subseteq{\uparrow_l}(Z)$. Take $Y = {\uparrow_l}(Z)$ and $X = {\downarrow_l}(Y)$ clearly $\langle X, Y\rangle \in \mathbf{C}_l(B, A, r)$. Hence $\langle X, Y, m\rangle \in \mathrm{PFC}(B, A, r)$ where $m = \bigvee\{k \in L : \langle X, Y\rangle \in \mathbf{C}_k(B, A, r)\}$.

Because $W \subseteq{\uparrow_l}(Y)$ and $Z \subseteq{\downarrow_l}{\uparrow_l}(Z) = {\downarrow_l}(Y) = X$, we have

$$\mathrm{pd}(Z, W) \geq m \geq l = \bigwedge_{(o,a)\in Z\times W} r(o, a).$$

$\boxed{\geq}$ Let $\langle X, Y, l\rangle \in \mathrm{PFC}(B, A, r)$, $Z \subseteq X$ and $W \subseteq Y$. Then $l = \bigvee\{k \in L : \langle X, Y\rangle \in \mathbf{C}_k(B, A, r)\}$ and it follows from Theorem 2 that $\langle X, Y\rangle \in \mathbf{C}_l(B, A, r)$. It means that, for all $o \in X$ and $a \in Y$ is $r(o, a) \geq l$
Hence $\bigwedge_{(o,a)\in Z\times W} r(o, a) \geq \bigwedge_{(o,a)\in X\times Y} \geq l$. It follows that $\bigwedge_{(o,a)\in Z\times W} r(o, a) \geq \bigvee\{l \in L : (\exists X \subseteq B)(\exists Y \subseteq A)\langle X, Y, l\rangle \in \mathrm{PFC}(B, A, r), Z \subseteq X, W \subseteq Y\} = \mathrm{pd}(Z, W)$. □

Theorem 5 *For above defined mappings holds*

$$\uparrow = \Uparrow \text{ and } \downarrow = \Downarrow.$$

Proof: Note these three facts:

- Because $\to$ is antitone in the first argument, $f(o) \geq l$ implies
$$f(o) \to r(o, a) \leq l \to r(o, a)$$
- Because $\{o \in B : f(o) \geq l\} \supseteq \{o \in B : f(o) = l\}$, we have
$$\bigwedge_{o\in B: f(o)\geq l} (l \to r(o, a)) \leq \bigwedge_{o\in B: f(o)=l} (l \to r(o, a))$$
- $$\bigcup_{l\in \mathrm{rng}(f)} \{o \in B : f(o) \geq l\} =$$
$$= \bigcup_{l\in \mathrm{rng}(f)} \bigcup_{m\in \mathrm{rng}(f): m\geq l} \{o \in B : f(o) = m\} =$$
$$= \bigcup_{l\in \mathrm{rng}(f)} \{o \in B : f(o) = l\}$$

Using the previous facts we obtain

$$\bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)=l} (f(o) \to r(o, a)) =$$
$$= \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)\geq l} (f(o) \to r(o, a)) \leq$$
$$\leq \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)\geq l} (l \to r(o, a)) \leq$$
$$\leq \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)=l} (l \to r(o, a)) =$$
$$= \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)=l} (f(o) \to r(o, a))$$

It follows that both inequalities are in fact equalities, hence (using Theorem 4)

$$\uparrow(f)(a) = \bigwedge_{o\in B} (f(o) \to r(o, a)) =$$

ISBN: 978-989-95079-6-8

$$= \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)=l} (f(o) \to r(o,a)) =$$

$$= \bigwedge_{l\in \mathrm{rng}(f)} \bigwedge_{o\in B: f(o)\geq l} (l \to r(o,a)) =$$

$$= \bigwedge_{l\in \mathrm{rng}(f)} (l \to \bigwedge_{o\in B: f(o)\geq l} r(o,a)) =$$

$$= \bigwedge_{l\in \mathrm{rng}(f)} (l \to \bigwedge_{o\in f_l} r(o,a)) =$$

$$= \bigwedge_{l\in \mathrm{rng}(f)} (l \to \mathrm{pd}(f_l, \{a\})) = \Uparrow (f)(a)$$

The second part can be proved dually. □

4.1 *One-sided fuzzy concepts*

In [8] we defined so-called one-sided fuzzy concepts, the pairs consisting of one classical set of objects and fuzzy set of attributes and defined mappings which are creating them.

Definition 16 *For L-context* $\langle B, A, r\rangle$ *define mappings* $\uparrow\uparrow$: ${}^B2 \to {}^AL$ *and* $\downarrow\downarrow$: ${}^AL \to {}^B2$. *For an arbitrary set of objects* $X \in {}^B2$ *and an L-fuzzy set of attributes* $g \in {}^AL$ *put*

$$\uparrow\uparrow (X)(a) = \bigwedge_{o\in X} r(o,a)$$

$$\downarrow\downarrow (g) = \{o \in B : (\forall a \in A) g(a) \leq r(o,a)\}.$$

By the theorem 4 we can write $\uparrow\uparrow (X)(a) = \mathrm{pd}(X, \{a\})$. Bělohlávek's theorem ([2]) say:

Theorem 6 *For* $X \in {}^B2$ *and* $g \in {}^AL$ *and mappings* $\uparrow, \downarrow, \uparrow\uparrow, \downarrow\downarrow$ *we have*

$$\uparrow\uparrow (X) = \uparrow (X') \text{ and } \downarrow\downarrow (g) = (\downarrow (g))_1$$

and X' *means L-fuzzy set of objects corresponding to* X, *and* $(\downarrow (g))_1$ *means* 1*-cut of* $\downarrow (g)$.

With the theorem 5 we can in previous theorem change the mappings $\uparrow, \downarrow$ by $\Uparrow, \Downarrow$.

5 Sketch of algorithm for generating all fuzzy concepts

```
Set<Object> B;
Set<Attribute> A;
Set<TrueDegree> L;
LFuzzy binary relation r;
Set<PFConcept> PC(B,A,r)
Set<FConcept> fcs;

public Set<FuzzyConcept> generateAllFC(
        Set<Proto-fuzzy concept> PC(B,A,r)
 ){

// generating all basic L-fuzzy concepts

 for ( PFConcept pc : PC(B,A,r) ){
  Set<Object> objs = pc.getObjects();
  Set<Attribute> atrs = pc.getAttributes();
  True Degree deg = pc.getTruthDegree();
   for ( TruthDegree m : L )
    for ( TruthDegree k : L )
     if ( m== k --> deg ){
      //new L-fuzzy set of Objects
      LFSObjs f = new LFSObjs(
        <obj,m> if obj : objs,
        <obj,0> if obj : B-objs
      );
      //new L-fuzzy set of Attributes
      LFSAtrbs g = new LFSAtrbs(
        <atr,k> if atr : atrs,
        <atr,0> if atr : A-atrs
      );
//every fuzzy concept will remeber of which
//of proto fuzzy concepts was created
      fcs.add(
       new FConcept(
                f ,
                g ,
                new Set<PFConcept>{ pfc } )
             );
    }
  }

// creating connected L-fuzzy concepts

 for ( LFuzzyConcept fc1 : fcs )
  for ( LFConcept fc2 : fcs ){
   Set<PFConcept> pfcs1 = fc1.getPFConcepts();
   Set<PFConcept> pfcs2 = fc2.getPFConcepts();
   boolean ordered = true;
   for ( PFConcept pfc1 : pfcs1 )
    for ( PFConcept pfc2 : pfcs2 )
     if ( !pfc1 <= pfc2 AND !pfc2 <= pfc2 )
      ordered = false;
   if ( ordered )
    fcs.add( createNewFC( fc1 , fc2 ) );
  }

  return fcs;
}

public FConcept createNewFConcept(
   FConcept fc1 ,
   FConcept fc2 )
{
//new L-fuzzy set of Objects
LFSObjs f = new LFSObjs(
 <obj,fc1.getDeg(obj)>
   if fc1.getDeg(obj)>=fc2.getDeg(obj),
 <obj,fc2.getDeg(obj)>
   if fc2.getDeg(obj)>=fc1.getDeg(obj)
);
//new L-fuzzy set of Attributes
LFSAtrbs g = new LFSAtrbs(
 <atr,fc1.getDeg(atr)>
   if fc1.getDeg(atr)>=fc2.getDeg(atr),
 <atr,fc2.getDeg(atr)>
   if fc2.getDeg(atr)>=fc1.getDeg(atr)
);
//Union of sets of proto-fuzzy concepts
Set<PFConcept> pfc =
  unite(
     fc1.getPFConcepts,
```

ISBN: 978-989-95079-6-8

```
      fc2.getPFConcepts
        )

return new FConcept( f , g , pfc );
}
```

6 Future work

Our future work will be to finish the sketched algorithm, to proove his good working and apply it.

We are grateful for precious comments of our colleague and friend Jozef Pócs.

Paper was created with support of grant 1/3129/06 Slovak grant agency VEGA.

References

[1] Bělohlávek R. Lattices Generated By Binary Fuzzy Relations Tatra Mountains Mathematical Publications 16 (1999),11-19

[2] Bělohlávek R., Vychodil V. What is fuzzy concept lattice? Proceedings of 3rd international conference on Concept Lattices and Their Applications 2005 (CLA 2005), 34 - 45

[3] Ganter, B. and Wille, R. 1997 Formal Concept Analysis: Mathematical Foundations. 1st. Springer-Verlag New York, Inc.

[4] Krídlo O., Krajči S. Proto-fuzzy concepts, their retrieval and usage Proceedings of Sixth International Conference on Concept Lattices and Their Applications (CLA 2008), 83 - 96

[5] Krajči, S., A Generalized Concept Lattice, Logic Journal of the IGPL, 2005, 543-550

[6] Krajči S., The Basic Theorem on Generalized Concept Lattice, CLA 2004,

[7] Snášel V., Ďuráková D., Krajči S., Vojtáš P.: Merging Concept Lattices of α-cuts Of Fuzzy Contexts Contributions To General Algebra 14 Proceedings of the Olomouc Conference 2002 (AAA 64) and the Postdam Conference 2003 (AAA 65) Verlag Johanes Heyn, Klagenfurt 2004

[8] Krajči, S.: Cluster Based Efficient Generation Of Fuzzy Concepts Neural Network World 5/03 521–530

ISBN: 978-989-95079-6-8

A Preliminary Analysis for Improving Model Structure of Fuzzy Habitat Preference Model for Japanese Medaka (*Oryzias latipes*)

Shinji Fukuda[1]

1. Institute of Tropical Agriculture, Kyushu University
Fukuoka, Japan
Email: shinji-fkd@agr.kyushu-u.ac.jp

Abstract—*The present study examined a preliminary analysis for improving model structure of fuzzy habitat preference model for Japanese medaka (Oryzias latipes) dwelling in agricultural canals in Japan. The present model employed a simplified fuzzy reasoning method for evaluating habitat preference of the fish based on the relationship with physical habitat characteristics observed in the field survey. The model parameter was optimized by using a simple genetic algorithm, in which number of fuzzy membership function was fixed. In the present analysis, number of fuzzy membership function was changed while the other methods were fixed as the original model. The model performance was evaluated based on mean square error between observed and predicted fish population density, and by using two different data sets. As a result, there was no clear tradeoff between number of fuzzy membership functions and prediction accuracy. By contrast, calibration and validation results showed a slight tendency of tradeoff. Further studies on clarifying the tradeoffs would be necessary for improving the model structure in an effective way.*

Keywords— fish habitat, genetic algorithm, habitat modelling, simplified fuzzy reasoning

1 Introduction

Since anthropogenic impacts largely affect and degrade aquatic environments, assessment based on geomorphologic and ecological data is the basis for evaluating the state of the environment and its management planning [1]. In the evaluation, a habitat preference approach, such as habitat evaluation procedure (HEP) [2, 3] and Instream Flow Incremental Methodology (IFIM) [4], is often applied for its simplicity and intuitiveness. Habitat suitability index (HSI), which was originally proposed for HEP and IFIM, is usually classified into three categories: Category I—professional judgements; Category II—habitat use indices; Category III—habitat preference indices [5]. Despite its generality, some studies concluded that HSI cannot describe habitat uses [6, 7, 8, 9]. This failure would be ascribed to non-linear and complex interactions between environmental factors affecting the habitat use and to uncertainty in ecology.

The uncertainty is increasingly recognized and gaining public interests in natural ecosystem research. In ecological modelling, the uncertainty of the model is caused by both the lack of knowledge (i.e. data imperfection) and the variability of models and parameters (models' sensitivity) [10]. For coping with the data imperfection, an approach using fuzzy logic was proposed and widely applied to expressing expert knowledge and dealing with incomplete and/or subjective information [11, 12, 13, 14, 15, 16]. This approach enables qualitative information consisting of linguistic terms to be used for quantitative evaluation of target systems that are required for decision-making at an ecosystem level. By contrast, [17] introduced fuzziness as a reliability bound so as to consider the subjective uncertainty in habitat evaluation by different interpretation. The author also applied fuzzy set theory to habitat evaluation for considering the vagueness of fish behaviour, measurement errors and dispersions of physical environment, in which models showed high ability to represent habitat use of the target fish by integrating fuzzy rule-based model with model optimization techniques of genetic algorithm (GA) [18, 19, 20, 21, 22, 23, 24] or artificial neural network [25, 26]. By applying GA, all the model parameters can be simultaneously optimized, which enables us to assess the nonlinearity in fish habitat preference [23]. This emerging property clarified from the modelling would make an appeal to researchers for applying artificial intelligence techniques. However, as reported in [22], the fuzzy habitat preference model showed weakness in terms of transferability which might be ascribed to the complexity of the models. Therefore, to assess the relationship between models' complexity and accuracy would be necessary for the model improvement without losing its accuracy. This could also promote the application of genetic fuzzy systems in ecological research.

The present study aims to examine a preliminary analysis for improving model structure of fuzzy habitat preference model (FHPM) for Japanese medaka (*Oryzias latipes*) dwelling in agricultural canals in Japan. Field surveys were conducted in the agricultural canals to establish a relationship between fish habitat preference and physical environments of water depth, current velocity, lateral cover ratio and percent vegetation coverage. All the FHPMs of different model structure were compared focusing on the tradeoffs between complexity of the model and accuracy of habitat prediction.

2 Methods

In this section, the author first gives a brief review on the application of fuzzy systems to ecological research. Second, an overview of the study site and target fish is described. And then, the author explains the field surveys and the modelling procedure of habitat prediction models together with the explanation on assessment and comparison of the models.

2.1 Fuzzy Modelling in Ecological Research

The pioneer work of fuzzy logic application to ecological study would be [27] in which fuzzy set theory was introduced to habitat quality evaluation, classification of wetlands, and formation of compartments of ecosystem components. Since

ISBN: 978-989-95079-6-8

then, many researchers have employed fuzzy logic-based approaches in ecological research (e.g., population dynamics [28, 29, 30, 31], cluster analysis [32], elicitation of expert knowledge [33], human decision behaviour [34]). Reference [35] reviewed applications of fuzzy logic for decision support in ecosystem management with focus on the identification, optimization, validation, interpretability and uncertainty aspects of rule-based models. As mentioned earlier, fuzzy modelling has also been applied to habitat assessment. This can be classified into two categories of knowledge-based modelling [12, 13, 14, 15] and data-driven modelling [16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 36]. The former cannot optimize model parameters, while the latter can do it. Several approaches were employed for data-driven optimization of model parameters. For instance, [16] employed a nearest ascent hill climbing algorithm, [22, 25, 26] incorporated fuzzy membership function into artificial neural network, and the others employed GA [18, 19, 20, 21, 22, 23, 24, 36]. According to [37], the latter two approaches can be classified into genetic fuzzy systems. The genetic fuzzy systems would gain more attentions in ecological research because of its interpretability and learning ability.

2.2 Study Site and Target Fish

The survey was carried out in an agricultural canal located in Kurume City, Fukuoka, Japan (33˚20’ N, 130˚42’ E; Fig. 1). The spring-fed canal runs through paddy fields, and is used for both irrigation and drainage purposes. It flows into the Kose River, which is a tributary of the Chikugo River.

The target fish, Japanese medaka (*Oryzias latipes*), is one of the most common freshwater fish in Japan. This fish has been considered as one of the symbols of rural environmental conservation and restoration because of the vulnerability to alterations of physical environment such as concrete lining of the earthen canal. For instance, Japanese medaka generally grows up to approximately 2 cm in length and thus is vulnerable to fast flowing water.

2.3 Field Survey

The field surveys were conducted on 14 October, and 5 and 9 November 2004. The surveys were conducted during a non-irrigation period, thus the discharge in the study reaches remained stable. Therefore, the habitat uses of Japanese medaka were not affected by any agricultural activities or agricultural chemicals. The water temperature remained stable (16.1–20.3˚C) during the surveys.

The habitat uses of Japanese medaka and the physical habitat characteristics of water depth (henceforth referred to as depth), current velocity (velocity), lateral cover ratio (cover), and percent vegetation coverage (vegetation) in the study reach were surveyed on sunny days. The four physical environments are found to be the primary factors affecting spatial distribution, i.e., habitat use, of the fish [20]. After mapping the reach, habitat use of the fish was first observed (11:00–14:30) and then the physical habitat characteristics within the reach were surveyed.

The habitat use of Japanese medaka was observed visually from the bank; the observer moved slowly and carefully to avoid disturbing the fish. The number of the fish was counted in units of five to take into consideration the habit of school formation, i.e., fish in a small school (less than five) were not counted. Observations were repeated eight times and the results were averaged to reduce observational error.

Immediately after completing the fish observation, the four physical habitat characteristics of depth, velocity, cover, and vegetation were surveyed to establish a relationship between physical environment and habitat preferences of Japanese medaka. First, depth and velocity were measured to divide the study reach into small water bodies having similar condition with regard to these two physical parameters. Depth was measured with a stainless steel ruler, and velocity with a portable propeller current meter (KENEK, V-303) at three lateral points comprising a midpoint and two near-shore points at longitudinal intervals of 1 m. By using the measurements of depth and velocity, the reach was divided into water bodies. Next, the other two factors of cover and vegetation were calculated from the schematic diagrams of the water bodies. The lateral cover ratio is defined as a function of the presence of lateral cover, which comprises the water’s edge, a dike, or anything that emerges from the water surface and surrounds the water body. The cover thus consists of four components (four lateral sides). The maximum cover ratio is 100%, and each of the cover components is assigned a score of 25%. In the definition of the cover, objects attached to more than 90% of the boundary between water bodies were regarded as cover components. That is, only instream and undersurface cover structures were considered as cover component because they may have had the same effects as the margin of the stream. Percent vegetation coverage is defined as the percentage of the area covered with aquatic vegetation in each water body. Both submerged and emergent vegetation were pooled because of their same roles in providing food and shelter from predators and fast-flowing currents.

Figure 1: Study reach

The habitat use data used in the following analyses were the observed fish population density obtained for the i^{th} water body $\rho_{o,i}$ (individuals per square metre), where i (i=1, 2, …, n) denotes the index of the water body and n the total number of water bodies.

2.4 Fuzzy Habitat Preference Model

In fuzzy habitat preference model (FHPM), a simplified fuzzy reasoning was introduced to relate physical habitat characteristic to habitat preference with the consideration of the uncertainties (Fig. 2), and a simple GA to optimize the

ISBN: 978-989-95079-6-8

model structure. GA was employed because it enables us to simultaneously optimize all the parameters in FHPM even under the nonlinear, complex interactions between physical environment and fish habitat preference. The premise part of FHPM (Fig. 2(i)), i.e. fuzzy membership functions, is defined for the purpose of reflecting the ecology of Japanese medaka. For instance, because the body length of an adult medaka is approximately 2 cm, the critical requirements for Japanese medaka with regard to depth and velocity thus would be relatively shallower and slowly flowing water. The habitat preference is calculated by taking the weighted mean of singletons in the consequent part (Fig. 2(ii)) by the membership value μ_i. That is,

$$P_j = \frac{\sum_{i=1}^{n_m} \mu_{j,i} \cdot c_{j,i}}{\sum_{i=1}^{n_m} \mu_{j,i}} \quad (1)$$

where P_j denotes the habitat preference to the environmental factor j (=depth, velocity, cover, and vegetation), c_i is the value assigned to each singleton (i.e., degree of habitat preference assigned to the corresponding membership function in the premise part), μ_i is the membership value (i.e., degree of fitness to the membership functions in the premise part), i is the index of the membership functions, and n_m is the number of membership functions (Fig. 2). By taking the weighted mean (1), the results are given in a simple linear form. The singletons in the consequence part are determined by using GA so as to minimize mean square error (MSE) between predicted and observed fish population density. The optimization procedure is summarized as follows.

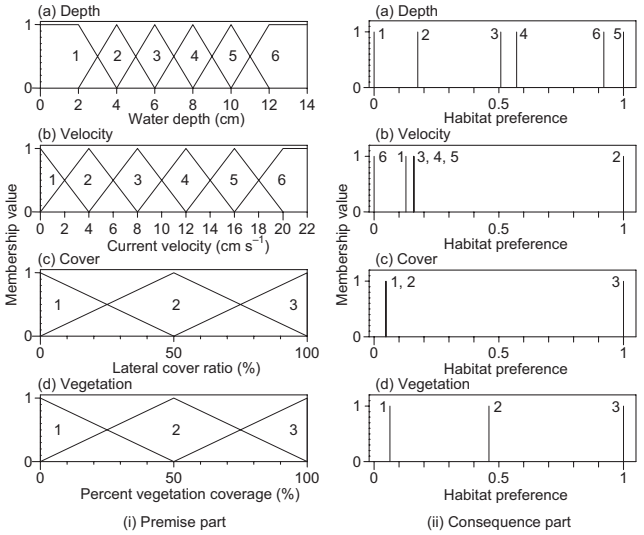


Figure 2: Membership functions of FHPM

First, GA proposed a set of initial model structures (i.e., consequence part) of FHPM. Second, the environmental measurements were given as input values. The input values, i.e., four physical environmental variables, are represented by symmetric triangular fuzzy numbers for considering the uncertainty originated from measurement errors and spatial dispersions. The fuzzy inputs are expressed by its centre a^c_j and spread a^s_j as (a^c_j, a^s_j), where j denotes the environmental factors of depth, velocity, cover, and vegetation. The spread a^s_j is determined from allowable variance when dividing the reach into water bodies: 1 cm is given for the spread of depth a^s_d, 2 cm s^{-1} for velocity a^s_v, 10% each for cover a^s_c and vegetation a^s_{veg}. Of these, the abbreviations of d, v, c, and veg indicate depth, velocity, cover, and vegetation, respectively. Third, habitat preferences for each environmental factor were calculated. Fourth, the habitat preferences to the four environmental factors were combined using (2).

$$P_i = P_{d,i} \times P_{v,i} \times P_{c,i} \times P_{veg,i} \quad (2)$$

Fifth, the habitat use of Japanese medaka was predicted by using (3).

$$\rho_{c,i} = \left(P_i \Big/ \sum_{i=1}^{n} P_i \right) \cdot \sum_{i=1}^{n} \rho_{o,i} \quad (3)$$

Table 1: Condition of GA optimization

Operation	Condition & Remarks
Selection	Elite preservation strategy
Crossover	Uniform
Mutation	0.05%
Number of individuals	100
Number of iteration	5000
Length of binary strings	6 bits per parameter

Table 2: Condition of model structure modification, in which filled circles indicate the factors modified.

Factor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Depth	●				●	●	●				●	●	●		●
Velocity		●			●			●	●		●	●		●	●
Cover			●			●		●		●	●		●	●	●
Vegetation				●			●		●	●		●	●	●	●
Number*	1	1	1	1	2	2	2	2	2	2	3	3	3	3	4

*Number of membership functions modified.

Next, MSE between the predicted and observed fish population density was calculated, and then GA repeatedly modified the singletons so as to minimize MSE. Finally, the optimized FHPM was obtained. The basic condition of GA optimization is summarised in Table 1. In the original setting, totalled 17 parameters were optimised (5 for depth, 6 for velocity, and 3 each for cover and vegetation). The number of parameters to be modified differed by the condition of model structure modification (see Table 2).

2.5 Model Development and Analysis

In the present study, a data set collected on 14 October and 5 November 2004 was used for calibration while the other data set collected on 9 November was used for validation. The condition given for comparing the results of model structure modification is summarised in Table 2. Totalled 31 sets of the condition were given in which one is given for original model, and 15 each for reducing and adding membership functions to the original model. Since model structures of the models vary in accordance with their initial conditions in the optimization, 20 different initial conditions

ISBN: 978-989-95079-6-8

were thus given to each model so as to evaluate the variance of model structure developed. For the comparison, MSE between predicted and observed fish population densities was calculated. Average and standard deviation of MSE on all the models were then used for the comparison. The relationship between MSE and number of fuzzy membership functions was compared in order to clarify the accuracy-complexity tradeoffs in FHPM.

3 Results

The results of field surveys on 14 October and 5 November 2004 were pooled and used for model calibration (Fig. 3(i)), while the result of survey on 9 November was used for model validation (Fig. 3(ii)). All the models were developed by using raw data shown in Fig. 3.

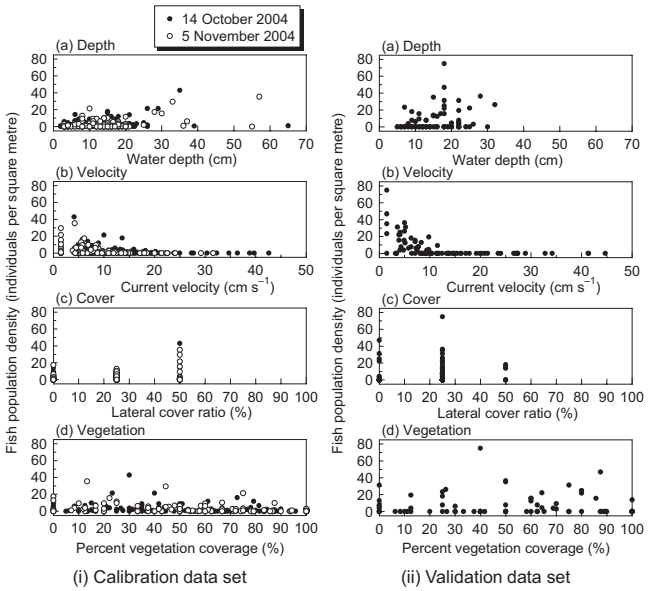


Figure 3: Result of field survey

Habitat preferences evaluated by all the models reflected their model structure, in which the variance in habitat preference curves would be ascribed to the difference in model structure among 20 different initial conditions. Of these, three habitat preference curves evaluated by the models of no modification (original), four-factor reduced and four-factor added are shown as an example (Fig. 4). Despite of the similar tendency in habitat preference, the larger variance in model structure was found in the case of four-factor added model.

By using the habitat preference models together with (2) and (3), habitat uses of Japanese medaka were predicted, of which MSE between predicted and observed fish population density (i.e., habitat use) were used for the comparison. Because some calibration turned out to be failure (especially in the case of adding fuzzy membership function), the best result (i.e., model) achieved at each condition was used in the following analysis. Fig. 5 shows the scatter diagrams of the averaged MSE and number of fuzzy membership functions in both calibration and validation, in which the number of 17 indicates the original model. The model structure could not be improved by modifying the number of fuzzy membership functions in calibration. By contrast, reducing number of fuzzy membership functions resulted in deterioration of prediction ability of the models. The degree of deterioration differed by factors that was modified. As a result, no tradeoff was found between prediction ability of the models (i.e., MSE) and number of fuzzy membership function (i.e., model complexity).

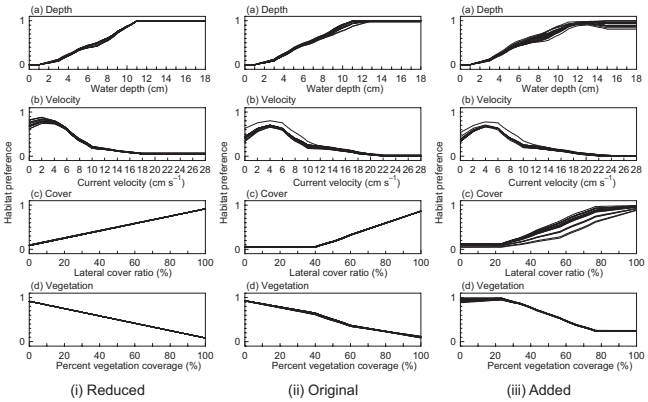


Figure 4: Habitat preference curves evaluated by three models of four-factor reduced (i), no modification (original) (ii), and four-factor added (iii).

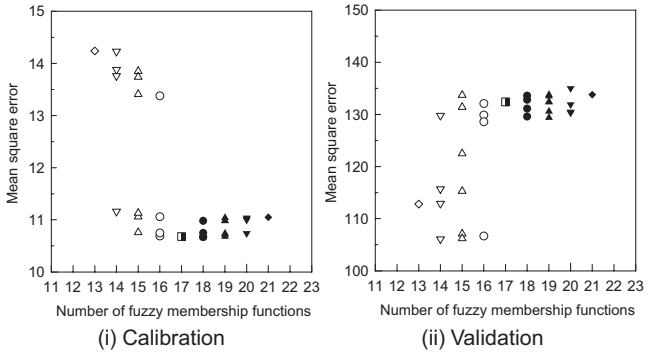


Figure 5: Scatter diagram between minimum mean square error and number of fuzzy membership functions in calibration (i) and validation (ii), in which open marks represent the case of reduced number of fuzzy membership function while filled marks represent that of added.

The MSEs of calibration and validation were compared for the deeper understandings of the relationship between prediction ability and model structure of FHPM. In Fig. 6, there were slight tradeoffs between calibration and validation results. For instance, models with smaller MSE in calibration resulted in larger MSE in validation. Fig. 6 also indicated that no improvement was achieved by adding fuzzy membership functions to the original one.

4 Discussion

The complex interactions between physical environment and habitat preference of the fish made it difficult to clarify the tradeoffs in FHPM. All the habitat preference curves of 31 different conditions showed similar trend of larger water depth, slower velocity, larger cover and smaller vegetation (Fig. 4). However, the prediction result differed between models (Figs. 5 & 6). In Fig. 5, reducing fuzzy membership

ISBN: 978-989-95079-6-8

function resulted in better prediction in the validation data set while prediction in the calibration data set resulted in deterioration. On one hand, this clearly indicates the effect of over-fitting to the data used in model calibration. This could be improved by modifying the conditions of GA optimization such as number of iteration and number of data used in model development. On the other hand, this would be ascribed to the different distribution of the physical environment and habitat use of the fish (Fig. 3). This is one of the major problems in habitat modelling and is often discussed as "transferability" of the habitat models [8, 9]. Adding fuzzy membership function thus could not improve prediction ability of the model. To achieve the model transferability, it is necessary to develop habitat models by using sufficient data which satisfy generality of the habitat use of the fish. In the present case, an approach to achieve generality would be the use of different sub-sets of data in model optimization. For instance, [38] employed different subsets of data to cope with the problems of overtraining and high sensitivity to outliers in the application of artificial neural network model. Reference [16] also applied similar approach in the optimization of fuzzy rule-based habitat model. Another approach would be to modify the fuzzy membership function into less precise definition as often applied in habitat modelling of riverine fish [12, 13, 14, 15, 16]. In this case, the advantage of consideration of critical requirement into model structure would however be lost at the same time. Besides the discussion on complexity and improvement of the model, the present result may suggest the significance of cover in the habitat use of the fish. In [20], the significance of cover was also supported by an approach using Akaike Information Criterion (AIC).

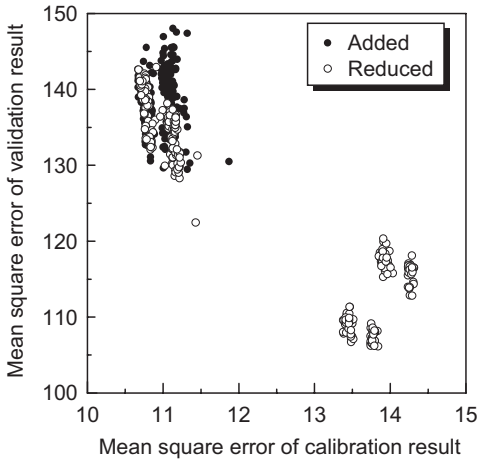


Figure 6: Scatter diagram between mean square errors of calibration and validation results, in which open circles represent the case of reduced number of fuzzy membership function while filled circles represent that of added.

As a concept, adding fuzzy membership function would improve model performance by precisely describing habitat preference of the fish. By contrast, increasing number of parameters to be optimised would put searching loads on GA. Although no clear trend was found between accuracy of the habitat prediction and model structure (i.e., number of fuzzy membership functions), further analyses should be carried out to assess the complex combination such as reducing and adding fuzzy membership functions at the same time. The use of AIC would be appropriate for the quantification of the complex relationship between model structure and prediction ability. In addition, different performance measures such as correctly fuzzy classified instances (%CCFI) and average deviation (AD), and interpretability-preserving optimization of the fuzzy models as presented in [36], can be a pathway for the accuracy improvement of the present model. Further studies would be needed to quantitatively understand the complexity-accuracy tradeoff of FHPM .

Acknowledgment

The author wishes his thanks to Prof. Dr. K. Hiramatsu (Faculty of Agriculture, Kyushu University) for his generous and capable assistance in all phases of this study. This study was partly supported by Grant-in-aid for Young Scientists (start-up) of Japan Society for the Promotion of Science (JSPS), Japan.

References

[1] P. W. Downs and K.J. Gregory, River Channel Management: Toward Sustainable Catchment Hydorosystems. Arnold, London. 2004.

[2] U.S. Fish and Wildlife Service, Habitat Evaluation Procedures (HEP): Ecological Service Manual 102. Washington, DC, USA. 1980.

[3] U.S. Fish and Wildlife Service, Standards for the Development of Habitat Suitability Index Models: Ecological Service Manual 103.Washington, DC, USA. 1980.

[4] K.D. Bovee et al., Stream habitat analysis using the instream flow incremental methodology. U.S. Geological Survey, Biological Resources Division Information and Technology Report. USGS/BRD-1998-0004. 1998.

[5] B. Ahmadi-Nedushan et al., A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Res. Appl. 2006.

[6] D. J. Orth and O. E Maughan, 1982. Evaluation of the incremental methodology for recommending instream flows for fishes. Trans. Am. Fish. Soc. 111, 413–445., 1982.

[7] J.C. Guay et al., Development and validation of numerical habitat models for juveniles of Atlantic salmon (*Salmo salar*). Can. J. Fish. Aquat. Sci. 57, 2065–2075., 2000.

[8] J.C. Guay et al., Assessment of the transferability of biological habitat models for Atlantic salmon parr (*Salmo salar*). Can. J. Fish. Aquat. Sci. 60, 1398–1408., 2003

[9] M. Nykänen and A. Huusko, Transferability of habitat preference criteria for larval European grayling (*Thymallus thymallus*). Can. J. Fish. Aquat. Sci. 61, 185–192, 2004.

[10] S. Lek, Uncertainty in ecological models. Ecol. Model. 207, 1–2., 2007

[11] W. Silvert, Fuzzy indices of environmental conditions. Ecol. Model. 130, 111–119., 2000.

[12] K. Jorde et al., Fuzzy based models for the evaluation of fish habitat quality and instream flow assessment. In: Proceedings of the 2001 International Symposium on Environmental Hydraulics., 2001;

[13] V. Adriaenssens et al., Fuzzy knowledge-based models for prediction of *Asellus* and *Gammarus* in watercourses in Flanders (Belgium). Ecol. Model. 195, 3–10, 2006.

[14] E. Van Broekhoven et al., Fuzzy rule-based macroinvertebrate habitat suitability models for running waters. Ecol. Model. 198, 71–84., 2006.

[15] A. S. Mouton et al., Fish habitat modelling as a tool for river management. Ecol. Eng. 29, 305–315, 2007.

ISBN: 978-989-95079-6-8

[16] A. S. Mouton et al., Optimisation of a fuzzy physical habitat model for spawning European grayling (*Thymallus thymallus* L.) in the Aare river (Thun, Switzerland). Ecol. Model. 215, 122–132, 2008.

[17] M. A. Burgman et al., 2001. Setting reliability bounds on habitat suitability indices. Ecol. Appl. 11 (1), 70–78., 2001

[18] K. Hiramatsu et al., Mathematical modeling of habitat preference of Japanese medaka for instream water environment using fuzzy inference. Trans. JSIDRE 228, 65–72 (in Japanese with English abstract)., 2003.

[19] S. Fukuda et al., Mathematical characterization of fuzziness in fish habitat preference of Japanese medaka (*Oryzias latipes*) in agricultural canal. Trans. JSIDRE 239, 43–49 (in Japanese with English abstract)., 2005.

[20] S. Fukuda et al., An information-theoretic approach for the model selection in habitat preference evaluation of Japanese medaka (*Oryzias latipes*). Bull. Inst. Trop. Agric. Kyushu Univ. 29, 105–118., 2006.

[21] S. Fukuda et al., Numerical quantification of the significance of aquatic vegetation affecting spatial distribution of Japanese medaka (*Oryzias latipes*) in an agricultural canal. Landscape Ecol. Eng. 2, 65–80, 2006.

[22] S. Fukuda and K. Hiramatsu, Prediction ability and sensitivity of artificial intelligence-based habitat preference models for predicting spatial distribution of Japanese medaka (*Oryzias latipes*). Ecol. Model. 215, 301-313., 2008.

[23] S. Fukuda and S. Okushima, Assessing nonlinearity in fish habitat preference of Japanese medaka (*Oryzias latipes*) using genetic algorithm-optimized habitat prediction models. JARQ-Jpn. Agric. Res. Quart. 42 (2), 97–107., 2008.

[24] S. Fukuda, Consideration of fuzziness: Is it necessary in modelling fish habitat preference of Japanese medaka (*Oryzias latipes*)? Ecol. Model., in press, 2009

[25] S. Fukuda et al., Fuzzy neural network model for habitat prediction and HEP for habitat quality estimation focusing on Japanese medaka (*Oryzias latipes*) in agricultural canals. Paddy Water Environ. 4 (3), 119–124, 2006.

[26] S. Fukuda, Uncertainty analysis on fuzzy neural network model for evaluating fish habitat preference of Japanese medaka (*Oryzias latipes*) in agricultural canals in Japan. Proceedings of 7th International Symposium on Ecohydraulics. Paper ID: conf187a145., 2009.

[27] R.W. Bosserman and R.K. Ragade, Ecosystem analysis using fuzzy set theory. Ecol. Model. 16, 191–208. 1982

[28] A. Salski, Fuzzy knowledge-based models in ecological research. Ecol. Model. 63, 103–112. 1992.

[29] W. Bock and A. Salski, A fuzzy knowledge-based model of population dynamics of the Yellow-necked mouse (*Apodemus flavicollis*) in a beech forest. Ecol. Model. 108, 155–161. 1998.

[30] D.G. Chen and S.R. Hare, Neural network and fuzzy logic models for pacific halibut recruitment analysis. Ecol. Model. 195, 11–19. 2006.

[31] S. Mackinson, An adaptive fuzzy expert system for predicting structure, dynamics and distribution of herring shoals. Ecol. Model. 126, 155–178. 2000.

[32] A. Salski, Fuzzy clustering of fuzzy ecological data. Ecol. Inform. 2, 262–269. 2007.

[33] A.M.F. Cornelissen et al., Elicitation of expert knowledge for fuzzy evaluation of agricultural production systems. Agric. Ecosyst. Environ. 95, 1–18. 2003.

[34] R. Bosma, et al., Fuzzy modelling of farmers motivations for integrated farming in the Vietnamese Mekong Delta. The 2005 IEEE International Conference on Fuzzy Systems 827–832. 2005.

[35] V. Adriaenssens et al., Fuzzy rule-based models for decision support in ecosystem management. Sci. Total Environ. 319, 1–12. 2004.

[36] E. Van Broekhoven et al., Interpretability-preserving genetic optimization of linguistic terms in fuzzy models for fuzzy ordered classification: An ecological case study. Int. J. Approx. Reasoning. 44, 65–90. 2007.

[37] F. Herrera, Genetic fuzzy systems: taxonomy, current research trends and prospects. Evol. Intel. 1, 27–46. 2008.

[38] S. C. Dekker et al., Analysing forest transpiration model errors with artificial neural networks. J. Hydrol. 246, 197–208., 2001

ISBN: 978-989-95079-6-8

Towards A Neural-Based Understanding of the Cauchy Deviate Method for Processing Interval and Fuzzy Uncertainty

Vladik Kreinovich[1] and Hung T. Nguyen[2]

[1]Department of Computer Science, University in the Texas at El Paso
500 W. University, El Paso, TX 79968, USA, vladik@utep.edu
[2]Department of Mathematical Sciences, New Mexico State University
Las Cruces, NM 88003, USA, hunguyen@nmsu.edu

Abstract— *One of the most efficient techniques for processing interval and fuzzy data is a Monte-Carlo type technique of Cauchy deviates that uses Cauchy distributions. This technique is mathematically valid, but somewhat counterintuitive. In this paper, following the ideas of Paul Werbos, we provide a natural neural network explanation for this technique.*

Keywords— Cauchy deviate method, fuzzy uncertainty, interval uncertainty, Monte-Carlo simulations, neural networks

1 Formulation of the Problem: Cauchy Deviate Method and Need for Intuitive Explanation

1.1 Practical Need for Uncertainty Propagation

In many practical situations, we are interested in the value of a quantity y which is difficult or even impossible to measure directly. To estimate this difficult-to-measure quantity y, we measure or estimate related easier-to-measure quantities $x_1, \ldots, x_n$ which are related to the desired quantity y by a known relation $y = f(x_1, \ldots, x_n)$. Then, we apply the relation f to the estimates $\widetilde{x}_1, \ldots, \widetilde{x}_n$ for x_i and produce an estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ for the desired quantity y.

In the simplest cases, the relation $f(x_1, \ldots, x_n)$ may be an explicit expression: e.g., if we know the current x_1 and the resistance x_2, then we can measure the voltage y by using Ohm's law $y = x_1 \cdot x_2$. In many practical situations, the relation between x_i and y is much more complicated: the corresponding algorithm $f(x_1, \ldots, x_n)$ is not an explicit expression, but a complex algorithm for solving an appropriate non-linear equation (or system of equations).

Estimates are never absolutely accurate:

- measurements are never absolutely precise, and
- expert estimates can only provide approximate values of the directly measured quantities $x_1, \ldots, x_n$.

In both cases, the resulting estimates $\widetilde{x}_i$ are, in general, different from the actual (unknown) values x_i. Due to these estimation errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$, even if the relation $f(x_1, \ldots, x_n)$ is exact, the estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ is different from the actual value $y = f(x_1, \ldots, x_n)$: $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y \neq 0$.

(In many situations, when the relation $f(x_1, \ldots, x_n)$ is only known approximately, there is an additional source of the approximation error in y caused by the uncertainty in knowing this relation.)

It is therefore desirable to find out how the uncertainty Δx_i in estimating x_i affects the uncertainty Δy in the desired quantity, i.e., how the uncertainties Δx_i *propagate* via the algorithm $f(x_1, \ldots, x_n)$.

1.2 Propagation of Probabilistic Uncertainty

Often, we know the probabilities of different values of Δx_i. For example, in many cases, we know that the approximation errors Δx_i are independent normally distributed with zero mean and known standard deviations σ_i; see, e.g., [16].

In this case, we can use known statistical techniques to estimate the resulting uncertainty Δy in y. For example, since we know the probability distributions, we can simulate them in the computer, i.e., use the Monte-Carlo simulation techniques to get a sample population $\Delta y^{(1)}, \ldots, \Delta y^{(N)}$ of the corresponding errors Δy. Based on this sample, we can then estimate the desired statistical characteristics of the desired approximation error Δy.

1.3 Propagation of Interval Uncertainty

In many other practical situations, we do not know these probabilities, we only know the upper bounds Δ_i on the (absolute values of) the corresponding measurement errors Δx_i: $|\Delta x_i| \leq \Delta$.

In this case, based on the known approximation $\widetilde{x}_i$, we can conclude that the actual (unknown) value of i-th auxiliary quantity x_i can take any value from the interval

$$\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]. \tag{1}$$

To find the resulting uncertainty in y, we must therefore find the range $\mathbf{y} = [\underline{y}, \overline{y}]$ of possible values of y when $x_i \in \mathbf{x}_i$:

$$\mathbf{y} = f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=}$$

$$\{f(x_1, \ldots, x_n) \,|\, x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}. \tag{2}$$

Computations of this range under interval uncertainty is called *interval computations*; see, e.g., [4, 5].

The corresponding computational problems are, in general, NP-hard [9]. Crudely speaking, this means that, in general, such problems require a large amount of computation time – and that therefore faster methods are needed.

1.4 Propagation of Fuzzy Uncertainty

In many practical situations, the estimates $\widetilde{x}_i$ come from experts. Experts often describe the inaccuracy of their estimates

ISBN: 978-989-95079-6-8

in terms of imprecise words from natural language, such as "approximately 0.1", etc. A natural way to formalize such words is to use special techniques developed for formalizing this type of estimates – specifically, the technique of fuzzy logic; see, e.g., [6, 15].

In this technique, for each possible value of $x_i \in \mathbf{x}_i$, we describe the degree $\mu_i(x_i)$ to which this value is possible. For each degree of certainty α, we can determine the set of values of x_i that are possible with at least this degree of certainty – the *α-cut* $\mathbf{x}_i(\alpha) = \{x \,|\, \mu(x) \geq \alpha\}$ of the original fuzzy set. Vice versa, if we know α-cuts for every α, then, for each object x, we can determine the degree of possibility that x belongs to the original fuzzy set [3, 6, 12, 13, 15]. A fuzzy set can be thus viewed as a nested family of its (interval) α-cuts.

We already know how to propagate interval uncertainty. Thus, to propagate this fuzzy uncertainty, we can therefore consider, for each α, the fuzzy set y with the α-cuts

$$\mathbf{y}(\alpha) = f(\mathbf{x}_1(\alpha), \ldots, \mathbf{x}_1(\alpha)); \tag{3}$$

see, e.g., [3, 6, 12, 13, 15]. So, from the computational viewpoint, the problem of propagating fuzzy uncertainty can be reduced to several interval propagation problems.

1.5 Need for Faster Algorithms for Uncertainty Propagation

Summarizing the above analysis, we can conclude that in principle, we need to consider three basic types of uncertainty propagation: situations when we propagate probabilistic, interval, and fuzzy uncertainty. It is also possible that some quantities are represented by fuzzy sets, while others may be represented by probabilities.

For probabilistic uncertainty, there exist reasonable efficient uncertainty propagation algorithms such as Monte-Carlo simulations. In contrast, the problems of propagating interval and fuzzy uncertainty are, in general, computationally difficult. It is therefore desirable to design faster algorithms for propagating interval and fuzzy uncertainty.

Once such methods are developed, we can then use these methods to propagate interval and fuzzy uncertainty components, and Monte-Carlo simulations to propagate the probabilistic uncertainty.

The computational problem of propagating fuzzy uncertainty can be naturally reduced to the problem of propagating interval uncertainty. Because of this reduction, in the following text, we will mainly concentrate on faster algorithms for propagating interval uncertainty.

1.6 Linearization Situations: Description

Due to the approximation errors $\Delta x_i = \widetilde{x}_i - x_i$, the unknown (actual) values $x_i = \widetilde{x}_i - \Delta x_i$ of the input quantities x_i are, in general, different from the approximate estimates $\widetilde{x}_i$. In many practical situations, the approximation errors Δx_i are small – e.g., when the approximations are obtained by reasonably accurate measurements. In such situations, we can ignore terms which are quadratic (and of higher order) in Δx_i.

1.7 Linearization Situations: Analysis

In the above situations, we can expand the expression for

$$\Delta y = \widetilde{y} - y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(x_1, \ldots, x_n) =$$

$$f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n) \tag{4}$$

in Taylor series in Δx_i and keep only the linear terms in this expansion. In this case, we get

$$\Delta y = c_1 \cdot \Delta x_1 + \ldots + c_n \cdot \Delta x_n, \tag{5}$$

where we denoted

$$c_i \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}(\widetilde{x}_1, \ldots, \widetilde{x}_n).$$

For a linear function, the largest possible value of Δy is obtained when each of the variables $\Delta x_i \in [-\Delta_i, \Delta_i]$ attains:

- either its largest value Δ_i (when $c_i \geq 0$)
- or its smallest value $-\Delta_i$ (when $c_i < 0$).

In both cases, the largest possible value of the corresponding term in Δy is equal to $|c_i| \cdot \Delta_i$. Thus, the largest possible value of Δy is equal to

$$\Delta = |c_1| \cdot \Delta_1 + \ldots + |c_n| \cdot \Delta_n. \tag{6}$$

Similarly, the smallest possible value of Δy is obtained when each of the variables $\Delta x_i \in [-\Delta_i, \Delta_i]$ attains

- either its smallest value $-\Delta_i$ (when $c_i \geq 0$)
- or its largest value Δ_i (when $c_i < 0$).

In both cases, the smallest possible value of the corresponding term in Δy is equal to $-|c_i| \cdot \Delta_i$. Thus, the smallest possible value of Δy is equal to

$$-\Delta = -|c_1| \cdot \Delta_1 - \ldots - |c_n| \cdot \Delta_n. \tag{7}$$

Can we transform these natural formulas into an algorithm? Due to the linearization assumption, we can estimate each partial derivative c_i as

$$c_i \approx \frac{f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) - \widetilde{y}}{h_i} \tag{8}$$

for some small values h_i. So, we arrive at the following algorithm.

1.8 Linearization Situations: Algorithm

To compute the range $\mathbf{y}$ of y, we do the following.

- First, we apply the algorithm f to the original estimates $\widetilde{x}_1, \ldots, \widetilde{x}_n$, resulting in the value $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.
- Second, for all i from 1 to n, we compute $f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n)$ for some small h_i and then compute

$$c_i = \frac{f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) - \widetilde{y}}{h_i}. \tag{9}$$

- Finally, we compute

$$\Delta = |c_1| \cdot \Delta_1 + \ldots + |c_n| \cdot \Delta_n$$

and the desired range $\mathbf{y} = [\widetilde{y} - \Delta, \widetilde{y} + \Delta]$.

ISBN: 978-989-95079-6-8

1.9 Linearization Situations: Computational Complexity

The main computation time is spent on calling the time-consuming algorithm f. In the above uncertainty propagation algorithm, after one call to f to compute $\widetilde{y}$, we need n calls to f to compute the corresponding partial derivatives c_i and then, we can estimate the desired uncertainty Δ in y by using the above simple formula.

Overall, we thus need $n+1$ calls to the algorithm f.

1.10 Cauchy Deviate Method

For large n, we can further reduce the number of calls to f if we use a special technique of Cauchy-based Monte-Carlo simulations, which enables us to use a fixed number of calls to f (≈ 200) for all possible values n; see, e.g., [7, 8].

1.11 Mathematics Behind the Cauchy Method

In our simulations, we use *Cauchy distribution* – i.e., probability distributions with the probability density

$$\rho(z) = \frac{\Delta}{\pi \cdot (z^2 + \Delta^2)}; \tag{10}$$

the value Δ is called the *(scale) parameter* of this distribution.

Cauchy distribution has the following property that we will use: if $z_1, \ldots, z_n$ are independent random variables, and each of z_i is distributed according to the Cauchy law with parameter Δ_i, then their linear combination

$$z = c_1 \cdot z_1 + \ldots + c_n \cdot z_n \tag{11}$$

is also distributed according to a Cauchy law, with a scale parameter $\Delta = |c_1| \cdot \Delta_1 + \ldots + |c_n| \cdot \Delta_n$.

Therefore, if we take random variables δ_i which are Cauchy distributed with parameters Δ_i, then the value

$$\delta \stackrel{\text{def}}{=} f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \delta_1, \ldots, \widetilde{x}_n - \delta_n) =$$

$$c_1 \cdot \delta_1 + \ldots + c_n \cdot \delta_n \tag{12}$$

is Cauchy distributed with the desired parameter

$$\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i. \tag{13}$$

So, repeating this experiment N times, we get N values $\delta^{(1)}, \ldots, \delta^{(N)}$ which are Cauchy distributed with the unknown parameter, and from them we can estimate Δ.

The bigger N, the better estimates we get.

1.12 Cauchy Method: Towards Implementation

To implement this idea, we must answer the following two questions:

- how to simulate the Cauchy distribution; and
- how to estimate the parameter Δ of this distribution from a finite sample.

Simulation can be based on the functional transformation of uniformly distributed sample values:

$$\delta_i = \Delta_i \cdot \tan(\pi \cdot (r_i - 0.5)), \tag{14}$$

where r_i is uniformly distributed on the interval $[0, 1]$.

In order to estimate Δ, we can apply the Maximum Likelihood Method

$$\rho(\delta^{(1)}) \cdot \rho(\delta^{(2)}) \cdot \ldots \cdot \rho(\delta^{(N)}) \to \max, \tag{15}$$

where $\rho(z)$ is a Cauchy distribution density with the unknown Δ. When we substitute the above-given formula for $\rho(z)$ and equate the derivative of the product with respect to Δ to 0 (since it is a maximum), we get an equation

$$\frac{1}{1 + \left(\dfrac{\delta^{(1)}}{\Delta}\right)^2} + \ldots + \frac{1}{1 + \left(\dfrac{\delta^{(N)}}{\Delta}\right)^2} = \frac{N}{2}. \tag{16}$$

The left-hand side of (16) is an increasing function that is equal to $0 (< N/2)$ for $\Delta = 0$ and $> N/2$ for $\Delta = \max\left|\delta^{(k)}\right|$; therefore the solution to the equation (16) can be found by applying a bisection method to the interval $\left[0, \max\left|\delta^{(k)}\right|\right]$.

It is important to mention that we assumed that the function f is reasonably linear within the box

$$[\widetilde{x}_1 - \Delta_1, \widetilde{x}_1 + \Delta_1] \times \ldots \times [\widetilde{x}_n - \Delta_n, \widetilde{x}_n + \Delta_n]. \tag{17}$$

However, the simulated values δ_i may be outside the box. When we get such values, we do not use the function f for them, we use a normalized function that is equal to f within the box, and that is extended linearly for all other values (we will see, in the description of an algorithm, how this is done).

As a result, we arrive at the following algorithm.

1.13 Cauchy Deviates Method: Algorithm

- Apply f to the results of direct measurements:

$$\widetilde{y} := f(\widetilde{x}_1, \ldots, \widetilde{x}_n); \tag{18}$$

- For $k = 1, 2, \ldots, N$, repeat the following:
  - use the standard random number generator to compute n numbers $r_i^{(k)}$, $i = 1, 2, \ldots, n$, that are uniformly distributed on the interval $[0, 1]$;
  - compute Cauchy distributed values

$$c_i^{(k)} := \tan(\pi \cdot (r_i^{(k)} - 0.5)); \tag{19}$$

  - compute the largest value of $|c_i^{(k)}|$ so that we will be able to normalize the simulated measurement errors and apply f to the values that are within the box of possible values: $K := \max_i |c_i^{(k)}|$;
  - compute the simulated measurement errors

$$\delta_i^{(k)} := \Delta_i \cdot c_i^{(k)} / K; \tag{20}$$

  - compute the simulated "actual values"

$$x_i^{(k)} := \widetilde{x}_i - \delta_i^{(k)}; \tag{21}$$

  - apply the program f to the simulated "actual values" and compute the simulated error of the indirect measurement:

$$\delta^{(k)} := K \cdot \left(\widetilde{y} - f\left(x_1^{(k)}, \ldots, x_n^{(k)}\right)\right); \tag{22}$$

ISBN: 978-989-95079-6-8

- Compute Δ by applying the bisection method to solve the equation (16).

Comment. To avoid confusion, we should emphasize that, in contrast to the Monte-Carlo solution for the probabilistic case, the use of Cauchy distribution in the interval case is a computational trick and *not* a truthful simulation of the actual measurement error Δx_i: indeed, we know that the actual value of Δx_i is always inside the interval $[-\Delta_i, \Delta_i]$, but a Cauchy distributed random attains values outside this interval as well.

1.14 Cauchy Deviate Method: Need for Intuitive Explanation

The above Cauchy deviate method is one of the most efficient techniques for processing interval and fuzzy data. However, this method has a serious drawback: while the corresponding technique is mathematically valid, it is somewhat counterintuitive – we want to analyze errors which are located *instead* a given interval $[-\Delta, \Delta]$, but this analysis use Cauchy simulated errors which are located, with a high probability, *outside* this interval.

It is therefore desirable to come up with an intuitive explanation for this technique. In this paper, we show that such an explanation can be obtained from neural networks. (For a general introduction to neural networks, see, e.g., [2, 11].)

2 Solution: Neural Explanation

2.1 Werbos's Idea: Use Neurons

Our explanation comes from the idea promoted by Paul Werbos, the author of the backpropagation algorithm for training neural networks. Traditionally, neural networks are used to simulate a deterministic dependence; Paul Werbos suggested that the same neural networks can be used to describe stochastic dependencies as well – if as one of the inputs, we take a standard random number r uniformly distributed on the interval $[0, 1]$; see, e.g., [18] and references therein.

In view of this idea, as a natural probability distribution, we can take the result of applying a neural network to this random number. The simplest case is when we have a single neuron. In this case, we apply the activation (input-output) function $f(y)$ corresponding to this neuron to the random number r.

So, let us see what will happen if we apply a neuron to the standard random number and get a value $f(r)$.

2.2 What is the Activation Function of a Neuron: Reminder

To answer the above question, let us recall what are the optimal choices of an activation function of a neuron. This problem was analyzed in detail in [14]; see also [10].

2.3 We Must Choose a Family of Functions, Not a Single Function

We talk about choosing f, but the expression for $f(y)$ will change if we change the units in which we measure all the signals (input, output and intermediate), so in mathematical terms, it is better to speak about choosing a *family* of functions f.

It is reasonable to suggest that if an f belongs to this family, then this family must contain $k \cdot f$ for positive real numbers k. This corresponds to changing units.

Also, it must contain $f + c$, where c is a constant. This is equivalent to adding a constant bias and therefore does not change the abilities of the resulting network.

Since we are talking about non-linear phenomena, we can also assume that some non-linear "rescaling" transformations $x \to g(x)$ are also applicable, i.e., the family must include the composition $g(f(y))$ for each of functions f.

This family must not be too big, therefore, it must be determined by finitely many parameters and should ideally be obtained from one function $f(y)$ by applying all these transformations. Without loss of generality, we can assume that this set of transformations is closed under composition and under inverse, i.e., if $z \to g_1(z)$ and $z \to g_2(z)$ are possible transformations, then $z \to g_1(g_2(z))$ and $z \to g_1^{-1}(z)$ are possible transformations, where by g_1^{-1} we denoted an inverse function $g_1^{-1}(z) = w$ if and only if $g_1(w) = z$. In mathematical terms this means that these transformations form a *group*, and therefore a family is obtained by applying to some function $f(y)$ all transformations from some finite-dimensional transformation group G that includes all linear transformations (and maybe some non-linear ones).

All these transformations correspond to appropriate "rescalings". Rescaling is something that is smoothly changing the initial scale. This means that if we have two different transformations, there must be a smooth transition between them. In mathematical terms, the existence of this continuous transition is expressed by saying that the group is *connected*, and the fact that both the transformations and the transitions are smooth is expressed by saying that this is a *Lie group*).

2.4 Which Family is the Best?

Among all such families, we want to choose the best one. In formalizing what "the best" means we follow the general idea described in [14].

The criteria to choose may be computational simplicity, efficiency of training, or something else. In mathematical optimization problems, numeric criteria are most frequently used, when to every family we assign some value expressing its performance, and choose a family for which this value is maximal. However, it is not necessary to restrict ourselves to such numeric criteria only. For example, if we have several different families that have the same training ability A, we can choose between them the one that has the minimal computational complexity C. In this case, the actual criterion that we use to compare two families is not numeric, but more complicated: *a family F_1 is better than the family F_2 if and only if either $A(F_1) > A(F_2)$ or $A(F_1) = A(F_2)$ and $C(F_1) < C(F_2)$*. A criterion can be even more complicated. What a criterion *must* do is to allow us for every pair of families to tell whether the first family is better with respect to this criterion (we'll denote it by $F_1 > F_2$), or the second is better ($F_1 < F_2$) or these families have the same quality in the sense of this criterion (we'll denote it by $F_1 \sim F_2$). Of course, it is necessary to demand that these choices be consistent, e.g., if $F_1 > F_2$ and $F_2 > F_3$ then $F_1 > F_3$.

Another natural demand is that this criterion must choose a *unique* optimal family (i.e., a family that is better with respect to this criterion than any other family). The reason for this demand is very simple. If a criterion does not choose any family at all, then it is of no use. If several different families are "the

ISBN: 978-989-95079-6-8

best" according to this criterion, then we still have a problem to choose among those "best". Therefore, we need some additional criterion for that choice. For example, if several families turn out to have the same training ability, we can choose among them a family with minimal computational complexity. So what we actually do in this case is abandon that criterion for which there were several "best" families, and consider a new "composite" criterion instead: F_1 is better than F_2 according to this new criterion if either it was better according to the old criterion or according to the old criterion they had the same quality and F_1 is better than F_2 according to the additional criterion. In other words, if a criterion does not allow us to choose a unique best family it means that this criterion is not ultimate; we have to modify it until we come to a final criterion that will have that property.

The next natural condition that the criterion must satisfy is connected with the following. Suppose that instead of a neuron with the transformation function $f(y)$ we consider a neuron with a function $\bar{f}(y) = f(y + a)$, where a is a constant. This new neuron can be easily simulated by the old ones: namely, the output of this new neuron is $\bar{f}(y) = f(y + a)$, so it is equivalent to an old neuron with an additional constant input a. Likewise, the old neuron is equivalent to the new neuron with an additional constant input $-a$. Therefore, the networks that are formed by these new neurons have precisely the same abilities as those that are built from the old ones.

We cannot claim that the new neurons have the same quality as the old ones, because adding a can increase computational complexity and thus slightly worsen the overall quality. But it is natural to demand that adding a does not change the *relative* quality of the neurons, i.e., if a family $\{f(y)\}$ is better that a family of $\{g(y)\}$, then for every a the family $\{f(y+a)\}$ must be still better than the family $\{g(y + a)\}$.

Now, we are ready for the formal definitions.

2.5 Definitions

By a *transformation* we mean a smooth (differentiable) function from real numbers into real numbers. By an *appropriate transformation group* G we mean a finite-dimensional connected Lie group of transformations. By a *family* of functions we mean the set of functions that is obtained from a smooth (everywhere defined) non-constant function $f(y)$ by applying all the transformations from some appropriate transformation group G. Let us denote the set of all the families by F.

A pair of relations $(>, \sim)$ is called *consistent* if it satisfies the following conditions: (1) if $a > b$ and $b > c$ then $a > c$; (2) $a \sim a$; (3) if $a \sim b$ then $b \sim a$; (4) if $a \sim b$ and $b \sim c$ then $a \sim c$; (5) if $a > b$ and $b \sim c$ then $a > c$; (6) if $a \sim b$ and $b > c$ then $a > c$; (7) if $a > b$ then $b > a$ or $a \sim b$ are impossible.

Assume a set A is given. Its elements will be called *alternatives*. By an *optimality criterion* we mean a consistent pair $(>, \sim)$ of relations on the set A of all alternatives. If $a > b$, we say that a is *better* than b; if $a \sim b$, we say that the alternatives a and b are *equivalent* with respect to this criterion. We say that an alternative a is *optimal* (or *best*) with respect to a criterion $(>, \sim)$ if for every other alternative b either $a > b$ or $a \sim b$.

We say that a criterion is *final* if there exists an optimal alternative, and this optimal alternative is unique.

In the present section we consider optimality criteria on the set F of all families.

By the *result of adding* a to a function $f(y)$ we mean a function $\bar{f}(y) = f(y + a)$. By the *result of adding* a to a family F we mean the set of the functions that are obtained from $f \in F$ by adding a. This result will be denoted by $F+a$. We say that an optimality criterion on F is *shift-invariant* if for every two families F and G and for every number a, the following two conditions are true:

i) if F is better than G in the sense of this criterion (i.e., $F > G$), then $F + a > G + a$;

ii) if F is equivalent to G in the sense of this criterion (i.e., $F \sim G$), then $F + a \sim G + a$.

2.6 Main Result

As we have already remarked, the demands that the optimality criterion is final and shift-invariant are quite reasonable. The only problem with them is that at first glance they may seem rather weak. However, they are not, as the following theorem shows:

Theorem. *If a family F is optimal in the sense of some optimality criterion that is final and shift-invariant, then every function f from F has the form $a + b \cdot s_0(K \cdot y + l)$ for some a, b, K and l, where $s_0(y)$ is either a linear function, or a fractional-linear function, or $s_0(y) = \exp(y)$, or the logistic (sigmoid) function $s_0(y) = 1/(1 + \exp(-y))$, or $s_0(y) = \tan(y)$.*

Comment. The logistic function is indeed the most popular activation function for actual neural networks, but others are also used. For our purpose, we will use the tangent function. As we have mentioned earlier, the application of the tangent function to the standard random number r indeed leads to the desired Cauchy distribution.

2.7 Proof: Main Idea

The idea of this proof is as follows: first we prove that the appropriate transformation group consists of fractionally-linear functions (in Part 1), then we prove that the optimal family is shift-invariant (in Part 2), and from that in Part 3 we conclude that any function f from F satisfies some functional equations, whose solutions are known.

2.8 Proof: Part 1

By an *appropriate group* we meant a connected finite-dimensional Lie group of transformations of the set of real numbers R onto itself that contains all linear transformations. Norbert Wiener asked [19] to classify such groups for an n-dimensional space with arbitrary n, and this classification was obtained in [17]. In our case (when $n = 1$) the only possible groups are the group of all linear transformations and the group of all fractionally-linear transformations $x \to (a \cdot x + b)/(c \cdot x + d)$. In both cases the group consists only of fractionally linear transformations.

2.9 Proof: Part 2

Let us now prove that the optimal family F_{opt} exists and is *shift-invariant* in the sense that $F_{opt} = F_{opt} + a$ for all real numbers a. Indeed, we assumed that the optimality criterion

ISBN: 978-989-95079-6-8

is final, therefore there exists a unique optimal family F_{opt}. Let's now prove that this optimal family is shift-invariant.

The fact that F_{opt} is optimal means that for every other F, either $F_{opt} > F$ or $F_{opt} \sim F$. If $F_{opt} \sim F$ for some $F \neq F_{opt}$, then from the definition of the optimality criterion we can easily deduce that F is also optimal, which contradicts the fact that there is only one optimal family. So for every F either $F_{opt} > F$ or $F_{opt} = F$.

Take an arbitrary a and let $F = F_{opt} + a$. If $F_{opt} > F = F_{opt} + a$, then from the invariance of the optimality criterion (condition ii) we conclude that $F_{opt} - a > F_{opt}$, and that conclusion contradicts the choice of F_{opt} as the optimal family. So $F_{opt} > F = F_{opt} + a$ is impossible, and therefore $F_{opt} = F = F_{opt} + a$, i.e., the optimal family is really shift-invariant.

2.10 Proof: Part 3

Let us now deduce the actual form of the functions f from the optimal family. If $f(y)$ is such a function, then the result $f(y + a)$ of adding a to this function f belongs to $F + a$, and so, due to 2., it belongs to F. But all the functions from f can be obtained from each other by fractionally linear transformations, so $f(y + a) = (A + B \cdot f(y))/(C + D \cdot f(y))$ for some A, B, C and D. So we arrive at a functional equation for f. Let us reduce this equation to a one with a known solution. For that purpose, let us use the fact that fractionally linear transformations are projective transformations of a line, and for such transformations the cross ratio is preserved ([1], Section 2.3), i.e., if $g(y) = (A + B \cdot f(y))/(C + D \cdot f(y))$, then

$$\frac{g(y_1) - g(y_3)}{g(y_2) - g(y_3)} \cdot \frac{g(y_2) - g(y_4)}{g(y_1) - g(y_4)} = \frac{f(y_1) - f(y_3)}{f(y_2) - f(y_3)} \cdot \frac{f(y_2) - f(y_4)}{f(y_1) - f(y_4)} \tag{23}$$

for all y_i. In our case this is true for $g(y) = f(y+a)$, therefore for all a the following equality is true:

$$\frac{f(y_1 + a) - f(y_3 + a)}{f(y_2 + a) - f(y_3 + a)} \cdot \frac{f(y_2 + a) - f(y_4 + a)}{f(y_1 + a) - f(y_4 + a)} = \frac{f(y_1) - f(y_3)}{f(y_2) - f(y_3)} \cdot \frac{f(y_2) - f(y_4)}{f(y_1) - f(y_4)}. \tag{24}$$

The most general continuous solutions of this functional equation are given by Theorem 2.3.2 from [1]: either f is fractionally linear, or $f(y) = (a+b\cdot\tan(k\cdot y))/(c+d\cdot\tan(k\cdot y))$ for some a, b, c, d, or $f(y) = (a+b\cdot\tanh(k\cdot y))(c+d\cdot\tanh(k\cdot y))$, where

$$\tanh(z) \stackrel{\text{def}}{=} \frac{\sinh(z)}{\cosh(z)}, \quad \sinh(z) \stackrel{\text{def}}{=} \frac{\exp(z) - \exp(-z))}{2},$$

$$\cosh(z) \stackrel{\text{def}}{=} \frac{\exp(z) + \exp(-z)}{2}. \tag{25}$$

The $\tanh(z)$ expression is equivalent to the logistic function.

The theorem is proven.

Acknowledgment

This work was supported in part by NSF grant HRD-0734825 and by Grant 1 T36 GM078000-01 from the National Institutes of Health. The authors are thankful to Paul Werbos for valuable discussions, and to the anonymous referees for valuable suggestions.

References

[1] J. Aczel, *Lectures on functional equations and their applications,* Dover, New York, 2006.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

[3] D. Dubois and H. Prade, Operations on fuzzy numbers, *International Journal of Systems Science*, 1978, Vol. 9, pp. 613–626.

[4] Interval computations website http://www.cs.utep.edu/interval-comp

[5] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[6] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*, Prentice Hall, New Jersey, 1995.

[7] V. Kreinovich, J. Beck, C. Ferregut, A. Sanchez, G. R. Keller, M. Averill, and S. A. Starks, "Monte-Carlo-type techniques for processing interval uncertainty, and their potential engineering applications", *Reliable Computing*, 2007, Vol. 13, No. 1, pp. 25–69.

[8] V. Kreinovich and S. Ferson, "A New Cauchy-Based Black-Box Technique for Uncertainty in Risk Analysis", *Reliability Engineering and Systems Safety*, 2004, Vol. 85, No. 1–3, pp. 267–279.

[9] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1998.

[10] V. Kreinovich and C. Quintana. "Neural networks: what non- linearity to choose?," *Proceedings of the 4th University of New Brunswick Artificial Intelligence Workshop*, Fredericton, N.B., Canada, 1991, pp. 627–637.

[11] J. Lawrence, *Introduction to Neural Networks*, California Scientific Software Press, 1994.

[12] R. E. Moore and W. Lodwick, Interval Analysis and Fuzzy Set Theory, *Fuzzy Sets and Systems*, 2003, Vol. 135, No. 1, pp. 5–9.

[13] H. T. Nguyen and V. Kreinovich, Nested Intervals and Sets: Concepts, Relations to Fuzzy Sets, and Applications, In: R. B. Kearfott and V. Kreinovich, eds., *Applications of Interval Computations*, Kluwer, Dordrecht, 1996, pp. 245–290.

[14] H. T. Nguyen and V. Kreinovich, *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht, 1997.

[15] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.

[16] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, American Institute of Physics, New York, NY, 2005.

[17] I. M. Singer and S. Sternberg, "Infinite groups of Lie and Cartan, Part 1", *Journal d'Analyse Mathematique*, 1965, Vol. XV, pp. 1–113.

[18] P. J. Werbos, "Intelligence in the brain: A theory of how it works and how to build it", *Neural Networks*, 2009, to appear.

[19] N. Wiener, *Cybernetics, or Control and Communication in the animal and the machine*, MIT Press, Cambridge, MA, 1962.

Fuzzy and PSO Based Algorithm for driver's behavior modeling

S. Khanmohammadi [1] M.A. Tinati[2] S. Ghaemi[2] M.A. Badamchizadeh[2]

1 Islamic Azad University, Science & Research Branch,
Poonak sq. Ashrafi Esfahani Blv., Hesarak, Tehran, Iran
2 Faculty of Electrical and Computer Engineering,
University of Tabriz, Tabriz, Iran
Email: khan@iaucss.org, {Tinati, ghaemi, mbadamchi}@tabrizu.ac.ir

Abstract— *The study of human behavior during driving is of primary importance for the improvement of drivers' security. This study is complex because of numerous situations in which the driver may be involved. In this paper, we propose a hierarchical fuzzy system for human in a driver-vehicle-environment system to model takeover by different drivers. The driver's behavior is affected by the environment. We include climate, road and car conditions in a fuzzy mode. For obtaining fuzzy rules, we have provided three separate questionnaires on the effects of climate; road and car condition on driver's performance. The number of fuzzy rules is optimized by Particle Swarm Optimization algorithm. Also the precision, age and driving individuality are used to model the driver's behavior under difference environments. We investigate the behavior of different drivers when a driver intends to pass the leading car. The comparative study showed that the simulation result is in good agreement with the real situations.*

Keywords— Fuzzy, PSO, Driver's behaviour.

1 Introduction

In recent years, by using new control ideas, safety in driving has been improved. MacAdam (1980) developed an optimal preview control algorithm; however, this algorithm could only be applied to single input single output systems. Fenton applied Linear Quadratic algorithm (1988) to design a controller for steering. Ackermann and Sienel (1990) used parameter space robust control to design the automatic steering controller. These models did not take the driver's preview behavior into consideration. Lee (1989) developed a discrete time preview control algorithm for four-wheel steering passenger vehicles and found that the control accuracy was improved substantially by taking the preview behavior into account. The fuzzy logic has proven to be a very effective tool for handling imprecision and uncertainty, which are both very important characteristics of driving environments. This makes fuzzy logic a powerful candidate tool in most traffic engineering studies [1]. In [2] Kamada et al. proposed a fuzzy logic lateral controller. Hessburg and Tomizuka [6] developed a fuzzy logic controller for vehicle lateral guidance which consisted of three sub-controllers: preview, feedback and gain scheduling. Cai, Lin and Mourant investigated the influence of driver emotion on performance through platoon driving simulated with multiple simulators [4]. They induced two kinds of emotion states (anger and excitation) through realistic driver-driver interaction by using networked driving simulators. Drivers' psycho physiological parameters changes were the indicators of emotions. In the anger and excitation states, drivers showed poorer lane control capability. In [4] the author investigated emotional behavior (anger, neutral, and excitation) of drivers by collecting driving performance data. The results demonstrated the feasibility and efficiency of using multiple networked driving simulators to study driver emotional behavior, e.g., road rage.

In this paper, we use new parameters to identify uniqueness of driving maneuver of each driver under different environment, the precision, age and driving individuality. In modeling, the decision making process is based on three positions, one position in equal lane and two positions in opposite lane. By considering three positions, the speed, direction of car and the steering angle, a fuzzy model is presented for steering angle and speed control. We used two levels for modeling. The low level control model is responsible for modeling the steering angle and the speed variations enforced by the driver. The high level control model, models the decision making process of the driver. For this purpose, first the fuzzified low level control models are illustrated in section 2. The high level decision making model is designed in section 3 and the simulation results are presented in section 4. Finally section 5 describes conclusions.

2 Low level control

In order to implement the low-level control model, a simple car's model is required [5]. The car states include the Cartesian position (x,y) centered mid-way between the rear wheels and the car's orientation denoted by equations 1 to 3, where φ is the steering angle, V is the car's speed and θ is the angle of the car with respect to the X axis and L is the distance between center of rear and front wheels.

$$x_{n+1} = x_n + V_n \times \cos(\theta_n) \times \Delta t \quad (1)$$

$$y_{n+1} = y_n + V_n \times \sin(\theta_n) \times \Delta t \quad (2)$$

$$\theta_{n+1} = \theta_n + \frac{V_n \times \tan(\varphi_n) \times \Delta t}{L} \quad (3)$$

The driver controls V and φ. In order to model the complete low level control procedure, we have assumed that apart from the information from the environment that the driver perceives, other information such as age and driver's individuality are influential in the driver's control procedure. This can be formulated as:

$$\Delta V = f(x_n, y_n, \varphi_n, \theta_n, Environment_n, Driver_n) \quad (4)$$

ISBN: 978-989-95079-6-8

$$\Delta\varphi = f(x_n, y_n, \varphi_n, \theta_n, Environment_n, Driver_n) \quad (5)$$

Where $(x_n, y_n, \theta_n, \varphi_n)$ determine car's state, *Environment* is the information of the condition of *Climate*, *Road* and *Car* and *Driver* is the individual driver influence.

2-1) Environmental condition

The *Climate* condition is the information of luminosity, field of view, rain, temperature and humidity. The *Road* condition is about the traffic surrounding the driver, road safety, quality of road materials, moving obstacle and having enough signs and finally the *Car* condition is the information of car's ergonomic, safety equipments performance and agent of distraction in car. The information is considered as the input variables of fuzzy system (Figure 1).

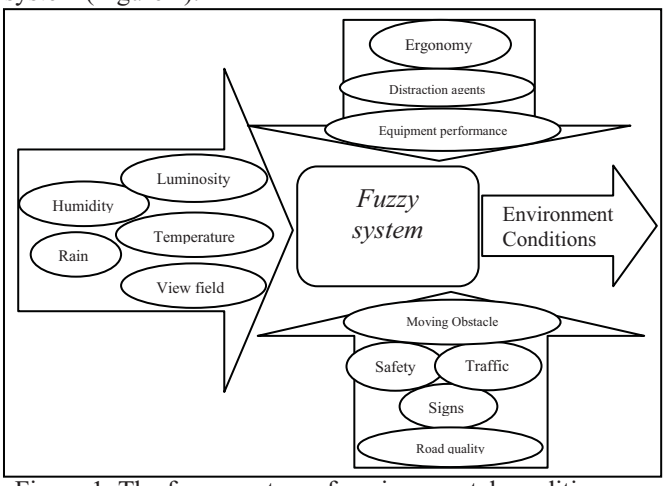


Figure 1. The fuzzy system of environmental condition

2-1-1) *Fuzzy rule base and membership functions*

Questionnaires have been provided to obtain the fuzzy rules, where the effects of climate, road and car condition in driving have been questioned separately. Each expert driver answered each question in three manners, percent value, graphic representation and linguistic term. We allocated a weight for each of them. For percent value $w_{per} = 1$, graphic representation $w_{gr} = 2$ and finally for linguistic term $w_{ling} = 0.5$ have been considered.

The membership functions of fuzzy values are supposed to be bell shape as follows:

$$\mu(x) = \frac{1}{1 + \frac{(x-c)^2}{d}} \quad (6)$$

where x is the member of universe, c is the median and d represents the shape factor.

By using collected data, the median value and shape factor for each expert driver is calculated as follows:

$$c_{ij} = \frac{w_{per} \times val_{per,j} + w_{gr} \times val_{gr,j} + w_{ling} \times val_{ling,j}}{w_{per} + w_{gr} + w_{ling}} \quad (7)$$

$$d_{ij} = \frac{\left(\left|val_{per,j} - val_{gr,j}\right|\right) + \left(\left|val_{per,j} - val_{ling,j}\right|\right) + \left(\left|val_{gr,j} - val_{ling,j}\right|\right)}{3} \quad (8)$$

where c_{ij} is median and d_{ij} is the shape factor suggested by i^{th} driver for the j^{th} parameter. So we have n values of c_{ij} and d_{ij} for n drivers.

Now we can obtain the overall membership function for each of environmental parameters by:

$$\mu_{P_j}(x) = \frac{1}{1 + \frac{(x - c_{P_j})^2}{d_{P_j}}} \quad (9)$$

where c_{P_j} and d_{P_j} are the proposed median and shape factor for the j^{th} parameter. c_{P_j} is obtained based on the number of linguistic terms which are optimized by PSO algorithm (section 2-1-2). For example, suppose the number of linguistic terms and range of universe for a variable are 5 and [0 1] respectively. Then values of medians for five verbal are: 0, 0.25, 0.5, 0.75 and 1 for *very bad*, *bad*, *medium*, *good* and *very good* respectively.

Also d_{P_j} is defined as:

$$d_{Pj} = \frac{1}{2} \times \frac{W_{univ_j} \times d_{ref_j}}{W_{ref\_univ_j} \times (N_{ling\_term_j} - 1)} \quad (10)$$

where W_{univ_j} is the width of universes for j^{th} parameter, $N_{ling\_term_j}$ is the number of linguistic terms for j^{th} parameter (obtained from PSO algorithm), $W_{ref\_univ_j}$ and d_{ref_j} are the width and the shape factor of reference universe for j^{th} parameter. The average of n value of the shape factors (obtained from eq. 8) is considered as d_{ref}. For all of parameters, $W_{ref\_univ_j}$ is fixed to 10.

The importance of the parameters of climate, road and car condition varies for different people. So we use a pair wise comparison table for determining the weights of different parameters. Table1 shows part of pair wise comparison table.

Table1. A part of compared table

Parameter	A	B	C	D
A		√	√	-

The above table shows that parameter A is more important than parameters B and C. But parameter D is more important than parameter A. The proposed method for determining weights of parameters in fuzzy rules is as follows:

First arrange the parameters. First parameter is the least important. Then a_i is assigned to i^{th} parameter such that $a_i = i \times w$. where i is the priority index of parameter and w is the priority weight which is considered as 2 in this paper. When k parameters have equal importance, $a = w \times \frac{\sum i}{k}$ for all of k parameters. After that, we form the priority table as follows:

- If i^{th} parameter is more important than the j^{th} parameter, value of i^{th} row and j^{th} column in table is equal to $T_{ij} = a_i - a_j$.

ISBN: 978-989-95079-6-8

- If the *jth* parameter is more important than the i^{th} one then $T_{ij} = \frac{1}{a_j - a_i}$
- For diagonal elements $T_{ii} = 1$

Finally the weight of each parameter is obtained as follow:

$$M_i = (\prod_{j=1}^{m} T_{ij})^{\frac{1}{m}} \quad , \quad W_i = \frac{M_i}{\sum_{j=1}^{m} M_j} \qquad (11)$$

where *m* is the number of parameters. The obtained weights for parameters are used to generate valid fuzzy rules in section 2-1-2.

The general form of fuzzy rules for climate condition is as follows:

If field of View is … *and* Luminosity is … *and* *Rain* is … *and* Temperature is … *and* Humidity is … *Then* climate condition is …

For road condition:

If Traffic is … *and* Road quality is … *and* Sign is … *and* Moving obstacle is … *and* Safety is … *Then* Road condition is …

Finally for car condition, we have:

If Safety equipment operation is … *and* Ergonomic is … *and* Distract agent is … *Then* Car condition is …

2-1-2) *Particle Swarm Optimization*

Particle Swarm Optimization (PSO) is an optimization algorithm which uses properties of a swarm to find an optimal solution [4]. In this work, the swarm is represented by 100 individuals (or *particles*) whose values change at each iteration. The performance of each particle is measured at each position using a "fitness" function. This function increases as the optimality of the solution increases; in this way, a particle with a higher fitness is considered to fit better than the one with a lower fitness. Also, a record of the best position (*pbest*) for each particle, as well as the best overall position (*gbest*) for all particles, is kept in memory. The entire swarm then searches around the gbest solution and each of the pbest solutions, all the while trying to find even better solutions.

a) particle representation

The number of fuzzy rules depends on the number of linguistic terms of input variables. So, in order to optimize number of fuzzy rules, the number of linguistic terms of each parameter of climate, road and car condition is optimized using PSO. The number of linguistic terms for each input parameter is randomly selected 3, 5 or 7, so maximum number of fuzzy rules will be 7^m (*m* is the number of input variables). The number of linguistic terms of output fuzzy variable is set to 5.

To optimize the number of fuzzy rules and linguistic terms, the following particle is considered for PSO:

Particle=[*par_set par_rules*]

In the presented particle, the first part *par*_set is related to number of linguistic terms and is defined as follows:

$par\_set = (N_{in1}, N_{in2}, \ldots, N_{in_j}, \ldots, N_{in_m})$

Where *m* is the number of input variables and N_{in_j} shows the number of linguistic terms of j^{th} input variable. The number of linguistic terms of output variable is always equal to 5. So the *par_set* has *m* elements.

The second part of particle *par_rules* is:

$par\_rules = (O_{R1}, O_{R2}, \ldots, O_{Ri}, \ldots, O_{R7^m})$

The allele value at each location in the second part of particle contains either zero or the label of an output linguistic value to be used for a given rule. In other words, if O_{Ri} represents the allele at position *i*, its nonzero value gives the consequent part (i.e., the label of the corresponding fuzzy set on the output variable) of the rule which corresponds to the i^{th} location of the rules particle. A particle containing a zero allele value at the i^{th} position (i.e., R_i=0) indicates that the rule set represented by the rule particle has not selected any rule with the i^{th} antecedent clause.

As mentioned before, maximum number of fuzzy rules is 7^m. So in each rule_par, when the number of fuzzy rules is less than 7^m, other elements are considered to be zero. A simple example is considered which has 2 input variables. Both of variables have 3 linguistic terms (named Lt1, Lt, Lt3). So the maximum of fuzzy rules is nine.

All fuzzy rules are as follows:

Rule1: *If* $input_1$ is Lt_1 *and* $input_2$ is Lt_1 *Then* output is O_{R1}
Rule2: *If* $input_1$ is Lt_1 *and* $input_2$ is Lt_2 *Then* output is O_{R2}
Rule3: *If* $input_1$ is Lt_1 *and* $input_2$ is Lt_3 *Then* output is O_{R3}
Rule4: *If* $input_1$ is Lt_2 *and* $input_2$ is Lt_1 *Then* output is O_{R4}
Rule5: *If* $input_1$ is Lt_2 *and* $input_2$ is Lt_2 *Then* output is O_{R5}
Rule6: *If* $input_1$ is Lt_2 *and* $input_2$ is Lt_3 *Then* output is O_{R6}
Rule7: *If* $input_1$ is Lt_3 *and* $input_2$ is Lt_1 *Then* output is O_{R7}
Rule8: *If* $input_1$ is Lt_3 *and* $input_2$ is Lt_2 *Then* output is O_{R8}
Rule9: *If* $input_1$ is Lt_3 *and* $input_2$ is Lt_3 *Then* output is O_{R9}

Since the number of linguistic terms of output variable is 5, then $O_{Ri} \in \{0, 1, 2, 3, 4, 5\}$. Suppose linguistic terms of output variable are {*very bad*, *bad*, *medium*, *good* and *very good*}, therefore, the allele value is 0 when the corresponding rule hasn't been selected as a fuzzy rule, 1 when output variable is *very bad*, 2 when it is *bad*, and so on.

The optimized *par_rules* is obtained by PSO algorithm as follows:

par_rules =(0, 1, 2, 0, 3, 0, 0, 3, 5, 0, … , 0)

where first nine values are related to fuzzy rules. By considering the optimized *par_rules* , We have five rules: Rule2, Rule3, Rule5, Rule8 and Rule9. So optimized fuzzy rules are:

If $input_1$ is Lt_1 *and* $input_2$ is Lt_2 *Then* output is *very bad*
If $input_1$ is Lt_1 *and* $input_2$ is Lt_3 *Then* output is *bad*
If $input_1$ is Lt_2 *and* $input_2$ is Lt_2 *Then* output is *medium*
If $input_1$ is Lt_3 *and* $input_2$ is Lt_2 *Then* output is *medium*
If $input_1$ is Lt_3 *and* $input_2$ is Lt_3 *Then* output is *very good*

b) The fitness function

In the fitness function, we need to use some valid fuzzy rules. In order to generate valid fuzzy rules, we select the worse (value=0) and the best (value=1) state (verbal value) of each input variable. The linguistic terms of them are *very bad* and *very good* respectively. For the middle state (*medium*), value of 0.5 is assigned. If the input variable is defined by 5 linguistic terms, the state between the worse

ISBN: 978-989-95079-6-8

and middle state (*bad*) is set to 0.25 and the state between the best and middle (*good*) is set to 0.75. Also the center of gravity defuzzifier is applied as follows:

$$z^* = \frac{\sum_{i=1}^{m} z_i \times w_i}{\sum_{i=1}^{m} w_i} \quad (12)$$

where z_i *and* w_i are the value and weight (section 2-1-1, eq.11) of i^{th} input variable respectively. All of the rules where z^* gets one of values 0, 0.25, 0.5, 0.75 and 1, corresponding to linguistic terms of input variables, are considered as valid rules.

For example if linguistic term of traffic, road quality, sign, moving obstacle and safety are considered to be medium, then values of all of them is 0.5. Then by considering appropriate weight for each input variable (section 2-1-1, eq.11), z^* will be obtained as follows:

$$z^* = \frac{0.5\times0.21+0.5\times0.13+0.5\times0.11+0.5\times0.21+0.5\times0.34}{0.21+0.13+0.11+0.21+0.34} = 0.5$$

The obtained value is corresponding to *medium* linguistic term for z^*. By this way the following valid rule is obtained:

Rule: If Traffic is medium *and* Road quality is medium *and* Sign is medium *and* Moving obstacle is medium *and* Safety is medium *Then* Road condition is medium

Finally the fitness function used for the optimization of the membership functions is defined by:

$$\text{Fitness} = W_b \times n_c + \frac{W_r}{n_R} \quad (13)$$

Where n_R is the number of rules and W_r is its weighting factor, n_c is the number of input data set which have same output for valid fuzzy rules and the rules generated by PSO and W_r is its weighting factor.

2-2) Driver behavior

The personality (individuality) of each person is very important in the task performance. Because it affects on driver's decisions and acts. We consider drivers personality into three groups, risky person, normal person and attentive person. Also, the people have different reactions in different ages. So we include age of driver in modeling of driver's behavior. In addition, the precision of each person affect on calculating of distance. So we include it by considering graduating diploma and the measure of necessary precision in his work. Figure 2 shows the hierarchical fuzzy system which models driver behavior in decision making. For obtaining fuzzy rules in hierarchical fuzzy system, we use the applied method to generating valid rules in PSO algorithm. In proposed model, only by applying one coefficient in the universes of speed change and steering angle change also the ideal low and high speed for each person, we can easily model the different behavior under take over conditions.

3 High Level Control

3-1) Structure of road and existence cars

Two lanes road is considered as the movement trajectory of cars. In the simulation, four cars are considered (Figure 3). The *A* car is the car which we intend to control it by fuzzy rules, it is called *Controlled car*. The *B* car is the back car which intends to pass *Controlled car* and it is called *BEDOL*. The *C* car is the front car which moves in opposite lane. If *C* car moves in the same direction as *Controlled car,* then it is called *FEDOL* and if direction of movement is in opposite direction then it is called *FODOL*. The *D* car is the front car which moves in the same lane of *Controlled car*. If *D* car moves in the same direction as controlled car then it is called *FEDEL* and if it is in opposite direction then it is called *FODEL*.

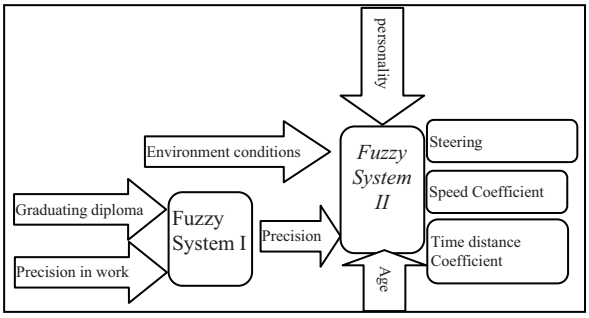


Figure 2. The driver's behavior hierarchy fuzzy system

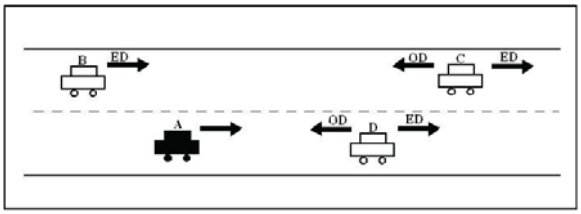
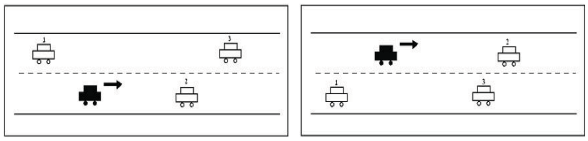


Figure 3. The movement trajectory of cars

3-2) Effective cars in the decision making of driver

When a person drives a car on the road, movements of other cars affect his/her decision making. For example, *FODEL* car is more important than *FEDEL* car and *FODOL* is more important than *FEDOL* car. When driver decides to pass *FEDEL* car, location of *BEDOL* and *FODOL* is very important. The three important cars in decision making of driver are shown in Figure 4.



(a) Controlled car is on the right lane

(b) Controlled car is on the left lane

Figure 4. The situations of important cars

3-3) Decision making process

By considering three positions, the speed and direction of car and the steering angle, the fuzzy variables for steering angle and speed control are presented as follows:

Td_{BOL}: (input variable) time distance of *Controlled car* with the back car in other lane, moving in the same direction.

Td_{FEL}: (input variable) Time distance with the front car, moving in the same lane

ISBN: 978-989-95079-6-8

Td_{FOL} : (input variable) Time distance with the car moving in opposite lane
V: (input variable) Speed of *Controlled car*
Dir: (input variable) *Controlled car*'s direction
Δv : (output variable) Change of speed
$\Delta\varphi$: (output variable) Direction change of steering.

Each of the input variables, Td_{BOL} , Td_{FEL} and Td_{FOL} has four linguistic values or fuzzy sets: The speed has three fuzzy sets. The car direction and the steering angle direction have seven fuzzy sets. Each of output variables is defined by using seven linguistic variables. We define a vector for Td_{BOL} , Td_{FEL} and Td_{FOL} separately. First element of each time distance vector represents the minimum time distance for collision with corresponding car. It is called very low. The second element is the lowest value of the middle time with corresponding car. We call it medium time. The last element is the lowest value as the highest time with corresponding car which is called high time. The above time distance vectors are defined for a *normal driver* who drives according to driving laws. But for modeling different behaviors of driver, we use the time distance coefficient which is obtained by hierarchical fuzzy system. The time distance coefficient is called *K_TD*. The purposed time distance vectors for different drivers are shown in table 2.
In table 2 we have

$$b=\begin{cases}(\sqrt{k\_TD})^{-1}, & K\_TD\geq 1\\ \sqrt{k\_TD}, & K\_TD<1\end{cases}$$

$$Td_{FEL}=\begin{cases}Td_{FEDEL}, & \text{if } FEDEL \text{ exict}\\ Td_{FODEL}, & \text{oherwise}\end{cases}$$

$$Td_{FOL}=\begin{cases}Td_{FEDOL}, & \text{if } FEDOL \text{ exict}\\ Td_{FODOL}, & \text{oherwise}\end{cases}$$

Table 2: Time distance vector for different drivers

Left Lane:	Right Lane:
Td_{BOL}=*K_TD**[0.5/b 2/b 7]	Td_{BOL}=*K_TD**[3 10 15]
Td_{FEDEL}=*K_TD**[1 3 5]	Td_{FEDEL}=*K_TD**[4/b 8/b 12]
Td_{FEDOL}=*K_TD**[3 5 8]	Td_{FEDOL}=*K_TD**[1 3 7]
Td_{FODEL}=*K_TD**[2 4 6]	Td_{FODEL}=*K_TD**[4 7 20]
Td_{FODOL}=*K_TD**[3 7 20]	Td_{FODOL}=*K_TD**[3 7 15]

3-3-1) Road partitions in decision making of driver
Drivers drive in different positions of road. When a driver intends to pass the leading car, he/she moves toward the middle of road while in low speed and dangerous conditions, the car moves toward the shoulder of road. So in decision making, we divide the road to four divisions: Left lane, Middle Lane, Right lane and Shoulder of lane. Among four decision lanes, the middle of lane and the shoulder of road lane are the transient lanes. A driver is driving in the right lane and only when he/she is passing the leading car, he/she drives in the left lane.

3-3-2) Fuzzy rule base for decision making of driver
The decision making process is categorized into four scenarios:
1. Right lane
- Staying in the same lane and continuing the path

a) No car is in front. The car continues its path.
b) A car is in front and the time distance with it is low, the driver starts to make a decision based on his/her desired speed, safety priorities but *FODOL* car is near or there is *BEDOL* car, then the driver decreases the car's speed and continues the path.

- Going to the shoulder of road

A car is in front and the time distance with it is very low; but *FODOL* car is near and *BEDOL* car is near. So the takeover from the left lane is impossible. In order to avoid collision, the driver must move the car to the shoulder of road.

- Going to the middle lane to passing

When the driver decides to change lane and perform a takeover from the left lane. So he/she moves the car to the middle lane.
2. Middle lane
As mentioned before, this lane is a transient lane. If the takeover conditions are satisfied, the driver increases the speed and continues with the same direction to enter the left lane. Otherwise he/she decides to stay in middle lane or to return to right lane.
After taking over, driver with the same speed from this lane enters to right lane
3. Left lane
- Before taking over

If the takeover conditions are satisfied, the driver continues the path with high speed. Otherwise he/she goes to middle lane and decides to stay there until the takeover conditions are provided or he/she returns to right lane.

- After taking over

The driver with the same speed from this lane goes to middle lane in order to go to the right lane.
4. Shoulder of road lane
As mentioned before, a driver in order to avoid collision goes to this lane. So he/she must stay in this lane until he/she could return.

4 Results

The obtained weights of each of environmental parameters from eq. (11) are summarized in Table 3. This table shows that among climate parameters, view field is the most important parameter. Safety is the most important among the road parameters and safety equipment operation is the most important among the car parameters.

Table 3: The weights of environmental parameters in fuzzy rules

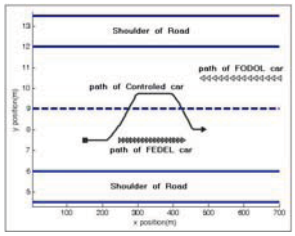
Climate parameter	*W*	Road parameter	*W*	Car parameter	*W*
View	0.37	Traff.	0.21	Eq.o.	0.47
Lum.	0.297	R.q.	0.13		
Rain	0.178	Sign	0.11	Er.	0.25
Temp.	0.098	M.obs	0.21	Dis.a.	0.28
Hum.	0.057	Safety	0.34		

ISBN: 978-989-95079-6-8

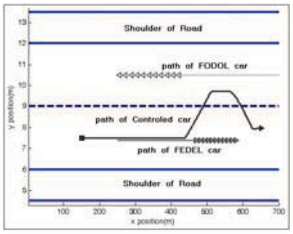
The obtained optimum number of fuzzy sets for environmental parameters input variables by PSO algorithm is shown in Table 4. The number of fuzzy sets of output variables (climate, road and car condition) is fixed and is equal to 5 sets.

Table 4. The optimum number of fuzzy sets for environmental parameters

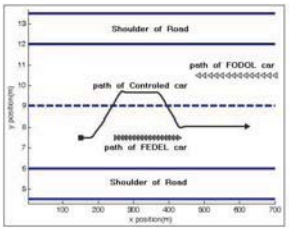
Climate parameter	number	Road parameter	number	Car parameter	number
View	5	Traff.	3	Eq.o.	5
Lum.	3	R.q.	5		
Rain	5	Sign	3	Er.	5
Temp.	3	M.obs.	5	Dis.a.	5
Hum.	3	Safety	3		



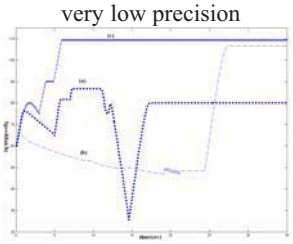
(a) Normal driver



(b) Attentive old driver with very low precision



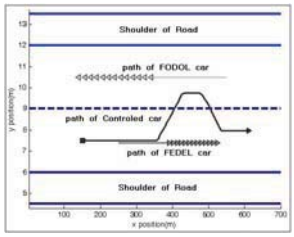
(c) Attentive old driver with very high precision



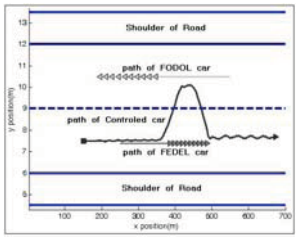
(d) Car's speed by (a), (b), (c) drivers

Figure 5: Comparison of different drivers in the same conditions

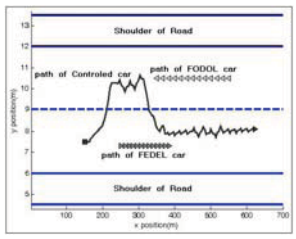
We investigate the behavior of different drivers under take over condition. The comparison of behavior of three drivers under equal passing conditions and the best environmental conditions is shown in Figures 5 and 6. As it is seen in Figure 5, *normal driver* and attentive old driver with very high precision pass the leading car. But attentive old driver with very low precision waits until the *FODOL* passes and then he/she passes the leading car. The paths of cars before passing are shown by solid lines. The very low precision person's error is higher than the very high precision person in calculating distance with other car. Because of being attentive, the driver waits for passing the leading car to collision avoidance. The car's speed of each driver shows the driver's behavior and decision under equal conditions. The *normal driver* is decided to passing and increases the speed. The movement of car is smooth. The old person with very high precision has more steering direction change because he/she is attentive and old and needs more time of passing than normal driver. The old person with very low precision is late in decision making and waits. After that he/she is passing the leading car and returns the right lane as soon as because of being attentive. Figure 6 compares a *normal driver* with a young driver. In equal conditions, the risky young driver passes the leading car. But the attentive young driver and normal driver waits until the *FODOL* car is passed. The speeds of cars are shown in figure (11- d). The results are obtained in good mental status.



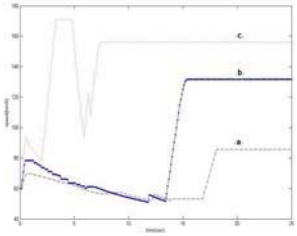
(a) Normal driver



(b) Attentive young driver with very low precision



(c) Risky young driver with very low precision



(d) Car's speed by (a), (b), (c) drivers

Figure 6: The comparison of different driving behavior in equal conditions

5 Conclusion

In this paper a fuzzy three positions model has been presented for the low-level control model and a fuzzy hierarchical system have been proposed for the high-level Control Model. The obtained results show that in equal conditions an attentive old driver with very high precision passes the leading car, but attentive old driver with very low precision waits until *FODOL* car passes and then he/she passes the leading car. The very low precision person's error is higher than that of the very high precision person in calculating distance with other car. This method also provides a basis for modeling individual driver behavior characteristics that may be tuned and used in automatically guided vehicles. It also provides a reference of natural driver behavior of each individual.

References

[1] Hoogendoom, S., Hoogendoom-lanser,S., Shouurman, H. "Perspectives of fuzzy logic in traffic engineering," TRB 78th Annual meeting, USA, 1999.

[2] M. Yoshida and H. Kamada, "A Visual Control System Using Image processing and fuzzy theory," IEEE round table discussion on vision based vehicle guidance, pp. 111-128, 1990.

[3] T. Hessburg and M. Tomizuka, "Fuzzy Logic Control for Lateral Vehicle Guidance," Proc. of IEEE Conference on Control Applications, Vancouver, pp. 581-586, Sept. 1993.

[4] H. Cai, Y. Lin and R. R. Mourant ,"Study on Driver Emotion in Driver-Vehicle-Environment Systems Using Multiple Networked Driving Simulators", DSC 2007 North America – Iowa City – September 2007.

[5] Mark W. Spong, Seth Hutchinson, M. Vidyasagar, "Robot Modeling and Control", John Wiley & Sons, Inc., 2006

[6] J, Kennedy and R. Eberhart, *Swarm Intelligence*, Morgan Kauffman Publishers, San Franscisco, CA. ISBN 1-55860-595-9, 2001.

ISBN: 978-989-95079-6-8

Evaluating Firms' Gender Equity by Fuzzy Logic

Tindara Addabbo[1] Gisella Facchinetti[1] Giovanni Mastroleo[2] Tiziana Lang[3]

1 University of Modena e Reggio Emilia, Department of Economia Politica, Modena, Italia
2 University of Calabria, Cosenza, Italia
3 Isfol, Roma, Italia
Email: tindara.addabbo@unimore.it, gisella.facchinetti@unimore.it, mastroleo@unical.it, t.lang@isfol.it

Abstract—This paper uses a Fuzzy Expert System (FES) in order to evaluate firms in terms of Gender Equity (GE). The scope for using FES is connected not only to the multidimensional nature of GE and to the need of providing a synthetic indicator of firm's GE without losing the its complexity, but even with the composite group of experts involved. The presence of sociologists, economists trade unions members, not used to a mathematics language, carried us to propose an instrument more user friendly like a FES.

Keywords— Fuzzy logic, Fuzzy expert system, gender equity, equal opportunity, discrimination.

1 Introduction

The Italian labour market shows remarkable gender inequalities notwithstanding the advanced labour market regulation in terms of Equal Opportunities. Italian women still experience many inequalities at their workplaces such as: wage differences, lower career paths, higher percentages of fixed-term and short-term contracts, etc. Women's employment rate on average in Italy was 46.6% in year 2007 (Istat, Labour Force Survey data) far below the 60% Lisbon target and far below men's employment rate (70.7%) Inequalities at the workplace are reinforced by and interact with the unequal distribution of unpaid work at home (care and domestic work). Italian women bear a higher share of unpaid work than their partners and when they enter the labour market this is going to produce an unequal share of total (paid and unpaid) work. [3],[10]. Though the Italian situation in terms of gender equity and access to paid labour is particularly weak gender segregation in employment is a feature that characterizes EU27 on average as the results by Burchell, Fagan, O'Brien and Smith [6], based on the European Foundation for the Improvement of living and working conditions 2005 Survey shows. The problem of firms' evaluation with regards to gender equity has been differently tackled across Europe in connection also to the different level of enforcement of equal opportunity laws and a different level of integration of gender equality policies in employment policies [5]. In order to improve women employment rates and to reduce the existing gender gaps in the workplaces the former Italian Ministry of Labour promoted the project "Bollino Rosa" together with Isfol (Istituto per la Formazione Professionale dei Lavoratori). The project started in June 2007 in the framework of the European Year of Equal Opportunities for All and has been co-financed by the European Commission. The test was carried out from November 2007 until February 2008. This study presents the fuzzy expert system (Section 2.1-3) proposed within the Bollino Rosa project and applied to a sample (section 2.4) of firms. Section 3 comments on the results of the application of the system to firms involved in the project while section 4 contains concluding remarks and proposals for the extension of the project.

2 A fuzzy expert system for evaluation

2.1 Why a fuzzy expert system

To face this complex problem and to reach an aggregated value of the certification level, we propose a Fuzzy Expert System (FES), which utilizes fuzzy sets and fuzzy logic to overcome some of the problems that occur when the data provided by the user are vague or incomplete.

This is not the natural framework of a FES, in fact engineering problems are more typical for FES, but recently economic and management researches have found in this instrument interesting applications, [2], [4], [7], [8], [9], [15]. In a multidisciplinary research, like the one we present, the power of FES shows its ability to describe linguistically a particular phenomenon or process, and then to represent that description with a small number of very flexible rules. In a FES, the knowledge is contained both in its rules and in fuzzy sets, which hold general description of the properties of the phenomenon under consideration. FES provides all possible solutions whose truth is above a certain threshold, and the user or the application program can then choose the appropriate solution depending on the particular situation. This fact adds flexibility to the system and makes it more powerful. FES uses fuzzy data, fuzzy rules, and fuzzy inference, in addition to the standard ones implemented in the ordinary Expert Systems. From the mathematical point of view a fuzzy system can be described as a function approximator. [16] More specifically it aims at performing an approximate implementation of an unknown mapping $f : A \subseteq R^n \to R^m$ where A is a compact of R^n. By means of

ISBN: 978-989-95079-6-8

variable knowledge relevant to the unknown mapping, it is possible to prove that that fuzzy systems are dense in the space of continuous functions on a compact domain and so can approximate arbitrarily well any continuous function on a compact domain. The following are the main phases of a FES design ([11], [17]):

- Identification of the problem and choice of the type of FES, which best suits the problem requirement. A modular system can be designed. It consists of several fuzzy modules linked together. A modular approach may greatly simplify the design of the whole system, dramatically reducing its complexity and making it more comprehensible. Fixed input and output variables, it is necessary to describe their linguistic attributes (fuzzy values) and their membership function (fuzzification of input and output);
- definition of the set of heuristic fuzzy rules. (IF-THEN rules);
- choice of the fuzzy inference method (selection of aggregation operators for precondition and conclusion);
- translation of the fuzzy output in a crisp value (defuzzification methods);
- test of the fuzzy system prototype, drawing of the goal function between input and output fuzzy variables, change of membership functions and fuzzy rules if necessary, tuning of the fuzzy system, validation of results.

2.2 System structure

The structure of the Fuzzy Expert System (FES) we present is very complex and contains sixty initial inputs. We try in the space we have to explain how it is born. Preparing the questionnaire, the economics and sociology experts fix the macro-indicators that produce the final evaluation called "certification". The two are "Gender Equity"(GE) and "Gender SusTainability" (GT). The two are macro-aggregations may be split in other sub macro-indicators: GE sums up two: Equity in the Firm (EF) and Equal Opportunity (EO). They include information on what firm makes to treat woman at the same level of man in the policy of entrance, haw firm respects welfare laws and so on. The EF intermediate variable, we present in detail, contains information on the firm's gender fairness in terms of wages and career opportunities. Firms are assessed with respect to employment equity (measured by horizontal segregation in terms of area or type of job), equity in earnings (measured by the gender wage gap and the share of women and men in top positions as a gross index of vertical segregation) and the participation of women as trade unions representatives. The other macro-indicator GT contains several types of information: how firm takes into account work life balance policy (WLB) that is how the way working time is scheduled and its balance with other use of time related to the family. The Human Resource Management (HRM) variables contain information on firm's recruitment methods for entrance into work and on employees leaving or being fired by the firm. The third Safety (S) produces information on security policies that the firm adopts and on the awareness of gender specific health risks at work. Due to limits of space we are going to present more in depth the "Equity in the Firm (EF)". Women tend to be severely underrepresented in Italy in apical positions [13], [14] and their low presence in apical positions in the firm is considered in this system as a negative element in firm's assessment with regards to gender equity both because:

1) it can lead to a higher wage gap at the disadvantage of women;

2) it can be a sign of employment discrimination by gender; 3) it can be a signal of an environment not favourable to women progression of career;

4) it can be a sign of a human resources evaluation system that rewards less women's competences than men's.

The wage gap has been computed by the ratio of women total cost of labour to men's. This is a gross measure of wage gap that, in further research, can be computed as a net wage gap by taking into account employee's characteristics (like work experience, tenure, education, hours of work). We have then considered the bargaining power of women inside the firm by computing how many women are representatives in trade unions with respect to women employed in the firm. An underrepresentation of women in unions representatives contributes to decrease the value of the final outcome since we deem that their voice cannot be enough listened in bargaining on job conditions if they are underrepresented amongst employees representatives. Equal opportunity, EO, takes into account the existence of agreements with unions and public institutions in terms of equal opportunities (the highest mark has been given to enacted agreements and the lowest if there are no agreements and the firm does not

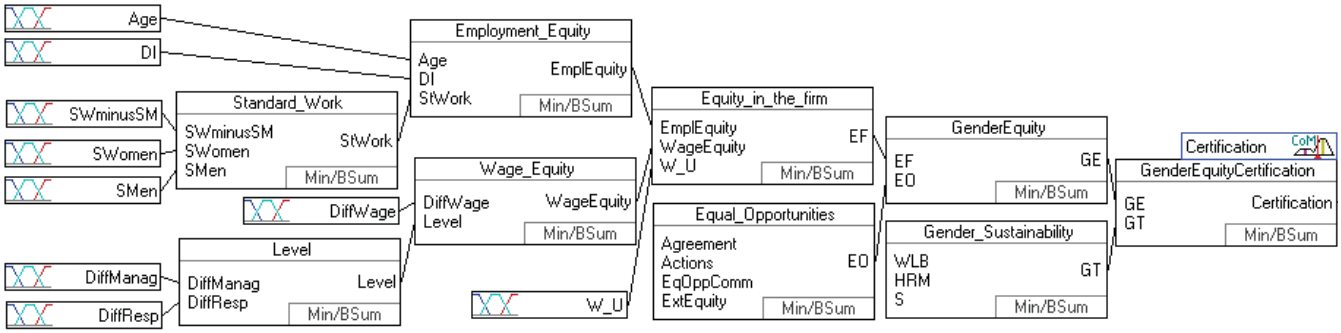


Figure 1: Reduced system layout

ISBN: 978-989-95079-6-8

intend to sign any in the future), of affirmative actions, of incentives to hire women,of the use of Italian Law 53/2000- that contains measures to increase work life balance in the firm - and equal opportunities and the presence of an Equal Opportunity Commission in the firm. We aim at improving the measurement of this last variable by collecting information on the type of activities performed by the Commission. This dimension contains also all strategies enacted by the firm to inform and make aware personnel about gender differences and all the actions that lead to an increase (in qualitative and quantitative terms) of women's participation in the firm. The equal opportunity dimension is also extended to include firm's awareness of the other firms of the network in which it operates.

2.3 Reduced system layout

We present the inputs involved in EF macro-indicator and the sub-system in which the variable GE is split in the original inputs.

Age. Age

SWminusSM. Gender gap in standard employment inside the firm

DI. Dissimilarity index defined as $\frac{1}{2}\sum_i \left|\frac{w_i}{w} - \frac{m_i}{m}\right|$ where m_i and w_i are the numbers of men and women in the i-th area.

SWomen. Women in standard jobs

SMen. Men in standard jobs

DiffMang. Gender gap in managerial position

DiffResp. Gender gap in high responsibility white-collar positions

DiffWage. Gender wage gap

W_U. Women share in union's representatives

Intermediate variables:

EmplEquity. Employment equity

EF. Equity in the firm

EO. Equal opportunities

GE. Gender equity

GT. Gender Sustainability

HRM. Human resources management

Level. Level in the job ladder as percentage of women in apical positions

S. Safety

StWork. Standard work

WageEquity. Wage equity

WLB. Work Life Balance

Now we present the fuzzyfied input variable DiffWage and a rule block in which this variable is involved. DiffWage measures the gross wage gap in the firm as the ratio of women's average labour cost with respect to men's. As it is possible to see, this variable enters directly at the second step and is aggregated with the intermediate variable Level. Level measures the share of women in apical positions with respect to men and can be considered as a measure of vertical segregation against women. Wage equity is then evaluated with respect to both variables to evaluate current wage gap and a cause of wage gap at disadvantage of women. All the cut points that define the fuzzifcation and the rules have been discussed in depth inside the research group where also the researchers who interviewed firm's representatives were involved.

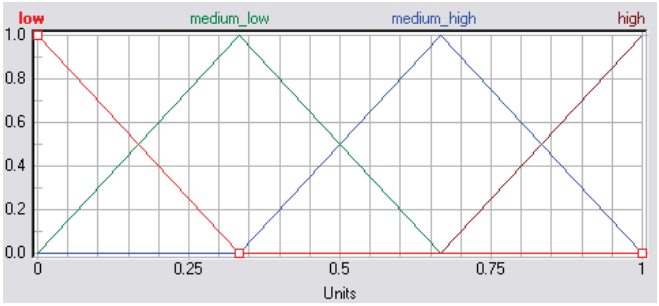


Figure 2: **DiffWage** variable layout

This research step can be deemed very important both in order to state the rules that are at the basis of evaluation in FES and in order to detect problems in the evaluation process and in the tools that have been used.

The following Rule Table is referred to Wage Equity: this intermediate variable is built with the input DiffWage and the intermediate variable Level.

Table 1: rule table of **WageEquity**

#	DiffRetrib	Level	WageEquity
1	low v medium_low	low	very_low
2	low	medium_low	very_low
3	low	medium	low
4	low	medium_high	medium_low
5	low	high	medium_high
6	medium_low	medium_low	low
7	medium_low	medium	medium_low
8	medium_low	medium_high	medium_high
9	medium_low v medium_high	high	high
10	medium_high	low	low
11	medium_high	medium_low	medium_low
12	medium_high	medium	medium_high
13	medium_high	medium_high	high
14	high	low	medium_low
15	high	medium_low	medium_high
16	high	medium	high
17	high	medium_high	very_high

The experts decided that if women's wage cost with respect to men's is low or medium and women's share in managerial and high responsibility white-collar positions are low the wage equity dimension assumes value 'very low'. The experts have suggested this rule that combines wage differentials and gender inequalities in access to apical positions since though the existence of a wage gap at the disadvantage of women can be related to vertical employment segregation one can observe wage differentials also in the same job position. Moreover the measure of the wage gap obtained from the sample of firms, given the structure of the questionnaire, is a gross wage gap (not corrected for employee's level of education or work experience). Inequalities in the access to apical positions can be considered as a proxy of vertical employment segregation by gender to which the experts' group assigns a high weigh in affecting firms' gender equity. This rule block shows that there is a function $f(DiffWage, Level): R^2 \rightarrow R$ increasing

ISBN: 978-989-95079-6-8

in all the two variables but they are weighted in a different way. Looking at rules 5 and 14 we have that: 5) *f(Low, High)→ medium_high;14) f(High, Low)→ medium_low* . This table translates in a simple way what the experts consider correct. This way to translate the experts' position is simpler than asking them numerical weights for the single variables and what type of function better represents their idea.

2.4 The sample of firms to which the system has been applied

The system introduced in this Section has been applied to the gender evaluation of 34 private firms of different size belonging to the different productive sectors that have been selected by the Italian Ministry of Labour through a public call. The sample was formed by: 14 joint-stock companies, 2 Ltds, 3 cooperatives Ltd, 10 social cooperatives and 5 others form of enterprise. According to the number of employees those firms can be classified as follows: 1 micro-enterprise (<10 employees), 12 small and medium enterprises (>10 employees <250), 21 big enterprises of which 4 employ more than 30.000 employees. 21 firms have got a turnover of max 50 millions euro (SMEs) while 5 companies exceed a turnover of 1 billion euro per year. Concerning the economic sectors, 17 firms belong to the service industry (research and training, welfare services, company services), 5 are utility companies (telecommunications and public transports), 3 of them belong to the pharmaceuticals industry, 3 to the retail trade, 1 is an energy producing industry and 1 is a bank. The last 4 firms of the sample belong to other sectors like: edutainment, food production and airport management. The selected firms filled in the self-assessment questionnaires and the sociologists conducted in-depth interviews to the relevant union representatives. The system has been presented and discussed inside a group of experts to whom the results of evaluation has also been submitted to validate the fuzzy outcome.

3 Results

We can start by observing that none of the 34 firms that applied to be evaluated and were selected in the Bollino Rosa project obtained a high result in terms of "Certification value", the final output of the FES presented in the previous section. The firm at the top position in terms reached a final evaluation of 0.68 against the maximum of 1 (Table 2) and the firm at the bottom has a value of 0.25, with 9 firms out of 34 with a value for the final outcome of the system lower than 0.40 (Table 3).At this stage of the project we cannot analyse in depth the situation of each firm in the sample and, for privacy reasons, we cannot provide more details about the firms involved in this experimental phase of the project. In the two tables we call them Ai and Ci if the firm is "not corporate" or "corporate" respectively. Table 2 shows the value of the intermediate and input variables of the first ten firms (in terms of the value of the final output of the system that we call certification) and Table 3 shows the last 10 firms. The tables show how similar output values can be derived by different situation in terms of the dimensions of gendered equity condensed in our intermediate variables. For instance the corporate firm C08 that performs as the not corporate firm A01 In terms of the final value of the FES output (0.62) shows a higher evaluation in terms of gender equity (GE=0.57) than in terms of gender sustainability (GT=0.65), whereas firm A01 performs better in terms of gender sustainability (GT=0.71) than in terms of gender equity (GE=0.53). The same result in terms of certification can therefore mean different outcomes in terms of policies suggestions to the firm in order to improve its 'Bollino Rosa' outcome. A further investigation on the reasons of a lower evaluation of firm A01 in terms of gender equity with respect to firm C08 reveals that this is connected to the high degree of vertical segregation in the firm (with no woman in apical positions) whereas the presence of women in apical position in firm C08 though lower (0.32) is higher than in firm A01. On the other hand both perform rather well in terms of gross wage gap.

Table 2: results the first ten best

Firm	DiffWage	W_U	GE	EF	EmplEquity	WageEquity	EO	Level	GT	Certification
A07	0,90	0,00	0,718	0,576	0,777	0,742	0,800	0,500	0,643	**0,67812**
A04	0,82	0,05	0,593	0,390	0,652	0,291	0,800	0,000	0,714	**0,65588**
A08	1,17	0,38	0,500	0,659	0,535	0,800	0,300	0,500	0,786	**0,65000**
A06	1,39	0,00	0,646	0,469	0,792	0,400	0,800	0,000	0,643	**0,63774**
C05	0,99	0,01	0,635	0,578	0,847	0,796	0,700	0,500	0,625	**0,63042**
A02	1,86	0,00	0,627	0,690	0,805	1,000	0,500	1,000	0,643	**0,62616**
A01	0,79	0,08	0,534	0,365	0,588	0,274	0,700	0,000	0,714	**0,62058**
C08	0,83	0,01	0,570	0,480	0,761	0,568	0,700	0,324	0,650	**0,62004**
C07	0,00	0,72	0,639	0,584	0,842	0,058	0,700	0,322	0,559	**0,60646**
A13	0,00	0,51	0,583	0,500	0,500	0,200	0,700	0,500	0,643	**0,60000**

Table 3: of results the last ten

Firm	DiffWage	W_U	GE	EF	EmplEquity	WageEquity	EO	Level	GT	Certification
A12	0,00	0,00	0,250	0,250	0,573	0,000	0,300	0,000	0,571	**0,40000**
A10	0,00	0,00	0,247	0,245	0,494	0,000	0,300	0,000	0,571	**0,39816**
A19	1,03	0,00	0,250	0,441	0,755	0,400	0,100	0,000	0,500	**0,35000**
A11	0,53	0,03	0,219	0,329	0,620	0,116	0,200	0,000	0,429	**0,32890**
A03	0,00	0,41	0,119	0,304	0,357	0,000	0,000	0,000	0,571	**0,32146**
A23	0,90	0,00	0,166	0,374	0,665	0,339	0,000	0,000	0,500	**0,31022**
A09	0,00	0,00	0,167	0,250	0,500	0,200	0,200	0,500	0,429	**0,30000**
A15	0,00	0,14	0,167	0,250	0,500	0,000	0,100	0,000	0,429	**0,30000**
A21	0,00	0,00	0,167	0,132	0,233	0,000	0,200	0,000	0,484	**0,30000**
A20	0,00	0,12	0,201	0,301	0,575	0,000	0,200	0,000	0,332	**0,24554**

Turning to the last ten firms in terms of the final assessment of gender certification, we can see that the bottom one performs very poorly both in terms of wage equity (0.00) and in terms of the equal opportunities (0.20) and gender sustainability (0.33) dimensions while it scores better (0.575) in terms of employment equity denoting a more equal presence of women and men in the areas of the firm and with regards to the type of contract. This firm is very far from a sufficient level and has to work a lot in all the dimensions to

ISBN: 978-989-95079-6-8

be certified as a 'Bollino Rosa' Firm.
The results shown in the previous section can be considered as a step in the whole process of certification that can use FES tools for assessing firms and addressing specific policies for each firm to reach a final assessment. The results of Fuzzy evaluation will be transmitted to the firms and commented upon by the relevant expert/auditor. According to the results obtained each firm will introduce the needed changes for correcting the gender discriminations. At the end of the evaluation path and only when the organizational situation within the firm will be improved, it will receive a Label that will identify it as a "Bollino Rosa firm", i.e. a gender friendly workplace in terms of equity and sustainability. The process towards certification of gender equity and sustainability involves also an assessment of the steps towards a higher values in the critical dimensions by means of the FES application to evaluate the changes occurred over time. Firms can therefore assess their level of 'Bollino Rosa' connected to the strategies enacted to improve its value.

4 Pursuing in the process of certification

The results shown in the previous section can be considered as a step in the whole process of certification that can use FES tools for assessing firms and addressing specific policies for each firm to reach a final assessment. The results of Fuzzy evaluation will be transmitted to the firms and commented upon by the relevant expert/auditor. According to the results obtained each firm will introduce the needed changes for correcting the gender discriminations. At the end of the evaluation path and only when the organizational situation within the firm will be improved, it will receive a Label that will identify it as a "Bollino Rosa firm", i.e. a gender friendly workplace in terms of equity and sustainability. The process towards certification of gender equity and sustainability involves also an assessment of the steps towards a higher values in the critical dimensions by means of the FES application to evaluate the changes occurred over time. Firms can therefore assess their level of 'Bollino Rosa' connected to the strategies enacted to improve its value. Further research involve considering how the level of gender equity and sustainability in the firm relates to the socioeconomic environment the firm is inserted in. As shown in the following chart the firm operates in connection to local, national government and European Commission that can set rules or adopt policies that are going to affect the firm's situation with regard to Gender Equity or Gender Sustainability. For instance the National Law 53/00 introduces the possibility for firms to receive funds dedicated to the implementation of family friendly policies that can improve gender sustainability in the firm. On the other hand the firm can supply childcare services for their employees but open to other families that can improve gender sustainability of employment in the area. Equal Opportunity Counsellor can take an action to remove discrimination in the firm. The firms that are suppliers or clients can be characterized by a different degree of gender equity and the workers in the firm are going to be affected by the equality or gender sustainability policies of the firm. Trade unions at local or at national level can bargain over working hours schedule to obtain a working time more that balances more with private and family life. The network of relations amongst private and public agents leading to the obtained results can itself be evaluated calling for an extension of the evaluation process to subcontractors, suppliers, clients. The system now includes firms' evaluation on the situation of clients and suppliers with regards to gender equity and this awareness is accounted for as a part of the gender equity dimension. However to have a more precise information on the interactions amongst firms in a gender perspective the whole network of firms should be reconstructed and subject to the evaluation process. Italy is characterized by a highly segmented labour market (by area and type of occupation) and by a different level of diffusion of public care services that can have an impact on gender equity and sustainability of work. Local labour market and institutional variables can be inserted to provide a comparative analysis upon different areas in Italy producing a more complex indicator of gender equity that can be inserted in the evaluation of gender human development.

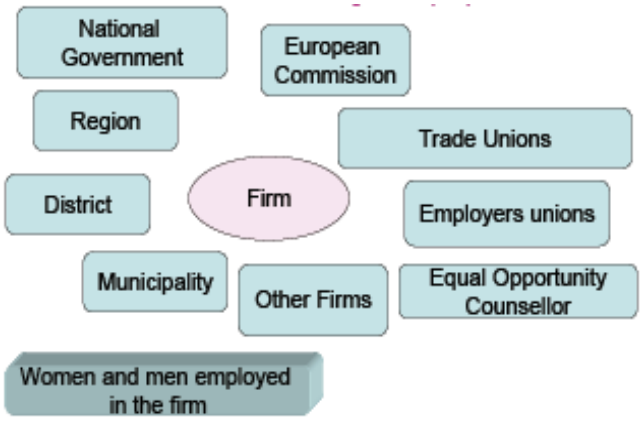


Figure 3: Toward firm's evaluation in a gendered perspective

5 Conclusions

The process of firm's evaluation in terms of gender equity pursued by the project 'Bollino Rosa' promoted by the Italian Ministry of Labour in year 2007 requires collecting and analysing indicators on different dimensions of gender equity. To provide a synthetic indicator on the firm's situation without loosing the complexity of the different dimensions of gender equity we have modelled and applied a Fuzzy Expert System. To our knowledge this is the first attempt of using Fuzzy Expert System to assess firms with respect to gender equity and gender sustainability, though broader issue of the gender perspective in the evaluation of quality of work using fuzzy expert system has been considered in [8]. Other experiences carried out in European countries implies self-evaluation like Total Equity Prize in Germany) or evaluation on the basis of the policies enacted by the firms in terms of equal opportunities, on their human resources management and on the presence of work life balance policies (like Label Egalité in France) however they do not use a system of evaluation structured as the one used

ISBN: 978-989-95079-6-8

in this application [5]. The micro data used have been collected through a structured questionnaire submitted to a sample of firms in the Private Sector that have proposed themselves by answering to a public competition on participation to the experiment, further extension to the public sector and a calibration of the questionnaire should follow its experimental phase. In this line we plan to let firms construct measures of net gender wage gap (by connecting wages to employees' individual characteristics on their work experience and education) that can improve the system of gender equity evaluation (currently based on gross wage gap). With respect to other systems of evaluation the proposed method entails an analytical model of evaluation that confers numerical indicators useful for the firm to assess its position as far as gender equity and sustainability are concerned allowing a backward process to detect causes of poor grades in terms of gender certification. Another advantage of the system is the transparency of the evaluation system reached by a clear statement of input variables and the rules used by the experts. Sometimes the transparency is not an advantage especially in situations in which politics is involved, but we think that a clear and transparent method is better than others in which the choices are expressed in "obscure" words in order to hide real choices from people involved. Public institutions and government can use the results of this model in terms of gender certification of firms to take policy decisions. They can also infer by comparative analyses on different areas how different policies and agreement carried out also at local level may interact with the gender assessment of firms in that area. Policies to consider firms' gender certification in gender equity and sustainability can also be used as an indicator for the assessment of firms when public institutions or firms contract out part of their production. This approach, more typical of engineering problems, finds here an interesting multi-disciplinary application. A very heterogeneous research unit containing mathematics, economics, sociology, political and trade union experts have faced this application. The fuzzy approach has given us all the possibility to work together, to have a common language. This language is due to the linguistic attributes of the variables, the rules given in a verbal way and the final interpretation of the results

Acknowledgment

The Authors thank G.Natoli and V.Cardinali for giving consent to the use of their contributions included in the volume "Strumenti per certificare e promuovere la parità di genere in azienda" [7]. The Authors thank the firms that have been involved in the experimental phase of the project and the Italian Ministry of Labour and other participants to Bollino Rosa project.

References

[1] Addabbo T.- Borghi, V. -Favaro, D."(2006) Differenze di genere nell'accesso a posizioni apicali. Risultati di una ricerca sul campo, in A. Simonazzi (ed.) Questioni di genere, questioni di politica. Trasformazioni economiche e sociali in una prospettiva di genere, Roma, Carocci, 2006.

[2] Addabbo T. – Facchinetti G. - Mastroleo G. (2005). "A fuzzy expert system to measure functionings. An application to child wellbeing". In K. Saeed, R. Mosdorf, J. Pejas, P. Hilmola, Z. Sosnowsky (eds) Proceeding of Image Analysis, Computer Graphics, Security Systems and Artificial Intelligence Applications. Bialystok Vol. 1, 29-42.

[3] Addabbo, T.: Unpaid work by gender in Italy,Ch.2 (2003) in'Unpaid work and the economy', A.Picchio (ed.) Routledge,London and New York (2003)

[4] Addabbo, T. - Di Tommaso, M.L. -Facchinetti, G.(2004): To what extent fuzzy set theory and structural equation modelling can measure functionings? An application to child well being. Materiali di Discussione del Dipartimento di Economia Politica n.468, Modena (2004)

[5] Bassanini, C. (2008) 'Modelli e buone pratiche per la certificazione di genere nelle imprese: un confronto internazionale', in Príncipe G. et al.,(2008). Strumenti per certificare e promuovere la parità di genere in azienda, ISFOL, 2008, Ch.10.

[6] Burchell, B., Fagan, C., O'Brien, C. and Smith, M. (2007) Working conditions in the European Union: The gender perspective, Dublin, European Foundation for the Improvement of living and working conditions.

[7] Addabbo T. – Facchinetti G -Mastroleo G. - Solinas G. (2006). "A fuzzy way to Measure Quality of Work in a multidimensional view" Plenary lecture in the ACS Conference Poland 0ctober 2006. Jerzy Pejas, Imed El Fray, Khalid Saeed Editors Vol I, 13-24. ISBN 83-87362-75-

[8] Addabbo T. – Facchinetti G -Mastroleo G. (2007). "Capability and functiongs: A Fuzzy Way to measure interaction between Father and Child" in Biometrics, Computer Security Systems and Artificial Intelligence Applications. Khalid Saeed, Jerzy Pejas, Romuald Mosdof Editors 185-195, Springer, ISBN 978-0-387-36232-8 (Print) 978-0-387-36503-9 (Online).

[9] Facchinetti G. –Franci F.- Mastroleo -G. Pagliaro V.-Ricci G. (2007) Illogica di un conflitto. La logica fuzzy applicata al conflitto tra Israele e Libano, Eurilink editor ISBN 978-88-95151-04-5

[10] Picchio A. (ed.) (2003) Unpaid work and the economy London and New York: Routledge.

[11] Piegat A. (2001), Fuzzy modelling and control. Springer-Verlag, Heidelberg-New York.

[12] Príncipe G. et al. (2008). Strumenti per certificare e promuovere la parità di genere in azienda, ISFOL, 2008.

[13] Rosti, L.(2006) La segregazione occupazionale in Italia, in A. Simonazzi (ed.) Questioni di genere, questioni di politica. Trasformazioni economiche e sociali in una prospettiva di genere, Roma, Carocci, 2006.

[14] Sabbadini L.L. (2004) Come cambia la vita delle donne, Ministero per le Pari Opportunità, 2004.

[15] Shun-Hsien L. (2005) Expert System Methodologies and application- a decade review from 1995 to 2004. Expert Systems with application. Vol 28 Isuue 1 93-103.

[16] Wang L., (1992), Fuzzy systems are universal approximators Proc. Of Int. Conf. On Fuzzy Engineering, 471-496, 1992.

[17] Von Altrock C. (1997), Fuzzy Logic and NeuroFuzzy applications in Business and Finance, Prentice-Hall Inc.

Morphisms in categories of sets with similarity relations

Jiří Močkoř

Institute for Research and Applications of Fuzzy Modelling, University of Ostrava
Ostrava, Czech Republic
Email: Jiri.Mockor@osu.cz

Abstract— *Morphisms of some categories of sets with similarity relations (Ω-sets) are investigated, where Ω is a complete residuated lattice. Namely a category $\mathbf{SetF}(\Omega)$ with morphisms $(A,\delta)\to(B,\gamma)$ defined as special maps $A\to B$ and a category $\mathbf{SetR}(\Omega)$ with morphisms defined as a special relations $A\times B\to\Omega$. It is proved that arbitrary maps $A\to\Omega$ and $A\times B\to\Omega$ can be extended onto morphisms $(A,\delta)\to(\Omega,\leftrightarrow)$ in $\mathbf{SetF}(\Omega)$ and morphisms $(A,\delta)\to(B,\gamma)$ in $\mathbf{SetR}(\Omega)$, respectively and that these extension processes are from categorical point of view special reflections. Moreover, if Ω is a complete Heyting algebra we also investigate morphisms in another category $\mathbf{Set}(\Omega)$ which consists of classical fuzzy sets $f:A\to\Omega$.*

Keywords— residuated lattice, Heyting algebra, similarity relations, fuzzy sets.

1 Introduction

In fuzzy set theory the category $\mathbf{SetF}(\Omega)$ of sets with similarity relations defined over a complete residuated lattice $\Omega=(\Omega,\to,\otimes,\vee,\wedge)$ is of principal importance ([5]-[10]). This category consists of objects (A,δ) (called Ω-sets), where A is a set and δ is a similarity relation, i.e. a map $\delta:A\times A\to\Omega$ such that

(a) $\delta(x,x)=1$,

(b) $\delta(x,y)=\delta(y,x)$,

(c) $\delta(x,y)\otimes(\delta(y,y)\to\delta(y,z))\le\delta(x,z)$.

A morphism $f:(A,\delta)\to(B,\gamma)$ in $\mathbf{SetF}(\Omega)$ is a map $f:A\to B$ such that $\gamma(f(x),f(y))\ge\delta(x,y)$ for all $x,y\in A$.

From historical point of view there is another category consisting of sets with similarity relations defined. This category $\mathbf{SetR}(\Omega)$ is an analogy of a category of sets with relations between sets as morphisms. Objects of this category $\mathbf{SetR}(\Omega)$ are the same as in the category $\mathbf{SetF}(\Omega)$ and morphism $f:(A,\delta)\to(B,\gamma)$ are maps $f:A\times B\to\Omega$ such that

(a) $(\forall x,z\in A)(\forall y\in B)\quad\delta(z,x)\otimes f(x,y)\le f(z,y)$,

(b) $(\forall x\in A)(\forall y,z\in B)\quad f(x,y)\otimes\gamma(y,z)\le f(x,z)$,

If $f:(A,\delta)\to(B,\gamma)$ and $g:(B,\gamma)\to(C,\omega)$ are two morphisms then their composition is a relation $g\circ f:A\times C\to\Omega$ such that

$$g\circ f(x,z)=\bigvee_{y\in B}(f(x,y)\otimes g(y,z)).$$

Finally we can consider a classical category $\mathbf{Set}(\Omega)$ of fuzzy sets over Ω with objects couples (A,f), where A is a set and f is a map $A\to\Omega$ and with morphisms $u:(A,f)\to(B,g)$ such that $u:A\to B$ is map and $f(a)\le g\circ u(a)$ for all $a\in A$.

With categories $\mathbf{SetF}(\Omega)$ and $\mathbf{SetR}(\Omega)$ we can consider new sets of objects:

(a) the set $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ of all morphisms in $\mathbf{SetR}(\Omega)$,

(b) the set $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ of all *weak morphisms* in $\mathbf{SetR}(\Omega)$, where $f:(A,\delta)\to(B,\gamma)$ is a weak morphism in $\mathbf{SetR}(\Omega)$ if $f:A\times B\to\Omega$ is an Ω-valued relation,

(c) the set $\mathbf{Map}(\mathbf{SetF}(\Omega))$ of all *weak fuzzy sets* in $\mathbf{SetF}(\Omega)$, where $f:(A,\delta)\to(\Omega,\leftrightarrow)$ is a weak fuzzy set in $\mathbf{SetF}(\Omega)$ if $f:A\to\Omega$ is a map and where $\leftrightarrow$ is a biresiduation in a lattice Ω.

(d) the set $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ of all *fuzzy sets* in $\mathbf{SetF}(\Omega)$, i.e. morphisms $(A,\delta)\to(\Omega,\leftrightarrow)$ in the category $\mathbf{SetF}(\Omega)$, where (A,δ) is an Ω-set.

It is clear that we have

$$\mathbf{Mor}(\mathbf{SetR}(\Omega))\hookrightarrow\mathbf{Rel}(\mathbf{SetR}(\Omega)),$$
$$\mathbf{Fuz}(\mathbf{SetF}(\Omega))\hookrightarrow\mathbf{Map}(\mathbf{SetF}(\Omega)).$$

In this paper we want to consider some relationships between these sets. We show that new categories (denoted by the same symbols as the above sets) can be defined with the above sets as object sets. Then we will receive the following functor commutative diagram:

$$\begin{array}{ccc}\mathbf{Fuz}(\mathbf{SetF}(\Omega)) & \xrightarrow{\ \hookrightarrow\ } & \mathbf{Map}(\mathbf{SetF}(\Omega))\\ {\scriptstyle F}\downarrow & & \downarrow{\scriptstyle F}\\ \mathbf{Mor}(\mathbf{SetR}(\Omega)) & \xrightarrow{\ \hookrightarrow\ } & \mathbf{Rel}(\mathbf{SetR}(\Omega))\end{array}$$

Finally, we show that

(a) the category $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ is a full reflective subcategory in the category $\mathbf{Rel}(\mathbf{SetR}(\Omega))$,

(b) the category $\mathbf{Map}(\mathbf{SetF}(\Omega))$ is a full reflective subcategory in the category $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$.

The reflections $f\to\tilde{f}$ and $g\to\widehat{g}$ of a weak morphism $f:(A,\delta)\to(B,\gamma)$ and a weak fuzzy set $g:(A,\delta)\to(\Omega,\leftrightarrow)$ can be then used for construction of a fuzzy logic formulas interpretation in corresponding categories $\mathbf{SetF}(\Omega)$ and $\mathbf{SetR}(\Omega)$.

ISBN: 978-989-95079-6-8

2 Categories of maps and relations

Let us firstly define two categories $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ and $\mathbf{Mor}(\mathbf{SetR}(\Omega))$. Objects of the category $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ are all weak morphisms from $\mathbf{SetR}(\Omega)$. A morphism from a weak morphism $f : (A,\delta) \to (B,\gamma)$ to a weak morphism $g : (C,\rho) \to (D,\omega)$ is a couple (u,v) of morphisms from the category $\mathbf{SetR}(\Omega)$ such that

(i) $u : (A,\delta) \to (C,\rho)$ is a morphism in $\mathbf{SetR}(\Omega)$,

(ii) $v : (B,\gamma) \to (D,\omega)$ is a morphism in $\mathbf{SetR}(\Omega)$,

(iii) $(v \circ f)(a,d) \le (g \circ u)(a,d)$ for all $a \in A, d \in D$, where the composition of a morphism and a weak morphism is formally the same as for morphisms in the category $\mathbf{SetR}(\Omega)$.

In that case we say that the diagram

$$\begin{array}{ccc} (A,\delta) & \xrightarrow{f} & (B,\gamma) \\ {\scriptstyle u}\downarrow & & \downarrow{\scriptstyle v} \\ (C,\rho) & \xrightarrow{g} & (D,\omega) \end{array}$$

fuzzy commutes. If $(u,v) : f \to g$ and $(u_1,v_1) : g \to h$ are two morphisms in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ such that $f : (A,\delta) \to (B,\gamma)$, $g : (C,\rho) \to (D,\omega)$, $h : (E,\sigma) \to (F,\tau)$ are weak morphisms then a composition $(u_1,v_1) \circ (u,v)$ is defined as $(u_1 \circ u, v_1 \circ v)$. Using a composition of weak morphisms we can easily prove that the following diagram fuzzy commutes

$$\begin{array}{ccc} (A,\delta) & \xrightarrow{f} & (B,\gamma) \\ {\scriptstyle u_1\circ u}\downarrow & & \downarrow{\scriptstyle v_1\circ v} \\ (E,\sigma) & \xrightarrow{h} & (F,\tau), \end{array}$$

i.e. the definition is correct.

Objects of the category $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ are all morphisms $f : (A,\delta) \to (B,\gamma)$ in the category $\mathbf{SetR}(\Omega)$ and morphisms between such objects are defined formally in the same way as for the category $\mathbf{Rel}(\mathbf{SetR}(\Omega))$. It is clear that $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ is a full subcategory of the category $\mathbf{Rel}(\mathbf{SetR}(\Omega))$.

Now, objects of the category $\mathbf{Map}(\mathbf{SetF}(\Omega))$ are all weak fuzzy sets $(A,\delta) \to (\Omega,\leftrightarrow)$ in the category $\mathbf{SetF}(\Omega)$, where (A,δ) are Ω-sets. If $f : (A,\delta) \to (\Omega,\leftrightarrow)$ and $g : (B,\gamma) \to (\Omega,\leftrightarrow)$ are objects in $\mathbf{Map}(\mathbf{SetF}(\Omega))$ then $u : f \to g$ is a morphism in $\mathbf{Map}(\mathbf{SetF}(\Omega))$ if $u : (A,\delta) \to (B,\gamma)$ is a morphism in $\mathbf{SetF}(\Omega)$ and $g \circ u(a) \ge f(a)$ for all $a \in A$. A composition of such morphisms is defined as a composition of corresponding morphisms in the category $\mathbf{SetF}(\Omega)$.

Finally objects of the category $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ are all fuzzy sets in the category $\mathbf{SetF}(\Omega)$, i.e. morphisms $(A,\delta) \to (\Omega,\leftrightarrow)$ in the category $\mathbf{SetF}(\Omega)$. Morphisms in the category $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ are defines similarly as in the category $\mathbf{Map}(\mathbf{SetF}(\Omega))$.

It is well known that there exists a functor $F : \mathbf{SetF}(\Omega) \to \mathbf{SetR}(\Omega)$ such that F is an identity on objects and if $f : (A,\delta) \to (B,\gamma)$ is a morphism in $\mathbf{SetF}(\Omega)$ then $F(f) : A \times B \to \Omega$ is such that $F(f)(a,b) = \gamma(f(a),b)$.

Lemma 1
Let $g : (A,\delta) \to (B,\gamma)$ be a weak morphism in $\mathbf{SetR}(\Omega)$. Let $\tilde{g} : A \times B \to \Omega$ be defined by the formula

$$\tilde{g}(a,b) = \bigvee_{x\in A}\bigvee_{y\in B} g(x,y) \otimes \delta(a,x) \otimes \gamma(b,y).$$

Then

(a) $\tilde{g}$ is a morphism in $\mathbf{SetR}(\Omega)$,

(b) $\tilde{g} = \bigwedge\{h : h$ is a morphism $(A,\delta) \to (B,\gamma)$ in $\mathbf{SetR}(\Omega), h \ge g\}$,

(c) If g is a morphism in $\mathbf{SetR}(\Omega)$, then $\tilde{g} = g$.

Lemma 2
Let $s : (A,\delta) \to (\Omega,\leftrightarrow)$ be a weak fuzzy set in $\mathbf{SetF}(\Omega)$. Let a map $\widehat{s} : A \to \Omega$ be defined such that $\widehat{s}(a) = \bigvee_{x\in A} \delta(a,x) \otimes s(x)$ for all $a \in A$. Then

(a) $\widehat{s} : (A,\delta) \to (\Omega,\leftrightarrow)$ is a morphism in $\mathbf{SetF}(\Omega)$,

(b) $\widehat{s} = \bigwedge\{t : t$ is a morphism $(A,\delta) \to (\Omega,\leftrightarrow)$ in $\mathbf{SetF}(\Omega), t \ge s\}$.

(c) If $s : (A,\delta) \to (\Omega,\leftrightarrow)$ is a fuzzy set in $\mathbf{SetF}(\Omega)$ then $\widehat{s} = s$,

Now
we define functors $G : \mathbf{Rel}(\mathbf{SetR}(\Omega)) \to \mathbf{Mor}(\mathbf{SetR}(\Omega))$ and $H : \mathbf{Map}(\mathbf{SetF}(\Omega)) \to \mathbf{Fuz}(\mathbf{SetF}(\Omega))$ which will be reflections.

Proposition 1
(a) Let a weak morphism f be an object from $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ and let $(u,v) : f \to g$ be a morphism in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$. Let $G(f) = \tilde{f}$ and $G(u,v) = (u,v)$. Then

$$G : \mathbf{Rel}(\mathbf{SetR}(\Omega)) \to \mathbf{Mor}(\mathbf{SetR}(\Omega))$$

is a functor.

(b) Let a weak fuzzy set f be an object in $\mathbf{Map}(\mathbf{SetF}(\Omega))$ and let $u : f \to g$ be a morphism in $\mathbf{Map}(\mathbf{SetF}(\Omega))$. Let $H(f) = \widehat{f}$ and $H(u) = u$. Then

$$H : \mathbf{Map}(\mathbf{SetF}(\Omega)) \to \mathbf{Fuz}(\mathbf{SetF}(\Omega))$$

is a functor.

Proof. (a) Let $f : (A,\delta) \to (B,\gamma)$ and $g : (C,\rho) \to (D,\tau)$ be weak morphisms. We show that (u,v) is a morphism $\tilde{f} \to \tilde{g}$. Since $(u,v) : f \to g$ is a morphism in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$, we have $v \circ f \le g \circ u$. Then since v and u are morphisms in $\mathbf{SetR}(\Omega)$, we have

$$\begin{aligned} v \circ \tilde{f}(a,d) = \\ \bigvee_{b,b',a'} f(a',b') \otimes \delta(a,a') \otimes \gamma(b,b') \otimes v(b,d) \le \\ \bigvee_{b',a'} f(a',b') \otimes \delta(a,a') \otimes v(b',d) = \\ \bigvee_{a'} v \circ f(a',d) \otimes \delta(a,a') \le \\ \bigvee_{a'} g \circ u(a',d) \otimes \delta(a,a') \le \bigvee_{c} u(a,c) \otimes g(c,d) \le \\ \tilde{g} \circ u(a,d). \end{aligned}$$

ISBN: 978-989-95079-6-8

(b) Let $f : (A,\delta) \to (\Omega,\leftrightarrow)$ and $g : (B,\gamma) \to (\Omega,\leftrightarrow)$ be weak fuzzy sets. We need to show that $\widehat{g} \circ u \geq \widehat{f}$. Since u is a morphism in $\mathbf{SetF}(\Omega)$, we have

$$\widehat{g} \circ u(a) = \bigvee_{b\in B} g(b) \otimes \gamma(b,u(a)) \geq \bigvee_{x\in A} g(u(x)) \otimes \gamma(u(x),u(a)) \geq \bigvee_{x\in A} g \circ u(x) \otimes \delta(x,a) \geq \bigvee_{x\in A} f(x) \otimes \delta(x,a) = \widehat{f}(a).$$

We now introduce a notion of a fuzzy reflective subcategory.

Theorem 1

(a) $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ *is a full reflective subcategory in the category* $\mathbf{Rel}(\mathbf{SetR}(\Omega))$ *and* G *is a reflection.*

(b) $\mathbf{Map}(\mathbf{SetF}(\Omega))$ *is a full reflective subcategory in the category* $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ *and* H *is a reflection.*

Proof. (a) Let $f : (A,\delta) \to (B,\gamma)$ be an object in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$. It is clear that $(id_A, id_B) : f \to \tilde{f}$ is a morphism in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$, where for an object (X,σ) from $\mathbf{SetR}(\Omega)$ the identity morphism $id_X : (X,\sigma) \to (X,\sigma)$ is defined such that $id_X : X \times X \to \Omega$ is such that $id_X(x,y) = 1_\Omega$ if $x = y$ and 0_Ω, otherwise. Let $g : (C,\rho) \to (D,\tau)$ be an object in $\mathbf{Mor}(\mathbf{SetR}(\Omega))$ and let $(u,v) : f \to g$ be a morphism in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$. Then we show that (u,v) is the unique morphism such that the diagram commutes:

$$\begin{array}{ccc} f & \xrightarrow{(id_A,id_B)} & \tilde{f} \\ \| & & \downarrow{\scriptstyle (u,v)} \\ f & \xrightarrow{(u,v)} & g \end{array}$$

In fact, it suffices to prove that $(u,v) : \widetilde{f} \to g$ is a morphism in $\mathbf{Rel}(\mathbf{SetR}(\Omega))$, i.e. that the diagram

$$\begin{array}{ccc} (A,\delta) & \xrightarrow{\widetilde{f}} & (B,\gamma) \\ {\scriptstyle u}\downarrow & & \downarrow{\scriptstyle v} \\ (C,\rho) & \xrightarrow{g} & (D,\tau) \end{array}$$

fuzzy commutes. We have $v \circ f(a,d) \leq g \circ u(a,d)$ for all $a \in A, d \in D$. Since u and v are morphisms in a category $\mathbf{SetR}(\Omega)$, we have

$$v \circ \tilde{f}(a,d) = \bigvee_{b,b'\in B, a'\in A} f(a',b') \otimes \delta(a,a') \otimes \gamma(b,b') \otimes v(b,d) \leq \bigvee_{b,b'\in B a'\in A} f(a',b') \otimes \delta(a,a') \otimes v(b',d) = \bigvee_{a'\in A} \delta(a,a') \otimes v \circ f(a',d) \leq \bigvee_{a'\in A} \delta(a,a') \otimes g \circ u(a',d) \leq g \circ u(a,d).$$

It is clear that (u,v) is the unique morphism with such property and that the diagram commutes.

(b) Let $f : (A,\delta) \to (\Omega,\leftrightarrow)$ be an object of $\mathbf{Map}(\mathbf{SetF}(\Omega))$. It is clear that $id_A : f \to \widehat{f}$ is a morphism in $\mathbf{Map}(\mathbf{SetF}(\Omega))$. Let $g : (B,\gamma) \to (\Omega,\leftrightarrow)$ be an object in $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ and let $u : f \to g$ be a morphism in $\mathbf{Map}(\mathbf{SetF}(\Omega))$. Then u is the unique morphism in $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$ such that the diagram commutes:

$$\begin{array}{ccc} f & \xrightarrow{id_A} & \widehat{f} \\ {\scriptstyle u}\downarrow & & \downarrow{\scriptstyle u} \\ g & = & g. \end{array}$$

In fact, since $g \circ u(a) \geq f(a)$ and $g \circ u$ is an object in $\mathbf{Fuz}(\mathbf{SetF}(\Omega))$, we have

$$\widehat{f}(a) = \bigvee_{x\in A} \delta(a,x) \otimes f(x) \leq \bigvee_{x\in A} \delta(a,x) \otimes g \circ u(x) \leq g \circ u(a).$$

3 Reflections and fuzzy logic models

The reflections G, H can be used for a definition of fuzzy logic formulas interpretation in categories $\mathbf{SetR}(\Omega)$ and $\mathbf{SetF}(\Omega)$, respectively (see [10]). Let J be a first order language of a fuzzy logic which consists (as classically) of a set of predicate symbols $P \in \mathscr{P}$, a set of functional symbols $f \in \mathscr{F}$ and a set of classical logical connectives $\{\wedge, \vee, \Rightarrow, \neg, \otimes\}$. Moreover J contains also a set Ω of logical constants.

Definition 1

Let $\mathbf{K}$ *be a category with products and with* Ω*-sets as objects. Then a model of a language* J *in a category* $\mathbf{K}$ *is*

$$\mathscr{D} = ((A,\delta), \{P_{\mathscr{D}} : P \in \mathscr{P}\}, \{f_{\mathscr{D}} : f \in \mathscr{F}\}),$$

where

(a) (A,δ) *is an* Ω*-set from a category* $\mathbf{K}$*,*

(b) $P_{\mathscr{D}} : (A,\delta) \times \cdots \times (A,\delta) \to (\Omega,\leftrightarrow)$ *is a morphism in* $\mathbf{K}$*,*

(c) $f_{\mathscr{D}} : (A,\delta) \times \cdots \times (A,\delta) \to (A,\delta)$ *is a morphism in a category* $\mathbf{K}$*.*

Further, let ψ (t, respectively) be a formula (term, respectively) with free variables contained in a set X of variables. Then an interpretation $\|\psi\|_{\mathscr{D},X}$ ($\|t\|_{\mathscr{D},X}$, respectively) of ψ (t, respectively) in a model $\mathscr{D}$ in a category $\mathbf{K}$ should be defined such that

(a) $\|\psi\|_{\mathscr{D},X} : (A,\delta)^X \to (\Omega,\leftrightarrow)$ is a morphism in $\mathbf{K}$,

(b) $\|t\|_{\mathscr{D},X} : (A,\delta)^X \to (A,\delta)$ is a morphism in $\mathbf{K}$,

where $(A,\delta)^X$ is a product $(A,\delta)^{|X|} = (A^{|X|}, \delta_X)$ in a category $\mathbf{K}$. We show shortly how by using reflections H and G the definition of a formula interpretation can be done in models $\mathscr{D}$ in categories $\mathbf{K} = \mathbf{SetF}(\Omega), \mathbf{SetR}(\Omega)$.

Let $\mathscr{D} = ((A,\delta), \{P_{\mathscr{D}} : P \in \mathscr{P}\}, \{f_{\mathscr{D}} : f \in \mathscr{F}\})$ be a model of a language J in a category $\mathbf{K}$, where $\mathbf{K} = \mathbf{SetF}(\Omega)$ or $\mathbf{K} = \mathbf{SetR}(\Omega)$, i.e.

ISBN: 978-989-95079-6-8

(i) (A,δ) is a Ω-set from $\mathbf{K}$,

(ii) $P_{\mathscr{D}} : (A,\delta)\times\cdots\times(A,\delta)\to(\Omega,\leftrightarrow)$ is a morphism in a category $\mathbf{K}$,

(iii) $f_{\mathscr{D}} : (A,\delta)\times\cdots\times(A,\delta)\to(A,\delta)$ is a morphism in a category $\mathbf{K}$.

Let t be a term with a set of variables contained in a set X. Then $\|t\|_{\mathscr{D},X} = \|t\|_X : (A,\delta)^X \to (A,\delta)$ is a morphism in $\mathbf{K}$ defined as follows.

(i) Let $t = x$, where $x \in X$. Then $\|t\|_X := pr_x : (A,\delta)^X \to (A,\delta)$ is a projection morphism in a category $\mathbf{K}$.

(ii) Let $t = f(t_1,\dots,t_n)$. Then $\|t\|_X$ is a composition (in $\mathbf{K}$) of morphisms

$$(A,\delta)^X \xrightarrow{\prod_i \|t_i\|_X} (A,\delta)^n \xrightarrow{f_{\mathscr{D}}} (A,\delta)\cdot$$

Hence, for $\mathbf{K} = \mathbf{SetR}(\Omega)$ we have $\|t\|_X(\mathbf{a},b) = \bigvee_{\mathbf{x}\in A^n}(\prod_i \|t_i\|_X(\mathbf{a},\mathbf{x})\otimes f_{\mathscr{D}}(\mathbf{x},b)$, where $\mathbf{a}\in A^X$.

Now let ψ be a formula with free variables contained in a set X. In a category $\mathbf{K}$ we will define firstly a weak morphisms (i.e. *weak fuzzy set* for $\mathbf{K} = \mathbf{SetF}(\Omega)$ or a *weak morphism* for $\mathbf{K} = \mathbf{SetR}(\Omega)$) $|\psi|_{\mathscr{D},X} = |\psi|_X : (A,\delta)^X \to (\Omega,\leftrightarrow)$. A definition will be done by induction principle on a structure of ψ.

(a) Let $\psi \equiv P(t_1,\dots,t_n)$. Then (according to induction assumption) $\|t_i\|_X, P_{\mathscr{D}}$ are defined and we define $|\psi|_X$ as a composition of the following morphisms in $\mathbf{K}$:

$$(A,\delta)^X \xrightarrow{\prod_i \|t_i\|_X} (A,\delta)^n \xrightarrow{P_{\mathscr{D}}} (\Omega,\leftrightarrow).$$

(b) Let $\psi \equiv t_1 = t_2$. Then $|\psi|_X$ is a composition of the following weak morphisms:

$$(A,\delta)^X \xrightarrow{\|t_1\|_X\times\|t_2\|_X} (A,\delta)^2 \xrightarrow{\Delta_{\mathbf{K},A}} (\Omega,\leftrightarrow).$$

(c) Let $\psi \equiv \psi_1\wedge\psi_2$. Then $|\psi|_X$ is a composition of the following weak morphisms:

$$(A,\delta)^X \xrightarrow{\|\psi\|_X\times\|\psi_2\|_X} (\Omega,\leftrightarrow)^2 \xrightarrow{\sqcap_{\mathbf{K}}} (\Omega,\leftrightarrow).$$

(d) Let $\psi \equiv \psi_1\vee\psi_2$. Then $|\psi|_X$ is a composition of the following weak morphisms:

$$(A,\delta)^X \xrightarrow{\|\psi\|_X\times\|\psi_2\|_X} (\Omega,\leftrightarrow)^2 \xrightarrow{\sqcup_{\mathbf{K}}} (\Omega,\leftrightarrow).$$

(e) Let $\psi \equiv \sigma\Rightarrow\tau$. Then $|\psi|_X$ is a composition of the following weak morphisms:

$$(A,\delta)^X \xrightarrow{\|\sigma\|_X\times\|\tau\|_X} (\Omega,\leftrightarrow)^2 \xrightarrow{\Rrightarrow_{\mathbf{K}}} (\Omega,\leftrightarrow).$$

(f) Let $\psi \equiv \neg\sigma$. Then $|\psi|_X$ is a composition of the following weak morphisms:

$$(A,\delta)^X \xrightarrow{\|\sigma\|_X} (\Omega,\leftrightarrow) \xrightarrow{\neg_{\mathbf{K}}} (\Omega,\leftrightarrow).$$

(g) Let $\psi \equiv (\exists x)\sigma$. Then $\|\sigma\|_{X\cup\{x\}}$ is already defined as a morphism $(A,\delta)^{X\cup\{x\}} = (A,\delta)^X\times(A,\delta)\to(\Omega,\leftrightarrow)$ in $\mathbf{K}$. Then we set

$$|\psi|_X(\mathbf{a},\alpha) = \bigvee_{x\in A}\|\sigma\|_{X\cup\{x\}}((\mathbf{a},x),\alpha), \quad \text{if } \mathbf{K} = \mathbf{SetR}(\Omega),$$
$$|\psi|_X(\mathbf{a}) = \bigvee_{x\in A}\|\sigma\|_{X\cup\{x\}}(\mathbf{a},x), \quad \text{if } \mathbf{K} = \mathbf{SetR}(\Omega).$$

The maps $\Rrightarrow_{\mathbf{K}}, \neg_{\mathbf{K}}, \sqcup_{\mathbf{K}}, \sqcap_{\mathbf{K}}, \Delta_{\mathbf{K},A}$ are defined as follows: For $\mathbf{K} = \mathbf{SetR}(\Omega)$, we have

(i) $\Delta_{\mathbf{SetR}(\Omega),A}((a,b),\alpha) = \alpha \leftrightarrow \delta(a,b)$ for all $a,b\in A, \alpha\in\Omega$.

(ii) $\sqcap_{\mathbf{SetR}(\Omega)}((\beta,\gamma),\alpha) = \alpha\leftrightarrow(\beta\wedge\gamma)$.

(iii) $\sqcup_{\mathbf{SetR}(\Omega)}((\beta,\gamma),\alpha) = \alpha\leftrightarrow(\beta\vee\gamma)$.

(iv) $\Rrightarrow_{\mathbf{SetR}(\Omega)}((\beta,\gamma),\alpha) = (\beta\otimes\alpha)\to\gamma$.

(v) $\neg_{\mathbf{SetR}(\Omega)}(\alpha,\beta) = \beta\leftrightarrow(\alpha\to 0)$.

For $\mathbf{K} = \mathbf{SetF}(\Omega)$, we have

(a) $\Delta_{\mathbf{SetF}(\Omega),A}(a,b) = \delta(a,b)$ for all $a,b\in A$.

(b) $\sqcap_{\mathbf{SetF}(\Omega)}(\beta,\gamma) = \beta\wedge\gamma$.

(c) $\sqcup_{\mathbf{SetF}(\Omega)}(\beta,\gamma) = \beta\vee\gamma$.

(d) $\Rrightarrow_{\mathbf{SetF}(\Omega)}(\beta,\gamma) = \beta\to\gamma$.

(e) $\neg_{\mathbf{SetF}(\Omega)}(\beta) = \neg\beta$.

Definition 2
Let ψ be a formula in J and let X be a set of variables containing all free variables of ψ. Let $\mathscr{D}$ be a model of J in a category $\mathbf{K} = \mathbf{SetF}(\Omega), \mathbf{SetR}(\Omega)$. Then we set

$$\|\psi\|_{\mathscr{D},X} = G(|\psi|_{\mathscr{D},X}) \quad \textit{if} \quad \mathbf{K} = \mathbf{SetR}(\Omega),$$
$$\|\psi\|_{\mathscr{D},X} = H(|\psi|_{\mathscr{D},X}) \quad \textit{if} \quad \mathbf{K} = \mathbf{SetF}(\Omega).$$

Theorem 2
For any formula ψ, $\|\psi\|_{\mathscr{D},X} : (A,\delta)^X \to (\Omega,\leftrightarrow)$ is a morphism in a category $\mathbf{K} = \mathbf{SetR}(\Omega), \mathbf{SetF}(\Omega)$.

4 Category $\mathbf{Set}(\Omega)$

To investigate a relationship between classical fuzzy sets (A,f), where $f : A\to\Omega$ is a map and Ω-sets (A,δ) it will be necessary to modify a definition of a similarity relation δ. There are several natural ways how to transform a classical fuzzy set into a similarity relation. For example, for a fuzzy set f we can define a similarity relation σ_f such that $\sigma_f(x,y) = f(x)\leftrightarrow f(y)$. We can also introduce another relation, namely $\Delta_f(x,y) = f(x)$ iff $x = y$ and $\Delta_f(x,y) = 0$, otherwise. It is clear that in that case Δ_f does not satisfies the condition $\Delta_f(x,x) = 1$ from the definition of similarity relations. Nevertheless to investigate a relationship between fuzzy sets and such similarity relations, we will introduce in this section a generalization of Ω-sets, namely we will define a new category $\mathbf{SetR}(\Omega)_*$ with objects (A,δ), where A is a set and δ satisfies only the conditions (b) and (c) from definition of a

ISBN: 978-989-95079-6-8

similarity relation. Moreover, we will require that morphisms of $\mathbf{SetR}(\Omega)_*$ are those of morphisms $f : (A,\delta) \to (B,\gamma)$ from $\mathbf{SetR}(\Omega)$ that satisfy the following additional conditions:

(a) $\gamma(x,x) = \bigvee_{y\in B} f(x,y)$, for all $x \in A$,

(b) $f(x,y) \otimes f(x,z) \le \gamma(y,z)$.

In this section we show a deeper relationship between morphisms in $\mathbf{Set}(\Omega)$ and $\mathbf{SetR}(\Omega)_*$ but only for a very special Ω, namely for a totally ordered Heyting algebra Ω.

Proposition 2
Let Ω be a totally ordered complete Heyting algebra and let $(A,\alpha),(B,\beta)$ be objects from a category $\mathbf{Set}(\Omega)$. Let $f : A \times B \to \Omega$ be a map. Then the following statement are equivalent.

(a) $f : (A,\Delta_\alpha) \to (B,\Delta_\beta)$ is a morphism in the category $\mathbf{SetR}(\Omega)_*$.

(b) There exist a morphism $g : (A,\alpha) \to (B,\beta)$ in a category $\mathbf{Set}(\Omega)$ such that

(i) $(\forall a \in A, b \in B) \quad f(a,b) \le \Delta_\beta(g(a),b)$,

(ii) $(\forall a \in A, b \in B) \quad f(a,g(a)) = \alpha(a)$.

(c) There exists a morphism $g : (A,\alpha) \to (B,\beta)$ in a category $\mathbf{Set}(\Omega)$ such that

$$(\forall a \in A, b \in B) \quad f(a,b) = \alpha(a) \wedge \Delta_\beta(g(a),b).$$

Proof. Recall that a *singleton* in $(X,\delta) \in \mathbf{SetR}(\Omega)_*$ is a map $s : X \to \Omega$ such that

$$(\forall x,y \in X) \quad s(x) \wedge \delta(x,y) \le s(y),$$
$$(\forall x,y \in X) \quad s(x) \wedge s(y) \le \delta(x,y).$$

The set $\mathrm{singl}\,(X,\delta)$ of all singletons of (X,δ) with a function $\tau_{(X,\delta)}$ is then an object of $\mathbf{SetR}(\Omega)_*$ if we set

$$(\forall s,t \in \mathrm{singl}\,) \quad \tau_{(X,\delta)}(s,t) = \bigvee_{a\in X} s(a) \wedge t(a).$$

(a)⇒ (b). We construct a map $h : A \to \mathrm{singl}\,(B,\Delta_\beta)$ such that

$$(\forall a \in A)(\forall y \in B) \quad h(a)(y) = \bigvee_{x\in A} (f(x,y) \wedge \Delta_\alpha(a,x)).$$

We have to show that $h(a) \in \mathrm{singl}\,(B,\Delta_\beta)$ for any $a \in A$. In fact, for $y \ne y' \in B$ we have

$$h(a)(y) \wedge h(a)(y') \le f(a,y) \wedge f(a,y') \le \Delta_\beta(y,y') = 0,$$
$$h(a)(y) \le f(a,y) \le \Delta_\alpha(a,a) = \alpha(a).$$

Moreover, $h : (A,\Delta_\alpha) \to (\mathrm{singl}\,(B,\Delta_\beta),\tau)$ is a morphism in a category $\mathbf{SetR}(\Omega)_*$. In fact, for any $a,a' \in A$ we have

$$\tau(h(a),h(a')) = \bigvee_{y\in B}\bigvee_{x,x'\in A} (f(x,y) \wedge \Delta_\alpha(a,x) \wedge f(x',y) \wedge \Delta_\alpha(a',x')) \ge$$
$$\ge \bigvee_{y\in B} (f(a,y) \wedge \Delta_\alpha(a,a) \wedge \Delta_\alpha(a',a)) = \Delta_\alpha(a',a),$$
$$\tau(h(a),h(a)) = \bigvee_{y\in B}\bigvee_{x\in A} f(x,y) \wedge \Delta_\alpha(a,x) \le \bigvee_{y\in B} f(a,y) = \Delta_\alpha(a,a).$$

Since $h(a)(y) \wedge h(a)(y') = 0$ for all $y \ne y' \in B$ and Ω is totally ordered, for any $a \in A$ there exists at most one element $g(a) \in B$ such that $h(a)(g(a)) > 0$. If $h(a)(y) = 0$ for all $y \in B$, let $g(a) \in B$ be an arbitrary element. Then $g : (A,\alpha) \to (B,\beta)$ is a morphism in $\mathbf{Set}(\Omega)$. In fact, if $h(a)(g(a)) > 0$, then we have

$$\alpha(a) = \Delta_\alpha(a,a) = \tau(h(a),h(a)) = \bigvee_{y\in B} h(a)(y)$$
$$= h(a)(g(a)) = \bigvee_{x\in A} f(x,g(a)) \wedge \Delta_\alpha(a,x) \le$$
$$f(a,g(a)) \le \Delta_\beta(g(a),g(a)) = \beta(g(a)).$$

If $h(a)(y) = 0$ for all $y \in B$, then $\alpha(a) = \Delta_\alpha(a,a) = 0 \le \Delta_\beta(g(a),g(a))$. We show that the function g satisfies the conditions (i) and (ii). Let $h(a)(g(a)) > 0$. Then we have

$$0 < h(a)(g(a)) = \bigvee_{x\in A} f(x,g(a)) \wedge \Delta_\alpha(a,x) \le f(a,g(a)).$$

Moreover, since f is a morphism in $\mathbf{SetR}(\Omega)_*$, we have

$$(\forall a,b) \quad f(a,b) \wedge f(a,g(a)) \le \Delta_\beta(b,g(a)),$$

and it follows that $f(a,b) = 0$ for all $b \in B$ such that $b \ne g(a)$. Hence, the condition (i) holds. Since $\alpha(a) = \Delta_\alpha(a,a) = \bigvee_{y\in B} f(a,y) = f(a,g(a))$, the condition (ii) holds. Now, if $h(a)(y) = 0$ for all $y \in B$, it follows that $f(a,y) = 0$ for all $y \in B$ and the conditions (i),(ii) hold as well.

(b)⇒(c). This is a trivial computation only.

(c)⇒(a). Let $g : (A,\alpha) \to (B,\beta)$ be a morphism in $\mathbf{Set}(\Omega)$ and let $f : A \times B \to \Omega$ be a function from the statement (c). Then for any $a \in A$ we have

$$\bigvee_{b\in B} f(a,b) = \bigvee_{b\in B} (\alpha(a) \wedge \Delta_\beta(g(a),b)) =$$
$$= \alpha(a) \wedge \Delta_\beta(g(a),g(a)) = \alpha(a) \wedge \beta(g(a)) = \alpha(a).$$

and $f : (A,\Delta_\alpha) \to (B,\Delta_\beta)$ is a morphism in $\mathbf{SetR}(\Omega)_*$. ■

It is clear that Proposition 2 introduces a map

$$\varphi : \mathbf{Hom}_{\mathbf{Set}(\Omega)}((A,\alpha),(B,\beta)) \to$$
$$\mathbf{Hom}_{\mathbf{SetR}(\Omega)_*}((A,\Delta_\alpha),(B,\Delta_\beta))$$

such that for any morphism $f : (A,\alpha) \to (B,\beta)$ in $\mathbf{Set}(\Omega)$, $\varphi(f) : (A,\Delta_\alpha) \to (B,\Delta_\beta)$ is a morphism in $\mathbf{SetR}(\Omega)_*$ such

ISBN: 978-989-95079-6-8

that $\varphi(f)(a,b) = \alpha(a) \wedge \Delta_\beta(f(a),b)$ for all $a \in A, b \in B$. On the other hand, there exists a relation

$$\psi \subseteq \mathbf{Hom}_{\mathbf{SetR}(\Omega)_*}((A,\Delta_\alpha),(B,\Delta_\beta)) \times \mathbf{Hom}_{\mathbf{Set}(\Omega)}((A,\alpha),(B,\beta))$$

such that $(f,g) \in \psi$ if g is a morphism from (c) (Proposition 2) for a morphism f. It should be observed that (on the contrary to φ) ψ is not a map, in general. In fact, let (A,α) be an object in $\mathbf{Set}(\Omega)$ such that $\alpha(a) = 0$ for some $a \in A$. Let $f : (A,\Delta_\alpha) \to (B,\Delta_\beta)$ be a morphism in $\mathbf{SetR}(\Omega)_*$ and let $g : (A,\alpha) \to (B,\beta)$ be a morphism in $\mathbf{Set}(\Omega)$ satisfying the condition (c) from Proposition 2. Let us define a map $g' : A \to B$ such that $g'(x) = g(x)$ for all $x \in A, x \neq a$ and $g'(a) \neq g(a)$. Then $g' : (A,\alpha) \to (B,\beta)$ is a morphism, $(f,g),(f,g') \in \psi$ and $g' \neq g$. Hence, ψ is not a map.

Acknowledgment

Author express his thanks to reviewers for their valuable remarks and recommendations. The research was supported by the project MSM6198898701 of the MŠMT ČR, grant 201/07/0191 of GAČR and grant 1M0572.

References

[1] Höhle, U., Presheaves over GL-monoids. *Non-Classical Logic and Their Applications to Fuzzy Subsets*, Kluwer Academic Publ., Dordrecht, New York(1995), 127-157.

[2] Höhle, U., M-Valued sets and sheaves over integral, commutative cl-monoids. *Applications of Category Theory to Fuzzy Subsets*, Kluwer Academic Publ., Dordrecht, Boston (1992), 33-72.

[3] Höhle, U., Classification of Subsheaves over GL-algebras. *Proceedings of Logic Colloquium 98 Prague*, Springer Verlag (1999).

[4] Höhle, U., Commutative, residuated l-monoids. *Non-Classical Logics and Their Applications to Fuzzy Subsets*, Kluwer Academic Publ. Dordrecht, New York (1995), 53–106.

[5] Močkoř, J., Complete subobjects of fuzzy sets over MV-algebras. *Czech.Math.J.* **129(54)**(2004), 379–392.

[6] Močkoř, J., Extensional subobjects in categories of Ω-fuzzy sets. *Czech.Math.J.* **57(132)**(2007), 631–645.

[7] Močkoř, J., Characteristic morphisms and models of fuzzy logic in a category of sets with similarities. *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing* Springer-Verlag Berlin (2007),832–840.

[8] Močkoř, J., Covariant functors in categories of fuzzy sets over MV-algebras. *Advances in Fuzzy Sets and Systems***1(2)**(2006), 83–109.

[9] Močkoř, J., Models of fuzzy logic in a category of sets with similarity relations. *International Journal of Innnovative Computing, Information and Control***4(5)**(2008),1063–1068.

[10] Močkoř, J.,Fuzzy logic models in a category of fuzzy relations. *to appear in Soft Computing 2009.*

[11] Novák,V., Perfilijeva, I., Močkoř,J., *Mathematical principles of fuzzy logic* Kluwer Academic Publishers,Boston,Dordrecht, London (1999).

Bipolar Queries: An Approach and its Various Interpretations

Sławomir Zadrożny[1] Janusz Kacprzyk[2]

1.Systems Research Institute, Polish Academy of Sciences
Warszawa, Poland
2.Systems Research Institute, Polish Academy of Sciences
Warszawa, Poland
Email: Slawomir.Zadrozny@ibspan.waw.pl, Janusz.Kacprzyk@ibspan.waw.pl

Abstract *— The concept of a bipolar query is studied in the framework of the flexible fuzzy querying of databases. The focus is on the aggregation of the negative and positive conditions forming a bipolar query. Three formal logical representations of such an aggregation are proposed and analyzed taking into account various possible interpretations of fuzzy logical connectives. The relation to other approaches known in the literature is shown.*

Keywords— aggregation operators, bipolar queries, flexible queries, fuzzy logic, logical connectives modeling.

1 Introduction

A query to a (relational) database may be identified, in a slightly simplified view, with a condition expressing what the user is looking for. A database management system returns in the response to a query a list of tuples satisfying this condition. Such a condition is usually composed of a few, simple, atomic conditions putting some constraints on the values of the attributes characterizing given relation (table). We will adopt such a simplified view, assuming moreover that the user is concerned with just one relation, i.e., using SQL's terminology, joins, subqueries etc. are excluded for the sake of clarity of the presentation of the main idea.

Atomic conditions are connected using the classical logical connectives of the conjunction, disjunction and negation. In many standard application scenarios this is exactly what is needed to retrieve required data from a database. However in some other scenarios more sophisticated forms of the queries seem to be useful. Thus, in the literature further extensions of this basic setting are proposed. Firstly, many authors advocate the convenience of using *linguistic terms* in queries (cf., e.g., [1, 2]). For example, it is more natural and comfortable for a user of a real-estate agency to state that she or he is looking for a "not very expensive house" rather than to use a precise interval of acceptable prices. The benefits of using fuzzy logic to model linguistic terms in queries are widely advocated and there is a large number of of both theoretical and practical results obtained by the researchers in this area (cf., e.g., [3, 1, 2]). Secondly, some (atomic) conditions may be for the user more important than the others. For example, a customer of a real-estate agency may be looking for a "not very expensive house located in a nice city district" and to treat the former condition as a much more important than the latter. Thus in the overall matching degree of a tuple against such a query the satisfaction of the former condition is crucial, while of the latter is to some extent secondary. Proper modeling of importance weights is a subject of many papers (cf., e.g., [4]).

The essence of the concept of a *bipolar query* is in a new way of differentiating the conditions than by assigning them with some *fixed* importance weights. Namely, a bipolar query is defined by two conditions: one – *negative (required)* – expresses the constraints that *have to be* satisfied while the second – *positive (preferred)* – expresses only what is *desired* and its violation does not necessarily lead to the rejection of given tuple. This combination of conditions may be meant and studied in a few different ways. Here we are interested in the ways the satisfaction of both types of conditions should be combined to obtain an overall matching degree of the whole query. From this point of view the most important is the question if there is a conflict between the conditions or they can be satisfied simultaneously. If there is a total conflict, i.e., satisfying one condition means totally failing to satisfy another, the bipolar query reduces itself to the required condition. On the other hand, if both conditions may be totally satisfied simultaneously then the bipolar query reduces to a simple conjunction. Thus, the most interesting are intermediate cases which may be characterized by a degree of conflict between the conditions.

In this paper we study three logical formulas which may be used as a representation of a bipolar query. They are equivalent in case of the classical predicate logic but as soon as we adopt the fuzzy (multivalued) logic context they become distinct and exhibit different properties depending on the assumed set of fuzzy operators used to model logical connectives. Our study may be seen as an attempt at justifying particular choices in this respect.

The structure of this paper is as follows. In Section 2 we briefly remind the concept of the bipolar query and introduce the notation used later on. We also discuss the concept of the *winnow* operator which is used to derive logical representations of bipolar queries. In Section 3 we compare particular representations under different choices of logical operators, i.e., operators used to model particular logical connectives.

2 The concept of a bipolar query

The very concept of bipolar queries has been introduced by Dubois and Prade [5]. Its basic idea is to distinguish two types of query conditions, which are related to the *negative* and *positive* preferences of a user. The former coincide with the traditional understanding of a condition as a constraint, which defines a set of *feasible* tuples or, equivalently, excludes all tuples that do not satisfy it. The latter, on the other hand, characterizes those tuples which are really desired, with such an un-

ISBN: 978-989-95079-6-8

derstanding that violating such a condition by a tuple does not necessarily exclude it from the consideration. Bipolar queries may be exemplified by the following one:

"Find a *non-expensive* house *and possibly* located *near* a railway station (1)

where the condition referring to the price is a negative one, excluding expensive houses, and the condition referring to the location is a positive one, expressing just a desire to get a house conveniently located, *if possible*.

Now there is a crucial question of the interplay between these two types of conditions which distinguishes different lines of research in this area. Namely, Dubois and Prade basically assume that these conditions should be consistent in such a sense that the set of desired tuples should be a subset of the set of feasible tuples. Then the main question is how to take into account the sets of negative and positive conditions. These are aggregated separately and if the resulting overall negative and positive conditions are not consistent some measures are undertaken to make them so. The answer to such a bipolar query is generated according to the strategy "first select (with respect to the negative condition) and then order (with respect to the positive condition)". This strategy requires a precisiation in case the first condition is fuzzy, i.e., is satisfied by tuples to a degree – what is, of course, the most interesting case, anyway. Then it is not that clear what does it mean to select tuples satisfying a fuzzy condition as they form, in fact, a fuzzy set. Dubois and Prade [6] propose to employ here a lexicographic order of the tuples represented by vectors of the degrees of matching of particular conditions. They propose also a comprehensive representation of bipolar preferences in the framework of the possibility theory [7].

Another line of research explicitly takes into account the conflict between the constraints and the desires and looks for the aggregation of both conditions directly referring to the degree of this conflict. This interpretation is emphasized by the use of the "and possibly" operator in (1). Thus, in general describing a bipolar query we will use the following notation:

$$C \textit{ and possibly } P \tag{2}$$

or, equivalently, an answer to a bipolar query may be defined as the set of tuples: $\{t : C(t) \textit{ and possibly } P(t)\}$. The above form puts emphasis on the question of a proper modeling of the aggregation of both types of conditions, which is expressed here with the use of the "and possibly" operator.

The interest of the database community in this type of queries dates back to the paper by Lacroix and Lavency [8]. They were the first to propose the use of a query comprising two categories of conditions: one which is mandatory (C) and another which expresses just mere preferences (desires) (P). The bipolarity of these conditions becomes evident as soon as one adopts the following interpretation. The former condition C may be seen as expressing the *negative* preferences: the tuples which do not satisfy it are definitely not matching the whole query. The latter condition P, on the other hand, has a *positive* character: a tuple satisfying it is preferred over another tuple not satisfying it, provided both tuples satisfy the mandatory condition C.

For the purposes of a further discussion we will use the following notation. We assume the queries are addressed against a set of tuples $T = \{t_j\}$ comprising a relation. We will identify the negative and positive conditions of a bipolar query with the predicates that represent them and denote them as C and P, respectively. For a tuple $t \in T$, $C(t)$ and $P(t)$ will denote that the tuple t satisfies respective condition (in crisp case) or the degrees of satisfaction, if the conditions are fuzzy. We will also denote the whole bipolar query as (C, P).

According to the original (crisp) approach by Lacroix and Lavency if there are no tuples meeting both conditions then the result of the aggregation is determined by the negative condition C alone. Otherwise the aggregation becomes a regular conjunction of both conditions. Thus the answer to such a query depends not only on the explicit arguments, i.e., $C(t)$ and $P(t)$, but also on the content of the database. This dependence is best expressed by the following logical formula [8]:

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge \exists s(C(s) \wedge P(s)) \Rightarrow P(t) \tag{3}$$

The characteristic feature of such an interpretation of bipolar queries is that if there is no conflict between the conditions P and C, i.e., there are tuples satisfying both of them, then the query turns into a conjunction of the conditions. On the other hand if there are no tuples satisfying both conditions then only condition C is used to select tuples.

Such an aggregation operator has been later proposed independently by Dubois and Prade [9] in the context of default reasoning and by Yager [10, 11] in the context of the multicriteria decision making for the case of so-called *possibilistically qualified criteria*. Yager [11] intuitively characterizes a possibilistically quantified criterion as such which should be satisfied unless it interferes with satisfaction of other criteria. This is in fact the essence of bipolar queries in the sense advocated in this paper. This concept was also applied by Bordogna and Pasi [12] for the textual information retrieval task.

Lacroix and Lavency [8] consider only the case of crisp conditions C and P. Then a bipolar query may be, in fact, processed using the "first select using C then order using P" strategy, i.e., the answer to the bipolar query (C, P) is obtained by, first, finding tuples satisfying C and, second, choosing from among them those satisfying the condition P, if any. If C is crisp and P is fuzzy then the second step consists in non-increasingly *ordering* the tuples satisfying C according to their degree of satisfaction of P. This understanding is predominant in the literature dealing with fuzzy extensions of the original concept of Lacroix and Lavency. Both, direct extensions proposed by Bosc and Pivert [13, 14] as well as sophisticated possibility theory-based interpretation of this concept by Dubois and Prade [6] focus, in fact, on the proper treatment of *multiple* required and preferred conditions, basically assuming the above strategy as the way of combining the negative and positive conditions.

In [15, 16] we propose a "fuzzification" of the formula (3) and study its basic properties. Here we focus on the question of combining fuzzy conditions C and P and will treat them in what follows as atomic. From this point of view it is worth mentioning some other approaches which are of relevance here. Dujmović [17] introduced the concept of the *partial absorption function* which may be used to combine the values of two variables in such a way that one variable controls the influence of the other on the result of their combina-

ISBN: 978-989-95079-6-8

tion. It makes it possible to express the requirement that for a high value of the result a high value of the first variable is mandatory while the high value of the second is desired but not mandatory. When applied to the aggregation of the values of $C(t)$ and $P(t)$ this is somehow similar to the idea of bipolar query expressed by (3), but lacks its dependence on the content of the whole database. This approach may be seen as based on a sophisticated, *dynamic* weighting of the importance of the combined values, where the weights itself depend on the combined values. A similar approach has been proposed by Dubois and Prade; cf., e.g., [4].

The operator "among", very close to the discussed here operator "and possibly", has been proposed by Tudorie [18]. She considers queries of the type "find tuples satisfying a condition P among those satisfying a condition C", which are in fact equivalent to the bipolar queries understood as in (2). The evaluation of a query with the "among" operator is very similar to the one adopted by us for the bipolar query, but is expressed in terms of the rescaling of the linguistic terms used in the condition P. Namely, first the set of tuples satisfying the condition C to a non-zero degree is selected. Then the membership functions of the fuzzy sets representing the linguistic terms appearing in P (such as "near" in (1)) are rescaled taking into account the actual range of the corresponding attributes in the set of tuples selected in the first step. For example, if originally the distance of 2 kilometers from the station has the membership value to the fuzzy set representing the term "near" equals 0.5, and it turns out that it is the shortest distance among the houses selected in the first step (i.e., among non-expensive houses, in case of the query (1)), then this membership degree may be changed to 1 (the actual algorithm of rescaling may, of course, take different forms). Finally, the overall matching degree is computed as a conjunction of the matching degrees against the condition C and the "modified" condition P, i.e., the one for which the rescaled membership function of the linguistic terms is used. Please note that if there is no interference between both conditions (in the sense discussed earlier) then there is no need for rescaling the membership functions and the query turns into a conjunction of both conditions, like in the case of the bipolar query.

Bipolar queries may be also seen a special case of *queries with preferences* proposed recently, for the crisp case, by Chomicki [19]. In the framework of this approach a new relational algebra operator called *winnow* is introduced. This unary operator selects from a set of tuples T those which are *non-dominated* with respect to a given *preference relation* R, $R \subseteq T \times T$. If two tuples $t, s \in T$ are in relation R, i.e., $R(t, s)$, then it is said that the tuple t *dominates* the tuple s with respect to the relation R. Then the *winnow* operator ω_R is defined as follows

$$\omega_R(T) = \{t \in T : \neg\exists_{s \in T}\ R(s, t)\} \tag{4}$$

Thus, for a given set of tuples it yields a subset of the *non-dominated* tuples with respect to R.

The concept of the *winnow* operator may be illustrated with the following example. Let us consider a database of a real-estate agency with a table `HOUSES` describing the details of particular real-estate properties offered by an agency. The schema of the relation `HOUSES` contains, among other, the attributes `city` and `price`. Let us assume that we are interested in the list of the *cheapest* houses in each city. Then the preference relation should be defined as follows

$$R(t, s) \Leftrightarrow (t.\texttt{city} = s.\texttt{city}) \wedge (t.\texttt{price} < s.\texttt{price})$$

where $t.A$ denotes the value of attribute A (e.g., `price`) at a tuple t. Then the *winnow* operator $\omega_R(\texttt{HOUSES})$ will select the houses that are sought. Indeed, according to the definition of the *winnow* operator, we will get as an answer a set of houses, which are non-dominated with respect to R, i.e., for which there is no other house in the same city which has a lower price.

In [16] we proposed a fuzzy counterpart of the *winnow* operator taking into account the fuzziness of the preference relation R and of the related concept of non-dominance as well as the fact that the set of tuples T is also a fuzzy set. It may be expressed as follows:

$$\mu_{\omega_R(T)}(t) = \text{truth}(T(t) \wedge \forall_s\ (T(s) \rightarrow \neg R(s, t))) \tag{5}$$

where $\mu_{\omega_R(T)}(t)$ denotes the value of the membership degree of the tuple t to the fuzzy set of tuples defined by $\omega_R(T)$.

A bipolar query (C, P) may be expressed using the concept of the fuzzy *winnow* operator as follows [19, 16]. Let R be a fuzzy preference relation of the following form (symbols R and P denote both the fuzzy predicates and the membership functions of corresponding fuzzy sets, depending on the context):

$$R(t, s) \Leftrightarrow P(t) \wedge \neg P(s) \tag{6}$$

Then the bipolar query may be expressed as the combination of the selection and fuzzy *winnow* operators $\omega_R(\sigma_C(T))$, i.e.,

$$\mu_{\omega_R(\sigma_C(T))} = \text{truth}(C(t) \wedge \forall_s\ (C(s) \rightarrow (\neg P(s) \vee P(t)))) \tag{7}$$

(where $\sigma_C(T)$ is a usual "fuzzy" extension of the standard relational algebra selection operator, i.e., $\mu_{\sigma_C(T)}(t) = C(t)$).

The definition of the fuzzy *winnow* operator (5) as well as the "fuzzification" of the formula defining a bipolar query (3) leave open the question of a choice of the logical operators which should be used to model particular logical connectives occurring in both formulas. In [16] we show that for a specific choice of them the fuzzy set of tuples obtained using (7) is identical with the fuzzy set defined by (3). In [20] we analyze the properties of the "fuzzified" version of (3) for the broader class of the logical operators. In the next section we further advance this study.

3 Various interpretations of bipolar queries and their properties

In our previous work [15, 16, 20] we studied a specific fuzzy version of the Lacroix and Lavency formula (3) representing an interpretation of the concept of bipolar queries. We have also shown its basic relation with a fuzzy version of the *winnow* operator we proposed. Here we extend this study comparing three formulas that may be used to represent the bipolar query and their properties under different interpretations of the logical connectives occurring in them.

We derive the logical formulas expressing the matching degree of a bipolar query in three different ways (we repeat here some formulas introduced in the previous section for the convenience of the reader):

ISBN: 978-989-95079-6-8

- making a direct "fuzzification" of the formula (3) proposed for the crisp case by Lacroix and Lavency [8]:

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge (\exists s\ (C(s) \wedge P(s)) \Rightarrow P(t)) \quad (8)$$

- making a direct "fuzzification" of the crisp *winnow* operator (4) and applying it with a specific preference relation (6) to obtain a bipolar query representation:

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge \neg\exists s\ ((C(s) \wedge P(s) \wedge \neg P(t))) \quad (9)$$

- using our fuzzy version of the *winnow* operator (5) and applying it as above:

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge \forall s\ (C(s) \Rightarrow (\neg P(s) \vee P(t))) \quad (10)$$

It may be easily seen that in the framework of the classical logic all three above formulas are equivalent. Here we study their properties in case of fuzzy (multivalued) interpretation, in particular taking into account various operators which may be used to model logical connectives. We follow usual approach of modeling conjunction and disjunction by the t-norm and t-conorm operators, respectively [21].

In order to carry out the analysis we consider so-called De Morgan Triples $(\wedge, \vee, \neg)$ that comprise a t-norm operator $\wedge$, a t-conorm operator $\vee$ and a negation operator $\neg$, where $\neg(x \vee y) = \neg x \wedge \neg y$ holds. Three following De Morgan Triples play the most important role in fuzzy logic (cf., e.g., [21] for a justification) $(\wedge_{min}, \vee_{max}, \neg)$, $(\wedge_{\Pi}, \vee_{\Pi}, \neg)$, $(\wedge_{W}, \vee_{W}, \neg)$, where particular t-norms and t-conorms are defined as follows:

$t-$ *norms*		
$x\ \wedge_{min}\ y$	$= \min(x, y)$	*minimum*
$x\ \wedge_{\Pi}\ y$	$= x \cdot y$	*product*
$x\ \wedge_{W}\ y$	$= \max(0, x + y - 1)$	*Łukasiewicz*

$t-$ *conorms*		
$x\ \vee_{max}\ y$	$= \max(x, y)$	*maximum*
$x\ \vee_{\Pi}\ y$	$= x + y - x \cdot y$	*probabilistic sum*
$x\ \vee_{W}\ y$	$= \min(1, x + y)$	*Łukasiewicz*

We will refer to these De Morgan Triples in what follows as, respectively, MinMax, Π and W triples. The negation operator $\neg$ in case of all the above De Morgan Triples is defined as: $\neg x = 1 - x$. Both t-norms and t-conorms are by definition associative and thus may be treated as m-ary operators, i.e., expressions like $x \wedge y \wedge \ldots$ and $x \vee y \vee \ldots$ are well defined.

Basically, the general and existential quantifiers are identified in fuzzy logic, for the case of a finite universe, with the maximum and minimum operators, respectively. They may be generalized via the use of other t-norms and t-conorms what leads to the concept of *t-quantifiers* and *s-quantifiers*; cf., e.g., [22]. The truth of a statement involving such a quantifier is computed as follows ($\{a_1, \ldots, a_m\}$ is a finite universe under consideration):

$$\text{truth}(\forall x A(x)) = \mu_A(a_1) \wedge \mu_A(a_2) \wedge \ldots \wedge \mu_A(a_m) \quad (11)$$

$$\text{truth}(\exists x A(x)) = \mu_A(a_1) \vee \mu_A(a_2) \vee \ldots \vee \mu_A(a_m) \quad (12)$$

We use generalized quantifiers while interpreting formulas (8)-(10) and particular t- and s-quantifiers will be denoted in what follows by the $\forall$ and $\exists$ symbol with a subscript indicating underlying t-norm or s-norm, e.g., $\exists_{max}$ denotes a fuzzy existential quantifier which obtains when the t-conorm "maximum" is used.

We consider two implication operators related to a given De Morgan Triple $(\wedge, \vee, \neg)$, so-called S-implications:

$$x \rightarrow_{S-\vee} y = \neg x \vee y \quad (13)$$

and R-implications:

$$x \rightarrow_{R-\wedge} y = \sup\{z : x \wedge z \leq y\} \quad (14)$$

Thus, for particular De Morgan Triples one obtains the following R-implication operators:

$$x\ \rightarrow_{R-min}\ y = \begin{cases} 1 & \text{for } x \leq y \\ y & \text{for } x > y \end{cases}$$

$$x\ \rightarrow_{R-\Pi}\ y = \begin{cases} 1 & \text{for } x = 0 \\ \min\{1, \frac{y}{x}\} & \text{for } x \neq 0 \end{cases}$$

$$x\ \rightarrow_{R-W}\ y = \min(1 - x + y, 1)$$

and the following S-implication operators:

$$x\ \rightarrow_{S-max}\ y = \max(1 - x, y)$$

$$x\ \rightarrow_{S-\Pi}\ y = 1 - x + x \cdot y$$

$$x\ \rightarrow_{S-W}\ y = \min(1 - x + y, 1)$$

Now let us consider the question of the choice of one of the formulas (8)-(10) to represent bipolar queries and an appropriate modeling of the logical connectives occurring therein, i.e., the choice of one of the De Morgan Triples.

In [20] we have shown certain basic properties of the fuzzified version of the original formula (8). Some of them are valid for any choice of logical operators, some are limited to some special cases. We study some of them here again, checking if they are valid also for the formulas (9) and (10). However, first we start with a property identifying the equivalence between formulas (8)-(10) for a certain choice of the logical operators.

Property 1 For a distributive De Morgan triple, i.e., when $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$, and a related S-implication all formulas (8)-(10) are equivalent.

We will show the equivalence between (8) and (9) and the remaining equivalences may be shown in a similar way (the equivalence of (8) and (10) was shown by us in [16] for the specific case of the MinMax De Morgan Triple). Formula (9) may be rewritten as follows:

$$C(t) \wedge \forall s\ (\neg C(s) \vee \neg P(s) \vee P(t))$$

what, due to the assumed distributivity of the triple is equivalent to:

$$C(t) \wedge ((\forall s\ (\neg C(s) \vee \neg P(s))) \vee P(t))$$

and via the following series of transformations leads to (8):

$$C(t) \wedge ((\forall s\ \neg(C(s) \wedge P(s))) \vee P(t)) \equiv$$
$$C(t) \wedge ((\neg\exists s\ (C(s) \wedge P(s))) \vee P(t)) \equiv$$
$$C(t) \wedge (\exists s\ (C(s) \wedge P(s)) \Rightarrow P(t))$$

ISBN: 978-989-95079-6-8

The last equivalence holds for "⇒" being an S-implication.

Among three De Morgan Triples that we consider in this paper only the MinMax triple is distributive. Thus for it and related S-implication we get equivalence of all three formulas (8)-(9). Property 1 shows only a sufficient condition for this equivalence, but for the other two triples some counterexamples for such an equivalence may be easily shown.

In [20] we have shown a property which is worth reminding as it best characterizes the understanding of the bipolar queries adopted here. Namely, if there is no conflict between the required (C) and preferred (P) conditions at all, i.e., there is a tuple fully (to a degree equal 1) satisfying both of them then the formula (8) turns into a regular conjunction of both conditions. This may be formally expressed as follows.

Property 2 If there exists a tuple t such that $C(t) = P(t) = 1$ then a bipolar query (C, P) turns into the conjunction $C \wedge P$.

In [20] we have shown that for any combination of a t-norm, t-conorm and S-implication or R-implication such that $(1 \Rightarrow x) = x$, Property 2 holds in case of the formula (8). Note that the conjunction mentioned in the Property is modeled then via assumed t-norm.

Here, in view of the Property 1, it is clear that in case of the MinMax De Morgan Triple and the S-implication, Property 2 holds also for the formulas (9) and (10). Moreover, it may be easily proved that this property also holds for the MinMax De Morgan Triple and the R-implication in case of the formula (10). Namely, we will show that:

$$\text{truth}(\forall s\ (C(s) \Rightarrow_{R-min} (\neg P(s) \vee_{max} P(t))) = \min_s \begin{cases} 1 & \text{if } C(s) \leq \neg P(s) \vee_{max} P(t) \\ \neg P(s) \vee_{max} P(t) & \text{otherwise} \end{cases} \tag{15}$$

is equal $P(t)$ for any tuple t, under the assumptions of the Property 2. If $P(t) = 1$ then for all tuples s the value of $C(s)$ is lower or equal $\max(1 - P(s), P(t))$ and thus (15) is equal 1, i.e., is equal to $P(t)$. Now let us assume that $P(t) < 1$ and let us denote with u a tuple for which $C(u) = P(u) = 1$ (the existence of such a tuple is assumed in the Property 2). Then, for any tuple t, the minimum over s in (15) is realized for $s = u$ and thus is equal $\max(1 - 1, P(t))$, i.e., $P(t)$, what completes our proof of the Property 2 also for the formula (10) and the R-implication operator.

Some examples may be easily found showing that Property 2 fails for the W and Π De Morgan triples.

Another property of the formula (8), shown in [20], is valid for two other formulas (9) and (10), and may be expressed as follows.

Property 3 If for a tuple t the value of $P(t)$ is equal to 1, then a bipolar query (C, P) turns into $C(x)$.

This property holds for all formulas (8)-(10), for any combination of a t-norm, t-conorm and S-implication or R-implication.

This property is a direct consequence of the general properties of the t-norm ($(x \wedge 0) = 0$), t-conorm ($(x \vee 1) = 1$ and $(0 \vee 0) = 0$) and implication operators ($(x \Rightarrow 1) = 1$).

Now we show some additional properties of the formulas (8)-(10).

Property 4 Let us assume that truth($\exists s\ (C(s) \wedge P(s))) > 0$. Then, for a De Morgan Triple with a t-norm without zero divisors, i.e., where $\forall x, y > 0\ (x \wedge y) \neq 0$, and the related R-implication, the matching degree computed using (8) for a tuple t fully satisfying the required condition and not satisfying the preferred condition at all (i.e., $C(t) = 1$ and $P(t) = 0$) is equal 0.

This is a property of the R-implication. Notice that this means that such a tuple t will get lower matching degree than a tuple s which satisfies both conditions to a degree ϵ, whatever small ϵ is. Thus it is surely a property we would like to avoid and which makes (8) under both MinMax and Π De Morgan Triples (whose t-norms do not have zero divisors) with related R-implications less appealing as models of the bipolar query.

A similar property, formulated below, is exhibited by (10).

Property 5 Let us assume that there is at least one tuple u such that $C(u) > 0$ and $P(u) = 1$. Then, for a De Morgan Triple with a t-norm without zero divisors and the related R-implication the matching degree computed using (10) for a tuple t fully satisfying the required condition and not satisfying the preferred condition at all (i.e., $C(t) = 1$ and $P(t) = 0$) is equal 0.

Thus also (10) under both MinMax and Π De Morgan Triples with related R-implications is not very appealing as a model of the bipolar query. This property seems to favor the W De Morgan Triple, at least in case of the R-implication and formulas (8) and (10).

Another negative property of (8) for a specific combination of logical operators may be expressed as follows.

Property 6 For the MinMax De Morgan Triple used with related S-implication, the aggregation scheme defined by (8) may lead to the same matching degree for two tuples t and u while t strongly Pareto dominates u, i.e., $C(t) > C(u)$ and $P(t) > P(u)$.

This may be demonstrated with the following example. Let us denote $\exists_s\ (C(s) \wedge P(s))$ with $\exists CP$. Let $\exists CP = 0.7$ and $C(t) = 1$, $P(t) = 0.3$, $C(u) = 0.3$ and $P(u) = 0$. Then the matching degree computed for both tuples is equal 0.3, while t strongly Pareto dominates u.

In fact this property may be supplemented by observing that all tuples t such that $P(t) \leq (1 - \exists CP)$ and $C(t) \geq (1 - \exists CP)$ obtain the same matching degree, equal $1 - \exists CP$. This fact has been observed by Dubois and Prade [5] for a formula similar to (8). However, it should be noted that even for the tuples not satisfying the above conditions, the Pareto domination may be not reflected by (8) used with the logical operators specified by Property 6. For example, for a tuple u, such that $C(u) = P(u) = 0.6$, still assuming $\exists CP = 0.7$, the matching degree is equal 0.6. The same matching degree obtains for t such that $C(t) = 1$ and $P(t) = 0.6$ as well as for t such that $C(t) = 0.6$ and $P(t) = 1.0$, while in both these cases the tuple t Pareto dominates u.

Note, that due to the Property 1, the Property 6 is also valid for two other formulas (9)-(10).

Certainly the list of properties discussed in the paper is not exhaustive and they should be seen as a first attempt at a more comprehensive analysis of the bipolar queries and their various representations proposed here. Such an analysis should provide some hints which representation should be used un-

ISBN: 978-989-95079-6-8

der which conditions and using which set of logical operators. However, already the properties discussed here provide some hints which may be summarized as follows.

We first list some general properties shared by all three formulas (8)-(10) under any combination of the logical operators. These properties may be expressed concisely as the properties of the "and possibly" operator, which is the essence of our understanding of bipolar queries (this operator is denoted below as $\wedge_{possibly}$, but it should be remembered that this operator is not truth-functional):

- monotonicity (but not strict) in both arguments,
- boundary properties: $1 \wedge_{possibly} 1 = 1$ and $x \wedge_{possibly} 1 = x$ (Property 3).

Now let us look at particular formulas (8)-(10) and summarize their properties in combinations with all considered logical operators.

Formula (8) exhibits Property 2 when used with any combination of logical operators. This is surely its advantage as this property is characteristic for our understanding of bipolar queries. Property 4 seems to suggest that the R-implication should be avoided in case of formula (8) (unless it is used in the framework of the W De Morgan Triple, but then both types of implication operators are identical, thus a general hint of avoiding R-implication may be still seen as valid). Property 6 suggests that the MinMax De Morgan Triple is generally not appropriate for the formula (8).

Formula (9) exhibits Property 2 only for the MinMax De Morgan Triple, but Property 6 makes this triple inappropriate to some extent.

Formula (10) also satisfies Property 2 only for the MinMax De Morgan Triple, which is on the other hand somehow inappropriate due to the Properties 5 and 6.

Concluding, if Property 2 is required, what seems to be a reasonable postulate, then the best choice of the representation of the bipolar queries and logical operators to model the logical connectives therein– according to studied properties– seems to be formula (8) with the Π De Morgan Triple and the S-implication operator. Such a choice saves the obtained representation from some negative properties discussed in this paper. Further studies are needed in order to identify a more comprehensive list of postulated properties.

References

[1] P. Bosc and O. Pivert. SQLf: A relational database language for fuzzy querying. *IEEE Transactions on Fuzzy Systems*, 3(1):1–17, 1995.

[2] J. Kacprzyk and S. Zadrożny. Computing with words in intelligent database querying: standalone and internet-based applications. *Information Sciences*, 134(1-4):71–109, 2001.

[3] S. Zadrożny, G. De Tre, R. De Caluwe, and J. Kacprzyk. An overview of fuzzy approaches to flexible database querying. In Galindo [23], pages 34–53.

[4] D. Dubois and H. Prade. Using fuzzy sets in flexible querying: why and how? In T. Andreasen, H. Christiansen, and H.L. Larsen, editors, *Flexible Query Answering Systems*, pages 45–60. Kluwer Academic Publishers, 1997.

[5] D. Dubois and H. Prade. Bipolarity in flexible querying. In T. Andreasen, A. Motro, H. Christiansen, and H.L. Larsen, editors, *Flexible Query Answering Systems, FQAS 2002*, volume 2522 of *Lecture Notes in Computer Science*, pages 174–182. Springer, Berlin, Heidelberg, 2002.

[6] D. Dubois and H. Prade. Handling bipolar queries in fuzzy information processing. In Galindo [23], pages 97–114.

[7] D. Dubois and H. Prade. An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. *Fuzzy Sets and Systems*, 160(10):1355–1366, 2009.

[8] M. Lacroix and P. Lavency. Preferences: Putting more knowledge into queries. In *Proceedings of the 13 International Conference on Very Large Databases*, pages 217–225, Brighton, UK, 1987.

[9] D. Dubois and H. Prade. Default reasoning and possibility theory. *Artificial Intelligence*, 35(2):243–257, 1988.

[10] R.R. Yager. Higher structures in multi-criteria decision making. *International Journal of Man-Machine Studies*, 36:553–570, 1992.

[11] R.R. Yager. Fuzzy logic in the formulation of decision functions from linguistic specifications. *Kybernetes*, 25(4):119–130, 1996.

[12] G. Bordogna and G. Pasi. Linguistic aggregation operators of selection criteria in fuzzy information retrieval. *International Journal of Intelligent Systems*, 10(2):233–248, 1995.

[13] P. Bosc and O. Pivert. Discriminated answers and databases: fuzzy sets as a unifying expression means. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 745–752, San Diego, USA, 1992.

[14] P. Bosc and O. Pivert. An approach for a hierarchical aggregation of fuzzy predicates. In *Proceedings of the Second IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'93)*, pages 1231–1236, San Francisco, USA, 1993.

[15] S. Zadrożny. Bipolar queries revisited. In V. Torra, Y. Narukawa, and S. Miyamoto, editors, *Modelling Decisions for Artificial Intelligence (MDAI 2005)*, volume 3558 of *LNAI*, pages 387–398. Springer-Verlag, Berlin, Heidelberg, 2005.

[16] S. Zadrożny and J. Kacprzyk. Bipolar queries and queries with preferences. In *Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA'06)*, pages 415–419, Krakow, Poland, 2006. IEEE Comp. Soc.

[17] J.J. Dujmović. Partial absorption function. *Journal of teh University of Belgrade, EE Dept.*, 659:156–163, 1979.

[18] C. Tudorie. Qualifying objects in classical relational database querying. In Galindo [23], pages 218–245.

[19] J. Chomicki. Querying with intrinsic preferences. *Lecture Notes in Computer Science*, 2287:34–51, 2002.

[20] S. Zadrożny and J. Kacprzyk. Bipolar queries using various interpretations of logical connectives. In *Foundations of Fuzzy Logic and Soft Computing*, Lecture Notes in Computer Science, pages 181–190. Springer, 2007.

[21] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Series D: System Theory, Knowledge Engineering and Problem Solving. Kluwer Academic Publishers, 1994.

[22] R. Mesiar and H. Thiele. On T-Quantifiers and S-Quantifiers. In V. Novak and I. Perfilieva, editors, *Discovering the World with Fuzzy Logic*, pages 310–326. Physica-Verlag, Heidelberg New York, 2000.

[23] J. Galindo, editor. *Handbook of Research on Fuzzy Information Processing in Databases*. Information Science Reference, New York, USA, 2008.

ISBN: 978-989-95079-6-8

A Fuzzy Variant of the Rand Index for Comparing Clustering Structures

Eyke Hüllermeier[1] and Maria Rifqi[2]

1. FB Mathematik/Informatik, Philipps-Universität Marburg
D-35032 Marburg, Germany
2. University Pierre et Marie Curie–Paris 6, CNRS UMR 7606, LIP6,
Paris, F-75016, France
Email: eyke@mathematik.uni-marburg.de, maria.rifqi@lip6.fr

***Abstract*— *In this paper, we introduce a fuzzy extension of the Rand index, a well-known measure for comparing two clustering structures. In contrast to an existing proposal, which is restricted to the comparison of a fuzzy partition with a non-fuzzy reference partition, our extension is able to compare two proper fuzzy partitions with each other. Elaborating on the formal properties of our fuzzy Rand index, we show that it exhibits desirable metrical properties.*

***Keywords*—** Clustering, Distance, Fuzzy Partition, Metric, Rand Index, Similarity

1 Introduction

The problem to compare two partitions of a set of objects occurs quite naturally in various domains, notably in data analysis and clustering. For example, one way to evaluate the result of a clustering algorithm is to compare the clustering structure produced by the algorithm with a correct partition of the data (which of course presumes that this information is available). In cluster analysis, so called *external evaluation measures* have been developed for this purpose [1, 2]. However, measures of that kind are not only of interest as evaluation criteria, i.e., for comparing a hypothetical partition with a true one. Instead, distance measures for partitions are interesting in their own right and can be used for different purposes.

In [3], for example, the authors consider the problem of clustering data in a very high-dimensional space. To increase efficiency, they propose to map the data into a low-dimensional space first and to cluster the transformed data thus obtained afterward. In this context, a distance measure for clustering structures (partitions) is useful to measure the loss of information incurred by the data transformation: If the transformation is (almost) lossless, the clustering structures in the two spaces should be highly similar, i.e., their distance should be small. On the other hand, a significant difference between the two partitions would indicate that the transformation does have a strong effect in the sense of distorting the structure of the data set.

Even though a large number of evaluation criteria and similarity indexes for clustering structures have been proposed in the literature, their extension to the case of fuzzy partitions has received much less attention so far. This is especially true for external evaluation criteria and measures comparing two clustering structures, whereas *internal criteria* for evaluating a single partition[1] have been studied more thoroughly (see, e.g., [4] and [5] for early proposals).

In a recent paper by Campello [6], the author has proposed an extension of the *Rand index* [7], a well-known measure of similarity between two partitions of a data set. Even though Campello's proposal is quite interesting, it also exhibits a number of disadvantages. Most notably, it is properly defined only for the comparison of a fuzzy partition with a non-fuzzy reference partition. It is true that this restriction can be tolerated if the index is used as an external evaluation criterion since, as correctly argued by the author, a reference partition provided by an external source is typically non-fuzzy. Yet, our example above has clearly shown that there is also a need for measures comparing two fuzzy partitions.

In this paper, we propose an alternative extension of the Rand index (which is, in principle, also applicable to related similarity measures for clustering structures). As opposed to Campello's proposal, our variant is able to compare two proper fuzzy partitions with each other. Moreover, we study our fuzzy Rand index from a formal point of view and show that it satisfies the desirable properties of a metric (when being used as a distance function).

The remainder of the paper is organized as follows. In the next section, we briefly review the proposal of Campello and discuss some of its properties in a critical way. In Section 3, we introduce our new measure and elaborate on its formal properties. The paper concludes with a short summary and an outlook on future work in Section 4.

2 Review of Campello's Proposal

Before reviewing Campello's proposal and discussing some of its properties, we briefly recall the original definition of the Rand index.

2.1 The Rand Index

Let $\mathbf{P} = \{P_1, \ldots, P_k\} \subset 2^X$ and $\mathbf{Q} = \{Q_1, \ldots, Q_\ell\} \subset 2^X$ be two (crisp) partitions of a finite set $X = \{x_1, x_2 \ldots x_n\}$ with n elements, which means that $P_i \neq \emptyset$, $P_i \cap P_j = \emptyset$ for all $1 \leq i \neq j \leq k$, and $P_1 \cup P_2 \cup \ldots \cup P_k = X$ (and analogously for $\mathbf{Q}$). Let

$$C = \{ (x_i, x_j) \in X \times X \mid 1 \leq i < j \leq n \}$$

denote the set of all tuples of elements in X.[2] We say that two elements $(x, x') \in C$ are *paired* in $\mathbf{P}$ if they belong to the same

[1] Typically, such criteria compare the intra-cluster variability, i.e., the variability among objects within the same cluster (which should be small) with the inter-cluster variability, i.e., the variability among objects from different clusters (which should be high).

[2] Since we consider unordered tuples, we should more correctly write $\{x_i, x_j\}$ instead of (x_i, x_j).

ISBN: 978-989-95079-6-8

cluster, i.e., if there is a cluster $P_i \in \mathbf{P}$ such that $x \in P_i$ and $x' \in P_i$. Moreover, we distinguish the following subsets of C:

- $C_1 \equiv$ the set of tuples $(x,x') \in C$ that are paired in $\mathbf{P}$ and paired in $\mathbf{Q}$;
- $C_2 \equiv$ the set of tuples $(x,x') \in C$ that are paired in $\mathbf{P}$ but not paired in $\mathbf{Q}$;
- $C_3 \equiv$ the set of tuples $(x,x') \in C$ that are not paired in $\mathbf{P}$ but paired in $\mathbf{Q}$;
- $C_4 \equiv$ the set of tuples $(x,x') \in C$ that are neither paired in $\mathbf{P}$ nor in $\mathbf{Q}$.

Obviously, $\{C_1,C_2,C_3,C_4\}$ is a partition of C, and $a+b+c+d = |C| = n(n-1)/2$, where

$$a = |C_1|,\ b = |C_2|,\ c = |C_3|,\ d = |C_4|. \tag{1}$$

The tuples $(x,x') \in C_1 \cup C_4$ are the *concordant* pairs, i.e., the pairs for which there is agreement between $\mathbf{P}$ and $\mathbf{Q}$, while the tuples $(x,x') \in C_2 \cup C_3$ are the *discordant* pairs for which the two partitions disagree. The Rand index is defined by the number of concordant pairs divided by the total number of pairs:

$$\mathrm{R}(\mathbf{P},\mathbf{Q}) = \frac{a+d}{a+b+c+d} \tag{2}$$

Thus defined, the Rand index is a similarity measure which assumes values between 0 and 1. It can easily be turned into a distance function by defining

$$\mathrm{D}_R(\mathbf{P},\mathbf{Q}) = 1 - \mathrm{R}(\mathbf{P},\mathbf{Q}) = \frac{b+c}{a+b+c+d}\ .$$

It is worth mentioning that D_R satisfies the classical properties of a distance (reflexivity, separation, symmetry, and triangular inequality).

2.2 Campello's Fuzzy Rand Index

The aim of Campello's paper is to extend the Rand index to the case of fuzzy partitions. To this end, he first reformulates it within a set-theoretic framework. An extension to the fuzzy case can then be accomplished in a straightforward way by using generalized set-theoretical operators. Recall that $k = |\mathbf{P}|$ and $\ell = |\mathbf{Q}|$, and consider the following sets:

- $V \equiv$ the set of pairs $(x,x') \in C$ that belong to the same cluster in $\mathbf{P}$; it can be expressed as $V = \bigcup_{i=1\ldots k} V_i$, where V_i is the set of pairs that both belong to the i-th cluster $P_i \in \mathbf{P}$.
- $W \equiv$ the set of pairs $(x,x') \in C$ that belong to different clusters in $\mathbf{P}$; it can be expressed as $W = \bigcup_{1\le i\ne j\le k} W_{ij}$, where W_{ij} is the set of pairs such that $x \in P_i$ and $x' \in P_j$.
- $Y \equiv$ the set of pairs $(x,x') \in C$ that belong to the same cluster in $\mathbf{Q}$; it can be expressed as $Y = \bigcup_{i=1\ldots \ell} Y_i$, where Y_i is the set of pairs that both belong to the i-th cluster $Q_i \in \mathbf{Q}$.
- $Z \equiv$ the set of pairs $(x,x') \in C$ that belong to different clusters in $\mathbf{Q}$; it can be expressed as $Z = \bigcup_{1\le i\ne j\le \ell} Z_{ij}$, where Z_{ij} is the set of pairs such that $x \in Q_i$ and $x' \in Q_j$.

The Rand index can directly be written in terms of the cardinalities of these sets, since the four quantities (1) are obviously given by

$$a = |V \cap W|, \qquad b = |V \cap Z|,$$
$$c = |W \cap Y|, \qquad d = |W \cap Z|.$$

In the fuzzy case, the above sets become fuzzy sets. Let $P_i(x) \in [0,1]$ denote the degree of membership of element $x \in X$ in the cluster $P_i \in \mathbf{P}$. The sets V, W, Y, and Z can then be defined through fuzzy-logical expressions involving a t-norm $\top$ and t-conorm $\bot$:

$$\begin{aligned} V(x,x') &= \bot_{i=1}^{k} \top(P_i(x),P_i(x')) \\ W(x,x') &= \bot_{1\le i\ne j\le k} \top(P_i(x),P_j(x')) \\ Y(x,x') &= \bot_{i=1}^{\ell} \top(Q_i(x),Q_i(x')) \\ Z(x,x') &= \bot_{1\le i\ne j\le \ell} \top(Q_i(x),Q_j(x')) \end{aligned} \tag{3}$$

Moreover, defining the intersection of sets by the t-norm combination of membership degrees and resorting to the commonly used sigma-count principle [8] for defining set cardinality, one obtains

$$\begin{array}{rcccl} a & = & |V \cap Y| & = & \sum_{(x,x')\in C} \top(V(x,x'),Y(x,x')) \\ b & = & |V \cap Z| & = & \sum_{(x,x')\in C} \top(V(x,x'),Z(x,x')) \\ c & = & |W \cap Y| & = & \sum_{(x,x')\in C} \top(W(x,x'),Y(x,x')) \\ d & = & |W \cap Z| & = & \sum_{(x,x')\in C} \top(W(x,x'),Z(x,x')) \end{array} \tag{4}$$

As before, the Rand index can then be defined as in (2), namely as the fraction

$$\frac{a+d}{a+b+c+d}\ .$$

2.3 Properties of Campello's Fuzzy Rand Index

Having defined a similarity or, equivalently, a distance function, it is natural to ask for desirable metrical properties of that function. In the case of the above fuzzy Rand index, however, this question has to be considered with caution, since Campello is actually only interested in comparing a fuzzy partition $\mathbf{P}$ with a non-fuzzy partition $\mathbf{Q}$. And indeed, formal properties of the measure are not investigated in his paper.

On the other hand, it is noted by Campello himself that, formally, the measure can in principle be applied to compare two fuzzy partitions. When doing so, however, it turns out quickly that it fails to be a proper metric. In fact, it does not even satisfy reflexivity, the perhaps most basic axiom: Even for two identical partitions $\mathbf{P}$ and $\mathbf{Q}$, the quantities b and c in (4) will generally not vanish, a necessary condition for having $\mathrm{R}(\mathbf{P},\mathbf{Q}) = 1$.

Consider, for example, the simple fuzzy partition $\mathbf{P}$ illustrated in Fig. 1, which consists of two clusters P_1 and P_2. Instead of a hard boundary, there is a "soft" transition between P_1 and P_2; the elements x_1, x_2, x_3, and x_4 partially belong to both clusters and have membership degrees, respectively, of $3/4$, $1/2$, $1/2$, $1/4$ in P_1 and $1/4$, $1/2$, $1/2$, $3/4$ in P_2. Comparing $\mathbf{P}$ to itself in terms of the fuzzy Rand index, we obtain $\mathrm{R}(\mathbf{P},\mathbf{P}) < 1$.

Upon closer examination, it seems that the core principle of Campello's extension is not suitable for comparing partitions in a fuzzy sense. In fact, despite being defined in terms

ISBN: 978-989-95079-6-8

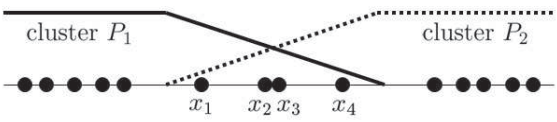


Figure 1: Illustration of a simple fuzzy partition of a subset of the reals (indicated by circles). The partition consists of two clusters, P_1 (left) and P_2 (right). While some elements definitely belong to only one of the clusters, some "critical" points in the middle have partial membership in both clusters.

of fuzzy logical formulas, the very idea of the approach has arguably a more "probabilistic flavor". This becomes especially obvious when using the product as a t-norm and the (bounded) sum as t-conorm. Then, if $P_i(x)$ is interpreted as the probability that x belongs to the i-th cluster, $V(x,x')$ is nothing else than the probability that x and x' are put in the same cluster, given that the two corresponding clusters are chosen *independently of each other* according to the distributions $(P_1(x),P_2(x)\dots P_k(x))$ and $(P_1(x'),P_2(x')\dots P_k(x'))$, respectively. Likewise, $W(x,x')$ is the probability that x and x' are put into different clusters.

Even if one accepts the probabilistic interpretation of a single membership degree, the additional assumption of independence is clearly not tenable. In fact, this property is obviously violated when comparing a partition with itself, since for each element $x \in X$, a cluster can then only be chosen once and not two times independently of each other. But even if $\mathbf{P}$ and $\mathbf{Q}$ are not identical, independence of cluster membership is in conflict with the topological relationships between the elements and clusters. In the example in Fig. 1, for instance, it is not reasonable to put x_1 and x_4 into cluster P_2 and x_2 and x_3 into cluster P_1. When putting elements independently of each other into clusters, however, this is a possible scenario. And indeed, this scenario contributes to Campello's fuzzy Rand index according to (3).

Seen from this point of view, one may even question the usefulness of the approach for its original purpose, namely the comparison of a fuzzy with a non-fuzzy partition. What the fuzzy partition in our example truly suggests is that we are uncertain about the boundary between the two clusters. More concretely, the fuzzy partition suggests four possible non-fuzzy partitions:

- $\mathbf{P}_1$ which puts the boundary left to x_1;
- $\mathbf{P}_2$ with boundary between x_1 and x_2;
- $\mathbf{P}_3$ with boundary between x_3 and x_4;
- $\mathbf{P}_4$ which puts the boundary right to x_4.

Thus, it seems reasonable to define an extension of the Rand index as an aggregation (e.g., weighted average) of the results of the non-fuzzy comparisons, namely

$$\mathrm{R}(\mathbf{P}_1,\mathbf{Q}), \mathrm{R}(\mathbf{P}_2,\mathbf{Q}), \mathrm{R}(\mathbf{P}_3,\mathbf{Q}), \mathrm{R}(\mathbf{P}_4,\mathbf{Q}).$$

In Campello's approach, there are not 4 but 16 scenarios which have an influence on the result, since each of the four cluster memberships is determined independently of each other. In general, the result will therefore be different. In fact, differences already occur for single pairs of elements. For example, since x_2 and x_3 are always in the same cluster in $\mathbf{P}_1,\dots,\mathbf{P}_4$, it is natural to say that they are paired with degree 1. According to Campello's approach, however, the degree to which x_2 and x_3 are in the same cluster in $\mathbf{P}$ is given by

$$V(x_2,x_3) = \bot(\top(1/2,1/2),\top(1/2,1/2)),$$

which corresponds to the truth degree of the proposition that "x_1 is put into P_1 AND x_2 is put into P_1 OR x_1 is put into P_2 AND x_2 is put into P_2". In general, this degree will be < 1 (except for special $(\top,\bot)$-combinations such as $\top = \min$ and $\bot =$ bounded sum).

3 A New Fuzzy Rand Index

In this section, we propose a new fuzzy variant of the Rand index which is able to compare any pair of fuzzy partitions and, moreover, has desirable metric properties. In the following, we focus on the view of the Rand index as a distance function. Thanks to the affine transformation $\mathrm{D}_R = 1 - \mathrm{R}$, all results can directly be transferred to the original conception as a measure of similarity.

3.1 Definition

Given a fuzzy partition $\mathbf{P} = \{P_1,P_2\dots P_k\}$ of X, each element $x \in X$ can be characterized by its membership vector

$$\mathbf{P}(x) = (P_1(x),P_2(x)\dots P_k(x)) \in [0,1]^k, \tag{5}$$

where $P_i(x)$ is the degree of membership of x in the i-th cluster P_i. We define a fuzzy equivalence relation on X in terms of a similarity measure on the associated membership vectors (5). Generally, this relation is of the form

$$E_{\mathbf{P}}(x,x') = 1 - \|\mathbf{P}(x) - \mathbf{P}(x')\|, \tag{6}$$

where $\|\cdot\|$ is a proper distance on $[0,1]^k$. The basic requirement on this distance is that it yields values in $[0,1]$. The relation (6) generalizes the equivalence relation induced by a conventional partition (where each cluster forms an equivalence class). In passing, we note that this definition is invariant toward a permutation (renumbering) of the clusters in $\mathbf{P}$, which is clearly a desirable property.

Now, given two fuzzy partitions $\mathbf{P}$ and $\mathbf{Q}$, the idea is to generalize the concept of concordance as follows. We consider a pair (x,x') as being concordant in so far as $\mathbf{P}$ and $\mathbf{Q}$ agree on their degree of equivalence. This suggest to define the *degree of concordance* as

$$1 - |E_{\mathbf{P}}(x,x') - E_{\mathbf{Q}}(x,x')| \in [0,1]. \tag{7}$$

Analogously, the *degree of discordance* is

$$|E_{\mathbf{P}}(x,x') - E_{\mathbf{Q}}(x,x')| \ .$$

Our distance measure on fuzzy partitions is then defined by the normalized sum of degrees of discordance:

$$d(\mathbf{P},\mathbf{Q}) = \frac{\sum_{(x,x')\in C} |E_{\mathbf{P}}(x,x') - E_{\mathbf{Q}}(x,x')|}{n(n-1)/2} \tag{8}$$

Likewise,

$$1 - d(\mathbf{P},\mathbf{Q}) \tag{9}$$

corresponds to the normalized degree of concordance and, therefore, is a direct generalization of the original Rand index.

ISBN: 978-989-95079-6-8

3.2 Formal Properties

In this section, we first show that our proposal is indeed a proper generalization of the Rand index. Afterward, we study the metrical properties of the measure.

Proposition: *In the case where* **P** *and* **Q** *are non-fuzzy partitions, the measure (9) reduces to the original Rand index.*

Proof: In the non-fuzzy case, the membership vectors (5) are 0/1-vectors. More specifically, each vector has a single entry $P_i(x) = 1$, while all other entries are 0. Consequently, the fuzzy equivalence (6) reduces to the conventional equivalence, that is, $E_{\mathbf{P}}(x,x') = 1$ if x and x' are in the same cluster and $E_{\mathbf{P}}(x,x') = 0$ otherwise. Likewise, (7) yields 1 if (x,x') is a concordant pair and 0 otherwise. Consequently, the measure (9) is the (normalized) sum of concordant pairs and, therefore, equals the original Rand index. □

Recall that a non-negative $U^2 \to \mathbb{R}$ mapping $d(\cdot)$ is called a metric on U if is satisfies the following properties for all $u,v,w \in U$:

- Reflexivity: $d(u,u) = 0$
- Separation: $d(u,v) = 0$ implies $u = v$
- Symmetry: $d(u,v) = d(v,u)$
- Triangle inequality: $d(u,w) \leq d(u,v) + d(v,w)$

The properties of reflexivity and symmetry are quite obviously valid for our measure (8). To show the triangle inequality, consider three fuzzy partitions **P**, **Q**, **R** and fix a single tuple $(x,x') \in C$. Let

$$a = E_{\mathbf{P}}(x,x'),\ b = E_{\mathbf{Q}}(x,x'),\ c = E_{\mathbf{R}}(x,x').$$

Since a, b, and c are real numbers (from the unit interval), and the simple difference on the reals satisfies the triangle inequality, we have $|a-c| \leq |a-b| + |b-c|$. Now, since this inequality holds for each pair $(x,x') \in C$, it remains valid when summing over all these pairs. In other words, it is also satisfied by (8), which means that

$$d(\mathbf{P},\mathbf{R}) \leq d(\mathbf{P},\mathbf{Q}) + d(\mathbf{Q},\mathbf{R}).$$

The separation property is not immediately valid for (8). Roughly speaking, this is due to the fact that, by mapping elements to their membership vectors (5), some information about the partition itself is lost. In particular, it is possible that two partitions, even though they are not identical, cannot be distinguished in terms of the distances between these vectors.

Nevertheless, we can guarantee the separation property by restricting to a reasonable subclass of fuzzy partitions. We call a fuzzy partition $\mathbf{P} = \{P_1, P_2 \ldots P_k\}$ *normal*, if it satisfies the following:

N1 For each $x \in X$: $P_1(x) + \ldots + P_k(x) = 1$.

N2 For each $P_i \in \mathbf{P}$, there exists an $x \in X$ such that $P_i(x) = 1$.

In other words, we consider Ruspini partitions [9] and assume that each cluster has a prototypical element. Moreover, we assume the following equivalence relation on X:

$$E_{\mathbf{P}}(x,x') = 1 - \frac{1}{2}\sum_{i=1}^{k} |P_i(x) - P_i(x')|. \tag{10}$$

Note that $0 \leq E_{\mathbf{P}}(x,x') \leq 1$ for all $(x,x') \in X^2$ under assumption N1.

Now, consider two normal fuzzy partitions **P** and **Q**, and suppose that $d(\mathbf{P},\mathbf{Q}) = 0$. According to our definition of $d(\cdot)$, this obviously means that

$$E_{\mathbf{P}}(x,x') = E_{\mathbf{Q}}(x,x') \tag{11}$$

for all $(x,x') \in C$. We call a set $\{p_1, p_2 \ldots p_k\} \subset X$ a prototype set for **P**, if $P_i(p_i) = 1$ for all $i = 1,\ldots,k$ (note that a prototype set is not necessarily unique). We distinguish two cases.

(a) There are no identical prototype sets for **P** and **Q** (note that this is necessarily the case if **P** and **Q** have a different number of clusters). Then, we can find elements $x,x' \in X$ such that x and x' are prototypes for **P** but not for **Q**. Note that N1 and N2 jointly imply that a prototype is represented by a 0/1 membership vector, and that $|\mathbf{P}(x) - \mathbf{P}(x')| = 1$ for two different prototypes x and x'. Moreover, these properties imply that the extreme distance of 1 can *only* be assumed for prototypes, whereas $|\mathbf{P}(x) - \mathbf{P}(x')| < 1$ if either x or x' is not a prototype. Thus, it follows that $E_{\mathbf{P}}(x,x') = 0$ and $E_{\mathbf{Q}}(x,x') > 0$, which means that condition (11) is violated. Hence, we have constructed a contradiction with the assumption that $d(\mathbf{P},\mathbf{Q}) = 0$.

(b) There are identical prototype sets $\{p_1,\ldots,p_k\} = \{q_1,\ldots,q_\ell\}$, respectively, for **P** and **Q** (which means that $k = \ell$, i.e., **P** and **Q** do have the same number of clusters). We can then establish a one-to-one correspondence between prototypes such that, without loss of generality, $p_i = q_i$ for $i = 1,\ldots,k$. From properties N1 and N2, it follows that the membership degree of any element x in the cluster P_i is a function of $E_{\mathbf{P}}(x,p_i)$. In fact, noting that $\mathbf{P}(p_i)$ is a 0/1 vector with a single 1 on position i, we get

$$\begin{aligned} E_{\mathbf{P}}(x,p_i) &= 1 - \frac{1}{2}\sum_{j=1}^{k} |P_j(x) - P_j(p_i)| \\ &= 1 - \frac{1}{2}\left((1-P_i(x)) - \sum_{j\neq i} P_j(x)\right) \\ &= 1 - \frac{1}{2}\left((1-P_i(x)) - (1-P_i(x))\right) \\ &= P_i(x). \end{aligned}$$

From (11), it thus follows that $P_i(x) = Q_i(x)$ for all $x \in X$, i.e., the i-th cluster in **P** and the i-th cluster in **Q** are identical. Since this holds for all $i \in \{1,2\ldots k\}$, we have shown that $\mathbf{P} = \mathbf{Q}$.

The above results can be summarized as follows.

Theorem: *The distance function (8) on fuzzy partitions is a pseudometric, i.e., it is reflexive, symmetric, and subadditive. Moreover, on the restricted class of normal fuzzy partitions or, more specifically, under the assumptions N1, N2, and (10), it also satisfies the separation property and, therefore, is a metric.*

4 Summary and Outlook

We have introduced a generalization of the Rand index for comparing two fuzzy clustering structures. Elaborating on the formal properties of our measure, we have shown that it is a

ISBN: 978-989-95079-6-8

pseudo-metric and, on a subclass of fuzzy partitions obeying certain normality assumptions, even a metric.

In future work, we plan to extent our approach to other similarity measures for (non-fuzzy) clustering structures which are related to the Rand index in the sense of being defined in terms of the same basic quantities, namely the numbers a, b, c, and d of concordant and discordant object pairs. An example of such a measure is the Jaccard coefficient, which is defined as $a/(a+b+c)$.

References

[1] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. Clustering validity checking methods: Part I. *ACM SIGMOD Record*, 31(2):40–45, 2002.

[2] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. Clustering validity checking methods: Part II. *ACM SIGMOD Record*, 31(3):19–27, 2002.

[3] J. Beringer and E. Hüllermeier. Fuzzy clustering of parallel data streams. In J. Valente de Oliveira and W. Pedrycz, editors, *Advances in Fuzzy Clustering and Its Application*, pages 333–352. John Wiley and Sons, 2007.

[4] MP. Windham. Cluster validity for the fuzzy c-means algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 4(4):357–363, 1982.

[5] XL. Xie and GA. Beni. Validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(8):841–846, 1991.

[6] R. J. G. B. Campello. A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment. *Pattern Recognition Letters*, 28(7):833–841, 2007.

[7] William M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

[8] A. DeLuca and S. Termini. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and Control*, 20:301–312, 1972.

[9] E.H. Ruspini. A new approach to clustering. *Information Control*, 15:22–32, 1969.

ISBN: 978-989-95079-6-8

Fuzzy Modeling of Labeled Point Cloud Superposition for the Comparison of Protein Binding Sites

Thomas Fober and Eyke Hüllermeier

FB Mathematik/Informatik, Philipps-Universität Marburg
D-35032 Marburg, Germany
Email: {thomas,eyke}@mathematik.uni-marburg.de

Abstract— *Geometric objects are often represented approximately in terms of a finite set of points in three-dimensional Euclidean space. In this paper, we extend this representation to what we call labeled point clouds. A labeled point cloud is a finite set of points, where each point is not only associated with a position in three-dimensional space, but also with a discrete class label that represents a specific property. This type of model is especially suitable for modeling biomolecules such as proteins and protein binding sites, where a label may represent an atom type or a physico-chemical property. Proceeding from this representation, we address the question of how to compare two labeled points clouds in terms of similarity. Using fuzzy modeling techniques, we develop a suitable similarity measure as well as an efficient evolutionary algorithm to compute it. Finally, an application study is presented in which the approach is used to classify protein binding sites.*

Keywords— structural bioinformatics, proteins, similarity, classification

1 Introduction

Geometric objects are often represented in terms of a set of points in three-dimensional Euclidean space. This type of representation is finite and hence approximate (even though the number of points can become very large, as for example in laser range scanning), focusing on the most important characteristics of the object while ignoring less important details. A well-known example of a representation of this kind is the *Molfile* format [1], where molecules are described in terms of the spatial coordinates of all atoms. However, since not only the position but also the type of an atom is of interest, this representation is not a simple point cloud. Likewise, other biomolecular structures, such as proteins and protein binding sites, are not only characterized by their geometry but also by additional features, such as physico-chemical properties. In this paper, we therefore introduce the concept of a *labeled point cloud*. A labeled point cloud is a finite set of points, where each point is not only associated with a position in three-dimensional space, but also with a discrete class label that represents a specific property.

Since theory formation in the biological sciences is largely founded on similarity-based and analogical reasoning principles, the comparison of two (or more) objects with each other is a fundamental problem in bioinformatics. To compare two point clouds, the authors in [2] make use of a measure based on the *Gromov-Hausdorff distance* of sets. This approach is limited to unlabeled point clouds, however. Another possibility is to transform a labeled point cloud into a (labeled) graph first, capturing, in one way or the other, geometrical information in terms of edges, and to apply graph matching techniques afterward. This strategy was recently proposed in [3], where the use of *graph kernels* as similarity measures [4, 5, 6] has been especially advocated. At first sight, this idea looks appealing, especially since methods for comparing graphs abound in the literature. Nevertheless, it also comes with a number of disadvantages. For example, many techniques for matching and comparing graphs capture aspects of similarity which are reasonable for graphs but not necessarily for geometric objects. Besides, graph matching techniques are typically quite complex from a computational point of view.

Perhaps most importantly, however, a graph representation captures the geometrical information only in an *implicit* way, namely through the presence, absence, and possibly the labeling of edges. Moreover, the transformation is often not even lossless. Matching objects while obeying geometrical constraints can then become troublesome, since the geometrical information is not explicitly available. Instead, it must be reconstructed from the graph representation whenever needed.

As an alternative to an indirect approach of that kind, we therefore propose the method of *labeled point cloud superposition* (LPCS), which operates on labeled point clouds directly. Thus, it preserves as much geometrical information as possible and facilitates the exploitation thereof. Related to the concept of an LPCS, we introduce a similarity measure which makes use of modeling techniques from fuzzy set theory. This measure proceeds from the idea of equivalence (inclusion) of point clouds in a set-theoretic sense, but is tolerant toward exceptions (on the level of label information) and geometric deformations.

The remainder of the paper if organized as follows. Subsequent to a brief introduction to protein binding sites and their representation in Section 2, we introduce the concept of LPCS in Section 3. The problem of computing an LPCS is then addressed in Section 4, where an evolution strategy is proposed for this purpose. Section 5 is devoted to the experimental validation of the approach, and Section 6 concludes the paper.

2 Modeling Protein Binding Sites

In this paper, our special interest concerns the modeling of protein binding sites. More specifically, our work builds upon CavBase [7], a database for the automated detection, extraction, and storing of protein cavities (hypothetical binding sites) from experimentally determined protein structures (available through the PDB). In CavBase, a set of points is used as a first approximation to describe a binding pocket. The database currently contains 113,718 hypothetical binding sites that have been extracted from 23,780 publicly available protein structures using the LIGSITE algorithm [8].

ISBN: 978-989-95079-6-8

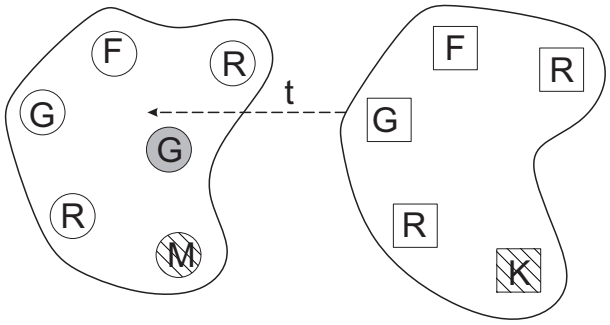


Figure 1: Two point clouds A (left, points as circle) and B (right, points as squares): The intra-point distances are the same in both point clouds, except for the additional gray point in A. Labels are depicted as letters within the circles and boxes, respectively.

The geometrical arrangement of the pocket and its physicochemical properties are first represented by predefined *pseudocenters* – spatial points that represent the center of a particular property. The type and the spatial position of the centers depend on the amino acids that border the binding pocket and expose their functional groups. They are derived from the protein structure using a set of predefined rules [7]. As possible types for pseudocenters, hydrogen-bond donor, acceptor, mixed donor/acceptor, hydrophobic aliphatic, metal ion, pi (accounts for the ability to form π–π interactions) and aromatic properties are considered.

Pseudocenters can be regarded as a compressed representation of areas on the cavity surface where certain protein-ligand interactions are experienced. Consequently, a set of pseudocenters is an approximate representation of a spatial distribution of physicochemical properties. Obviously, just like in the case of Molfile, this representation is already in the form of a labeled point cloud: pseudocenters are given with their coordinates and labels, so that no further transformation is needed.

3 Labeled Point Cloud Superposition

Intuitively, two labeled point clouds are similar if they can be spatially superimposed. That is, by fixing the first and "moving" the second one (as a whole, i.e., without changing the internal arrangement of points) in a proper way, an approximate superposition of the two structures is obtained. More specifically, we will say that two point clouds are well superimposed if, for each point in one of the structures, there exists a point in the other cloud which is spatially close and has the same label. As an illustration, the example in Fig. 1 shows two point clouds A and B, for simplicity only in two dimensions. By moving B to the left (or A to the right), a superposition can be found so that, except for the hatched and gray nodes, all points in A spatially coincide with a corresponding point in B having the same label, and vice versa. So, A and B can be considered as being similar, at least to some extent.

More formally, let

$$A = \{(x_1, \ell(x_1)), \ldots, (x_m, \ell(x_m))\}$$

be a point cloud consisting of m points $x_i = (x_{i1}, x_{i2}, x_{i3}) \in \mathbb{R}^3$ with associated label $\ell(x_i) \in \mathcal{L}$, where $\mathcal{L}$ is a discrete set of labels (in the context of modeling protein binding sites, as discussed in the previous section, $\mathcal{L}$ is given by the seven types of pseudocenters). Moreover, let

$$B = \{(y_1, \ell(y_1)), \ldots, (y_n, \ell(y_n))\}$$

be a second point cloud to be compared with A. In the following, we define a function $\mathrm{SIM}(\cdot,\cdot)$ that returns a degree of similarity between two such structures A and B.

Roughly speaking, we consider similarity as a generalized (fuzzy) equivalence, which we in turn reduce to two inclusion relations, namely the inclusion of A in B and, vice versa, of B in A. Thus, we are first of all interested in whether each point $y \in B$ is also present in A (and each point $x \in A$ also present in B). For a fixed $y \in B$, we define the membership degree of this point in A by

$$\mu_A(y) = \exp\left(-\gamma \cdot d(y, A)\right) \quad , \tag{1}$$

where

$$d(y, A) = \min_{\substack{x \in A \\ \ell(x) = \ell(y)}} \|y - x\|_1$$

is the distance between a point $y \in B$ and the closest point $x \in A$ having the same label ($d(y,A) = \infty$ and hence $\mu_A(y) = 0$ if no such point exists); for $x \in A$, $\mu_B(x)$ and $d(x,B)$ are defined analogously.

In its proper sense, the inclusion of a set B in a set A means that *each* point $y \in B$ is also contained in A or, stated differently, if a point y is in B, then it is also present in A. If membership is a matter of degree, i.e., if A and B are fuzzy sets, this condition is often formalized in terms of a fuzzy implication [9]:

$$\min_{y \in B} \left(\mu_B(y) \to \mu_A(y)\right) \quad .$$

Here, the minimum operator plays the role of a generalization of the universal quantifier. In our case, $\mu_B(y) \equiv 1$, so that the above expression can be simplified as follows:

$$inc(B, A) = \min_{y \in B} \mu_A(y) \quad . \tag{2}$$

However, a universal quantification (modeled by the min operator) is too strict in our biological context, where data is typically inexact and noisy. To relax this definition of fuzzy inclusion, we replace the minimum by a fuzzy quantifier Q, which is specified in the form of a non-decreasing $[0,1] \to [0,1]$ mapping [10, 11]. This leads to

$$inc(B, A) = \min_{i=1\ldots|B|} \max\{Q(i/|B|), m_i\} \quad ,$$

where m_i is the i-th largest membership degree in the fuzzy set $\{\mu_A(y) \,|\, y \in B\}$. (Note that we recover (2) for Q defined by $Q(1) = 1$ and $Q(t) = 0$ for $0 \leq t < 1$.) Here, we simply take Q as the identical mapping $t \mapsto t$. Roughly speaking, $inc(B,A)$ thus defined can be interpreted as the generalized truth degree of the proposition that A is *almost* contained in B. The degree of inclusion of A in B, $inc(A,B)$, is defined analogously.

As mentioned above, the idea of our approach is to define the similarity between two labeled point clouds in terms of the best superposition of these two clouds. Therefore, let $\mathrm{TF}(\cdot, t)$ be a function that moves a point cloud via rotation and translation, as specified by the six-dimensional vector $t = (\theta_1, \theta_2, \theta_3, \delta_1, \delta_2, \delta_3) \in [0, 2\pi]^3 \times \mathbb{R}^3$. Thus,

$$B^* = \mathrm{TF}(B, t) = \{(y_1^*, \ell(y_1^*)), \ldots, (y_n^*, \ell(y_n^*))\}$$

ISBN: 978-989-95079-6-8

is the point cloud obtained by translating the point cloud B by $\delta = (\delta_1, \delta_2, \delta_3)$ (which means adding δ to each point $y \in B$) and rotating the result thus obtained by the angles θ_1, θ_2, and θ_3. Note that this operation leaves the label information unchanged (i.e., $\ell(y_i) = \ell(y_i^*)$). The position-invariant degree of inclusion of B in A is then given by

$$\mathrm{INC}(B,A) = \max_{t \in [0,2\pi]^3 \times \mathbb{R}^3} inc(\mathrm{TF}(B,t),A) \ , \qquad (3)$$

and $\mathrm{INC}(A,B)$ is defined analogously.

Based on these degrees, the similarity between A and B, in the sense of a generalized equivalence, can be defined as

$$\mathrm{SIM}(A,B) = \min\{\,\mathrm{INC}(A,B), \mathrm{INC}(B,A)\,\} \ . \qquad (4)$$

It is worth mentioning, however, that (4) is not always appropriate, especially if A and B greatly differ in size. In some applications, it makes sense to have a high similarity degree even if A is only a substructure of B, for example if A is a subpocket of B containing the most important catalytic residues (while the rest of the binding site B is functionally less important). Obviously, this is not guaranteed by (4). An interesting generalization, therefore, is to let

$$\begin{aligned} \mathrm{SIM}(A,B) = & \alpha \cdot \min\{\mathrm{INC}(A,B), \mathrm{INC}(B,A)\} + \\ & + (1-\alpha) \cdot \max\{\mathrm{INC}(A,B), \mathrm{INC}(B,A)\} \ . \end{aligned} \qquad (5)$$

Formally, this similarity measure can be motivated from a fuzzy logical point of view as follows. Considering the min (max) operator as a generalized conjunction (disjunction), the first (second) combination of the two inclusion degrees is the truth degree of the proposition that A is contained in B AND (OR) B is contained in A. A conjunctive combination of the two degrees of inclusion is obviously more demanding than a disjunctive one, as the former requires equality between A and B while the latter only requires inclusion of A in B or B in A. The measure (5), which formally corresponds to an OWA (ordered weighted average) combination of the two degrees of inclusion [12], achieves a trade-off between these two extreme aggregation modes, which is controlled by the parameter $\alpha \in [0,1]$: The closer α is to 0, the closer the aggregation is to the maximum, i.e., the less demanding it becomes. The optimal α is application-specific and depends on the purpose of the similarity measure.

4 Solving the LPCS Problem

The computation of the similarity (5) involves the solution of a real-valued optimization problem, namely the problem of finding an optimal vector t in (3) and, thus, an optimal point cloud superposition. The objective function to be maximized here is highly non-linear and multimodal. As an illustration, Fig. 2 shows the objective function obtained for the superposition of a randomly generated two-dimensional point cloud A (in which all points have the same label) with itself. This function maps each two-dimensional translation vector $t = (x,y)$ to the corresponding similarity degree between $\mathrm{TF}(A)$ and A (where we used $\alpha = 0.5$ in (5) and did not consider rotation). As can be seen, there is a sharp peak at $t = (0,0)$, which corresponds to the optimal superposition. Surrounding this solution, however, there are also many local optima.

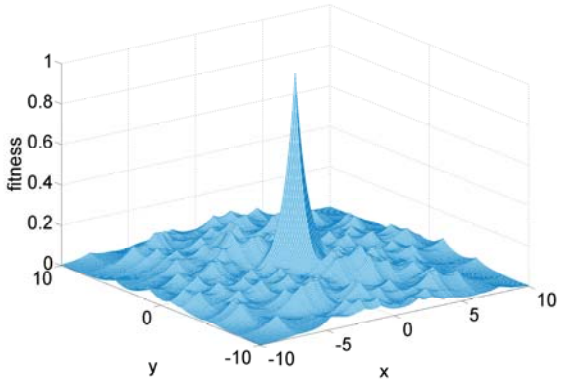


Figure 2: Example of an LPCS objective function.

The problem of local optima also becomes clear from the small example in Fig. 1. Moving the point cloud A from left to right, into the direction of B, has the following effect: First, a good superposition of two sub-clouds will be found, namely the right part of cloud A and the left part of cloud B. This results in a local maximum. Moving A further to the right leads to a larger local maximum (sub-clouds are growing), until the global maximum will eventually be reached.

4.1 Evolution Strategies

To solve the LPCS problem, we resort to *evolution strategies* (ES), a population-based, stochastic optimization method inspired by biological evolution and specifically developed for real-valued optimization problems [13]. An evolution strategy is based on a population, a set of μ (sub-optimal) candidate solutions that are initially spread randomly over the search space. In each generation, new solutions are generated by applying the genetic operators *recombination* and *mutation*. Recombination randomly selects ρ individuals from the current population and combines them to a new solution. Mutation takes this solution and shifts it randomly in the search space. An ES produces $\lambda = \lceil \mu \cdot \nu \rceil$ offsprings per iteration, so that this procedure has to be repeated λ times. A selection operator implements the "survival of the fittest" principle by picking the best individuals for the new population. There are two kinds of selection: The *plus*-selection chooses the best μ individuals among the offsprings plus the parents, while the *comma*-selection ignores the parent generation (this requires $\nu > 1$). A main advantage of the ES is its self-adaptation mechanism that controls the step sizes used in the mutation operator. One property of this mechanism (the advantage during optimization is obvious) is that step sizes decrease dramatically if the optimization reached a maximum. This property can be used as a qualitative termination criterion (stop when the largest step size falls below a given threshold).

Population-based optimization methods are especially advantageous for highly multimodal problems. Using a large population leads to an increased probability to generate a candidate solution in a region where the direction of descent points to the global maximum. Choosing the membership function (1) as a strictly monotone decreasing function which converges to zero ensures to have this direction in each point $t \in [0,2\pi]^3 \times \mathbb{R}^3$ and thus greatly simplifies the maximization problem. However, our experiments indicated that the solu-

ISBN: 978-989-95079-6-8

tion we found was most often only a local maximum. Therefore, we propose to use *fast restarts* of the ES. This means that the ES is started n times using comma-selection and weak termination criteria to achieve a large and quick but inexact exploration of our search space. We thus obtain n results in total. In a last step, we use the ES with plus-selection and strong termination criterion. Additionally, we include the best solution so far in the start population. The last run of the ES usually yields a globally optimal degree of similarity.

4.2 Complexity

Even though evolution strategies are generally known to be quite efficient solvers, the concrete complexity does of course depend on the application at hand. The application-specific part is the fitness function, i.e., the objective function to be optimized. This function has to be evaluated frequently and, therefore, is an important factor for the runtime. In our case, this function is given by the similarity measure (5), and its evaluation is strongly dominated by the nearest neighbor search which has to be conducted for each single point in both structures (recall that, according to (1), membership degrees are determined by the distance to closest points with the same label).

There exist a lot of data structures for supporting nearest neighbor search; see e.g. [14]. The most efficient among them need time $O(n\log^2 n)$ for construction and $O(\log^3 n)$ for answering a query. Unfortunately, we are not aware of an approach that allows for updating a data structure in an efficient and dynamic way. This would be desirable for our problem, in which the point clouds permanently change (the point cloud associated with an individual changes in each iteration). Instead, conventional approaches necessitate a construction from scratch in every iteration.

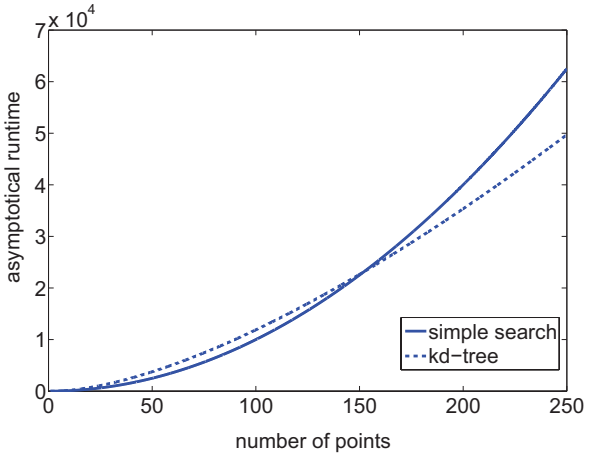


Figure 3: Runtime of a simple procedure and a more complex data structure as a function of the number of points.

Fig. 3 compares the runtimes, as a function of the number of points, for two approaches: (1) The use of a kd-tree data structure, which is reconstructed in each iteration and then used for query processing. (2) The use of a simple linear data structure, in which the points are stored in a fixed order. It needs linear instead of logarithmic time to answer a query but, on the other hand, does not cause additional costs for reconstruction. As can be seen, the use of a more complex approach pays off only for sufficiently large point clouds: The kd-tree reaches a break-even point at approximately 150 points.

In our application, we are mainly concerned with protein binding sites, which are characterized by around 180 points on average (even though much larger structures do of course exist). The use of a complex data structure did therefore not pay off. Nevertheless, we increased efficiency by hashing the points x_i of a point cloud, using the label $\ell(x_i) \in \mathcal{L}$ as a key. Since nearest neighbors are only searched among points having the same label, this obviously reduces runtime by a factor of approximately $|\mathcal{L}|$.

5 Experimental Results

5.1 Methods

In our experiments, we compared our novel method (LPCS) with existing graph-based approaches, namely the random walk (RW) kernel [6], the shortest path (SP) kernel [5], and the method of multiple graph alignment (MGA) recently introduced in [15]. Given two labeled points clouds as input, all these methods produce a degree of similarity as an output. Yet, for the graph-based approaches, it is of course necessary to transform a point cloud into a graph representation in a preprocessing step. This was done as as proposed in [15]:

1. each point is transformed into a node with corresponding node label
2. for each pair of nodes:
   (a) the Euclidean distance between both nodes is calculated
   (b) if the distance is below a certain threshold (here 11 Å to ensure connected graphs), an edge with weight equal to this distance is added

Our ES was restarted $n = 5$ times. The parameterization was optimized with the *sequential parameter optimization toolbox* [16] and was chosen as follows:

- inexact ES: $\mu = 30, \nu = 4, \rho = 2$, comma-selection, termination criteria: largest step size < 0.05, discrete recombination for strategy- and object-component.
- exact ES: $\mu = 30, \nu = 4, \rho = 6$, plus-selection, termination criteria: largest step size < 0.00001, intermediate recombination for object and discrete recombination for strategy-component.

A comprehensive explanation of the different ES parameters and operators can be found in [13].

For both variants we initialized the object-component in $[-150, 150]^3$ for translation and $[0, 2\pi]^3$ for rotation: The step sizes were initialized in $[5, 15]^3$ and $[1, \pi]^3$, respectively. The SP-kernel is parameter-free, the RW-kernel expects a parameter λ that is set to the largest degree of a node in the data set to ensure a geometric series during calculation, which results in a simpler evaluation [4]. Since the geometric information of real-world data is noisy, we also need a tolerance parameter ε to decide whether two edges have equal length (difference $\leq \varepsilon$) or not; in our experiments, we used $\varepsilon = 0.2$. For MGA, we chose the parameterization proposed in [15].

ISBN: 978-989-95079-6-8

The assessment of a similarity measure for biomolecular structures, such as protein binding sites, is clearly a non-trivial problem. In particular, since the concept of similarity by itself is rather vague and subjective, it is difficult to evaluate corresponding measures in an objective way. To circumvent this problem, we propose to evaluate similarity measures in an indirect way, namely by means of their performance in the context of nearest neighbor (NN) classification. The underlying idea is that, the better a similarity measure is, the better should be the predictive performance of an NN classifier using this measure for determining similar cases.

5.2 Data

One important problem in pharmaceutical chemistry is the identification of protein binding sites that bind a certain ligand. We selected two classes of binding sites that bind, respectively, to NADH or ATP. This gives rise to a binary classification problem: Given a protein binding site, predict whether it binds NADH or ATP.

More concretely, we compiled a set of 355 protein binding pockets representing two classes of proteins that share, respectively, ATP and NADH as a cofactor. To this end, we used CavBase to retrieve all known ATP and NADH binding pockets that were co-crystallized with the respective ligand. Subsequently, we reduced the set to one cavity per protein, thus representing the enzymes by a single binding pocket. As protein ligands adopt different conformations due to their structural flexibility, it is likely that the ligands in our data set are bound in completely different ways, hence the corresponding binding pocket does not necessarily share much structural similarity. We thus had to ensure the selection of binding pockets with ligands bound in similar conformation. To achieve this, we used the Kabsch algorithm [17] to calculate the root mean square deviation (RMSD) between pairs of ligand structures. Subsequently, we combined all proteins whose ligands yielded a RMSD value below a threshold of 0.2, thereby ensuring a certain degree of similarity. This value was chosen as a trade-off between data set size and similarity. Eventually, we thus obtained a two-class data set comprising 214 NADH-binding proteins and 141 ATP-binding proteins.

5.3 Results

The results of a leave-one-out cross validation, using the simple 1-NN classifier for prediction, are summarized in Table 1. As can be seen, the kernel-based methods (SP and RW) perform very poorly and are hardly better than random guessing. In terms of accuracy, MGA is much better, though still significantly worse than LPCS. In fact, LPCS performs clearly best on this problem.

Table 1: Accuracy and runtimes (in seconds with standard deviation, referring to a single comparison) of LPCS ($\alpha = 0.5$, with restarts like described above), MGA, RW, and SP on the NADH/APT data set.

Method	Accuracy	Runtime
MGA	0.7662	121.74 ± 418.02
SP	0.6056	$\mathbf{9.75 \pm 97.77}$
RW	0.5972	65.51 ± 89.07
LPCS	**0.9352**	20.04 ± 24.65

Table 2 furthermore shows how the performance of LPCS depends on the choice of the trade-off parameter α in (5). As can be seen, this parameter does indeed have an influence, even though the differences are not extreme. For this data set, α-values around 0.5 yield better results than extreme values close to 0 or 1; the optimal choice would be $\alpha = 0.7$. In practice, α can be considered as a tuning parameter to be adapted to the problem at hand (e.g., by means of a cross-validation on the training data).

Table 2: Accuracy of LPCS for different values of α in (5).

α	accuracy	α	accuracy
0	0.9042	0.6	0.9352
0.1	0.9183	0.7	0.9380
0.2	0.9126	0.8	0.9239
0.3	0.9154	0.9	0.9267
0.4	0.9267	1	0.9183
0.5	0.9352		

Regarding runtime, the experiment shows that LPCS is quite efficient, abandon restarts of the ES it would be the fastest of all alternatives, however with an increasing risk to get stuck in a local optimum. Using restarts, only SP is still a bit faster on average, however, it completely fails in terms of predictive accuracy. Even though we did not investigate this issue in a systematic way so far, our experience has shown that LPCS scales much better than typical graph-based approaches. This is hardly surprising, since the dimensionality of the LPCS optimization problem is constant (six parameters have to be optimized) and does not depend on the number of data points. It is true that the size of the point clouds does have an influence on the evaluation of the objective function, which involves a nearest neighbor search for each point. The increase in runtime is at most quadratic, however. As opposed to this, the complexity of graph-based methods such as MGA, in which graph matching is reduced to a combinatorial optimization problem, grows exponentially with the number of nodes.

6 Conclusions

In this paper, we have introduced labeled point cloud superposition (LPCS) as a novel tool for structural bioinformatics, namely as a method for comparing biomolecules on a structural level. Besides, using fuzzy modeling techniques, we have defined a related similarity measure. The concept of a labeled point cloud appears to be a quite natural representation for biological structures, especially since it is closely leaned on existing database formats. In comparison to other approaches, such as the prevalent graph-based methods, the modeling is hence simplified and does not involve any complex transformations. More importantly, a labeled point cloud preserves the full geometric information and makes it easily accessible to computational procedures.

A labeled point cloud superposition is a spatial "alignment" of two point clouds which is optimal in the sense of a given scoring (similarity) function. As for related problems in bioinformatics, such as sequence alignment, the computation of the similarity between two objects hence involves the solution of an optimization problem. To this end, we have proposed the

ISBN: 978-989-95079-6-8

use of an evolution strategy, an approach from the family of evolutionary algorithms, which appears to be especially suitable for this problem.

First experimental results with classification data are quite promising and suggest that our approach is able to compare protein binding sites in a reasonable way. In terms of classification accuracy, LPCS turned out to be significantly better than existing (graph-based) methods used for comparison. Moreover, even though it is computationally more complex than these methods for small data sets, it scales much better and becomes more efficient for larger data sets. This is due to the fact that, in contrast to graph-based methods, the search space does not depend on the size of the point clouds and remains low-dimensional.

References

[1] Johann Gasteiger and Thomas Engel. *Chemoinformatics*. Wiley-Vch, Weinheim, 2003.

[2] F. Mémoli and G. Sapiro. Comparing point clouds. In *Eurographics / ACM SIGGRAPH symposium on Geometry processing*, pages 32–40, Nice, France, 2004.

[3] Francis R. Bach. Graph kernels between point clouds. In *International Conference on Machine Learning*, pages 25–32, Helsinki, Finland, 2008.

[4] K. M. Borgwardt. *Graph Kernels*. PhD thesis, Ludwig-Maximilians-Universität München, Germany, 2007.

[5] K. M. Borgwardt and H. P. Kriegel. Shortest-path kernels on graphs. In *International Conference on Data Mining*, pages 74–81, Houston, Texas, 2005.

[6] Thomas Gärtner. A survey of kernels for structured data. *SIGKKD Explorations*, 5(1):49 – 58, 2003.

[7] S. Schmitt, D. Kuhn, and G. Klebe. A new method to detect related function among proteins independent of sequence and fold homology. *Journal of Molecular Biology*, 323(2):387–406, 2002.

[8] M. Hendlich, F. Rippmann, and G. Barnickel. LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins. *Journal of Molecular Graphics and Modelling*, 15:359–363, 1997.

[9] D. Sinha and E.R. Dougherty. Fuzzification of set inclusion: theory and applications. *Fuzzy Sets and Systems*, 55(1):15–42, 1993.

[10] L.A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Comput. Math. Appl.*, 9:149–184, 1983.

[11] J. Fodor and R.R. Yager. Fuzzy set-theoretic operators and quantifiers. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy Sets*, pages 125–194. Kluwer Academic Publishers, Boston/London/Dordrecht, 2002.

[12] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. 18(1):183–190, 1988.

[13] Hans-Georg Beyer and Hans-Paul Schwefel. Evolution strategies: A comprehensive introduction. *Journal Natural Computing*, 1(1):3–52, 2002.

[14] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf. *Computational Geometry*. Springer, New York, 2000.

[15] N. Weskamp, E. Hüllermeier, D. Kuhn, and G. Klebe. Multiple graph alignment for the structural analysis of protein active sites. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(2):310–320, 2007.

[16] Thomas Bartz-Beielstein. *Experimental research in evolutionary computation: The new experimentalism*. Springer, 2006.

[17] Wolfgang Kabsch. A solution of the best rotation to relate two sets of vectors. *Acta Crystallographica*, 32:922–923, 1976.

ISBN: 978-989-95079-6-8

IFSA-EUSFLAT 2009

An Interval Approach for Fuzzy Linear Regression with Imprecise Data

Amory Bisserier[1] Reda Boukezzoula[1] Sylvie Galichet[1]

1 Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance (LISTIC)
Université de Savoie BP 80439 74941 Annecy le Vieux, Cedex France
Email: {amory.bisserier, reda.boukezzoula, sylvie.galichet}@univ-savoie.fr

Abstract *In this paper, a revisited approach for fuzzy regression linear model representation and identification is introduced. By adopting the commonly used principle of α-cuts, the fuzzy regression implementation is reduced to the handling of conventional intervals, for inputs, parameters and outputs. Using the Midpoint-Radius representation of intervals, the uncertainty attached to linear models becomes more interpretable. Actually, it is possible to determine the output uncertainty origin (model parameters and/or inputs). In this context, a possibilistic regression method is proposed to identify models of minimal global uncertainty, that is with respect to all possible inputs.*

Keywords Interval Regression, Fuzzy Regression, Uncertainty Representation, Fuzzy Inputs-Fuzzy Outputs.

1 Introduction

Fuzzy regression, a fuzzy type of conventional regression analysis, has been proposed to evaluate the functional relationship between input and output variables in a fuzzy environment. Indeed, unlike statistical regression modeling based on probability theory, fuzzy regression is based on possibility theory and fuzzy set theory [14].

In the fuzzy literature, the regression problem with fuzzy data has been previously treated from different points of view. According to [5], fuzzy regression techniques can be classified into two distinct areas. The first proposed by Tanaka ([10], [12]) which minimizes the total spread of the output is named possibilistic regression. In this case, the problem is viewed as finding fuzzy coefficients of a regression model according to a mathematical programming problem. The second approach developed by Diamond [4], which minimizes the total square error of the output is called the fuzzy least square method. In this paper a possibilistic approach is adopted.

In the chosen possibilistic context, different kinds of input/output data can be considered. Actually, the complete specification of regression problems highly depends on the nature of input-output data [5]. Some researches are thus devoted to Crisp-Input Crisp-Output (CICO) data [9] while others [6], [7], [8], [13] consider the regression problem using Fuzzy-Inputs and Fuzzy-Outputs (FIFO) data. Most commonly, a mixed approach with Crisp-Inputs and Fuzzy-Outputs (CIFO) is used [1], [2], [10], [11], [12].

In this framework, the fuzzy regression with CICO and CIFO data aroused a major interest. However, the FIFO model regression remains a little studied field. Indeed, only a few papers concerning this problem have been published ([6], [7], [8], [13]). For instance, in [6] and [13], different distances are proposed in order to extend the least square method to FIFO data. In [8], a multiobjective optimization technique for FIFO systems is proposed, whereas in [7] the optimization is based on the minimization of the deviation between observations and predictions.

All above mentioned methods lead to interesting results but present a lack of representativity and an awkward illustration of the uncertainties in the regression model. In order to overcome this problem, an extension of the regression methodology initially developed for CIFO systems in [1], [2] to FIFO ones is proposed. The key idea of the proposed approach resides in using a Midpoint-Radius regression model representation. By doing so, the regression problem can be viewed as a conventional linear optimization problem where a new criterion is proposed.

From a practical point of view and as commonly used in fuzzy regression literature ([5], [7], [8], [12]), the a-cut principle is adopted. In this case, for a specified a-cut, the fuzzy interval becomes a conventional interval, which states that a fuzzy interval representation is a generalization of a conventional one. In order to take into account the maximum of uncertainty in the system (the pessimistic case), an a-cut equal to 0 (0-cut) is considered. In this case, the fuzzy regression problem can be reduced to an interval regression one. That is the approach adopted in this paper where a new regression methodology is proposed.

The structure of this paper is as follows. In section 2, relevant concepts and notations are introduced. Section 3 is devoted to fuzzy regressive models representation. An identification method of such models based on a linear problem optimization is proposed in section 4. An application is shown in section 5. Finally, conclusion and perspectives are presented in section 6.

2 Relevant Concepts and Notations

For the sake of rigor and clarity, let us define the basic notions and notations used in this paper.
A fuzzy trapezoidal interval A is a particular case of a fuzzy number where a possibility distribution, represented by a membership function μ_A, is associated with the conventional intervals of the support and the kernel (see Figure 1).

ISBN: 978-989-95079-6-8

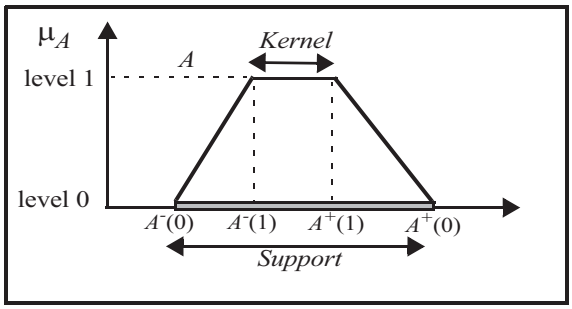


Figure 1: Trapezoidal fuzzy interval representation

In this case, for a given α-cut (α∈[0, 1]) on the fuzzy interval A, a conventional interval is obtained:

$$A(\alpha) = [A^{-}(\alpha), A^{+}(\alpha)] \tag{1}$$

Using the α-cut representation, a symmetrical fuzzy interval is viewed as a weighted family of nested intervals. In the sequel, in order to consider the maximum uncertainty, only α-cuts at level 0 will be considered. Consequently, all computations are performed with conventional intervals.

In this framework, an interval A is defined by the set of elements lying between its lower and upper limits (endpoints) as:

$$A = [A^{-},\ A^{+}] = \{x \mid A^{-} \le x \le A^{+}, x \in \Re\} \tag{2}$$

The interval A is called thick if $A^{-} < A^{+}$ and thin (or point) if $A^{-} = A^{+}$.

In order to manipulate and exhibit the uncertainty with intervals, an efficient representation is needed. Indeed, the latter should be able to represent arbitrary intervals and their sets in a single and uniform graphical way. In this framework, it is obvious that to uniquely define an interval, two parameters are necessary, hence the space of intervals is two-dimensional. There are many possible representations of an interval. The most used one is the endpoints representation given in equation (2). However, the main drawback of this representation resides in the fact that the uncertainty (the Radius) is awkward and not directly exhibited in the interval. In order to overcome this difficulty, another representation is adopted. Indeed, in order to facilitate the uncertainty representation, Midpoint (M) and Radius (R) coordinates are used (see Figure 2). In this case, an interval is represented as a point in the space (M, R).

For an interval A, its Midpoint and its Radius are defined by:

$$M_A = (A^{-} + A^{+})/2 \ \text{ and: } \ R_A = (A^{+} - A^{-})/2 \tag{3}$$

Moreover, the set of all the intervals contained in the interval A can be represented by a domain $C(A)$. As shown in Figure 2, the domain $C(A)$ is represented by the upper and lower bound lines of A, forming a triangle.

In the (M, R) representation, it is clear that the most imprecise element of $C(A)$ is the interval A itself which corresponds to the highest value of Radius, i.e. the top of the triangle. On the opposite, the triangle base represents all the crisp values between the lower and the upper bounds (the set of all the thin intervals).

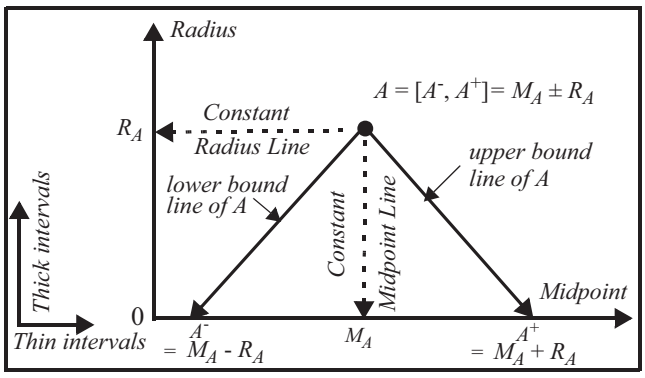


Figure 2: Midpoint-Radius representation of A and $C(A)$

For two intervals $A = [A^{-}, A^{+}]$ and $B = [B^{-}, B^{+}]$, the addition and multiplication operations are defined as follows:

$$\begin{cases} A \oplus B = [A^{-} + B^{-}, A^{+} + B^{+}] \\ A \otimes B = [min(Z), max(Z)] \end{cases} \tag{4}$$

where:

$$Z = \{A^{-}B^{-}, A^{-}B^{+}, A^{+}B^{-}, A^{+}B^{+}\} \tag{5}$$

In this case, it can be stated that the Midpoint and Radius of these operations are given by:

$$\begin{cases} M_{A \oplus B} = M_A \oplus M_B \\ R_{A \oplus B} = R_A \oplus R_B \end{cases} \tag{6}$$

and:

$$\begin{cases} M_{A \otimes B} = M_A M_B + \varphi_{A \otimes B} \cdot sign(M_A M_B) \\ R_{A \otimes B} = R_A R_B + |M_A| R_B + R_A |M_B| - \varphi_{A \otimes B} \end{cases} \tag{7}$$

where:

$$\varphi_{A \otimes B} = min\{|M_A| R_B, R_A |M_B|, R_A R_B\} \tag{8}$$

For two intervals [3] A and B, an inclusion relation of A in B (see Figure 3) is defined as follows:

$$A \subseteq B \Leftrightarrow \begin{cases} B^{-} \le A^{-} \\ A^{+} \le B^{+} \end{cases} \Leftrightarrow \begin{cases} M_B - R_B \le M_A - R_A \\ M_A + R_A \le M_B + R_B \end{cases} \tag{9}$$

$$\Leftrightarrow \begin{cases} M_B - M_A \le R_B - R_A \\ M_A - M_B \le R_B - R_A \end{cases} \tag{10}$$

From equation (10) it follows:

$$A \subseteq B \Leftrightarrow |M_B - M_A| \le R_B - R_A \tag{11}$$

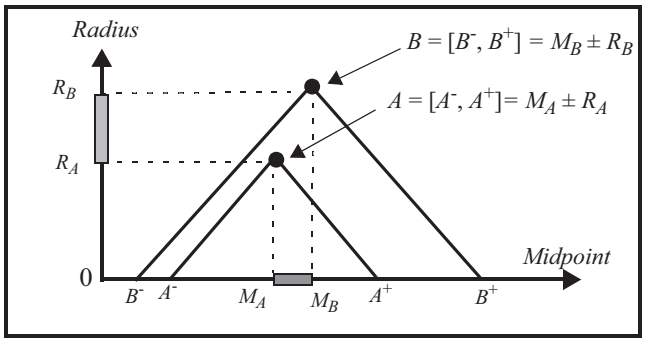


Figure 3: Inclusion of two intervals ($A \subseteq B$)

3 Fuzzy Regressive Model Representation

ISBN: 978-989-95079-6-8

The fuzzy linear regression model is the most frequently used form in regression analysis for expressing the relationship between one or more explanatory variables and response. For the sake of simplicity and as commonly assumed in the fuzzy regression literature, the case of simple linear regression model involving a single independent variable is considered. The case of multiple inputs is a straightforward generalization of this methodology but more complex in a computational point of view.

An interval regression model is mathematically expressed in the following form:

$$Y(X) = A_0 \oplus A_1 \otimes X \tag{12}$$

where Y is the interval output (dependant variable), A_0 and A_1 the interval parameters, i.e. unknowns to be estimated from observed data. According to (3), the parameters A_0 and A_1 are denoted as follows:

$$\begin{aligned} A_0 &= [A_0^-, A_0^+] = [M_{A_0} - R_{A_0}, M_{A_0} + R_{A_0}] \\ A_1 &= [A_1^-, A_1^+] = [M_{A_1} - R_{A_1}, M_{A_1} + R_{A_1}] \end{aligned} \tag{13}$$

The input X can take any value in a definition domain to be chosen.

As mentioned in the introduction, the regression problem has been previously treated from different points of view and by considering different kinds of input/output data. Indeed, according to the nature of X (crisp or imprecise) two regression problems can been distinguished. The first one deals with Crisp-Input Imprecise-Output data ([12], [10], [1], [2]). In this case, as the model is linear and the inputs are considered as crisp values, the model output (12) will also be an interval. So, for a crisp input $X = X_*$, equation (12) becomes:

$$Y(X_*) = [Y_*^-, Y_*^+] = A_0 \oplus A_1 \cdot X_* \tag{14}$$

In this case, the definition domain is the interval $D = [inf(D), sup(D)]$, which means that the input X can take any crisp value in D.

For the sake of illustration simplicity, in the sequel, the particular case when the parameter A_1 and the input are positive is considered. The other cases are a straightforward generalization of this particular case. The evolution of the model output (14) is illustrated in Figure 4. Actually, as the input X is a crisp value of D, the model output uncertainty is only due to the uncertainty of the two parameters A_0 and A_1.

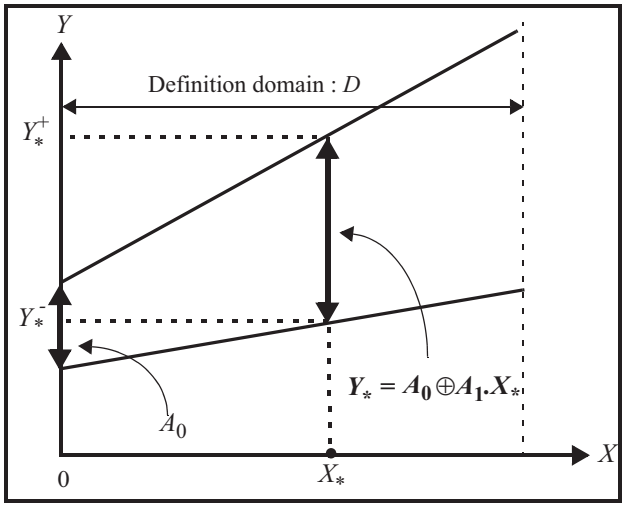


Figure 4: Interval model representation with a crisp input

The second regression problem is devoted to systems with Imprecise-Inputs and Imprecise-Outputs. For an interval input $X = X_* = [X_*^-, X_*^+]$, the interval model output is given by the following equation with Z defined according to (5):

$$\begin{aligned} Y(X_*) = [Y_*^-, Y_*^+] &= [A_0^- + min(Z), A_0^+ + max(Z)] \\ &= [A_0^- + A_1^- \cdot X_*^-, A_0^+ + A_1^+ \cdot X_*^+] \end{aligned} \tag{15}$$

In this case, the input X can be any interval included in the interval D previously defined for crisp inputs. Consequently, the model definition domain is now $\Delta = C(D)$ (see Figure 2). The model output evolution is presented in Figure 5.

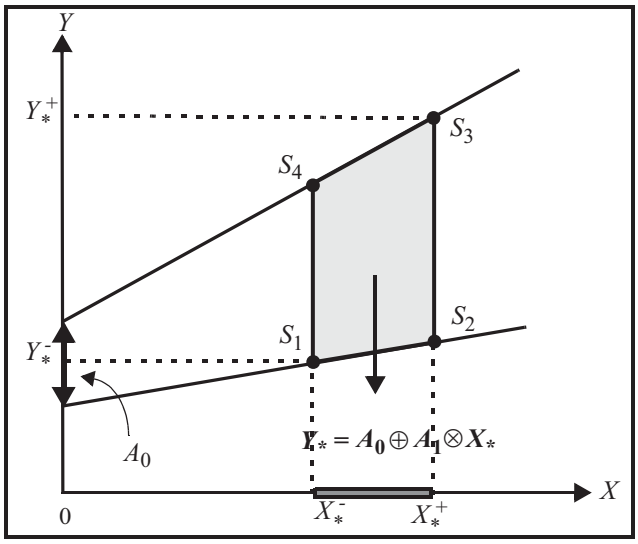


Figure 5: Conventional interval model representation

Several main drawbacks can be underlined:

- The interval input X_* is represented by its endpoints coordinates on a real line with only one dimension. It leads to a lack of information about the input uncertainty (Radius). In this case, neither the definition domain Δ nor the output uncertainty on it can be exhibited in a conventional 2D space.
- According to Figure 5, the model output is defined by the 4-sided area delimited by the four vertices (S_1, S_2, S_3, S_4). If the vertices S_1 and S_3 exhibit the model output bounds, the others (S_2 and S_4) have no interpretability in the model output. In this situation, two intervals are defined to represent the model output, which limits the interpretability of this kind of representation ([7], [8]).

In order to deal with these drawbacks, the Midpoint-Radius representation is used to define all handled intervals. In this context, the output Y is expressed by:

$$\begin{aligned} Y(X_*) &= [Y_*^-, Y_*^+] \\ &= [M_{Y(X_*)} - R_{Y(X_*)}, M_{Y(X_*)} + R_{Y(X_*)}] \end{aligned} \tag{16}$$

where:

$$\begin{cases} M_{Y(X_*)} = M_{A_0} + M_{A_1} \otimes X_* \\ R_{Y(X_*)} = R_{A_0} + R_{A_1} \otimes X_* \end{cases} \tag{17}$$

For positive parameter A_1 and input X_*, we have:

$$M_{A_1 \otimes X_*} = M_{A_1} M_{X_*} + \varphi_{A_1 \otimes X_*} \cdot sign(M_{A_1} M_{X_*}) \tag{18}$$

As:

ISBN: 978-989-95079-6-8

$$\varphi_{A_1 \otimes X_*} = min\{M_{A_1}R_{X_*}, R_{A_1}M_{X_*}, R_{A_1}R_{X_*}\} = R_{A_1}R_{X_*}, \quad (19)$$

the equation (18) becomes:

$$M_{A_1 \otimes X_*} = M_{A_1}M_{X_*} + R_{A_1}R_{X_*} \quad (20)$$

By adopting the same principle, the following Radius expression is determined:

$$R_{A_1 \otimes X_*} = R_{A_1}R_{X_*} + M_{A_1}R_{X_*} + R_{A_1}M_{X_*} - R_{A_1}R_{X_*} = M_{A_1}R_{X_*} + R_{A_1}M_{X_*} \quad (21)$$

By substitution of equations (20) and (21) in (17), it follows:

$$\begin{cases} M_{Y(X_*)} = M_{A_0} + M_{A_1}M_{X_*} + R_{A_1}R_{X_*} \\ R_{Y(X_*)} = R_{A_0} + M_{A_1}R_{X_*} + R_{A_1}M_{X_*} \end{cases} \quad (22)$$

The Midpoint and Radius output evolutions on the domain Δ are illustrated in Figure 6 and Figure 7.

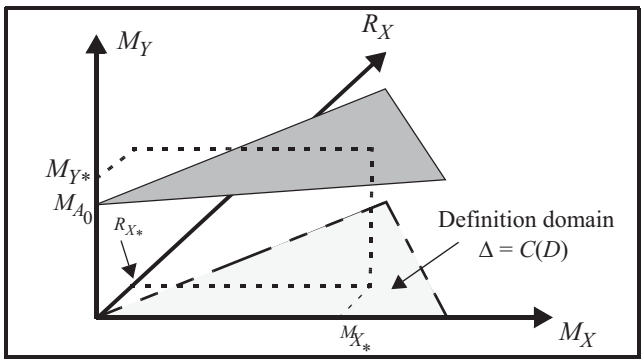


Figure 6: The midpoint output evolution

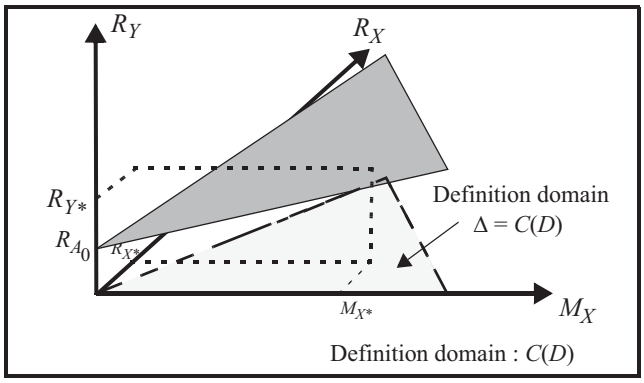


Figure 7: The radius output evolution.

From equations (16) and (22), it is obvious that:

$$\begin{cases} Y_*^- = M_{A_0} - R_{A_0} + (M_{A_1} - R_{A_1})(M_{X_*} - R_{X_*}) \\ Y_*^+ = M_{A_0} + R_{A_0} + (M_{A_1} + R_{A_1})(M_{X_*} + R_{X_*}) \end{cases} \quad (23)$$

A global representation of the model output evolution on the domain Δ is illustrated in Figure 8.

This Midpoint-Radius representation presents several advantages:

- In this representation, only one interval output is defined for a considered input. Indeed, the output expression represents as well the contribution of the input Midpoint and Radius.
- According to this representation, the origin of uncertainty in the output can be precisely determined. In other words, the contributions of the parameters A_0, A_1 and the input X are well defined in the model output (22).

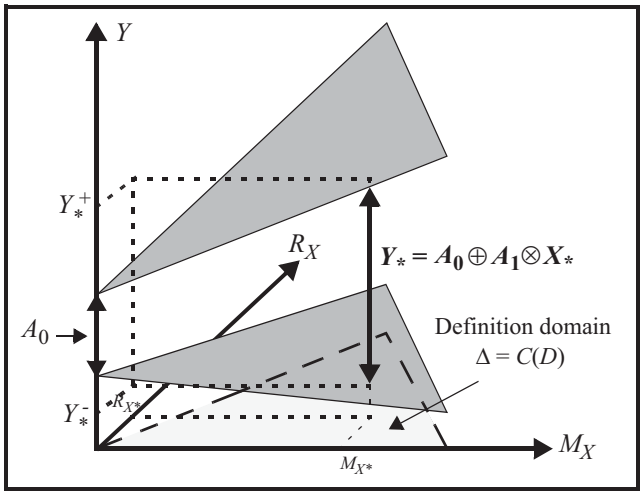


Figure 8: The model output evolution

4 Interval Regressive Model Identification

In the context of regression model identification, two main problems come up, namely, the model structure specification and the estimation of the given model. The first problem is focused on the choice of a suitable model structure for a data set. This problem is traditionally addressed *a priori.* In this paper, as commonly used, the model structure is assumed to be linear. In this case, the regression problem is reduced to an estimation problem of the model parameters.

Let us consider a set of N observed data samples. Let the j^{th} sample be represented by the couple $(X_j, Y_j), j = 1,..., N$ where X_j are interval inputs and the Y_j are the corresponding interval output.

All the observed input intervals are assumed to be included in the interval $D = [D^-, D^+]$ given by:

$$D^- = min(X_j^-) \text{ and } D^+ = max(X_j^+) \quad (24)$$

Like any regression technique, the fuzzy regression objective is to determine a predicted functional relationship between inputs X and outputs Y:

$$\hat{Y}(X) = A_0 \oplus A_1 \otimes X \quad (25)$$

In this case, the definition domain of the identified model is the set of all the intervals included in D, i.e. $\Delta = C(D)$.

As discussed in the paper introduction, a possibilistic regression approach is adopted where the objective is to determine the model parameters such that the observed data are included in the predicted ones (possibility model). In this framework, another kind of model (necessity one) could be identified [11].

In the proposed identification methodology, two points have to be considered for its implementation:

- the interval inclusion constraints to be introduced in the optimization problem,
- the identification criterion to be optimized.

4.1 The optimization constraints

In this case, the inclusion constraints are given by:

ISBN: 978-989-95079-6-8

$$Y_j \subseteq \hat{Y}_j \Leftrightarrow \left|M_{\hat{Y}_j} - M_{Y_j}\right| \leq R_{\hat{Y}_j} - R_{Y_j} \quad (26)$$

with:

$$M_{\hat{Y}_j} = M_{A_0} + M_{A_1} \otimes X_j \text{ and: } R_{\hat{Y}_j} = R_{A_0} + R_{A_1} \otimes X_j \quad (27)$$

By substitution of (27) in (26), the constraints inclusion can be written as:

$$\begin{aligned} &\left|M_{A_0} + M_{A_1} M_{X_j} + R_{A_1} R_{X_j} - M_{Y_j}\right| \leq \\ &R_{A_0} + M_{A_1} R_{X_j} + R_{A_1} M_{X_j} - R_{Y_j} \end{aligned} \quad (28)$$

Moreover, the optimization constraints must ensure the identification of intervals, whose bounds are ordered. Moreover, the sign of the parameter A_1 must be constrained as well. In the case of a model whose parameter A_1 is identified as positive, it follows:

$$R_{A_0} \geq 0, \text{ and } : M_{A_1} \geq R_{A_1} \geq 0 \quad (29)$$

It is important to note that all the considered constraints in the identification problem are linear according to the identification variables (Midpoint and Radius of A_0 and A_1).

4.2 The used criterion

The objective is to identify the less imprecise model on its definition domain, as it has been proposed in our previous works [1]. The computation of the global uncertainty on the whole domain D proposed in [1] is here extended to the whole domain Δ, whose elements are intervals. It follows that the criterion J to be optimized, i.e. the global uncertainty on Δ, is expressed as::

$$\begin{aligned} J &= \iint_{\Delta} (\hat{Y}^{+} - \hat{Y}^{-}) dM_X dR_X = 2 \iint_{\Delta} R_{\hat{Y}} dM_X dR_X \\ &= 2 \iint_{\Delta} R_{A_0} + M_{A_1} R_X + R_{A_1} M_X) dM_X dR_X \end{aligned} \quad (30)$$

The substitution of equation (23) in (30) yields to:

$$J = 2R_{A_0} \cdot R_D^2 + 2R_{A_1} \cdot R_D^2 \cdot M_D + \frac{2}{3} M_{A_1} \cdot R_D^3 \quad (31)$$

To sum up, the identification method is performed by minimizing the criterion (31) under the constraints (28) and (29).

At this stage, several points can be discussed about the proposed approach:

- As the interval D is defined with the observed inputs, its Radius and Midpoint are known numerical values. So, the criterion J is linear according to the optimization variables. Moreover, J is independent of the observed inputs, i.e. the optimization is made on the whole domain Δ of the model, not only on the observed data. This point contributes to improve the model robustness.
- It is obvious that the criterion must be optimized under constraints in order to ensure inclusion property. If the constraints are released, the identified model corresponds to the minimal possible uncertainty, achieved with null parameters, and has no sense.
- As the minimization of the model uncertainty is performed on Δ, the input uncertainty is considered in the model output expression in the identification process.

5 Illustrative example

In this part, the proposed identification method is applied on the illustrative example given in [8], [7]. Given the data set [8], [7] illustrated in Table 1, corresponding of α-cuts of level 0 of proposed fuzzy data set, the objective is to identify the model parameters A_0 and A_1.

Table 1: Observed data set and predicted one

inputs X_j	observed outputs Y_j	predicted outputs $\hat{Y}_j$
[1.5 ; 2.5]	[3.5 ; 4.5]	[3.5 ; 6.5]
[3 ; 4]	[5 ; 6]	[4.16 ; 7.25]
[4.5 ; 6.5]	[6.5 ; 8.5]	[4.82 ; 8.5]
[6.5 ; 7.5]	[6 ; 7]	[5.69 ; 9]
[8 ; 9]	[8 ; 9]	[6.35 ; 9.75]
[9.5 ; 11.5]	[7 ; 9]	[7 ; 11]
[10.5 ; 11.5]	[10 ; 11]	[7.45 ; 11]
[12 ; 13]	[9 ; 10]	[8.1 ; 11.75]

The considered definition domain is D = [1.5, 13] and represented in Figure 9. This domain corresponds to a set of positive inputs, and, as discussed previously, the parameter A_1 is assumed to be positive too.

The identified parameters corresponding to the less imprecise model according to the criterion J defined in (31) are:

$$A_0 = [2.844 ; 5.25] \text{ and } A_1 = [0.438 ; 0.5] \quad (32)$$

In this case, the optimal value of uncertainty obtained is $J = 153.95$.

The identified model output representation is given in Figure 9, whereas the predicted outputs are presented in Table 1.

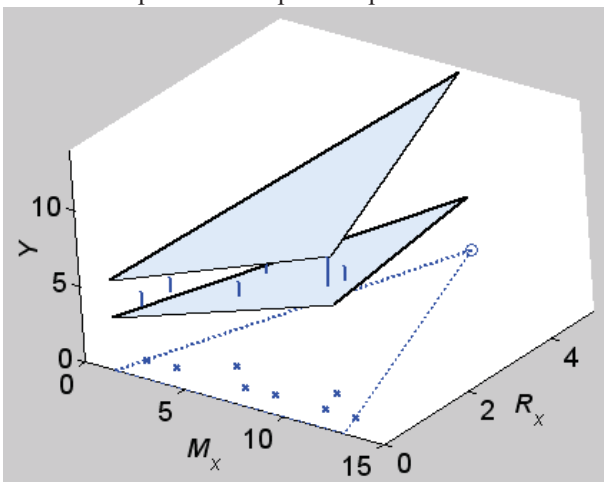


Figure 9: Identified model representation.

As presented in Table 1, the main advantage of the proposed approach compared to those presented in [7] and [8] is that only one interval output corresponds to a considered input. In Figure 9, it can be stated that the desired inclusion is respected.

Another advantage of the representation using the (Midpoint, Radius) space for the interval inputs concerns the exploitation

ISBN: 978-989-95079-6-8

of the model characteristics. According to (22), the output Radius evolution on Δ can be represented (see Figure 10). On the axe (M_X, $R_X = 0$), i.e. the triangle basis, only the parameter uncertainty is exhibited, whereas at the top of the triangle, the output uncertainty is maximum, corresponding to the most uncertain input in Δ, i.e. D. In this example (Figure 10), the output radius is nearly constant on the basis of the triangle, corresponding to the case of a nearly crisp parameter A_1. The output radius increasing on Δ is essentially due to the input uncertainty.

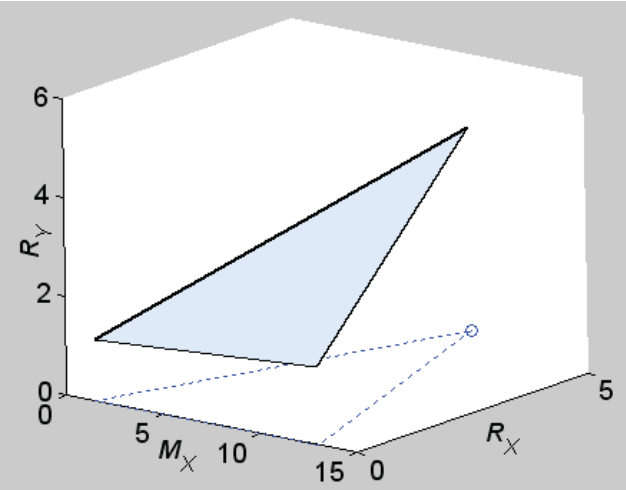


Figure 10: Identified model output Radius evolution.

The same analysis can be performed on the output Midpoint evolution on Δ, corresponding to the global tendency of the output on the domain.

In [8], the identified model, whose output respects the inclusion statement, corresponds to a criterion value $J = 281.03$, according to (31). Its parameters are the following:

$$A_0 = [1.25 ; 6.75] \text{and } A_1 = [0.4 ; 0.56] \quad (33)$$

In fact, the vertical dimension α is introduced into the proposed multiobjective optimization problem. It is known that the higher the α value is set for optimization, the more uncertain the identified model is and inclusion is respected for any lower α levels. So, this compromise between higher level of inclusion and lower uncertainty leads to the identification of a model whose uncertainty is not minimal on the domain.

On the other hand, the proposed method in [7], which minimizes deviations between observations and predictions without constraining inclusion, leads to the identification of a model which corresponds to a criterion value $J = 102.6$, and whose parameters are:

$$A_0 = [3 ; 3.82] \text{and } A_1 = [0.5 ; 0.54] \quad (34)$$

It is obvious that guaranteeing the inclusion increase the model uncertainty.

6 Conclusion

In this paper, a revisited representation of interval regressive model is proposed and used in an identification problem.

By adopting the Midpoint-Radius formalism for intervals, uncertainty representation is improved. Applied to regressive interval models, this formalism allows a better handling of the input uncertainty in the definition domain establishment and in the model output definition. In an identification problem, based on the optimization of a linear expression of the model output global uncertainty on its definition domain, the desired inclusion is ensured.

Further work will focus on the extension of this approach to fuzzy linear systems, by considering the vertical dimension α. In the proposed space, a symmetrical trapezoidal fuzzy interval can be interpreted by a zooming operation as illustrated in the following Figure 11:

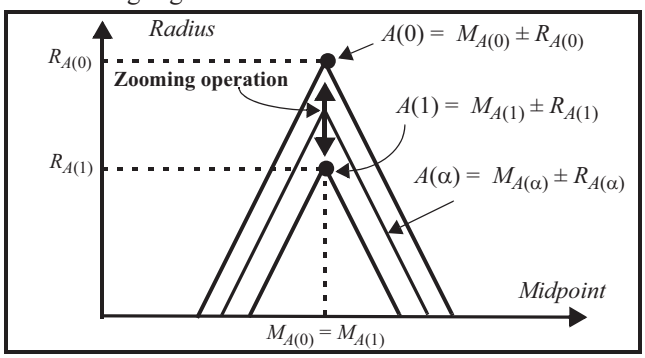


Figure 11: Midpoint-Radius representation of symmetrical trapezoidal fuzzy interval α-cuts.

A particular attention should be paid to the loss of linearity of the membership function of the model output, due to the fuzzy product between the input X and the parameter A_1.

References

[1] A. Bisserier, S. Galichet, R. Boukezzoula, Fuzzy Piecewise Linear Regression, *5th IEEE World Congress on Computational Intelligence (WCCI 2008 / FUZZ-IEEE 2008)*, Hong-Kong, China, 2008, 2089-2094.

[2] A. Bisserier, R. Boukezzoula, S. Galichet, Linear fuzzy regression using trapezoidal fuzzy intervals, *Information Processing and Management of Uncertainty (IPMU2008)*, Malaga, Spain, 2008, 181-188

[3] R.Boukezzoula, S. Galichet and L. Foulloy, MIN and MAX Operators for Fuzzy Intervals and their Potential Use in Aggregation Operators, *IEEE Trans. on Fuzzy Systems* 15(6): 1135-1144, 2007

[4] P. Diamond, Fuzzy Least Squares, *Information Sciences* 46: 141-157, 1988.

[5] P. Diamond , H. Tanaka, Fuzzy regression analysis, *Fuzzy sets in decision analysis, operations research and statistics*, Kluwer Academic Publishers, Norwell, MA, 1999.

[6] P. D'Urso, Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data, *Computational Statistics & Data Analysis* 42 (1-2), 2003, 47-72.

[7] M. Hojati, C.R. Bector, K. Smimou, A simple method for computation of fuzzy linear regression, *European Journal of Operational Research* 166: 172-184, 2005.

[8] M. Sakawa, H. Yano, Multiobjective fuzzy linear regression analysis for fuzzy input-output data, *Fuzzy Sets and Systems* 47: 173-181, 1992.

[9] D.A. Savic, W. Pedrycz, Evaluation of fuzzy linear regression models, *Fuzzy Sets and Systems* 39, 1991, 51-63.

[10] H.Tanaka, I. Hayashi, J. Watada, Possibilistic linear regression analysis for fuzzy data, *European Journal of Operational Research* 40: 389-396, 1989.

[11] H. Tanaka, H. Lee, Interval regression analysis by quadratic programming approach, *IEEE Trans. on Fuzzy Systems* 6(4): 473-481, 1998.

[12] H.Tanaka, S. Uejima, K. Asai, Linear regression analysis with fuzzy model, *IEEE, Systems, Trans. Systems Man Cybernet*. SMC-2: 903-907, 1982.

[13] H.-C. Wu, Fuzzy estimates of regression parameters in linear regression models for imprecise input and output data, *Computational Statistics & Data Analysis* 42 (1-2), 2003, 203-217.

[14] L.A. Zadeh, Fuzzy Sets, *Information and Control* 8: 338-353, 1965.

ISBN: 978-989-95079-6-8

A Predator-Prey Model with Fuzzy Initial Populations

Muhammad Zaini Ahmad Bernard De Baets

Department of Applied Mathematics, Biometrics and Process Control, Ghent University
Coupure links 653, B-9000 Ghent, Belgium
Email: mzaini_ahmad@yahoo.com, Bernard.DeBaets@ugent.be

Abstract— *The aim of this paper is to study a predator-prey population model which takes into account the uncertainty that arises when determining the initial populations of predator and prey . This model is solved numerically by means of a 4^{th}-order Runge-Kutta method. Simulations are made and a graphical representation is also provided to show the evolution of both populations over time. In addition to that, a new phase-plane in this fuzzy setting, referred to as fuzzy phase-plane, is introduced. The stability of the equilibrium points is also described. Finally, this paper points out a new research direction on fuzzy dynamical systems, especially in non-linear cases.*

Keywords— Equilibrium points, Fuzzy predator-prey population model, Fuzzy phase-plane, Fuzzy stability, Uncertainty.

1 Introduction

Nowadays, differential equations have proved a successful modelling paradigm. In a wide range of disciplines, the behaviour of an idealised version of a problem under study has been adequately described by one or more ordinary differential equations. Real-world problems, however, are pervaded with uncertainty. According to Diniz *et al.* [1], the uncertainty can arise in the experimental part, the data collection, the measurement process, as well as when determining the initial conditions. These are patently obvious when dealing with "living" material, such as soil, water, microbial populations, etc. In order to handle these problems, the use of fuzzy sets may be seen as an effective tool for a better understanding of the studied phenomena. It is therefore not surprising that there is a vast literature dealing with fuzzy differential equations (FDEs) for modelling this kind of perception. In the study of differential equations in a fuzzy environment, the term fuzzy differential equations is used for referring to differential equations with fuzzy coefficients, differential equations with fuzzy initial values or fuzzy boundary values, or even differential equations dealing with functions on the space of fuzzy intervals (see [2, 3, 4, 5, 6, 7, 8]).

In population dynamics, Bassanezi *et al.* [9] have started using FDEs to study the stability of fuzzy dynamical systems. The variables and/or the parameters in the systems under consideration were supposed to be uncertain and modelled by fuzzy sets. The authors have suggested a new definition of fixed point: if f is a continuous function, then u is a fixed point if $f([u]^\alpha) = [u]^\alpha$ for $\alpha \in [0,1]$, where $[u]^\alpha = [u_1^\alpha, u_2^\alpha] = [\min_{x\in[u_1^\alpha,u_2^\alpha]} f(x), \max_{x\in[u_1^\alpha,u_2^\alpha]} f(x)]$ (see [9] for more details). According to Ortega *et al.* [10], the proposed idea is not easy to implement in practice because several details should be treated carefully. Guo *et al.* [11] have adopted Hüllermeier's approach [7] to establish fuzzy impulsive functional differential equations and some results were applied in the logistic model and the Gompertz model. In [12], the author has established existence and uniqueness results for fuzzy functional differential equations. These results were applied in population models whose solutions define fuzzy intervals. Recently, Peixoto *et al.* [13] have studied the fuzzy predator-prey population model. The authors have elaborated the classical deterministic model by means of a fuzzy rule-based system. The fuzzy rule-based system considered in their work consists of two inputs and two outputs. The fuzzy rule base is given by 30 rules of the type: "If the number of preys is large AND the potential of predation is very small, THEN the variation of preys increase a little AND the variation of the potential of predation increase a lot". The words such as "large", "very small", "increase a little" and "increase a lot" are modelled by fuzzy sets. This might be a possible way to treat the model in a fuzzy setting. In this paper, we propose a new way of studying the fuzzy predator-prey population model. For this purpose, we only consider the initial populations of predator and prey to be fuzzy intervals.

2 Preliminary Concepts

In this section, let us first introduce the notations and some theoretical background we will be using throughout this paper. Denote $\mathcal{F}(\mathbb{R}) = \{U \mid U : \mathbb{R} \to [0,1]\}$, where U satisfies the following conditions:

1. U is normal: there exists $x_0 \in \mathbb{R}$ such that $U(x_0) = 1$;
2. U is convex: for $x, y \in \mathbb{R}$ and $0 \leq \lambda \leq 1$, one has
$$U(\lambda x + (1-\lambda)y) \geq \min\left(U(x), U(y)\right);$$
3. U is upper semi-continuous: for any $x_0 \in \mathbb{R}$, one has
$$U(x_0) \geq \lim_{x \to x_0^{\pm}} U(x);$$
4. $[U]^0 = \overline{\{x \in \mathbb{R} \mid U(x) > 0\}}$ is a compact subset of $\mathbb{R}$.

If U satisfies (1)–(4), then it is called a fuzzy interval. We can define its α-cuts as follows:

$$[U]^\alpha = \{x \in \mathbb{R} \mid U(x) \geq \alpha\}, \;\; 0 < \alpha \leq 1$$

For a fuzzy interval U, its α-cuts are closed intervals in $\mathbb{R}$. Let us denote them by $[U]^\alpha = [u_1^\alpha, u_2^\alpha]$.

In what follows, we introduce some results we adopt from [14] concerning the notion of interactivity of fuzzy intervals.

ISBN: 978-989-95079-6-8

Definition 1 *Two fuzzy intervals A and B are said to be interactive fuzzy intervals if there exist $q, r \in \mathbb{R}$, $q \neq 0$, such that their joint possibility distribution is given by*

$$\begin{aligned} C(x_1, x_2) &= A(x_1) \cdot \chi_{\{qx_1+r=x_2\}}(x_1, x_2) \\ &= B(x_2) \cdot \chi_{\{qx_1+r=x_2\}}(x_1, x_2) \end{aligned} \tag{1}$$

where $\chi_{\{qx_1+r=x_2\}}(x_1, x_2)$ stands for the characteristic function of the line

$$\left\{(x_1, x_2) \in \mathbb{R}^2 \mid qx_1 + r = x_2\right\}.$$

In this case we have

$$[C]^\alpha = \left\{(x, qx + r) \in \mathbb{R}^2 \mid x = (1-t)a_1^\alpha + ta_2^\alpha, t \in [0,1]\right\}$$

where $[A]^\alpha = [a_1^\alpha, a_2^\alpha]$ and $[B]^\alpha = q[A]^\alpha + r$ for any $\alpha \in [0,1]$, and, finally

$$B(x) = A\left(\frac{x-r}{q}\right)$$

for all $x \in \mathbb{R}$.

Definition 2 *Two fuzzy intervals A and B are said to be completely positively interactive if q is positive in (1).*

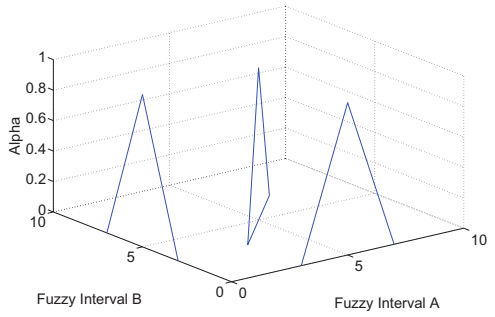


Figure 1: Completely positively interactive fuzzy intervals.

Definition 3 *Let C be a joint possibility distribution with marginal possibility distributions A and B, and let $f : \mathbb{R}^2 \to \mathbb{R}$ be a continuous function. If A and B are interactive fuzzy intervals, then the extension of f via C is defined as*

$$f(A, B)(y) = \sup_{y=f(x_1,x_2)} C(x_1, x_2), \quad \forall y \in \mathbb{R}. \tag{2}$$

Remark 1 *If A and B are non-interactive fuzzy intervals, that is, their joint possibility distribution is defined by*

$$C(x_1, x_2) = \min\left(A(x_1), B(x_2)\right),$$

then (2) turns into the extension principle introduced by Zadeh [15].

Definition 4 *If A and B are two fuzzy intervals, then the distance D between A and B is defined as*

$$D(A, B) = \sup_{\alpha \in [0,1]} d_H([A]^\alpha, [B]^\alpha)$$

where

$$d_H([A]^\alpha, [B]^\alpha) =$$

$$\max\left\{\sup_{a \in [A]^\alpha} \inf_{b \in [B]^\alpha} d(a,b), \sup_{b \in [B]^\alpha} \inf_{a \in [A]^\alpha} d(a,b)\right\}$$

3 Two-Dimensional Autonomous Systems

In this section, we first consider a general two-dimensional autonomous system with crisp initial values:

$$\begin{cases} x'(t) = f(x,y), & x(t_0) = x_0 \\ y'(t) = g(x,y), & y(t_0) = y_0 \end{cases} \tag{3}$$

where $f, g : \mathbb{R}^2 \to \mathbb{R}$ are continuous functions and $x_0, y_0 \in \mathbb{R}$. This model is called an autonomous system because the functions appearing on the right-hand side of (3) do not depend on the independent variable t. Suppose that the initial values in (3) are uncertain and modelled by fuzzy intervals, then we have the following fuzzy autonomous system [16]:

$$\begin{cases} x'(t) = f(x,y), & x(t_0) = X_0 \\ y'(t) = g(x,y), & y(t_0) = Y_0 \end{cases} \tag{4}$$

where $f, g : \mathbb{R}^2 \to \mathbb{R}$ are continuous functions and $X_0, Y_0 \in \mathcal{F}(\mathbb{R})$ with $[X_0]^\alpha = [x_{01}^\alpha, x_{02}^\alpha]$ and $[Y_0]^\alpha = [qx_{01}^\alpha + r, qx_{02}^\alpha + r]$ for $\alpha \in [0,1]$, $q \in \mathbb{R}^+$ and $r \in \mathbb{R}$. Here we assume that X_0 and Y_0 are completely positively interactive fuzzy intervals.

A solution to (4) can be approximated by any numerical method existing in the literature. For any fixed value of t, the solution is represented as a joint possibility distribution in the xy-plane. As t varies, the solution traces out the joint possibility distributions in the plane. By plotting a collection of such joint possibility distributions, we gain graphical insight into the behaviour of solutions. In this paper, the set of joint possibility distributions and the xy-plane are called fuzzy trajectories and the fuzzy phase-plane, respectively.

4 Numerical Methods

In many cases, model (4) presented in Section 3 cannot be solved analytically and therefore we need a numerical scheme to approximate the exact solution. We consider the following 4^{th}-order Runge-Kutta method:

$$x_{i+1} = x_i + \frac{h}{6}\left[K_1 + 2K_2 + 2K_3 + K_4\right] \tag{5}$$

$$y_{i+1} = y_i + \frac{h}{6}\left[L_1 + 2L_2 + 2L_3 + L_4\right] \tag{6}$$

where

$$\begin{aligned} K_1 &= f\left(x_i, y_i\right) \\ L_1 &= g\left(x_i, y_i\right) \\ K_2 &= f\left(x_i + \frac{h}{2}K_1, y_i + \frac{h}{2}L_1\right) \\ L_2 &= g\left(x_i + \frac{h}{2}K_1, y_i + \frac{h}{2}L_1\right) \\ K_3 &= f\left(x_i + \frac{h}{2}K_2, y_i + \frac{h}{2}L_2\right) \\ L_3 &= g\left(x_i + \frac{h}{2}K_2, y_i + \frac{h}{2}L_2\right) \\ K_4 &= f\left(x_i + hK_3, y_i + hL_3\right) \\ L_4 &= g\left(x_i + hK_3, y_i + hL_3\right) \end{aligned}$$

ISBN: 978-989-95079-6-8

Since the arguments x_i and y_i on the right-hand side of (5) and (6) are interactive, we can define new functions as follows:

$$F_h(x_i, y_i) = x_i + \frac{h}{6}\left[K_1 + 2K_2 + 2K_3 + K_4\right] \quad (7)$$

$$G_h(x_i, y_i) = y_i + \frac{h}{6}\left[L_1 + 2L_2 + 2L_3 + L_4\right] \quad (8)$$

By using the concept of interactivity of fuzzy intervals proposed by Majlender [14], we have the following generalised 4^{th}-order Runge-Kutta method for solving (4), where X_0 and Y_0 are interactive fuzzy intervals.

$$X_{i+1} = F_h(X_i, Y_i) \quad (9)$$

$$Y_{i+1} = G_h(X_i, Y_i). \quad (10)$$

The idea is that if $X_i(u) \geq \alpha$ for some $u \in \mathbb{R}$, then there exists a unique $v \in \mathbb{R}$ that Y_i can take. Furthermore, if u is moved to the left (right), then the corresponding value (that Y_i can take) will also move to the left (right). Vice versa, if u is moved to the left (right), then the corresponding value (that Y_i can take) will move to the right (left) [14]. From this explanation, we can compute the numerical approximations as follows:

$$x^{\alpha}_{1,i+1} = F_h(x^{\alpha}_{1,i}, y^{\alpha}_{1,i}) \quad (11)$$

$$x^{\alpha}_{2,i+1} = F_h(x^{\alpha}_{2,i}, y^{\alpha}_{2,i}) \quad (12)$$

$$y^{\alpha}_{1,i+1} = G_h(x^{\alpha}_{1,i}, y^{\alpha}_{1,i}) \quad (13)$$

$$y^{\alpha}_{2,i+1} = G_h(x^{\alpha}_{2,i}, y^{\alpha}_{2,i}) \quad (14)$$

Our purpose here is to generate accurate approximations at each α-cut. We begin by making a partition of the form $t_0 < t_1 < t_2 < ... < t_{N-1} < t_N = T$ on the interval $[t_0, T]$. This partition is uniformly spaced, that is the partition points are $t_i = t_0 + ih$, $i = 0, 1, 2, ..., N$ and the partition spacing $h = \frac{T - t_0}{N} > 0$ is sufficiently small and we called it the step-size or step-length.

5 Stability

In this section, we consider the following fuzzy autonomous system in vector form [16]:

$$\mathbf{z}'(t) = \mathbf{f}(\mathbf{z}), \quad \mathbf{z}(t_0) = \mathbf{Z}_0 \quad (15)$$

where $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ and $\mathbf{Z}_0 \in \mathcal{F}(\mathbb{R}^2)$ is a joint possibility distribution of which the marginal possibility distributions are interactive fuzzy intervals.

Next, we give several definitions and sufficient conditions for stability of the model studied.

Definition 5 *A point $\mathbf{z}_e$ is called an equilibrium point of (15) at time t_0, if for all $t \geq t_0$, it holds that $\mathbf{f}(\mathbf{z}_e) = 0$.*

This definition asserts that an equilibrium point $\mathbf{z}_e$ is a constant solution of (15), since $\mathbf{f}(\mathbf{z}_e) = 0$ and thus $\mathbf{z}'(t) = 0$ at such a point.

Lemma 1 *If $\mathbf{Z}$ is joint possibility distribution and $\mathbf{z}_e$ is an equilibrium point, then*

$$D(\mathbf{Z}, \mathbf{z}_e) = \sup_{\mathbf{s} \in [\mathbf{Z}]^0} d(\mathbf{s}, \mathbf{z}_e)$$

Definition 6 *An equilibrium point $\mathbf{z}_e$ of (15) is defined to be fuzzy stable if for all $\epsilon > 0$, there exists $\delta > 0$ such that if $D(\mathbf{Z}(t_0), \mathbf{z}_e) < \delta$ then $D(\mathbf{Z}(t), \mathbf{z}_e) < \epsilon$ for all $t \geq t_0$.*

This definition states that all solutions of (15) that start "sufficiently close" to an equilibrium point $\mathbf{z}_e$ stay "close" to $\mathbf{z}_e$. Note that this definition of fuzzy stability does not require the solution to approach the equilibrium point.

Proposition 1 *If an equilibrium point $\mathbf{z}_e$ is stable, then it is also fuzzy stable.*

Definition 7 *An equilibrium point $\mathbf{z}_e$ of (15) is defined to be asymptotically fuzzy stable if it is fuzzy stable and there exists $\delta > 0$ such that if $D(\mathbf{Z}(t_0), \mathbf{z}_e) < \delta$ then $\lim_{t \to \infty} D(\mathbf{Z}(t), \mathbf{z}_e) = 0$ for all $t \geq t_0$.*

This definition asserts that solutions of (15) that start "sufficiently close" to an equilibrium point $\mathbf{z}_e$ not only stay "close" to $\mathbf{z}_e$, but must eventually approach the equilibrium point $\mathbf{z}_e$ as t approaches infinity. Note that this is a stronger requirement than fuzzy stability, since the equilibrium point $\mathbf{z}_e$ must be stable before one can talk about whether or not it is asymptotically fuzzy stable.

Proposition 2 *If an equilibrium point $\mathbf{z}_e$ is asymptotically stable, then it is also asymptotically fuzzy stable.*

Definition 8 *An equilibrium point $\mathbf{z}_e$ of (15) is defined to be fuzzy unstable if it is not fuzzy stable. Hence, there exists an $\epsilon > 0$ such that for any choice of $\delta > 0$, there exists $\mathbf{Z}(t)$ satisfying $D(\mathbf{Z}(t_0), \mathbf{z}_e) < \delta$ such that $D(\mathbf{Z}(t), \mathbf{z}_e) \geq \epsilon$ for some t.*

This definition states that all solutions of (15) that start "sufficiently close" to an equilibrium point $\mathbf{z}_e$ move away from $\mathbf{z}_e$ when t increases.

Proposition 3 *If an equilibrium point $\mathbf{z}_e$ is unstable, then it is also fuzzy unstable.*

6 Fuzzy Predator-Prey Population Model

Consider the following fuzzy predator-prey population model

$$\begin{cases} x'(t) = x - 0.1xy \\ y'(t) = -0.5y + 0.02xy \end{cases}$$

where the initial populations of prey and predator are interactive fuzzy intervals with $q = 1$ and $r = -5$, and they are defined as follows, respectively:

$$X_0 = \begin{cases} 0, & \text{if } x < 14, \\ x - 14, & \text{if } 14 \leq x < 15, \\ -x + 16, & \text{if } 15 \leq x < 16, \\ 0, & \text{if } x > 16 \end{cases}$$

and

$$Y_0 = \begin{cases} 0, & \text{if } y < 9, \\ y - 9, & \text{if } 9 \leq y < 10, \\ -y + 11, & \text{if } 10 \leq y < 11, \\ 0, & \text{if } y > 11 \end{cases}$$

ISBN: 978-989-95079-6-8

First we find the equilibrium points of the model by setting the derivatives equal to zero. One can note that the model has two equilibrium points: $(0, 0)$ and $(25, 10)$. The equilibrium point $(0, 0)$ is uninteresting because there are no populations to observe in the model. However, the second equilibrium point is of interest. From the numerical approximation, we can see that the point $(25, 10)$ is fuzzy stable since the supports of the joint possibility distributions move around the equilibrium point when t increases (see Fig. 4). This is true since the prey population increases when the predator population is minimum, and the prey population decreases when the predator population is maximal (see Figs. 2 and 3).

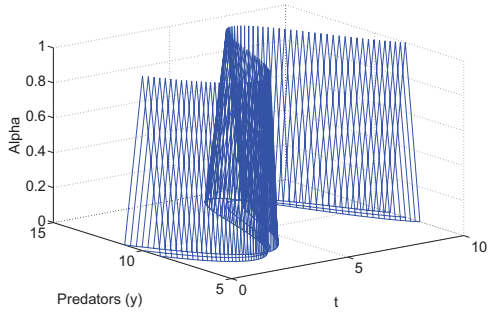


Figure 2: The evolution of the predator population over time.

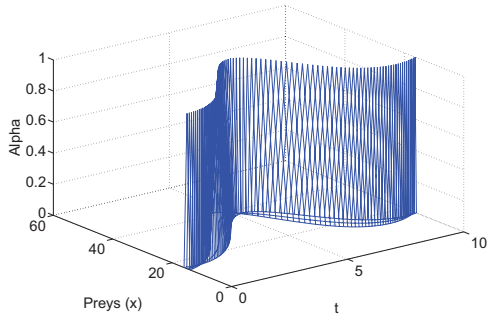


Figure 3: The evolution of the prey population over time.

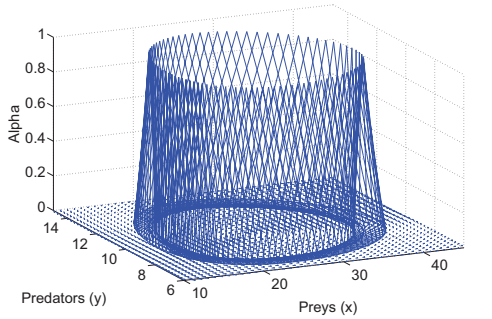


Figure 4: Fuzzy phase-plane with direction field ($N = 100$, $t = 10$, $h = 0.1$).

7 Conclusions

In this paper, we have proposed a new way of studying predator-prey population models, which takes into account the uncertainty that arises when determining the initial populations of predator and prey. By using the concept of interactivity of fuzzy intervals proposed by Majlender [14], we have developed a generalised numerical method, based on a 4^{th}-order Runge-Kutta method, to approximate the number of predators and preys over time.

Acknowledgment

This research was co-funded by the Ministry of Higher Education of Malaysia (MOHE) and partially supported by Universiti Malaysia Perlis (UniMAP) under the program "Skim Latihan Akademik IPTA (SLAI)".

References

[1] G.L. Diniz, J.F.R. Fernandes, J.F.C.A. Meyer, and L.C. Barros. A fuzzy Cauchy problem modelling the decay of the biochemical oxygen demand in water. *Proc. Joint 9th IFSA World Congress and 20th NAFIPS International Conferences*, 1:512–516, 25-28 July 2001.

[2] S. Abbasbandy and T. Allahviranloo. Numerical solution of fuzzy differential equation by Runge-Kutta method. *Nonlinear Studies*, 11(1):117–129, 2004.

[3] M. Ma, M. Friedman, and A. Kandel. Numerical solution of fuzzy differential equations. *Fuzzy Sets and Systems*, 105:133–138, 1999.

[4] B. Bede and S.G. Gal. Generalizations of the differentiability of fuzzy-number-valued functions with applications to fuzzy differential equations. *Fuzzy Sets and Systems*, 151:581–599, 2005.

[5] J.J. Buckley and T. Feuring. Fuzzy differential equations. *Fuzzy Sets and Systems*, 110:43–54, 2000.

[6] Y. Chalco-Cano and H. Román-Flores. On new solutions of fuzzy differential equations. *Chaos, Solitons and Fractals*, 38 (1):112–119, 2008.

[7] E. Hüllermeier. An approach to modelling and simulation of uncertain dynamical systems. *Int. J. Uncertainty Fuzziness Knowledge-Based Systems*, 5:117–137, 1997.

[8] M.T. Mizukoshi, L.C Barros, Y. Chalco-Cano, H. Román-Flores, and R.C. Bassanezi. Fuzzy differential equations and the extension principle. *Information Sciences*, 177:3627–3635, 2007.

[9] R. C Bassanezi, L.C. Barros, and P.A. Toneli. Attractor and asymptotic stability for dynamical systems. *Fuzzy Sets and Systems*, 113:473–483, 2000.

[10] N. R. S. Ortega, P. C. Sallum, and E. Massad. Fuzzy dynamical systems in epidemic modelling. *Kybernetes*, 29 (2):201–218, 2000.

[11] M. Guo, X. Xu, and R. Li. Impulsive functional differential inclusions and fuzzy population models. *Fuzzy Sets and Systems*, 138:601–615, 2003.

[12] V. Lupulescu. On a class of fuzzy functional differential equations. *Fuzzy Sets and Systems*, 160:1547–1562, 2009.

[13] M.S. Peixoto, L.C. Barros, and R.C. Bassanezi. Predator-prey fuzzy model. *Ecological Modelling*, 214:39–44, 2008.

[14] C. Carlsson, R. Fullér, and P. Majlender. Additions of completely correlated fuzzy numbers. *Proceeding of IEEE International Conference on Fuzzy Systems*, 1:353–539, 25-29 July 2004.

[15] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[16] S. Seikkala. On the fuzzy initial value problem. *Fuzzy Sets and Systems*, 24:319–330, 1987.

ISBN: 978-989-95079-6-8

Local semantic context analysis for automatic ontology matching

Giuseppe Fenza, Vincenzo Loia and Sabrina Senatore

Dipartimento di Matematica e Informatica,
Università degli Studi di Salerno,
via ponte don Melillo, 84084 - Fisciano (SA), Italy
E-mails: {gfenza, loia, ssenatore}@unisa.it.

Abstract—*This paper proposes an algorithm for concept matching, applied in the ontology mapping domain. Basic idea is to seek the effective semantics embedded in the concept name by analyzing the context in which it appears. Through simple interactions with the known lexicon WordNet, the right meaning associated to a concept is unequivocally elicited by exploring their local semantic context, viz. the surrounding concepts.*

This approach reveals interesting results for the word sense disambiguation, when polysemy problems requires a semantic interpretation.

Keywords— Semantic Web, concept matching, ontology mapping, WordNet, concept similarity.

1 Introduction

The Web comprises huge and disparate collections of information, whose size is estimated to overcome 11.5 billion web pages [1], though the search engines do not seem to cover the whole size (for instance, Google indexes approximately 9 billion pages). The diffusion of the Semantic Web has promised new models to support integrated access to web resources and services as well as intelligent applications for information processing on the Web.

Ontologies represents a conceivable solution for data representation as well as the knowledge sharing, aimed at the integration of the Web content in a unique and coherent view. Nevertheless, due to the decentralized nature of the Web, a plethora of ontologies has been defined and disseminated on the net; often they describe overlapped application domains; sometimes are specialized for specific domain.

It is evident the exigency to find some semantic correspondence among concepts which refer to different ontologies in order to get a "semantic reconciliation", aimed at establishing interoperability between semantic Web applications and a more homogeneous integration of information [22].

Ontology mapping has been proposed as an effective way of handling the semantic heterogeneity problem. It plays a central role in many application domains, such as e-commerce, semantic web services matchmaking [22], information integration, query mediation [23], etc..

It is the process that discovers a set of semantic correspondence between some entities of different ontologies. Many research studies have yield approaches and tools for (semi) automatic ontology matching [2], [3]; structural and linguistic matching has been taken into account; often combined approaches for the matching valuation provide efficient results too.

This paper presents a simple approach to concept matching, based on the linguistic similarity. The main idea is to discriminate the effective meaning of a word by analyzing the context in which it appears, in order to overcome the polysemy problems and the lexical ambiguity.

The approach achieves a primitive ontology mapping, by discovering semantic correspondences between concepts of two ontologies (implemented as graph-like structures). A graphical interface presents the discovered mapping.

The paper is organized as follows. Section 2 sketches a state of art in the ontology matching domain. Section 3 introduces our approach, describing the basic algorithm and then in Section 4, implementation details and some relevant case studies are presented in order to validate its applicability. Finally, conclusions close the paper.

2 Related works

In the last years, ontologies are increasingly being used to support the integration of information. Yet, their diffusion in many Web areas emphasizes impressive heterogeneity among information sources and in particular in the formal model exploited to encode the domain conceptualizations.

In literature, different types of heterogeneity have been identified, mainly split in the following classes.

- *Syntactic heterogeneity*: represents the heterogeneity due to differences at the language level; for instance, when two ontologies are defined by using different knowledge representation formalisms even though the meaning is the same [4].
- *Terminological heterogeneity*: occurs when different names are given to similar ontology entities [6].
- *Semantic heterogeneity*: is often called conceptual heterogeneity. It occurs when the same domain of interest is modelled in different way, for instance, exploiting different types of axioms for defining concepts or just giving diverse expressive values to them [5].

Many efforts have been done from the research communities to find a solution to the matching problem by developing a variety of tools and mostly providing a well-founded formalization [7, 8, 9]. Matching ontologies (or schemas) has been recognized as a critical operation in many

ISBN: 978-989-95079-6-8

application domains. It takes as input two ontologies, each one consisting of a set of entities and produces as output the "connections" (which may represent equivalences or just relations such as consequence, subsumption, or disjointness) holding between these entities.
Systems for ontology matching such as Cupid [11], OntoBuilder [10], Similarity Flooding [2], Clio [12], Glue [13], S-Match [3], OLA [14] and Prompt [15] aim at providing an alignment, namely a set of correspondences between semantically related entities of different ontologies. Finding the correspondences which are effective from the user's viewpoint is often not easy.
In [17] a tree-based classification of elementary matching approaches is given. At granularity level, matching techniques can be distinguished into the following classification criteria:
– Element-level vs. structure-level: Element-level matching techniques discover correspondences by analyzing entities or their relative instances, without considering their relations with other entities or their instances. Structure-level techniques instead compute correspondences by studying the structure in which the concepts or their instances appear [16].
– Syntactic vs. external vs. semantic: The syntactic techniques consider the input with regard to its sole structure, according to some well stated algorithm. On the other hand, external techniques exploit external resources (such as human support, thesaurus, etc.) and common knowledge in order to interpret the input [17]. Finally, semantic techniques use some formal semantics (i.e. model-theoretic approaches) to interpret the input and justify their results. Semantic matching [3] is based on the idea that labels at concept nodes, which are written in natural language, are translated into propositional formulas which codify the intended meaning of the labels themselves. This way the matching problem becomes a propositional unsatisfiability problem, which can then be efficiently implemented through propositional satisfiability (SAT) solvers [18].
Although many advances in the ontology matching research topic have been addressed, nevertheless, there are still open issues to investigate in order to get a more efficient and effective integration of ontology matching tools in the web applications domain.

3 Ontology Matching

The ontology matching problem can be seen as a problem of concept matching. Two concepts match, if they are similar. In this section some formal notions are given to delineate a common background, useful to describe our ontology matching approach. Then the ontology matching method is detailed.

3.1 Scope of a concept into an ontology

Concept matching requires the evaluation of the similarity between two concepts. This approach aims at discovering linguistic similarities between the involved entities. In general, linguistic similarities are based on morphology and semantics, which are associated to the words that describe the relative entities. Thus, in this approach the similarity between two entities of different ontologies is evaluated not only by investigating the semantics of the entities names, but also taking into account the local context, through which the effective meaning is described.
Often the same word placed in different textual contexts assumes completely different meanings. In addition, lexicons are not able to disambiguate situations in which homonyms occur.
In order to deal with lexical ambiguity, this approach introduces the notion of "scope" of a concept which represents the context where the concept is placed.

Definition 1: *Let O be ontology and c is a concept in O. The scope of c, with radius r,* ***scope(c, r)*** *is a set of all the concepts outgoing from c included in a path of length r, with centre c. More formally:*

$$scope(c,r) = \{c' |\, c' \in O\ , dist(c,c') < r\}$$

where $dist(c,c')$ *is the number of edges that are in the path from concept c to concept c'. Let us note that* $dist(c,c') = 0\ iff\ c = c'$.

Figure 1 sketches the idea. The scope defines a round area composed of all concepts that are connected directly or indirectly to the central node. This area represents the context. Increasing the radius means to enlarge the scope (i.e. this area) and, consequently, the set of neighbour concepts that intervene in the description of the context.
Fixed the radius of the scope of two concepts α and β, belonging to different ontologies, our goal is to find out a semantic relationship exist between them.

Definition 2. Let *α be a concept in the ontology O. The concept name of α,* ***a= label(α)*** *is the linguistic label associated to the concept α.*

Definition 3. Given two concepts *α , β belonging respectively to* ontologies *O and O', let* $\mathcal{L}$ *be a lexicon and a= label(α) and b= label(β). Then, let lex (a, b) ∈ [0, 1] be a lexical similarity associated to the pair of concept names (α, β), with α∈O , β∈O'. The set L is composed of all pairs, defined as follows:*

$$L = \{(\alpha', \beta') | \forall\ \alpha \in scope(\alpha', r)\ and\ \forall\ \beta \in scope(\beta', r)\ and\ \exists lex(a, b) \neq 0)\}$$

WordNet is the lexicon exploited in this approach. It is one of the most common lexical databases, used for research studies in computational linguistics, text analysis, etc. The function *lex(a, b)* is the similarity metric.
Many similarity functions have been implemented in the WordNet package. In particular, a measure specifically developed for WordNet is the similarity proposed by Wu and Palmer [19]. It is simple to compute and results an efficacy metric for ontology matching.

ISBN: 978-989-95079-6-8

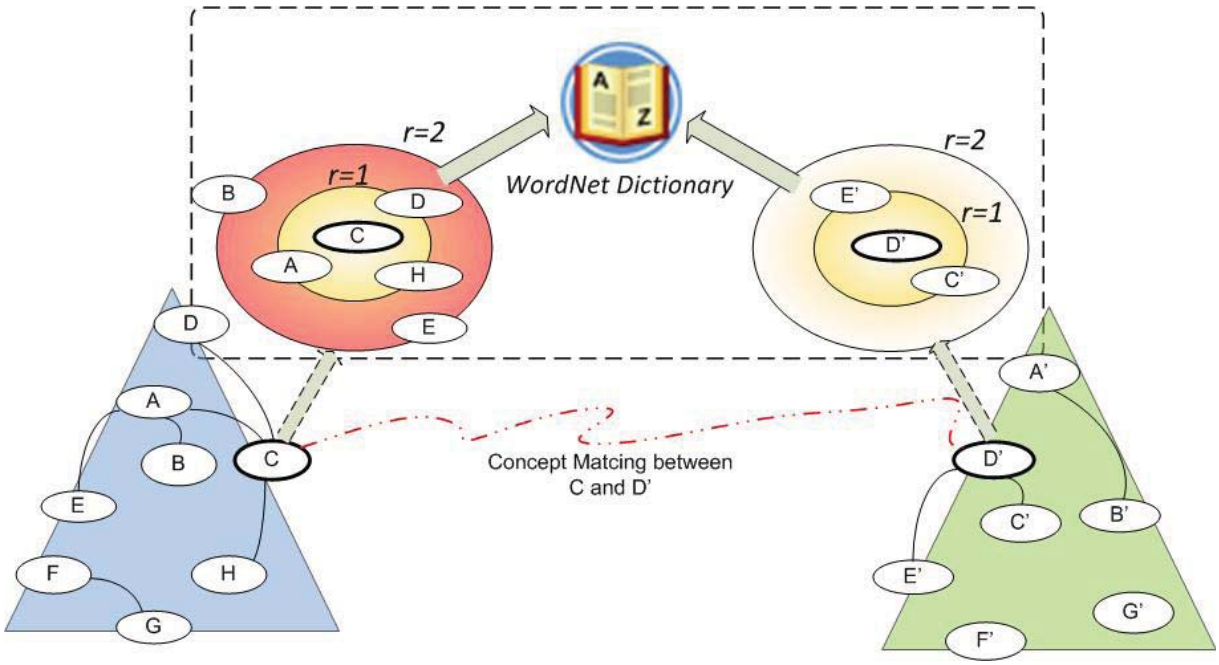


Figure 1: Intuitive idea of concept matching, exploiting the *scope* and the effective *sense* of the concept names

The Wu-Palmer similarity is based on the observation that although two concepts near the root of a hierarchy are close to each other in terms of edges, they can be very different semantically; conversely, two classes under one concept that are distant as number of edges, could be more similar conceptually. The idea is to find a common closest hyperonym in the path from them to the root.
The principle of similarity computation between two nodes is based on the distance which separates these two nodes from the root node and the distance which separates their closest common ancestor from the root too.
Let us assume that the arcs in ontology represent uniform distances (i.e. all the semantic edges have the same weight).

Definition 4. *Let c_1 and c_2 be two ontology elements in the ontology O. The distance between two nodes $\delta(c_1, c_2)$ is represented by the minimum number of edges that connect them.*

Definition 5. *Given an ontology O formed by a set of nodes and a root node R. Let c_1 and c_2 be two ontology concepts of which we will calculate the similarity. Then, g is the common ancestor of c and c'. The Wu–Palmer similarity is defined by the following expression[19]:*

$$sim(c_1, c_2) = \frac{2 \times \delta(g, R)}{\delta(c_1, R) + \delta(c_2, R)}$$

3.2 Concept Sense Discrimination

Through the given definitions, it is possible to individuate the meaning of a name associated to an ontology concept. Given a word, WordNet provides a list of all the synsets and word senses, related to that word. Just to give an example, submitting the word "table" to WordNet, the result is the following list.

```
The noun table has 6 sense (first 3 form tagged
texts)

1. (57) table, tabular array -- (a set of data
   arranged in rows and columns; "see table 1")
2. (25) table -- (a piece of furniture having a
   smooth flat top that is usually supported by
   one or more vertical legs; "it was a sturdy
   table")
3. (5) table -- (a piece of furniture with
   tableware for a meal laid out on it; "I
   reserved a table at my favorite restaurant")
4. mesa, table -- (flat tableland with steep
   edges; "the tribe was relatively safe on the
   mesa but they had to descend into the valley
   for water")
5. table -- (a company of people assembled at a
   table for a meal or game; "he entertained the
   whole table with his witty remarks")
6. board, table -- (food or meals in general;
   "she sets a fine table"; "room and board")
```

Let us note all the senses are ranked with respect to the frequency of the term (shown in the parenthesis for the first three senses) in the reference context (according to the spoken common sense).
The following pseudo-code details the algorithm for discriminate the actual sense of a word associated to a given concept. The algorithm takes as input an ontology O, a reference concept α in that ontology and a word *w*. The word *w* represents the name associated to the concept α (i.e. w=label(α)).
Let us note the semantic difference between the concept (or class) in an ontology and the label associated to that concept. The algorithm replies to question like "I would like to know the effective sense of the word w, placed in the context (or scope) of the concept α".

ISBN: 978-989-95079-6-8

Algorithm1

```
Input: Ontology O, concept α ∈ O, radius r and the
word w

Output: sense number of w

1: build an array N[|synset(w)|];
2: for each t1 in synset(w)
3:    initialize N[senseNumber(t1)] = 0;
4:    for i = 1 to r
5:       set M = ∅;
6:       for each c in scope(α,i) - scope(α,i-1)
7:          set S = ∅;  //set of similarity values
8:          for each t2 in synset(label(c))
9:             sim = similarity_WP(t1, t2)
10:            S = S U {sim};
11:         end for each
12:
13:         set max = maximum in S;
14:         max = max / |scope(α, i)|;
15:         M = M U {max};
16:      end for each
17:      set Sum = summation of all the values in M
18:      Sum = Sum /i;
19:      N[senseNumber(t1)] += Sum;
20:   end for
21: end for each
22: i1 = sense number with the highest
                              frequency in N.
23: return i1
```

First step (line 1) is to declare a vector structure whose size corresponds to the number of synsets (or senses) associated to the given word w. Goal is to maintain in each cell of the vector a pertinence value that represents how much the word w is semantically related to that sense (or belongs to that synset).
The algorithm selects all the concepts in the scope of α (belonging to the reference ontology O) by varying the radius (lines 4-6), in order to get different set of terms. Then compute the Wu-Palmer similarity between two terms coming from the concept name of α and the word *w* (line 9).
Just to give an idea about how the algorithm works, let us look at Figure 2. For simplicity, let us suppose *w* = *label(α)* = "Table". The algorithm evaluates the similarity between all the terms coming from both the WordNet synset of "Table" and the scope of concept "Table" in the ontology.
In particular, Figure 2 shows the flat ring shaped areas in different colours, by varying the radius.
For each concept *c* in the ring shaped area (computed as the difference of the areas between two successive radii, see line 6), the max similarity values between the name associated to *c* and a concept name in synset of α are maintained in M (line 15).
At the end of the two loops (lines 6-11) the variable *Sum* contains the sum of all the max similarity values computed for each couple of terms coming from the fixed term t1 of synsets of α and all the terms in the synset of *w*.
The *Sum* is "weighted" with respect to the current radius (line 18). The final value of *Sum* is added to the value stored in the cell associated to the sense of some term t1. This is repeated for each t1 in the synset of *w*. At the end, in each cell of the vector N there is a value, associated to each term in the synset of *w*.

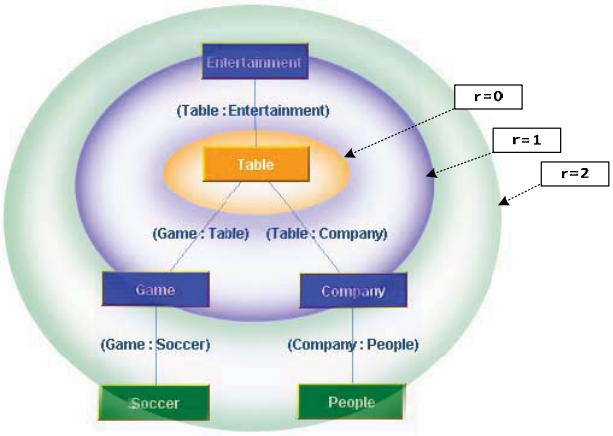


Figure 2. A sketched ontology and the scope of concept "Table" at different radius r.

The index of the vector cell, whose value is the maximum, represents the sense number to give as output.
Applying the algorithm on the ontology of Figure 2, the retuned index value is 5.
Abstractly speaking, the approach helps the user to individuate the right meaning of a word, given the context.

3.3 Ontology mapping

As said, the ontology mapping outlines correspondences or matches between concepts coming from two different ontology. In this approach, the concept matching is measured by computing a similarity between concepts at linguistic level.
The algorithm described in the previous section has been exploited to evaluate the concept matching. More formally, given two ontologies *O* e *O'* and two concepts *c* and *c'*, belonging respectively to these two ontologies, there is a match between *c* and *c'* if a similarity between them exists, computed as follows:

1. Algorithm1 is invocated twice: one time it takes as input the concept *c* and the ontology *O* and another time, by giving the concept *c'* and the ontology *O'*. Outputs of these two independent executions of Algorithm1 are two indexes, *i* for the synset of *c* and *i'* for the synset of c'. As said, they identify the sense number associated to each concept.

2. Once discovered the sense of involved concepts, the affinity between the concepts is computed by a similarity measure (for instance, the Wu-Palmer similarity defined in Section 3.1) between *c* and *c'*.

In order to obtain all the semantic correspondences among the concepts in the two ontologies, this procedure can be applied for each couple of concepts coming from two reference ontologies. Final result is an ontology mapping; a similarity value is assigned to each discovered correspondence between concepts in the two ontologies.

ISBN: 978-989-95079-6-8

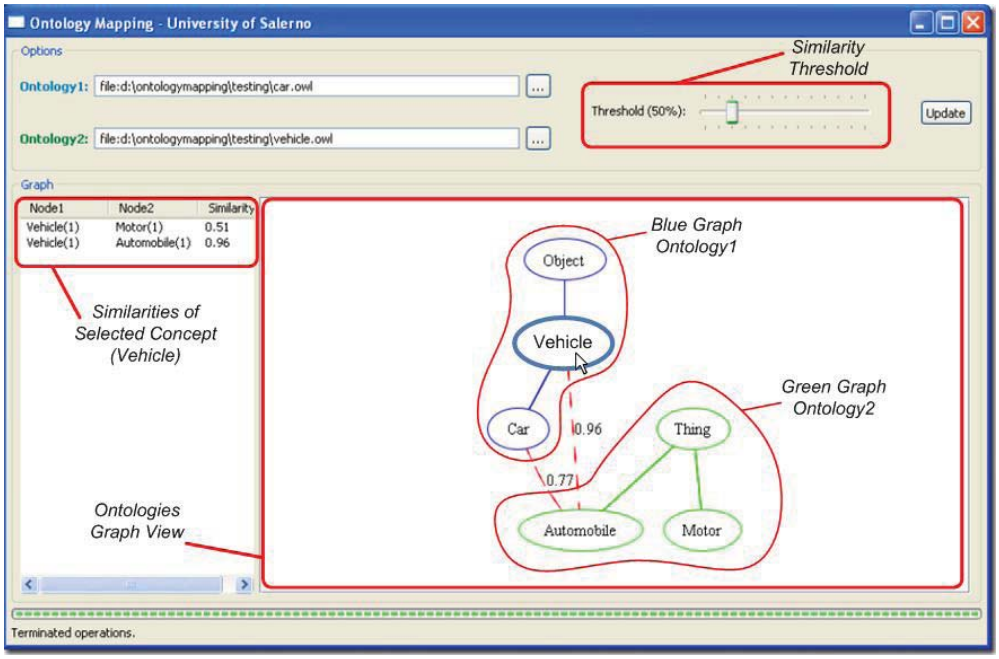


Figure 3. Ontology Mapping GUI

4 User interface and application details

The algorithm has been implemented in Java language. It exploits Java WordNet Similarity Library (JWSL) [20] for accessing the WordNet database. JSWL implements the most commons similarity and relatedness measures between words and get information about *synsets*.

A simple graphical interface has been designed for the user interaction (see Figure 3). It has been implemented by exploiting Graphviz library [21]. It is an open source graph visualization software which provides several main graph layout programs.

Let us note the ontologies have been represented (and implemented) as graphs, whose nodes represent the concepts whereas the edges connects two nodes when a relationship exists between them. In this version, a restriction on just hierarchies or taxonomies (i.e. only IS-A relationships) is considered.

The user can load the two ontologies and the final mapping is drafted in the *Ontology Graph View* panel (see Figure 3). The correspondences discovered by the mapping are sketched in Figure 3 as dotted lines between the concepts. A weight (value of similarity) is associated to each arc.

Figure 3 shows on the right hand of the interface, a sliding bar for setting the similarity threshold. Moving the cursor, the threshold is modified, returning in the graph view panel just the arcs whose similarity values are greater than this threshold.

Moreover, when the user clicks on a concept (node) in the graph view panel, all its connections with other concepts and the relative similarity values are shown on the left of the interface.

4.1 Some case studies

The algorithm has been applied on some sketched ontologies. The application acquires the ontologies in OWL format, although this prototypal version is limited to work on hierarchical structure.

In the interface of Figure 3, a simple example of ontology mapping is shown. The case study has been built ad-hoc; just the relevant portions of two ontologies are shown.

The Ontology1 (in blue color in Figure 3) is described by the concepts named “Object”, “Vehicle” and “Car” , whereas the Ontology2 (green color) is composed of the concept with names “Thing”, “Automobile” and “Motor”.

The application adds an arc between two concepts of different ontologies, when a similarity value (in the range [0, 1]) is found.

Let us note there exists an arc between the concept names “Vehicle” and “Automobile” with an high similarity (0.96), but the arc between “Vehicle” and “Motor” is not inserted in the graph view, because the similarity between these concept names assumes values lower than the selected threshold.

Similar considerations can be done on the ontology mapping shown in Figure 4 (although the GUI is not shown).

Two ontologies are drafted in different colors. Both of them describe a similar domain (the living beings).

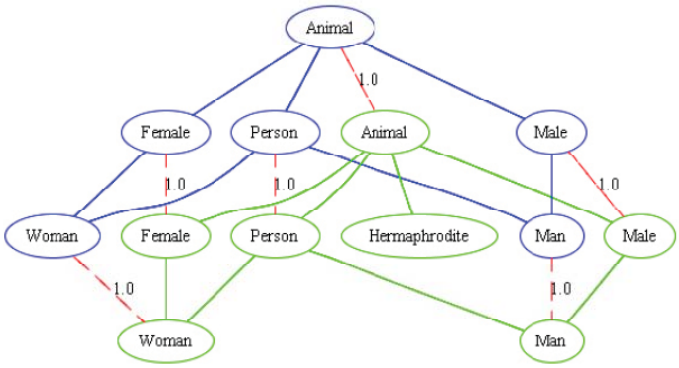


Figure 4: Example of Ontology Matching on the human beings domain.

ISBN: 978-989-95079-6-8

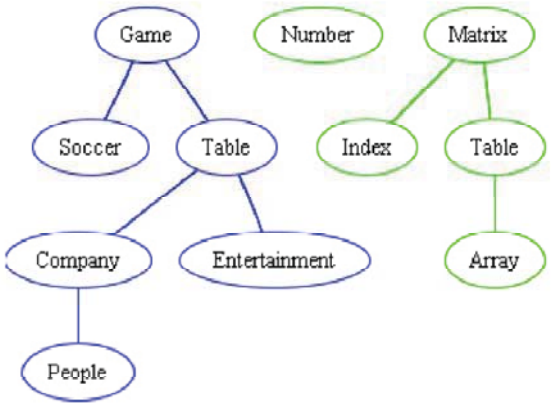


Figure 5: No concept matching occurs between these two ontologies, due to different meanings of the reference ontology domains.

Because the ontologies present concepts with similar names and meanings, the similarity values on the added arcs (in red) are equal to 1 (the maximum similarity).
Finally, Figure 5 shows a case of no concept match. The two ontologies share the same concept name "Table", but the reference context is completely different. In fact, Algorithm1 returns the sense number 5 for the ontology on the left (colored blue) and the sense number 1 for the ontology on the right, respectively (see Section 3.2 for details). Thus no arcs can be placed between them.
Let us note, in the example in Figure 5, the node with concept name "Number" is disconnected by the others, because the algorithm does not analyzes the OWL ontology relationships that are not hierarchical.

5 Conclusions

The paper describes a study for ontology mapping based on the discovering of linguistic similarity. A graphical interface allows the user to see the final mapping, presenting a similarity value for each discovered concept matching.
This algorithm represents an initial version of an ontology matching method and needs some extension for processing all the ontology relationships. Its efficacy would be visible if it was exploited in combination with structure based ontology matching. Further developments are addressed to reach these issues. Nevertheless, its applicability may not restricted to ontology matching problems. The basic algorithm could be used for Information Retrieval problems, for instance when the textual analysis requires the word sense disambiguation. The context of a word could be described by defining a "sliding window" on the text surrounding the analyzed word, in order to discover its appropriate sense. Furthermore, the algorithm could be exploited for discriminating common sense words, in specialized contexts where the word meaning depends on the application domain (or the reference ontology).

References

[1] Gulli, A. and Signorini, A. 2005. The indexable web is more than 11.5 billion pages. In *Special interest Tracks and Posters of the 14th international Conference on World Wide Web* (Chiba, Japan, May 10 - 14, 2005). WWW '05. ACM, New York, NY, 902-903.

[2] S. Melnik, H. G. Molina, E. Rahm, Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching, *In Proceedings of the 18th International Conference on Data Engineering*, San Jose, CA, 2002, pp.117-128.

[3] F. Giunchiglia, M. Yatskevich, E. Giuchilglia, Efficient Semantic Matching, *In Proceedings of ESWC, Heraklion*, Greece, 2005, pp.272-289.

[4] J. Euzenat and H. Stuckenschmidt. The 'family of languages' approach to semantic interoperability. In Borys Omelayenko and Michel Klein, editors, Knowledge transformation for the semantic web, pages 49–63. IOS press,Amsterdam (NL), 2003.

[5] M. Klein. Combining and relating ontologies: an analysis of problems and solutions. In Proc. IJCAI Workshop on Ontologies and Information Sharing, Seattle (WA US), 2001.

[6] N. Noy and M. Klein. Ontology evolution: Not the same as schema evolution. Knowledge and Information Systems, 6(4):428–440, 2004.

[7] P. Bernstein, A. Halevy, and R. Pottinger. A vision of management of complex models. *ACM SIGMOD Record*, 29(4):55–63, 2000

[8] P. Bouquet, M. Ehrig, J. Euzenat, E. Franconi, P. Hitzler, M. Krotzsch, L. Serafini, G. Stamou, Y. Sure, and S. Tessaris. Specification of a common framework for characterizing alignment. Deliverable D2.2.1, Knowledge web NoE, 2004

[9] A. Zimmermann, M. Krotzsch, J. Euzenat, and P. Hitzler. Formalizing ontology alignment and its operations with category theory. In Proc. 4th International Conference on Formal Ontology in Information Systems (FOIS),pages 277–288, Baltimore (MD US), 2006

[10] A. Gal, G. Modica, H. Jamil, and A. Eyal. Automatic ontology matching using application semantics. AI Magazine, 26(1):21{32, 2005.

[11] J. Madhavan, P. Bernstein, and E. Rahm. Generic schema matching with Cupid. In Proceedings of the 27th International Conference on Very Large Data Bases (VLDB), pages 48-58, Rome, Italy, 2001.

[12] R. Miller, M. Hernandez, L. Haas, L.-L. Yan, C. Ho, R. Fagin, and L. Popa. The Clio project: Managing heterogeneity. SIGMOD Record, 30(1):78-83,2001.

[13] A. Doan, J. Madhavan, P. Domingos, and A. Halevy. Learning to map between ontologies on the semantic web. In Proceedings of the 11th International Conference on World Wide Web (WWW), pages 662{673, Honolulu, USA, 2002.

[14] J. Euzenat and P. Valtchev. Similarity-based ontology alignment in OWLlite. In Proceedings of the 15th European Conference on Artificial Intelligence (ECAI), pages 333-337, Valencia, Spain, 2004.

[15] N. Noy and M. Musen. The PROMPT suite: interactive tools for ontology merging and mapping. International Journal of Human-Computer Studies, 59(6):983-1024, 2003.

[16] J. Kang and J. Naughton. On schema matching with opaque column names and data values. In Proc. 22nd International Conference on Management *of Data (SIGMOD)*, pages 205–216, San Diego (CA US), 2003.

[17] P. Shvaiko and J. Euzenat. A survey of schema-based matching approaches. Journal on Data Semantics, IV:146–171, 2005.

[18] D. Le Berre JSAT: the java satisfiability library http://www.sat4j.org/

[19] Z. Wu and M. Palmer. Verb semantics and lexical selection. In Proc. 32nd Annual Meeting of the Association for Computational Linguistics (ACL), pages 133–138, Las Cruces (NM US), 1994.

[20] Java WordNet Similarity Library (JWSL) Available: http://grid.deis.unical.it/similarity/

[21] Graphviz - Graph Visualization. Available: http://www.graphviz.org/

[22] G. Fenza, V. Loia, S. Senatore, A hybrid approach to semantic web services matchmaking, International Journal of Approximate Reasoning, Volume 48, Issue 3, August 2008, pp. 808-828.

[23] G. Fenza, V. Loia, S. Senatore: *Improving Fuzzy Service Matchmaking through Concept Matching* Discovery. FUZZ-IEEE 2007: pp. 1-6

ISBN: 978-989-95079-6-8

Using fuzzy linguistic summaries for the comparison of time series: an application to the analysis of investment fund quotations

Janusz Kacprzyk[1,2] Anna Wilbik[1*]

[1] Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
[2] WIT – Warsaw School of Information and Technology
Newelska 6, 01-447 Warsaw, Poland
Email: kacprzyk@ibspan.waw.pl, wilbik@ibspan.waw.pl

Abstract— *We propose a new, human consistent method for the evaluation of similarity of time series that uses a fuzzy quantifier base aggregation of trends (segments), within the authors' (cf. Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9]) approach to the linguistic summarization of trends based on Zadeh's protoforms and fuzzy logic with linguistic quantifiers. The results obtain are very intuitively appealing and justified by valuable outcomes of similarity analyses between quotations of an investment fund and the two main indexes of the Warsaw Stock Exchange.*

1 Introduction

Among various application areas of advanced data mining and knowledge discovery tools and techniques, all kinds of finance related areas are of primordial and growing importance. For instance, in an interesting statistical report [10] on top data mining applications in 2008, the first two positions are, in the sense of yearly increase:

• *Investment/Stocks*, up from 3% of respondents in 2007 to 14% of respondents in 2008% (350% increase),

• *Finance*, up form 7.2% in 2007 to 16.8% in 2008 (108% increase).

This trend will presumably continue in view of serious difficulties on financial/investment market which is expected to continue well after 2009.

This paper is a continuation of our previous works (cf. Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9]) which deal with the problem of how to effectively and efficiently support a human decision maker in making decisions concerning investments. We deal mainly with investments in investment (mutual) funds. Clearly, decision makers are here concerned with possible future gains/losses, and their decisions is related to what might happen in the future. However, our aim is not the forecasting of the future daily prices, which could have been eventually used directly for a purchasing decision. Instead, in our works, we follow a decision support paradigm, that is we try to provide the decision maker with some information that can be useful for his/her decision on whether and how many units to purchase. We do not intend to replace the human decision maker.

This problem is very complex. First of all, there may be two general approaches. The first one, which may seem to be the most natural is to provide means to derive a price forecast for an investment unit so that the decision maker could "automatically' purchase what has been forecast, and as much as he/she could afford. Unfortunately, the success in such a straightforward approach is much less than expected. Basically, statistical methods employed usually for this purpose are primitive in the sense that they just extrapolate the past and do not use domain knowledge, intuition, some inside information, etc. A natural solution may be to try to just support the human decision maker in making those investment decisions by providing him/her with some additional useful information, and not getting involved in the actual investment decision making.

Various philosophies in this respect are possible. Basically, from our perspective, the following one will be followed. In all investment decisions the future is what really counts, and the past is irrelevant. But, the past is what we know, and the future is (completely) unknown. Behavior of the human being is to a large extend driven by his/her (already known) past experience. We usually assume that what happened in the past will also happen (to some, maybe large extent) in the future. This is basically, by the way, the very underlying assumption behind the statistical methods too!

This clearly indicates that the past can be employed to help the human decision maker find a good solution. We follow here this path, i.e. we present a method to subsume the past, to be more specific the past performance of an investment (mutual) fund, by presenting results in a very human consistent way, using natural language statements.

It should be noted that this line of reasoning has often been articulated by many well known investment practitioners, and one can quote here some more relevant opinions. In any information leaflets of investment funds, one may always notice a disclaimer stating that "Past performance is no indication of future returns" which is true. However, on the other hand, for instance, in a well known posting "Past Performance Does Not Predict Future Performance" [11], they state something that may look strange in this context, namely: "... according to an Investment Company Institute study, about 75% of all mutual fund investors mistakenly use short-term past performance as their primary reason for buying a specific fund". But, in an equally well known posting "Past performance is not everything" [12], they state: "... disclaimers apart, as a practice investors continue to make investments based on a schemes past performance. To make matters worse, fund houses are only too pleased to toe the line by actively advertising the past performance of their schemes leading investors to conclude that

*Partially supported by the Ministry of Science and Higher Education Grant no. NN 516 4309 33.

ISBN: 978-989-95079-6-8

it is the single-most important parameter (if not the most important one) to be considered while investing in a mutual fund scheme".

As strange as this apparently is, we may ask ourselves why it is so. Again, in a well known posting "New Year's Eve: Past performance is no indication of future return" [13], they say "...if there is no correlation between past performance and future return, why are we so drawn to looking at charts and looking at past performance? I believe it is because it is in our nature as human beings ...because we don't know what the future holds, we look toward the past ...".

And, continuing along this line of reasoning, we can find many other examples of similar statements supporting our position. For instance, Myers [14] says: "...Does this mean you should ignore past performance data in selecting a mutual fund? No. But it does mean that you should be wary of how you use that information ...While some research has shown that consistently good performers continue to do well at a better rate than marginal performers, it also has shown a much stronger predictive value for consistently bad performers ...*Lousy performance in the past is indicative of lousy performance in the future...*". And, further: Bogle [15] states: "... there is an important role that past performance can play in helping you to make your fund selections. While you should disregard a single aggregate number showing a fund's past long-term return, you can learn a great deal by studying the *nature of its past returns*. Above all, look for consistency.". In [16], we find: "While past performance does not necessarily predict future returns, it can tell you how volatile a fund has been". In the popular "A 10-step guide to evaluating mutual funds" [17], they say in the last, tenth, advise: "Evaluate the funds performance. Every fund is benchmarked against an index like the BSE Sensex, Nifty, BSE 200 or the CNX 500 to cite a few names. Investors should compare fund performance over varying time frames vis-a-vis both the benchmark index and peers. Carefully evaluate the funds performance across market cycles particularly the downturns".

One can give many more quotations from all kinds of investment guides, analyses, etc. by leading experts. Virtually all of them emphasize the importance of looking at the past to help make future decisions. Moreover, they also generally advocate a more comprehensive look in the sense that what might be useful would rather be not particular single values but some deeper meaning, even nature of past behavior and returns.

We have followed this line of reasoning in our past papers (cf. Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9]), i.e. to try to find a human consistent, fuzzy quantifier based scheme for a linguistic summarization of the past in terms of various aspects of how the time series representing daily quotations of the investment fund(s) behave. However, we have mainly concentrated on a sheer absolute performance, i.e. the time evolution of the quotations themselves. This may be relevant, and sometimes attractive to the users who can see a summary of their gains/loses and their temporal evolution.

However, for many more detailed analyses we need to relate the performance of an investment fund to the performance (time evolution) of some benchmark(s) against which the particular investment fund is to be compared, as stated in its prospectus. These benchmarks are usually some stock exchange indexes, or a composition thereof.

One of more interesting and relevant questions that may be posed by both a professional analyst or a customer whose money is being invested may be: how similar was the temporal evolution of quotations of the particular investment fund and its benchmark(s)? This has clearly to do with the measuring of similarity of time series.

The problem of similarity of time series or finding similar subsequences of time series is an important issue in many problems, e.g., in indexing [18, 19], clustering [20] or motif discovery [21, 22]. A similarity measure should allow for the imprecise match, i.e., should indicate time series sequences that are similar to some extent. Moreover, the algorithm should be very efficient [23], as we may consider very long time series.

Different approaches to the similarity measures of time series were considered and proposed in the literature. A simple solution is to view each sequence as a point in an n-dimensional Euclidean space, where n is the length of the segment (trend). The similarity or dissimilarity is based on the Euclidean distance, and computed as the norm in this space [24]. This approach can be used for various time series but normally some extensions should be applied providing various transformations such as scaling or shifting were proposed [21].

Another methods use, for instance, the Dynamic Time Wraping (DTW) [19] and Longest Common Subsequence (LCS) measures [24], and they allows gaps in sequences. They are successfully used in other domains such as speech recognition and text pattern matching [25].

Other approaches developed for similarity search in time series data are based on dimension reduction. This can be achieved in numerous ways, e.g. using the Discrete Fourier Transform (DFT) [23], wavelets [26], etc.

One can also mention here approaches based on the piecewise linear representation of time series [27]. In this approach time series are represented by segments (trends), not the particular values. The method proposed here falls into this category.

In this paper we propose new method of calculating the similarity of time series; as we have already indicated, this concerns the quotations of an investment fund and some benchmark. The degree of similarity of two time series is meant here as the degree to which, for instance, "most" of the long, overlapping segments are similar (for instance, such that "almost all" of their features are similar). Notice that this has very much to do with a soft definition of consensus meant as a degree to which, for instance, "*most* of the *important* individuals agree as to *almost all* of the *relevant* options". This was introduced by Kacprzyk and Fedrizzi [28, 29, 30], Kacprzyk, Fedrizzi and Nurmi [31], and recently by Kacprzyk and Zadrożny [32] and has been found useful in many areas.

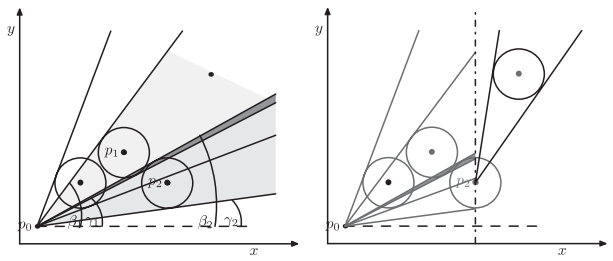
We will now discuss first the segmentation of time series, as the similarity will proceed in terms of segments (trends), not all the consecutive values. Then, we will propose some new, linguistic quantifier based measures of similarity, followed by examples on real data, and conclusions.

ISBN: 978-989-95079-6-8

2 Segmentation of time series

In our approach (cf. Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9]) we summarize the trends (segments) extracted from a time series, and we have to extract these segments assumed to be represented by a fragment of a straight line. There are many algorithms for a piecewise linear segmentation of time series, including, e.g., on-line (sliding window) algorithms, bottom-up or top-down strategies (cf. Keogh [33, 34]). We may either divide the time series into k segments (using, e.g., bottom-up or top-down strategies), where k is specified by the user, or divide it so that the maximum error for any segment does not exceed some threshold value ε, also specified by the user. In our works [1, 2, 3, 4, 5, 6, 7, 8, 9] we used a simple on-line algorithm, a modification of one proposed by Sklansky and Gonzalez [35] in the contest of image analysis. The method is simple and fast, and fully satisfies our needs though many other methods are possible.

Our algorithm constructs the segments so that the maximum distance between the segment line and the observed time point is smaller than a user specified ε value. It works by constructing the intersection of cones starting from point p_i of the time series and including a circle of radius ε around the subsequent data points p_{i+j}, $j = 1, 2, \ldots$, until the intersection of all cones starting at p_i is empty. If for p_{i+k} the intersection is empty, then we construct a new cone starting at p_{i+k-1}. Figure 1 present the idea of the algorithm. The family of possible solutions is indicated as a gray area. Clearly other algorithms can also be used, and there is a lot of them in the literature (cg. [18, 36]).



(a) the intersection of the cones is indicated by the dark grey area

(b) a new cone starts in point p_2

Figure 1: Illustration of the segmentation algorithm used

To present details of the algorithm, let us first denote:
- `p_0` – a point starting the current cone,
- `p_1` – the last point checked in the current cone,
- `p_2` – the next point to be checked,
- `Alpha_01` – a pair of angles (γ_1, β_1), meant as an interval, that defines the current cone as shown in Figure 1(a),
- `Alpha_02` – a pair of angles of the cone starting at the point `p_0` and inscribing the circle of radius ε around the point `p_2` (cf. (γ_2, β_2) in Figure 1(a)),
- function `read_point()` reads a next point of data series,
- function `find()` finds a pair of angles of the cone starting at the point `p_0` and inscribing the circle of radius ε around the point `p_2`.

A pseudocode of the algorithm that extracts trends is depicted in Figure 2.

```
read_point(p_0);
read_point(p_1);
while(1)
{
  p_2=p_1;
  Alpha_02=find();
  Alpha_01=Alpha_02;
  do
  {
    Alpha_01 = Alpha_01 ∩ Alpha_02;
    p_1=p_2;
    read_point(p_2);
    Alpha_02=find();
  } while(Alpha_01 ∩ Alpha_02 ≠ ∅);
  save_found_trend();
  p_0=p_1;
  p_1=p_2;
}
```

Figure 2: Pseudocode of the modified Sklansky and Gonzalez [35] algorithm for extracting trends

The bounding values of `Alpha_02` (γ_2, β_2), computed by function `find()` correspond to the slopes of two lines that (1) are tangent to the circle of radius ε around point $p_2 = (x_2, y_2)$ and (2) start at the point $p_0 = (x_0, y_0)$. Thus

$$\beta_2, \gamma_2 = arctg\left(\frac{\Delta x \cdot \Delta y - \pm\varepsilon\sqrt{(\Delta x)^2 + (\Delta y)^2 - \varepsilon^2}}{(\Delta x)^2 - \varepsilon^2}\right)$$

where $\Delta x = x_0 - x_2$ and $\Delta y = y_0 - y_2$.

The resulting linear ε-approximation of a group of points `p_0, ..., p_1` is either a single segment, chosen as, e.g., a bisector of the cone, or one that minimizes the distance (e.g., the sum of squared errors, SSE) from the approximated points, or the whole family of possible solutions, i.e., rays of the cone.

This method is effective and efficient as it requires only a single pass through the data. From now on we will identify (partial) *trends* with the line segments of the constructed piecewise linear function.

3 Linguistic summarization of time series

First, we will outline the very essence of our approach to the linguistic summarization of time series which is based on the two types of protoforms of linguistic summaries of trends:
- a short form:

$$\text{Among all segments, } Q \text{ are } P \tag{1}$$

e.g.: "Among all segments, *most* are *slowly increasing*".
- an extended form:

$$\text{Among all } R \text{ segments, } Q \text{ are } P \tag{2}$$

e.g.: "Among all *short* segments, *most* are *increasing*".
whose truth values are, respectively:

$$\mathcal{T}(\text{Among all } y\text{'s, } Q \text{ are } P) = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_P(y_i)\right) \tag{3}$$

$$\mathcal{T}(\text{Among all } Ry\text{'s, } Q \text{ are } P) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_R(y_i) \wedge \mu_P(y_i)}{\sum_{i=1}^{n}\mu_R(y_i)}\right) \tag{4}$$

where $\wedge$ is the minimum operation (more generally it can be another appropriate operator, notably a t-norm).

ISBN: 978-989-95079-6-8

For more information on this approach, and on the advantages of using protoforms as general forms of linguistic summaries, see Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9].

4 A new fuzzy quantifier based method for the evaluation of similarity of time series

The degree of similarity of two time series is meant here as the degree to which, say "most" of the long, simultaneous segments are similar (i.e., "majority" of their features are similar).

The time series is represented by its segments (trends). Each segment is described by the following four values:
• starting time, called here a partition point,
• duration of the segment (expressed in time units),
• dynamics of change (expressed as the slope of the segment),
• variability occurring in the segment, describing the variability of differences between the trend (segment line) and the particular time series values over the duration of this trend.

Let us assume, we wish to compare two time series A and B. As the result of the segmentation procedure time series A was divided into k segments, represented by a_j, $j = 1, \ldots, k$ and time series B was divided into l segments, represented by b_j, $j = 1, \ldots, l$.

Next we create the set of partition points of the two time series considered, i.e. this set contains the time points when at least one segment has started. There are at most $k + l$ such partition points. Those times are sorted, so that the earlier ones come first.

Now between any consecutive partition points (e.g. p_i and p_{i+1}) there is only one segment (or a part of it) of each time series. We may denote those segments as a_{p_i} and b_{p_i}, i.e., a_{p_i} is the segment representing time series A that take place between p_i and p_{i+1}. Now we can compute the similarity of such simultaneous segments.

We compute the degree of similarity of two segments (trends), as the degree to which "most" of their features are similar. Here we consider for simplicity the three features only: duration, dynamics and variability. The values of the features of segment a_{p_i} are $dur_{a_{p_i}}$ for duration, $dyn_{a_{p_i}}$ for dynamics, and $var_{a_{p_i}}$ for variability. R_{dur}, R_{dyn} and R_{var} are the ranges of possible values of duration, dynamics and variability. The degree of similarity of segments a_{p_i} and b_{p_i} with respect to duration is computed as

$$sim^{dur}(a_{p_i}, b_{p_i}) = \mu^{dur}_{sim}\left(\frac{|dur_{a_{p_i}} - dur_{b_{p_i}}|}{R_{dur}}\right) \quad (5)$$

Similarly, separate functions can be defined for the similarity for each feature, here μ^{dyn}_{sim} and μ^{var}_{sim}. Let $\mu_{Q_1}(.)$ be the membership function of quantifier "most", i.e regular, nondecreasing and monotone. Thus the degree of similarity of segments a_{p_i} and b_{p_i} is computed as

$$\begin{aligned} sim(a_{p_i}, b_{p_i}) = \mu_{Q_1}\Bigg(& w_{dur}\left(\mu^{dur}_{sim}\left(\frac{|dur_{a_{p_i}} - dur_{b_{p_i}}|}{R_{dur}}\right)\right) + \\ &+ w_{dyn}\left(\mu^{dyn}_{sim}\left(\frac{|dyn_{a_{p_i}} - dyn_{b_{p_i}}|}{R_{dyn}}\right)\right) + \\ &+ w_{var}\left(\mu^{var}_{sim}\left(\frac{|var_{a_{p_i}} - var_{b_{p_i}}|}{R_{var}}\right)\right)\Bigg) \end{aligned} \quad (6)$$

where $w_{dur} + w_{dyn} + w_{var} = 1$, here all are equal, $w_{dur} = w_{dyn} = w_{var} = \frac{1}{3}$.

Next, to obtain the similarity of A and B, we need to aggregate above similarity values between the overlapping segments, and again we use a linguistic quantifier driven aggregation.

$$sim(A, B) = \mu_Q\left(\frac{\sum_{i=1}^{n} \frac{p_{i+1}-p_i}{T} sim(a_{p_i}, b_{p_i})}{\sum_{i=1}^{n} \frac{p_{i+1}-p_i}{T}}\right) \quad (7)$$

where $p_{i+1} - p_i$ is the difference between the two consecutive partition points p_i and p_{i+1}, and T is the total time span considered.

As $\sum_{i=1}^{n} p_{i+1} - p_i = T$ we may simplify this formula and we obtain

$$sim(A, B) = \mu_Q\left(\sum_{i=1}^{n} \frac{p_{i+1} - p_i}{T} sim(a_{p_i}, b_{p_i})\right) \quad (8)$$

5 Numerical results

The method proposed in this paper was tested on data on quotations of an investment (mutual) fund that invests at most 50% of assets in shares listed at the Warsaw Stock Exchange (WSE). We have used two benchmark time series, the index of WSE companies (WIG), that is the benchmark for this fund mentioned in the fund prospectus, and the index of 20 biggest and most liquid companies (WIG20). More information can be found on the WSE Web page (`http://www.gpw.pl`).

Data shown in Fig. 3 were collected from the beginning of 2007 until the end of 2008 (501 quotations); notice that one can see when world wide financial problems begun. They are measured in different units, and in order to compare them, we have scaled them using the following formula. The new value of the quotation

$$\bar{v}_i = \frac{v_i - v_0}{v_0} 10 + 10$$

where v_0 is the quotation value of the first observation, i.e. on January, 2, 2007. Those scaled time series may be interpreted as the amount of money that would be earned if we invested PLN 10 (PLN stands for the Polish Zloty) in the mutual fund or stocks of the index on January 2, 2007.

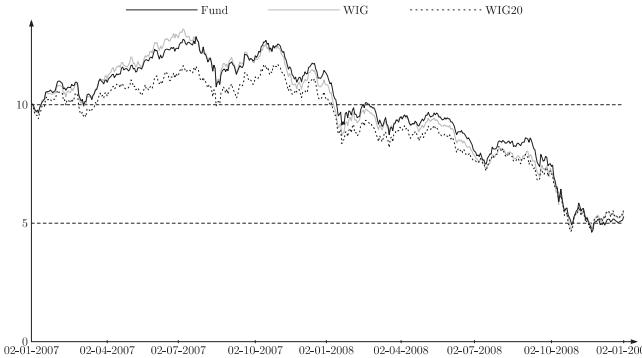


Figure 3: Daily quotations of an investment fund considered and the WIG and WIG20 indexes

The value of the mutual fund time series was equal PLN 5.28 at the end of the time span considered. The minimal value

ISBN: 978-989-95079-6-8

recorded was PLN 4.63 while the maximal one during this period was PLN 12.86.

The values of the benchmark time series were equal to PLN 5.46 and PLN 5.61 at the end of the time span considered for the WIG and WIG20, respectively. The minimum value of WIG recorded was PLN 4.71 while the maximum one during this period was PLN 13.20. The minimum and maximum values for WIG20 were PLN 4.62 and PLN 11.70, respectively.

Using the modified Sklansky and Gonzalez algorithm (cf. [2]) and $\varepsilon = 0.25$ we obtained 43 extracted trends (segments) for the fund time series, 51 for the WIG time series and 49 for the WIG 20 time series.

We can describe each time series independently using linguistic summaries of time series (cf. Kacprzyk, Wilbik, Zadrożny [1, 2, 3, 4, 5, 6] or Kacprzyk, Wilbik [7, 8, 9]). As those data sets serve only the purpose of example and are small, we applied the granulation with 3 labels for each feature only, namely *decreasing*, *constant* and *increasing* for the dynamics of change; *short*, *medium length* and *long* for duration and *low*, *moderate* and *high* for the variability.

The summaries obtained for the fund time series along with their truth values $\mathcal{T}$ are shown in Table 1.

Table 1: Linguistic summaries for the mutual fund time series

linguistic summary	$\mathcal{T}$
Among all y, most are constant	1
Among all medium y, most are constant	1
Among all moderate y, most are constant	1
Among all low y, most are constant	1
Among all medium and moderate y, most are constant	1
Among all medium and low y, most are constant	1
Among all long y, most are constant	1
Among all long and moderate y, most are constant	1
Among all short y, most are low	0.8107
Among all short y, most are constant	0.7819
Among all moderate y, most are medium	0.7462
Among all moderate y, most are medium and constant	0.7462
Among all long y, most are constant and moderate	0.7302
Among all long y, most are moderate	0.7302
Among all long and constant y, most are moderate	0.7302

The summaries obtained for the WIG index time series along with their truth values $\mathcal{T}$ are shown in Table 2.

Table 2: Linguistic summaries for the WIG index time series

linguistic summary	$\mathcal{T}$
Among all y, most are constant	1
Among all medium y, most are constant	1
Among all low y, most are constant	1
Among all moderate y, most are constant	1
Among all medium and low y, most are constant	1
Among all medium and moderate y, most are constant	1
Among all long y, most are constant	1
Among all short and moderate y, most are constant	1
Among all short y, most are constant	0.9950
Among all short and low y, most are constant	0.7856

The summaries obtained for the WIG20 index time series along with their truth values $\mathcal{T}$ are shown in Table 3.

Table 3: Linguistic summaries for the WIG20 index time series

linguistic summary	$\mathcal{T}$
Among all y, most are constant	1
Among all y, most are constant and low	1
Among all y, most are low	1
Among all constant y, most are low	1
Among all low y, most are constant	1
Among all medium y, most are constant	1
Among all medium y, most are constant and low	1
Among all medium y, most are low	1
Among all medium and constant y, most are low	1
Among all medium and low y, most are constant	1
Among all short y, most are low	1
Among all short and constant y, most are low	1
Among all long y, most are constant	1
Among all long y, most are constant and low	1
Among all long y, most are low	1
Among all long and constant y, most are low	1
Among all long and low y, most are constant	1
Among all short and low y, most are constant	0.9713
Among all short y, most are constant	0.9465
Among all short y, most are constant and low	0.8850
Among all constant y, most are medium	0.7333

The degree of similarity between the fund quotation time series and the WIG time series is equal 0.8622, while the degree of similarity of the fund quotation time series and the WIG20 time series is equal to 0.7529. This result clearly confirms our intuition. This may be viewed, on the one hand, as a proof of usefulness of our method, and on the other hand is quite natural because the investment fund considered in based on the WIG index. Luckily enough, the similarity with the WIG20 index, that of the biggest and most liquid companies, is also at a similarly high level.

6 Concluding remarks

We have proposed a new, human consistent method for the evaluation of similarity of time series, within our approach to the linguistic summarizatioon of time series that uses a fuzzy quantifier based aggregation of trends. The results obtained are very intuitively appealing. This is justified by the analysis of similarity between quotations of an investment fund and the two main indexes of the Warsaw Stock Exchange, WIG and WIG20. It also seems that a relation to natural language generation (NLG), along the lines of Kacprzyk and Zadrożny [37], may be very promising.

References

[1] J. Kacprzyk, A. Wilbik, and S. Zadrożny. Linguistic summarization of trends: a fuzzy logic based approach. In *Proceedings of the 11th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 2166–2172, 2006.

[2] J. Kacprzyk, A. Wilbik, and S. Zadrożny. On some types of linguistic summaries of time series. In *Proceedings of the 3rd International IEEE Conference "Intelligent Systems"*, pages 373–378. IEEE Press, 2006.

[3] J. Kacprzyk, A. Wilbik, and S. Zadrożny. Linguistic summarization of time series by using the choquet integral. In P. Melin, O. Castillo, L.T. Aguilar, J. Kacprzyk, and W. Pedrycz, editors,

ISBN: 978-989-95079-6-8

Foundations of Fuzzy Logic and Soft Computing - IFSA 2007, pages 284–294. Springer-Verlag, Berlin and Heidelberg, 2007.

[4] J. Kacprzyk, A. Wilbik, and S. Zadrożny. Linguistic summaries of time series via an owa operator based aggregation of partial trends. In *Proceedings of the FUZZ-IEEE 2007 IEEE International Conference on Fuzzy Systems*, pages 467–472. IEEE Press, 2007.

[5] J. Kacprzyk, A. Wilbik, and S. Zadrożny. Mining time series data via linguistic summaries of trends by using a modified sugeno integral based aggregation. In *IEEE Symposium on Computational Intelligence and Data Mining CIDM 2007*, pages 467–472. IEEE Press, 2007.

[6] J. Kacprzyk, A. Wilbik, and S. Zadrożny. Linguistic summarization of time series using a fuzzy quantifier driven aggregation. *Fuzzy Sets and Systems*, 159(12):1485–1499, 2008.

[7] J. Kacprzyk and A. Wilbik. An extended, specificity based approach to linguistic summarization of time series. In *Proceedings of the 12th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 551–559, 2008.

[8] J. Kacprzyk and A. Wilbik. A new insight into the linguistic summarization of time series via a degree of support: Elimination of infrequent patterns. In D. Dubois, M.A. Lubiano, H. Prade, M. A. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Soft Methods for Handling Variability and Imprecision*, pages 393–400. Springer-Verlag, Berlin and Heidelberg, 2008.

[9] J. Kacprzyk and A. Wilbik. Linguistic summarization of time series using linguistic quantifiers: augmenting the analysis by a degree of fuzziness. In *Proceedings of 2008 IEEE World Congress on Computational Intelligence*, pages 1146–1153. IEEE Press, 2008.

[10] Subject: Poll results: Top data mining applications in 2008 and trends. http://www.kdnuggets.com/news/2009/n01/1i.html.

[11] Past performance does not predict future performance. www.freemoneyfinance.com/2007/01/past_performanc.html.

[12] Past performance is not everything. www.personalfn.com/detail.asp? date=9/1/2007&story=3.

[13] New year's eve:past performance is no indication of future return. stockcasting.blogspot.com/2005/12/new-years-evepast-performance-is-no.html.

[14] R. Myers. Using past performance to pick mutual funds. *Nation's Business*, Oct, 1997. findarticles.com/p/articles/mi_m1154/is_n10_v85/ai_19856416.

[15] J. C. Bogle. *Common Sense on Mutual Funds: New Imperatives for the Intelligent Investor*. Wiley, New York, 1999.

[16] U.S. Securities and Exchange Commission. Mutual fund investing: Look at more than a fund's past performance. www.sec.gov/investor/pubs/mfperform.htm.

[17] A 10-step guide to evaluating mutual funds. www.personalfn.com/detail.asp?date=5/18/2007&story=2.

[18] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra. Locally adaptive dimensionality reduction for indexing large time series databases. In *Proceedings of ACM SIGMOD Conference on Management of Data*, pages 151–162, Santa Barbara, CA, 2001.

[19] E. Keogh and C. A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge Information Systems*, 7(3):358–386, 2005.

[20] P. Geurts. Pattern extraction for time series classification. In *PKDD '01: Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery*, London, UK, 2001. Springer-Verlag.

[21] B. Chiu, E. Keogh, and S. Lonardi. Probabilistic discovery of time series motifs. In *Proceedings of the the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 493–498, 2003.

[22] A. Mueen, E. Keogh, Q. Zhu, S. Cash, and B. Westover. Exact discovery of time series motifs. In *SDM 2009, in press*, 2009.

[23] R. Agrawal, C. Faloutsos, and A. Swami. Efficient similarity search in sequence databases. In *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms*, pages 69–84, 1993.

[24] G. Das, D. Gunopulos, and H. Mannila. Finding similar time series. In Jan Komorowski and Jan Zytkow, editors, *Principles of Data Mining and Knowledge Discovery (PKDD'97)*, pages 88–100. Springer-Verlag, 1997.

[25] N. Nakatsu, Y. Kambayashi, and S. Yajima. A longest common subsequence algorithm suitable for similar text strings. *Acta Informatica*, 18:171–179, 1982.

[26] K.-P. Chan and W.-C. Fu. Efficient time series matching by wavelets. In *Proceedings of the 15th International Conference on Data Engineering*. IEEE Computer Society, 1999.

[27] H. Shatkay and S. B. Zdonik. Approximate queries and representations for large data sequences. In *Proceedings 12th International Conference on Data Engineering*, pages 536–545, 1996.

[28] J. Kacprzyk and M. Fedrizzi. 'soft' consensus measures for monitoring real consensus reaching processes under fuzzy preferences. *Control and Cybernetics*, 15:309–323, 1986.

[29] J. Kacprzyk and M. Fedrizzi. A 'soft' measure of consensus in the setting of partial (fuzzy) preferences. *European Journal of Operational Research*, 34:315–325, 1988.

[30] J. Kacprzyk and M. Fedrizzi. A 'human-consistent' degree of consensus based on fuzzy logic with linguistic quantifiers. *Mathematical Social Sciences*, 18:275–290, 1989.

[31] J. Kacprzyk, M. Fedrizzi, and H. Nurmi. Group decision making and consensus under fuzzy preferences and fuzzy majority. *Fuzzy Sets and Systems*, 49:21–31, 1992.

[32] J. Kacprzyk and S. Zadrożny. Towards a general and unified characterization of individual and collective choice functions under fuzzy and nonfuzzy preferences and majority via the ordered weighted average operators. *International Journal of Intelligent Systems*, 24(1):4–26, 2009.

[33] E. Keogh, S. Chu, D. Hart, and M. Pazzani. An online algorithm for segmenting time series. In *Proceedings of the 2001 IEEE International Conference on Data Mining*, 2001.

[34] E. Keogh, S. Chu, D. Hart, and M. Pazzani. Segmenting time series: A survey and novel approach. In M. Last, A. Kandel, and H. Bunke, editors, *Data Mining in Time Series Databases*. World Scientific Publishing, 2004.

[35] J. Sklansky and V. Gonzalez. Fast polygonal approximation of digitized curves. *Pattern Recognition*, 12(5):327–331, 1980.

[36] E. Keogh and M. Pazzani. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*, pages 239–241, New York, NY, 1998.

[37] J. Kacprzyk and S. Zadrożny. Data mining via protoform based linguistic summaries: Some possible relations to natural language generation. In *2009 IEEE Symposium Series on Computational Intelligence Proceedings*, pages 217–224, Nashville, TN, 2009.

ISBN: 978-989-95079-6-8

Exchanging iterated expectations of random upper semicontinuous functions: an application to decision theory

Miguel López-Díaz Luis J. Rodríguez-Muñiz

Department of Statistics, Operations Research and Mathematical Teaching, University of Oviedo
c/ Calvo Sotelo s/n, E33071 Oviedo (Spain)
Email: {mld,luisj}@uniovi.es

Abstract— *In this paper we present a procedure to deal with a kind of single-stage decision problems with imprecise utilities. In this type of problems the product measurability of the utility function is not required. So that, the involved expectations are calculated by means of iterated integrals instead of integrals over product spaces.*

Keywords— Bayesian decision analysis, Iterated expectation, Kudo-Aumann's integral, Random upper semicontinuous function, Uncertainty modeling.

1 Introduction

There are situations in which product measurability of certain mappings is not satisfied, but iterated integrals are well-defined (see [10]). This type of situations appear in single-stage decision problems where we need conditions which allows us to exchange iterated expectations in order to perform a Bayesian analysis. This paper studies the case of Bayesian analysis of single-stage decision problems with imprecise utilities and non product-measurable utility funtion.

Several studies have been developed before to evaluate imprecise utilities, see for instance Watson *et al.* [27], Tong and Bonissone [26], Dubois aand Prade [7, 8, 9], Gil and Jain [11], Billot [1], Chen and Klein [3], Gil *et al.* [13], Krätschmer [16], Bordley [2], Rébillé [22] or Rodríguez-Muñiz and López-Díaz [25].

Here we model the imprecise utilities by means of fuzzy-valued utility functions (based on the concept of random upper semicontinuous functions or fuzzy random variable). However, only in [25] the non-product measurable case has been analyzed for this type of furrzy utilities, based on the theoretical results on iterated integrals of random upper semicontinuous functions given in [24].

In this paper we gather theoretical and applied results about how to deal with the type of problems referred above.

The paper is organized as follows: Preliminaries and notation constitutes Section 2, Section 3 contains theoretical results and Section 4 gather the statistical decision analysis results.

2 Preliminaries

Let us consider $\mathcal{K}_c$ the class of nonempty compact convex subsets of $\mathbb{R}$, endowed with a semilinear structure by means of the Minkowski addition and the product by a scalar. Also, consider the *Hausdorff metric* on $\mathcal{K}_c$ (see [6]). On a measurable space $(\Omega, \mathcal{A})$ we can define $S : \Omega \to \mathcal{K}_c$ a randon set as a $\mathcal{A}|\mathcal{B}_{d_H}$-measurable mapping ([14]).

A random set S is said to be *integrably bounded with respect to measure* μ, if $\|S\| \in L^1(\Omega, \mathcal{A}, \mu)$, where $\|S\|(\omega) = \sup_{x \in S(\omega)} \|x\|$. The *integral*, or *expected value* in case of μ being a probability, of S, is given by the *Kudo-Aumann integral* and it will be denoted either by $\int_\Omega S(\omega)\, d\mu(\omega)$ or $E(S|\mu)$ ([14]).

Let $\mathcal{F}_c$ denote the class of upper semicontinuous functions or *fuzzy sets* $U : \mathbb{R} \to [0,1]$ such that $U_\alpha \in \mathcal{K}_c$ for all $\alpha \in [0,1]$, where $U_\alpha = \{x \in \mathbb{R} : U(x) \geq \alpha\}$ for $\alpha \in (0,1]$, and $U_0 = \mathrm{cl}\,\{x \in \mathbb{R} : U(x) > 0\}$, cl denoting the topological closure.

The class $\mathcal{F}_c$ can be endowed with a semilinear structure, where addition and product by a scalar can be defined by means of Zadeh's extension principle ([28, 19]). On $\mathcal{F}_c$ we consider the d_∞ metric ([19]). The *magnitude* of $U \in \mathcal{F}_c$ is given by $\|U\| = d_\infty(U, \mathbf{1}_{\{0\}}) = d_H(U_0, \{0\})$.

Given a measurable space $(\Omega, \mathcal{A})$, a mapping $X : \Omega \to \mathcal{F}_c$ is said to be a *random upper semicontinuous function* (r.u.s.f. for short) if $X_\alpha : \Omega \to \mathcal{K}_c$ with $X_\alpha(\omega) = (X(\omega))_\alpha$ for all $\omega \in \Omega$, is a random set for all $\alpha \in [0,1]$ ([21, 4]).

A r.u.s.f. X is said to be *integrably bounded* with respect to a measure $\mu : \mathcal{A} \to \mathbb{R}$, if the mapping $\|X\| \in L^1(\Omega, \mathcal{A}, \mu)$, where $\|X\| : \Omega \to \mathbb{R}$ is given by $\|X\|(\omega) = \|X(\omega)\|$ for all $\omega \in \Omega$.

For an integrably bounded r.u.s.f., in [21] is defined its *integral*, denoted by $\int_\Omega X(\omega)\, d\mu(\omega)$ or $E(X|\mu)$, as the unique set in $\mathcal{F}_c$ such that $E(X|\mu)_\alpha = E\big(X_\alpha \big| \mu\big)$ for every $\alpha \in [0,1]$. When $\Omega = [a,b]$, we will use also the notation $\int_a^b X(\omega)\, d\mu(\omega)$.

If μ is a probability measure, an r.u.s.f. is also referred to as a *fuzzy random variable* and its integral as *the fuzzy expected value of* X.

It is possible to extend to upper semicontinuous functions the concept of *Hukuhara difference* or *Minkowski difference* for subsets ([15, 20]), so given $U, V \in \mathcal{F}_c$, its *Hukuhara difference*, denoted by $U -_h V$, is the set $W \in \mathcal{F}_c$ (if it exists) such that $U = V + W$.

Let T be a nonempty open subset of $\mathbb{R}$. A mapping $G : T \to \mathcal{F}_c$ is said to be *Hukuhara differentiable* at $t_0 \in T$ if there exists $G'(t_0) \in \mathcal{F}_c$, which is called the *Hukuhara differential of* G *at* t_0, such that

$$\lim_{\Delta t \to 0^+} d_\infty\left(\frac{G(t_0 + \Delta t) -_h G(t_0)}{\Delta t}, G'(t_0)\right)$$
$$= \lim_{\Delta t \to 0^+} d_\infty\left(\frac{G(t_0) -_h G(t_0 - \Delta t)}{\Delta t}, G'(t_0)\right) = 0.$$

The above definition ([20, 23]) is an extension of the Hukuhara's differentiability criterion for set-valued mappings ([15]).

ISBN: 978-989-95079-6-8

If a mapping G depends on more than one argument, we will make use of the usual symbol of partial derivative to indicate with respect to which argument the Hukuhara differential is considered.

Throughout the paper, for any set $\Omega \subset \mathbb{R}^k$ with $k \in \mathbb{N}$, $\mathcal{B}_\Omega$ will denote the Borel σ-field on Ω. Given $(\Omega, \mathcal{B}_\Omega)$ and $m_1, m_2 : \mathcal{B}_\Omega \to [0, \infty]$ two σ-finite measures, $m_1 \ll m_2$ will indicate that m_1 is absolutely continuous with respect to m_2, and $\frac{dm_1}{dm_2}$ will denote a Radon-Nikodym derivative of m_1 with respect to m_2. If it is supposed that there exists a continuous Radon-Nikodym derivative, then $\frac{dm_1}{dm_2}$ will denote this particular function.

As we will model the imprecise utilities by means of fuzzy-valued functions it should be necessary to rank fuzzy sets. For this purpose the ranking criterion introduced by De Campos and González [5] will be considered. Therefore, we will say that $U \in \mathcal{F}_c$ is *greater than or equal to* $W \in \mathcal{F}_c$ *in the* λ, μ*-average sense*, we will denote it by $U \geq_{\lambda,\mu} W$, if $V_\mu^\lambda(U) \geq V_\mu^\lambda(W)$, where $\lambda \in [0,1]$ represents a kind of degree of optimism/pessimism and μ is a measure on $[0,1]$ (see [18] for more details).

3 Theoretical results

We include in this section those theoretical results regarding the exchange of iterated integrals that are necessary to prove the applied results in Section 4.

The first result is about differentiating under the integral sign, by using Hukuhara derivative.

Proposition 3.1 *Let $(\Omega, \mathcal{A}, P)$ be a probability space with $\Omega \subset \mathbb{R}^k$, and let m denote the Borel measure on the interval $[a,b]$. For every $\omega \in \Omega$, let P_ω be a probability measure on $([a,b], \mathcal{B}_{[a,b]})$ with $P_\omega \ll m$, such that there exists a continuous Radon-Nikodym derivative.*

If $X : \Omega \times [a,b] \to \mathcal{F}_c$ satisfies the following conditions:

i) for every $\omega \in \Omega$, the mapping $X_\omega : [a,b] \to \mathcal{F}_c$, with $X_\omega(t) = X(\omega, t)$, is an integrably bounded r.u.s.f. with respect to P_ω, and X_ω is continuous $a.s.\,[P]$,

ii) there exists $h \in L^1(\Omega, \mathcal{A}, P)$, such that $\left\|X(\omega,t)\frac{dP_\omega}{dm}(t)\right\| \leq h(\omega)\ a.s.\,[P]$ for every $t \in [a,b]$, and the mapping $\omega \mapsto X(\omega,t)\frac{dP_\omega}{dm}(t)$ is continuous $a.e.\,[m]$,

iii) there exists $g \in L^1([a,b], \mathcal{B}_{[a,b]}, m)$ with $\left\|X(\omega,t)\frac{dP_\omega}{dm}(t)\right\| \leq g(t)\ a.e.\,[m]$ for every $\omega \in \Omega$,

then, the mapping

$$t \in [a,b] \mapsto \int_\Omega \Big(\int_a^t X(\omega,s)\, dP_\omega(s) \Big)\, dP(\omega)$$

is Hukuhara differentiable on (a,b), and for every $t \in (a,b)$ it holds that

$$\frac{\partial}{\partial t} \int_\Omega \left(\int_a^t X(\omega,s)\, dP_\omega(s) \right)\, dP(\omega)$$

$$= \int_\Omega X(\omega,t) \frac{dP_\omega}{dm}(t)\, dP(\omega).$$

Starting from last result, we can go one step further to prove a exchange of iterated integrals:

Theorem 3.2 *Let $(\Omega, \mathcal{B}_\Omega, P)$ be a probability space with $\Omega \subset \mathbb{R}^k$ and let m denote the Borel measure on the interval $T = [a,b]$. For every $t \in T$, let P_t be a probability measure on $(\Omega, \mathcal{B}_\Omega)$ such that $P_t \ll P$ and there exists a continuous Radon-Nikodym derivative. For every $\omega \in \Omega$, let P_ω be a probability on $(T, \mathcal{B}_T)$ such that $P_\omega \ll m$ and there exists a continuous Radon-Nikodym derivative.*

Let $X : \Omega \times T \to \mathcal{F}_c$ be a mapping satisfying that:

i) for every $t \in T$, X_t is an integrably bounded r.u.s.f. with respect to P_t,

ii) for every $\omega \in \Omega$, X_ω is an integrably bounded r.u.s.f. with respect to P_ω and it is continuous $a.s.\,[P]$,

iii) there exists $h_1 \in L^1(\Omega, \mathcal{B}_\Omega, P)$ such that $\left\|X(\omega,t)\frac{dP_\omega}{dm}(t)\right\| \leq h_1(\omega)\ a.s.\,[P]$ for every $t \in T$, and the mapping $\omega \mapsto X(\omega,t)\frac{dP_\omega}{dm}(t)$ is continuous $a.e.\,[m]$,

iv) there exists a mapping $g \in L^1([a,b], \mathcal{B}_{[a,b]}, m)$ such that for every $\omega \in \Omega$, $\|X(\omega,t)\frac{dP_\omega}{dm}(t)\| \leq g(t)\ a.e.\,[m]$ for every $\omega \in \Omega$,

v) the mapping $t \mapsto X(\omega,t)\frac{dP_t}{dP}(\omega)$ is continuous on T $a.s.\,[P]$,

vi) there exists $h_2 \in L^1(\Omega, \mathcal{B}_\Omega, P)$ such that $\|X(\omega,t)\frac{dP_t}{dP}(\omega)\| \leq h_2(\omega)\ a.s.\,[P]$ for every $t \in T$.

Let m' be a probability measure on $(T, \mathcal{B}_T)$ such that $m' \ll m$ and there exists a continuous Radon-Nikodym derivative. If for every $t \in T$, the equality

$$\frac{dP_\omega}{dm}(t) = \frac{dP_t}{dP}(\omega)\frac{dm'}{dm}(t) \quad a.s.\,[P]$$

holds, then

$$\int_\Omega \left(\int_a^t X(\omega,s)\, dP_\omega(s) \right) dP(\omega)$$

$$= \int_a^t \left(\int_\Omega X(\omega,s)\, dP_s(\omega) \right) dm'(s)$$

for every $t \in T$.

And also an unbounded version of previous theorem can be obtained.

Theorem 3.3 *Assume the conditions in Theorem 3.2 with the interval T being not necessarily bounded, and suppose that there exists $g' \in L^1(\Omega, \mathcal{B}_\Omega, P)$ such that*

$$\int_T \|X(\omega,s)\| dP_\omega(s) \leq g'(\omega)\ \ a.s.\,[P]\,.$$

Then, the following equality holds,

$$\int_\Omega \left(\int_T X(\omega,s)\, dP_\omega(s) \right) dP(\omega)$$

$$= \int_T \left(\int_\Omega X(\omega,s)\, dP_s(\omega) \right) dm'(s)\,.$$

It should be remarked that the conditions in Theorems 3.2 and 3.3 do not imply that X is an r.u.s.f. on the product measurable space as is illustrated in [24].

ISBN: 978-989-95079-6-8

4 Applied results

In this section we are using the theoretical results obtained in Section 3 to obtain applied results to the single-stage decision problem under non-product measurability conditions. Obviously, different conditions than in the Fubini-applicable case will be obtained, but it is again very important to underline that under the conditions stated here product measurability is not implied.

We first introduce the concept of fuzzy utility function considered in this communication. We will use the following notation: Θ is the state space and it will be considered an interval of $\mathbb{R}$, $\mathcal{B}_\Theta$ is the Borel σ-field on Θ, with m the Borel measure and $\mathbf{A}$ is the action space.

Definition 4.1 *A mapping $U : \Theta \times \mathbf{A} \to \mathcal{F}_c$ is said to be a fuzzy utility function on $\Theta \times \mathbf{A}$ if*

i) for every $a \in \mathbf{A}$, the projection $U_a : \Theta \to \mathcal{F}_c$ is an r.u.s.f. on $(\Theta, \mathcal{B}_\Theta)$,

ii) for every pair $a_1, a_2 \in \mathbf{A}$, a_1 will be considered preferred or indifferent to a_2 with respect to a probability distribution ξ on $(\Theta, \mathcal{B}_\Theta)$, if $E(U_{a_1}|\xi) \geq_{\lambda,\mu} E(U_{a_2}|\xi)$ (for fixed $\lambda \in [0,1]$ and measure μ).

The decision problem with fuzzy utilities will be denoted by $(\Theta, \mathbf{A}, U)$.

On the other hand, it will be considered a Bayesian context, so the existence of a probability distribution π on $(\Theta, \mathcal{B}_\Theta)$, *the prior distribution*, will be assumed. Then the "value" of the decision problem will be the fuzzy value $E(U_{a^\pi}|\pi)$, where a^π is a *prior Bayes action* in the λ, μ-average sense, this is, $a^\pi \in \mathbf{A}$ verifies $E(U_{a^\pi}|\pi) \geq_{\lambda,\mu} E(U_a|\pi)$ for all $a \in \mathbf{A}$.

Similar to the case of real-valued utilities, it is useful for increasing the expected utility to incorporate *sample information*. Let $\mathbf{X}$ be a statistical experiment characterized by a probability space $(\mathbb{X}, \mathcal{B}_\mathbb{X}, P_\theta)$, where $\theta \in \Theta$, $\mathcal{B}_\mathbb{X}$ is the Borel σ-field on $\mathbb{X} \subset \mathbb{R}^k$ and the experimental distribution P_θ depends on the true unknown state θ. We will denote by P the *marginal (predictive)* distribution of the experiment.

After the experiment is performed, if $\mathbf{X} = x$ is the available sample information, the fuzzy expected utility associated with an action $a \in \mathbf{A}$ is given by $E(U_a|\pi_x)$, where π_x is the *posterior distribution* of θ given $\mathbf{X} = x$, obtained on the basis of Bayes' formula. So, a *posterior Bayes action* is any $a^{\pi_x} \in \mathbf{A}$ such that $E(U_{a^{\pi_x}}|\pi_x) \geq_{\lambda,\mu} E(U_a|\pi_x)$ for every $a \in \mathbf{A}$.

In order to generalize the choice of an action for each possible sample, the concept of *decision rule,* as a mapping from $\mathbb{X}$ to $\mathbf{A}$ satisfying several conditions (based on Theorems 3.2 and 3.3) is formalized. These conditions will allow the proper extension of the normal and extensive forms of the Bayesian analysis.

Definition 4.2 *Let $(\mathbb{X}, \mathcal{B}_\mathbb{X}, P_\theta)$ be the probability space of a statistical experiment $\mathbf{X}$ associated with the decision problem $(\Theta, \mathbf{A}, U)$. A decision rule is a mapping $d : \mathbb{X} \to \mathbf{A}$ satisfying that*

i) for every $\theta \in \Theta$, $U(\theta, d()) : \mathbb{X} \to \mathcal{F}_c$ is an integrably bounded r.u.s.f. with respect to P_θ,

ii) for every $x \in \mathbb{X}$, $U(, d(x)) : \Theta \to \mathcal{F}_c$ is an integrably bounded r.u.s.f. with respect to π_x, moreover, it is continuous a.s. $[P]$,

iii) there exists $h_1 \in L^1(\mathbb{X}, \mathcal{B}_\mathbb{X}, P)$ such that $\left\|U(\theta, d(x))\frac{d\pi_x}{dm}(\theta)\right\| \leq h_1(x)$ a.s. $[P]$ for every $\theta \in \Theta$, and the mapping $x \mapsto U(\theta, d(x))\frac{d\pi_x}{dm}(\theta)$ is continuous a.e. $[m]$,

iv) there exists $g \in L^1(\Omega, \mathcal{B}_\Omega, m)$ such that for every $x \in \mathbb{X}$, it holds that $\|U(\theta, d(x))\frac{d\pi_x}{dm}(\theta)\| \leq g(\theta)$ a.e. $[m]$ for every $x \in \mathbb{X}$,

v) the mapping $\theta \mapsto U(\theta, d(x))\frac{dP_\theta}{dP}(x)$ is continuous on Θ a.s. $[P]$,

vi) there exists $h_2 \in L^1(\mathbb{X}, \mathcal{B}_\mathbb{X}, P)$ such that $\left\|U(\theta, d(x))\frac{dP_\theta}{dP}(x)\right\| \leq h_2(x)$ a.s. $[P]$ for every $\theta \in \Theta$,

vii) there exists $g' \in L^1(\mathbb{X}, \mathcal{B}_\mathbb{X}, P)$ with $\int_\Theta \|U(\theta, d(x))\| \, d\pi_x(\theta) \leq g'(x)$.

Now, on the one hand we can consider the *normal Bayesian analysis*. In this case we should find a *Bayes decision rule*, that is, a rule d_B such that

$$\int_\Theta \left(\int_\mathbb{X} U(\theta, d_B(x)) \, dP_\theta(x) \right) d\pi(\theta)$$

$$\geq_{\lambda,\mu} \int_\Theta \left(\int_\mathbb{X} U(\theta, d(x)) \, dP_\theta(x) \right) d\pi(\theta)$$

for every decision rule d. In this case, the "value" of the problem is

$$\int_\Theta \left(\int_\mathbb{X} U(\theta, d_B(x)) \, dP_\theta(x) \right) d\pi(\theta). \qquad (1)$$

On the other hand, we can consider the *extensive Bayesian analysis*. We should obtain for each sample information x a posterior Bayes action a^{π_x}, and consider the decision rule which associates with each x an action a^{π_x}. In this analysis, the "value" of the experiment $\mathbf{X}$ is quantified by the fuzzy expected terminal utility, defined as follows:

Definition 4.3 *Given $(\Theta, \mathbf{A}, U)$ a decision problem and $\mathbf{X} = (\mathbb{X}, \mathcal{B}_\mathbb{X}, P_\theta)$, an associated experiment, the fuzzy expected terminal utility of $\mathbf{X}$ is given by*

$$U_t(\mathbf{X}) = \int_\mathbb{X} \left(\int_\Theta U(\theta, a^{\pi_x}) \, d\pi_x(\theta) \right) dP(x). \qquad (2)$$

We are now showing the equivalence between the two forms of the Bayesian analysis in the sense that (1) and (2) are equal in the λ, μ-average sense. Previously, the following result for the exchange of the integrals is stated.

Theorem 4.4 *Let $(\Theta, \mathbf{A}, U)$ be a decision problem, let $\Theta \subset \mathbb{R}$ and let π be a prior probability on $(\Theta, \mathcal{B}_\Theta)$ such that $\pi \ll m$ with a continuous Radon-Nikodym derivative. Let $\mathbf{X} = (\mathbb{X}, \mathcal{B}_\mathbb{X}, P_\theta)$ be an associated experiment, and let P be the marginal distribution. For every $\theta \in \Theta$, suppose that $P_\theta \ll P$ and there exists a continuous Radon-Nikodym derivative. For every $x \in \mathbb{X}$, let π_x be the posterior distribution on $(\Theta, \mathcal{B}_\Theta)$ such that $\pi_x \ll m$ with a continuous Radon-Nikodym derivative.*

ISBN: 978-989-95079-6-8

If for every $\theta \in \Theta$*, it holds that* $\frac{d\pi_x}{dm}(\theta) = \frac{dP_\theta}{dP}(x)\frac{d\pi}{dm}(\theta)$ $a.s.\,[P]$*, then*

$$\int_{\mathbb{X}} \left(\int_{\Theta} U(\theta, d(x))\, d\pi_x(\theta) \right) dP(x)$$
$$= \int_{\Theta} \left(\int_{\mathbb{X}} U(\theta, d(x))\, dP_\theta(x) \right) d\pi(\theta)$$

whatever the decision rule $d : \mathbb{X} \to \mathbf{A}$ *may be.*

As a consequence, the following key result, which states the equivalence between the normal and extensive forms of Bayesian analysis, is obtained.

Theorem 4.5 *Assume the conditions of Theorem 4.4. Let us consider the mapping which associates with each sample* $x \in \mathbb{X}$ *a posterior Bayes action* a^{π_x}*. If this mapping satisfies the definition of decision rule, then it is a Bayes decision rule. Moreover,* $U_t(\mathbf{X})$ *is equal, in the* λ, μ*-average sense, to the fuzzy expected utility associated with any Bayes decision rule, this is*

$$U_t(\mathbf{X}) =_{\lambda,\mu} \int_{\mathbb{X}} \left(\int_{\Theta} U(\theta, d_B(x))\, d\pi_x(\theta) \right) dP(x)$$

Thus, the fuzzy expected terminal utility, so calculated, can be interpreted as the "value" of the decision problem once the experiment $\mathbb{X}$ is performed and one Bayes decision rule is calculated. This lead us not only to express the information of the problem by a value but also as a criterium to rank experiments in order to obtain the more informative.

5 Conclusions

By using a theoretical result about exchanging iterated integrals of $[0,1]^{\mathbb{R}}$-valued r.u.s.f., the model established in this paper provides a framework for single-stage decision problems in which both forms of Bayesian analysis (normal and extensive) are proved to be equivalent when imprecise utilities are not necessarily product measurable. Thus, these results together with those in [12, 17, 13] cover most of the situations which one can found when analyzing single-stage decision problems with imprecise utilities modeled by fuzzy random variables.

Acknowledgment

Authors want to acknowledge the financial support by Grant MTM2008-01519 from Ministry of Science and Innovation, Government of Spain.

References

[1] A. Billot. An existence theorem for fuzzy utility functions: A new elementary proof. *Fuzzy Sets and Systems*, 74 271-276, 1995

[2] F. Bordley. Reformulating decision theory using fuzzy set theory and Shafer's theory of evidence. *Fuzzy Sets and Systems*, 139: 243-266, 2003.

[3] C.B. Chen, C.M. Klein. A simple approach to ranking a group of aggregated fuzzy utilities. *IEEE Transactions on Systems, Man, and Cybernetics* 27: 26-35, 1997.

[4] A. Colubi, J.S. Domínguez-Menchero, M. López-Díaz, D.A. Ralescu. A $D_E[0,1]$ representation of random upper semicontinuous functions. *Proceedings of the American Mathematical Society*, 130: 3237-3242, 2002.

[5] L.M. De Campos, A. González. A subjective approach for ranking fuzzy numbers. *Fuzzy Sets and Systems*, 29: 145-153, 1989.

[6] G. Debreu. Integration of correspondences, *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability*, 1965/66, Vol. II: Contributions to Probability Theory, Part 1. University of California Press, Berkeley, pp. 351-372, 1967.

[7] D. Dubois, H. Prade. Additions of interactive fuzzy numbers. *IEEE Transactions on Automatic Control*, 26: 926-936, 1981.

[8] D. Dubois, H. Prade. *The use of fuzzy numbers in decision analysis. Fuzzy Information and Decision Processes*. North-Holland, Amsterdam, 309-321, 1982.

[9] D. Dubois, H. Prade. *Possibility Theory. An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.

[10] H. Friedman. A consistent Fubini-Tonelli theorem for nonmeasurable functions. *Illinois Jouranl of Mathematics*, 24: 390-395, 1980.

[11] M.A. Gil, P. Jain. Comparison of experiments in statistical decision problems with fuzzy utilities. *IEEE Transactions on Systems, Man, and Cybernetics* 22: 662-670, 1992.

[12] M.A. Gil, M. López-Díaz. Fundamentals and Bayesian analyses of decision problems with fuzzy-valued utilities. *International Journal of Approximate Reasoning*, 15: 203-224, 1996.

[13] M.A. Gil, M. López-Díaz, L.J. Rodríguez-Muñiz. An improvement of a comparison of experiments in statistical decision problems with fuzzy utilities. *IEEE Transactions on Systems, Man, and Cybernetics*, 28: 856-864, 1998.

[14] F. Hiai, H. Umegaki. Integrals, conditional expectations and martingales of multivalued functions. *Journal of Multivariate Analysis*, 7: 149-182, 1997

[15] M. Hukuhara. Intégration des applications mesurables dont la valeur est un compact convexe. *Funkcialaj Ekvacioj*, 10: 205-223, 1967.

[16] V. Krätschmer. Coherent lower previsions and Choquet integrals. *Fuzzy Sets and Systems*, 138: 469-484, 2003.

[17] M. López-Díaz, M.A. Gil. Reversing the order of integration in iterated expectations of fuzzy random variables, and statistical applications. *Journal of Statistical Planning and Inference*, 74: 11-29, 1998.

[18] M. López-Díaz, M.A. Gil. The λ-average value and the fuzzy expectation of a fuzzy random variable. *Fuzzy Sets and Systems*, 99: 347-352, 1998.

[19] M.L. Puri, D.A. Ralescu. Différentielle d'une fonction floue. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique* 293: 237-239, 1981.

[20] M.L. Puri, D.A. Ralescu. Differentials of fuzzy functions. *Journal of Mathematical Analysis and Applications* 91: 552-558, 1983.

[21] M.L. Puri, D.A. Ralescu. Fuzzy random variables. *Journal of Mathematical Analysis and Applications*, 114: 409-422, 1986.

[22] Y. Rébillé. Decision making over necessity measures through the Choquet integral criterion. *Fuzzy Sets and Systems* 157: 3025-3039, 2006.

ISBN: 978-989-95079-6-8

[23] L.J. Rodríguez-Muñiz, M. López-Díaz. Hukuhara derivative of the fuzzy expected value. *Fuzzy Sets and Systems*, 138: 593-600, 2003.

[24] L.J. Rodríguez-Muñiz, M. López-Díaz. On the exchange of iterated expectations of random upper semicontinuous functions. *Statistics and Probability Letters*, 77: 1628-1635, 2007.

[25] L.J. Rodríguez-Muñiz, M. López-Díaz. A new framework for the Bayesian analysis of single-stage decision problems with imprecise utilities. *Fuzzy Sets and Systems* 159: 3271-3280, 2008.

[26] R.M. Tong, P.P. Bonissone. A linguistic approach to decision-making with fuzzy sets. *IEEE Transactions on Systems, Man, and Cybernetics* 10: 716-723, 1980.

[27] S.R. Watson, J.J. Weiss, M.L. Donnell. Fuzzy decision analysis. *IEEE Transactions on Systems, Man, and Cybernetics*, 9: 1-9, 1979.

[28] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. Parts I, II and III. *Information Science*, 8: 199-249; 8: 301-357; 9: 43-80, 1975.

ISBN: 978-989-95079-6-8

60 years "A Mathematical Theory of Communication" – Towards a "Fuzzy Information Theory"

Rudolf Seising

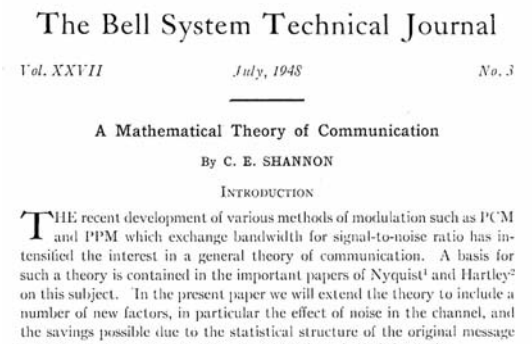
European Centre for Soft Computing
Mieres, Asturias, Spain
Email: rudolf.seising@softcomputing.es

Abstract*—When 60 years ago Shannon established "A Mathematical Theory of Communication" nobody could know the consequences for science and technology in the second half of the century. Shannon published his article in two parts in the July and October 1948 editions of the* Bell System Technical Journal*. However, it is very probable that this article wouldn't have become famous without the help of Weaver, whose popular text "The Mathematics of communication" re-interpreted Shannon's work for broader scientific audiences. Weaver's "preface" and Shannon's article were published together in the book* The Mathematical Theory of Communication *(1949) that represents the beginning of the then so-called "Information theory". However, in his "introduction" Weaver went over and above Shannon´s mathematical theory mentioning not only the technical but also the semantic and influential problems of communication. This classification is very similar to the foundations of the* Theory of Signs *(1938) that was established by Morris. This paper deals with the connectivity between this Information theory and the Theory of Fuzzy sets and systems that appeared in the first half of the 1950s.Then the paper focuses to the non-technical but philosophical aspects of information theory and it advocates a fuzzy information theory that has to be appropriate to cover the concept of information – particularly with regard to the philosophical aspects.*

***Keywords*—** communication, fuzziness, information, philosophy, semiotics, signs, signals

1 Introduction

60 years ago, in July 1948, the mathematician and electrical engineer Claude Elwood Shannon (1916-2001) published "A Mathematical Theory of Communication" in the *Bell System Technical Journal* [1] (Fig. 1) and in July of the following year "The Mathematics of Communication" by the mathematician and physicist Warren Weaver (1894-1978) appeared in the *Scientific American* [2] (Fig. 2).



The Bell System Technical Journal

Vol. XXVII July, 1948 No. 3

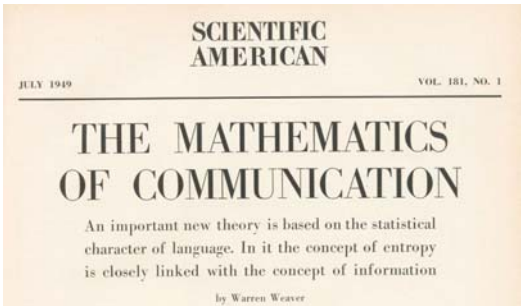
A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message

Figure 1: Title of Shannon's publication in July 1948 [1].



SCIENTIFIC AMERICAN

JULY 1949 VOL. 181, NO. 1

THE MATHEMATICS OF COMMUNICATION

An important new theory is based on the statistical character of language. In it the concept of entropy is closely linked with the concept of information

by Warren Weaver

Figure 2: Title of Weaver's publication in July 1949 [2].

However, other roots of the later so-called "Information Theory" can be found also in the *Cybernetics* of Norbert Wiener (1894-1964) [3], in the work of the Russian mathematician Andrei N. Kolmogorov (1903-1987) [4] and of the British statistician Ronald A. Fisher (1890-1962) [5].
Wiener wrote: "This idea occurred at about the same time to several writers, among them the statistician R. A. Fisher, Dr. Shannon of the Bell Telephone Laboratories, and the author. Fisher's motive in studying this subject is to be found in classical statistical theory; that of Shannon in the problem of coding information; and that of the author in the problem of noise and message in electrical filters. Let it be remarked parenthetically that some of my speculations in this direction attach themselves to the earlier work of Kolmogorov in Russia, although a considerable part of my work was done before my attention was called to the work of the Russian school." ([3], p. 18.)[1]
Shannon's "Mathematical Theory of Communication" and Wiener's *Cybernetics* had appeared simultaneously, but in both cases the publication was delayed by the war. However, Wiener mentioned the fact that, since Shannon was a Bell employee, his research projects were geared toward realizing and marketing his findings as quickly as possible, whereas Wiener, as a college professor, was able to approach his research freely. He had "found the new realm of communication ideas a fertile source of new concepts not only in communication theory, but in the study of the living organism and in many related problems".[2] In the manuscript

[1] Wiener cited Kolmogorov's article [4].
[2] Wiener in conversation with Bello (technology editor of the journal *Fortune*) on 13 October 1953, Box 4.179, MC22, Wiener Papers, MIT. Quoted from [6], p. 138.

ISBN: 978-989-95079-6-8

for his book *Invention: The Care and Feeding of Ideas*, which he never finished, Wiener wrote in 1957:
"Once I had alerted myself and the public in general to the statistical element in communication theory, conformation began to flow in from all sides. At the Bell Telephone Laboratories there was, and is, a young mathematical physicist by the name of Claude Shannon. He had already applied mathematical logic to the design of switching systems, and throughout all his work he has shown a love for the discrete problems, the problems with a small number of variable quantities, which come up in switching theory. I am inclined to believe that from the very start, a large part of his ideas in communication theory and its statistical basis were independent of mine, but whether they were or not, each of us appreciated the significance of the work of the other. The whole subject of communication began to assume a new statistical form, both for his sorts of problems and for mine. This is not the point for me to give the genealogy of every single piece of apparatus which this new statistical communication theory has fostered, but I can say that the impact of the work has gone from one end of communication theory to the other, until now there is scarcely a recent communication invention which has not been touched by statistical considerations. Thus, this whole wide spreading branch of science represents a subtle working out of concepts implicit in Gibbs and in the Lebesque-Borel team, but if I may say so, implicitly implicit, so that until some forty years had passed, no one could have seen the direction in which the earlier thought was bound to lead. This is, in my mind, a key example of a change in intellectual climate, and of the effect it has had both in discovery and in invention." ([7], p. 22f)
Today Shannon's name is associated almost unanimously with mathematical information and communication theory – these terms are often considered synonyms – although people had spoken at first of the "Wiener-Shannon Communication Theory".

Figure 3: C. E. Shannon, W. Weaver and N. Wiener.

However, it was not least Warren Weaver's article that popularized Shannon's theory and therefore it became – with little changes and the new title "Recent Contributions to the Mathematical Theory of Communication" – as an "introduction" to Shannon's text in their joint book-publication *The Mathematical Theory of Communication* [8]. Even more Weaver's manuscript was not only a brief and understandable sketch of Shannon's theory and their technical problems – in this text Weaver presented self-contained and more philosophical considerations on the semantic and influential problems with the concept of information that is in the core of Shannon's theory and that was called in the following years "Information theory". We will deal with these philosophical subjects in section 3. Beforehand we will give a few remarks on the genesis of the theory of fuzzy sets and systems, especially against the background of the developments of information theory in the 1950s.

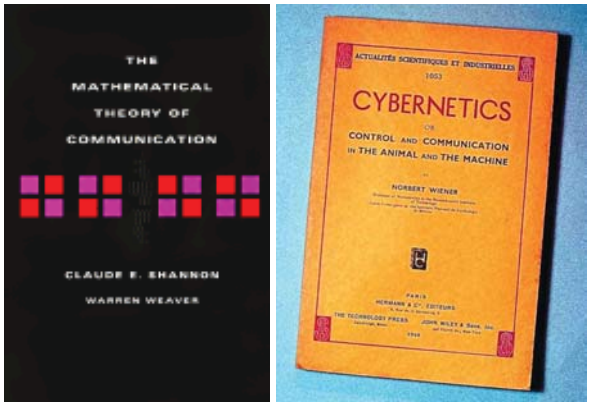


Figure 4: The books of Shannon and Weaver and Wiener.

2 Information theory and the theory of Fuzzy Sets and Systems

2.1 A Historical Sketch - in General

Inspired by Wiener's *Cybernetics*, Shannon's and Weaver's *Mathematical Theory of Communication* and the digital computer era that started during the wartime with the *Electronic Numerical Integrator and Computer* (ENIAC) and the *Electronic Discrete Variable Computer* (EDVAC), both designed by John P. Eckert (1919-1995) and John W. Mauchly (1907-1980), the young emigrant Lotfi A. Zadeh continued his studies in electrical engineering at MIT in Boston after his emigration from Iran into the USA in the 1940s. When he started his doctoral studies at the Columbia University in 1946, he became acquainted with these new milestones in science and technology. Shannon and Wiener delivered lectures in New York about the new theories they had developed during the War. In 1950 Zadeh acted as a moderator at a debate on digital computers at Columbia University, held between Shannon, Edmund Callis Berkeley (1923-1988), the author of the book *Giant Brains or Machines That Think* published in 1949 [9], and Francis Joseph Murray (1911-1996), mathematician and consultant to IBM.
15 years later Zadeh, who was then a professor of electrical engineering at Berkeley, established the new mathematical theory of Fuzzy Sets and Systems [10-12]. Already in 1962 he described the basic necessity of a new scientific tool to handle very large and complex systems in the real world: ''we need a radically different kind of mathematics, the mathematics of fuzzy or cloudy quantities which are not describable in terms of probability distributions. Indeed, the

ISBN: 978-989-95079-6-8

need for such mathematics is becoming increasingly apparent even in the realm of inanimate systems, for in most practical cases the a priori data as well as the criteria by which the performance of a man-made system are judged are far from being precisely specified or having accurately-known probability distributions'' [13].

The potential of the new techniques of the theory of Fuzzy sets and Fuzzy systems urged Ebrahim H. Mamdani, a professor of electrical engineering at Queen Mary College in London, to attempt the implementation of a fuzzy system under laboratory conditions. He expressed the intention to his doctor student Sedrak Assilian, who designed a fuzzy algorithm to control a small steam engine within a few days. The concepts of so-called linguistic variables and Zadeh's max-min composition were suitable to establish fuzzy control rules because *input*, *output* and *state* of the steam engine system range over fuzzy sets. Thus, Assilian and Mamdani designed the first real fuzzy application when they controlled the system by a fuzzy rule base system [14]. In 1974, Assilian completed his Ph. D. thesis on this first fuzzy control system [15].

The steam engine heralded the Fuzzy boom that started in the 1980s in Japan and later pervaded the Western hemisphere. Many fuzzy applications, such as domestic appliances, cameras and other devices appeared in the last two decades of the 20th century. Of greater significance, however, was the development of fuzzy process controllers and fuzzy expert systems that served as trailblazers for scientific and technological advancements of fuzzy sets and systems.[3]

Very little is known about the historical connectivity between the Information theory and the theory of Fuzzy Sets and Systems. In the following paragraph we review some links across the boundaries of these fields.

2.2 *In detail: Signals, Noise and Uncertainty*

Shannon intended to establish a general theory of communication and a basis for such a theory was already given in two former papers of other Bell-engineers, Harry Nyquist and Ralph V. L. Hartley [17, 18].[4] 20 years later, Shannon extended this theoretical basis. He included "new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information." ([1], p. 379) To illustrate this mathematical theory, Shannon had drawn a diagram of a general communication model (Fig. 5):

- An information source produces a series of messages that are to be delivered to the receiver side. Transmission can occur via a telegraph or teletypewriter system, in which case it is a series of letters. Transmission can also occur via a telephone or radio system, in which case it is a function of time $f(t)$, or else it is a function $f(x, y, t)$, such as in a black and white television system, or it consists of complicated functions.
- The transmitter transforms the message in some way so that it can produce signals that it can transmit via the channel. In telegraphy, these are dotdash codes; in telephony, acoustic pressure is converted into an electric current.
- The channel is the medium used for transmission. Channels can be wires, light beams and other options.
- The receiver must perform the opposite operation to that of the transmitter and in this way reconstructs the original message from the transmitted signal.
- The destination is the person or entity that should receive the message.

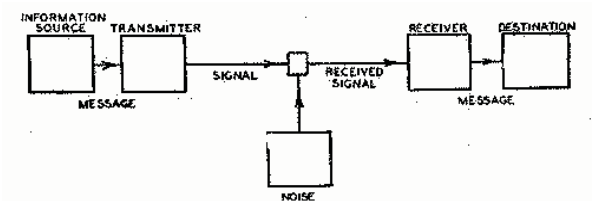


Figure 5: Shannon's diagram ([1], p. 381).

According to this, Shannon conceived of communication purely as the transmission of messages – completely detached from the meaning of the symbols!

In the second paragraph of his paper Shannon states: "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point." ([1], p. 379) Eventually, in chapter 12 that is entitled "Equivocation and Channel Capacity", he applies his theoretical framework to this problem: "If the channel is noisy it is not in general possible to reconstruct the original message or the transmitted signal with *certainty* by any operation on the received signal *E*. There are, however, ways of transmitting the information which are optimal in combating noise." His solution of the problem is as follows: "We consider a communication system and an observer (or auxiliary device) who can see both what is sent and what is recovered (with errors due to noise). This observer notes the errors in the recovered message and transmits data to the receiving point over a "correction channel" to enable the receiver to correct the errors." ([1], p. 408) Shannon also indicated this situation schematically (Fig. 6).

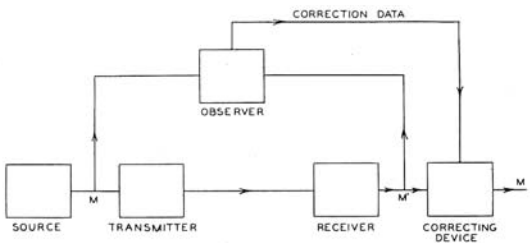


Figure 6: Shannon's diagram of a communication system with correction device ([1], p. 409).

[3] A more detailed presentation of the history of the theory of Fuzzy sets and systems can be found in [16].

[4] A more detailed presentation of the aspect of these papers is given in my paper for the 2009 annual meeting of NAFIPS [19].

ISBN: 978-989-95079-6-8

Shannon's introduction of the observer is of great theoretical interest but – of course – the system has to transmit the data on errors and the correction data via communication channels to the observer. When Warren Weaver explained the mathematics of Communication in the first version of his paper [2], he wrote: "If noise is introduced, then the received message contains certain distortions, certain errors. Certain extraneous material, that would certainly lead to increased uncertainty. […] To get the useful information in the received signal we must subtract the spurious portion. This is accomplished, in the theory, by establishing a quantity known as the "equivocation", meaning the amount of ambiguity introduced by noise." ([2], p. 13) And some paragraphs later he emphasized: "However clever one is with the coding process, it will always be true that after the signal is received there remains some undesirable uncertainty about what the message was". ([2], p. 13)

Let's see how Lotfi Zadeh dealt with this fundamental problem of communication in the 1950s!

In 1949, the year when the book of Shannon and Weaver appeared, Zadeh wrote the Ph. D. thesis on *Frequency Analysis of Variable Networks*, under supervision of Professor John Ralph Ragazzini, and in 1950 he was appointed to an assistant professor. He was enhanced by Wiener's cybernetics and Shannon's information theory. In February 1952, he presented *Some Basic Problems in Communication of Information* at the meeting of the Section of Mathematics and Engineering of the New York Academy of Sciences and in the following, one of these problems will be sketched, that deals with the *recovery process* of transmitted signals. In the proceedings of this meeting, Zadeh wrote: "Let $X=\{x(t)\}$ be a set of signals. An arbitrarily selected member of this set, say $x(t)$, is transmitted through a noisy channel Γ and is received as $y(t)$. As a result of the noise and distortion introduced by Γ, the received signal $y(t)$ is, in general, quite different from $x(t)$. Nevertheless, under certain conditions it is possible to recover $x(t)$ – or rather a time-delayed replica of it – from the received signal $y(t)$." ([20], p. 201)

In this paper, he didn't examine the case where $\{x(t)\}$ is an ensemble; he restricted his view to the problem to recover $x(t)$ from $y(t)$ "irrespective of the statistical character of $\{x(t)\}$"[5, p. 201]. Corresponding to the relation $y = \Gamma x$ between $x(t)$ and $y(t)$ he represented the recovery process of $y(t)$ from $x(t)$ by $x = \Gamma^{-1}y$, where $\Gamma^{-1}y$ is the inverse of Γ, if it exists, over $\{y(t)\}$.

Zadeh represented signals as ordered pairs of points in a signal space Σ, which is imbedded in a function space with a delta-function basis, and to measure the disparity between $x(t)$ and $y(t)$ he attached a distance function $d(x, y)$ with the usual properties of a metric. Then he considered the special case in which it is possible to achieve a perfect recovery of the transmitted signal $x(t)$ from the received signal $y(t)$. He supposed that "$X = \{x(t)\}$ consist of a finite number of discrete signals $x_1(t)$, $x_2(t)$, …, $x_n(t)$, which play the roles of symbols or sequences of symbols. The replicas of all these signals are assumed to be available at the receiving end of the system. Suppose that a transmitted signal x_k is received as y. To recover the transmitted signal from y, the receiver evaluates the 'distance' between y and all possible transmitted signals x_1, x_2, …, x_n, by the use of a suitable distance function $d(x, y)$, and then selects that signal which is 'nearest' to y in terms of this distance function (Fig. 7). In other words, the transmitted signal is taken to be the one that results in the smallest value of $d(x, y)$. This in brief, is the basis of the reception process." ([20], p. 201) By this process the received signal x_k is always 'nearer' – in terms of the distance functional $d(x, y)$ – to the transmitted signal $y(t)$ than to any other possible signal x_i, i.e. $d(x_k, y) < d(x_i, y)$, $i \neq k$,*for all k and i.*

But at the end of his reflection of this problem Zadeh conceded that "in many practical situations it is inconvenient, or even impossible, to define a quantitative measure, such as a distance function, of the disparity between two signals. In such cases we may use instead the concept of neighborhood, which is basic to the theory of topological spaces" ([20], p. 202). Spaces such as these, Zadeh surmised, could be very interesting with respect to applications in communication engineering. Therefore, this problem of the recovery process of transmitted signals which is a special case of Shannon's *fundamental problem of communication*, the problem "of reproducing at one point either exactly or approximately a message selected at another point", was one of the problems that initiated Zadeh's thoughts about not precisely specified quantitative measures, i. e. or cloudy or fuzzy quantities. – About 15 years later he proposed his new 'concept of neighborhood' which is now basic to the theory of fuzzy systems!

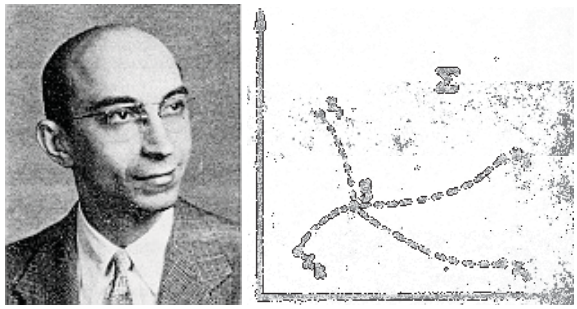


Figure 7: Lotfi A. Zadeh (born 1921) and his illustration: "Recovery of the input signal by means of the comparison of the distances between the received signal y and all possible transmitted signals" [20].

3 Fuzziness and Information Theory

In this section we concentrate our attention from the problem of communication to some philosophical considerations in the first half of the 20th century by Charles William Morris (1901-1976) and Warren Weaver. However, Morris was an engineer by training and he did his Ph. D. in philosophy under George Herbert Mead (1863-1931), the founder of social psychology, and Weaver was a mathematician and physicist. Nevertheless, both scientists wrote philosophical

ISBN: 978-989-95079-6-8

works on the subjects of signs and their transmission. Moreover, there is a distinct similarity between the two theoretical works that has to be underlined.

3.1 Theory of Signs, semiotics, and Information theory

The study of sign processes, or signification and communication is called "semiotics" since the 1930s. Great work to formalize this field was done by members of the Vienna Circle but also by the philosophers and linguists Charles Sanders Peirce (1839–1914), Ferdinand de Saussure (1857–1913) and Louis Hjelmslev (1899–1965), but here we will limit our considerations to the fundamental work of semiotics by Morris.

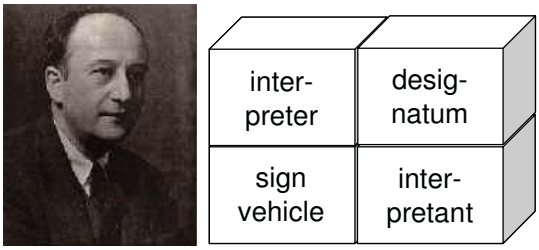


Figure 8: Ch. W. Morris; an illustration of the four components of the semiosis process.

Morris was in touch with some members of the Vienna Circle and he was a member of the Unity of Science movement. When he intended in his 1938 published *Foundations of the Theory of Signs* [21] a science of signs "on a biological basis and specifically with the framework of the science of behavior", he defined *semiotics* as a *universal* theory of signs and an *interdisciplinary* undertaking. In his view, the mission of semiotics as a science of signs is analyzing language as a social system of signs. Language is a system of signs which produces dispositions to social behavior, and in order to understand the uses and effects of signs we have to understand that and how signs influence social behavior. The process by which a sign-vehicle may function as a sign is called *semiosis*.[5] In Morris' foundation there are four components of the semiosis (Fig. 8):

1) the *sign vehicle* − this is the object or event which functions as a sign,
2) the *designatum* − this is the kind of object or class of objects which the sign designates,
3) the *interpretant* − this is the disposition of an interpreter to initiate a response-sequence as a result of perceiving the sign, and
4) the *interpreter* − this is the person for whom the sign-vehicle functions as a sign.

He also divided *semiotics* into three interrelated disciplines:

1) syntactics − the study of the methods by which signs may be combined to form compound signs,
2) semantics − the study of the signification of signs,
3) pragmatics − the study of the origins, uses, and effects of signs.

3.2 Weaver's three levels in Information theory

It seems that Warren Weaver was familiar with Morris' 10 years old classification of the semiosis when he wrote his paper on Shannon's Mathematical theory of communication [2]. Already in the third paragraph of his paper he wrote "In communication there seem to be problems at three levels: 1) technical, 2) semantic, and 3) influential. The technical problems are concerned with the accuracy of transference of information from sender to receiver. They are inherent in all forms of communication, whether by sets of discrete symbols (written speech), or by a varying two-dimensional pattern (television). The semantic problems are concerned with the interpretation of meaning by the receiver, as compared with the intended meaning of the sender. This is a very deep and involved situation, even when one deals only with the relatively simple problems of communicating through speech. […] The problems of influence or effectiveness are concerned with the success with which the meaning conveyed to the receiver leads to the desired conduct on his part. It may seem at the first glance undesirable narrow to imply that the purpose of all communication is to influence the conduct of the receiver. But with any reasonably broad definition of conduct, it is clear that communication either affects conduct or is without any discernible and provable effect at all." ([2], p. 11)

In the revised version of the paper that was published in [8], Weaver explained the trichotomy of the communication problem in extenso and he divided it into three levels:

- **Level *A*** contains the purely technical problem involving the exactness with which the symbols can be transmitted,
- **Level *B*** contains the semantic problem that inquires as to the precision with which the transmitted signal transports the desired meaning,
- **Level *C*** contains the pragmatic problem pertaining to the effect of the symbol on the destination side: What influence does it exert?

He underscored very clearly the fact that Shannon's theory did not even touch upon any of the problems contained in levels *B* and *C*, that the concept of information therefore must not be identified with the "meaning" of the symbols: "In fact, two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent, from the present viewpoint, as regards information." [8] However, there is plenty of room for fuzziness in the levels *B* and *C*. The interpretation of meaning of signs, e. g. linguistic signs, names, words, is obviously a fuzzy process, and influence or effectiveness the exerted to the receiver's side is a fuzzy process, too. We will have this fuzziness at the back of our mind following Weaver's continuing considerations.

[5] The term *semiosis* was introduced by Ch. S. Peirce to describe "a process that interprets signs as referring to their objects" [22].

ISBN: 978-989-95079-6-8

3.3 Fuzziness in the diagram of a communication system

It is quite plain: Weaver went over and above Shannon´s theory: "The theory goes further. Though ostensibly applicable only to problems at the technical level, it is helpful and suggestive at the levels of semantics and effectiveness as well." Weaver stated, that Shannon's formal diagram of a communication system (Fig. 5) "can, in all likelihood, be extended to include the central issues of meaning and effectiveness. […] One can imagine, as an addition to the diagram, another box labeled "Semantic Receiver" interposed between the engineering receiver (which changes signals to messages) and the destination. This semantic receiver subjects the message to a second decoding the demand on this one being that it must match the statistical semantic characteristics of the message to the statistical semantic capacities of the totality of receivers, or of that subset of receivers which constitutes the audience one wishes to affect.

Similarly one can imagine another box in the diagram which inserted between the information source and the transmitter, would be labeled "Semantic Noise" (not to be confused with "engineering noise". This would represent distortions of meaning introduced by the information source, such as a speaker, which are not intentional but nevertheless affect the destination, or listener. And the problem of semantic decoding must take this semantic noise into account. It is also possible to think of a treatment or adjustment of the original message that would make the sum of message meaning plus semantic noise equal to the desired total message meaning at the destination." ([2], p. 13)

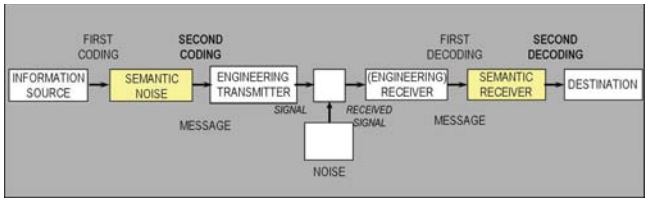


Figure 9: Shannon's diagram with Weaver's addition of the two boxes "Semantic Receiver" and "Semantic Noise".

Fig. 9 shows Weaver's two additional "fuzzy" boxes in Shannon's diagram of a communication system. We will interpret the "first coding" between the information source and the "Semantic Noise" as a *fuzzification* and the "second decoding" between the "Semantic Receiver" and the destination as a *defuzzification*.

4 Outlook

In future work we will proceed with this generalization of Weaver's philosophical considerations on Shannon's Mathematical theory of communication to a "Fuzzy information theory".

References

[1] C. E. Shannon, A Mathematical Theory of Communication, *The Bell System Technical Journal*, 27, pp. 379-423 & 623-656, 1948.

[2] W. Weaver, The mathematics of communication, *Scientific American*, 181: 11-15, 1949.

[3] N. Wiener, *Cybernetics or Control and Communications in the Animal and the Machine*. Cambridge, Massachusetts: MIT Press 1948.

[4] A. N. Kolmogorov, Interpolation and Extrapolation. In: *Bulletin de l'Académie des Sciences de l'URSS*, Serie Mathématique, Vol. 5, pp. 3-14, 1941.

[5] R. A. Fisher, *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd, 1925.

[6] L. E. Kay, *Who Wrote the Book of Life?: A History of the Genetic Code*. Stanford, Calif.: Stanford University Press 2000.

[7] N. Wiener, *Invention. The Care and Feeding of Ideas*, London: The MIT Press, 1993.

[8] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, Illinois: University of Illinois Press, 1949.

[9] E. C. Berkeley, *Giant Brains or Machines that Think*. New York, London: John Wiley & Sons, Chapman & Hall, 1949.

[10] L. A. Zadeh, Fuzzy Sets, *Information and Control*, 8, 1965, pp. 338-353.

[11] L. A. Zadeh, Fuzzy Sets and Systems. In: Fox, J. (ed.): *System Theory*. Microwave Research Institute Symposia Series XV, Brooklyn, New York: Polytechnic Press 1965, p. 29-37.

[12] R. E. Bellman, R. Kalaba, L. A. Zadeh, Abstraction and Pattern Classification. *Journal of Mathematical Analysis and Applications* 13 1966, pp. 1-7.

[13] L. A. Zadeh, From Circuit Theory to System Theory. *Proceedings of the IRE*, Vol. 50, 1962, pp. 856-865.

[14] E. H. Mamdani, S. Assilian: An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. In: *International Journal of Man-Machine Studies*, Vol. 7, No. 1, pp. 1-13, 1975.

[15] S. Assilian, *Artificial Intelligence in the Control of Real Dynamic Systems*. Ph.D. Thesis, No. DX193553, University London, August 1974.

[16] R. Seising, *The Fuzzification of Systems. The Genesis of Fuzzy Set Theory and Its Initial Applications – Developments up to the 1970s* (Studies in Fuzziness and Soft Computing Vol. 216) Berlin, New York, [et al.]: Springer 2007.

[17] H. Nyquist, Certain Factors Affecting Telegaph Speed, *The Bell System Technical Journal* 3, 1924, pp. 324-346.

[18] R. V. L. Hartley, Transmission of information, *The Bell System Technical Journal* 7, 1928, pp. 535-563.

[19] R. Seising, *On two 60 years old theories and the Theory of Fuzzy Sets and Systems: Cybernetics and Information Theory*. Proceedings of the 2009 Annual Meeting of the North American Fuzzy Information Society – NAFIPS-09 Annual Conference, University of Cincinnati, Ohio, USA, June 14-17, 2009 (to appear)

[20] L A. Zadeh, Some Basic Problems in Communication of Information *The New York Academy of Sciences,* Series II, Vol. 14, No. 5, 1952, pp. 201-204

[21] Ch. W. Morris, *Foundations of the Theory of Signs,* (International Encyclopedia of Unified Science, ed. Otto Neurath, vol. 1 no. 2.) Chicago: University of Chicago Press, 1938. Rpt, Chicago: University of Chicago Press, 1970-71. Reprinted in Charles Morris, *Writings on the General Theory of Signs* (The Hague: Mouton, 1971), pp. 13-71.

[22] N. Houser and Chr. Kloesel (eds.), *The Essential Peirce. Selected Philosophical Writings*, Vol. 2 (1893-1913), Peirce Edition Project, Bloomington/Indianapolis 1998 pp. 411.

ISBN: 978-989-95079-6-8

On classical, fuzzy classical, quantum, and fuzzy quantum systems

Rudolf Seising

European Centre for Soft Computing
Mieres, Asturias, Spain
Email: rudolf.seising@softcomputing.es

***Abstract*—*In this paper we consider physical systems and the concept of their states in the context of the theory of fuzzy sets and systems. In section 1 we give a brief sketch on the fundamental difference between the theories of classical physics and quantum mechanics. In section 2 and 3 we introduce very shortly systems and their states in classical and quantum mechanics, respectively. Section 4 presents the concept of fuzzy systems. We propose to fuzzify the classical systems in section 5 and quantum systems in section 6. In section 7 we start to consider a fuzzy interpretation of the uncertainty principle.*

Keywords— Fuzzy sets and systems, quantum mechanics, system theory, philosophy.

1 Introduction

Due to the scientific revolution brought about by the discovery of quantum mechanics in the first third of the 20th century, a basic change took place in the relationship between the exact scientific theory of physics and the phenomena observed in basic experiments. Systems of quantum mechanics – "quantum systems" – do not behave like systems of classical theories in physics – their elements are not particles and they are not waves, they are different. This change led to a new mathematical conceptual fundament in physics.

Niels Bohr, Max Born, Louis de Broglie, Paul A. M. Dirac, Werner Heisenberg, Pacual Jordan, Wolfgang Pauli, Erwin Schrödinger, John von Neumann and others introduced new objects and theoretical terms to the new mathematical theory of atomic physics and rather quantum physics, then so-called "quantum mechanics" that differ significantly from those in classical physics. Their properties are completely new and not comparable with those of observables in classical theories such as Newton's mechanics or Maxwell's electrodynamics.

The new theoretical term is the quantum mechanical state function ψ that is an element of the abstract Hilbert space H. The theory of quantum mechanics is completely abstract: it is a theory of mathematical state functions that have no exact counterpart in reality. This means that per se ψ is not observable but, nonetheless, we can experiment with a quantum mechanical object having a state function in order to measure its position value, and we can also experiment with this object in order to measure its momentum value. However, we cannot conduct both experiments simultaneously and thus are not able to get both values for the same point in time. But we can predict these values as outcomes of experiments at this point in time. Since predictions are targeted on future events, we cannot valuate them with the logical values "true" or "false," but must use probabilities. The probability distribution to measure a certain position, e.g. value $\boldsymbol{r}$ (t) = (x (t), y (t), z (t)) at point t in time is given by $|\psi(\boldsymbol{r}, t)|^2$ and the probability distribution to measure a certain momentum value $\boldsymbol{p}$ at time point t is given by $|\psi(\boldsymbol{p}, t)|^2$, where $\psi(\boldsymbol{x}, t)$ or $\psi(\boldsymbol{p}, t)$, are representtations of the abstract Hilbert space element ψ in the position or momentum representation respectively.

2 Classical Systems

In classical physics, the state of a "system" or an "object" is represented by a set of observables. For example, in Newtonian mechanics the state of an object (a particle with mass m) is given by the pair of 3-component vector values of the object's position vector $\boldsymbol{r}$ and its momentum vector $\boldsymbol{p}$. These two vectors imply all other properties of the object that are relevant in the Newtonian theory of mechanics. We can formulate that the state of a physical object is the collection of all the object's properties P_i. In order to represent these properties P_i in terms of the physical theory, we must determine the formally possible functions F_i in this mechanical theory, and in order to know the object's properties at a given point in time t, we must measure the values of these functions F_i. Thus, the representation of the "state of a classical object" is related to the measurement process or the perception process of the observer.

Due to the possible errors of measurement and the systematic errors occurring in every experiment, we can attribute their probability of this being the real value to all measured values of observables. Thus, the state of an object in Newtonian mechanics is given by the pair of the probability distributions of position $\boldsymbol{r}$ and momentum $\boldsymbol{p}$.

3 Quantum Systems

3.1 The concept of state in quantum mechanics

The state of a quantum mechanical system is much more difficult to determine than that of classical systems as we cannot measure sharp values for both variables simultaneously. This is the meaning of Heisenberg's uncertainty principle.

However, we can experiment with quantum mechanical objects in order to measure a position value, and we can also experiment with these systems in order to measure their momentum value. But: we cannot conduct both experiments

ISBN: 978-989-95079-6-8

simultaneously and thus are not able to get both values for the same point in time respectively. We can predict these values as outcomes of experiments at this point in time. Since predictions are targeted on future events, we cannot valuate them with logical values "true" or "false", but with probabilities.
Accordingly, in quantum mechanics, we have to use a modified concept of the state: the state of a quantum mechanical system consists of the probability distributions of all the object's properties that are formally possible in this physical theory.

- Max Born ([1, 2]) proposed an interpretation of this non-classical peculiarity of quantum mechanics – the quantum mechanical wave function is a "probability-amplitude": The absolute square of its value equals the probability of it having a certain position or a certain momentum if we measure the position or momentum respectively. The higher the probability of the position value, the lesser that of the momentum value and vice versa.
- In 1932, John von Neumann published the *Mathematical Foundations of Quantum Mechanics* [3], in which he defined the quantum mechanical wave function as a one-dimensional subspace of an abstract Hilbert space, which is defined as the state function of a quantum mechanical system (or object). Its absolute square equals the probability density function of it having a certain position or a certain momentum in the position or momentum representation of the wave function respectively.

Unfortunately there is no joint probability distribution for events in which both variables have a certain value simultaneously, as there is no classical probability space that comprises these events. Such pairs would describe classical states. Thus, the quantum mechanical system's state function embodies the probabilities of all properties of the object, but it delivers no joint probability distribution for all these properties. Therefore we claim here: "We need a radically different kind of mathematics, the mathematics of […] quantities which are not describable in terms of [classical] probability distributions.[1]

3.2 Quantum logic and Quantum probability theory

After the establishment of quantum mechanics there appeared some approaches to achieve a new logic and later also a new probability theory to handle quantum mechanical propositions and quantum mechanical events.

- 1936, Garett Birkhoff and John von Neumann proposed the introduction of a "quantum logic", as the lattice of quantum mechanical propositions is not distributive, and therefore not Boolean [5].
- In 1963, George Whitelaw Mackey attempted to provide a set of axioms for the propositional system of predictions of experiment's outcomes. He was able to show that this system is an orthocomplemented partially ordered set. [6]

In these logico-algebraic approaches, the "probabilities" of evaluating the predictions of the properties of a quantum mechanical system do not satisfy Kolmogorov's well-known axioms. The double-slit experiment shows that they are not additive and together with their non-distributivity it is indicated that the probabilistic structure of quantum mechanics is more complicated than that of the classical probability space as it was defined by Kolmogorov.

- Already in the 1960s, the philosopher and statistician Patrick Suppes discussed the "probabilistic argument for a non-classical logic of quantum mechanics" [7, 8]. He introduced the concept of a "quantum mechanical σ-field" as an "orthomodular partial ordered set" covering the classical σ-fields as substructures.
- In the 1980s, a "quantum probability theory" was proposed and developed by Stanley Gudder and Imre Pitowski [9, 10].[2]

The quantum mechanical lattice of predictions is Suppes' "quantum mechanical σ-field", which can be restricted to a Boolean lattice corresponding to a given observable. The quantum probabilities became classical probabilities again, only applying to predictions of compatible observables.
The theory of Fuzzy sets and systems pertains to "quantities which are not describable in terms of probability distributions" and theory of Quantum mechanics pertains to quantities which are not describable in terms of classical probability distributions.
Quantum logic and Quantum probability theory represent important approaches to manage quantum mechanical uncertainties within limits of usual mathematics and the developments in the last decades show numerous and also very difficult results. Thus, quantum logic and quantum probability theory are new theories in classical mathematics that became more and more complex. On the other hand there was the new theory of fuzzy sets and systems available at the same time and the question arose in the 1980s and 1990s of whether fuzzy sets could be useful in the interpretation of quantum mechanics. However, at that time this approach was not successful. The disappointing results may have stemmed from the fact that fuzzy set theory was not as well accepted as a mathematical tool the as it is today and from a lack of interest in using the new theory on the part of theoretical physicists. Moreover, until recently there was also no interest in fuzzy set theory in the philosophy of science. But now the theory of fuzzy sets is broadly accepted, particularly in applied sciences and technology and in the history of science the theory of fuzzy sets attracts

[1] This is analogue to Zadeh's requirement: ". . . we need a radically different kind of mathematics, the mathematics of fuzzy or cloudy quantities which are not describable in terms of probability distributions" in his 1962 article [4], p. 857.

[2] These developments regarding a theory of probabilistic structures of quantum mechanics became very complex, as the reader can see in the authors Ph. D thesis [11].

ISBN: 978-989-95079-6-8

attention [12, 13, 14]. In physics, the results of new experiments (Alain Aspect's test of Bell's inequality in 1982 [15, 16] and Anton Zeilinger's experiments on quantum teleportation since 1997 [17] have sparked a new debate on the interpretation of quantum mechanics and there is growing interest in the theory of fuzzy sets in the field of scientific history [29]. In the next sections we will evolve some ideas towards a "fuzzy view" on quantum systems.

4 Fuzzy Systems

In his talk on "A New View on System Theory," for the *Symposium on System Theory* that took place in Brooklyn in April 1965, Lotfi Zadeh defined fuzzy systems as follows:

Definition: A system S is a fuzzy system if input $u(t)$, output $y(t)$, or state $x(t)$ of S or any combination of them ranges over fuzzy sets. ([18], p. 33)[3]

He explained that "these concepts relate to situations in which the source of imprecision is not a random variable or a stochastic process but rather a class or classes which do not possess sharply defined boundaries." ([18], p. 29

Eight years later, in "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," he introduced "linguistic variables" that are variables whose values may be sentences in a specific natural or artificial language. [19] To illustrate, the values of the linguistic variable "age" might be expressible as *young*, *very young*, *not very young*, *somewhat old*, *more or less young*.

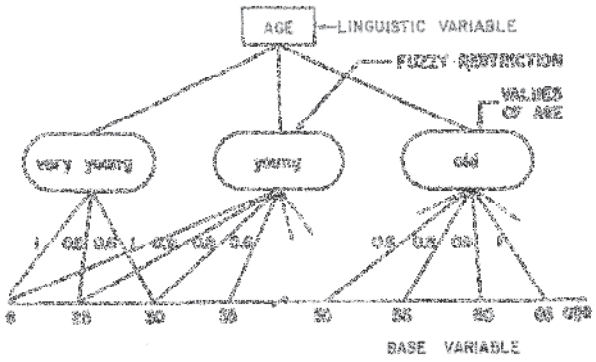


Figure 1: Example of the linguistic variable "Age" [20].

These values are formed with the label *old*, the negation *not*, and the hedges *very*, *somewhat*, and *more or less*. In this sense the variable "age" is a linguistic variable (see Fig. 1).

Linguistic variables became a proper tool for reasoning without exact values. Since in many cases, it is either impossible or too time-consuming (and therefore too expensive) to measure or compute exact values, the concept of linguistic variables has been successful in many fuzzy application systems, e.g., in control and (medical) decision making. In the next section we seek to apply the concept of linguistic variables in quantum mechanics where exact values of observables do not exist. The situation is that outcomes of a physicist's experiments have to be values of observables, i.e., an observing physicist assigns sharp values (and their classical probability distributions of the classical variables (e.g. its position $\boldsymbol{r}$ or its momentum $\boldsymbol{p}$) due to the possible errors of measurement and the systematic errors occurring in every experiment)[4] to a quantum theoretical object. This value is not sufficient to determine the quantum object's state. This is only one representation – among many others – and none of these representations of the state of the quantum system is complete!

5 Fuzzy Classical Systems

In this section a new approach to the interpretation of uncertainty in quantum mechanics using the methodology of fuzzy sets and systems is outlined.

We will interpret the "fuzzy state variable" of a physical system as a vector of linguistic variables instead of numerical variables. This interpretation yields to the interpretation of the "fuzzy state" of a classical physical object as a pair of the two linguistic variables – position and momentum

In a manner similar to Zadeh's extension of systems to fuzzy systems, the definition of a linguistic variable operating on a fuzzy set, and assignment of membership degrees and elements of the term set of the linguistic variable, the "fuzzy state" of a physical system is interpreted as a vector of linguistic variables instead of numerical variables.

A concrete system a has a certain number of properties P^i, $i \in \{1, \ldots, n\}$. In classical physics we attach classical numerical variables (observables) $V_P{}^i$ to these properties that can be measured. To use the methods of fuzzy set theory, now we attach can find also a linguistic variable $LV_P{}^i$ to representing the property P^i. These linguistic variables operate on fuzzy sets and assign membership degrees and elements of a term set, for example:

$$T(LV_P{}^i) = \{\text{very small, small, big, very big, … etc .}\}$$

We can imagine the n-tuple $\boldsymbol{LVP_n} = \langle LV_P{}^1, LV_P{}^2, \ldots, LV_P{}^n \rangle$ to be a "linguistic vector" in an n-dimensional Cartesian space. The value $\boldsymbol{LVP_n}$ (a, t) for a system a at time point t is called the "linguistic state" or "fuzzy state" of this system at this time. During this time, the linguistic state of a system moves in the "linguistic state space" or "fuzzy state space" $\Sigma_{\boldsymbol{n}}{}^L(\mathrm{a}) = \{\boldsymbol{LVP_n}(\mathrm{a}, t) \mid t \in T\}$ of the system.

In the case of a classical particle in Newtonian mechanics, the "fuzzy state" is the pair (2-tupel) $(LV_{\boldsymbol{r}}, LV_{\boldsymbol{p}})$ of the two linguistic variables position $LV_{\boldsymbol{r}}$ and momentum $LV_{\boldsymbol{p}}$ that operate on fuzzy sets and assign membership degrees and elements of a term set, for example:

$$T(LV_{\boldsymbol{r}}) = \{\text{very small, small, null, big, very big, … etc.}\}$$

and

$$T(LV_{\boldsymbol{p}}) = \{\text{very small, small, null, big, very big, … etc.}\}$$

Usually the shape of the fuzzy set's membership functions is subjectively chosen or dependant on the problem. In a very special case, the membership function may have the shape of the classical Gaussian probability distribution and thus the

[3] We assume that the reader is familiar with the basic principles of the theory of fuzzy sets.

[4] In the next sections we sometimes will omit this completion about the classical errors of measurement and the systematic errors.

ISBN: 978-989-95079-6-8

fuzzy state variable yields the probabilities of measuring the observables position ***r*** and momentum ***p*** due to the calculation of errors. However, in general, membership functions of fuzzy sets do not represent probability distributions of measurement errors or randomness, but more general uncertainties that are deeply rooted in the absence of the theoretical concept's strict boundaries.

We know already that classical concepts such as position and momentum, having strict boundaries in Newtonian mechanics, do not have such boundaries in the theory of quantum mechanics. Therefore this pair of concepts does not match the quantum mechanical state variable – to represent the quantum mechanical "state", position and momentum are in use with some uncertainty. This concept of uncertainty in quantum mechanics is often represented by classical probability, but, in the strict sense, the quantum mechanical uncertainty is different from the concept of classical probability.

6 Fuzzy Quantum Systems

Let's try to extend our approach of "fuzzy states" of systems to quantum systems to include the assumption that the classical theoretical concepts are not the right concepts, but that we have no better concepts to interpret the outcomes of classical experiments. Thus, we can use fuzzy sets and linguistic variables to convert classical observables to systems of quantum mechanics.

In the case of a quantum system, the "fuzzy quantum state" is an infinite dimensional vector $\boldsymbol{LVP_\infty}\ (a,t)$ of linguistic variables in the abstract Hilbert space, with an infinite tuple $\boldsymbol{LVP_\infty} = \langle LV_P^1, LV_P^2, \ldots, LV_P^n, \ldots \rangle$ of linguistic variables LV_i, but not all linguistic variables LV_i and LV_j are compatible, i.e. they are in an uncertainty relation with each other, e.g., $LV_1 = LV_{\boldsymbol{r}}$ (position observable) and $LV_j = LV_{\boldsymbol{p}}$ (momentum observable).

In general, at one point in time we can measure one of these LV_i and this measurement reduces the membership function to a numerical value of this observable. This is an effect that is known in usual quantum mechanics as the "collapse" of the quantum mechanical state function.

Let's use the fuzzy state concept for quantum systems! After the measurement of an observable, say V_P^i, there is no "collapse" of the fuzzy quantum state. We still have it's complete representation as the infinite tuple of linguistic variables $\boldsymbol{LVP_\infty} = \langle LV_P^1, LV_P^2, \ldots, LV_P^n, \ldots \rangle$. However, after the measurement of observable V_P^i, this component of the fuzzy quantum state is not any longer a linguistic but a numeric term: the measurement, just a number (associated with a unit).

7 Fuzzy Uncertainty Principle

First we consider the classical uncertainty principle that was found by Heisenberg in 1925. This uncertainty principle states that the values of certain pairs of conjugate variables[5] (e.g. position ***r*** and momentum ***p***) cannot both be known with arbitrary precision. That is, the more precisely one variable is known, the less precisely the other is known. We already noticed that this is not the uncertainty of the measurement of particular observables of a system.

Heisenberg's uncertainty principle was very often formulated in usual mathematics (classical probability theory and statistics) as follows: every quantum state has the property that the *root-mean-square deviation* of the position ***r*** from its mean (the standard deviation of the ***r***-distribution) times the root-mean-square deviation of the momentum ***p*** from its mean (the standard deviation of ***p***) can never be smaller than a small fixed fraction of Planck's constant:

$$\Delta r \cdot \Delta p \geq \tfrac{\hbar}{2}, \tag{1}$$

where $\Delta r = \sqrt{\left\langle (r - \langle r \rangle)^2 \right\rangle}$ and $\Delta p = \sqrt{\left\langle (p - \langle p \rangle)^2 \right\rangle}$.

Any measurement of the position with accuracy $\Delta\boldsymbol{r}$ collapses the quantum state making the standard deviation of the momentum $\Delta\boldsymbol{p}$ larger than $\frac{\hbar}{2}$.

This view is using classical probability theory what is adequate for each one of the two variables ***r*** and ***p*** but what is not appropriate to the combination of the two "pictures". There is no joint probability distribution for the two variables ***r*** and ***p*** and therefore the product of their *root-mean-square deviations* has no meaning in classical probability theory. That is, what is meant by the logico-algebraic result that there is no Boolean lattice of quantum mechanics.

Let's consider the uncertainty principle in our approach of fuzzy quantum states! In our view position and momentum, ***r*** and ***p***, are associated with linguistic variables, $LV_{\boldsymbol{r}}$ and $LV_{\boldsymbol{p}}$. Hence we can consider $\Delta r = \sqrt{\left\langle (r - \langle r \rangle)^2 \right\rangle}$ and $\Delta p = \sqrt{\left\langle (p - \langle p \rangle)^2 \right\rangle}$ as being linguistic variables as well.

Because Heisenberg's uncertainty relation (1) is a proposition of the possible values of $\Delta\boldsymbol{r}\cdot\Delta\boldsymbol{p}$, we can also consider this product of the two linguistic variables $\Delta\boldsymbol{r}$ and $\Delta\boldsymbol{p}$ as being a linguistic variable ***S*** ("action"[6]).

Heisenberg's inequality claims that $\boldsymbol{S} = \Delta\boldsymbol{r}\cdot\Delta\boldsymbol{p}$ can never be smaller than $\frac{\hbar}{2}$, i.e. the minimal value of ***S*** equals $\frac{\hbar}{2}$, but this result was never derived by Heisenberg himself. Moreover, in his famous 1930 lecture in Chicago he refined his principle to the following inequality [21]:

$$\Delta x \Delta p \gtrsim h. \tag{2}$$

The formulation (1) of the uncertainty relations was proved by E. H. Kennard in 1927 but in this prove it was postulated that $\Delta\boldsymbol{r}$ and $\Delta\boldsymbol{p}$ are the standard deviations of position and momentum, therefore this formulation uses classical probability theory [21]

What does this mean? – The value of the linguistic variable ***S*** at a given point of time is the composition of the values of the linguistic variables $\Delta\boldsymbol{r}$ and $\Delta\boldsymbol{p}$ but to get both of the two

[5] In physics, conjugate variables are pair of variables mathematically defined in such a way that they become Fourier transform duals of one-another.

[6] "Action" is the name of this particular quantity (product of position and momentum or of energy and time) in physics.

ISBN: 978-989-95079-6-8

values is not possible in the theory of quantum mechanics. (Remark: To compute the value of $\Delta \boldsymbol{r}$ or $\Delta \boldsymbol{p}$ we have to measure the value of $\boldsymbol{r}$ or $\boldsymbol{p}$ respectively.)
Thus, it is only possible to know one of the linguistic values of $\Delta \boldsymbol{r}$ or $\Delta \boldsymbol{p}$ and therefore it is not possible to compute the value of the linguistic variable $\boldsymbol{S} = \Delta \boldsymbol{r} \cdot \Delta \boldsymbol{p}$.

8 Conclusions

During the last decade the literature on combining the theories of quantum mechanics and fuzzy sets was growing, e.g. [23- 29].[7] Of course there might be many more papers on the meeting of the two theories but unfortunately – as one of the reviewers to this paper, "it seems to be sometimes the application of ideas from quantum theory to fuzzy sets than the other way around". This paper was written to plea for applications of ideas from Fuzzy set theory to quantum mechanics. Of course, we gave only some preliminary ideas to establish a fuzzy approach to physics and particularly to quantum mechanics.
We presented a rough idea of what the methodologies of fuzzy sets and systems, along with linguistic variables, could contribute to the project to represent phenomena in physics that can not be represented in classical mechanics. Quantum mechanics is a very successful theory to represent these phenomena but there are difficulties with the using of the concept of probability. It seems that the concepts of fuzzy sets and linguistic variables can be more appropriate for the representation of quantum phenomena in the framework of quantum mechanics than the classical concept of probability. Our definition of the fuzzy state of a quantum mechanical system avoids the difficulties that arise in classical probability theory in attempts to define this state. On the other hand, with this fuzzy approach new difficulties arise, e.g. with the experimental side of physics and also with its interpretative side. Of course, scientists should take such problems seriously. At the other hand, these are the usual and well-known problems of all fuzzy-approaches in the field of science and technology of the last 40 year. From the point of view of a historian and philosopher of science this is also a big challenge!

References

[1] M. Born, Zur Quantenmechanik der Stoßvorgänge, *Zeitschrift für Physik*, 37: 863-867, 1926

[2] M. Born, Das Adiabatenprinzip in der Quantenmechanik, *Zeitschrift für Physik*, 40: 167-191, 1926.

[3] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton Univ. Press, Princeton, New Jersey, 1955.

[4] L. A. Zadeh: From Circuit Theory to System Theory, *Proceedings of the IRE*, 50, 5: 856-865, 1962.

[5] G. Birkhoff, J. von Neumann, The Logic of Quantum Mechanics, *Annals of Mathematics*, Series 2, 37: 823 ff, 1936.

[6] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics*, New York: W. A. Benjamin, 1963.

[7] Thank you very much to the anonymous referee who mentioned these papers in his/her review to this paper.

[7] P. Suppes, Probability Concepts in Quantum Mechanics, *Philosophy of Science*, 28: 378-389, 1961.

[8] P. Suppes, The Probabilistic Argument for a Non-Classical Logic of Quantum Mechanics, *Philosophy of Science*, 33, 14-21, 1966.

[9] S. P. Gudder, *Quantum Probability*, Academic Press: San Diego, 1988.

[10] I. Pitowski, *Quantum Probability, Quantum Logic*, Lecture Notes in Physics, 321, Berlin: 1989.

[11] A. Aspect, P. Grangier, and G. Roger, Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities, *Physical Review Letters*, 49, 2: 91-94, 1982.

[12] R. Seising, From Vagueness in Medical Thought to the Foundations of Fuzzy Reasoning in Medical Diagnosis. *Artificial Intelligence in Medicine* 38: 237-256, 20'06.

[13] R. Seising, *The Fuzzification of Systems. The Genesis of Fuzzy Set Theory and Its Initial Applications – Its Development to the 1970s*. Berlin [et al.]: Springer (Studies in Fuzziness and Soft Computing 216) 2007.

[14] R. Seising, On the Absence of Strict Boundaries – Vagueness, Haziness, and Fuzziness in Philosophy, Medicine, and Science, *Applied Soft Computing* (Special Issue: Forging the Frontiers) 8, 3: 1232-1242, 2008.

[15] A. Aspect, Dalibard, G. Roger, Experimental test of Bell's inequalities using time-varying analyzers, *Physical Review Letters,* 49: 1804-1807, 1982.

[16] Dik Bouwmeester, Jan-Wei Pan, Klaus Mattle, Manfred Eibl, Harald Weinfurter, Anton Zeilinger, Experimental Quantum Teleportation, *Nature,* 390: 575 ff, 1997.

[17] R. Seising, *Probabilistische Strukturen der Quantenmechanik*, Frankfurt: Lang 1996. (in German)

[18] L. A. Zadeh, Fuzzy Sets and Systems. In: J. Fox (ed.): *System Theory*, Microwave Res. Inst. Symp. Series XV, Brooklyn, New York: Polytechnic Press, 1965, 29-37.

[19] L. A. Zadeh, Outline of a New Approach to the Analysis of Complex Systems and Decision Processes, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3, 1: 28-44, 1973.

[20] L. A. Zadeh, On the Analysis of Large-Scale Systems. Memorandum No. ERL-M418, Electronic Research Laboratory, College of Engineering, University of California, Berkeley, January 8, 1974.

[21] W. Heisenberg, *Physikalische Prinzipien der Quantentheorie* Leipzig: Hirzel, 1930. English translation *The Physical Principles of Quantum Theory*, Chicago: University of Chicago Press, 1930.

[22] E. H. Kennard, Zur Quantenmechanik einfacher Bewegungstypen *Zeitschrift für Physik*, 44: 326, 1927.

[23] S. Ishikawa, A quantum mechanical approach to a fuzzy theory, *Fuzzy Sets and Systems*, 90, 3: 277-306, 1997.

[24] Mirko Navara, Boolean representations of fuzzy quantum spaces, Fuzzy Sets and Systems, 87, 2: 201-207, 1997.

[25] Anatolij Dvurecenskij, Fuzzy set representations of some quantum structures, Fuzzy Sets and Systems, 101, 1: 67-78, 1999.

[26] Stan Gudder, Non-disturbance for fuzzy quantum measurements, Fuzzy Sets and Systems, 155, 1: 18-25, 2005.

[27] G.G. Rigatos, S.G. Tzafestas, Quantum learning for neural associative memories, Fuzzy Sets and Systems, 157, 13: 1797-1813, 2006.

[28] G. Cattaneo, A. Leporati, R. Leporini, Quantum conservative many-valued computing, Fuzzy Sets and Systems, 159, 9: 1001-1030, 2008.

[29] Roberto Giuntini, Hector Freytes, Antonio Ledda, Francesco Paoli, A discriminator variety of Gödel algebras with operators arising in quantum computation, Fuzzy Sets and Systems, 160, 8: 1082-1098, 2009.

ISBN: 978-989-95079-6-8

Possibility and Necessity Evaluations based on Ordinal Comparability

Masaaki Ida

National Institution for Academic Degrees and University Evaluation
1-29-1, Gakuen-nishimachi, Kodaira, Tokyo 187-8587 Japan
Email: ida@niad.ac.jp

Abstract— *Subjective uncertainty is one of the most essential subjects for evaluation that was the reason Fuzzy theory was proposed. Related research areas are widely spread such as decision making, data analysis, information retrieval, psychology, and human computer interaction and so on. As a basis for various researches mathematical interpretation of evaluation methods occupies an important position. In Possibility theory, possibility and necessity evaluations have been utilized as ways for evaluating a vague achievement in an ambiguous situation. However, mathematical interpretations for these evaluations have been insufficient. In this article, possibility and necessity evaluations are examined based on the ordinal comparability in social choice theory. Equivalent conditions to the evaluations are deduced so that axiomatic interpretations for the evaluations become possible, and order relations related to the two evaluations are deduced.*

Keywords— evaluation, measure, necessity, ordinal comparability, ordinal utility, possibility.

1 Introduction

In Possibility theory [1, 2, 3, 4, 5], subjective value of attainment or satisfaction of an evaluation item in state x is represented by the membership function $\mu(x)(0 \leqq \mu(x) \leqq 1)$. And occurrence degree of subjective possibility is represented by the membership function $\pi(x)(0 \leqq \pi(x) \leqq 1)$. Tuple of these functions, $(\mu(x), \pi(x))$, can be regarded as a representation of state x with two kinds of uncertainties (vagueness and ambiguity) and be applied for evaluation and decision making.

An evaluation value for (μ, π) is defined as follows:

Definition 1

$$\Pi(\mu, \pi) = \sup_x \min(\mu(x), \pi(x)) \tag{1}$$

This can be regarded as an extension of possibility measure for fuzzy sets. This is also interpreted as a fuzzy integral for fuzzy set on possibility measure.

In Possibility theory, necessity measure is defined as dual for possibility evaluation:

Definition 2

$$N(\mu, \pi) = \inf_x \max(\mu(x), 1 - \pi(x)) \tag{2}$$

These two evaluations (possibility and necessity) also can be interpreted as optimistic (positive) and pessimistic (negative) evaluations respectively.

In this paper, we investigate generalized possibility and necessity evaluations. As seen in these definitions, generally in Fuzzy theory, *max* operation and *min* operation are used frequently. For example, they are used in the synthetic rule of fuzzy reasoning.

Moreover, some restrictions, e.g., *normality*, are assumed on membership functions in many cases. A continuity and convexity are sometimes assumed on membership functions. [3].

Definition 3 *A membership function $\pi(x)$ is normal means that there exists x such that $\pi(x) = 1$.*

In order to perform the mathematical interpretation on possibility and necessity evaluations and to examine the validity of the evaluations, it is necessary to consider on the property of functions $\mu(x)$ and $\pi(x)$ determined subjectively. For example, whether it is a cardinal utility function or ordinal utility function.

In Possibility theory, considerations for improving qualitative possibilistic criteria are discussed. Axiomatic approach has been made in order to examine decision theoretic foundation [6]. Generalized concordance rules that request comparisons within possible values for components and comparisons between sets of attributes is proposed [7]. The injection of lexicographic ingredients in the possibilistic criteria has been studied [8].

It is necessary to consider the conditions at the time of unifying the functions and performing synthetic evaluation, e.g., evaluation level comparability on evaluation items. In social choice theory or voting theory, the comparability of utility functions was introduced with ordinal number based on extended ordering [9, 10]. and examination has been made about how synthetic evaluation should be performed [11].

This paper newly defines possibility and necessity evaluations based on ordinal comparability and deduces equivalent conditions to them. Therefore, the applicable conditions of possibility and necessity evaluations are clarified. Moreover, relations between these two evaluations are discussed.

Hereafter, Section 2 examines necessary and sufficient condition of the evaluations based on extended ordering. In Section 3, interpretations of the conditions for evaluations, possibility, and necessity evaluation, are discussed. Section 4 considers order relations between possibility evaluation and necessity evaluation.

2 Extended ordering and preference relation

We denote that X is a set of states or a set of alternatives. We define *preference relation* as a binary relation.

Definition 4 *Preference relation R on X ($x, y \in X, xRy$: x is preferred to y) is a binary relation which is reflective*

ISBN: 978-989-95079-6-8

($\forall x \in X : xRx$), transitive ($\forall x, y, z \in X : (xRy \wedge yRz) \rightarrow xRz$). and total ($\forall x, y \in X : (x \neq y) \rightarrow (xRy \vee yRx)$). Strong preference relation P and indifference relation I are defined by: $xPy \equiv (xRy \wedge \neg(yRx))$, and $xIy \equiv (xRy \wedge yRx)$.

We denote H as a set of evaluation items. In social choice theory [9], X is called *a set of social states*, and H is called *a set of individuals*. In this paper, we make fundamental investigations on preference relation, so that the number of elements of the set H is restricted to two.

We consider $(x, i) \in X \times H$. A binary relation $\tilde{R}$ on $X \times H$ is assumed to be a preference relation. This relation is called *extended ordering* [9]. In similar meaning, it is also called extended sympathy [10].

Definition 5 *Preference relation $(x, i)\tilde{R}(y, j)$ means that an evaluation item i on a state x is preferred to j on a state y. Strong preference relation $\tilde{P}$ and indifference relation $\tilde{I}$ on $X \times H$ are defined by: $(x, i), (y, j) \in X \times H$ and $\tilde{R}$, $(x, i)\tilde{P}(y, j) \equiv ((x, i)\tilde{R}(y, j) \wedge \neg((y, j)\tilde{R}(x, i)))$, $(x, i)\tilde{I}(y, j) \equiv ((x, i)\tilde{R}(y, j) \wedge (y, j)\tilde{R}(x, i))$.*

Here, we consider generalized function $f : \tilde{R} \rightarrow R$, which corresponds the preference relation $\tilde{R}$ ($\tilde{P}$ and $\tilde{I}$) on $X \times H$ to the preference relation R (P and I) on X. The function f is called *Generalized Social Welfare Function* (GSWF) in social choice theory [9].

In the following section, higher evaluation on state x is denoted as $2(x) \in \{1, 2\}$ and lower evaluation is denoted as $1(x) \in \{1, 2\}$. This means that $(x, 2(x))\tilde{R}(x, 1(x))$.

Definition 6 *GSWF f^P is defined as follows. For all $x, y \in X$,*
(1) if for all $i \in \{1, 2\}$ $(x, i(x))\tilde{I}(y, i(y))$, then xIy,
(2) if $(x, 1(x))\tilde{I}(y, 1(y))$ and $(x, 2(x))\tilde{P}(y, 2(y))$, then xPy,
(3) if $(x, 1(x))\tilde{P}(y, 1(y))$, then xPy.

GSWF f^P corresponds to *leximin rule*, because we compare preference from lower level to higher level [9, 11].

Next, we consider following five conditions on the two preference relations $\tilde{R}$ and R. We denote $(x, i)\tilde{R}(y, i)\tilde{R}(z, i)$ for short as $(x, i)\tilde{R}(y, i)$ and $(y, i)\tilde{R}(z, i)$. We denote for $\tilde{P}$ and $\tilde{I}$ in similar way.

(C1) *Preference relation $\tilde{R}$ can take any logically possible relation.*

(C2) *If the restrictions of $\tilde{R}$ and $\tilde{R}'$ on any pair in X are the same, then the restrictions of R and R' on that pair are also same.*

(C3) *For any x, y in X, if $(x, 1)\tilde{R}(y, 1) \wedge (x, 2)\tilde{R}(y, 2)$, then xRy. If one of the two relations is a strong preference relation $\tilde{P}$, then xPy.*

(C4) *For any x, y in X, if $(y, 2)\ \tilde{P}\ (x, 2)\ \tilde{R}\ (x, 1)\ \tilde{P}\ (y, 1)$ or if $(y, 1)\ \tilde{P}\ (x, 1)\ \tilde{R}\ (x, 2)\ \tilde{P}\ (y, 2)$, then xPy.*

(C5) *For any x, y in X, if $(x, 1)\tilde{I}(y, 2) \wedge (x, 2)\tilde{I}(y, 1)$ and $\neg((x, 1)\tilde{I}(x, 2))$, then xIy.*

We note that if $(x, 1)\tilde{R}(y, 1) \wedge (x, 2)\tilde{R}(y, 2)$ and $(y, 1)\tilde{R}(x, 1) \wedge (y, 2)\tilde{R}(x, 2)$, then xIy by (C3). Furthermore we note that the conditions (C3), (C4), and (C5) are mutually independent.

For GSWF f^P and the five conditions, following theorem is stated [12].

Theorem 1 *Assume that $|X| > 2$, only GSWF that satisfies (C1), (C2), (C3), (C4), and (C5) is f^P.*

Next, from the view of symmetry, we consider a different condition only for (C4) in the above five conditions. We define the following converse condition (C4').

(C4') *For any x, y in X, if $(x, 2)\ \tilde{P}\ (y, 2)\ \tilde{R}\ (y, 1)\ \tilde{P}\ (x, 1)$ or if $(x, 1)\ \tilde{P}\ (y, 1)\ \tilde{R}\ (y, 2)\ \tilde{P}\ (x, 2)$, then xPy.*

In this case we consider the followings.

Definition 7 *GSWF f^N is defined as follows. For all $x, y \in X$,*
(1) for any $i \in \{1, 2\}$, if $(x, i(x))\tilde{I}(y, i(y))$, then xIy,
(2) if $(x, 2(x))\tilde{I}(y, 2(y))$ and $(x, 1(x))\tilde{P}(y, 1(y))$, then xPy,
(3) if $(x, 2(x))\tilde{P}(y, 2(y))$, then xPy.

For GSWF f^N and the five conditions, following theorem is stated.

Theorem 2 *Assume $|X| > 2$, only GSWF that satisfies conditions (C1), (C2), (C3), (C4'), (C5) is f^N.*

By the above two theorems, conditions equivalent to two GSWFs, f^P and f^N, were obtained. Based on these results we can discuss what kind of conditions are assumed when each GSWF is applied for evaluation.

3 Interpretation of possibility and necessity evaluation based on extended ordering

3.1 Ordinal utility function

In this section, as an application of the knowledge obtained in the previous section to Possibility theory, we consider the case that $H = \{\mu, \pi\}$. In the case we define GSWF f^P as a possibility evaluation.

We can say from (1) and the definition of f^P that if xPy for f^P, then $\min(\mu(x), \pi(x)) \geqq \min(\mu(y), \pi(y))$.

In the following, we discuss the interpretation of possibility evaluation based on the theorem obtained in the previous section.

We consider that $(x, i) \in X \times H$ corresponds to (x, μ) and (x, π). The relations between the preference relation $\tilde{R}$ and membership functions are defined as follows;
$(x, \mu)\tilde{R}(y, \mu) \equiv \mu(x) \geqq \mu(y)$,
$(x, \pi)\tilde{R}(y, \pi) \equiv \pi(x) \geqq \pi(y)$,
$(x, \mu)\tilde{R}(x, \pi) \equiv \mu(x) \geqq \pi(x)$,
$(x, \pi)\tilde{R}(x, \mu) \equiv \pi(x) \geqq \mu(x)$.
We denote strong preference relation and indifference relation as "$>$" and "$=$" respectively.

ISBN: 978-989-95079-6-8

We should note that membership functions μ and π can be regarded as **ordinal level comparable utility functions** based on extended ordering.

3.2 Possibility evaluation

In this subsection we discuss the conditions in Theorem 1.

Conditions (C1) and (C2) are naturally acceptable conditions.

Condition (C3) (if $\mu(x) \geqq \mu(y)$ and $\pi(x) \geqq \pi(y)$, then xRy) is so-called Pareto condition, so that this is naturally acceptable for the functions μ and π.

However, (C4) (if $\mu(y) > \mu(x) \geqq \pi(x) > \pi(y)$ or $\pi(y) > \pi(x) \geqq \mu(x) > \mu(y)$, then xRy) means that we must compare two evaluation elements of μ and π on one state.

And condition (C5) (if $\mu(x) = \pi(y)$ and $\pi(x) = \mu(y)$, then xIy) means that in indifferent two states, two evaluation elements of μ and π can be exchanged as same.

Therefore conditions (C4) and (C5) for the functions μ and π are restrictions on two evaluations, then these restrictions are strongly imposed on setting these two membership functions.

If conditions (C1), (C2), (C3), (C4), (C5) are acceptable for μ and π, then by Theorem 1, we can utilize the function f^P as a preference functions on $x, y \in X$.

On the other hand, in the case that we use the function f^P as a preference relation, the conditions (C1), (C2), (C3), (C4), (C5) should be satisfied.

Consequently, for $x, y \in X$ we clarify the conditions for evaluation of (1) based on comparing $\min(\mu(x), \pi(x))$ and $\min(\mu(y), \pi(y))$.

3.3 Necessity evaluation

Similarly we can discuss necessity evaluation. In case of necessity, for π in (2) we can see $1 - \pi(x)$. Therefore, as a converse order relation we defined an ordinal utility function $\pi^*(x)$ as follows.

Definition 8 *$\pi^*(x)$ is defined by:*

$$\pi^*(x) < \pi^*(y) \quad \equiv \quad \pi(x) > \pi(y), \tag{3}$$

$$\pi^*(x) = \pi^*(y) \quad \equiv \quad \pi(x) = \pi(y). \tag{4}$$

Similarly as possibility evaluation we can discuss the interpretation of necessity evaluation based on Theorem 2 obtained in the previous section.

We consider that $(x, i) \in X \times H$ corresponds to (x, μ) and (x, π^*) in the same way as possibility. The relation between the preference relation $\tilde{R}$ and membership functions are defined as follows;

$(x, \mu)\tilde{R}(y, \mu) \equiv \mu(x) \geqq \mu(y)$,
$(x, \pi^*)\tilde{R}(y, \pi^*) \equiv \pi^*(x) \geqq \pi^*(y)$,
$(x, \mu)\tilde{R}(x, \pi^*) \equiv \mu(x) \geqq \pi^*(x)$,
$(x, \pi^*)\tilde{R}(x, \mu) \equiv \pi^*(x) \geqq \mu(x)$.

We denote strong preference relation and indifference relation as "$>$" and "$=$" respectively.

We should note that membership functions μ and π^* can be regarded as ordinal comparable utility functions based on extended ordering. As similar with possibility, we can discuss μ, π^*, and conditions (C1), (C2), (C3), (C4'), (C5).

4 Order relation between possibility evaluation and necessity evaluation

In this section, as discussed in earlier sections we assume that $H = \{\mu, \pi\}$ and that extended ordering on (x, μ) and (x, π) satisfies conditions (C1), (C2), (C3), (C4), and (C5) for preference relations $\tilde{R}$ and R.

Definition 9 *$x^P \in X$ is defined by: for all $y(\neq x^P) \in X$, $x^P R y$.*

Similarly we assume that $H = \{\mu, \pi^*\}$ and that extended ordering on (x, μ) and (x, π^*) satisfies conditions (C1), (C2), (C3), (C4'), and (C5) for preference relations $\tilde{R}^*$ and R^*. We discriminate $\tilde{R}, R$ and $\tilde{R}^*, R^*$.

Definition 10 *$x^N \in X$ is defined by: for all $y(\neq x^N) \in X$, $y R^* x^N$.*

We should recall that $\mu(x)$ and $\pi(x)$ are ordinal functions.

In Possibility theory, optimistic evaluation and pessimistic evaluation are discussed. In this section, order relation between possibility evaluation and necessity evaluation is discussed.

We obtain the following result with respect to $\mu(x^P)$ and $\mu(x^N)$.

Lemma 3

$$\mu(x^P) \geqq \mu(x^N). \tag{5}$$

(Proof)
Assuming $\mu(x^P) < \mu(x^N)$ we deduce contradiction. In this case, $x^P \neq x^N$. If $\pi(x^P) \leqq \pi(x^N)$, then $(x^N, \mu)\tilde{P}(x^P, \mu)$ and $(x^N, \pi)\tilde{R}(x^P, \pi)$. Therefore, by (C3) of GSWF f^P, $x^N P x^P$. This contradicts the definition of x^P. Next, we consider the case that $\pi(x^P) > \pi(x^N)$. In this case, $\pi^*(x^P) < \pi^*(x^N)$ and $\mu(x^P) < \mu(x^N)$, then $(x^N, \mu)\tilde{P}^*(x^P, \mu)$ and $(x^N, \pi)\tilde{P}^*(x^P, \pi)$. By (C3) of GSWF f^N, $x^N P^* x^P$, so that this contradicts the definition of x^N.

For order relations with respect to $\max(\mu(x^N), \pi^*(x^N))$ and $\min(\mu(x^P), \pi(x^P))$, we obtain the following results.

Lemma 4 *If $\pi^*(x) > \pi(x)$, then*

$$\max(\mu(x^N), \pi^*(x^N)) > \min(\mu(x^P), \pi(x^P)). \tag{6}$$

Lemma 5 *If there exists x' which satisfies the condition $\pi(x') \geqq \mu(x') \geqq \pi^*(x')$, then*

$$\min(\mu(x^P), \pi(x^P)) \geqq \mu(x') \geqq \max(\mu(x^N), \pi^*(x^N)). \tag{7}$$

(Proof)
By the definition of x^P, we have $\min(\mu(x^P), \pi(x^P)) \geqq \min(\mu(x'), \pi(x')) \geqq \mu(x')$. Similarly with respect to x^N, $\max(\mu(x^N), \pi^*(x^N)) \leqq \max(\mu(x'), \pi^*(x')) \leqq \mu(x')$. Consequently we obtain the result.

Next we discuss in the framework of conventional Possibility theory, i.e., we consider $\pi^*(x) = 1 - \pi(x)$. x^P is most preferable on X with respect to GSWF f^P. Then x^P takes the optimal value of (1) in Definition 1. Similarly, x^N is most unpreferable on X with respect to GSWF f^N. Then x^N takes the optimal value of (2) in Definition 2.

ISBN: 978-989-95079-6-8

Some assumptions are sometimes taken in Possibility theory. In this section we assume that the membership function $\pi(x)(0 \leqq \pi(x) \leqq 1)$ satisfies normality condition, i.e., there exists the $x' \in X$ such that $\pi(x') = 1$. Then we obtain the following result by Lemma 5.

Theorem 6 *If $\pi(x)$ is normal, then*

$$\Pi(\mu, \pi) \geqq N(\mu, \pi). \tag{8}$$

(Proof) By the definition of normality there exists x' that $\pi^*(x') = 0$ and $\pi(x') = 1$, and $0 \leqq \mu(x') \leqq 1$, then the prerequisite condition of Lemma 5 is satisfied.

5 Conclusion

This paper described the mathematical foundation of subjective uncertainty. Axiomatic interpretations on possibility evaluation and necessity evaluation, and the relation between these evaluations were examined based on the extended ordering in social choice theory.

The obtained results are as follows:

In section 2, based on extended ordering we obtained the equivalent conditions for two kinds of GSWF f^P and f^N; conditions (C1), (C2), (C3), (C4), (C5), and conditions (C1), (C2), (C3), (C4'), (C5).

In section 3, we examined generalized possibility evaluation and necessity evaluation, and discussed the conditions obtained in section 2.

In section 4, we defined x^P and x^N, and discussed the order relations between possibility evaluation and necessity evaluation, and obtained some results on the relations.

By these results obtained in this paper, mathematical interpretations and applicable conditions of the evaluations became clear. These are important when the evaluations are applied in various practical situations.

References

[1] L.A. Zadeh. Fuzzy Sets as a Basis for a Theory of Possibility, *Fuzzy Sets and Systems*, 1:3–28, 1978.

[2] D. Dubois, and H. Prade. *Possibility Theory*, Plenum Press, 1988.

[3] H. Tanaka, and P. Guo. *Possibilistic Data Analysis for Operations Research*, Physica-Verlag, 1999.

[4] T. Murofushi, and M. Sugeno. An Interpretation of Fuzzy Measures and the Choquet Integral as an Integral with Respect to a Fuzzy Measure, *Fuzzy Sets and Systems*, 29:201–227, 1989.

[5] G. Klir. *Uncertainty and Information*, Wiley, 2006.

[6] D. Dubois, H. Prade, and R. Sabbadin. Decision-theoretic Foundations of Qualitative Possibility Theory, *European Journal of Operational Research*, 128:459–476, 2001.

[7] D. Dubois, H. Fargier, P. Perny, and H. Prade. A Characterization of Generalized Concordance Rules in Multicriteria Decision Making. *International Journal of Intelligent Systems*, 18:751–774, 2003.

[8] H. Fargier and R Sabbadin. Qualitative decision under uncertainty: Back to expected utility. *Artificial Intelligence*, 164:245–280, 2005.

[9] A. Sen. *Collective Choice and Social Welfare*, Holden-Day, 1970.

[10] K.J. Arrow, A. Sen, and K. Suzumura. *Handbook of Social Choice and Welfare*, Elsevier, 2002.

[11] P.J. Hammond. Equity, Arrow's Conditions, and Rawls' Difference Principle, *Econometrica*, 44(4):793–804, 1976.

[12] M. Ida. Possibility Evaluation based on Extended Ordering, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 19:41–46, 2007.

ISBN: 978-989-95079-6-8

Fuzzy answer set programming with literal preferences

Jeroen Janssen[1*] Steven Schockaert[2†] Dirk Vermeir[1] Martine De Cock[3‡]

1. Department of Computer Science, Vrije Universiteit Brussel, Belgium
2. Department of Applied Mathematics and Computer Science, Universiteit Gent, Belgium
3. Institute of Technology, University of Washington, Tacoma, USA
Email: {jeroen.janssen,dvermeir}@vub.ac.be, steven.schockaert@ugent.be, mdecock@u.washingon.edu

Abstract— In the current approaches to fuzzy answer set programming (FASP) one can state preferences amongst rules to denote their relative importance. However, in many situations we need more complex preferences such as those in the answer set optimization framework proposed by Brewka for crisp answer set programming. Unfortunately, these complex preferences do not readily fit into the current FASP approaches. In this paper we propose a language to state such preferences and show that programs with these preferences can be translated into equivalent general fuzzy answer set programming (gFASP) programs. This not only provides an implementation method, but also shows that this extension can be added as syntactic sugar on top of general fuzzy answer set programming solvers.

Keywords— Logic Programming, Fuzzy Logic, Answer Set Programming, Preference-based Logic Programming.

1 Introduction

Answer set programming [1] is a logic programming language based on the stable model semantics [2]. Roughly speaking, in answer set programming a program contains a set of *rules* of the form $a \leftarrow b_1, \ldots, b_n, \mathit{not}\ c_1, \ldots, \mathit{not}\ c_m$ describing a problem whose solution is generated by an *answer set solver*. For example, if one is interested in finding solutions to the problem of assigning a seat s to either of two persons p_1 and p_2, the following program could be used:

$$\begin{array}{lcl} sit(p_1, s) & \leftarrow & \mathit{not}\ sit(p_2, s) \\ sit(p_2, s) & \leftarrow & \mathit{not}\ sit(p_1, s) \end{array}$$

This program has two *answer sets*, which correspond to the solutions of the modeled problem, $S_1 = \{sit(p_1, s)\}$ and $S_2 = \{sit(p_2, s)\}$.

In recent work, answer set programming and, more general, logic programming have been extended to deal with fuzzy [3, 4], probabilistic [5] and many-valued logics [6, 7, 8, 9], allowing for a more flexible description of both the problem domain and the proposed solutions of the problem domain. Specifically, in [4] degrees are attached to answer sets describing the quality of prospective solutions. For example, in case one is interested in only allowing solutions in which people are seated close to friends, one could write a program containing fuzzy *constraints*, as follows:

$$0 \leftarrow sit(p_1, s_1) \wedge sit(p_2, s_2) \wedge \mathit{friend}(p_1, p_2) \wedge \sim near(s_1, s_2)$$

where $sit(p, s)$ means that person p sits on seat s, $\mathit{friend}(p_1, p_2)$ that person p_1 is a friend of p_2 (to a certain degree), $near(s_1, s_2)$ denotes the proximity of seats s_1 and s_2 and $\sim$ is a negator. The constraint above would then attach a higher degree of quality to answer sets in which friends are close together. Note that $\sim near(s_1, s_2)$ should be used and not $near(s_1, s_2)$ since the satisfaction of the constraint rule should increase with the value of $near(s_1, s_2)$. More details can be found in Section 2.

Although the quality degree allows to express preference amongst solutions, the current proposals only allow this preference to be based on the degree of rule fulfillment. However, in many practical situations, we would like to express preferences based on the truth degrees of literals. Furthermore, we might want to express that the importance of some preferences depends on the context, i.e. we would like to have conditional preferences. For example, we wish to be able to state that if we are sitting on a seat, we prefer it to be by the window, but only consider this important when the scenery is nice, as follows:

$$scene : (sit(p, s) \rightarrow nearWin(s)) \triangleleft nicescenery \qquad (1)$$

where $nearWin(s)$ denotes that seat s is near the window, the $nicescenery$ literal determines the weight of importance of the rule and $\rightarrow$ is an implicator. The exact semantics of such rules are given in Section 3.

In the crisp case, Brewka proposed rules like (1) in [10]. However, to the best of our knowledge no such proposal exists for fuzzy answer set programs, which we remedy in this paper. The main contributions of this paper are the following:

- In Section 3 we propose a generalisation of the *Preference Description Language* (PDL) of Brewka to the fuzzy case.
- In Section 4 we show how programs containing these preferences can be translated to standard general fuzzy answer set programming (gFASP) programs [11], providing an implementation method and hence also showing that this extension is actually syntactic sugar on the general fuzzy answer set programming language.

2 General Fuzzy Answer Set Programming

General fuzzy answer set programming (gFASP) programs over a complete lattice $\mathcal{L}$ consist of *rules*, which are objects of the form

$$r\colon a \leftarrow f(b_1, \ldots, b_n)$$

*Funded by a joint Research Foundation–Flanders (FWO) project

†Postdoctoral fellow of the Research Foundation–Flanders (FWO)

‡On leave from Ghent University

ISBN: 978-989-95079-6-8

where r is the label of the rule, a and b_i, $1 \leq i \leq n$, are either elements from $\mathcal{L}$ or propositional symbols called *literals* (where $b_i = b_j$ for $i \neq j$ is possible), f is a (computable) function from $\mathcal{L}^n$ to $\mathcal{L}$ such that for $b_k, b'_k \in \mathcal{L}$ it holds that if $b_k \leq b'_k$, either $f(b_1, \ldots, b_k, \ldots, b_n) \leq f(b_1, \ldots, b'_k, \ldots, b_n)$ or $f(b_1, \ldots, b_k, \ldots, b_n) \geq f(b_1, \ldots, b'_k, \ldots, b_n)$ and $\leftarrow$ denotes a residual implicator over $\mathcal{L}$. If $a \in \mathcal{L}$, the rule is called a *constraint*. The head a of a rule r is denoted as r_h and the body $f(b_1, \ldots, b_n)$ is denoted as r_b.

The semantics are given by *interpretations*, i.e. fuzzy sets over the literals of a program. These are extended to give meaning to rules in a straightforward fashion, i.e. if I is an interpretation, $[a \leftarrow f(b_1, \ldots, b_n)]_I = [a]_I \leftarrow [f(b_1, \ldots, b_n)]_I$, where $[a]_I = I(a)$ if a is a literal and $[a]_I = a$ if $a \in \mathcal{L}$, and $[f(b_1, \ldots, b_n)]_I = f([b_1]_I, \ldots, [b_n]_I)$.

Models of a program P are defined using an *aggregator expression* $\mathcal{A}_P$ which maps *rule propositions* (corresponding to rule labels) to a value in a quasi-ordered set $\mathcal{Q}$. For example, given a program P_1 with two rules r_1 and r_2, a possible aggregator would be $\mathcal{A}_{P_1} = r_1 + r_2$. Such an expresssion can be evaluated by an interpretation, e.g. for I an interpretation of P we obtain $[\mathcal{A}_{P_1}]_I = \gamma([r_1]_I) + \gamma([r_2]_I)$, with γ an order-preserving $\mathcal{L} \rightarrow \mathcal{Q}$ mapping. A *k-model*, $k \in \mathcal{Q}$, is then an interpretation for which $[\mathcal{A}_P]_I \geq k$.

k-Answer sets of a program P are k-models that contain the maximal amount of knowledge inferrable from a program through forward chaining, without resorting to external hypotheses, i.e. they are models for which the truth value of every literal is supported by the application of a rule in the program.

As an example, consider program P_{seat}, describing seating arrangements, with the following rulebase $\mathcal{R}_{P_{seat}}$ over the lattice $([0,1], \leq)$:

$$
\begin{array}{rrll}
gen\colon & sit(P,S) & \leftarrow_g & 1 \\
c_1\colon & 0 & \leftarrow_g & sit(P,S) \wedge sit(P',S) \wedge P \neq P' \\
c_2\colon & 0 & \leftarrow_g & sit(P,S) \wedge sit(P,S') \wedge S \neq S' \\
cr\colon & 0 & \leftarrow_g & sit(P,S) \wedge (1 - sit(P,S)) \\
u\colon & unh(P) & \leftarrow_g & sit(P,S) \wedge friend(P,P') \\
 & & & \wedge sit(P',S') \wedge (1 - near(S,S')) \\
q\colon & 0 & \leftarrow_l & unh(P) \\
s\colon & s(P) & \leftarrow_g & sit(P,S) \\
a\colon & 0 & \leftarrow_g & 1 - s(P)
\end{array}
$$

where $\wedge$ is the minimum t-norm over $([0,1,\leq)$ and $\leftarrow_l$, $\leftarrow_g$ are resp. the Łukasiewicz and Gödel implicator. The aggregator is defined as $\mathcal{A}_{P_{seat}} = \inf_{r \in \mathcal{R}_{P_{seat}} \setminus \{gen\}} r$. Due to the fact that the gen-rule is not incorporated in the aggregation expression, this rule generates random truth degrees for $sit(P,S)$. To ensure that $sit(P,S)$ is either 0 or 1, we use the cr (crispify) rule. Rules c_1 and c_2 guarantee that a seat is occupied by a single person, resp. that a person only occupies one seat. The a and s rules ensure that everyone has a seat and the u and q rules are used to prefer solutions where friends are seated close together. Note that due to grounding a rule such as $s(P) \leftarrow_g sit(P,S)$ actually denotes a set of rules $\{s(p_i) \leftarrow_g sit(p_i, s_j)\}$ for a set of persons $(p_i)_{i \in I}$ and a set of seats $(s_j)_{j \in J}$.

If we have three seats s_1, s_2 and z, only two of which are near, viz. $near(s_1, s_2)^{\cdot 8}$ and $near(s_2, s_1)^{\cdot 8}$ and there are three persons a, b and c connected by friendship degrees $friend(a,b)^{\cdot 8}$, $friend(a,c)^{\cdot 5}$, and $friend(b,c)^0$, then from program P_{seat} one could e.g. derive the .2-answer set $A_1 = \{sit(a,s_1)^1,\ sit(b,z)^1,\ sit(c,s_2)^1,\ unh(a)^{\cdot 8},\ \ldots\}$ and the (better) .5-answer set $A_2 = \{sit(a,s_1)^1,\ sit(b,s_2)^1,\ sit(c,z)^1,\ unh(a)^{\cdot 5}, \ldots\}$.

For more details on the gFASP framework, we refer the reader to [11].

3 Complex Preferences on Literals

3.1 Fuzzy preference rules: syntax

In this section, we build a fuzzy preference framework based on a *fuzzy preference description language* ($FPDL$) which is an extension of the language PDL proposed by Brewka in [10]. Formally, a gFASP program with literal preferences consists of a tuple (P, e), where P is a gFASP program and e is an expression from $FPDL$. This expression determines a preference ordering on answer sets based on truth degrees of literals.

The basic elements of this language are *fuzzy preference rules*. Note that although they are called *rules*, they should not be confused with the rules of a fuzzy answer set program introduced in Section 2: the former are used for creating preference expressions (the e in (P, e)), whereas the latter are used to create (general) fuzzy answer set programs (the P in (P, e)). Formally the fuzzy preference rules are defined over a set of literals L and the complete lattice $\mathcal{L}$ as expressions of the following form:

$$r : g(a_1, \ldots, a_n) \triangleleft f(b_1, \ldots, b_m) \tag{2}$$

In this rule, the a_is ($1 \leq i \leq n$) and b_js ($1 \leq j \leq m$) are literals from L, the function g is a mapping from $\mathcal{L}^n$ to $\mathcal{L}$, f is a mapping from $\mathcal{L}^m$ to $\mathcal{L}$ and r is the label of the rule. For convenience, for a given fuzzy preference rule r we also refer to $g(a_1, \ldots, a_n)$ as r_h (the *head* of the rule) and to $f(b_1, \ldots, b_m)$ as r_b (the *body* of the rule).

Intuitively, the g-function denotes an expression determining the *suitability* of an answer set with respect to a certain preference and the f-function denotes the *applicability* of the rule, meaning the conditions under which this preference rule is considered important or relevant. For example, consider the following preference rule:

$$scene : (sit(p,s) \rightarrow nearWin(s)) \triangleleft nicescenery$$

This rule states that by default we (with "we" being person p) would like to have a seat near the window. However, we only consider this requirement important to the degree that the scenery is actually nice. The reason for choosing g to be an implication in this rule is that we want to state that this specific rule is also fulfilled (thus the answer set considered is still "suitable") if we are not sitting at this seat. Note that there is a difference between the $\rightarrow$ in the head and the $\triangleleft$ of the rule: if we would rewrite the scene rule as

$$scene' : (sit(p,s) \wedge nicescenery \rightarrow nearWin(s)) \triangleleft 1$$

it would denote that the nicer the scenery, the closer to the window we would like to sit rather than that we only find it important to sit near the window when the scenery is nice.

ISBN: 978-989-95079-6-8

This can also be seen from the fact that if $\rightarrow$ is a residual implicator and $nicescenery$ is true to a degree of 0.5, the $scene$ rule is only fully fulfilled when $nearWin(s)$ is 1 (or $sit(p,s)$ is 0), whereas in the $scene'$ rule it suffices that $nearWin(s)$ is 0.5 (or $sit(p,s)$ is 0).

Intuitively we thus prefer answer sets with a better suitability score, but can forgive the answer set a low suitability score if the applicability degree is rather low. Formally the semantics of rules are dependent on a function $suit$ and app defined for a given fuzzy preference rule r and an arbitrary answer set A as

$$suit(A,r) = [g(a_1,\ldots,a_n)]_A = g(A(a_1),\ldots,A(a_n))$$

$$app(A,r) = [f(b_1,\ldots,b_m)]_A = f(A(b_1),\ldots,A(b_m))$$

In the approach of Brewka [10], the semantics of preference rules are defined by attaching a *penalty* score in $\mathbb{R}$ to each applicable rule. Such a penalty reflects how good an answer set fulfills a preference rule, where lower penalties correspond to better fulfillment. Inapplicable rules are not considered important and are therefore attached penalty 0, i.e. they are always optimally fulfilled.

Unfortunately, the same approach is not possible when generalizing from crisp to fuzzy answer set programming: in the fuzzy case the applicability value is in $[0,1]$ and by reducing the values $suit(A,r)$ and $app(A,r)$ to a single number (i.e. penalty score in $\mathbb{R}$), we lose the important semantic distinction between suitability of answer sets and applicability of preference rules. In the following sections, we define an alternative semantics for fuzzy preference rules which does adhere to this distinction. We will first give the semantics for rules with a crisp applicability score and then define penalty expressions built from rules; starting from the semantics thus obtained we then introduce the semantics for rules with a fuzzy applicability score using the recently introduced concept of gradual number [12]; finally we introduce the general penalty expressions built from rules with fuzzy applicabilities.

3.2 Rules with crisp applicability

Suppose r is a rule with a crisp applicability condition (i.e. $app(A,r)$ is either 1 or 0), then we define the penalty score as:

$$pen^c(A,r) = \begin{cases} \Xi(suit(A,r)) & \text{if } app(A,r) = 1 \\ 0 & \text{otherwise} \end{cases}$$

where Ξ is a decreasing $\mathcal{L}$ to $[0,+\infty[$ mapping, such that $\Xi(1) = 0$, in the examples in this paper usually assumed to be $\Xi(x) = 1 - x$.

As an example, consider the following rule:

$$VIP_p : goodSeat(p) \triangleleft vip(p)$$

where $goodSeat(p) \in [0,1]$ reflects how good the seat of person p is and $vip(p) \in \{0,1\}$ denotes whether person p is a VIP. This rule encodes that we find it important that VIPs have good seats. If $A(goodSeat(p)) = 0.8$ and $A(vip(p)) = 1$, we know from the definition of pen^c that $pen^c(A,r) = 1 - 0.8 = 0.2$. Hence, answer set A has a low penalty score and thus represents a seating configuration that satisfies our wishes.

Starting from rules with crisp applicability, we can define the set of complex penalty expresssions $FPDL^p$ as follows:

1. If r is a fuzzy preference rule whose applicability condition takes only crisp values, r is in $FPDL^p$;
2. If $e_1, e_2, \ldots, e_k$ are in $FPDL^p$, then so are $(sum\ e_1,\ldots,e_k)$ and $(max\ e_1,\ldots,e_k)$.

We extend the definition of pen^c to cover complex penalty expressions in a straightforward way, e.g.

$$pen^c(A, sum\ e_1, ..., e_k) = pen^c(A,e_1) + ... + pen^c(A,e_k)$$

For example, suppose we have two VIP_p rules for persons p_1 and p_2, then we can combine these into one penalty expression summing the penalties as

$$e_v : (sum\ VIP_{p_1}, VIP_{p_2})$$

If we then have an answer set A for which $pen^c(A, VIP_{p_1}) = 0.2$ and $pen^c(A, VIP_{p_2}) = 0.5$, the total penalty of this expression is $pen^c(A, (sum\ VIP_{p_1}, VIP_{p_2}) = 0.7$.

3.3 Rules with vague applicability

In the general case, where applicability conditions are vague, we define penalties as gradual numbers in the sense of [12], i.e. as mappings from $]0,1]$ to $\mathbb{R}$. For each threshold λ in $]0,1]$, we convert preference rules r of the form (2) to preference rules r_λ with a crisp applicability condition:

$$r_\lambda : g(a_1,\ldots,a_n) \triangleleft (f(b_1,\ldots,b_m) \geq \lambda) \tag{3}$$

The (gradual) penalty of A w.r.t. a preference rule r is then given by the gradual number $pen(A,r)$, defined for each λ in $]0,1]$ as

$$pen(A,r)(\lambda) = pen^c(A, r_\lambda) \tag{4}$$

In Figure 1, one can see an example of the penalty thus induced by an answer set A on the rule VIP_p where $A(vip(p)) = app(A, VIP_p) = 1/2$ and $A(goodSeat(p)) = suit(A, VIP_p) = 0.75$. Note that for each rule the set $\{\lambda \mid pen(A,r)(\lambda) = \Xi(suit(A,r))\}$ is the half-open interval $]0, app(A,r)]$ and the set $\{\lambda \mid pen(A,r)(\lambda) = 0\}$ is the half-open interval $]app(A,r), 1]$. Hence, an answer set is only accountable (gets a penalty) w.r.t. a rule to the extent that the rule is applicable for the answer set.

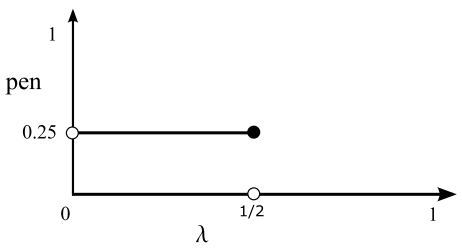


Figure 1: Example rule penalty

These penalties can easily be extended to complex preference expressions from a set $FPDL^{gp}$, defined analogously as $FPDL^p$ above, but starting from arbitrary preference rules. To define the semantics for expressions from $FPDL^{gp}$, we rely on the semantics of corresponding expressions from

ISBN: 978-989-95079-6-8

$FPDL^p$. Specifically, for e in $FPDL^{gp}$ and λ in $\mathcal{L}$, we define

$$pen(A, e)(\lambda) = pen^c(A, e^\lambda)$$

where e^λ is obtained from e by replacing all occurrences of preference rules r by the corresponding preference rules r_λ. Note that in principle, aggregation strategies that operate directly on gradual penalties could be defined as well.

Figure 2 depicts $pen(A, e)$ (the solid line) and $pen(A', e)$ (the dashed line) for an expression of the form $e = (sum\ dance, win)$ where the rule

$$dance : (sit(p, s) \rightarrow nearDanceFl(s)) \triangleleft nearFriends(p)$$

encodes that person p would like to sit near the dancefloor, but only finds this important if he is sitting close to his friends. The other rule is defined as

$$win : (sit(p, s) \rightarrow nearWin(s)) \triangleleft nearColleagues(p)$$

and denotes that person p likes a window seat, but only finds this important if he is sitting close to his colleagues. The answer sets A and A' take on the following applicability and suitability values:

app	$dance$	win
A	0.375	0.625
A'	0.75	0.5

$suit$	$dance$	win
A	0.18	0.82
A'	0.8	0.95

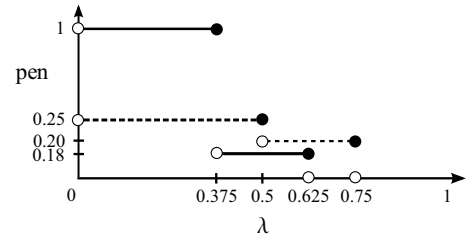


Figure 2: Example complex penalties

Now, given Figure 2, which answer set is preferred: A or A'? This depends on the application: if one only wants to look at the higher (more important) lambdas, we would prefer A as for $\lambda = 0.75$ this answer set has penalty 0; however if the global penalty score over all lambdas is important, we would prefer A' as A has a high penalty for values lower than 0.375 (viz. 1). Hence many options are viable, depending on the application. Furthermore there is a need to state preferences amongst expressions themselves so that we can e.g. state that only satisfying the $dance$ rule is better than only satisfying the win rule. In the next section we show how such orderings of answer sets based on the penalties and the preferences amongst preference expressions can be stated.

3.4 Preference strategies

Using penalties, we can define preference orderings on answer sets, which are referred to as *strategies*. Various such strategies can be specified. A natural strategy if we are comparing answer sets on the basis of one expression is to define the preference ordering $\leq$ between answer sets, based on an ordering $\preceq$ of gradual numbers. This strategy is denoted by $(\preceq\ e)$, and the corresponding preference ordering $\leq$ is defined for answer sets A_1 and A_2 as

$$A_1 \leq A_2 \equiv pen(A_1, e) \succeq pen(A_2, e)$$

Several partial and total orderings between gradual numbers can be used to this end. When the applicability values are interpreted as priorities, the total ordering $\preceq_1$ can be used, where $g_1 \preceq_1 g_2$ for gradual numbers g_1 and g_2 whenever

1. $g_1 = g_2$; or
2. $g_1(\lambda) < g_2(\lambda)$ for some λ in $]0, 1]$, and $g_1(\lambda') = g_2(\lambda')$ for all $\lambda' > \lambda$.

A useful partial ordering is the pointwise extension of the natural ordering $\leq$ on $\mathbb{R}$, denoted by $\preceq_2$, i.e. $g_1 \preceq_2 g_2$ iff $g_1(\lambda) \leq g_2(\lambda)$ for all λ in $]0, 1]$. Finally, an averaging strategy $\preceq_3$ could be used where $g_1 \preceq_3 g_2$ iff $\int_0^1 g_1(\lambda)w(\lambda)\,\mathrm{d}\lambda \leq \int_0^1 g_2(\lambda)w(\lambda)\,\mathrm{d}\lambda$. Intuitively, we may want to give more weight to the values of g_1 and g_2 for larger values of λ. This is encoded by using an increasing $]0, 1]$–$]0, 1]$ mapping w, defined, for instance, by $w(\lambda) = \lambda$ for all λ in $]0, 1]$.

As an example, consider Figure 2 with expression e from Section 3.3 again. From the picture one can see that if $\preceq_1$ is used, answer set A would be preferred over A' as for $\lambda = 0.75$ we have that $pen(A, e)(\lambda) < pen(A', e)(\lambda)$ and for any $\lambda > 0.75$ it holds that $pen(A, e)(\lambda) = 0 = pen(A', e)(\lambda)$. Hence $\preceq_1$ encodes a form of priority where we discriminate answer sets on their penalties for rules with a high applicability. Using $\preceq_2$, we would obtain that A and A' are incomparable. Thus $\preceq_2$ takes a more global approach, only considering preference when one of the two answer sets globally imposes a lower penalty score. With $\preceq_3$ and $w(\lambda) = \lambda$ for each $\lambda \in]0, 1]$ then, we obtain that A' would be preferred over A, due to the fact that A has a very high penalty at lower λ scores and is close to the penalty of A at the more important λ values. We can thus conclude that $\preceq_3$ can be used when one is not only interested in the fact that one penalty is higher than the other, but also in the exact value of these penalties.

The $\preceq_i$ $(1 \leq i \leq 3)$–strategies are not sufficient however: if we have two preference rules r_1 and r_2 and we want r_1 to be more important than r_2, we cannot readily encode this using a combination of the sum or max expressions and one of the orderings on the gradual penalties. To this end, we introduce other strategies, encoded as expressions from the set $FPDL$. First we have $(\preceq\ e) \in FPDL$ for $e \in FDPL^{gp}$ and $\preceq$ a partial ordering between gradual numbers. In analogy with Brewka [10], we also consider the following expressions:

1. If $e_1, \ldots, e_k$ are in $FPDL$, then $(pareto\ e_1, \ldots, e_k)$ and $(lex\ e_1, \ldots, e_k)$ are in $FPDL$
2. If $e_1, \ldots, e_k$ are in $FPDL^{gp}$ then $(rinc\ e_1, \ldots, e_k)$ and $(rcard\ e_1, \ldots, e_n)$ are in $FPDL$

The semantics of $(pareto\ e_1, \ldots, e_k)$ and $(lex\ e_1, \ldots, e_k)$ are defined analogously to the crisp case. In particular, the ordering $\leq$ induced by $(pareto\ e_1, \ldots, e_k)$ is defined by

$$A_1 \leq A_2 \equiv A_1 \leq_1 A_2 \wedge \ldots \wedge A_1 \leq_k A_2$$

ISBN: 978-989-95079-6-8

where $\preceq_i$ is the ordering induced by e_i. As an example, consider the expression $(pareto\ \ (\preceq_1\ \ dance), (\preceq_1\ \ win))$ with $dance$ and win as defined in Section 3.3. The preference ordering induced by this expression does not favor either of the expressions $(\preceq_1\ \ dance)$ and $(\preceq_1\ \ win)$.

Similarly, the ordering induced by $(lex\ e_1, \ldots, e_k)$ is given by

$$A_1 \leq A_2 \equiv \forall j \in 1..k \cdot A_1 \preceq_j A_2 \\ \vee \exists j \in 1..k \cdot A_1 \prec_j A_2 \wedge \forall j' < j \cdot A_1 \preceq_{j'} A_2$$

Expression $(lex\ (\preceq_1\ \ dance), (\preceq_1\ \ win))$ states that we first need to look whether we can prefer one of two answer sets A and A' by the expression $(\preceq_1\ \ dance)$. If this is not the case, we look whether $(\preceq_1\ \ win)$ can be used to prefer one of the two. Intuitively we thus prefer answer sets where person p is close to the dancefloor, even if that means he is sitting far from the window and close to his colleagues.

The $pareto$ and lex strategies both take arbitrary $FPDL$ expressions as their arguments. The next two strategies we introduce, $rinc$ and $rcard$, however, have expressions from $FPDL^{gp}$ as arguments. The difference with the $\preceq_i$ ($1 \leq i \leq 3$) strategies introduced earlier is that $rinc$ and $rcard$ base the preferences on more than one penalty expression and thus allow to have some interplay between rules and penalty expressions.

The semantics of $(rinc\ e_1, \ldots, e_k)$ and $(rcard\ e_1, \ldots, e_k)$ are defined in terms of the interval-valued fuzzy set P^a_A in the finite universe $\{1, \ldots, k\}$, where A is an answer set, $a \in [0, +\infty[$ and $L^a_A(i) = \{\lambda \mid pen(A, e_i)(\lambda) = a\}$:

$$P^a_A(i) = \begin{cases} [\inf L^a_A(i), \sup L^a_A(i)] & \text{if } L^a_A(i) \neq \emptyset \\ [0, 0] & \text{otherwise} \end{cases}$$

Intuitively, P^a_A is the set of (indices of) subexpressions e_i that attach a penalty a to an answer set A. Due to penalties being gradual numbers and because of the specific form of $pen(A, e)(\lambda)$ for any penalty expression e, P^a_A is an interval-valued fuzzy set.

The preference ordering induced by $(rinc\ \ e_1, \ldots, e_k)$ is defined as $A_1 \leq A_2$ iff

1. $P^a_{A_1} = P^a_{A_2}$ for all a in $[0, +\infty[$; or
2. $P^a_{A_1} \subset P^a_{A_2}$ for some a in $[0, +\infty[$, and $P^b_{A_1} = P^b_{A_2}$ for all $b < a$

where $\subset$ is an inclusion measure on interval-valued fuzzy sets, such as those proposed in [13]. As an example, consider $(rinc\ \ dance, win, boss)$ with $dance$ and win as in Section 3.4 and $boss$ the rule

$$boss : (sit(p, s) \rightarrow nearSpeechStand(s)) \triangleleft nearBoss(p)$$

Furthermore, we have answer sets A and A' with the following applicability and suitability values

app	r_1	r_2	r_3
A	0.25	0.5	0.5
A'	0.75	0.25	1

$suit$	r_1	r_2	r_3
A	0.5	0.5	0.25
A'	1	0.1	0.5

By definition of $rinc$ we first need to check whether we can discriminate between A and A' on the basis of P^0_A and $P^0_{A'}$. From the definition we easily obtain that:

P^0	r_1	r_2	r_3
A	$[0.25, 1]$	$[0.5, 1]$	$[0.5, 1]$
A'	$[0, 1]$	$[0.25, 1]$	$[0, 0]$

Note that $P^0_{A'}(1) = [0, 1]$ as the penalty of A' is 0 over the whole interval. Likewise $P^0_{A'}(3) = [0, 0]$ as the rule is fully applicable and penalty 0 is thus never reached. We use the inclusion measure $A \subseteq B \equiv \forall x \in X \cdot A(x) \leq B(x)$ for interval-valued fuzzy sets A and B over a universe X where $[a, b] \leq [a', b'] \equiv b \leq b' \wedge (b - a) \leq (b' - a')$ in this example, i.e. we take both the more important high λs and the amount of lambdas into account. By this definition $P^0_A(2) < P^0_{A'}(2)$ and $P^0_{A'}(3) < P^0_A(3)$ meaning that using $rinc$ the answer sets A and A' are incomparable.

For $(rcard\ e_1, \ldots, e_k)$, which uses a cardinality based approach, the ordering is given by $A_1 \leq A_2$ iff

1. $|P^a_{A_1}| = |P^a_{A_2}|$ for all a in $[0, +\infty[$; or
2. $|P^a_{A_1}| < |P^a_{A_2}|$ for some a in $[0, +\infty[$, and $|P^b_{A_1}| = |P^b_{A_2}|$ for all $b < a$

where $|U|$ denotes a measure of cardinality for the interval-valued fuzzy set U, such as those defined in [14]. Note that $rcard$ induces a partial order, whereas $rinc$ induces a total order.

As an example, consider $(rcard\ dance, win, boss)$, A and A' defined as before and a cardinality measure of the form $|A| = \sum_{x \in X} f(A(x))$ on an interval-valued fuzzy set A. When interval-valued membership degrees are used to encode uncertainty or bipolarity, $f_1([a, b]) = \frac{a+b}{2}$ is a natural choice. However, in our case, we need a notion of cardinality which is increasing in both b and $b - a$: a given penalty value is more representative for a penalty expression when it corresponds to more λ values, and to higher λ values. A suitable function is therefore given by $f_2([a, b]) = b(b - a)$. To further justify this choice, we provide the outcomes of both f_1 and f_2 in this example. For $rcard$ we obtain with f_1 that $|P^0_A| = \frac{4.25}{2} > \frac{2.25}{2} = |P^0_{A'}|$ whereas with f_2 we get $|P^0_A| = 1.75 = |P^0_{A'}|$, hence with f_1 answer set A would be preferred over A', whereas with f_2 no difference can be made by looking at penalty 0 and thus we must look for a higher penalty that does make a difference. The first penalty we should consider is 0.5 as this is the first penalty actually occurring for some rules. Computing $P^{0.5}$ gives:

$P^{0.5}$	r_1	r_2	r_3
A	$[0, 0.25]$	$[0, 0.5]$	$[0, 0]$
A'	$[0, 0]$	$[0, 0]$	$[0, 1]$

From this table we see that with f_2 we get $|P^0_A| = 0.25 * 0.25 + 0.5 * 0.5 = 0.3125$ and $|P^0_{A'}| = 1$, hence $A \leq A'$ and thus A' is preferred over A.

4 Translating Literal Preferences to a gFASP Program

In this section, we show that the gFASP framework we summarized in Section 2 is sufficiently general to encode $FPDL$–

ISBN: 978-989-95079-6-8

rules, an observation which immediately leads to an implementation method. Consider the gFASP program P, comprised of a rule base $\mathcal{R}_P$ and an aggregator expression $\mathcal{A}_P$ over the quasi-order $\mathcal{Q}_P$, and the expression e from $FPDL$. Specifically, we will show that a set of rules $\mathcal{R}_{P'}$ can be found, and an aggregator expression $\mathcal{A}_{P'}$ over a quasi-order $\mathcal{Q}_{P'}$, such that for any two answer sets A_1 and A_2

$$[\mathcal{A}_{P'}]_{A_1} \leq [\mathcal{A}_{P'}]_{A_2} \equiv [\mathcal{A}_P]_{A_1} \leq [\mathcal{A}_P]_{A_2} \wedge A_1 \leq_e A_2$$

Intuitively, $\mathcal{A}_P$ encodes which rules are more important, and how answer sets should be penalized when some rules are violated, while e encodes preferences between the actual solutions, i.e. preferences on the truth values of the literals.

As a first step, we need to provide the means to refer to the value of a certain function over literals $f(a_1, \ldots, a_n)$ from the aggregator expression. This is done by creating a constraint of the form $c_f : 0 \leftarrow \sim f(a_1, \ldots, a_n)$ where $\sim$ is an involutive negator and $\leftarrow$ is an implicator satisfying $0 \leftarrow x = \sim x$. It is easy to see that, in this way, the interpretation of rule proposition c_f, which we can refer to in the aggregator expression, will always be equal to the corresponding interpretation of $f(a_1, \ldots, a_n)$. Since the aggregator expression is increasing in all rule propositions, this procedure only works if we prefer answer sets with higher values of $f(a_1, \ldots, a_n)$. If, on the other hand, lower values of $f(a_1, \ldots, a_n)$ are preferred, we introduce the constraint $c_f : 0 \leftarrow f(a_1, \ldots, a_n)$ and use $\sim c_f$ in the aggregator.

Let $\mathcal{R}_e = \{r_1, \ldots, r_n\}$ be the set of preference rules occurring in e. Given that the preference of answer sets increases with increasing suitability of the rules and decreasing applicability, the aforementioned procedure leads to the rulebase $\mathcal{R}_{P'}$, defined by

$$\mathcal{R}_{P'} = \mathcal{R}_P \cup \{g_i : 0 \leftarrow \sim r_{ih} \mid i \in 1..n\} \cup \{f_i : 0 \leftarrow r_{ib}\}$$

The second part of our embedding consists of constructing a suitable aggregator expression $\mathcal{A}_{P'}$ and corresponding quasi-order $\mathcal{Q}_{P'}$. The aggregator $\mathcal{A}_{P'}$ is straightforwardly obtained from $\mathcal{A}_P$ by extending it with the various suitability and applicability values, i.e.:

$$\mathcal{A}_{P'} = (\mathcal{A}_P, (g_1, \ldots, g_n, \sim f_1, \ldots, \sim f_n))$$

The corresponding quasi-order is given by $\mathcal{Q}_{P'} = (\mathcal{Q}_P \times \mathcal{L}^{2n}, \leq_{\mathcal{Q}_{P'}})$ where

$$(a, b) \leq_{\mathcal{Q}_{P'}} (a', b') \equiv (a \leq_{\mathcal{Q}_P} a') \wedge (b \leq_e b')$$

Note how we can discriminate between answer sets on the basis of their rule fulfillment, using the first element of tuples from $\mathcal{Q}_{P'}$, as well as on the basis of the fulfillment of preference expression e, using the second element. The ordering $\leq_e$ is defined in the same way as the orderings on preference expressions, taking as arguments tuples with applicability and suitability scores, instead of answer sets.

The following proposition shows that the proposed embedding indeed preserves the intended semantics:

Proposition 1 *Let P be a gFASP program and e a fuzzy preference expression from FPDL. For the program P', obtained in the way described above, it holds for any two answer sets A_1 and A_2 that*

$$[\mathcal{A}_{P'}]_{A_1} \leq [\mathcal{A}_{P'}]_{A_2} \equiv [\mathcal{A}_P]_{A_1} \leq [\mathcal{A}_P]_{A_2} \wedge A_1 \leq_e A_2$$

5 Concluding remarks

In this paper, we generalized Brewka's answer set optimization framework to the fuzzy domain. The resulting formalism combines the flexibility of fuzzy answer set programming with the expressiveness of conditional preferences. We emphasized the conceptual difference between applicability and suitability of preference rules. This difference, which collapses in the two-valued case, played a crucial role in our fuzzification, leading to a generalization of penalty values using the notion of gradual numbers.

Although our extension is independent of any particular approach to fuzzy answer set programming, we showed that the recently introduced gFASP framework can provide a convenient implementation. One of the core elements in this approach is the use of an aggregator expression to discriminate between answer sets. We have shown how fuzzy preference expressions can be seen as particular choices of this aggregator expression. Hence, fuzzy preference expressions should be regarded as a convenient way to encode complex quasi-orders between answer sets, in a truely declarative way, rather than as a fundamental extension to this earlier work.

References

[1] C. Baral. *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press, 2003.

[2] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *Proc. of ICLP1988*, pages 1081–1086. MIT Press, 1988.

[3] T. Lukasiewicz and U. Straccia. Tightly integrated fuzzy description logic programs under the answer set semantics for the semantic web. In *Proc. of RR2007*, pages 289–298. Springer-Verlag, 2007.

[4] D. Van Nieuwenborgh, M. De Cock, and D. Vermeir. An introduction to fuzzy answer set programming. *Annals of Mathematics and Artificial Intelligence*, 50(3-4):363–388, 2007.

[5] T. Lukasiewicz. Probabilistic logic programming. In *Proc. of ECAI1998*, pages 388–392. J. Wiley and Sons, 1998.

[6] U. Straccia. Annotated answer set programming. In *Proc. of IPMU2006*, 2006.

[7] C.V. Damásio, J. Medina, and M. Ojeda-Aciego. Sorted multi-adjoint logic programs: termination results and applications. In *Proc. of JELIA2004*, pages 260–273. Springer-Verlag, 2004.

[8] C. V. Damásio and L. M. Pereira. Antitonic logic programs. In *Proc. of LPNMR2001*, pages 379–392. Springer-Verlag, 2001.

[9] U. Straccia. Managing uncertainty and vagueness in description logics, logic programs and description logic programs. In *Reasoning Web, 4th International Summer School, Tutorial Lectures*, number 5224, pages 54–103. Springer Verlag, 2008.

[10] G. Brewka. Complex preferences for answer set optimization. In *Proc. of KR2004*, pages 213–223, 2004.

[11] J. Janssen, S. Schockaert, D. Vermeir, and M. De Cock. General fuzzy answer set programs. Accepted for WILF2009.

[12] J. Fortin, D. Dubois, and H. Fargier. Gradual numbers and their application to fuzzy interval analysis. *Fuzzy Systems, IEEE Transactions on*, 16(2):388–402, 2008.

[13] C. Cornelis and E.E. Kerre. Inclusion measures in intuitionistic fuzzy set theory. In *Proc. of ECSQARU2003*, pages 345–356, 2003.

[14] G. Deschrijver and P. Král'. On the cardinalities of interval-valued fuzzy sets. *Fuzzy Sets and Systems*, 158(15):1728–1750, 2007.

ISBN: 978-989-95079-6-8

Two Results About Optimization of Fuzzy Variable Functions.

Andrés David Báez-Sánchez
Ph.D. Student
Department of Applied Mathematics
Institute of Mathematics, Statistics and Scientific Computation, Campinas State University
Campinas, SP. , Brazil
Email: davidbaez@ime.unicamp.br

Abstract— *We discuss some optimization problems for fuzzy variable functions and show two interesting results. First result is related to conditions for existence of global optimal solutions for some general class of optimization problems on $\mathbb{E}^n$, the space of compacts, upper semicontinuous and normal fuzzy sets of $\mathbb{R}^n$. The second result is about preservation of local optimality points, via Zadeh's principle of extension.*

Keywords— Coercitivity of fuzzy variable functions, compacity in fuzzy metric spaces, fuzzy optimization, optimization of fuzzy variable functions, principle of extension.

1 Introduction

In most optimization problems, the coefficients, parameters and variable decisions involved are assumed to be precise elements of some real space $\mathbb{R}^n$. However in many real situations there is not such certainty about those values and is necessary to consider some models or methods to deal with optimization problems under uncertainty. Two main approaches have been deeply developed when we consider uncertainty in the parameters, in the constraints or in the values of the objective function, these are stochastic programming and fuzzy optimization. We may refer [1] or [2] for stochastic programming and [3], [4] or [5] for an overview of fuzzy optimization.

By the other hand, in connection with applications and modelling using fuzzy theory, a different kind of fuzzy optimization problems has been appeared in which some of the decision variables are itself fuzzy, this means, optimization of fuzzy variable functions. Some situations like this can be found in problems of depend or constraints programming as is shown in [6] or [7]), in estimation and regression models for fuzzy random variables as in [8], [9], [10], [11] or [12], or in problems of optimal control for fuzzy differential equations as in [13] or [14].

In most of the cases those fuzzy variable optimization problem was faced successfully using some *ad hoc* techniques or properties, some heuristic algorithm, or considering just specific classes of fuzzy numbers such as triangular or L-R fuzzy numbers. We believe that fuzzy variable optimization problems have not been thoroughly studied yet, and they deserve a systematic study in itself, because a general approach to this issue can leave a fertile field for their application and theoretical developments.

In this work we are dealing with fuzzy quantities in $\mathbb{R}^n$, specifically we are considering $\mathbb{E}^n$ the spaces of compacts, upper semicontinuous and normal fuzzy sets of $\mathbb{R}^n$. It is well known that $\mathbb{E}^n$ with usual arithmetic between fuzzy numbers is not a vector space, so many of general results in optimization theory are not availables immediately on this context, precisely because most of them use the vector space structure. Taking advantage of the special characteristics and properties of $\mathbb{E}^n$ we will be able to show two interesting theoretical results in optimization of fuzzy variable functions.

In section 2 we set the two problems under consideration and we recall some basic notation and definitions. In section 3 we recall some results about compacity in $\mathbb{E}^n$. Section 4 presents our first main result related to conditions for existence of optimal global solutions for some general class of optimization problems on $\mathbb{E}^n$. In section 5 we recall the Zadeh's principle of extension and some related properties and in section 6 we present our second main result related to preservation of local optimality points, via Zadeh's principle of extension.

2 Preliminaries

We are going to consider first the following optimization problem:

$$\min f(u) \qquad \text{subject to} \qquad u \in D \tag{1}$$

with $D \subset \mathbb{E}^n$ and $f : \mathbb{E}^n \to \mathbb{R}$. Later we will consider a given function $f : \mathbb{R}^n \to \mathbb{R}$ and study the problem of optimal local points of the function $\hat{f} : \mathbb{E}^n \to \mathbb{E}$ defined by Zadeh's principle of extension from f.

Besides of the space $\mathbb{E}^n$ we will also use the metric space $K(\mathbb{R}^n)$ of compacts subsets of $\mathbb{R}^n$ with the Hausdorff metric h defined by $h(A,B) = \max\{d(A,B), d(B,A)\}$ where A, B are compacts subsets of $\mathbb{R}^n$, $d(A,B) = \sup_{a\in A} d(a,B)$ and $d(a,B) = \inf_{b\in B} d(a,b)$.

Let u be an element of $\mathbb{E}^n$. For every $\alpha \in (0,1]$ the α-level of u is defined by $u_\alpha = \{x \in \mathbb{R}^n : u(x) \geq \alpha\}$ and $u_0 = supp(u) = \overline{\{x \in \mathbb{R}^n : u(x) > 0\}}$.

ISBN: 978-989-95079-6-8

The supremum metric d_∞ on $\mathbb{E}^n$ is defined in terms of Hausdorff metric between α-level sets:

$$d_\infty(u,v) = \sup_{\alpha\in[0,1]} h(u_\alpha, v_\alpha),$$

where $u, v \in \mathbb{E}^n$.

3 Compacity in $\mathbb{E}^n$

Definition 3.1. *A set $U \subset \mathbb{E}^n$ is said to be **compact-supported** if there is a compact subset K of $\mathbb{R}^n$ such that $supp(u) \subset K$ for every $u \in U$.*

If a subset U of $\mathbb{E}^n$ is compact-supported then is bounded. We explain this in the next theorem to be used later.

Theorem 3.2. *Let U be a subset of $\mathbb{E}^n$. If U is not compact-supported then exist a sequence $\{u^k\} \subset U$ such that $||u^k|| := d_\infty(u^k, 0) \to \infty$ as $k \to \infty$.*

Proof: If U is not compact-supported then for every compact set $K \subset \mathbb{R}^n$ there exist some $u^K \in U$ such that $supp(u^K) \not\subset K$. If we take $K = \overline{B_k(0)}$ the closed ball with center 0 an radius k, then for each $k \in \mathbb{N}$ there is a $u^k \in U$ such that $supp(u^k) \not\subset \overline{B_k(0)}$.

We have that $d(0, \overline{B_k(0)}) = 0$ and $d(\overline{B_k(0)}, 0) = k$ so $h(0, \overline{B_k(0)}) = k$, and by the other hand we have that at least one element x of $supp(u^k)$ is not an element of $\overline{B_k(0)}$ so $d(x, 0) > k$ and this implies that $d(supp(u^k), 0) = d((u^k)_0, 0) > k$ and furthermore $h((u^k)_0, 0) > k$.

By definition of supremum metric we have that

$$d_\infty(u^k, 0) = \sup_{\alpha\in[0,1]} h((u^k)_\alpha, 0) > k,$$

so, $||u^k|| := d_\infty(u^k, 0) \to \infty$ as $k \to \infty$ □

We consider also the level-set mapping $u_{(\cdot)} : [0,1] \to K(\mathbb{R}^n)$ which associate with each α the level-set u_α.

Definition 3.3. *A set $D \subset \mathbb{E}^n$ is called **equicontinuous** if the family of level-sets functions $\{u_{(\cdot)} : u \in D\}$ is a family of equicontinuous functions of [0,1] in $K(\mathbb{R}^n)$.*

The following is a well known result of metric spaces writen in terms of our fuzzy optimization problem:

Theorem 3.4. *Let D be a compact subset of $\mathbb{E}^n$ and f a continuous function on D. Then exists an optimal global solution for problem 1.*

An also well known corollary it follows, but first we need an additional notation. For $c \in \mathbb{R}$ and $f : D \to \mathbb{R}$ the set $L_{f,D}(c)$ is defined by

$$L_{f,D}(c) := \{x \in D : f(x) \leq c\}$$

Corollary 3.5. *If f is continuous and exist $c \in \mathbb{R}$ such that $L_{f,D}(c)$ is not empty and compact then exists an optimal global solution for problem 1.*

Proof: If $L_{f,D}(c)$ is not empty and compact then previous theorem implies that the function f reaches a minimum at some $x^* \in L_{f,D}(c)$ when f is restricted to $L_{f,D}(c)$. By definition if $x \notin L_{f,D}(c)$ then $f(x) > c$. But $f(x^*) \leq c < f(x)$ so x^* is an optimal global solution of problem 1 □

So far what we have just write are old compacity results. Those results can be easy to understand but maybe hard to apply because until now we have not talked about the meaning of compacity on $(\mathbb{E}^n, d_\infty)$.

An interesting characterizations of compacity in $(\mathbb{E}^n, d_\infty)$ in terms of concepts previously defined is the following.

Theorem 3.6. *A subset D of $(\mathbb{E}^n, d_\infty)$ is relatively compact if D is compact-supported and equicontinuous.*

Proof: See Corollary 10 in [15] □

Corollary 3.7. *If a subset D of $(\mathbb{E}^n, d_\infty)$ is closed, compact-supported and equicontinuous then D is compact.*

4 A Coercitivity Condition

In this section we will prove our first main result. We need to recall some definitions.

Definition 4.1. *A sequence $\{x^k\} \subset \mathbb{E}^n$ is called critic in relation to a set D if $\{x^k\} \subset D$ and $||x^k|| := d_\infty(0, x^k) \to \infty$ or $x^k \to x$ with $x \in \overline{D}$ but $x \notin D$.*

Definition 4.2. *A function f is coercive in a set D if for all sequence $\{x^k\}$ critics in relation to D, $\limsup_{k\to\infty} f(x^k) = +\infty$.*

Previous ones are well known definitions on normed vector spaces but even if $\mathbb{E}^n$ is not exactly a vector space, this concepts can be used in this context because of the special characteristics of $\mathbb{E}^n$. Let's see how this can be done.

Theorem 4.3. *If f is continuous and coercive in D and exist $c \in \mathbb{R}$ such that $L_{f,D}$ is not empty and equicontinuous, then problem 1 has an optimal global solution.*

Proof: We will show that under this hypothesis $L_{f,D}(c)$ is compact and then we use corollary 3.5.

ISBN: 978-989-95079-6-8

If $L_{f,D}(c)$ is not compact then because corollary 3.7 and hypothesis, $L_{f,D}(c)$ is not closed or is not compact-supported.

If $L_{f,D}(c)$ is not closed, as we assume that $L_{f,D}(c)$ is not empty, then exist a sequence $\{u^k\} \subset L_{f,D}(c)$ such that $u^k \to u$ and $x \notin L_{f,D}(c)$. If $u \in D$ then continuity of f in D and the fact that $f(u^k) \leq c$ for all $k = 0, 1, ...$, allow us to conclude $f(u) \leq c$ but $u \notin L_{f,D}(c)$ by definition of u, so $u \notin D$.

This means that $\{u^k\}$ is a critic sequence in relation to D and by coercitivity of f in D we have that $f(u^k) \to \infty$, which is a contradiction because $f(u^k) \leq c$.

Now, if $L_{f,D}(c)$ is not compact-supported then for theorem 3.2 there is a sequence $\{u^k\} \subset L_{f,D}(c)$ such that $||u^k|| := d_\infty(u^k, 0) \to \infty$ and again because of coercitivity of f we conclude that $f(u^k) \to \infty$ which is again a contradiction because $f(u^k) \leq c$.

So $L_{f,D}(c)$ has to be compact and the desired result it follows from corollary 3.5 □

5 Zadeh's Principle of Extension and Ordering in $\mathbb{E}$

We recall now for the case of $f : \mathbb{R}^n \to \mathbb{R}$ one of the cornerstones of fuzzy theory.

Zadeh's Principle of Extension: Given a function $f : \mathbb{R}^n \to \mathbb{R}$ we define the Zadeh's extension of f by:

$$\hat{f} : \mathbb{E}^n \to \mathbb{E},$$

$$\hat{f}(u)(x) = \begin{cases} \sup_{z \in f^{-1}(x)} u(z), & \text{if } f^{-1}(x) \neq \emptyset, \\ 0, & \text{if } f^{-1}(x) = \emptyset, \end{cases}$$

for each $u \in \mathbb{E}^n$.

Next theorem is now a well known result and establishes conditions on f in order to $\hat{f}$ be a well-defined and continuous function.

Theorem 5.1. *If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous then $\hat{f} : (\mathbb{E}^n, d_\infty) \to (\mathbb{E}, d_\infty)$ is a well-defined continuous function and*

$$[\hat{f}(u)]_\alpha = f(u_\alpha),$$

for all $u \in \mathbb{E}^n$ and $\alpha \in [0, 1]$.

Proof: Proof of this result is an easy adaptation of theorems 2.1 and 3.3 at [16] and a famous result of Nguyen at [17] □

A usual partial order on $\mathbb{E}^n$ is defined by α-cuts inclusion in the following way:

$$u \leq v \Leftrightarrow u_\alpha \subseteq v_\alpha \quad \text{for all } \alpha \in [0, 1]],$$

with $u, v \in \mathbb{E}^n$.

Nevertheless, even that this order has been used succesfully for theoretical developments as in [18], it is still a little unsatisfactory at least in the case $n = 1$ because in this case this order is not a generalization of usual order on $\mathbb{R}$. In fact, number 2 is not longer greater than 1, they are not even comparable using this order.

There is, however, a generalization of natural order on $\mathbb{R}$ that we are going to consider now. Before, is necessary to note that if $u, v \in \mathbb{E}$ this implies that all their α-cuts are bounded closed interval of real line, this means that for all $\alpha \in [0, 1]$

$$u_\alpha = [a_\alpha, b_\alpha], v_\alpha = [c_\alpha, d_\alpha]$$

with $a_\alpha, b_\alpha, c_\alpha, d_\alpha \in \mathbb{R}$.

Definition 5.2. *Given $u, v \in \mathbb{E}$ and $\alpha \in [0, 1]$ we consider $u_\alpha = [a_\alpha, b_\alpha]$ and $v_\alpha = [c_\alpha, d_\alpha]$. We define a partial order relation on $\mathbb{E}$ by:*

$$u \leq_1 v \Leftrightarrow \quad a_\alpha \leq c_\alpha \quad \textit{and} \quad b_\alpha \leq d_\alpha$$

for all $\alpha \in [0, 1]$.

It is easy to show that previous order is in fact a partial order on $\mathbb{E}$ and that it is indeed a generalization of natural order on $\mathbb{R}$.

6 Local Optimality Preservation

Now we are going to prove our second main result. We are still considering the order defined in definition 5.2 and the supremum metric on $\mathbb{E}^n$

Theorem 6.1. *If $f : \mathbb{R}^n \to R$ is continuous and has a local minimum (maximum) at x^* then $\hat{f} : (\mathbb{E}^n, d_\infty) \to (\mathbb{E}, d_\infty, \leq_1)$ has also a local minimum (maximum) at x^*.*

Proof: We will just make the proof for local minimum. The other part is identical.

If f has a local minimum at x^* then exist a closed ball $B_r(x^*) \subset \mathbb{R}^n$ such that $f(x^*) \leq f(x)$ for all $x \in B_r(x^*)$.

Because $\hat{f}$ is just an extension of f to $\mathbb{E}^n$ and the order $\leq_1$ is a generalization or usual order of $\mathbb{R}$ to $\mathbb{E}$ then obviously $\hat{f}(x^*) \leq_1 \hat{f}(x)$ for all $x \in B_r(x^*)$ so we need to find a neighborhood of x^* in $(\mathbb{E}^n, d_\infty)$ with the same property.

We consider simply the closed ball $\tilde{B}_r(x^*) \subset (\mathbb{E}^n, d_\infty)$ with the same radius and center than the previous one, this means

$$\begin{aligned} \tilde{B}_r(x^*) =& \{u \in \mathbb{E}^n : d_\infty(u, x^*) \leq r\} \\ =& \{u \in \mathbb{E}^n : \sup_{\alpha \in [0,1]} h(u_\alpha, x^*) \leq r\} \\ =& \{u \in \mathbb{E}^n : \forall \alpha \in [0, 1] \quad h(u_\alpha, x^*) \leq r\} \\ \tilde{B}_r(x^*) =& \{u \in \mathbb{E}^n : \forall \alpha \in [0, 1] \quad u_\alpha \subset B_r(x^*), \} \end{aligned} \tag{2}$$

ISBN: 978-989-95079-6-8

and we have that $B_r(x^*) \subset \tilde{B}_r(x^*)$ when elements of $B_r(x^*)$ are considered as fuzzy as well.

If $u \in \tilde{B}_r(x^*)$, we have that for all $\alpha \in [0,1]$, $u_\alpha \subset B_r(x^*)$ so $f(u_\alpha) \subset f(B_r(x^*))$ and by definition of $B_r(x^*)$ this implies that for all $\alpha \in [0,1]$, $f(x^*) \leq y$ for all $y \in f(u_\alpha)$.

By theorem 5.1 we have that $[\hat{f}(u)]_\alpha = f(u_\alpha)$ so, for all $\alpha \in [0,1]$

$$f(x^*) \leq y \quad \text{for all} \quad y \in [\hat{f}(u)]_\alpha. \tag{3}$$

If we write $[\hat{f}(u)]_\alpha$ as an interval, this means, $[\hat{f}(u)]_\alpha = [a_\alpha, b_\alpha]$ for some $a_\alpha, b_\alpha \in \mathbb{R}$ the previous expression in the specific case of a_α and b_α becomes

$$\forall \alpha \in [0,1], \quad f(x^*) \leq a_\alpha \quad \text{and} \quad f(x^*) \leq b_\alpha. \tag{4}$$

Because $f(x^*) = \hat{f}(x^*)$ and because $f(x^*)$ considered as a fuzzy set has α-level sets always equals to $[f(x^*), f(x^*)]$, the previous considerations and the expression 4 implies that for $u \in \tilde{B}_r(x^*)$

$$\hat{f}(x^*) \leq_1 \hat{f}(u),$$

so $\hat{f}$ has a local minimum at x^*. □.

7 Conclusions and Future Work

In this work has been shown that a classical result of vector space optimization about coercitivity can be extended also for optimization problems on $\mathbb{E}^n$, even if this spaces is not a vector space. We believe that many other results can be generalized in this way and we will continue to study this subject and its applications.

This kind of problem can be seen also as general metric spaces optimization problems, a subject that is lately getting attention again (See [19]), however because of the fact that $\mathbb{E}^n$ is indeed more than just a metric space (after all it has some arithmetic structure) we will try to take advantage of its particular structure to obtain better results and applications.

We also believe that there are interesting questions about the way that a fuzzy extension of a function and orderings are related to optimal points and others qualitative characteristics of a function. We will keep study this subject deeply and we hope to find interesting interpretations and future applications.

Acknowledgment

I want to thank my advisor professor Marko Rojas-Medar and my dear colleague Johanna Garzón. This work was supported by CNPq, National Council for Scientific and Technological Development - Brazil.

References

[1] J.R. Birge, F. Louveaux, *Introduction to Stochastic Programming*, Springer-Verlag, (1997).

[2] A. Prékopa, *Stochastic Programming*, Kluwer Academic Publishers, (1995).

[3] C.R. Bector, S. Chandra, Fuzzy Mathematical Programming and Fuzzy Matrix Games,*Studies in Fuzziness and Soft Computing*, Springer-Verlag, Berlin, (2005).

[4] R. Slowinski (Ed.)A. Prkopa, *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*, (The Handbooks of Fuzzy Sets) Springer (1998).

[5] Y.J. Lai, C.L. Hwang, Fuzzy Multiple Objective Decision Making: Methods and Applications, *Lecture Notes in Economics and Mathematical Systems* 404, Springer-Verlag, (1994).

[6] J. Gao, J. Zhao, X. Ji, Fuzzy Chance-Constrained Programming for Capital Budgeting Problem with Fuzzy Decisions, *L. Wang and Y. Jin (Eds.): FSKD 2005, LNAI 3613*, pp. 304-311, Springer-Verlag Berlin Heidelberg, (2005).

[7] B. Liu, Dependent-chance programming with fuzzy decisions, *IEEE Transactions on Fuzzy Systems*, 7, 354-360, (1999).

[8] P. Diamond, R. Körner, Extended Fuzzy Linear Models and Least Squares Estimates, *Computers Math. Applic.* Vol. 33, No. 9, pp. 15-32, (1997).

[9] R. Körner, W. Näther, Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates, *Journal of Information Sciences* 109 95–118, (1998).

[10] A. Wünsche, W. Näther, Least-squares fuzzy regression with fuzzy random variables, *Fuzzy Sets and Systems* 130, 43-50, (2005).

[11] A. Colubi, et. al., A determination coefficient for fuzzy random variables in a fuzzy arithmetic-based linear model. *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE, London, United Kingdom)*, pp. 473-476, (2007).

[12] G. González-Rodríguez, et. al., Estimation of a simple linear regression model for fuzzy random variables, *Fuzzy Sets and Systems* 160, 357-370, (2009).

[13] P. Diamond, P. Kloeden, Optimization under uncertaintly. *Bouchon-Meunier, B., Zadeh, L.A., Yager, R.R. (eds.) IPMU 1990. LNCS 521*, pp. 247-249. Springer, Heidelberg (1991)

ISBN: 978-989-95079-6-8

[14] Y.C. Kwun, et. al., The Existence of Fuzzy Optimal Control for the Semilinear Fuzzy Integrodifferential Equations with Nonlocal Conditions, *B.-Y. Cao (Ed.): Fuzzy Information and Engineering (ICFIE), ASC*(40) , pp. 657665, (2007).

[15] Greco, G.H., Moschen M.P.: Supremum metric and relatively compact sets of fuzzy sets. *Nonlinear Analysis* **64**, 1325–1335 (2006).

[16] H. Román-Flores, et. al., A note on Zadeh's Extension, *Fuzzy Sets And Systems* 117, 327–331, (2002).

[17] H.T. Nguyen, A note on the extension principle for fuzzy sets, *J. Math. Anal.* Appl. 64, 369–380, (1978).

[18] M.A. Rojas-Medar, et.al., Fuzzy quasilinear spaces and applications, *Fuzzy Sets and Systems* 152 2, 173–190, (2005).

[19] M. McAsey, L. Mou L, A multiplier rule on a metric spaces, *J. Math. Anal. Appl.*, 337, 1064–1071 (2008)

[20] A. Izmailov, M. Solodov, *Otimização Vol. 1* IMPA, (2005).

[21] P. Diamond, P. Kloeden *Metric Spaces of Fuzzy Sets: Theory and Applications* World Scientific Publishing Co., Inc., River Edge, (1994).

ISBN: 978-989-95079-6-8

Industrial Monitoring by Evolving Fuzzy Systems

Lara Aires[1], João Araújo[2], António Dourado[1]

[1] Center for Informatics and Systems, Department of Informatics Engineering,
University of Coimbra, Polo II 3030-290 Coimbra Portugal
[2] Galp Sines Refinery, Process Control,
Sines, Portugal
Email: lccaires@gmail.com, dourado@dei.uc.pt, joao.araujo@galpenergia.com

Abstract *— Industrial monitoring of complex processes with hundreds or thousands of variables is a hard task faced in this work through evolving fuzzy systems. The Visbreaker process of the Sines Oil Refinery is the case studied. Firstly dimension reduction is performed by multidimensional scaling, obtaining the process evolution in a three dimensional space. Then an evolving fuzzy system (eFS) is developed to detect eventual malfunction of sensors. This eFS takes the data in the three reduced dimensions as antecedents and classifies the process state into normal and abnormal states. A software platform- the eFSLab (Evolving Fuzzy Systems Laboratory) - , with which this work has been developed, is presented and discussed. Several strategies for rule creation and evolution of rules, for Takagi-Sugeno (and for Mamdani system obtained from these) , are implemented in eFSLab. The obtained eFS shows a promising performance in the case studied, classifying in some simulations the state of the process into abnormal-normal condition in about 95% of the cases, with a number of rules between 5 and 8.*

Keywords — Evolving fuzzy systems; multidimensional scaling; industrial monitoring; fuzzy logic software.

1 Introduction

The growing complexity of industrial processes, the market competition, the more strict quality and safety rules, enhance the needs for sophisticated maintenance functions, including online process monitoring, failure detection and failure diagnosis [1] [11] [12]. This turns to be a complex problem mainly because of its non-linear high dimensional characteristics, since industrial systems are described using a large number of variables, which can be determined by various parameters and measurements.

According to [2] the main problem is to find the clusters of states in order to reflect the measurements and determine the general behavior of the system. Fuzzy systems may play here an important role, if iterative, real time techniques for rule base creation and evolution can be found. The eFSLab, Evolving Fuzzy Systems Laboratory, is a computational framework with that aim, briefly presented in this work.

The Visbreaker process of the Sines Refinery (Galp) is used to obtain real data for testing the performance of eFSLab to classify the state of the process into abnormal-normal condition. This classification is done through fuzzy systems produced by an evolving algorithm which is implemented in the platform.

An analysis is made of the performance of the fuzzy systems as classifier for this problem concluding that this algorithm has the potential to support maintenance activities as a failure detector.

The paper is organized as follows. Section 2 reviews briefly the eFS algorithm and in section 3 the eFSLab environment will be described. In section 4 the results obtained with eFSLab for data from Sines Refinery are presented and in section 5 these results are discussed and some conclusions and directions for future work are proposed.

2 The eFS algorithm

The algorithm to create Takagi-Sugeno fuzzy models is based on the approach proposed by Angelov et al. for online learning [3] and improved by [4] [5]. This algorithm allows creating zero or first orders Takagi-Sugeno fuzzy models with global or local recursive estimation parameters.

Fuzzy models are created with a data-driven approach, composed of an iterative process with 7 main stages [3] .

The procedure begins with the initialization of the rule-base structure (antecedent part of the rules) with a single rule, based on the first data sample. The second stage is to read the next data sample. Then, the potential of each new data sample is recursively calculated and in stage 4 occurs a recursively update of the potentials of the focal points (centers) of the existing rules/clusters.

After this, the potential of the new data sample is compared with the updated potential of existing centers, in order to decide if the rule-base is modified or upgraded. In the next stage consequence's parameters are recursively updated by Recursive Least Squares (RLS) or weighted RLS for globally or locally optimal parameters, respectively. Finally, the last stage is to predict the output of the system. By reading the next data sample at the next time step, the procedure returns to stage 2.

The algorithm intends to approach the subtractive clustering technique applied in a batch onset. Since it is to be applied in real-time environments, the data appears successively and at each time the new available data is used to improve the fuzzy system, in such a way that the new informative content of data should be reflected in the fuzzy rules, whether by change some of them, whether by creating new rules if there is sufficiently novelty in the process behavior. There are here some scientific challenges. The main one is how to

ISBN: 978-989-95079-6-8

measure the informative novelty in the new data. The second is how to change/create the rules consequently. The proposed techniques by [3] and [4] are based basically on heuristics and experimentation and as a consequence are case dependent. In the present work the same approach is followed. More conditions for rule update/create have been investigated and implemented in the computational framework eFSLab.

eFSLab allows to develop a Mamdani fuzzy model from a zero order Takagi-Sugeno (TS) fuzzy model. The Mamdani system is made based on the same antecedents of TS system previously created and by centering the fuzzy sets of Mamdani consequents in the Takagi-Sugeno constant consequent (only zero order models can be considered).

3 The eFSLab interface

The eFSLab interface is presented in Fig. 1(at the end of the paper). It was intended to create an interface as complete as possible, where the user could set the great majority of parameters needed to create a Takagi-Sugeno fuzzy system and transform it into a Mamdani system (in case of zero order TS systems), in a simple way.
The first step to use eFSLab interface is to select the file to import data. This file can be an Excel or a text file.
Then the user can choose TS order (zero or first), the recursive parameter estimations method, and set some other parameters needed to the recursive process.
Conditions for creation or modification of rules can also be chosen or defined by the user. There is a set of predefined conditions or the user can build up his (her) own conditions, as shown in Fig. 2

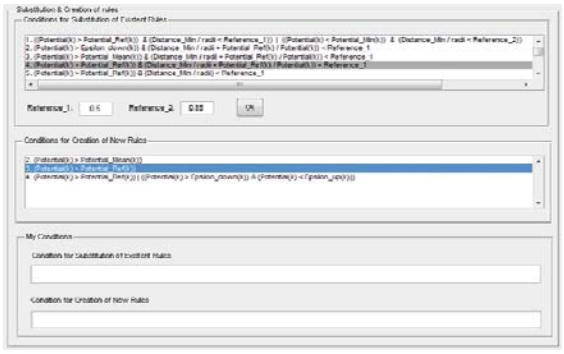
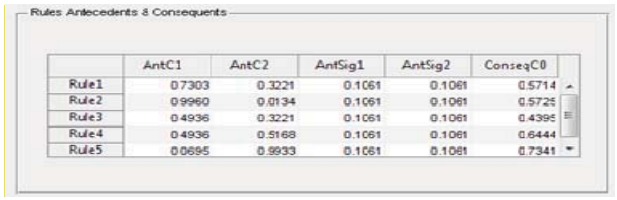


Figure 2. Creating or modifying rules in eFSLab.

The fuzzy system is created and its characteristics can be analysed using the Matlab Fuzzy Logic Toolbox, with all its functionalities (viewing the graphics, consulting antecedents and consequents fuzzy sets, etc). A diagnosis text file is created by eFSLab. The created rules are shown in a table in the interface, shown in Figure 3 for the case of two antecedents, Gaussian membership functions and five rules.



Rules Antecedents & Consequents

	AntC1	AntC2	AntSig1	AntSig2	ConseqC0
Rule1	0.7303	0.3221	0.1061	0.1061	0.5714
Rule2	0.9960	0.0134	0.1061	0.1061	0.5725
Rule3	0.4936	0.3221	0.1061	0.1061	0.4395
Rule4	0.4936	0.5168	0.1061	0.1061	0.6444
Rule5	0.0695	0.9933	0.1061	0.1061	0.7341

Figure 3. Table containing antecedents and consequents for each rule created by eFS procedure.

eFSLab produces a set of graphical information about the evolution of the TS systems, such as the number of rules, the instants of modification or creation of rules, the potential evolution of the data samples, the model output versus real output for validation data, the norm of covariance matrix and rules parameters evolution, the obtained cluster in input space. Some of them are shown in figures below.
The algorithm is based on recursive subtractive clustering [3]. At the end of the learning phase, the collection of points that has been used, together with the obtained clusters, can be observed as in Fig. 4.

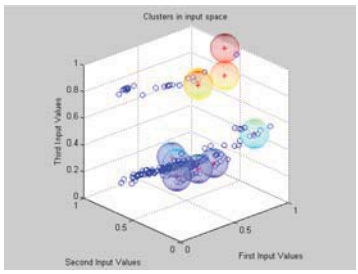


Figure 4. Clusters in input space after training. The dimensions of the spheres depend on the parameters of the clusters in a subtractive clustering environment.

In an on-line (iterative) implementation, the new rules can be created or the existent rules can be modified, and the instant when that happens can be shown as in Fig. 5

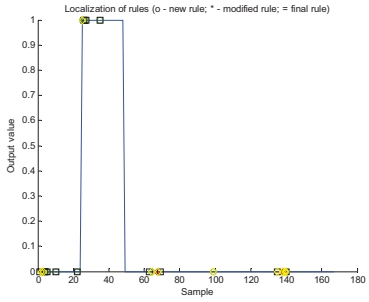


Figure 5. Localization of rules (new rules, rules modified and rules deleted).

The evolution of the number of rules during the process can be visualized as in Fig. 6.
To transform a zero-order TS fuzzy system into a Mamdani one, the *SugenoToMamdani* interface is opened as Fig. 7.
The user can choose the type of membership functions, the correspondent parameters and can preview their shape.

ISBN: 978-989-95079-6-8

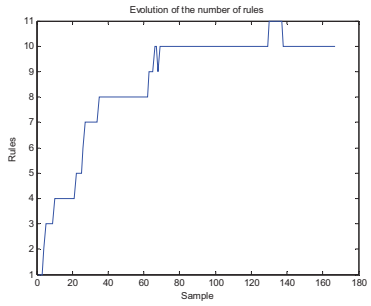


Figure 6. Evolution of the number of rules.

The Mamdani fuzzy system produced can be opened in the Matlab Fuzzy Logic Toolbox, benefiting from its functionalities for a deeper analysis .

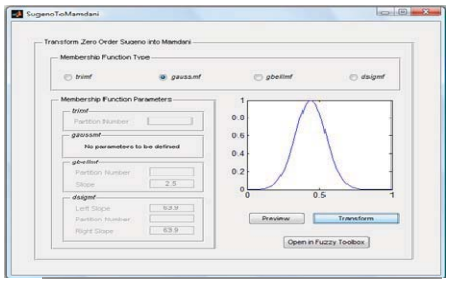


Figure 7.The *SugenoToMamdani* interface.

4 Applicatiom to the Visbreaker process

The Visbreaker Unit is intended to reduce the viscosity of the residual coming from the vacuum column. With this objective a thermal cracking process is used with a relatively low temperature, and a long residence time. As a result of the thermal process a low viscosity Visbreaker residual is obtained, as well as lighter products, such as hydrocarbonets (gas oil diesel, gasoline and gases).The great economical advantage of the Visbreaker process is in fact that it produces a residual with a lower viscosity that the load feed. By this way, it is possible to use a lower quantity of "cutterstocks" (some of them of high benefit) for the production of fuel oil.
Fig. 8 [9] presents its flow sheet. It is composed of several sub-processes, and its main part is a kiln operating at about 310 ºC.

Data averaged for each hour, between the 1st of January and the 5th of June of 2008 was collected. Actually, data from 160 tags is available. From these, after correlation analysis and process expertise, 59 were selected as sufficiently representative of the process. Multidimensional scaling is then applied to those 59 dimensions to obtain a three dimensional representation by using VisRed [6]. The three dimensional data is then fed to eFSLab that finds a zero-order TS fuzzy model.

Multidimensional scaling reduces an *n*-dimensional data set to *p*-dimensional data set, with $p<<n$, such that the distance (for example Euclidian) between each pair of points in the reduced space is similar to the distance in the high dimensional space. This means that the topology of the original data set is preserved, as well as the information embedded in the topology. The technique consists in the minimization of the Euclidian distance between two matrices: the dissimilarity matrix in the original space and the dissimilarity matrix in the reduced space. The (*i,j*) element of a dissimilarity matrix is the distance between points *i* and *j*.

Three simulated scenarios have been considered with increasing time horizon and number of faults: one week with one fault in one sensor (Case 1), two weeks with two faults in two sensors (Case 2), and one month with two sensors faults during one day (Case 3). In Case 1 a 20% reduction in data from one temperature sensor in the kiln, during one day, was introduced into the data between the 1st and the 6th of June 2008. In Case 2 a 20% reduction in data from two temperature sensors in the kiln, during one day, was introduced into the data between the 22nd of May and the 4th of June 2008. In Case 3, the process data between the 5th of May and the 5th of June 2008 was changed in the same way in the same two sensors. These scenarios allow to measure the diagnosis capability of the eFS, as a classifier, as the size of data set increase.

For each test, there are some parameters of eFSLab that had to be set in order to produce the fuzzy system. The first parameter to be set is the order of the TS fuzzy model (zero) and the parameter estimation (Global estimation) for all the three cases. Other parameters controlling the rule base evolutions were fixed at their default values in eFSLab.
Radii value (controlling the region of influence of each cluster in input space) changed from test to test in order to evaluate its influence on the process. The conditions for creation and modification of rules were chosen to have the capability to produce rules in regions with less data points, improving the classification performance.
Some results of the three cases are presented in the following. In Case 1 and Case 3was introduced perturbation both in test and train data sets. In Case 2 the perturbation was only introduced in test data set

Case 1. Data is from the 1st to the 6th of June 2008 (one week). In train data set it was added a reduction of 20% to data from the 2nd of June and in test data set the same perturbation was added to data from the 4th of June, in a temperature sensor.
Dimension reduction by multidimensional scaling was performed with cosine dissimilarity metric without normalization. Fig. 9 represents data after dimension reduction. Two different groups corresponding to normal and abnormal data can be observed.
The reduced data is then presented to eFSLab, obtaining a zero order TS system with 8 rules to classify the data into two classes 1 and 2. A Radii value of 0.3 was used. These rules can be visualized in Fig. 10 (using Matlab Fuzzy Logic Toolbox).
In Fig. 11 shows the real output and the fuzzy system output It can be seen that the fault was detected. This classifier shows here 95% of successful classification, although there

ISBN: 978-989-95079-6-8

are also some false faults detected, reducing the performance of the system.

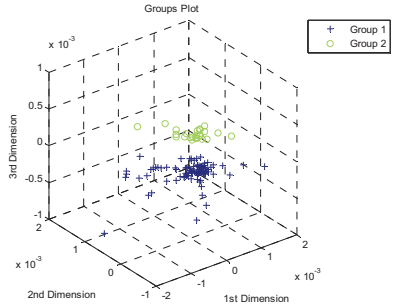


Figure 9. The three-dimensional state space for Case 1. Blue points (Group 1) are normal operating point. Group 2 are points of a simulated sensor fault.

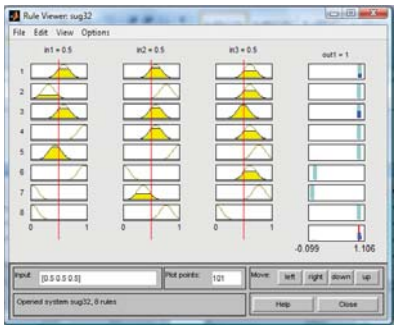


Figure 10. The 8 rules Takagi-Sugeno eFS obtained to classify the points in Case 1.

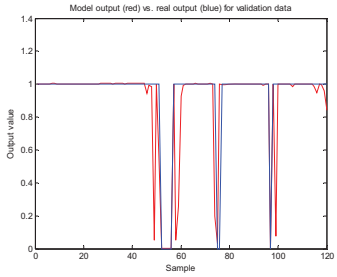


Figure 11. Real output (blue) and simulation output (red) obtained with zero-order TS system created for Case 1.

Case 2. Data from 22nd of May to 4th of June 2008 (two weeks) is considered, with a perturbation of 20% in data from two sensors in 23 of May and 4 of June.
Dimension reduction was made with Euclidean dissimilarity metric without normalization. Fig. 12 is a three dimensional representation of the reduced data.
Giving to eFSLab the reduced data, with Radii of 0.1, a 10 rules zero order TS system was obtained, shown in Fig. 13.

Case 3. Now all the data from 5th of May to the 5th of June (one month) is considered. Perturbation is the same as in the Case 2, but in 20th of May and 4th of June. Also the Radii value and the dissimilarity metric for dimension reduction are the same as in Case 2. In Fig. 15 is represented data set after dimension reduction.
Fig.14 compares the real output with the fuzzy system output. The classification is successful in 98.8% of the points.
The obtained zero-order TS system has now 11 rules, presented in Fig. 16.

Fig. 17 compares the output of the fuzzy system with the prepared data. The performance of the created fuzzy system is now substantially decreased: although the fault was detected, so have been many false positives. Overall 80.80% of the points have been correctly classified.

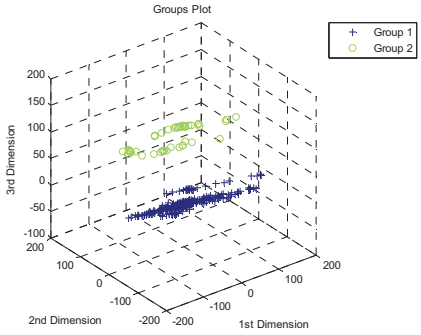


Figure 12: The three-dimensional state space representation of the Visbreaker process in the last week of May and the first week of June 2008. Blue points are normal operating point. Group 2 are points of a simulated sensor malfunction situation.

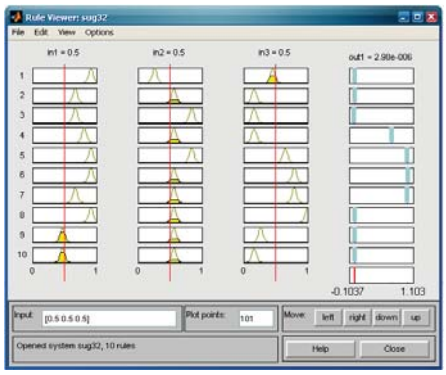


Figure 13: The 10 rules Takagi-Sugeno eFS obtained to classify the points in Case 2.

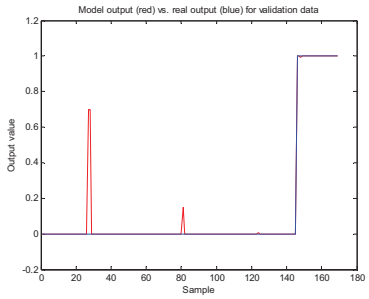


Figure 14. Real output (blue) and simulation output (red) obtained with fuzzy system created for Case 2.

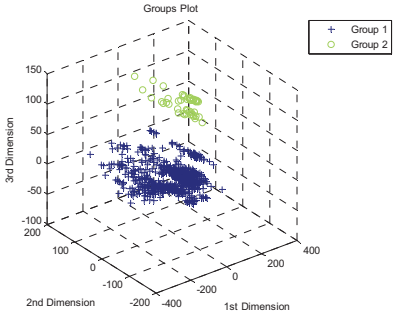


Figure 15: The three-dimensional state space representation of the Visbreaker process during the 5th of May and the 5th of June 2008. Blue points are normal operating point. Group 2 are points of a simulated sensor malfunction situation.

ISBN: 978-989-95079-6-8

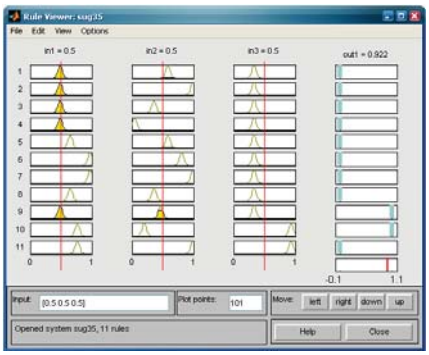


Figure 16: The 11 rules Takagi-Sugeno eFS for Case 3.

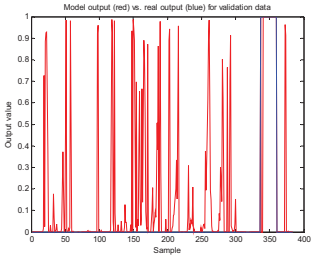


Figure 17. Case 3 :real output (blue) and fuzzy system output in Case 3.

Table 1 presents a summary of the results.

Table 1: Summary of results for the three cases.

Test	Data Set Size	MD Scaling Dissimilarity Metric	Radii	Output Matching (%)
1	1 week	Cosine	0.3	95%
2	2 weeks	Euclidean	0.1	98.8%
3	1 month	Euclidean	0.1	80.80%

When the time horizon increases (the data sets size increases proportionally), the performance of the fuzzy system as classifier seems to be worse. The way how the rules are created probably needs improvements. This is also supported by the distribution of the final clusters in the input space (see for example Fig. 4), that does not cover all the relevant space. However the results show a promising performance of the fuzzy system in terms of output (classification) matching. Furthermore, the low number of rules produced in each test contributes to a lower fuzzy sets superposition, increasing eventually interpretability and transparency, which is a main concern in fuzzy modelling [7] [8] [9]. Further work is needed to improve the iterative clustering technique, looking for better approximations to the subtractive clustering method.

5 Conclusions

Development of fuzzy systems from data is still an unattained aim of the fuzzy logic communities. Accuracy, interpretability and transparency are characteristics with an important practical relevance, allowing to give sense and credibility to the fuzzy systems. In the case of industrial monitoring, in big complexes of basic process industry, they may play an important role supporting the decision making in daily operation of the plants. *eFSLab* can be a useful computational framework to this goal. It is built with friendly user interfaces and allows an easy configuration and application to any data set built in proper form.

Evolving fuzzy systems, where rules are based on clusters, need more elaborated techniques for iterative clustering methods. This is not an easy task, since here neighbour points do not appear at the same time. Some kind of manageable memory must be developed to approach the iterative results to the non-iterative ones. This is one research direction of the authors.

New ways to measure novelty in data is needed. In a real time set the size of datasets tends to infinity. It is possible to maintain in memory only a finite time window of data Deleted points are forgotten history and new points may simply repeat pas behaviour but leading to new rules that are not only unnecessary but also prejudicial. This is another challenge to future work. The fuzzy community can have here a rich field of research to make fuzzy systems a credible and useful tool for data mining and knowledge discovery, whether in batch, whether in real-time.

Acknowledgments. This work was supported by Project CLASSE POSC/EIA/58162/2004 including EU FEDER Support.

References

[1] N. Palluat *et al*. A neuro-fuzzy monitoring system application to flexible production systems. *Computers in Industry*, 57:528-538, 2006.

[2] O. Simula *et al*, Monitoring and Modeling of Complex Processes using Hierarchical Self-Organization Maps. Helsinki University of Technology, Laboratory of Computer and Information Science. Espoo, Finland. 1996.

[3] P. Angelov and D. Filev. An Approach to Online Identification of Takagi-Sugeno Fuzzy Models. *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, 34(1), 2004.

[4] J. Victor and A. Dourado, "Evolving Takagi-Sugeno fuzzy models," Technical Report, CISUC, (2003, September). Available: http://cisuc.dei.uc.pt/acg/view_pub.php?id_p=760.

[5] J. V. Ramos and A. Dourado, "On line interpretability by rule base simplification and reduction," Eunite Symposium, Aachen, 2004.

[6] Dourado, A. and Ferreira, E. and Barbeiro, P. , "VISRED - Numerical Data Mining with Linear and Nonlinear Techniques", in Advances in Data Mining. Theoretical Aspects and Applications,, pp. 93-106, LNAI, Vol. 4597, Springer, Industrial Conference on Data Mining 2007, Leipzig, Germany. Available for download from

[7] Mikut, R., J. Jäkel, L. Gröll "Interpretability issues in data-based learning of fuzzy systems", Fuzzy Sets and Systems , vol. 150, 179–197, 2005.

[8] C. Mencar and A.M. Fanelli. Interpretability constraints for fuzzy information granulation. *Information Sciences*, 178:4585–4618, 2008.

[9] González, J., I.Rojas, H. Pomares, L. J. Herrera, A. Guillén, J. M. Palomares, F. Rojas, "Improving the accuracy while preserving the interpretability of fuzzy function approximators by means of multi-objective evolutionary algorithms", International Journal of Approximate Reasoning ,Volume 44 , Is.1, , pp32-44, 2007.

[10] Galp (2006), Sines Refinery Flow Sheets.

[11] Filev D. P. and F. Tseng, Novelty Detection Based Machine Health Prognostics, Proc. of the 2006 International Symposium on Evolving Fuzzy Systems, Lake District, UK, pp. 193--199, 2006

[12] Lughofer, E. and C. Guardiola, On-Line Fault Detection with Data-Driven Evolving Fuzzy Models. Control and Intelligent Systems, Vol. 36 (4), pp. 307-317, 2008.

ISBN: 978-989-95079-6-8

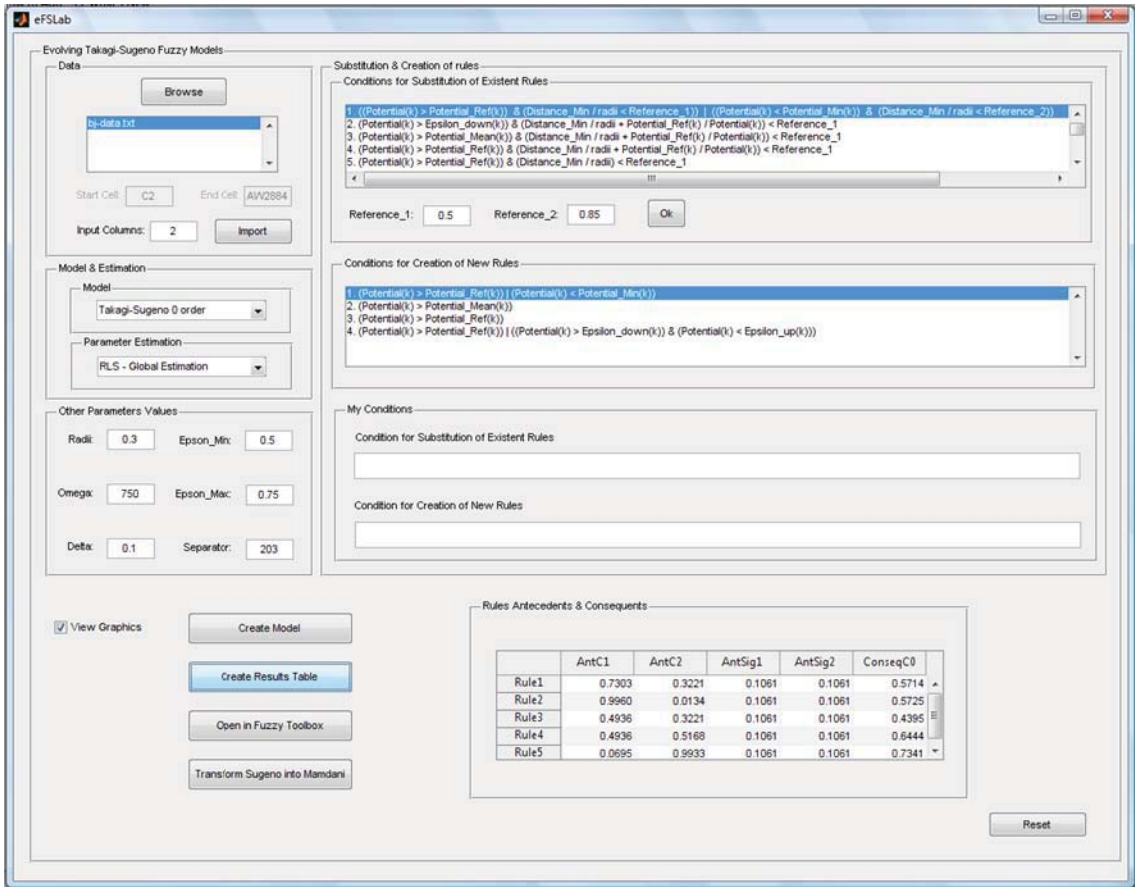


Figure 1: *eFSLab* interface.

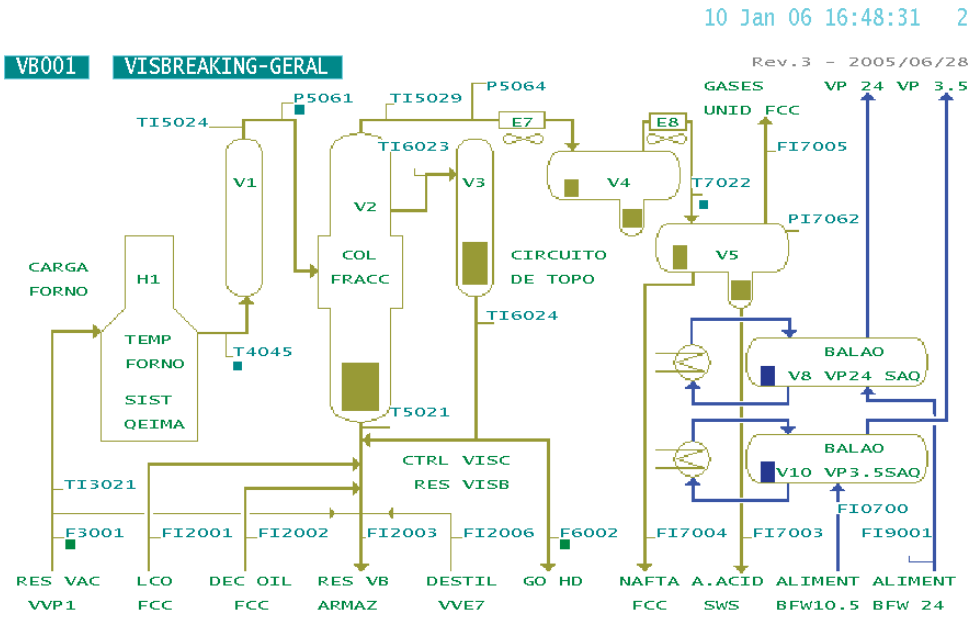


Figure 8: The Visbreaker Process [10].

ISBN: 978-989-95079-6-8

Probabilistic Decision Making with the OWA Operator and its Application in Investment Management

José M. Merigó

Department of Business Administration, University of Barcelona
Av. Diagonal 690, 08034 Barcelona, Spain
Email: jmerigo@ub.edu

Abstract—*We develop a new model for decision making under risk environment and under uncertainty. We introduce a new aggregation operator that unifies the probabilities and the ordered weighted averaging (OWA) operator in the same formulation. We call it the probabilistic ordered weighted averaging (POWA) operator. This aggregation operator provides a more complete representation of the decision problem because it is able to consider probabilistic information and the attitudinal character of the decision maker. We study different properties and families of the POWA operator. We also develop an illustrative example of the new approach in a decision making problem about selection of investments.*

Keywords— Decision making; OWA operator; Probabilities; Investment selection.

1 Introduction

Decision making problems [2-8,14,16,18] are very common in the literature. There are different ways and methods for solving the decision process. Usually, the method used for solving the decision problem depends on the available information. For example, when the decision maker has probabilistic information he will solve the problem calculating the expected value for each alternative. Thus, he will be in a problem of decision making under risk environment. However, in other problems the decision maker may not have probabilistic information. Therefore, he must use another approach for solving the problem such as the use of the ordered weighted averaging (OWA) operator [13] that reflects the attitudinal character (degree of optimism) of the decision maker. In this case, we are in a situation of decision making under uncertainty.

The OWA operator is a very useful technique for aggregating the information. It provides a parameterized family of aggregation operators between the maximum and the minimum. In decision making it is very useful for representing the attitudinal character of the decision maker. Since its appearance, the OWA operator has been studied and applied in a wide range of problems [1-4,6-7,9-18].

In [4], they presented a new model that tried to unify the concept of probability with the OWA operator. They introduced the immediate probabilities. This method formulates an aggregation operator that uses probabilities and OWA operators at the same time. Therefore, this approach permits to consider decision making problems under risk environment and under uncertainty at the same time. The concept of immediate probabilities has been further studied in [6-7,16]. It is very useful in some particular situations. However, further analysis (as those shown in the paper) show that it can only unify probabilities and OWAs giving a neutral degree of importance to each case. But the real situation is that sometimes the decision maker believes more on the probabilistic information or on the OWAs. Therefore, the real unification must consider the degree of importance of these to concepts in the aggregation process. The aim of this paper is to present a new formulation that unifies the probabilities and the OWA operators considering the degree of importance of each case in the aggregation process. We call this new approach as the probabilistic ordered weighted averaging (POWA) operator. The main advantage of this approach is that it provides more complete information of the decision process because we can unify the probabilistic information and the attitudinal character according to the available information in the decision process. We study some of its main properties and we see that the OWA operator and the usual probabilistic decision making are particular cases of this approach. We study other particular cases of the POWA operator such as the Hurwicz-POWA, the olympic-POWA and the S-POWA.

We study the applicability of the POWA operator and we see that it is incredibly broad because it can be used in most of the problems where it appears the probability. In this paper, we focus on a decision making problem about selection of investments. We see the usefulness of the POWA operator because we can use situations of risk and uncertainty at the same time.

This paper is organized as follows. In Section 2 we briefly review the OWA operator and the immediate probability. Section 3 presents the POWA operator and Section 4 different particular cases. In Section 5 we present a decision making problem with the POWA operator in investment selection and in Section 6 we give the main conclusions.

2 Preliminaries

2.1 The OWA operator

The OWA operator [13] is an aggregation operator that provides a parameterized family of aggregation operators between the minimum and the maximum. It can be defined as follows.

ISBN: 978-989-95079-6-8

Definition 1. An OWA operator of dimension n is a mapping OWA: $R^n \to R$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, then:

$$\text{OWA}(a_1, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j \quad (1)$$

where b_j is the jth largest of the a_i.
Note that different properties can be studied such as the distinction between descending and ascending orders, different measures for characterizing the weighting vector and different families of OWA operators. Note that it is commutative, monotonic, bounded and idempotent. For further reading, refer, e.g., to [1-4,6-7,9-18].

2.2 Immediate probabilities

An immediate probability is a type of probability that includes the attitudinal character of the decision maker by using the OWA operator in the aggregation process. Therefore, it is an attemp to unify the probabilities with the OWA operator. It can be defined as follows.

Definition 2. An IPOWA operator of dimension n is a mapping IPOWA: $R^n \to R$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, according to the following formula:

$$\text{IPOWA}(a_1, \ldots, a_n) = \sum_{j=1}^{n} \hat{v}_j b_j \quad (2)$$

where b_j is the jth largest of the a_i, each argument has associated a probability v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_j = (w_j v_j / \sum_{j=1}^{n} w_j v_j)$ and v_j is the probability v_i ordered according to b_j, that is, according to the jth largest of the a_i.
Note that the IPOWA operator is a good approach for unifying probabilities and OWAs in some particular situations. But it is not always useful, especially in situations where we want to give more importance to the OWA operators or to the probabilities. One way to see why this unification does not seem to be a final model is considering other ways of representing $\hat{v}_j$. For example, we could also use $\hat{v}_j = [w_j + v_j / \sum_{j=1}^{n} (w_j + v_j)]$ or other similar approaches.

3 The probabilistic OWA operator

The probabilistic ordered weighted averaging (POWA) operator is an extension of the OWA operator for situations where we find probabilistic information. It can also be seen as a unification between decision making problems under uncertainty (with OWA operators) and under risk (with probabilities). This approach seems to be complete, at least as an initial real unification between OWA operators and probabilities.

However, note that some previous models already considered the possibility of using OWA operators and probabilities in the same formulation. The main model is the concept of immediate probability explained in Section 2 [4,6-7,16]. Although it seems to be a good approach it is not so complete than the POWA because it can unify OWAs and probabilities in the same model but it can not take in consideration the degree of importance of each case in the aggregation process. Other methods that could be considered are the hybrid averaging (HA) operator and the weighted OWA (WOWA) operator. These methods are focused on the weighted average (WA) but it is easy to extend them to probabilities because sometimes the WA is used as a subjective probability. As said before, these an other approaches are useful for some particular situations but they does not seem to be so complete than the POWA because they can unify OWAs with probabilities (or with WAs) but they can not unify them giving different degrees of importance to each case. Note that in future research we will also prove that these models can be seen as a special case of a general POWA operator (or its respective model with WAs) that uses quasi-arithmetic means. Obviously, it is possible to develop more complex models of the IP-OWA, the HA and the WOWA that takes into account the degree of importance of the OWAs and the probabilities (or WAs) in the model but they seem to be artificial and not a natural unification as it will be shown below.
In the following, we are going to analyze the POWA operator. It can be defined as follows.

Definition 3. A POWA operator of dimension n is a mapping POWA: $R^n \to R$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, according to the following formula:

$$\text{POWA}(a_1, \ldots, a_n) = \sum_{j=1}^{n} \hat{v}_j b_j \quad (3)$$

where b_j is the jth largest of the a_i, each argument a_i has an associated probability v_i with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, $\hat{v}_j = \beta w_j + (1 - \beta) v_j$ with $\beta \in [0, 1]$ and v_j is the probability v_i ordered according to b_j, that is, according to the jth largest of the a_i.
Note that it is also possible to formulate the POWA operator separating the part that strictly affects the OWA operator and the part that affects the probabilities. This representation is useful to see both models in the same formulation but it does not seem to be as a unique equation that unifies both models.

Definition 4. A POWA operator is a mapping POWA: $R^n \to R$ of dimension n, if it has an associated weighting vector W, with $\sum_{j=1}^{n} w_j = 1$ and $w_j \in [0, 1]$ and a probabilistic vector V, with $\sum_{i=1}^{n} v_i = 1$ and $v_i \in [0, 1]$, such that:

ISBN: 978-989-95079-6-8

$$\text{POWA}(a_1, \ldots, a_n) = \beta \sum_{j=1}^{n} w_j b_j + (1-\beta) \sum_{i=1}^{n} v_i a_i \qquad (4)$$

where b_j is the jth largest of the arguments a_i and $\beta \in [0, 1]$.
In the following, we are going to give a simple example of how to aggregate with the POWA operator. We consider the aggregation with both definitions.

Example 1. Assume the following arguments in an aggregation process: (20, 40, 50, 30). Assume the following weighting vector W = (0.2, 0.2, 0.2, 0.4) and the following probabilistic weighting vector P = (0.4, 0.2, 0.3, 0.1). Note that the probabilistic information has a degree of importance of 60% while the weighting vector W a degree of 40%. If we want to aggregate this information by using the POWA operator, we will get the following. The aggregation can be solved either with (3) or (4). With (3) we calculate the new weighting vector as:

$$\hat{v}_1 = 0.4 \times 0.2 + 0.6 \times 0.3 = 0.26$$
$$\hat{v}_2 = 0.4 \times 0.2 + 0.6 \times 0.2 = 0.2$$
$$\hat{v}_3 = 0.4 \times 0.2 + 0.6 \times 0.1 = 0.14$$
$$\hat{v}_4 = 0.4 \times 0.4 + 0.6 \times 0.4 = 0.4$$

And then, we calculate the aggregation process as follows:

POWA = 0.26 × 50 + 0.2 × 40 + 0.14 × 30 + 0.4 × 20 = 33.2.

With (4), we aggregate as follows:

POWA = 0.4 × (0.2 × 50 + 0.2 × 40 + 0.2 × 30 + 0.4 × 20) + 0.6 × (0.4 × 20 + 0.2 × 40 + 0.3 × 50 + 0.1 × 30) = 33.2.

Obviously, we get the same results with both methods.
From a generalized perspective of the reordering step, it is possible to distinguish between the descending POWA (DPOWA) and the ascending POWA (APOWA) operator by using $w_j = w^*_{n-j+1}$, where w_j is the jth weight of the DPOWA and w^*_{n-j+1} the jth weight of the APOWA operator.
If B is a vector corresponding to the ordered arguments b_j, we shall call this the ordered argument vector and W^T is the transpose of the weighting vector, then, the POWA operator can be expressed as:

$$\text{POWA}(a_1, \ldots, a_n) = W^T B \qquad (5)$$

Note that if the weighting vector is not normalized, i.e., $W = \sum_{j=1}^{n} w_j \neq 1$, then, the POWA operator can be expressed as:

$$\text{POWA}(a_1, \ldots, a_n) = \frac{1}{W} \sum_{j=1}^{n} \hat{v}_j b_j \qquad (6)$$

The POWA is monotonic, commutative, bounded and idempotent. It is monotonic because if $a_i \geq u_i$, for all a_i, then, POWA(a_1, a_2,…, a_n) ≥ POWA(u_1, u_2…, u_n). It is commutative because any permutation of the arguments has the same evaluation. That is, POWA(a_1, a_2,…, a_n) = POWA(u_1, u_2,…, u_n), where (u_1, u_2,…, u_n) is any permutation of the arguments (a_1, a_2,…, a_n). It is bounded because the POWA aggregation is delimitated by the minimum and the maximum. That is, Min$\{a_i\}$ ≤ POWA(a_1, a_2,…, a_n) ≤ Max$\{a_i\}$. It is idempotent because if $a_i = a$, for all a_i, then, POWA(a_1, a_2,…, a_n) = a.
Another interesting issue to analyze are the measures for characterizing the weighting vector W. Following a similar methodology as it has been developed for the OWA operator [7,13] we can formulate the attitudinal character, the entropy of dispersion, the divergence of W and the balance operator. Note that these measures affect the weighting vector W but not the probabilities because they are given as some kind of objective information.

4 Families of POWA operators

First of all we are going to consider the two main cases of the POWA operator that are found by analyzing the coefficient β. Basically, if $\beta = 0$, then, we get the probabilistic approach and if $\beta = 1$, the OWA operator.
By choosing a different manifestation of the weighting vector in the POWA operator, we are able to obtain different types of aggregation operators. For example, we can obtain the probabilistic maximum, the probabilistic minimum, the probabilistic average and the probabilistic weighted average.

Remark 1. The probabilistic maximum is found when $w_1 = 1$ and $w_j = 0$ for all $j \neq 1$. The probabilistic minimum is formed when $w_n = 1$ and $w_j = 0$ for all $j \neq n$.

Remark 2. More generally, the step-POWA is formed when $w_k = 1$ and $w_j = 0$ for all $j \neq k$. Note that if $k = 1$, the step-POWA is transformed to the probabilistic maximum, and if $k = n$, the step-POWA becomes the probabilistic minimum operator.

Remark 3. The probabilistic average is obtained when $w_j = 1/n$ for all j, and the probabilistic weighted average is obtained when the ordered position of i is the same as the ordered position of j.

Remark 4. For the median-POWA, if n is odd we assign $w_{(n+1)/2} = 1$ and $w_{j^*} = 0$ for all others. If n is even we assign for example, $w_{n/2} = w_{(n/2)+1} = 0.5$ and $w_{j^*} = 0$ for all others.

Remark 5. For the weighted median-POWA, we select the argument b_k that has the kth largest argument such that the sum of the weights from 1 to k is equal or higher than 0.5 and the sum of the weights from 1 to $k - 1$ is less than 0.5.

Remark 6. The olympic-POWA is generated when $w_1 = w_n = 0$, and for all others $w_{j^*} = 1/(n - 2)$. Note that it is possible

ISBN: 978-989-95079-6-8

to develop a general form of the olympic-POWA by considering that $w_j = 0$ for $j = 1, 2, \ldots, k, n, n-1, \ldots, n-k+1$, and for all others $w_{j^*} = 1/(n-2k)$, where $k < n/2$. Note that if $k = 1$, then this general form becomes the usual olympic-POWA. If $k = (n-1)/2$, then this general form becomes the median-POWA aggregation. That is, if n is odd, we assign $w_{(n+1)/2} = 1$, and $w_{j^*} = 0$ for all other values. If n is even, we assign, for example, $w_{n/2} = w_{(n/2)+1} = 0.5$ and $w_{j^*} = 0$ for all other values.

Remark 7. Note that it is also possible to develop the contrary case, that is, the general olympic-POWA operator. In this case, $w_j = (1/2k)$ for $j = 1, 2, \ldots, k, n, n-1, \ldots, n-k+1$, and $w_j = 0$, for all other values, where $k < n/2$. Note that if $k = 1$, then we obtain the contrary case for the median-POWA.

Remark 8. Another interesting family is the S-POWA operator. It can be subdivided into three classes: the "or-like," the "and-like" and the generalized S-POWA operators. The generalized S-POWA operator is obtained if $w_1 = (1/n)(1-(\alpha+\beta)) + \alpha$, $w_n = (1/n)(1-(\alpha+\beta)) + \beta$, and $w_j = (1/n)(1-(\alpha+\beta))$ for $j = 2$ to $n-1$, where $\alpha, \beta \in [0, 1]$ and $\alpha + \beta \leq 1$. Note that if $\alpha = 0$, the generalized S-POWA operator becomes the "and-like" S-POWA operator, and if $\beta = 0$, it becomes the "or-like" S-POWA operator.

Remark 9. Another family of aggregation operator that could be used is the centered-POWA operator. We can define a POWA operator as a centered aggregation operator if it is symmetric, strongly decaying and inclusive. Note that these properties have to be accomplished for the weighting vector W of the OWA operator but not necessarily for the weighting vector P of the probabilities. It is symmetric if $w_j = w_{j+n-1}$. It is strongly decaying when $i < j \leq (n+1)/2$ then $w_i < w_j$ and when $i > j \geq (n+1)/2$ then $w_i < w_j$. It is inclusive if $w_j > 0$. Note that it is possible to consider a softening of the second condition by using $w_i \leq w_j$ instead of $w_i < w_j$. We shall refer to this as softly decaying centered-POWA operator. Another particular situation of the centered-POWA operator appears if we remove the third condition. We shall refer to it as a non-inclusive centered-POWA operator.

Remark 10. Another type of aggregation that could be used is the E-Z POWA weights. In this case, we should distinguish between two classes. In the first class, we assign $w_{j^*} = (1/q)$ for $j^* = 1$ to q and $w_{j^*} = 0$ for $j^* > q$, and in the second class, we assign $w_{j^*} = 0$ for $j^* = 1$ to $n-q$ and $w_{j^*} = (1/q)$ for $j^* = n-q+1$ to n.

Remark 11. A further interesting type is the non-monotonic-POWA operator. It is obtained when at least one of the weights w_j is lower than 0 and $\sum_{j=1}^{n} w_j = 1$. Note that a key aspect of this operator is that it does not always achieve monotonicity. Therefore, strictly speaking, this particular case is not a POWA operator. However, we can see it as a particular family of operators that is not monotonic but nevertheless resembles a POWA operator.

Remark 12. Note that other families of POWA operators could be used following the recent literature about different methods for obtaining OWA weights. Some of these methods are explained in [1,7,9,15,18].

5 Decision making with the POWA operator in investment selection

The POWA operator is applicable in a wide range of situations where it is possible to use probabilistic information and OWA operators. Therefore, we see that the applicability is incredibly broad because all the previous models, theories, etc., that uses the probability can be extended by using the POWA operator. The reason is that all the problems with probabilities deal with uncertainty. Usually, in most of the problems it is assumed a neutral attitudinal character against the probabilistic information but we are still under uncertainty. Thus, sometimes we may prefer to be more or less optimistic against the probabilistic information. Note that this problem can be proved by looking to utility theory. In this theory, people prefer more safety results independently that the expected results (obtained with probabilities) give better results to a more risky alternative. For example, people prefer 100.000 euros with probability 1 instead of 90% probability of obtaining 1.000.000 euros and 10% probability of obtaining 0 euros. If we calculate the expected result (with probabilities) we get that the first alternative gives 100.000 euros and the second one 900.000 euros. However, most of the people will prefer the first alternative because there is no risk involved. This problem has been solved by the economists by using utility theory but as we can see, the use of the POWA operator is also useful for solving this type of problems.

Summarizing some of the main fields where it is possible to apply the POWA operator, we can mention:

- Statistics (especially in probability theory).
- Mathematics
- Economics
- Decision theory
- Engineering
- Physics
- Etc.

In this paper, we focus on an application in decision making about selection of investments. The main reason for using the POWA operator is that we are able to assess the decision making problem considering probabilities and the attitudinal character of the decision maker. Thus, we get a more complete representation of the decision problem.

In the following, we present a numerical example of the new approach in investment selection. We analyze a company that operates in Europe and North America that wants to invest some money in a new market. They consider five alternatives.

ISBN: 978-989-95079-6-8

- A_1 = Invest in the Asian market.
- A_2 = Invest in the South American market.
- A_3 = Invest in the African market.
- A_4 = Invest in all three markets.
- A_5 = Do not invest money in any market.

In order to evaluate these investments, the investor has brought together a group of experts. This group considers that the key factor is the economic situation of the world economy for the next period. They consider 5 possible states of nature that could happen in the future: S_1 = Very bad economic situation, S_2 = Bad economic situation, S_3 = Regular economic situation, S_4 = Good economic situation, S_5 = Very good economic situation.
The results of the available investments, depending on the state of nature S_i and the alternative A_k that the decision maker chooses, are shown in Table 1.

Table 1: Available investment alternatives

	S_1	S_2	S_3	S_4	S_5
A_1	20	40	60	70	80
A_2	50	70	90	40	20
A_3	30	40	50	80	60
A_4	40	40	50	60	70
A_5	80	60	50	40	30

In this problem, the experts assume the following weighting vector: W = (0.3, 0.2, 0.2, 0.2, 0.1). They assume that the probability that each state of nature will happen is: P = (0.1, 0.3, 0.3, 0.2, 0.1). Note that the OWA operator has an importance of 40% and the probabilistic information an importance of 60%. With this information, we can calculate how the attitudinal character of the decision maker may affect the probabilistic information. For doing so, we will use (3) to calculate the "attitudinal probabilities". The results are shown in Table 2.

Table 2: Attitudinal probabilities

	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$	$\hat{v}_5$
A_1	0.18	0.2	0.26	0.26	0.1
A_2	0.3	0.26	0.14	0.2	0.1
A_3	0.24	0.14	0.26	0.26	0.1
A_4	0.18	0.2	0.26	0.26	0.1
A_5	0.18	0.26	0.26	0.2	0.1

With this information, we can aggregate the expected results for each state of nature in order to make a decision. In Table 3, we present different results obtained by using different types of IOWAAC operators.

Table 3: Aggregated results

	Prob.	OWA	PAM	POWA
A_1	54	60	54	56.4
A_2	63	61	59.4	62.2
A_3	52	57	52	54
A_4	50	55	50.8	52
A_5	52	57	52	54

Note that we can also obtain these results by using (4). Then, we will calculate separately the OWA and the probabilistic approach as shown in Table 4.

Table 4: First aggregation process

	Prob.	AM	OWA
A_1	54	54	60
A_2	63	54	61
A_3	52	52	57
A_4	50	52	55
A_5	52	52	57

After that, we will aggregate both models in the same process considering that the OWA model has a degree of importance of 40% and the probabilistic information 60% as shown in Table 5.

Table 5: Final aggregated results

	Prob.	OWA	PAM	POWA
A_1	54	60	54	56.4
A_2	63	61	59.4	62.2
A_3	52	57	52	54
A_4	50	55	50.8	52
A_5	52	57	52	54

Obviously, we get the same results with both methods. If we establish an ordering of the alternatives, a typical situation if we want to consider more than one alternative, then, we get the results shown in Table 6. Note that the first alternative in each ordering is the optimal choice.

Table 6: Ordering of the investments

	Ordering
Probabilistic	$A_2 \succ A_1 \succ A_3 = A_5 \succ A_4$
OWA	$A_2 \succ A_1 \succ A_3 = A_5 \succ A_4$
PAM	$A_2 \succ A_1 \succ A_3 = A_5 \succ A_4$
POWA	$A_2 \succ A_1 \succ A_3 = A_5 \succ A_4$

As we can see, depending on the aggregation operator used, the ordering of the investments may be different. Therefore, the decision about which investment select may be also different. Note that in this example, we get the same results for all the cases but in other problems or other types of aggregation operators we may find different results and decisions.

6 Conclusions

We have presented an approach for decision making with probabilities and OWA operators. We have developed a new aggregation operator that unifies the probabilities with the OWA operators. We have called it the POWA operator. The main advantage is that it gives a more complete representation of the decision problem because we are taking into account the probabilistic information and the attitudinal character of the decision maker. We have studied this new formulation and we have seen that it includes decision making problems under risk environment and under uncertainty as particular cases. We have studied some of its main properties and different families of POWA operators.

ISBN: 978-989-95079-6-8

We have also studied the applicability of the POWA operator and we have seen that there are a lot of potential applications that can be developed because in almost all the studies where it appears the probabilities, it is possible to extend the analysis by using the POWA operator. The reason is that the probabilistic information is always affected by the uncertainty. Therefore, it is always possible to consider the attitudinal character of the decision maker by using the OWA operator. We have focused on a decision making problem about selection of investments. We have seen the usefulness of using the POWA operator because we are able to consider probabilities and OWAs at the same time.

In future research, we expect to develop further extensions to this approach by adding new characteristics in the problem such as the use of order inducing variables, uncertain information (interval numbers, fuzzy numbers, linguistic variables, etc.), generalized and quasi-arithmetic means and distance measures. We will also extend this approach to situations where we use the WA instead of probabilities and further developments that have been initially developed in [7]. We will also consider different applications giving special attention to business decision making problems such as strategic and product management.

Acknowledgment

This paper is partly supported by the Spanish *Ministerio de Asuntos Exteriores y de Cooperación*, *Agencia Española de Cooperación Internacional para el Desarrollo* (AECID) (project A/016239/08).

References

[1] G. Beliakov, A. Pradera and T. Calvo, *Aggregation Functions: A Guide for Practitioners*. Berlin: Springer-Verlag, 2007.

[2] H. Bustince, F. Herrera and J. Montero, *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*. Berlin: Springer-Verlag, 2008.

[3] T. Calvo, G. Mayor and R. Mesiar, *Aggregation Operators: New Trends and Applications*. New York: Physica-Verlag, 2002.

[4] K.J. Engemann, D.P. Filev and R.R. Yager, Modelling decision making using immediate probabilities. International Journal of General Systems, 24:281-294, 1996.

[5] J. Gil-Aluja, *The interactive management of human resources in uncertainty*. Dordrecht: Kluwer Academic Publishers, 1998.

[6] J.M. Merigó, Using immediate probabilities in fuzzy decision making. In: *Proceedings of the 22nd ASEPELT Conference*, Barcelona, pp. 1650-1664, 2008.

[7] J.M. Merigó, *New extensions to the OWA operators and its application in decision making* (In Spanish). PhD Thesis, Department of Business Administration, University of Barcelona, 2008.

[8] J.M. Merigó and A.M. Gil-Lafuente, Unification point in methods for the selection of financial products. *Fuzzy Economic Review*, 12:35-50, 2007.

[9] J.M. Merigó and A.M. Gil-Lafuente, The induced generalized OWA operator. *Information Sciences*, 179:729-741, 2009.

[10] V. Torra, The weighted OWA operator. *International Journal of Intelligent Systems*, 12:153-166, 1997.

[11] V. Torra and Y. Narukawa, *Modeling Decisions: Information Fusion and Aggregation Operators*. Berlin: Springer-Verlag, 2007.

[12] Z.S. Xu and Q.L. Da, An overview of operators for aggregating information. *International Journal of Intelligent Systems*, 18:953-968, 2003

[13] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics B*, 18:183-190, 1988.

[14] R.R. Yager, Decision making under Dempster-Shafer uncertainties. *International Journal of General Systems*, 20:233-245, 1992.

[15] R.R. Yager, Families of OWA operators. *Fuzzy Sets and Systems*, 59:125-148, 1993.

[16] R.R. Yager, Including decision attitude in probabilistic decision making. *International Journal of Approximate Reasoning*, 21:1-21, 1999.

[17] R.R. Yager and D.P. Filev, Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man and Cybernetics B*, 29:141-150, 1999.

[18] R.R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators: Theory and Applications*. Norwell: Kluwer Academic Publishers, 1997.

ISBN: 978-989-95079-6-8

Some results on Lipschitz quasi-arithmetic means

Gleb Beliakov[1], Tomasa Calvo[2], Simon James[3]

[1,3]School of Information Technology, Deakin University
221 Burwood Hwy, Burwood, 3125, Australia.
[2]Departamento de Ciencias de la Computación, Universidad de Alcalá
Crta. Barcelona Km 33,6. 28871-Alcalá de Henares (Madrid), Spain.
Email: {[1]gleb, [3]sgj}@deakin.edu.au, tomasa.calvo@uah.es

Abstract— *We present in this paper some properties of k-Lipschitz quasi-arithmetic means. The Lipschitz aggregation operations are stable with respect to input inaccuracies, what is a very important property for applications. Moreover, we provide sufficient conditions to determine when a quasi–arithemetic mean holds the k-Lipschitz property and allow us to calculate the Lipschitz constant k.*

Keywords— k-Lipschitz aggregation functions, quasi-arithmetic means, stability, triangular norms.

1 Introduction

Aggregation of several input values into a single output value is an indispensable tool in many disciplines and applications such as decision making [1], pattern recognition, expert and decision support systems, information retrieval, etc [2]. There is a wide range of aggregation functions which provide flexibility to the modeling process, including different types of aggregation functions. There are several recent books that provide details of many aggregation methods [3, 4, 5, 6, 7].

For applications it is important to design aggregation functions that are stable with respect to small perturbations of inputs (e.g., due to input inaccuracies). Such aggregation functions need not only be continuous, but Lipschitz continuous [8]. Kernel and 1-Lipschitz aggregation functions have been studied in [9, 10, 11]. It is known, for instance, that 1-Lipschitz triangular norms are copulas [12, 3, 7]. More recently, k-Lipschitz t-norms and t-conorms were studied [13, 14, 15]. k-Lipschitz t-norms do not increase the perturbation of inputs by more than a factor of k, which is suitable for many applications.

There are many other generated functions constructed similarly to the Archimedean triangular norms with the help of additive generators. In this article we examine quasi-arithmetic means and establish conditions under which these functions are Lipschitz or not Lipschitz.

Firstly, in Section 2 we recall some basic notions to develop the rest of the work. Section 3 contains the main results involving quasi–arithmetic means. At the end we provide some conclusions.

2 Preliminaries and related works

We restrict ourselves to aggregation functions defined on $[0,1]^n$.

Definition 1 *A function $f : [0,1]^n \to [0,1]$ is called an aggregation function if it is monotone non-decreasing in each variable and satisfies $f(0,\ldots,0) = 0$, $f(1,\ldots,1) = 1$.*

Now, we will pay attention to a special class of aggregation function, -the class of weighted quasi–arithmetic means-, for this we need to consider a continuous strictly monotone function $g : [0,1] \to [-\infty,\infty]$, which we call a *generating function* or generator. Of course, g is invertible, but it is not necessarily a bijection (i.e., its range may be $Ran(g) \subset [-\infty,\infty]$). Other two examples of generated functions are Archimedean t-norms and t-conorms. Further there exists a class of uninorms, known as *representable uninorms* or *generated uninorms*, that can also be built by means of additive generators. Further, a vector $\vec{w} = (w_1,\ldots,w_n)$ is called a weighting vector if $w_i \in [0,1]$ and $\sum_{i=1}^{n} w_i = 1$.

Definition 2 *For a given generating function g, and a weighting vector $\vec{w}$, the weighted quasi-arithmetic mean is the function*

$$M_{\vec{w},g}(\vec{x}) = g^{-1}\left(\sum_{i=1}^{n} w_i g(x_i)\right). \qquad (1)$$

From this definition, we have the following particular quasi–arithmetic means:

Arithmetic mean	$M(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$
Geometric mean	$G(\vec{x}) = (\prod_{i=1}^{n} x_i)^{\frac{1}{n}}$
Harmonic mean	$H(\vec{x}) = n(\sum_{i=1}^{n} \frac{1}{x_i})^{-1}$
Power mean	$M_r(\vec{x}) = \frac{1}{n}(\sum_{i=1}^{n} x_i^r)^{\frac{1}{r}}$, if $r \neq 0$ and $M_0(\vec{x}) = G(\vec{x})$.

Another class of aggregation operators is the following

Definition 3 *Let $g : [0,1] \to [-\infty,\infty]$ be a continuous strictly monotone function and let $\vec{w}$ be a weighting vector. The function*

$$GenOWA_{\vec{w},g}(\vec{x}) = g^{-1}\left(\sum_{i=1}^{n} w_i g(x_{(i)})\right) \qquad (2)$$

is called a generalized OWA (also known as ordered weighted quasi-arithmetic mean [4]). As for OWA, $x_{(i)}$ denotes the i-th largest value of $\vec{x}$.

Another aggregation operator that include the previous one for the case of a symmetric fuzzy measure is the generalized discrete Choquet integral which is defined as follows

ISBN: 978-989-95079-6-8

Definition 4 *Let $g : [0,1] \to [-\infty,\infty]$ be a continuous strictly monotone function. The generalized Choquet integral with respect to a fuzzy measure v is the function*

$$C_{v,g}(\vec{x}) = g^{-1}\left(C_v(g(\vec{x}))\right),$$

where C_v is the discrete Choquet integral with respect to v and $g(\vec{x}) = (g(x_1),\ldots,g(x_n))$.

Now, we consider the crucial concept of this work

Definition 5 *An aggregation function f is called Lipschitz continuous if there is a positive number k, such that for any two vectors $\vec{x},\vec{y}$ in the domain of definition of f:*

$$|f(\vec{x}) - f(\vec{y})| \leq kd(\vec{x},\vec{y}), \tag{3}$$

where $d(\vec{x},\vec{y})$ is a distance between $\vec{x}$ and $\vec{y}$. The smallest such number k is called the Lipschitz constant of f (in the distance d). We shall call such functions k-Lipschitz.

Typically the distance is chosen as a p-norm $d(\vec{x},\vec{y}) = ||\vec{x}-\vec{y}||_p$, with $||\vec{x}||_p = \left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p}$, for finite p, and $||\vec{x}||_\infty = \max_{i=1,\ldots,n}|x_i|$. In this work we concentrate on the 1–norm.

Definition 6 *A function f is called locally Lipschitz continuous on Ω if for every $x \in \Omega$ there exists a neighbourhood $D(x)$ such that f restricted to $D(x)$ is Lipschitz.*

Of course, duality w.r.t. standard negation preserves Lipschitz property (and the Lipschitz constant). It is easy to see that if an aggregation function A is k–Lipschitz, it is also m–Lipschitz for any $m \geq k$. Also any convex combination of $k-$Lipschitz aggregation functions $f = \alpha f_1 + \beta f_2$, $\alpha+\beta = 1, \alpha,\beta \geq 0$, is $k-$Lipschitz.

The class of k–Lipschitz t–norms, whenever $k > 1$, has been already characterized (see [13]). Note that 1–Lipschitz t–norms are copulas, see, e.g., [3, 5]. A strictly decreasing continuous function $g : [0,1] \to [0,1]$ with $g(1) = 0$ is an additive generator of a 1–Lipschitz Archimedean t–norm if and only if g is convex.

The k–Lipschitz property implies continuity of the t–norm. Recall that a continuous t–norm can be represented by means of an ordinal sum of continuous Archimedean t–norms, and that a continuous Archimedean t–norm can be represented by means of a continuous additive generator [3, 16]. Characterization of all k–Lipschitz t–norms can be reduced to the problem of characterization of all Archimedean k–Lipschitz t–norms.

Definition 7 *Let $g : [0,1] \to [0,+\infty]$ be a strictly monotone function and let $k \in]0,+\infty[$ be a real constant. Then g will be called k–convex if*

$$g(x+k\varepsilon) - g(x) \leq g(y+\varepsilon) - g(y)$$

holds for all $x \in [0,1[, y \in]0,1[$, with $x \leq y$ and $\varepsilon \in]0, \min(1-y, \frac{1-x}{k})]$.

Obviously, if $k = 1$ the function g is convex.

Observe that, a k–convex monotone function is also continuous in $]0,1[$, as was earlier. A decreasing function g can be k–convex only for $k \geq 1$. Moreover, when a decreasing function g is k–convex, it is also m–convex for all $m \geq k$. In the case of a strictly increasing function $g^\star$, it can be k-convex only for $k \leq 1$. Moreover, when $g^\star$ is k–convex, it is m–convex for all $m \leq k$.

Considering $k \geq 1$ and a strictly decreasing function g, we provide the following characterization given in [13].

Proposition 1 *Let $T : [0,1]^2 \to [0,1]$ be an Archimedean t–norm and let $g : [0,1] \to [0,+\infty], g(1) = 0$ be an additive generator of T. Then T is k–Lipschitz if and only if g is k–convex.*

Another useful characterizations are the following

Corollary 1 *(Y.-H. Shyu) [17] Let $g : [0,1] \to [0,\infty]$ be an additive generator of a t–norm T which is differentiable on $]0,1[$ and let $g'(x) < 0$ for $0 < x < 1$. Then T is k–Lipschitz if and only if $g'(y) \geq kg'(x)$ whenever $0 < x < y < 1$.*

Corollary 2 *Let $T : [0,1]^2 \to [0,1]$ be an Archimedean t–norm and let $g : [0,1] \to [0,\infty]$ be an additive generator of T such that g is differentiable on $]0,1[\backslash S$, where $S \subset [0,1]$ is a discrete set. Then T is k–Lipschitz if and only if $kg'(x) \leq g'(y)$ for all $x,y \in [0,1], x \leq y$ such that $g'(x)$ and $g'(y)$ exist.*

The following useful results follow from Corollary 1, with it we can determine whether a given piecewise differentiable t–norm is k–Lipschitz.

Corollary 3 *Let $T : [0,1]^2 \to [0,1]$ be an Archimedean t–norm and let $g : [0,1] \to [0,\infty]$ be its additive generator differentiable on $]0,1[$, and $g'(t) < 0$ on $]0,1[$. If*

$$\inf_{t\in]x,1[} g'(t) \geq k \sup_{t\in]0,x[} g'(t)$$

holds for every $x \in]0,1[$ then T is k–Lipschitz.

Corollary 4 *Let $g : [0,1] \to [0,\infty]$ be a strictly decreasing function, differentiable on $]0,a[\cup]a,1[$. If g is k-convex on $[0,a[$ and on $]a,1]$, and if*

$$\inf_{t\in]a,1[} g'(t) \geq k \sup_{t\in]0,a[} g'(t),$$

then g is k–convex on $[0,1]$.

Remark 1 *Generated uninorms are not k–Lipschitz since that they are not continuous at (0,1) and (1,0) (in the binary case). Nullnorms are aggregation functions related to t–norms and t–conorms. In this case, it is clear that a nullnorm V is k–Lipschitz if and only if the underlying t-norm and t–conorm are k–Lipschitz, and the k–Lipschitz constant of V is the maximum of Lipschitz constants of the underlying t–norm and t–conorm.*

ISBN: 978-989-95079-6-8

3 Quasi-arithmetic means

Consider a univariate continuous strictly monotone function $g : [0,1] \to [-\infty, \infty]$, called generator.

For a given g, the quasi-arithmetic means are defined in Definition 2, and are denoted by M_g. We start with bivariate quasi-arithmetic means of such operations. Quasi-arithmetic means are continuous if and only if $Ran(g) \neq [-\infty, \infty]$ [18]. Moreover, its generator is not defined uniquely, i.e., if $g(t)$ is a generating function of some weighted quasi-arithmetic mean, then $ag(t) + b$, $a, b \in \Re$, $a \neq 0$ is also a generating function of the same mean provided $Ran(g) \neq [-\infty, \infty]$. For this reason, one can assume that g is monotone increasing, as otherwise we can simply take $-g$.

We shall consider two cases: I) $g(0) = 0$, $g(1) = 1$, and II) $g(0) = -\infty, g(1) = 0$. Of course, by duality we also cover the case $g(1) = \infty, g(0) = 0$, and by using appropriate linear transformations, all generators can be reduced to the mentioned cases.

Let us make some preliminary remarks on convexity. As opposed to the case of convex additive generators of t-norms, where the resulting t-norms are 1-Lipschitz, convexity of the generator g does not play any role by itself for quasi-arithmetic means. Since both g and $-g$ are generators of the same mean, and obviously when g is convex $-g$ is concave, convexity of g by itself does not lead to the Lipschitz condition. Also note that $g(x) = -\ln(x)$ is a convex generator of the geometric mean $G(x, y) = \sqrt{xy}$, which is not Lipschitz. Further, even if g is convex and increasing or convex and decreasing, this does not imply Lipschitz condition either: note that $g_d(x) = 1 - g(1 - x)$ is a generator of a quasi-arithmetic mean dual to the one generated by g, and Lipschitz condition is preserved under duality. If g is convex increasing, then g_d is convex decreasing and vice versa. Thus we will look for a different condition.

3.1 Case of finite generators

We start with the case I) of g finite. First, let us show that g must be Lipschitz on [0,1]. For short, we will denote $M_{\vec{w},g}$ by M_g.

Lemma 1 *Let g be finite locally Lipschitz and continuously differentiable except at a point $a \in [0;1]$. Then M_g is not k–Lipschitz for any k.*

Proof. Suppose that M_g is k-Lipschitz, which means it is differentiable almost everywhere in its domain (Rademacher's theorem), and we must have

$$\frac{\partial M_g}{\partial x}(x, y) \leq k, \ x, y \neq a$$

whenever such a derivative exists, and similarly for the other partial derivative. Since M_g is symmetric, only the derivative with respect to x is needed.

$$\frac{\partial M_g}{\partial x} = \frac{1}{g'\left(M_g(x, y)\right)} \cdot \frac{1}{2} g'(x) \leq k.$$

Since g is strictly increasing we must have

$\frac{1}{2} g'(x) \leq k g'(M_g(x, y))$ for all $x, y \in [0, 1]$ such that $x \neq a$ and $M_g(x, y) \neq a$ or

$$\frac{1}{2} g'(x) \leq k \cdot \inf_{y \in [0,1]} g'(M_g(x, y)). \tag{4}$$

Since g is finite, M_g does not have an absorbing element. Let $\lim_{x \to a} g'(x) = \infty$.

$\exists y \neq a : z = M_g(a, y) \neq a$ such that $g'(z) \leq M < \infty$ (because g is locally Lipschitz). Then inequality (4) fails, because we can always choose such $x \neq a$ that $g'(x) > 2kM$, which would give us

$$kM < \frac{1}{2} g'(x) \leq k g'(z) \leq kM,$$

which is false. Then $\frac{\partial M_g}{\partial x} > k$, hence M_g is not Lipschitz. □

Now, since g is Lipschitz on $[0, 1]$, it is differentiable almost everywhere, which means that the left- and right-derivatives exist in $[0, 1]$. We start with the case of g differentiable on $[0, 1]$, and then adapt it to g differentiable almost everywhere by using left- and right-derivatives. Let M_g be k-Lipschitz. Then we must have

$$\frac{\partial M_g}{\partial x}(x, y) \leq k.$$

Following the same procedure as in Lemma 1, we get the condition

$$\frac{1}{2} g'(x) \leq k \cdot \inf_{y \in [0,1]} g'(M_g(x, y)) \tag{5}$$

for all $x \in [0, 1]$. Finally, by using left- and right- derivatives g'_-, g'_+ we obtain a general condition for non-smooth increasing generators

$$\begin{aligned} \frac{1}{2} g'_-(x) &\leq k \inf_{z \in [M(x,0), M(x,1)]} g'_-(z) \\ \frac{1}{2} g'_+(x) &\leq k \inf_{z \in [M(x,0), M(x,1)]} g'_+(z) \end{aligned} \tag{6}$$

for all $x \in]0, 1[$, and only one of the above inequalities for $x = 0$ and $x = 1$.

Remark 2 *If g is finite and concave increasing, then it is sufficient to check*

$$\frac{1}{2} g'(x) \leq k g'(M(x, 1)) = k(g' \circ g^{-1}) \left(\frac{g(x)}{2} + \frac{1}{2} \right),$$

(and similarly for left- and right-derivatives if g is not smooth). If g is finite and convex increasing, it is sufficient to check

$$\frac{1}{2} g'(x) \leq k g'(M(x, 0)) = k(g' \circ g^{-1}) \left(\frac{g(x)}{2} \right).$$

Let us provide some examples of Lipschitz and non-Lipschitz quasi-arithmetic means.

Example 1 *If g is linear (M_g is the arithmetic mean), $g'(x) = const$, and M_g is k-Lipschitz for $k = \frac{1}{2}$.*

Example 2 *If $g(x) = x^p, p > 1$ (M_g is a power mean $M_{[p]}$), $g'(x) = px^{p-1}$, and M_g is k-Lipschitz for $k = \left(\frac{1}{2}\right)^{\frac{1}{p}}$. It follows from*

$$\frac{1}{2} p x^{p-1} \leq kp \left(\frac{x^p}{2} \right)^{\frac{p-1}{p}} = kp x^{p-1} \left(\frac{1}{2} \right)^{\frac{p-1}{p}}.$$

Example 3 *If $g(x) = x^p, 0 < p < 1$ (M_g is a power mean $M_{[p]}$), M_g is not Lipschitz by Lemma 1.*

ISBN: 978-989-95079-6-8

3.2 Case of generators infinite at 0

Now we turn to the case II), g increasing with $g(0) = -\infty$, which entails that 0 is the absorbing element of M_g. We have an analogue of Lemma 1. The proof is similar, except that it fails for $a = 0$, hence the modification.

Let g be finite locally Lipschitz and continuously differentiable except at a point a 2 [0; 1]. Then Mg is not kLipschitz for any k

Lemma 2 *Let g be locally Lipschitz and continuously differentiable except at a point $a \in]0,1]$. Then M_g is not k-Lipschitz for any k.*

Let g be finite locally Lipschitz and continuously differentiable except at a point $a \in [0;1]$.

For $x \in]0,1]$ we have condition (6), to which we add condition

$$\frac{1}{2} \lim_{x \to 0^+} \frac{g'_+(x)}{g'_+(M(x,y))} \leq k \tag{7}$$

for any fixed $y \in]0,1]$. This condition may or may not be satisfied depending on the rate at which $M(x,y) \to 0$ as $x \to 0$. The choice of $y > 0$ is irrelevant as $g(y)$ is finite and disappears under the limit.

Example 4 *If $g(x) = -x^p, p < -1$ (M_g is a power mean $M_{[p]}$), $g'(x) = -px^{p-1}$, and M_g is k-Lipschitz for $k = \left(\frac{1}{2}\right)^{\frac{1}{p}}$. Differentiating M_g*

$$\begin{aligned}
\tfrac{\partial M_g}{\partial x} &= \tfrac{1}{p}\left(\tfrac{x^p+y^p}{2}\right)^{\frac{1}{p}-1} \tfrac{p}{2}x^{p-1} \\
&= \left(\tfrac{1}{2}\right)^{\frac{1}{p}} x^{-p(\frac{1}{p}-1)}(x^p+y^p)^{\frac{1}{p}-1} \\
&= \left(\tfrac{1}{2}\right)^{\frac{1}{p}} (1+x^{-p}y^p)^{\frac{1}{p}-1}.
\end{aligned}$$

Given $p < -1$,

$$k = \left(\tfrac{1}{2}\right)^{\frac{1}{p}} \lim_{x\to 0} (1+x^{-p}y^p)^{\frac{1}{p}-1} = \left(\tfrac{1}{2}\right)^{\frac{1}{p}}.$$

Example 5 *Let $M_{[p]}$, $-1 < p < 0$ be a power mean with a generator given by $g(x) = -x^p = -x^{-\frac{1}{q}}$, $q > 1$). The Lipschitz constant will be $k = \sup \frac{\partial M}{\partial x} = 2^q$. To see this*

$$\begin{aligned}
\tfrac{\partial M_g}{\partial x} &= \tfrac{-q}{2}\left(\tfrac{x^{-\frac{1}{q}}+y^{-\frac{1}{q}}}{2}\right)^{-q-1} \cdot \left(-\tfrac{1}{q}x^{-\frac{1}{q}-1}\right) \\
&= 2^q x^{\frac{1}{q}(-q-1)}(x^{-\frac{1}{q}}+y^{-\frac{1}{q}})^{-q-1} \\
&= 2^q\left(1+\left(\tfrac{x}{y}\right)^{\frac{1}{q}}\right)^{-q-1}. \\
k &= \sup\left\{2^q\left(1+\left(\tfrac{x}{y}\right)^{\frac{1}{q}}\right)^{-q-1}\right\} \\
&= 2^q = (\tfrac{1}{2})^{\frac{1}{p}}.
\end{aligned}$$

Condition 7 deals with the asymptotic behavior of the additive generators near 0. Its direct verification for a given g may be difficult. In the remainder of this section we will establish two sufficient conditions that guarantee that a quasi-arithmetic mean is not Lipschitz (although it is continuous). These conditions are easier to verify, and they provide a tool for a quick screening of additive generators with respect to their suitability for applications. One sufficient condition involves an inequality on the derivatives of the inverse of an additive generator. The other condition is that a decreasing additive generator cannot decrease slower than a certain rate (1/polynomial) when $x \to 0$. We will express this rate through the growth of an auxiliary function $1/g^{-1}$, for which the growth is expressed in traditional terms (e.g., polynomial) when $x \to \infty$. First, two simple auxiliary results.

Lemma 3 *If two functions f, g are continuous and differentiable at $x = 0$ and, $f(0) = g(0)$ and $f(x) \geq g(x)$ for $x > 0$, then $f'(0) \geq g'(0)$.*

Proof: Follows directly from the definition of the derivative.

The next result is a well-known condition for comparability of quasi-arithmetic means, see, e.g., [19].

Theorem 1 *Let g_1, g_2 be the generators of quasi-arithmetic means M_{g_1} and M_{g_2},and g_1 decreasing. Then $M_{g_1} \leq M_{g_2}$ if and only if $g_1 \circ g_2^{-1}$ is convex.*

Theorem 2 *Let g be an increasing (decreasing) twice continuously differentiable on $]0,1]$ generator of a quasi-arithmetic mean M_g where $g^{-1} = h$, and $\lim_{x\to 0} g(x) = -\infty$ ($\lim_{x\to 0} g(x) = +\infty$). If $h'^2 - hh'' \geq 0$ then M_g is not Lipschitz.*

Proof. We will show that $M_{[p]} \leq M_g$ for any $-1 < p < 0$ decreasing, and hence by Lemma 3 is not Lipschitz. If $x^p \circ g^{-1}$ is convex, for $-1 < p < 0$ by Theorem 1, with $g_1(x) = x^p$, $M_{[p]} \leq M_g$. Let us show that $(x^p \circ h)'' \geq 0$.

$$\begin{aligned}
(x^p \circ h)' &= ph^{p-1}h' \\
(x^p \circ h)'' &= p(p-1)h^{p-2}h'^2 + ph^{p-1}h'' \\
&= ph^{p-2}\left((p-1)h'^2 + hh''\right) \geq 0.
\end{aligned}$$

Given $ph^{p-2} < 0$ for $p < 0, h > 0$, convexity will hold if for all $p < 0$

$$(1-p)h'^2 - hh'' \geq 0. \tag{8}$$

Therefore $h'^2 - hh'' \geq 0$ implies $(x^p \circ h)'' \geq 0$ and $M_{[p]} \leq M_g$, and by Lemma 3 the Lipschitz constant of M_g is greater than that of $M_{[p]}$, which is $2^{-\frac{1}{p}}$, and $p \to 0^-$. □

Remark 3 *The generator g can be either increasing or decreasing. Clearly when changing g to $-g$, we change $h(x)$ to $h(-x)$. Then $h^{'}$ changes the sign but $h^{''}$ does not, hence the inequality in Theorem 2is the same for either increasing or decreasing generators.*

Example 6 *Using the geometric mean M_g, take $g(x) = \ln x$ with $h(x) = h^{'}(x) = h^{''}(x) = e^x$. Then $(h'^2 - hh'')(x) = e^{2x} - e^{2x} = 0$. Therefore M_g is not Lipschitz.*

For the sake of convenience, we will formulate our next result for decreasing additive generators satisfying $g(0) = \infty$. To obtain the respective condition on the increasing generators, we simply invert the sign of g.

Theorem 3 *Let $h = g^{-1}$ be the inverse of a decreasing generator g of a quasi-arithmetic mean M_g. If the function $\tilde{h} = \frac{1}{h}$ grows faster than any power $x^q, q > 0$, then M_g is not Lipschitz.*

ISBN: 978-989-95079-6-8

Proof. Fix y so that $g(y) = h^{-1}(y) = 0$, which is always possible (we remind that g is defined up to an arbitrary linear transformation).

$$\begin{aligned}\lim_{x\to 0}\frac{\partial M_g(x,y)}{\partial x} &= \lim_{x\to 0}\frac{dh\left(\frac{h^{-1}(x)}{2}\right)}{dx},\\ &= \lim_{x\to 0}\tfrac{1}{2}h'\left(\frac{h^{-1}(x)}{2}\right)(h^{-1})'(x)\\ &= \tfrac{1}{2}\lim_{x\to 0}h'\left(\frac{h^{-1}(x)}{2}\right)\tfrac{1}{h'(h^{-1}(x))}.\end{aligned}$$

Let $z = \frac{h^{-1}(x)}{2}$. Then

$$\lim_{x\to 0}\frac{\partial M_g(x,y)}{\partial x} = \frac{1}{2}\lim_{z\to\infty}\frac{h'(z)}{h'(2z)}.$$

Since h decreases faster than the power function $p(z) = Cz^r$, l'Hôpital's rule gives

$$0 = \lim_{z\to\infty}\frac{h(z)}{p(z)} = \lim_{z\to\infty}\frac{h'(z)}{p'(z)} = \lim_{z\to\infty}\frac{h'(2z)}{p'(2z)}.$$

For convenience of notation take p such that $p'(z) = \frac{1}{z^q}$. Then $p'(z) = p'(2z)2^q$.

$$\begin{aligned}\lim_{x\to 0}\frac{\partial M_g(x,y)}{\partial x} &= \tfrac{1}{2}\lim_{z\to\infty}\tfrac{h'(z)}{h'(2z)} = \tfrac{1}{2}\lim_{z\to\infty}\tfrac{h'(z)}{h'(2z)}\tfrac{p'(2z)2^q}{p'(z)}\\ &= 2^{q-1}.\end{aligned}$$

Since q can be arbitrarily large, the derivative is unbounded and M_g is not Lipschitz.

□

Example 7 *Let the generator be $g(x) = -\ln x$ as in Example 6. Clearly its inverse is $\exp(-x)$, and the auxiliary function $\frac{1}{h(x)} = \exp(x)$, which grows faster than any polynomial, hence the corresponding geometric mean is not Lipschitz.*

Further take any power of the logarithm $g(x) = (-\ln x)^r, r > 1$. The auxiliary function $\frac{1}{h(x)} = \exp(x^{\frac{1}{r}})$, it grows faster than a polynomial, hence the resulting mean is not Lipschitz either. Note that this quasi-arithmetic mean is related to the Aczél-Alsina family of t-norms [3, 5] by the equation

$$M_g = (T_r^{AA})^{\frac{1}{\sqrt{2}}},$$

which shows directly that M_g is not Lipschitz ($f(x) = M_g(x,1) = T_r^{AA}(x,1)^{\frac{1}{\sqrt{2}}} = x^{\frac{1}{\sqrt{2}}}$ is not Lipschitz).

Example 8
Consider the generator $g(x) = (-\ln x)^2, \tilde{h}(x) = e^{\sqrt{x}}$. From the previous example, $r = 2$ and we know the resulting mean is not Lipschitz, however this would not have been apparent from the application of Theorem 2, as

$$\begin{aligned}h'^2 - hh'' &= \tfrac{1}{4x}e^{-2\sqrt{x}} - \tfrac{1}{4x}e^{-2\sqrt{x}} - \tfrac{1}{4\sqrt{x^3}}e^{-2\sqrt{x}}\\ &= -\tfrac{1}{4\sqrt{x^3}}e^{-2\sqrt{x}} < 0.\end{aligned}$$

3.3 Weighted quasi-arithmetic means

We adapt conditions (6) and (7) for the case of unequal weights. For this we take partial derivatives with respect to all arguments. The Lipschitz constant is the largest, hence we have conditions

$$\begin{aligned}g'_-(x) &\le \frac{k}{\max w_i}\min_z g'_-(z) \qquad (9)\\ g'_+(x) &\le \frac{k}{\max w_i}\min_z g'_+(z)\end{aligned}$$

where
the minimum for z is over $[M(x,0,\dots,0), M(x,1,\dots,1)]$, and

$$\lim_{x\to 0^+}\frac{g'_+(x)}{g'_+(M(x,c,\dots,c))} \le \frac{k}{\max w_i} \qquad (10)$$

with $c \in]0,1]$.

Conditions (9) and (10) can also be used for symmetric means in the multivariate case, where $\max w_i = \frac{1}{n}$. It is clearly seen that the higher the number of variables, the smaller is the Lipschitz constant, if it exists.

Remark 4 *Similar, results can be obtained for generalized OWAs and generalized Choquet integrals.*

4 Conclusions

k-Lipschitz aggregation functions are important for applications because they can control the changes in the outputs due to input inaccuracies, to a fixed factor of k. k-Lipschitz triangular norms and conorms have been already characterized by k-convex additive generators, however no analogous results were available for quasi-arithmetic means. We have found verifiable conditions which guarantee that an aggregation function is $k-$Lipschitz for a given k, or alternatively, not Lipschitz. We also presented various examples of both Lipschitz and non-Lipschitz aggregation functions. Our results will benefit those who design aggregation functions for practical applications, as they allow one to make an informed choice on suitability of specific functions for these applications.

Acknowledgements

This work was supported by the Spanish project MTM2006-08322, PR2007-0193 and the European project 143423-2008-LLP-ES-KA3-KA3MP.

References

[1] H.-J. Zimmermann and P. Zysno. Latent connectives in human decision making. *Fuzzy Sets and Systems*, 4:37–51, 1980.

[2] D. Dubois and H. Prade. On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems*, 142:143–161, 2004.

[3] E.P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer, Dordrecht, 2000.

[4] T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar. Aggregation operators: properties, classes and construction methods. In T. Calvo, G. Mayor, and R. Mesiar, editors, *Aggregation Operators. New Trends and Applications*, pages 3–104. Physica-Verlag, Heidelberg, New York, 2002.

[5] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners*. Springer, Heidelberg, Berlin, New York, 2007.

[6] V. Torra and Y. Narukawa. *Modeling Decisions. Information Fusion and Aggregation Operators*. Springer, Berlin, Heidelberg, 2007.

[7] C. Alsina, M.J. Frank, and B. Schweizer. *Associative Functions: Triangular Norms And Copulas*. World Scientific, Singapore, 2006.

[8] T. Calvo and R. Mesiar. Stability of aggregation operators. In *2nd Conf. of the Europ. Soc. for Fuzzy Logic and Technology*, pages 475–478, Leicester, 2001.

[9] E.P. Klement and A. Kolesárová. Extension to copulas and quasi-copulas as special 1-Lipschitz aggregation operators. *Kybernetika*, 41:329–348, 2005.

[10] A. Kolesárová. 1-Lipschitz aggregation operators and quasi-copulas. *Kybernetika*, 39:615–629, 2003.

[11] A. Kolesárová, E. Muel, and J. Mordelová. Construction of kernel aggregation operators from marginal values. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 Suppl.:37–49, 2002.

[12] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North-Holland, New York, 1983.

[13] A. Mesiarová. *Special Classes of Triangular Norms*. Phd thesis, Slovak Technical University, 2005.

[14] A. Mesiarová. Extremal k-Lipschitz t-conorms. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 14:247–257, 2006.

[15] A. Mesiarová. Approximation of k-Lipschitz t-norms by strict and nilpotent k-Lipschitz t-norms. *Int. J. General Syst*, 36:205 – 218, 2007.

[16] C.M. Ling. Representation of associative functions. *Publ. Math. Debrecen*, 12:189–212, 1965.

[17] Y.-H. Shyu. *Absolute continuity in the τ-operations*. Phd thesis, Illinois Institute of Technology, 1984.

[18] M. Komorníková. Aggregation operators and additive generators. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:205–215, 2001.

[19] P.S. Bullen. *Handbook of Means and Their Inequalities*. Kluwer, Dordrecht, 2003.

ISBN: 978-989-95079-6-8

Modeling Position Specificity in Sequence Kernels by Fuzzy Equivalence Relations

Ulrich Bodenhofer, Karin Schwarzbauer, Mihaela Ionescu, and Sepp Hochreiter

Institute of Bioinformatics, Johannes Kepler University Linz, 4040 Linz, Austria
Email: {bodenhofer,schwarzbauer,ionescu,hochreit}@bioinf.jku.at

***Abstract*— *This paper demonstrates that several known sequence kernels can be expressed in a unified framework in which the position specificity is modeled by fuzzy equivalence relations. In addition to this interpretation, we address the practical issues of positive semi-definiteness, computational complexity, and the extraction of interpretable features from the final support vector machine classifier.*

***Keywords*—** fuzzy equivalence relation, kernel, sequence classification, support vector machines.

1 Introduction

The classification of biological sequences — in particular, nucleotide sequences (DNA and RNA) and amino acid sequences (proteins) — is one of the fundamental tasks in computational biology [1]. In the early days of computational biology, sequence statistics were employed. These approaches were later improved by Hidden Markov Models (HMM) and Artificial Neural Networks (ANN). In the last decade, *Support Vector Machines* (SVM) have become increasingly popular for sequence classification [2]. In many tasks — in computational biology, but also in many other domains — SVMs have outperformed all competing methods. It is justified to state that SVMs are nowadays widely considered the most powerful class of classification methods.[1] SVMs have been applied successfully in a wide spectrum of sequence classification tasks, ranging from promoter and splice site detection (both on DNA data) [3, 4, 5] to protein fold and secondary structure prediction (both on amino acid sequences) [6, 7, 8].

SVM classifiers, in their simplest form, are nothing else but linear classifiers. What distinguishes SVMs from other linear classifiers like perceptrons or logistic regressors is the fact that SVMs determine the classification function by maximizing the margin between the classes — a principle that is known to be optimal in terms of bounds on the generalization error [9, 10]. Since linear SVMs use input vectors only to compute scalar products, SVMs facilitate the so-called *kernel trick*, i.e. the replacement of the scalar products by a *kernel*, i.e. a nonlinear two-place function that is positive semi-definite. The use of kernels enables support vector machines to be applicable to almost any kind of data, including raw sequence data, signals, images, or graphs — only an appropriate kernel is needed. It is not surprising, therefore, that the design of appropriate sequence kernels is one of the central research topics in sequence classification with SVMs. Most of these sequence kernels are based on comparing sub-sequences and do not take the positions of these occurrences into account. In many applications, however, the occurrence of a specific sub-sequence is only indicative if it is at a specific position or region.

[1] although one has to admit that there are applications and circumstances/requirements under which other methods may be preferable

This paper repeats established sequence kernels that are based on occurrences of specific patterns along with some position-specific variants. We demonstrate that these kernels can be expressed in a unified framework in which the position specificity is modeled by fuzzy equivalence relations. This framework, on the one hand, allows for an interpretation of existing position-specific sequence kernels from the viewpoint of fuzzy equivalence relations. On the other hand, and more importantly from the practical viewpoint, this framework gives rise to new position-specific sequence kernels by allowing to use fuzzy equivalence relations that have not been considered for modeling position specificity previously.

2 Support vector classification

Suppose that we have to do binary classification of samples from a given arbitrary non-empty set $\mathcal{X}$. The two classes are denoted with $+1$ (positive class) and -1 (negative class). For a given training set

$$\{(x_i, y_i) \mid 1 \leq i \leq l\}$$

with $x_i \in \mathcal{X}$ and $y_i \in \{-1,+1\}$ for all $i = 1,\ldots,l$, the support vector machine classifier is represented as the following discriminant function:

$$f(x) = b + \sum_{i=1}^{l} \alpha_i \cdot y_i \cdot k(x, x_i) \tag{1}$$

In this formula, b is a real value, α_i are non-negative factors, and $k(.,.)$ is the so-called *kernel*, that is, a symmetric $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$ mapping fulfilling *positive semi-definitenes*, i.e.

$$\sum_{i=1}^{n} \sum_{j=1}^{n} z_i \cdot z_j \cdot k(x_i, x_j) \geq 0 \tag{2}$$

for all $n \in \mathbb{N}$, all $(z_1,\ldots,z_n) \in \mathbb{R}^n$, and all $(x_1,\ldots x_n) \in \mathcal{X}^n$. With the notations $\mathbf{z} = (z_1,\ldots,z_n)^T$ and $\mathbf{K} = (k(x_i,x_j))_{i=1,\ldots,n}^{j=1,\ldots,n}$, Eq. (2) can be written as $\mathbf{z}^T\mathbf{K}\mathbf{z} \geq 0$. Hence, the positive semi-definiteness of k corresponds to the positive semi-definiteness of any kernel matrix $\mathbf{K}$.

The factors $\alpha_1,\ldots,\alpha_n$ are the Lagrange multipliers of a convex quadratic optimization problem arising from margin maximization. For details, the reader is referred to introductory tutorials [11, 12] and standard SVM literature [9, 10, 13, 14]. We only note shortly that samples contributing to the final classifier, obviously those x_i for which $\alpha_i > 0$, are called *support vectors*.

The positive semi-definiteness of the kernel k serves for two purposes. First, it ensures that the SVM optimization problem is a convex quadratic one, which ensures the existence

ISBN: 978-989-95079-6-8

of a global solution that can be determined efficiently. More importantly, by the famous *Mercer theorem* [15], it guarantees that there exists a Hilbert space $(\mathcal{H}, \langle .,. \rangle)$ and a mapping $\varphi : \mathcal{X} \to \mathcal{H}$ such that the representation

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle \tag{3}$$

holds for all $x, y \in \mathcal{X}$. This means that the kernel k can be understood as a scalar product in a feature space $\mathcal{H}$.

3 An overview of sequence kernels

In this paper, we restrict ourselves to sequence kernels that are based on the occurrence of specific patterns. Assume in the following that we have a certain finite alphabet $\mathcal{A}$ and that $\mathcal{X}$ is a set of finite strings over the alphabet $\mathcal{A}$. Further assume that $\mathcal{M}$ is a finite set of patterns. A pattern can either be a string over $\mathcal{A}$ itself (i.e. an *exact pattern*) or contain *wildcards*, i.e. positions that match to an arbitrary symbol or a subset of symbols from the alphabet $\mathcal{A}$. We consider sequence kernels that can be expressed as

$$k(x,y) = \sum_{m \in \mathcal{M}} N(m,x) \cdot N(m,y), \tag{4}$$

where $N(m,x)$ denotes the number of occurrences/matches of pattern m in string x. It is obvious that, according to the representation (3), such a kernel is nothing else but a scalar product in an $|\mathcal{M}|$-dimensional real feature space, where the feature mapping φ is explicitly given as

$$\varphi(x) = (N(m,x))_{m \in \mathcal{M}}. \tag{5}$$

In order to accommodate different matching criteria used in existing sequence kernels, let us express $N(m,x)$ as

$$N(m,x) = \sum_{p=1}^{\text{length}(x)} \mathbf{1}(m,x,p), \tag{6}$$

where $\mathbf{1}(m,x,p)$ is a kernel-specific indicator function whose value is 1 if pattern m matches string x at position p and 0 if not. So, in the framework presented here, a sequence kernels is basically given by its pattern set $\mathcal{M}$ and its match indicator function $\mathbf{1}$. This representation accommodates the following well-known sequence kernels:

Spectrum kernel [6]: $\mathcal{M} = \mathcal{A}^K$, i.e. the set of patterns is the set of all K-length strings; the indicator function is given as[2]

$$\mathbf{1}(m,x,p) = \begin{cases} 1 & \text{if } p+K-1 \leq \text{length}(x) \text{ and} \\ & m = x[p \ldots p+K-1], \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the spectrum kernel maps each of the two sequences to the numbers of occurrences of all strings of length K and computes the scalar product of these two feature vectors.

[2]For a given string x, we define $x[r]$ to be the r-th character in x and $x[r \ldots q]$ to be the substring that starts with the r-th and ends with the q-th position.

Mismatch kernel [7]: this kernel is analogous to the spectrum kernel with the exception that matches are considered up to a maximal number of M mismatches, i.e. $\mathcal{M} = \mathcal{A}^K$ and

$$\mathbf{1}(m,x,p) = \begin{cases} 1 & \text{if } p+K-1 \leq \text{length}(x) \text{ and} \\ & |\{1 \leq r \leq K \mid m[r] \neq x[p+r-1]\}| \leq M, \\ 0 & \text{otherwise.} \end{cases}$$

Motif kernel [16]: $\mathcal{M}$ is a set of predefined patterns known or assumed to be related to the given classification task; the indicator mapping $\mathbf{1}$ is given as follows:[3]

$$\mathbf{1}(m,x,p) = \begin{cases} 1 & \text{if } p+\text{length}(m)-1 \leq \text{length}(x) \text{ and} \\ & m[r] = x[p+r-1] \text{ for all } 1 \leq r \leq \text{length}(m) \\ & \text{such that } m[r] \neq \text{“}*\text{”} \\ 0 & \text{otherwise} \end{cases}$$

Beside these three prominent kernels, also some others comply with the representation (4), e.g. spatial sample kernels [8].

The class of kernels (4) facilitates easy feature extraction, which is not surprising given the fact that the feature mapping φ is known explicitly. Suppose we are given a support vector machine with Lagrange multipliers $\alpha_1, \ldots, \alpha_l$ and an offset b. Then we can make the following rearrangements:

$$\begin{aligned} f(x) &= b + \sum_{i=1}^{l} \alpha_i \cdot y_i \cdot k(x,x_i) \\ &= b + \sum_{i=1}^{l} \alpha_i \cdot y_i \cdot \sum_{m \in \mathcal{M}} N(m,x) \cdot N(m,x_i) \\ &= b + \sum_{m \in \mathcal{M}} N(m,x) \cdot \underbrace{\sum_{i=1}^{l} \alpha_i \cdot y_i \cdot N(m,x_i)}_{=w(m)} \end{aligned} \tag{7}$$

So we obtain $w(m)$ as the linear scaling factor with which every occurance/match of m in the sequence x is scored. Obviously, the higher the absolute value $w(m)$, the more important the pattern m is for the final classification. If $w(m)$ is positive, the pattern m is indicative for the positive class and, if $w(m)$ is negative, m is indicative for the negative class.

Note that the representation (7) is not only beneficial for feature extraction, i.e. for determining which patterns are indicative for the classification tasks. If the total number of patterns actually occurring in the training set is not excessively large, the explicit representation (7) also has strong advantages in terms of computational complexity: the weights $w(m)$ can be stored in a simple hash table indexed by the patterns. For a new sample x, it is then only necessary to sum up the weights for all occurrences of all patterns in x. For the spectrum kernel, for instance, this means adding up $\text{length}(x) - K + 1$ numbers, whereas the direct implementation of the general SVM (1) requires the computation of all $k(x,x_i)$ for which $\alpha_i > 0$ and, therefore, much more computational effort.

[3]We restrict to fixed-length motifs with only regular characters and the wildcard “$*$” that matches all single characters, although the generalization to more sophisticated patterns, e.g. by regular expressions, is straightforward.

ISBN: 978-989-95079-6-8

4 Fuzzy position preference

The position-independent kernels presented in Section 3 have been used successfully in protein classification. Many protein classification tasks are concerned with inferring the structure and/or function of proteins from their amino acid sequences. Structure and function of proteins are largely determined by shorter sub-sequences (patterns/motifs), mostly irrespective of their exact position in the sequence, so position specificity is not an important issue.

When analyzing DNA or RNA nucleotide sequences, the situation is quite different. There is a rather small alphabet (four nucleic acids instead of twenty amino acids), so some patterns (in particular, shorter ones) are less indicative as they may occur by chance. The position where a pattern occurs, for example, relative to the start or end of a gene, plays a much bigger role. Therefore, there is a strong need for position-specific sequence classification.

We now want to illustrate how position specificity can be integrated into the class of kernels introduced in the previous section. So let us assume that all sequences we consider have the same length L, i.e. $\mathcal{X} = \mathcal{A}^L$, and are *aligned in a way meaningful for the given classification task*.

The simplest way to include position dependence is to consider each pattern at each position completely independent of the other occurrences at other positions (similar to the philosophy behind the *weighted degree kernel* [3]). This means that, instead of mapping a sequence to the total numbers of occurrences of all patterns (cf. (5)), each pattern induces L features, one for each position. The explicit feature mapping can be formulated as follows:

$$\varphi(x) = \Big(\underbrace{(\mathbf{1}(m,x,1),\ldots,\mathbf{1}(m,x,L))}_{\text{binary list of occurrences of pattern } m} \Big)_{m\in\mathcal{M}} \tag{8}$$

Then the corresponding position-specific kernel is given as follows:

$$\bar{k}(x,y) = \sum_{m\in\mathcal{M}} \sum_{p=1}^{L} \mathbf{1}(m,x,p)\cdot\mathbf{1}(m,y,p) \tag{9}$$

So this kernel basically counts the number of positions where the same pattern matches both sequences x and y. It is clear that the number of features (i.e. the dimension of the feature space $\mathcal{H}$) grows by a factor of L compared to the position-independent kernel (4) — potentially increasing the risk of overfitting. Moreover, this approach is not ideally suited for tasks in which patterns are indicative if they appear in certain regions (but not necessarily at exact positions).

The *oligo kernel* [5] solves this challenge by replacing the binary indicators in the feature mapping (8) by the sum of proximity functions around the occurrences of the pattern, with Gaussian bell functions being the standard choice. We leave this variant aside for a moment and concentrate on a different approach in line with the *shifted weighted degree (SWD) kernel* [4]. As in the kernel $\bar{k}$, a pattern m occurring at the same position in both sequences contributes 1 to the sum, but also an occurrence of a pattern m at position p in the sequence x and an occurrence of the same pattern m at position q in the sequence y may contribute a non-zero value to the sum. This is solved by a two-place weighting function $E : \{1,\ldots,L\}^2 \to [0,1]$ which corresponds to the closeness of positions p and q. The farer p and q are apart, the lower the value $E(p,q)$ should be. If $p = q$, i.e. if the positions coincide exactly, $E(p,q) = 1$ holds. So we may generalize the kernel $\bar{k}$ in the following way:

$$\tilde{k}(x,y) = \sum_{m\in\mathcal{M}} \sum_{p=1}^{L} \sum_{q=1}^{L} \mathbf{1}(m,x,p)\cdot E(p,q)\cdot\mathbf{1}(m,y,q) \tag{10}$$

If we leave some constant factors and the SWD kernel's option to define importance factors for each position aside, the position preference weights used by the SWD kernel essentially come down to

$$E_{\mathrm{SWD}}(p,q) = 2^{-|p-q|}.$$

It is easy to see that the kernel $\bar{k}$ given in (9) is a special case of $\tilde{k}$ with the position weighting function

$$E_{=}(p,q) = \begin{cases} 1 & \text{if } p = q, \\ 0 & \text{otherwise,} \end{cases}$$

i.e. the ordinary equality relation, and the position-independent variant defined in (4) is also a special case of (10) with the trivial position-independent weighting function $E_1(p,q) = 1$ (for all $p,q \in \{1,\ldots,L\}$).

Since the range of the weighting function E is the unit interval and since it is intended to be a model of the closeness of the two positions, it is immediate to ask the question whether fuzzy equivalence relations [17, 18] would be reasonable choices for E. As usual, a mapping $E : \mathcal{X}^2 \to [0,1]$ is called *fuzzy equivalence relation* with respect to a given triangular norm T if it has the following properties (for all $x,y,z \in \mathcal{X}$):

(i) Reflexivity: $E(x,x) = 1$

(ii) Symmetry: $E(x,y) = E(y,x)$

(iii) T-transitivity: $T(E(x,y),E(y,z)) \le E(x,z)$

It is trivial that the two special cases $E_=$ and E_1 highlighted above are fuzzy equivalence relations, no matter which triangular norm (t-norm) we consider.

It is also straightforward to prove that E_{SWD} is a fuzzy equivalence relation with respect to the product t-norm $T_{\mathbf{P}}(x,y) = x\cdot y$. This can be proved directly, but it is essentially a consequence of the well-known correspondence between pseudo-metrics and fuzzy equivalence relations with respect to continuous Archimedean t-norms [19]. As further consequences of this result, we can deduce some more examples (generally for $x,y \in \mathbb{R}$, containing the special case of natural numbers $1,\ldots,L$):

- $E_{\mathrm{lin},\sigma}(x,y) = \max(1 - \frac{1}{\sigma}|x-y|, 0)$ is a fuzzy equivalence relation with respect to the Łukasiewicz t-norm $T_{\mathbf{L}}(x,y) = \max(x+y-1,0)$.
- $E_{\exp,\sigma}(x,y) = \exp(-\frac{|x-y|}{\sigma})$ is a fuzzy equivalence relation with respect to the product t-norm $T_{\mathbf{P}}$.
- $E_{\mathrm{Gauss},\sigma}(x,y) = \exp(-\frac{(x-y)^2}{\sigma^2})$ is a fuzzy equivalence relation with respect to the t-norm

  $$T(x,y) = \exp\big(-(\sqrt{-\ln x} + \sqrt{-\ln y})^2\big),$$

  which is nothing else but the Aczél-Alsina t-norm with parameter $\lambda = \frac{1}{2}$ [20].

ISBN: 978-989-95079-6-8

The question arises which choices of E are feasible. In principle, every choice for which $E(p,q)$ is a non-increasing transformation of $|p-q|$ seems reasonable, and all examples mentioned above are of that kind.[4] However, there is also a technical requirement that should not be forgotten: the resulting function $\tilde{k}$ must be positive semi-definite. In order to deduce meaningful requirements on the fuzzy equivalence relation E, let us make a simple reformulation of $\tilde{k}$. It is trivial that the two inner sums are just a vector-matrix-vector product

$$\sum_{p=1}^{L}\sum_{q=1}^{L}\mathbf{1}(m,x,p)\cdot E(p,q)\cdot\mathbf{1}(m,y,q)=\varphi_m(x)^T\cdot\mathbf{E}\cdot\varphi_m(y),$$

with

$$\varphi_m(x)=\big(\mathbf{1}(m,x,1),\ldots,\mathbf{1}(m,x,L)\big)^T,$$
$$\varphi_m(y)=\big(\mathbf{1}(m,y,1),\ldots,\mathbf{1}(m,y,L)\big)^T,$$
$$\mathbf{E}=\begin{pmatrix}E(1,1)&\cdots&E(1,L)\\ \vdots&\ddots&\vdots\\ E(L,1)&\cdots&E(L,L)\end{pmatrix}.$$

Hence, we can rewrite $\tilde{k}$ as

$$\tilde{k}(x,y)=\sum_{m\in\mathcal{M}}\underbrace{\varphi_m(x)^T\cdot\mathbf{E}\cdot\varphi_m(y)}_{=\tilde{k}_m(x,y)}.$$

The sum of kernels is again a kernel [21]. So if we can guarantee that each $\tilde{k}_m$ is a kernel, $\tilde{k}$ is guaranteed to be a kernel. It is clear that, if the matrix $\mathbf{E}$ is positive semi-definite, $\tilde{k}_m$ is a scalar product, hence a kernel. Then the positive semi-definiteness of E, i.e. that E can be understood as a kernel itself, is a sufficient criterion for $\mathbf{E}$ to be positive semi-definite — and finally for $\tilde{k}$ to be a kernel.

For the first two examples above, the situation is clear: $E_=$ induces the identity matrices, E_1 induces matrices containing only 1's; both classes of matrices are trivially positive semi-definite. For non-trivial situations, fortunately, several results are known as well:

- Every fuzzy equivalence relation with respect to the minimum t-norm $T_{\mathbf{M}}(x,y)=\min(x,y)$ is positive semi-definite [22].
- $E_{\exp,\sigma}$ and $E_{\text{Gauss},\sigma}$ are long known to be positive semi-definite [23], where the latter is one of the most used kernels — the RBF kernel.
- $E_{\text{lin},\sigma}$ has recently been proved to be positive semi-definite [24].
- E_{SWD} is also positive semi-definite, as obviously every matrix $\mathbf{E}$ induced by E_{SWD} is diagonally dominant.

So we can conclude that the three choices $E_{\exp,\sigma}$, $E_{\text{Gauss},\sigma}$, $E_{\text{lin},\sigma}$ and E_{SWD} are reasonable and technically feasible. Further note that the former three choices have in common that the resulting kernels tend to the position-independent variant (4) as σ goes to infinity, and the resulting kernels tend to the most position-specific variant (9) as $\sigma\to 0$. So the parameter σ allows for adjusting the degree of position specificity in a continuous manner.

We further note that also the oligo kernel [5] can be accommodated in the framework introduced above. This kernel uses an explicit feature representation for each pattern that convolves the occurrences with Gaussian neighborhood (for some a priori chosen σ). So we can integrate the oligo kernel into the above framework with

$$\mathbf{E}_{\text{oligo},\sigma}=\mathbf{E}_{\text{Gauss},\sigma}\cdot\mathbf{E}_{\text{Gauss},\sigma}.$$

So, from the computational point of view, everything said here is also valid for the oligo kernel. However, since entries of the matrix $\mathbf{E}_{\text{oligo},\sigma}$ can exceed the unit interval, this distance weighting matrix cannot be interpreted as a fuzzy equivalence relation.

5 Feature extraction and computational issues

There is no doubt that the position-specific variant introduced in the previous section is more complicated than the position-independent variant presented in Section 3. The question arises whether the position-specific framework also allows for simple feature extraction and computational efficiency like the position-independent variant. Fortunately, this is the case. The basis of this result is the following rearrangement of the SVM discriminant function (analogous to Section 3):

$$\begin{aligned}f(x)&=b+\sum_{i=1}^{l}\alpha_i\cdot y_i\cdot\tilde{k}(x,x_i)\\&=b+\sum_{i=1}^{l}\alpha_i\cdot y_i\cdot\sum_{m\in\mathcal{M}}\sum_{p=1}^{L}\sum_{q=1}^{L}\mathbf{1}(m,x,p)\cdot E(p,q)\cdot\mathbf{1}(m,x_i,q)\\&=b+\sum_{m\in\mathcal{M}}\sum_{p=1}^{L}\mathbf{1}(m,x,p)\cdot\underbrace{\sum_{i=1}^{l}\sum_{q=1}^{L}\mathbf{1}(m,x_i,q)\cdot\alpha_i\cdot y_i\cdot E(p,q)}_{=\tilde{w}(m,p)}\end{aligned}$$

A value $\tilde{w}(m,p)$ can be interpreted as the weight to which an occurrence of pattern m at position p contributes to the final result. Analogously to Section 3, $\tilde{w}(m,p)>0$ means that an occurrence of pattern m at position p is indicative for the positive class, whereas $\tilde{w}(m,p)<0$ tells us that an occurrence of pattern m at position p is indicative for the negative class. The absolute value of $\tilde{w}(m,p)$ corresponds to the strength of this influence.

It is worth to point out that the way the weights $\tilde{w}(m,p)$ are computed has a very intuitive interpretation too. Suppose that pattern m occurs at position q in a support vector x_i. Obviously, if we consider p a free variable, $E(p,q)$ is the proximity (equivalence class) of position q. In case $E=E_{\text{lin},\sigma}$ is used, it is nothing else but a triangle with width σ; in the case $E=E_{\exp,\sigma}$, it is a Gaussian bell. So the term $\alpha_i\cdot y_i\cdot E(p,q)$ is nothing else but this proximity scaled by the Lagrange multiplier α_i and the sign y_i. Thus, the list $(\tilde{w}(m,p))_{p=1,\ldots,L}$ is the superimposition of proximities of all occurrences of pattern m in the training set scaled by corresponding Lagrange multipliers and signs/classes.

The list $(\tilde{w}(m,p))_{p=1,\ldots,L}$ can be plotted as a graph over the sequence showing the influence of pattern m at each position, indicating in which regions of the sequence the pattern is indicative for either class or not indicative at all. Thus,

[4] More generally, we could also apply a transformation to the positions first.

ISBN: 978-989-95079-6-8

the framework presented here allows for the same simple feature extraction as the position-independent variant. For other methods to extract features from SVM-based sequence classifiers, see [25, 26].

Moreover, it is obvious from the reformulation above that the discriminant function $f(x)$ can be evaluated in the same way as the position-independent variant — by adding up weights over all patterns occurring in the sequence x. So there is no difference in computational complexity between the position-specific and the position-independent variant. However, we have to store L times as many weights. This may be an obstacle if the number of patterns occurring in the training set and the sequence length L are both high.

For the position-specific spectrum kernel, for instance, the situation is very convenient. A sequence contains exactly $L-K+1$ sub-strings. Thus, the computation of $f(x)$ can be done by summing up $L-K+2$ values (again assuming that the lists of weights are stored in a convenient hash table). Obviously, the choice of K is of little influence on this complexity, which is is not true if we consider memory requirements. For low K's, the memory needed to store the weights grows exponentially (as $|\mathcal{M}| = \mathcal{A}^K$), but the total number of patterns that need to be considered is bounded above by the total number of patterns occurring in all support vectors.

All said above is valid for the spectrum kernel only. The mismatch kernel, for instance, is a much more difficult matter, as every sub-string of length K can match several patterns (where this number grows exponentially with M). Thus, the total number of patterns to be considered can be much higher and, more importantly, the number of patterns that need to be taken into account for computing $f(x)$ according to the above principle is much higher. Moreover, the extraction of features in the above manner is possible in principle, but impeded by the fact that the sets of sub-strings matching two different patterns may have large overlaps.

6 A DNA classification example

In this section, we demonstrate the concept introduced above by means of a case study. We consider the task of characterizing long nucleosome-free DNA segments.

Nucleosomes are a DNA packing mechanism in eukaryotic genomes. A nucleosome basically consists of a protein complex (a histone octamer) around which approximately 147 DNA base pairs (bp) are wrapped in $1\frac{2}{3}$ turns. Nonchalantly speaking, the histone complex acts as a reel around which DNA is wrapped. The DNA segment around the histone is mostly inaccessible for interactions with other molecules, most importantly, RNA polymerase. That is why the positions of nucleosomes and possible change of those positions play an essential role in transcription regulation. Therefore, it is of essential interest for the systems biology of eukaryotic cells which mechanisms determine the positioning and repositioning of nucleosomes. An interesting sub-topic is whether there are specific feature of the DNA that favor or hamper the positioning of nucleosomes.

For the yeast sub-species *Saccharomyces cerevisiae*, a commonly used model organism, data are available about where nucleosomes are located [27]. Instead of trying to find out which DNA features favor nucleosomes [28, 29], we tried to elicit DNA features that hamper the positioning of nucleosomes. To this end, we selected a total number of 10226 nucleosome-free sub-sequences of 100 bp length which occur before (on either strand) a well-positioned nucleosome (with the end of the sequence aligned to the start of the nucleosome coming thereafter). This was done using the data published along with [27]. These sequences were labeled as $+1$. The data set was complemented by the same number of random 100 bp long DNA fragments occurring in nucleosomes or in between two nucleosomes that are less than 100 bp apart. These sequences were labeled as -1.

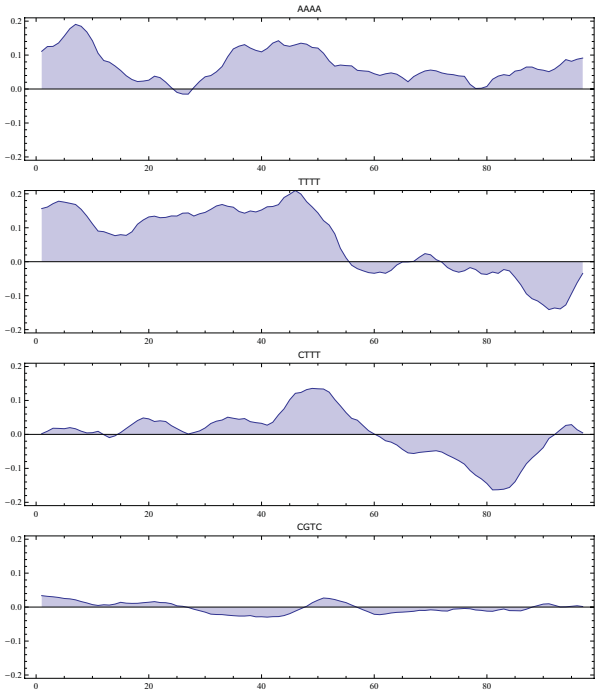


Figure 1: Weight profiles for the four patterns "AAAA", "TTTT", "CTTT" and "CGTC" (from top to bottom).

Our investigations were carried out as follows. The choices $\mathcal{A} = \{\text{"A"}, \text{"G"}, \text{"C"}, \text{"T"}\}$ and $L = 100$ were given by the data set under consideration. We used the spectrum kernel, both position-independent and position-specific, with different values for K ranging from 2 to 6. For the position-specific variant, we used $E = E_{\text{lin},\sigma}$ with σ's of 5, 10, and 20. We employed libSVM [30] to train standard soft margin SVMs. The best cross validation accuracy of 79.2% was obtained for the position-specific variant with $K = 4$ and $\sigma = 10$.

Let us have a closer look at the best setting $K = 4$. Obviously, we have $|\mathcal{M}| = 4^4 = 256$ different patterns. Hence, in order to realize the principle described in the previous section we have to store at most $256 \cdot 100 = 25,600$ numbers, which is easily manageable. Computing the discriminant function $f(x)$ for a new sequence x requires only to slide a window of length $K = 4$ over the sequence and to sum up the corresponding weights, hence only 98 additions are necessary, which is negligible compared to directly applying Eq. (1).

Figure 1 shows four examples of such weight profiles obtained for the given classification task. It is obvious that "AAAA" is indicative for the positive class (longer nucleosome-free positions), more or less regardless of the position. This is in line with the result of Peckham et al. [29]

ISBN: 978-989-95079-6-8

who have found out that poly-A occurrences are indicative for non-nucleosome sequence regions. One would expect the same situation for the complementing pattern "TTTT", but the profile in Figure 1 shows that this pattern is indicative for the positive class only in the first half of the sequence, whereas it is not indicative for the last 45 bases before a nucleosome. The pattern "CTTT" is not indicative for either class in the first half of the sequence, it seems to occur frequently around 50 bases before a nucleosome, and it occurs with less-than-random probability in the last 40 bases before a nucleosome (note that the negative class is randomly sampled from nucleosomes and shorter nucleosome-free fragments). The last graph shows that pattern "CGTC" seems to be of no significance at all. Although the biological interpretation of these results is a more advanced topic, the examples demonstrate how easily claims about the significance of specific patterns can be deduced from the weight profiles.

7 Concluding remarks

In this paper, we have formulated a generalization of position-specific sequence kernels by using fuzzy equivalence relations for modeling position specificity. This is not only an interesting link, but also gives rise to new kernels, e.g. the one based on $E_{\mathrm{lin},\sigma}$. We have obtained that these kernels facilitate an explicit representation which allows for computational efficiency and easy feature extraction.

References

[1] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

[2] B. Schölkopf, K. Tsuda, and J.-P. Vert, editors. *Kernel Methods in Computational Biology*. MIT Press, Cambridge, MA, 2004.

[3] G. Rätsch and S. Sonnenburg. Accurate splice site detection for *Caenorhabditis elegans*. In B. Schölkopf, K. Tsuda, and J.-P. Vert, editors, *Kernel Methods in Computational Biology*, pages 277–298. MIT Press, Cambridge, MA, 2004.

[4] G. Rätsch, S. Sonnenburg, and B. Schölkopf. RASE: recognition of alternatively spliced exons in *C.elegans*. *Bioinformatics*, 21(Suppl. 1):i369–i377, 2005.

[5] P. Meinicke, M. Tech, B. Morgenstern, and R. Merkl. Oligo kernels for datamining on biological sequences: a case study on prokaryotic translation initiation sites. *BMC Bioinformatics*, 5:169, 2004.

[6] C. Leslie, E. Eskin, and W. S. Noble. The spectrum kernel: a string kernel for SVM protein classification. In R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale, and T. E. D. Klein, editors, *Pacific Symposium on Biocomputing 2002*, pages 566–575. World Scientific, 2002.

[7] C. Leslie, E. Eskin, A. Cohen, J. Weston, and W. S. Noble. Mismatch string kernels for discriminative protein classification. *Bioinformatics*, 1(1):1–10, 2003.

[8] P. Kuksa, P.-H. Huang, and V. Pavlovic. A fast, large-scale learning method for protein sequence classification. In *8th Int. Workshop on Data Mining in Bioinformatics*, pages 29–37, Las Vegas, NV, 2008.

[9] C. Cortes and V. N. Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1986.

[10] V. N. Vapnik. *Statistical Learning Theory*. Adaptive and Learning Systems. Wiley Interscience, 1998.

[11] C. J. M. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2:121–167, 1998.

[12] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Networks*, 12(2):181–201, 2001.

[13] N. Christianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[14] B. Schölkopf and A. J. Smola. *Learning with Kernels*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2002.

[15] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc. London*, A 209:415–446, 1909.

[16] A. Ben-Hur and D. Brutlag. Remote homology detection: a motif based approach. *Bioinformatics*, 19:i26–i33, 2003.

[17] L. A. Zadeh. Similarity relations and fuzzy orderings. *Inform. Sci.*, 3:177–200, 1971.

[18] L. Valverde. On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, 17(3):313–328, 1985.

[19] B. De Baets and R. Mesiar. Pseudo-metrics and T-equivalences. *J. Fuzzy Math.*, 5(2):471–481, 1997.

[20] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*, volume 8 of *Trends in Logic*. Kluwer Academic Publishers, Dordrecht, 2000.

[21] C. H. FitzGerald, C. A. Micchelli, and A. Pinkus. Functions that preserve families of positive semidefinite matrices. *Linear Alg. Appl.*, 221:83–102, 1995.

[22] B. Moser. On the T-transitivity of kernels. *Fuzzy Sets and Systems*, 157(1):1787–1796, 2006.

[23] C. A. Micchelli. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constr. Approx.*, 2:11–22, 1986.

[24] L. Belanche, J. L. Vázquez, and M. Vázquez. Distance-based kernels for real-valued data. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker, editors, *Data Analysis, Machine Learning and Applications*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 3–10. Springer, Berlin, 2008.

[25] T. Lingner and P. Meinicke. Remote homology detection based on oligomer distances. *Bioinformatics*, 22(18):2224–2231, 2006.

[26] S. Sonnenburg, A. Zien, P. Philips, and G. Rätsch. POIMs: positional oligomer importance matrices—understanding support vector machine-based signal detectors. *Bioinformatics*, 24:i6–i14, 2008.

[27] W. Lee, D. Tillo, N. Bray, R. H. Morse, R. W. Davis, T. R. Hughes, and C. Nislow. A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genetics*, 39(10):1235–1244, September 2007.

[28] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. C. Thåström, Y. Field, I. K. Moore, J.-P. Z. Wang, and J. Widom. A genomic code for nucleosome positioning. *Nature*, 442(17):1235–1244, August 2006.

[29] H. E. Peckham, R. E. Thurman, Y. Fu, J. A. Stamatoyannopoulos, W. S. Noble, K. Struhl, and Z. Weng. Nucleosome positioning signals in genomic DNA. *Genome Res.*, 17:1170–1177, 2007.

[30] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

ISBN: 978-989-95079-6-8

Universal Approximation of a Class of Interval Type-2 Fuzzy Neural Networks Illustrated with the Case of Non-linear Identification

Juan R. Castro[1], Oscar Castillo[2], Patricia Melin[2], Antonio Rodriguez-Diaz[1], Olivia Mendoza[1]

[1]Baja California Autonomous University, UABC, Tijuana, B. C., Mexico.
2 Division of Graduate Studies and Research, Tijuana Institute of Technology, Tijuana, B. C., Mexico.
Email: jrcastror@uabc.mx, ocastillo@tectijuana.mx

Abstract—*Neural Networks (NN), Type-1 Fuzzy Logic Systems (T1FLS) and Interval Type-2 Fuzzy Logic Systems (IT2FLS) are universal approximators, they can approximate any non-linear function. Recent research shows that embedding T1FLS on an NN or embedding IT2FLS on an NN can be very effective for a wide number of non-linear complex systems, especially when handling imperfect information. In this paper we show that an Interval Type-2 Fuzzy Neural Network (IT2FNN) is a universal approximator with some precision using a set of rules and Interval Type-2 membership functions (IT2MF) and the Stone-Weierstrass Theorem. Also, simulation results of non-linear function identification using the IT2FNN for one and three variables with 10-fold cross-validation are presented.*

Keywords— Interval Type-2 Fuzzy Logic Systems, Interval Type-2 Fuzzy Neural Networks, Neural Networks, Universal Approximation.

1 Introduction

It has been shown that a three layer NN can approximate any real continuous function [1]. The same has been shown for a T1FLS [2, 3] using Stone-Weierstrass Theorem [6]. A similar analysis was made by Kosko [4, 5] using the concept of fuzzy regions. In [6, 7] Buckley shows that, with a Sugeno model [9], a T1FLS can be built with the ability to approximate any non-linear continuous function. Also, combining neural and fuzzy logic paradigms [10, 11], an effective tool can be created for approximating any non-linear function [8]. In this sense, an expert can use a Type-1 Fuzzy Neural Network (T1FNN) [13, 14] or IT2FNN systems and find interpretable solutions [12, 17-20]. In general, Takagi-Sugeno-Kang (TSK) T1FLS's approximate well by the use of polynomial consequent rules [9, 23].

This paper uses the Levenberg-Marquardt backpropagation learning algorithm for adapting antecedent and consequent parameters for an adaptive IT2FNN, since its efficiency and soundness characteristics made them fit for these optimizing problems.

An Adaptive IT2FNN is used as a universal approximator of any non-linear functions. A set Ψ of IT2FNN is universal if and only if (iff), given any process Ω, there is a system $\Phi \in \Psi$ such that the difference between the output from IT2FNN and output from Ω is less than a given ε.

2 Interval Type-2 Fuzzy Neural Networks

An IT2FNN [12] combines a TKS interval type-2 fuzzy inference system (TSKIT2FIS) [15,16] with an adaptive NN in order to take advantage of both models best characteristics. Even though Mamdami and Tsukamoto are much complex models [13, 14, 16], TSK is preferred due to the lower computational cost in type reduction.

2.1 IT2FNN-2:A2C0 Architecture

An IT2FNN-2:A2C0 (IT2FNN-0) is a seven layer IT2FNN, which integrates a first order TSKIT2FIS (interval type-2 fuzzy antecedents and real consequents), with an adaptive NN. The IT2FNN-2:A2C0 (Figure 1) layers are described as follows:

Layer 0: Inputs

$$o_i^0 = x_i \quad ; i = 1,...,n$$

Layer 1: Adaptive type-1 fuzzy neuron (T1FN)

for i=1,…,n
  for k=s_i+1,…,s_{i+1}
    $net_k^1 = w_{k,i}^{1,0}\, x_i + b_k^1$
    $o_k^1 = \mu(net_k^1)$
  end
end

Layer 2: Non-adaptive T1FN

$o_{2k-1}^2 = o_{2k-1}^1 \cdot o_{2k}^1 \quad \forall\, k = 1,...,\vartheta$
$o_{2k}^2 = o_{2k-1}^1 + o_{2k}^1 - o_{2k-1}^2$

for k=1,…,r
  for i=1,…,n
    $\tau = 1_{k,i} \quad ; \quad \pi = \sum_{j=1}^{i-1} v_j + |\tau|$
    if $\tau > 0$
      $\underline{\mu}_{k,i} = o_{i_l(\pi)}^2 \,;\, \overline{\mu}_{k,i} = o_{i_u(\pi)}^2$
    else
      $\underline{\mu}_{k,i} = null \,;\, \overline{\mu}_{k,i} = null$
    end
  end
end

Layer 3: Lower-upper firing strength rule normalization. Nodes in this layer are non-adaptive and the output is defined as a the ratio between the k^{th} lower-upper firing strength rule and the sum of lower-upper firing strength of all rules:

ISBN: 978-989-95079-6-8

$$o_{2k-1}^3 = \underline{w}^k = \frac{\underline{f}^k}{\sum_{l=1}^{r} \underline{f}^l} \quad ; \quad k = 1,...,r$$

$$o_{2k}^3 = \overline{w}^k = \frac{\overline{f}^k}{\sum_{l=1}^{r} \overline{f}^l}$$

Layer 4: Rule consequents. Each node is adaptive and its parameters are $\{c_i^k, c_0^k\}$. The node's output corresponds to partial output of k[th] rule y^k.

$$y^k = \sum_{i=1}^{n} c_i^k x_i + c_0^k \quad ; \quad k = 1,...,r$$

$$o_{2k-1}^4 = y^k$$

$$o_{2k}^4 = y^k$$

Layer 5: Type reduction. Estimates left-right interval values $[\hat{y}_l, \hat{y}_r]$, nodes are non-adaptive with outputs $\hat{y}_l, \hat{y}_r$. Layer 5 output is defined by:

$$o_1^5 = \hat{y}_l = \sum_{k=1}^{r} \underline{w}^k \cdot y^k$$

$$o_2^5 = \hat{y}_r = \sum_{k=1}^{r} \overline{w}^k \cdot y^k$$

Layer 6: Defuzzification. This layer's node is adaptive, where the output $\hat{y}$ is defined as weighted average of left-right values and parameter γ. Parameter γ (default value 0.5) adjusts the uncertainty interval defined by left-right values $[\hat{y}_l, \hat{y}_r]$.

$$o_1^6 = \hat{y} = \gamma \hat{y}_l + (1-\gamma)\hat{y}_r$$

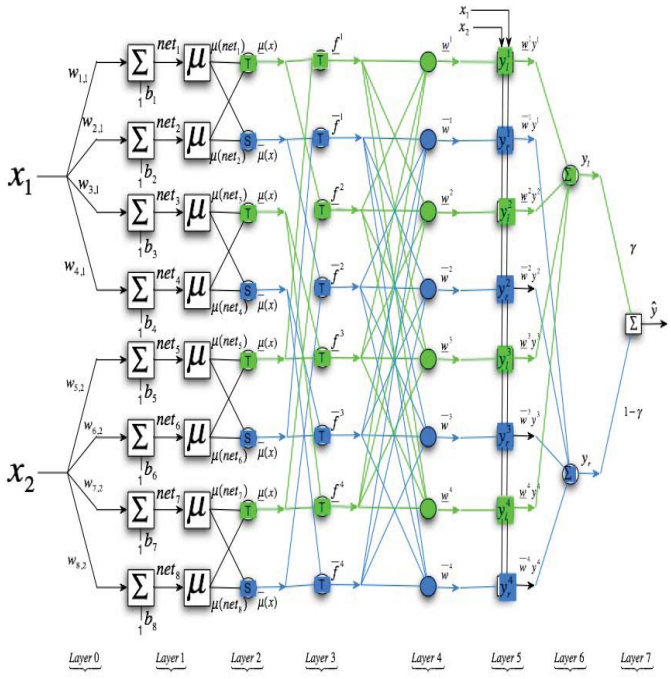


Figure 1: IT2FNN-2:A2C0 (IT2FNN-0) Architecture

2.2 IT2FNN-2:A2C1 Architecture

An IT2FNN-2:A2C1 (IT2FNN-2) [12] is a six layer IT2FNN, which integrates a first order TSKIT2FIS (interval type-2 fuzzy antecedents and interval type-1 fuzzy consequents), with an adaptive NN. The IT2FNN-2:A2C1 (Figure 2) layers are described as follows:

Layers 0-2 of IT2FNN-2:A2C1 architecture are the same as layers 0-2 of IT2FNN-2:A2C0 architecture.

Layer 3: Rule consequents. Each node is adaptive with parameters $\{c_i^k, c_0^k\}$ and $\{s_i^k, s_0^k\}$. Node's output corresponds to k[th] rule's partial output, $\tilde{y}^k \in [y_l^k, y_r^k]$.

$$o_{2k-1}^3 = y_l^k = \sum_{i=1}^{n} c_i^k x_i + c_0^k - \sum_{i=1}^{n} s_i^k |x_i| - s_0^k \quad ; \quad k = 1,...,r$$

$$o_{2k}^3 = y_r^k = \sum_{i=1}^{n} c_i^k x_i + c_0^k + \sum_{i=1}^{n} s_i^k |x_i| + s_0^k$$

Layer 4: Type reduction. Estimates left-right values of interval $[\hat{y}_l, \hat{y}_r]$ and left(f_l^k)-right(f_r^k) firing strength that are used for computing $[\hat{y}_l, \hat{y}_r]$ using KM algorithm [15, 16].

$$f_l^k = leftpoint(\underline{f}^k, \overline{f}^k, y_l^k) \quad ; \quad f_r^k = rightpoint(\underline{f}^k, \overline{f}^k, y_r^k)$$

$$o_1^4 = \hat{y}_l = \frac{\sum_{k=1}^{r} f_l^k y_l^k}{\sum_{k=1}^{r} f_l^k} \quad ; \quad o_2^4 = \hat{y}_r = \frac{\sum_{k=1}^{r} f_r^k y_r^k}{\sum_{k=1}^{r} f_r^k}$$

Layer 5: Defuzzification. This layer's node is adaptive with output $\hat{y}$ defined by:

$$o_1^5 = \hat{y} = \gamma \hat{y}_l + (1-\gamma)\hat{y}_r$$

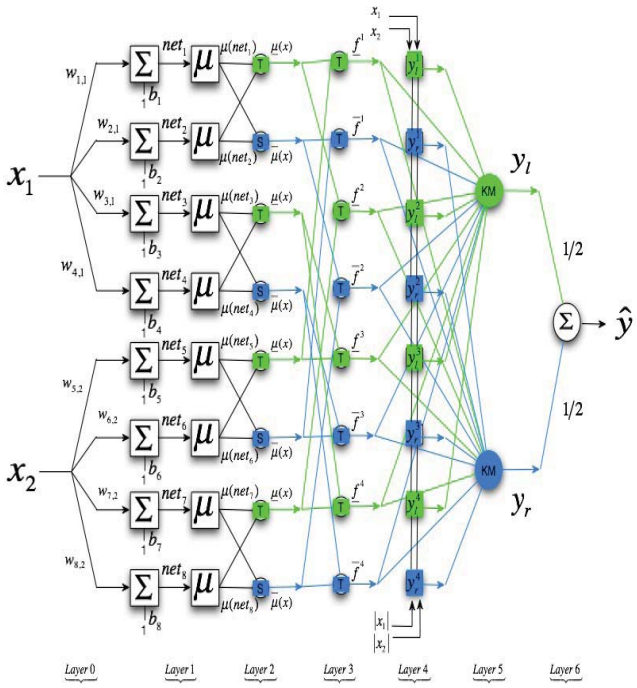


Figure 2: IT2FNN-2:A2C1 (IT2FNN-2) Architecture

ISBN: 978-989-95079-6-8

3 IT2FNN as a Universal Approximator

With this simplified interval type-2 fuzzy if-then rule, it is possible to prove that under certain circumstances, the resulting IT2FIS has unlimited approximation power to match any non-linear functions on a compact [24, 25] set using the Stone-Weierstrass theorem [13, 14, 21, 22] as follows:

3.1 Stone-Weierstrass Theorem.

Let the domain **D** be a compact space of **N** dimensions, and let $\mathfrak{I}$ be a set of continuous real-valued function on **D**, satisfying the following criteria [24, 25]:

1. **Identity function:** The constant $f(x) = 1$ is in $\mathfrak{I}$.
2. **Separability:** For any two points $x_1 \neq x_2$ in ***D***, there is an ***f*** in $\mathfrak{I}$ such that $f(x_1) \neq f(x_2)$.
3. **Algebraic closure:** If ***f*** and ***g*** are any two functions in $\mathfrak{I}$, then ***fg*** and ***af + bg*** are in $\mathfrak{I}$ for any two real numbers ***a*** and ***b***.

Then $\mathfrak{I}$ is dense in ***C***(***D***), the set of continuous real-value functions on **D**. In other words, for any $\varepsilon > 0$, and any function ***g*** in ***C***(***D***), there is a function ***f*** in $\mathfrak{I}$ such that $|g(x) - f(x)| < \varepsilon$ for all $\mathbf{x} \in \mathbf{D}$.

3.2 Applying Stone-Weierstrass Theorem used on IT2FNN-2:A2C0 Architecture.

In the IT2FNN, the domain in which we operate is almost always compact. It is a standard result in real analysis that every closed and bounded set in $\Re^n$ is compact. Now we shall apply the Stone-Weierstrass theorem to show the representational power of IT2FNN with simplified fuzzy if-then rules.

3.2.1 Identity Function

The first requirement of Stone-Weierstrass theorem needs our IT2FNN to be able to compute the identity function f(x)=1. An obvious way to compute the function is to set the consequence part of each rule equal to [1-δ, 1+δ], where δ is the uncertainty spread. In fact, an IT2FNN with only one rule suffices to satisfy this requirement.

3.2.2 Separability

The second requirement of the Stone-Weierstrass theorem needs the IT2FNN to be able to compute functions that have different values for different points. Without this requirement, the trivial set of functions f: f(x)= [c-δ, c+δ] c $\in \Re$ would satisfy Stone-Weierstrass theorem. Separability is satisfied whenever IT2FNN can compute strictly monotonic functions of each input variable. This can easily be achieved by adjusting the membership functions of the premise part.

3.2.3 Algebraic Closure-Additive

The third requirement of Stone-Weierstrass theorem needs the IT2FNN to be able to approximate sums and products of functions. Suppose we have two IT2FNNs S and $\overline{S}$ each of which has two rules. The output of each system can be expressed as

$$S: z = \gamma \cdot y_l + (1-\gamma) y_r = \gamma \frac{\underline{f}^1 y^1 + \underline{f}^2 y^2}{\underline{f}^1 + \underline{f}^2} + (1-\gamma) \frac{\overline{f}^1 y^1 + \overline{f}^2 y^2}{\overline{f}^1 + \overline{f}^2}$$

$$\overline{S}: \overline{z} = \overline{\gamma} \cdot \overline{y}_l + (1-\overline{\gamma}) \overline{y}_r = \overline{\gamma} \frac{\underline{g}^1 \overline{y}^1 + \underline{g}^2 \overline{y}^2}{\underline{g}^1 + \underline{g}^2} + (1-\overline{\gamma}) \frac{\overline{g}^1 \overline{y}^1 + \overline{g}^2 \overline{y}^2}{\overline{g}^1 + \overline{g}^2}$$

then the sum of z and $\overline{z}$ is equal to

$$az + b\overline{z} = \frac{\underline{f}^1 \overline{f}^1 \underline{g}^1 \overline{g}^1}{fg}\left[a y^1 + b\overline{y}^1\right] + \frac{\underline{f}^1 \overline{f}^1 \underline{g}^1 \overline{g}^2}{fg}\left[a y^1 + b\left(\overline{\gamma}\overline{y}^1 + (1-\overline{\gamma})\overline{y}^2\right)\right] +$$

$$\frac{\underline{f}^1 \overline{f}^1 \overline{g}^1 \underline{g}^2}{fg}\left[a y^1 + b\left(\overline{\gamma}\overline{y}^2 + (1-\overline{\gamma})\overline{y}^1\right)\right] + \frac{\underline{f}^1 \overline{f}^1 \underline{g}^2 \overline{g}^2}{fg}\left[a y^1 + b\overline{y}^2\right] +$$

$$\frac{\underline{f}^1 \overline{f}^2 \underline{g}^1 \overline{g}^1}{fg}\left[a\left(\gamma y^1 + (1-\gamma)y^2\right) + b\overline{y}^1\right] + \frac{\underline{f}^1 \overline{f}^2 \underline{g}^1 \overline{g}^2}{fg}\left[a\left(\gamma y^1 + (1-\gamma)y^2\right) + b\left(\overline{\gamma}\overline{y}^1 + (1-\overline{\gamma})\overline{y}^2\right)\right] +$$

$$\frac{\underline{f}^1 \overline{f}^2 \overline{g}^1 \underline{g}^2}{fg}\left[a\left(\gamma y^1 + (1-\gamma)y^2\right) + b\left(\overline{\gamma}\overline{y}^2 + (1-\overline{\gamma})\overline{y}^1\right)\right] + \frac{\underline{f}^1 \overline{f}^2 \underline{g}^2 \overline{g}^2}{fg}\left[a\left(\gamma y^1 + (1-\gamma)y^2\right) + b\overline{y}^2\right] +$$

$$\frac{\overline{f}^1 \underline{f}^2 \underline{g}^1 \overline{g}^1}{fg}\left[a\left(\gamma y^2 + (1-\gamma)y^1\right) + b\overline{y}^1\right] + \frac{\overline{f}^1 \underline{f}^2 \underline{g}^1 \overline{g}^2}{fg}\left[a\left(\gamma y^2 + (1-\gamma)y^1\right) + b\left(\overline{\gamma}\overline{y}^1 + (1-\overline{\gamma})\overline{y}^2\right)\right] +$$

$$\frac{\overline{f}^1 \underline{f}^2 \overline{g}^1 \underline{g}^2}{fg}\left[a\left(\gamma y^2 + (1-\gamma)y^1\right) + b\left(\overline{\gamma}\overline{y}^2 + (1-\overline{\gamma})\overline{y}^1\right)\right] + \frac{\overline{f}^1 \underline{f}^2 \underline{g}^2 \overline{g}^2}{fg}\left[a\left(\gamma y^2 + (1-\gamma)y^1\right) + b\overline{y}^2\right] +$$

$$\frac{\underline{f}^2 \overline{f}^2 \underline{g}^1 \overline{g}^1}{fg}\left[a y^2 + b\overline{y}^1\right] + \frac{\underline{f}^2 \overline{f}^2 \underline{g}^1 \overline{g}^2}{fg}\left[a y^2 + b\left(\overline{\gamma}\overline{y}^1 + (1-\overline{\gamma})\overline{y}^2\right)\right] +$$

$$\frac{\underline{f}^2 \overline{f}^2 \overline{g}^1 \underline{g}^2}{fg}\left[a y^2 + b\left(\overline{\gamma}\overline{y}^2 + (1-\overline{\gamma})\overline{y}^1\right)\right] + \frac{\underline{f}^2 \overline{f}^2 \underline{g}^2 \overline{g}^2}{fg}\left[a y^2 + b\overline{y}^2\right]$$

therefore we can construct an IT2FNN that computes $az + b\overline{z}$.

3.2.4 Algebraic Closure-Multiplicative

Modeling the product of $z\overline{z}$ of two IT2FNNs is the last capability we must demonstrate before we can conclude that the Stone-Weierstrass theorem can be applied to the proposed reasoning mechanism. The product $z\overline{z}$ can be expressed as

$$z\overline{z} = \frac{\underline{f}^1 \underline{g}^1 \left(\overline{f}^1 + \overline{f}^2\right)\left(\overline{g}^1 + \overline{g}^2\right)}{fg} \gamma y^1 \overline{\gamma}\overline{y}^1 + \frac{\underline{f}^1 \underline{g}^2 \left(\overline{f}^1 + \overline{f}^2\right)\left(\overline{g}^1 + \overline{g}^2\right)}{fg} \gamma y^1 \overline{\gamma}\overline{y}^2 +$$

$$\frac{\underline{f}^2 \underline{g}^1 \left(\overline{f}^1 + \overline{f}^2\right)\left(\overline{g}^1 + \overline{g}^2\right)}{fg} \gamma y^2 \overline{\gamma}\overline{y}^1 + \frac{\underline{f}^2 \underline{g}^2 \left(\overline{f}^1 + \overline{f}^2\right)\left(\overline{g}^1 + \overline{g}^2\right)}{fg} \gamma y^2 \overline{\gamma}\overline{y}^2 +$$

$$\frac{\underline{f}^1 \overline{g}^1 \left(\overline{f}^1 + \overline{f}^2\right)\left(\underline{g}^1 + \underline{g}^2\right)}{fg} \gamma y^1 (1-\overline{\gamma})\overline{y}^1 + \frac{\underline{f}^1 \overline{g}^2 \left(\overline{f}^1 + \overline{f}^2\right)\left(\underline{g}^1 + \underline{g}^2\right)}{fg} \gamma y^1 (1-\overline{\gamma})\overline{y}^2 +$$

ISBN: 978-989-95079-6-8

$$\frac{\underline{f}^2\underline{g}^1\left(\overline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}\gamma y^2(1-\bar{\gamma})\bar{y}^1+\frac{\underline{f}^2\underline{g}^2\left(\overline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}\gamma y^2(1-\bar{\gamma})\bar{y}^2+$$

$$\frac{\overline{f}^1\underline{g}^1\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}{fg}(1-\gamma)y^1\bar{\gamma}\bar{y}^1+\frac{\overline{f}^1\underline{g}^2\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}{fg}(1-\gamma)y^1\bar{\gamma}\bar{y}^2+$$

$$\frac{\overline{f}^2\underline{g}^1\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}{fg}(1-\gamma)y^2\bar{\gamma}\bar{y}^1+\frac{\overline{f}^2\underline{g}^2\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}{fg}(1-\gamma)y^2\bar{\gamma}\bar{y}^2+$$

$$\frac{\overline{f}^1\overline{g}^1\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}(1-\gamma)y^1(1-\bar{\gamma})\bar{y}^1+\frac{\overline{f}^1\overline{g}^2\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}(1-\gamma)y^1(1-\bar{\gamma})\bar{y}^2+$$

$$\frac{\overline{f}^2\overline{g}^1\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}(1-\gamma)y^2(1-\bar{\gamma})\bar{y}^1+\frac{\overline{f}^2\overline{g}^2\left(\underline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}{fg}(1-\gamma)y^2(1-\bar{\gamma})\bar{y}^2$$

therefore we can construct an IT2FNN that computes $z\bar{z}$.

3.3 Applying Stone-Weierstrass theorem to the IT2FNN-2:A2C1 architecture

As in the case of IT2FNN-2:A2C0, let there be two distinct IT2FNN-2:A2C1, S and $\overline{S}$, each with two rules; we can proceed in this case as follows:

3.3.1 Algebraic Closure-Additive

$$S: z = \frac{1}{2}\left[\frac{\overline{f}^1 y_l^1+\underline{f}^2 y_l^2}{\overline{f}^1+\underline{f}^2}+\frac{\underline{f}^1 y_r^1+\overline{f}^2 y_r^2}{\underline{f}^1+\overline{f}^2}\right]$$

$$\overline{S}: \bar{z} = \frac{1}{2}\left[\frac{\overline{g}^1 \bar{y}_l^1+\underline{g}^2 \bar{y}_l^2}{\overline{g}^1+\underline{g}^2}+\frac{\underline{g}^1 \bar{y}_r^1+\overline{g}^2 \bar{y}_r^2}{\underline{g}^1+\overline{g}^2}\right]$$

then the sum of z and $\bar{z}$ is equal to

$$az+b\bar{z}=\frac{1}{2}\frac{\overline{f}^1\overline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}\left(ay_l^1+b\bar{y}_l^1\right)+\frac{1}{2}\frac{\overline{f}^1\underline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}\left(ay_l^1+b\bar{y}_l^2\right)+$$

$$\frac{1}{2}\frac{\underline{f}^2\overline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}\left(ay_l^2+b\bar{y}_l^1\right)+\frac{1}{2}\frac{\underline{f}^2\underline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}\left(ay_l^2+b\bar{y}_l^2\right)+$$

$$\frac{1}{2}\frac{\underline{f}^1\underline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}\left(ay_r^1+b\bar{y}_r^1\right)+\frac{1}{2}\frac{\underline{f}^1\overline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}\left(ay_r^1+b\bar{y}_r^2\right)+$$

$$\frac{1}{2}\frac{\overline{f}^2\underline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}\left(ay_r^2+b\bar{y}_r^1\right)+\frac{1}{2}\frac{\overline{f}^2\overline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}\left(ay_r^2+b\bar{y}_r^2\right)$$

therefore we can construct an IT2FNN that computes $az+b\bar{z}$.

3.3.2 Algebraic Closure-Multiplicative

The product $z\bar{z}$ can be expressed as

$$z\bar{z}=\frac{1}{4}\frac{\overline{f}^1\overline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}y_l^1\bar{y}_l^1+\frac{1}{4}\frac{\overline{f}^1\underline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}y_l^1\bar{y}_l^2+$$

$$\frac{1}{4}\frac{\underline{f}^2\overline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}y_l^2\bar{y}_l^1+\frac{1}{4}\frac{\underline{f}^2\underline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\overline{g}^1+\underline{g}^2\right)}y_l^2\bar{y}_l^2+$$

$$\frac{1}{4}\frac{\overline{f}^1\underline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_l^1\bar{y}_r^1+\frac{1}{4}\frac{\overline{f}^1\overline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_l^1\bar{y}_r^2+$$

$$\frac{1}{4}\frac{\underline{f}^2\underline{g}^1}{\left(\overline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_l^2\bar{y}_r^1+\frac{1}{4}\frac{\underline{f}^2\overline{g}^2}{\left(\overline{f}^1+\underline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_l^2\bar{y}_r^2+$$

$$\frac{1}{4}\frac{\underline{f}^1\overline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}y_r^1\bar{y}_l^1+\frac{1}{4}\frac{\underline{f}^1\underline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}y_r^1\bar{y}_l^2+$$

$$\frac{1}{4}\frac{\overline{f}^2\overline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}y_r^2\bar{y}_l^1+\frac{1}{4}\frac{\overline{f}^2\underline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\underline{g}^2\right)}y_r^2\bar{y}_l^2+$$

$$\frac{1}{4}\frac{\underline{f}^1\underline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_r^1\bar{y}_r^1+\frac{1}{4}\frac{\underline{f}^1\overline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_r^1\bar{y}_r^2+$$

$$\frac{1}{4}\frac{\overline{f}^2\underline{g}^1}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_r^2\bar{y}_r^1+\frac{1}{4}\frac{\overline{f}^2\overline{g}^2}{\left(\underline{f}^1+\overline{f}^2\right)\left(\underline{g}^1+\overline{g}^2\right)}y_r^2\bar{y}_r^2$$

therefore we can construct an IT2FNN that computes $z\bar{z}$.

The IT2FNN architectures that compute $az+b\bar{z}$ and $z\bar{z}$ have the same S and $\overline{S}$ if and only if the class of membership functions is invariant under multiplication. This is loosely true if the class of interval type-2 membership functions is the set of all bell-shaped functions and Gaussian-shaped (igbellmtype2, igbellstype2 and igbelltype2 for bell-shaped and igaussmtype2, igaussstype2 and igausstype2 for Gaussian-shaped, all of which can be found in the Matlab's Interval Type-2 Fuzzy Logic Toolbox [26]).

Therefore by choosing an appropriate class of membership functions, we can conclude that the IT2FNN with simplified fuzzy if-then rules satisfy the four criteria of Stone-Weierstrass theorem. Consequently, for any given $\varepsilon > 0$ and any real-valued function g, there is an IT2FNN S such that $|g(x)-S(x)|<\varepsilon$ for all x in underlying compact set. Moreover since the simplified IT2FNN is a proper subset of two types of architecture, we can conclude that the IT2FNN architecture has unlimited approximation power to match any given data set.

4 Application Examples

Two application examples are used to illustrate the proofs of universality, as follows.

Experiment 1. Identification of a one variable non-linear function.

In this experiment we approximate a non-linear function $f:\Re\to\Re$:

ISBN: 978-989-95079-6-8

$$f(u)=0.6\sin(\pi u)+0.3\sin(3\pi u)+0.1\sin(5\pi u)+\eta$$

(where η is a uniform noise component) using a one input-one output IT2FNN, 50 training data sets with 10 fold cross-validation with uniform noise levels, 6 IT2MF type igaussmtype2, 6 rules and 50 epochs. Once ANFIS and IT2FNN models are identified a comparison was made, taking into account RMSE statistic values with 10 fold cross-validation. Table 1 and Figure 3 show resulting RMSE values for ANFIS and IT2FNN, it can be seen that IT2FNN architectures [12] perform better than ANFIS.

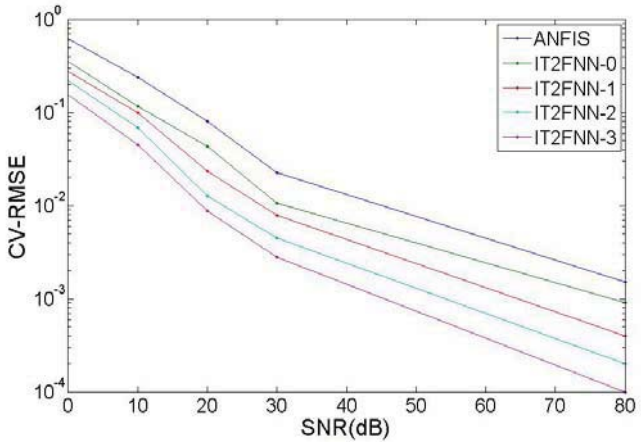


Figure 3: RMSE values of ANFIS and IT2FNN using 10-fold cross-validation for identifying non-linearity in Experiment 1.

Table 1: RMSE values of ANFIS and IT2FNN with 10-fold cross-validation for identifying non-linearity of Experiment 1.

SNR(dB)	ANFIS	IT2FNN-0	IT2FNN-1	IT2FNN-2	IT2FNN-3
0	0.6156	0.3532	0.2764	0.2197	0.1535
10	0.2375	0.1153	0.0986	0.0683	0.0453
20	0.0806	0.0435	0.0234	0.0127	0.0087
30	0.0225	0.0106	0.0079	0.0045	0.0028
free	0.0015	0.0009	0.0004	0.0002	0.0001

Experiment 2. Identification of a three variable non-linear function.

A three input – one output IT2FNN is used to approximate non-linear Sugeno [9] function $f:\Re^3\rightarrow\Re$:

$$f(x_1,x_2,x_3)=\left(1+\sqrt{x_1}+\frac{1}{x_2}+\frac{1}{\sqrt{x_3^3}}\right)^2+\eta$$

216 training data sets are generated with 10-fold cross-validation and 125 for tests; 2 igaussmtype2 IT2MF for each input, 8 rules and 50 epochs. Once the ANFIS and IT2FNN models are identified, a comparison is made with RMSE statistic values and 10-fold cross-validation. Table 2 and Figure 4 show the resultant RMSE values for ANFIS and IT2FNN. It can be seen that IT2FNN architectures [12] perform better than ANFIS.

Table 2: Resulting RMSE values in ANFIS and IT2FNN for non-linearity identification in Experiment 2 with 10-fold cross-validation.

SNR(dB)	ANFIS	IT2FNN-0	IT2FNN-1	IT2FNN-2	IT2FNN-3
0	1.0432	0.7203	0.6523	0.5512	0.5267
10	0.3096	0.2798	0.2583	0.2464	0.2344
20	0.1703	0.1637	0.1592	0.1465	0.1387
30	0.1526	0.1408	0.1368	0.1323	0.1312
free	0.1503	0.1390	0.1323	0.1304	0.1276

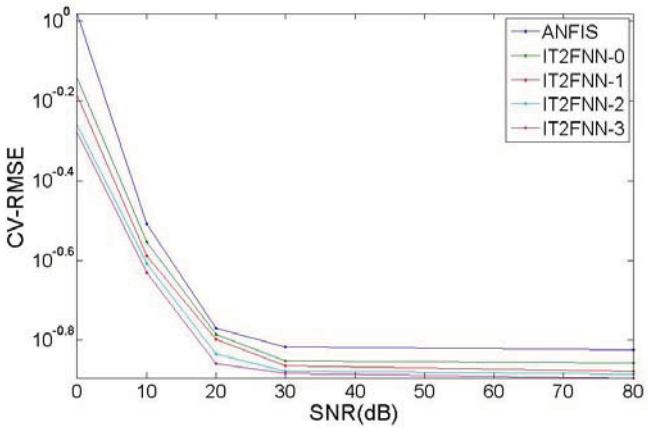


Figure 4: Resulting RMSE values obtained by ANFIS and IT2FNN for non-linearity identification in Experiment 2 with 10-fold cross-validation.

5 Conclusions

In these experiments, the estimated RMSE values for non-linear function identification with 10-fold cross-validation for the hybrid architectures IT2FNN-2:A2C0 and IT2FNN-2:A2C1 illustrate the proof based on Stone-Weierstrass theorem that they are universal approximators for efficient identification of non-linear functions. Also, it can be seen that while increasing the Signal Noise Ratio (SNR), IT2FNN architectures handle uncertainty more efficiently.

Acknowledgment

The authors would like to thank CONACYT and DGEST for the financial support given for this research project. The student (Juan R. Castro) is supported by a scholarship from UABC-CONACYT.

References

[1] K. Hornick, M., Stinchcombe and H. White, Multilayer feedforward networks are universal approximators. Neural Networks, 2(5):359-366, 1989.

ISBN: 978-989-95079-6-8

[2] L. Wang, and J. Mendel, Fuzzy basis functions, universal approximation, and orthogonal least squares learning. IEEE Trans. Neural Networks, 3:807-813, 1992.

[3] L. Wang and J. Mendel, Generating fuzzy rules by learning from examples. IEEE Trans. Syst. Man and Cyber., 22(6):1414-1427, 1992.

[4] B. Kosko, Neural Networks and Fuzzy Systems. Englewood Cliffs, NJ, Prentice Hall, 1992.

[5] B. Kosko, Fuzzy systems as universal approximators. IEEE Trans. Computers, 43(1):1329-1333, 1994.

[6] J. Buckley, Universal fuzzy controllers. Automatica, 28(6):1245-1248, 1992.

[7] J. Buckley, Sugeno type controllers are universal controllers. Automatica, 53:299-304, 1993.

[8] J. Buckley and Y. Hayashi, Can fuzzy neural networks approximate continuos fuzzy functions?, Fuzzy Sets and Systems, 61:43-52, 1994.

[9] T. Takagi and M. Sugeno, Fuzzy Identification of Systems and its applications to modelling and Control, *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1): 116-132, Jan-Feb 1985.

[10] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338-353, 1965.

[11] K. Hirota and W. Pedrycz, OR/AND neuron in modeling fuzzy set connectives, IEEE Transactions on Fuzzy Systems, 2(994) 151-161.

[12] J.R. Castro, O. Castillo, P. Melin, A. Rodriguez-Diaz, A Hybrid Learning Algorithm for Interval Type-2 Fuzzy Neural Networks: The Case of Time Series Prediction; In Soft Computing for Hybrid Intelligent Systems. Book Chapter, Volume 154/2008: 363-386, Springer Berlin / Heidelberg. ISBN: 978-3-540-70811-7, ISSN: 1860-949X (Print) 1860-9503 (Online), September 2008.

[13] J.-S.R. Jang, "ANFIS: adaptive-network-based fuzzy inference system". IEEE Trans. Syst., Man, Cybern., vol. 23, no. 3, pp. 665-684, 1993.

[14] J.-S.R. Jang, C.T. Sun, and E. Mizutani, Neuro-fuzzy and Soft Computing. New York: Prentice-Hall, 1997.

[15] Q. Liang and J.M. Mendel, "Interval type-2 logic systems: theory and design," IEEE Transactions on Fuzzy Systems, vol. 8 no. 5, pp. 535-550, Oct. 2000.

[16] J.M. Mendel, Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions, Prentice Hall, NJ. 2001.

[17] Lee, C-H., Hu, T-W, Lee, C-T, Lee, Y-C, "A Recurrent Interval Type-2 Fuzzy Neural Network with Asymmetric Membership Functions for Nonlinear System Identification". In FUZZ-IEEE 2008, Hong Kong, June 2008.

[18] J.R. Castro, O. Castillo, P. Melin, A. Rodriguez-Diaz, Hybrid learning algorithm for interval type-2 fuzzy neural networks, In Proceedings of Granular Computing 2007, pages 157-162, Silicon Valley, CA, USA, ISBN: 0-7695-3032-X, November 2007.

[19] C.H. Lee, J.L. Hong, Y.C. Lin, and W.Y. Lai, "Type-2 Fuzzy Neural Network Systems and Learning," International Journal of Computational Cognition, Vol. 1, No. 4, pp. 79-90, 2003.

[20] C.H. Wang, C.S. Cheng, and Tsu-Tian Lee, "Dynamical optimal training for interval type-2 fuzzy neural network (T2FNN)," IEEE Trans. on Systems, Man, and Cybernetics Part B: Cybernetics, Vol. 34, No. 3, pp. 1462-1477, 2004.

[21] L.V. Kantorovich and G.P. Akilov. Functional analysis. Pergamon, Oxford, 2nd edition, 1982.

[22] H.L. Royden. Real analysis. Macmillan, New York, 2nd edition 1968.

[23] V. Kreinovich, G.C. Mouzouris and H.T. Nguyen, "Fuzzy Rule Based Modelling as a Universal Aproximation Tool," in Fuzzy Systems, Modeling and Control (H. T. Nguyen and M. Sugeno, Eds.), pp. 135-195, Kluwer Ac. Publ., Boston, 1998.

[24] C.T. Scarborough, A.H. Stone, Products of nearly compact spaces, Transactions of the American Mathematical Society 124: 131-147, 1966.

[25] R.E. Edwards, Functional Analysis: Theory and Applications, Dover Publications, 1995.

[26] J.R. Castro, O. Castillo, P. Melin, A. Rodriguez-Diaz, Building fuzzy inference systems with a new interval type-2 fuzzy logic toolbox, In Transactions on Computational Science. Book Chapter, Volume 4750:104-114, Springer-Verlag. ISBN: 978-3-540-79298-7, ISSN: 0302-9743 (Print) 1611-3349 (Online), June 2008.

ISBN: 978-989-95079-6-8

Spectral Learning with Type-2 Fuzzy Numbers for Question/Answering System

Asli Celikyilmaz[1] I. Burhan Turksen[2]

1. Computer Sciences Division, University in California, Berkeley, CA, USA
2. TOBB Economy and Technology University, Ankara, Turkey & University of Toronto, Canada

Abstract— *Graph-based semi-supervised learning has recently emerged as a promising approach to data-sparse learning problems in natural language processing. They rely on graphs that jointly represent each data point. The problem of how to best formulate the graph representation remains an open research topic. In this paper, we introduce a type-2 fuzzy arithmetic to characterize the edge weights of a formed graph as type-2 fuzzy numbers. The fuzzy numbers are identified by the changing parameters of the fuzzy kernel nearest neighbor algorithm, namely the degree of fuzziness and the hyper-parameter of the Gaussian kernel function, both of which have an effect on the uncertainty in forming the affinity matrix of the graph. We introduce a new graph-based semi-supervised learning with the type-2 arithmetic operations. We apply this technique in the framework of label propagation and evaluate on a question answering task. We demonstrate that the type-2 SSL can improve the prediction accuracy and can be considered to be the an alternative tool for text mining applications of computational linguistics.*

Keywords— Graph-based semi-supervised learning, kernel fuzzy k-nearest neighbor, type-2 fuzzy numbers.

1 Introduction and Motivation

In building reliable models for real systems, identification of exact values of variables of model equations are required. In real life practices, precise values of parameters may not be obtained due to imprecise, noisy, vague, or incomplete nature of information. Fuzzy logic provides explanatory tools for such tasks, mainly because of its capability to manage imprecise categories to represent imperfect information, by means of fuzzy sets, graduality, measures of resemblance or aggregation methods. Type-1 fuzzy sets may not be enough to explain the whole spectrum of possible results, mainly because the values used to characterize the membership functions of type-1 fuzzy numbers are usually overly precise. Usually the level of information is improperly set to define membership functions, thus it is rather necessary to use type-2 fuzzy numbers to represent uncertainties in model parameters.

In this work, we mainly focus on uncertainties in finding similarities in text mining. We consider one of the most commonly used learning methods, namely the semi-supervised learning (SSL) method [1]. It is often the case in the areas of machine learning for classification problems such as the problem of text classification on web pages, automatic translation or online question/ answering systems, etc. that one needs to deal with a very small portion of labeled data and vast amounts of unlabeled data. For such cases, graph-based SSL methods (spectral learning methods) have proved to outperform other learning methods. In graph- based methods the data is represented by the nodes of a graph (Fig. 1), the edges of which are labeled with the pairwise distance of the incident nodes. One problem with spectral learning methods is that the procedure is highly sensitive to the choice of the kernel, for example it is very sensitive to the choice of the spread (variance) of a Gaussian kernel, which naturally effects the similarity matrix defined for the given dataset.

As in the phrase of "words can mean different things to different people", an entailment relation between a candidate sentence and a question posed by the user may be evaluated differently by different people. For instance, a different degree of entailment may be assigned by different people for pairs of question "Who bought Overture?" and candidate sentences such as "Yahoo bought Overture", "Yahoo owns Overture", "Overture acquisition by Yahoo", using linguistic terms such $strict$, $loose$, or $direct$ entailment. Current methods can only use crisp values to define such relations, which cannot be explained to a full extent. Type-2 fuzzy logic is the best fit to define the entailment relations between each sentence. To our knowledge, characterization of edge weights of a graph as type-2 fuzzy numbers, as presented in this paper, is a new approach. The novel type-2 SSL defines such uncertain entailment relation between two sentences by characterizing soft linked graphs.

In this paper we concentrate on characterization of the uncertainties in similarity measure when discovering knowledge from unstructured text using graph-based SSL algorithm. A common way to construct the affinity matrix of a graph is by application of a nearest neighbor method. We use a fuzzy k-nearest neighbor (FKNN) to allow fuzzy decisions based on fuzzy labels. In addition we use its kernel extension [2] to enable solving possible non-linearly separable problems and get non-linear fuzzy boundaries instead of linear boundaries when necessary. In addition kernel methods have proven to prevent over-fitting in high dimensional feature spaces. For these reasons, we consider applying type-2 fuzzy arithmetic to situations where the similarity between two objects, i.e, two sentences, is imprecise.

Thus, the novel type-2 SSL method learns the edge link weights via kernel fuzzy k-nearest neighbor algorithm [2] (KFKNN). We use the arithmetic operations on type-2 fuzzy numbers defined in [3]. For ease of calculations, we graduate the interval valued degree of fuzziness and the kernel parameter and obtain bounded discrete valued weights (interval valued) with associated type-2 membership grades, enabling to represent each weight link with a type-2 fuzzy number. In a way each membership value is further stretched [1] based on fuzziness of the model to capture uncertainty interval of membership values (Fig. 2). Using the interval type-2 fuzzy

[1] Zadeh[4] defines the membership values as elastic constraints that has to be stretched to get their full meaning.

ISBN: 978-989-95079-6-8

weights, we construct fuzzy graph Laplacian. The novel type-2 SSL uses label propagation algorithm [1] to obtain interval valued output values. Defuzzification follows to measure the performance of the new classifier.

In this paper we mainly focus on the application of the novel spectral classification method on a specific area of natural language processing. Text-based question answering (QA) is the process of automatically finding the answers to arbitrary questions in plain English by searching collection of text files. Recently intensive research has been continuing to grow in this area fostered by evaluation-based conferences, one of which is the Text REtrieval Conference (TREC) [5]. Among different question types, most current research focuses on the $factoid$ type questions, e.g., "*How tall is Eiffel Tower*", where the answer is short string indicating a fact with a named entity. We construct a textual entailment model to find the degree of match between the question posed by the user and the candidate answers retrieved by the search engine and then rank the sentences likely to contain the correct answer at the top.

We will briefly review graph-based SSL methods, KFKNN, and type-2 fuzzy numbers. Then we will present the novel type-2 SSL method followed by the benchmark analysis on the TREC dataset. Finally conclusions are drawn.

2 Graph Based Semi-Supervised Learning

We begin with notation and a brief summary of SSL algorithms [1], [6]. Let $X = \{x_1, ..., x_n\}$ represent n data points in $\Re^d$ and $Y = \{y_1, ..., y_n\}$ be their output targets. The labeled part of X is represented with $X_l = \{x_1, ..., x_l\}$ with associated labels $Y_l = \{y_1, ..., y_l\}^T$. For ease of presentation, we concentrate on binary classification, where y_i can take on either of $\{-1, +1\}$. X has also unlabeled part, i.e., $X = X_u \cup X_l$. The aim is to predict labels for the remaining unlabeled points, $X_u = \{x_1, ..., x_u\}$.

In this section we review the most general graph-based semi-supervised learning method for binary classification problem. A graph is denoted with $g = (V, E)$, where $V = X_l \cup X_u$ is set of nodes. E represents the edges connecting two vertexes and is represented by the $n \times n$ symmetric affinity matrix W, where an edge $W(x_i, x_j) = e(i, j)$ represents the similarity between any vertexes x_i and x_j and 0 if there is no similarity between them. The most common similarity measure is the Gaussian kernel of width σ,

$$W_{ij} = K(x_i, x_j) = e^{-\|x_i - x_j\|^2/2\sigma^2} \tag{1}$$

Earlier research [1] suggests that the choice of σ can strongly influence the results (it is one of the imprecise parameters we use to characterize uncertainty interval of membership functions in our approach.) The diagonal degree matrix D is defined for g by D=$\sum_j W_{ij}$.

In graph-based SSL, a function over the graph is estimated such that it satisfies two conditions: 1) it should be close to the given labels (L), and 2) it should be smooth (S) on the whole graph. These two conditions are presented in a regularization form to minimize the following basic algorithm. Let **f**=$[f_1, ..., f_l, f_{l+1}, ..., f_n]$ denote the predicted labels of X where $f \in \Re^n$. According to the mathematical representation of graph based approach, the objective is to minimize

$$\arg min_f \; E_L(\mathbf{f}) + \lambda E_S(\mathbf{f}). \tag{2}$$

In (2) $E_L(\mathbf{f})$ is a loss function to penalize the deviation from the given labels,

$$E_L(\mathbf{f}) = \sum_{i \subset L} (f_i - y_i)^2 \tag{3}$$

and $E_S(\mathbf{f})$ is a regularizer to represent the label smoothness,

$$E_S(\mathbf{f}) = \frac{1}{2} \sum_{i,j \in L \cup U} W_{ij}(f_i - f_j)^2 = f^T \mathbf{L} f \tag{4}$$

In (4), $L = D - W$ is the graph Laplacian. To satisfy the local and global consistency [1], the normalized combinatorial Laplacian is used such that the $E_S(\mathbf{f})$ is replaced with normalized Laplacian, $\mathcal{L} = I - D^{-1/2} L D^{-1/2}$, as follows:

$$E_S(\mathbf{f}) = \sum_{i,j \in L \cup U} W_{ij}(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}})^2 = f^T \mathcal{L} f \tag{5}$$

The classification function **f** is learned through any of the $label\ propagation$(LP) algorithms, e.g., [1], [7] on the graph. From consistency approach [1] $f*$ can be found as :

$$f^* = (I - \lambda \left(D^{-1/2} W D^{-1/2} \right)^{-1} Y \tag{6}$$

Nearest neighbors are most commonly used method to identify the affinity matrix of the graph. In this work, we use the KFKNN method by capturing the uncertainty interval of similarities via perturbations on learning parameters.

3 Kernel Fuzzy k-Nearest Neighbor

The k-nearest neighbor approach (KNN) is commonly used in spectral learning to obtain the link weights of the graph Laplacian. It does not depend on the distribution of selected k objects, whereas fuzzy KNN [8] deals with the distribution of the selected k objects to determine the sum of similarities between labeled data points. They both use Euclidean distance function. To handle with more complex real systems, instead of Euclidean distance, we use kernel function to calculate distance between two objects. Through some non-linear mapping the input data is mapped onto a higher-dimensional feature space, i.e., $X \rightarrow \phi(X)$ to conduct the fuzzy KNN. With the kernel trick, instead of mapping the data point and calculating the distance in the feature space, we can use the kernel function and compute the dot product in some feature space, $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. The distances are calculated in this space as follows:

$$\|\phi(x) - \phi(x_j)\|^2 = 2 - 2K(x, x_j) \tag{7}$$

Gaussian kernel functions as in (1) is used. Let $C_i,\ i = 1, ...c$ represent each class c of distinct labels. The FKNN algorithm [8] assigns memberships as a function of vector distance from their nearest neighbors and those neighbor memberships in the possible classes. The membership values of x_i^u in each labeled class $c = 1, ..., k$ is calculated with

$$\mu_i(x^u) = \frac{\sum_{j=1}^k \mu_{ij} \left(1/\left(1 - K(x - x_j)^{1/m-1}\right) \right.}{\sum_{j=1}^k \left(1/\left(1 - K(x - x_j)^{1/m-1}\right) \right.} \tag{8}$$

ISBN: 978-989-95079-6-8

Algorithm 1 Kernel Fuzzy κ-Nearest Neighbor

procedure KFKNN(k, X^l, X^u, m, σ)
  for $j \leftarrow 1, n_u$, $x_j^u \in X^u$ **do**
    $w_{x_j^u, x_l^l} \leftarrow K(x_j, x_l^l)$ and σ using (1).
    Select k nearest neighbors among all labeled objects, $N_j = \left\{x_{1,j}^l, ...x_{k,j}^l\right\}$
    Calculate the membership values of x_j^u, $\mu_i(x_j^u)$, in each class $i = 1, ..., c$ using N_j data points via (8)
    If required, classify x_j^u into any of C_i using criteria in (9)
  end for
end procedure

The initial class memberships of labeled points are calculated using the function in [2]. Class separation criteria is based on;

$$\hat{y}^u = \arg max_{i=1...c} \ \mu_i(x) \tag{9}$$

The input parameters of KFKNN are the k-nearest neighbors, the labeled X^l data points, and unlabeled X^u data points, the degree of fuzziness parameter, m and kernel distance variance parameter σ in (1). The parameter m determines how heavily the distance is weighted when calculating each neighbor contribution to the membership value. It has been recently shown [9], [10], [11] that changing the m parameter in fuzzy clustering methods [12] such as fuzzy c-means and fuzzy c-regression have an effect on the outcome of the partition matrix, as well as the classification results. With some perturbations on the fuzziness parameter, we can identify uncertainty interval of membership values distributions. In this work, we identify the parameter m as an interval valued parameter, $m := [m^l, m^u]$, $l : lower$, $u : upper$. Next we graduate the interval manually into crisp values, $m^r = \left\{m^1, ..., m^{nr}\right\}$. We do the same discretization for the Gaussian kernel hyperparameter, $\sigma^s = \left\{\sigma^1, ..., \sigma^{ns}\right\}$. Each pair of parameter corresponds to one discrete affinity matrix obtained from KFKNN. We obtain the degree of membership values of relations, degree of similarity of an edge, via fuzzy $max - min$ rule, to be explained in section 5.

4 Type-2 Fuzzy Numbers and Operations

Before we present the novel type-2 SSL method, we give a brief review of type-2 fuzzy numbers in the next. Introduced by Zadeh [13] in the trilogy of papers, a type-2 fuzzy set $\tilde{A}$ is characterized in [14] with a type-2 membership function, $\mu_{\tilde{A}}(x, u)$, which maps elements $x \subset X$ to their membership values $u \subset J_x \subseteq [0, 1]$ as follows:

$$\tilde{A} = \left\{\left((x, u), \mu_{\tilde{A}}(x, u)\right) | \forall x \subset X, \forall u \subset J_x \subseteq [0, 1]\right\} \tag{10}$$

where J_x is a type-1 fuzzy set, i.e., primary membership value, and $0 \leq \mu_{\tilde{A}}(x, u) \leq 1$ is called secondary membership grade. Any type-2 fuzzy set can be represented as a collection of embedded type-2 fuzzy sets [14], as follows:

$$\tilde{A}_e = \sum_{i=1}^{N} \left[f_{x_i}(u_i)/u_i\right]/x_i; u_i \subset J_{x_i} \subseteq [0, 1] \tag{11}$$

where u_i is the primary and $f_{x_i}(u_i)$ is the secondary membership grade. The total number of type-2 sets embedded in $\tilde{A}$ is given by $n_e = \prod_{i=1}^{N} M_i$, where M_i is the cardinality of each primary membership grade on $x_i \in X$. When X has a numeric domain, $\tilde{A}$ can be characterized as fuzzy number. Thus for $\tilde{A}_e^j = \left\{\left(u_i^j, f_{x_i}(u_i^j)\right), j = 1..N\right\}$ and $u_i^j \in \{u_{ik}, k = 1, ..., M\}$, we can define a fuzzy number with

$$\tilde{A} = \sum_{j=1}^{n_e} A_e^j \tag{12}$$

Zadeh [13] introduced the union and intersection of type-2 fuzzy sets. Mendel and John [14] then defined the type-2 operations with embedded type-2 fuzzy sets. The union of discrete [1] type-2 fuzzy sets $\tilde{A}$ and $\tilde{B}$ is given as:

$$\tilde{A} \cup \tilde{B} = \sum_{j=1}^{n_A} \tilde{A}_e^j \sum_{i=1}^{n_B} \tilde{B}_e^i = \sum_{j=1}^{n_A} \sum_{i=1}^{n_B} \tilde{A}_e^j \cup \tilde{B}_e^i \tag{13}$$

The union of two type-2 embedded sets in (13) is defined as:

$$\tilde{A}_e^j \cup \tilde{B}_e^i = \bigcup_{k=1}^{N} \left[F_{x_k}\left(u_k^j, w_k^j\right) / \left(u_k^j \vee w_k^i\right)\right] / x_k \tag{14}$$

where $F_{x_k}\left(u_k^j, w_k^j\right)$ is the flag that is computed by the t-norm of the secondary membership grades of any kth embedded set, $\mu_{A_e^j}\left(x_k, f_{x_k}\left(u_k^j\right)\right)$ and $\mu_{B_e^i}\left(x_k, f_{x_k}\left(w_k^i\right)\right)$ [3]. When the secondary membership grades are unity, i.e., interval type-2 fuzzy sets, the union of type-2 embedded sets are defined as:

$$\tilde{A}_e^j \cup \tilde{B}_e^i = \bigcup_{k=1}^{N} \left[1 / \left(u_k^j \vee w_k^i\right)\right] / x_k \tag{15}$$

In [15] the arithmetic operations with type-2 fuzzy numbers such as addition of $\tilde{about}$ 2 and $\tilde{about}$ 5 is defined based on (14) in series of steps. Later in [3], computationally less expensive arithmetic operations, e.g., addition and multiplication with type-2 fuzzy numbers, is defined, which will be used in this paper to define the similarity between the sentences to construct the type-2 affinity matrix. with unity secondary membership grades.

Here we will only give a brief description of addition between two interval type-2 fuzzy numbers,

$$\tilde{about}\ 3 = \{0.3, 0.6, 0.7\}/2, \{1\}/3, \{0.3, 0.6, 0.7\}/4$$
$$\tilde{about}\ 5 = \{0.6, 0.7\}/4, \{1\}/5, \{0.6, 0.7\}/6$$

Since the secondary membership grades are unity, we obtain for each discrete primary membership value:

$$\mu_{about A_e^j + \tilde{about} B_e^i}(z) = sup\left(u_x^j \wedge w_y^i\right) \tag{16}$$

where $z = x \otimes y$ is performed for any discrete number defined in $\tilde{about}\ A$ and $\tilde{about}\ B$, on all embedded numbers $u_x^j = \mu_{about\ A_e^j}$, $j = 1..n_A$ and $w_x^i = \mu_{about\ B_e^i}$, $i = 1..n_B$. In (16), $\wedge$ is t-norm(minimum). Then the union of all the pairs are captured along domains of $z \in Z = X \otimes Y$. For potential duplicate values from the resultant of the $supremum$ operation in (16), the best practice is to select pair with the highest

[1] In this paper we deal with discrete fuzzy sets as naturally the fuzzy sets are discretized to do inference.

ISBN: 978-989-95079-6-8

membership value to form the resultant interval type-2 fuzzy number. From the addition operation on two type-2 fuzzy numbers, the resulting type-2 fuzzy number would yield:

$$\tilde{about}\ 8 = \{0.3, 0.6, 0.7\}/6, \{0.6, 0.7\}/7, \{1\}/8, \{0.6, 0.7\}/9, \{0.3, 0.6, 0.7\}/10$$

5 Novel Type-2 Semi-Supervised Learning

5.1 Motivation

In graph based SSL methods, one of the main problems is to construct the graph. Typically, k-nearest neighbor of each data point is identified and then each node is connected to their k nearest neighbors according to the edge weights using a Gaussian distance function as in (1). In cases where it is difficult to identify numerical characterization of similarities between two data points, one could identify imprecise similarities as type-2 fuzzy numbers (as explained in the previous section) and characterize these fuzzy numbers based on small perturbations on the learning parameters. Applications of Type-2 fuzzy modeling tools, e.g., [16], [17], [18] has shown to be effective methods for such cases.

In this work, we mainly focus on the imprecision of model parameters to define the edge weights. It is in this sense that we characterized the edge weights of a graph with type-2 fuzzy numbers, to be explained next.

5.2 The Learning Algorithm

We assume that the dataset $X = \{x_1, ..., x_n\}$ contain set of labeled and unlabeled data points which are represented as $X = \{X^u \cup X^l\}$. There are c number of classes based on the class labels of the labeled data and k represents the nearest neighbor count. For the labeled training data points the degree of memberships to any of classes is either 0 or 1 depending on the label of the class. After we obtain the partition matrix of the overall dataset X, we compute type-2 link membership values of each pair in the dataset.

We construct one KFKNN model and obtain one partition matrix $U^{sr} \subset \Re^{n \times k}$ for each pair of imprecise parameters, $\{\sigma^s, m^r\}$, $s = 1..ns$, $r = 1..nr$, where $\mu_{ij}^{sr} \in U^{sr}$ represents the degree of membership of data point x_j to any class i obtained using parameter set of $\{\sigma^s, m^r\}$. The next step is to merge all these partition matrices to construct type-2 fuzzy graph. Each vertex is connected to its k-nearest neighbor obtained from each KFKNN model. First we measure the similarities between each connected node, $w_{pq}^s \in W^s$ in each graph with Gaussian weighted distance measure as in (1). Note that the edge similarities are effected by the Gaussian variance parameter, σ^s. As a result, we characterize each edge weight, $\tilde{w}_{pq} = \{w_{pq}^1, ..., w_{pq}^{ns}\}$, $s = 1, .., ns$, with a discrete interval valued number based on σ^s.

The next step is to associate degree of similarity to each edge weight. Any vertex v_j, representing each data point x_j holds a degree of membership μ_{ij}^{sr} to each class label $i = 1...c$ (conditional degree of possibility given class labels). To construct the fuzzy graph, μ_i^{sr} we calculate the degree of relationship between each node using conditional possibilities via min-max rule. Let $\mu_{R(x_p,x_q)}$ be the degree of relationship (similarity) between paired data points, x_p and x_q, in the graph which is measured from the conditional possibility values of

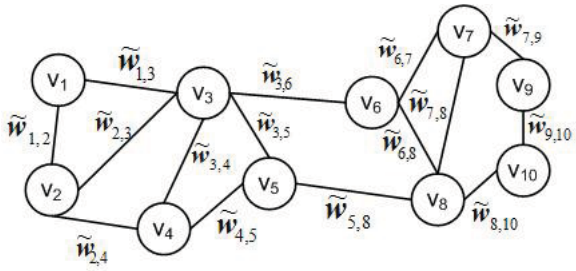


Figure 1: A type-2 fuzzy graph.

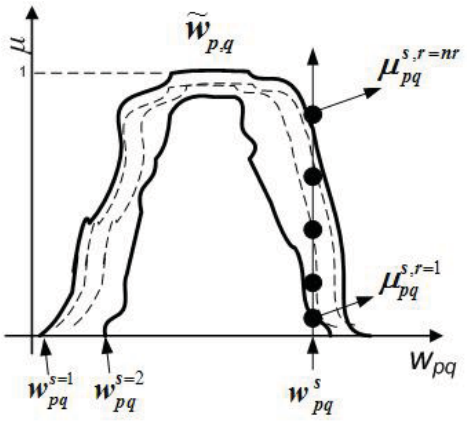


Figure 2: The interval type-2 fuzzy edge weight.

each vertex. Using the fuzzy $max-min$ rule of relations [13], we calculate the joint degree of fuzzy relations as follows:

$$\mu_{pq}(x_p, x_q) = \vee_{i=1,...,c} \left(\mu_i(x_p) \wedge \mu_i(x_q)\right) \qquad (17)$$

where $\vee$ and $\wedge$ denote max and min operations, respectively. Using (17) we define as many degree of similarity values, μ_{pq}^{sr}, as the number of discrete m^r, $r = 1..nr$, to characterize each discrete edge weight, w_{pq}^s. This would identify the edge weights as type-2 fuzzy numbers, see Figure 2.

In summary, we defined ns, $s = 1, ..., ns$ different variance values, σ^s to calculate ns different similarity values, w_{pq}^s, $s = 1.., ns.$ for any pair of connected nodes. These form the soft weight, $\tilde{w}_{pq} = w_{pq}^{s=1..ns}$ of the corresponding edge $e_{pq}, p, q = 1, ..., n$ of the graph. In addition, each discrete similarity value w_{pq}^s is represented with a degree of similarity from the $max-min$ operation, $\mu_{pq}^{sr}(x_p, x_q)$ using (17). These stretch out the degree of membership of each w_{pq}^s, and we identify discrete membership values for each weight, w_{pq}^s.

The fuzzy graph where the edges are represented with the soft type-2 weights is shown in Fig. 1, and Fig. 2 displays a sample edge weight that is in the form of type-2 fuzzy number.

The pseudo code for the type-2 fuzzy semi-supervised graph learning method is shown in Algorithm 2. Initially the partition matrix U is computed for unlabeled data points using the labeled data (Step-3). Note that the membership values of the labeled data points $\mu_i(x_j^l)$ to any of the c clusters are either 1 or 0 so they are merged to the partition matrix (Step-7).

After the conditional possibility values of each data point x_j to each class $\mu_i^{sr}(x_j)$ is obtained from every pair of parameter values, $\{\sigma^s, m^r\}$ we measure the joint possibilities of paired data points in step 9-10. We use the $max-min$ rule to obtain single degree of similarity between two nodes, x_p and x_q. In

ISBN: 978-989-95079-6-8

Algorithm 2 Forming Type-2 SSL Graph

procedure LEARNGRAPH(k, X^l, X^u, $X = \{X^l + X^u\}$)
  For $\sigma^s = \{\sigma^1, ..., \sigma^{ns}\}$ and $m^r = \{m^1, ..., m^r\}$
    -Execute Algorithm 1 using $\{\sigma^s, m^r\}$ to obtain
    each partition matrix $U^{sr} \subset \Re^{n \times c}$
    -Calculate similarity matrix $W^s = \{w^s_{pq}\}$ using σ^s
  EndFor
  $\mu^{sr}_j \in U^{sr} = U^{sr} + \{\mu_{i=1..c}(x^l_{i=1..n_l})\}$ where
  For $p, q = 1, ..., n,\ p \neq q$
    -Calculate degree of fuzzy relation, $\mu^{sr}_{pq}(x_p, x_q)$
    using (17)
    -Characterize each edge weight as type-2 fuzzy
    number $\tilde{w}_{pq} = \{w^{s=1..ns}_{pq}\} \in \tilde{W} \subset \Re^{n \times n}$
    using each μ^{sr}_{pq}.
    $\tilde{d}_p \in \tilde{D} \subset \Re^{n \times n} \leftarrow \sum_{q=1}^{n} \tilde{w}_{pq}$
  EndFor
  $\tilde{Ln} \leftarrow \tilde{D}^{-1/2} \tilde{W} \tilde{D}^{-1/2}$
end procedure

this way we represent the discrete membership values of the type-2 fuzzy number of the corresponding edge similarity.

Using the type-2 fuzzy number operations from section 4 [3], we obtain the fuzzy degree of each connected edge. Note that, instead of crisp edge weights, W, and graph node degrees, D, we characterized type-2 fuzzy edge matrix $\tilde{W}$ and degrees, $\tilde{D}$. To obtain the soft normalized graph Laplace, we compute the Step-16.

5.3 Inference with Soft Graphs

In the previous section we presented the algorithm to form a soft graph Laplacian matrix $\tilde{L}_n$, the members of which are represented with interval type-2 fuzzy numbers. It has been shown in recent literature [**?**, 9, 11, 10] that some of the uncertainties in system models can be captured during learning when we define interval values for the learning parameters instead of crisp values. For instance when fuzzy clustering methods are used to construct membership functions (fuzzy sets) defining an interval valued fuzziness parameter, m, instead of crisp value, we can identify the uncertainty interval of the membership functions as well as the perturbation effects on the local functions. The same is also true for the parameters of similarity functions. The variance parameter of a Gaussian kernel has a significant effect on the outcome [7]. By the analogy of these approaches, we use the interval valued Gaussian variance distance measures and fuzziness parameter of the KFKNN for identification of the soft graph Laplacian as explained in the previous section. The classifier function is learnt using the graph Laplacian. In standard SSL methods, briefly explained in section 2, one can use a label prorogation method to find the classifier function, f^*. In Type-2 SSL algorithm of this paper, we learn classifier by first type-reducing the type-2 graph Laplacian and then defuzzifying its values to obtain a single graph Laplace.

The graph Laplacian is a sparse matrix $L := \left(l^{n \times n}_{pq}\right)$ representing the difference between the degree matrix and weight matrix. The matrix elements are zero if there is no edge link, and 1 for the diagonal elements. Hence the matrix is sparse representing the graph where there are edge

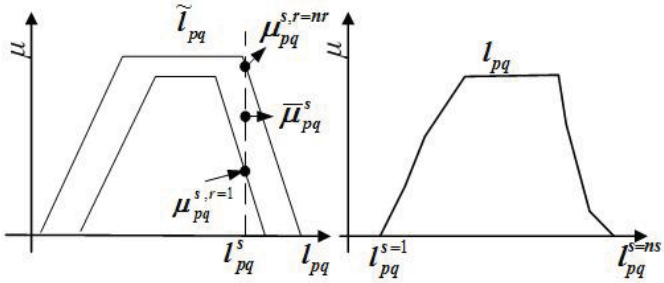


Figure 3: (Left Graph)Interval type-2 fuzzy laplace value. (Right Graph)Type-1 fuzzy laplace value.

links. We consider soft normalized graph Laplacian $\tilde{\mathcal{L}} = I - \tilde{D}^{-1/2} \tilde{W} \tilde{D}^{-1/2}$. Each matrix element $\tilde{l}_{pq}$ is an interval type-2 fuzzy number, represented as:

$$\tilde{l}_{pq} = \left\{\mu^{s=1,r=1}_{pq}, .., \mu^{s,r=nr}_{pq}\right\} / l^{s=1}_{pq} + \cdots + \left\{\mu^{s=ns,r=1}_{pq}, .., \mu^{s=ns,r=nr}_{pq}\right\} / l^{s=ns}_{pq}$$

We consider type-reducing the Laplace values via center of gravity (COG) rule as shown in Figure 3 as

$$COG(\mu^s_{pq}) = \sum_{j=1}^{nr} \mu^{s,r=j}_{pq} / nr \tag{18}$$

Then we defuzzify to obtain a single Laplace values:

$$l_{pq} = \frac{\sum_{t=1}^{ns} COG\left(\mu^s_{pq}\right) l^{s=t}_{pq}}{\sum_{i=1}^{ns} COG\left(\mu^s_{pq}\right)} \tag{19}$$

Using the defuzzified graph Laplacian, we obtain the classifier function using (6).

6 Textual Entailment Model for Question Answering

In this section we use a real question and answering (Q/A) dataset to test the performance of the type-2 SSL in comparison to the SSL method [7], and well-known SVM [19]. We used the TREC2001-2003 factoid questions for training and TREC04 questions for testing. The training and testing datasets comprise of 20 relevant sentences for each question that may or may not contain the true answer. We build the training and testing datasets using the relevant documents provided by NIST [2] and extracted 20 relevant sentences from TREC corpus using search engine [3]. We manually classified each retrieved sentence as true/false based on the provided gold-factoid-answers. In addition, we used additional question-sentence pairs from RTE-3, and RTE-4 datasets of RTE challenges [4]. We used only Q/A type sentence pairs and converted the hypothesis sentences into question forms to build the Q/A pairs. We sustained the true/false distributions in the training and testing datasets around 25% true-75% false.

Each of the Q/A pairs, i.e., from TREC and RTE, are used to extract features to indicate true/false entailment. Thus the

[2] available at http://trec.nist.gov/data/qa.html

[3] available at http://lucene.apache.org/java/docs/

[4] Recognizing Textual Entailment Challenge, datasets available at http://www.nist.gov/tac/tracks/2008/rte/index.html

ISBN: 978-989-95079-6-8

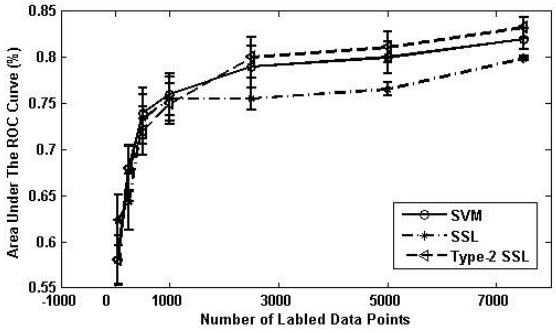


Figure 4: AUC performance on Testing Dataset.

aim of the model is to identify if the candidate sentence entails the question. To measure entailment we extracted 20 different features based on syntactic and semantic match between the question and answer sentences. Some of our features include syntactic matching (keyword and phrase match), named-entity match, headword match using hypernym, hyponym extensions of words, answer type match, which is determined from the question. An automatic hypernym relation extraction algorithm is implemented to capture semantic structures in sentences, such as 'a.k.a', 'such as', apposition relations, to list a few. In addition, we used semantic features such as sentence structure match, i.e., subject, object, headword match, named entity match based on our named entity module to match 50 fine objects in sentence pairs such as time, place, location, person, descriptions, reason, numbers, etc.

In this experiment we used around 600 training questions to compile around 7500 training data points. Similarly, using the entire 202 TREC04 factoid questions as testing, we compiled around 4000 testing data points. Among the training datasets, we used different number of labeled data points to compile series of experiments. Since the data points are randomly selected from the pool, we repeated the experiments 10 times and captured the standard errors to measure robustness. We measured the accuracy of the true/false entailments based on **A**rea **U**nder the Receiver Operating **C**urve statistic. We repeated the experiment using different training sets of different sizes. We built type-2 SSL models with parameters, i.e., $k = \{3, 5, 10, 50\}$, $m = \{1.4, 1.5, ..., 2.6\}$, $\sigma = \left\{10^{-2}, ..10^{2}\right\}$ and tested the AUC performance on the same testing dataset. We compared the performance of our models with the ones obtained from application of the same experiments using standard SSL method [20] with the same learning parameters as well as the SVM methods using learning parameters $C = \left\{2^{-1}, ..., 2^{8}\right\}$ and $\gamma = \{1 - default\}$. The AUC performance comparison with the well-known SSL and SVM method is shown Figure 4. The results of the experiment on QA datasets reveals that the new soft SSL outperforms the standard SSL method. The results are comparable with the state-of the art SVM tool. Based on the current experiment we can suggest that the presented method is an alternative method for linguistic analysis when there are not enough labeled data but more unlabeled data present.

7 Conclusions

We presented a new uncertainty modeling tool for graph-based semi-supervised learning methods using type2 fuzzy number arithmetic. We learn the interval valued degree of similarity values based on fuzzy k-nearest approach. Each edge linking weight is represented with a type-2 fuzzy number and arithmetic operations on type-2 fuzzy numbers are implemented to build the type-2 fuzzy graph Laplacian and learn a classifier function. Experiments on information extraction from unstructured text for question/answering task have shown promising results.

References

[1] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Scholkopf. Learning with local and global consistency. *Advances in Neural Information Processing Systems*, 16:321–328, 2004.

[2] X.-H. Wu and J.-J. Zhou. Kernel-based fuzzy k-nearest neigbor algorithm. 2:159–162, 2005.

[3] W. Blewitt, S.M. Zhou, and S. Coupland. A novel approach to type-2 fuzzy addition. *Proc. Intern. Conference on Fuzzy Systems*, pages 1–6, 2007.

[4] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[5] E.-M. Voorhees. The trec question answering track. *Natural Language Engineering*, 7(4):361–378, 2001.

[6] D. Zhou and B. Scholkopf. Learning from labeled and unlabeled data using random walks. pages 237–244, 2004.

[7] X. Zhu, J. Lafferty, and Z. Ghahramani. *Semi-supervised learning: From Gaussian Fields to Gaussian processes*. Technical Report, Carnegie Mellon University, 2003.

[8] J.M. Keller, M.R. Gray, and J.A. Givens Jr. A fuzzy k-nearest neighbor algorithm. *IEEE Trans. Systems Man and Cybernetics*, 15:580–585, 1985.

[9] A. Celikyilmaz and I.B. Turksen. Uncertainty bounds of fuzzy c-regression method. *IEEE-FUZZ Int. Conf. on Fuzzy Systems, Hong Kong*, pages 1098–7584, June 2008.

[10] A. Celikyilmaz and I. B. Turksen. Uncertainty modeling with evolutionary improved fuzzy functions approach. *IEEE Systems, Man, and Cybernetics- Part B*, 38(4):1098–1110, 2008.

[11] C. Hwang and F. C.-H. Rhee. Uncertainty fuzzy clustering: Interval type-2 fuzzy approach to c-means. *IEEE Trans. On Fuzzy Systems*, 15(1):107–120, 2007.

[12] J.C. Bezdek. *Pattern recognition with fuzzy objective functions*. Plenum Press, New York, 1981.

[13] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning - (i-ii-iii). *Information Sciences*, 8-9, 1975.

[14] J.M. Mendel and R.I. John. Type-2 fuzzy sets made simple. 10(2):117–127, 2002.

[15] S. Coupland and R. John. An approach to type-2 fuzzy arithmetic. *Proc. UK Workshop on Computational Intelligence*, pages 107–114, 2003.

[16] A. Celikyilmaz and I.B. Turksen. Uncertainty modeling with evolutionary improved fuzzy functions approach. *IEEE Systems, Man, and Cybernetics- Part B*, 38(4):1098–1110, 2008.

[17] O. Castillo and P. Melin. *Type-2 Fuzzy Logic: Theory and Applications*. Springer-Verlag, Heidelberg, Germany, 2008.

[18] M. B. Begian, W. Melek, and J.M. Mendel. Parametric design of stable type-2 tsk fuzzy systems. 2008.

[19] B.E. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. *Proc. of Fifth Annual Workshop on Computational Learning Theory*, 3:144–152, 1992.

[20] X. Zhu and Z. Ghahramani. *Learning from labeled and unlabeled data with label propagation*. Technical Report CMU-CALD-02-107, Carnegie Mellon University, Pittsburgh, 2002.

ISBN: 978-989-95079-6-8

Can We Learn Algorithms from People Who Compute Fast: An Indirect Analysis in the Presence of Fuzzy Descriptions

Olga Kosheleva[1]

1.Department of Teacher Education
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
olgak@utep.edu

Abstract— *In the past, mathematicians actively used the ability of some people to perform calculations unusually fast. With the advent of computers, there is no longer need for human calculators – even fast ones. However, recently, it was discovered that there exist, e.g., multiplication algorithms which are much faster than standard multiplication. Because of this discovery, it is possible than even faster algorithm will be discovered. It is therefore natural to ask: did fast human calculators of the past use faster algorithms – in which case we can learn from their experience – or they simply performed all operations within a standard algorithm much faster? This question is difficult to answer directly, because the fast human calculators' self-description of their algorithm is very fuzzy. In this paper, we use an indirect analysis to argue that fast human calculators most probably used the standard algorithm.*

Keywords— Fast Fourier transform, fast human calculators, fast multiplication, fuzzy description

1 People Who Computed Fast: A Historical Phenomenon

In history, several people have been known for their extraordinary ability to compute fast. Before the 20 century invention of computers, their computational abilities were actively used.

For example, in the 19 century, Johann Martin Zaharias Dase performed computations so much faster than everyone else that professional mathematicians hired him to help with their calculations; see, e.g., [1, 3]. Dase:

- computed π with a record-breaking accuracy of 200 digits,
- calculated the logarithms table with 7 digit accuracy, and
- performed many other computational tasks.

Karl Friedrich Gauss himself recommended that Dase be paid by the Academy of Science to perform the calculations.

2 People Who Computed Fast: How They Computed?

Calculations have been (and are) important in many practical problems. Because of this practical importance, people have therefore always been trying to speed up computations. One natural way to speed up computations is to learn from the people who can do computations fast.

3 People Who Computed Fast: Their Self-Explanations Were Fuzzy and Unclear

In spite of numerous attempts to interview the fast calculators and to learn from them how they compute, researchers could not extract a coherent algorithm. One of the main reasons is that most of the fast calculators were *idiots savants*: their intellectual abilities outside computations were below average. Their explanations of their algorithms were always fuzzy and imprecise, formulated in terms of words of natural language rather than in precise mathematical terms.

To us familiar with fuzzy logic and its applications this fuzziness should not be surprising. It is normal that people in general – and not necessarily idiots savants – cannot precisely describe

- how they drive,
- how they operate machines,
- how they walk,
- how they translate from one language to another,
- how they recognize faces,
- how they control different situations, etc.

– this is why fuzzy control and other fuzzy techniques, techniques for translating this fuzzy description into a precise strategy, have so many practical applications; see, e.g., [4].

What may be somewhat unusual about fast computations is that while the self-description of fast calculators was fuzzy and imprecise, the results of the computations were always correct and precise.

4 With the Appearance of Computers, the Interest in Fast Human Calculators Waned

With the appearance of computers, the need for human calculators disappeared, and the interest in their skills waned. Such such folks may provide a good entertainment, they may be of interest to psychologists who study how we reason and how we perform menial tasks – but mathematicians were no longer interested.

Yes, fast human calculators can perform calculations faster than an average human being – but electronic computers can perform the same computations much much faster. From this viewpoint, fast human calculators remained a curiosity.

ISBN: 978-989-95079-6-8

5 Why Interest Waned: Implicit Assumptions

One of the main reasons why in the 1940s and 1950s the interest in fast human calculators waned is that it was implicitly assumed that the standard techniques of addition and multiplication are the best.

For example, it was assumed that the fastest way to add the two n-digit numbers is to add them digit by digit, which requires $O(n)$ operations with digits.

It was also implicitly assumed that the fastest way to multiply two n-digit numbers is the standard way to multiply the first number by each of the digits of the second numbers, and then to add the results. Each multiplication by a digit and each addition requires $O(n)$ steps, thus multiplication by all n digits and the addition of all n results require $n \cdot O(n) = O(n^2)$ computational steps.

From this viewpoint, the only difference between a normal human calculator, a fast human calculator, and an electronic computer is in the speed with which we can perform operations with digits. From this viewpoint, the only thing we can learn from fast human calculators is how to perform operations with digits faster. Once the electronic computers became faster than fast human calculators, the need to learn from the fast human calculators disappeared.

6 A Surprising 1960s Discovery of Fast Multiplication Algorithms

The above implicit assumption about arithmetic operations was not questioned until a surprising sequence of discoveries was made in the 1960s; see, e.g., [2].

These discoveries started with the discovery of the Fast Fourier Transform algorithm, an algorithm that enables us to compute the Fourier Transform

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \cdot \int f(t) \cdot \exp(\mathrm{i} \cdot \omega \cdot t)\, dt, \tag{1}$$

where $\mathrm{i} = \sqrt{-1}$, in time $O(n \cdot \log(n))$ – instead of the $n \cdot O(n) = O(n^2)$ time needed for a straightforward computation of each of n values of $\hat{f}(\omega)$ as an integral (i.e., in effect, a sum) over n different values $f(t)$.

The ability to compute Fourier transform fast lead to the ability to speed the computation of the convolution of two functions:

$$h(t) \stackrel{\text{def}}{=} \int f(s) \cdot g(t-s)\, ds. \tag{2}$$

A straightforward computation of the convolution requires that for each of the n values $h(t)$, we compute the integral (sum) of n different products $f(s) \cdot g(t-s)$ corresponding to n different values s. Thus, the straightforward computation requires $O(n^2)$ computational steps.

Indeed, it is known that the Fourier transform of the convolution is equal to the product of Fourier transforms. Thus, to compute convolution, we can do the following:

- first, we compute Fourier transforms $\hat{f}(\omega)$ and $\hat{g}(\omega)$;
- then, we compute the Fourier transform of h as

$$\hat{h}(\omega) = \hat{f}(\omega) \cdot \hat{g}(\omega); \tag{3}$$

- finally, we apply the inverse Fourier transform to the function $\hat{h}(\omega)$ and compute the desired convolution $h(t)$.

What is the computation time of this algorithm?

- Both Fourier transform and inverse Fourier transform can be computed in time $O(n \cdot \log(n))$.
- The point-by-point multiplication $\hat{h}(\omega) = \hat{f}(\omega) \cdot \hat{g}(\omega)$ requires n computational steps.

Thus, the overall computation time requires

$$O(n \cdot \log(n)) + n + O(n \cdot \log(n)) = O(n \cdot \log(n)) \tag{4}$$

steps, which is much faster than $O(n^2)$.

V. Strassen was the first to notice, in 1968, than this idea can lead to fast multiplication of long integers. Indeed, an integer x in a number system with base b can be represented as a sum

$$x = \sum_{i=1}^{n} x_i \cdot b^i. \tag{5}$$

In these notations, the product $z = x \cdot y$ of two integers $x = \sum\limits_{i=1}^{n} x_i \cdot b^i$ and $y = \sum_{i=1}^{n} y_i \cdot b^i$ can be represented as

$$z = \left(\sum_{i=1}^{n} x_i \cdot b^i\right) \cdot \left(\sum_{j=1}^{n} y_j \cdot b^j\right) =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_i \cdot y_j \cdot b^i \cdot b^j = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i \cdot y_j \cdot b^{i+j}. \tag{6}$$

By combining terms at different values b^k, we conclude that

$$z = \sum_{k=1}^{n} z_k \cdot b^k, \tag{7}$$

where

$$z_k = \sum_{i} x_i \cdot y_{k-i}. \tag{8}$$

This is convolution, and we know that convolution can be computed in time $O(n \cdot \log(n))$.

The values z_k are not exactly the digits of the desired number z, since the sum (8) can exceed the base b. Thus, some further computations are needed. However, even with these further computations, we can multiply two numbers in almost the same time

$$O(n \cdot \log(n) \cdot \log(\log(n))). \tag{9}$$

The corresponding algorithm, first proposed by A. Schönhage and V. Strassen in their 1971 paper [5], remains the fastest known – and it is much faster than the standard $O(n^2)$ algorithm.

7 Fast Multiplication: Open Problems

Fast algorithms drastically reduced the computation time. Fast Fourier transform is one of the main tools for signal processing. These successes has led to the need to find faster and faster algorithms. From this viewpoint, it is is desirable to look for even faster algorithms for multiplying numbers.

The fact that researchers succeeded in discovering algorithms which are much faster than traditional multiplication gives us hope that even faster algorithms can be found.

ISBN: 978-989-95079-6-8

8 Interest in Fast Human Calculators Revived

Where to look for these algorithms? One natural source is folks who did compute fast. From this viewpoint, the fact that they were unable to clearly explain what algorithm they used becomes an advantage: maybe the algorithm that they actually used is some fast multiplication algorithm? This possibility revived an interest in fast human calculators.

So, the question is:

- did fast human calculators use fast multiplication algorithm(s), or
- they used the standard algorithm but simply performed operations with digits faster?

9 Direct Analysis Is Impossible

The most well-known fast human calculator, Johann Martin Zacharias Dase, died almost 150 years ago. Even when he was alive, his self-descriptions were not sufficient to find out how exactly he performed the computations. There might have been hope that our knowledge of fast multiplication algorithms can help in this understanding, but it did not work out. In other words, the direct analysis of Dase's behavior has been impossible – and it is still impossible. We therefore need to perform an *indirect* analysis.

10 Indirect Analysis: Main Idea

A natural way to check which algorithm is used by a computational device – be it a human calculator or an electronic computer – is to find out how the computation time changes with the size n (= number of digits) of the numbers that we are multiplying.

- If this computation time grows with n as n^2, then it is reasonable to conclude that the standard algorithm is used – since for this algorithm, the computation time grows as n^2.
- On the other hand, if the computation time grows with n as $\approx n \cdot \log(n)$ (or even slower), then it is reasonable to conclude that a fast multiplication algorithm is used.
  - It maybe a Strassen-type algorithm, for which the computation time grows as $n \cdot \log(n)$.
  - It may be a (yet unknown) faster algorithm in which case the computation time grows even slower than $n \cdot \log(n)$.

11 Data That We Can Use

Interestingly, there is a data on the time that Dase needed to perform multiplication of numbers of different size. This data comes from fact that Dase's performance was analyzed and tested by several prominent mathematicians of his time – including Gauss himself. Specifically:

- Dase multiplied two 8-digit numbers in 54 seconds;
- he multiplied two 20-digit numbers in 6 minutes;
- he multiplied two 40-digit numbers in 40 minutes; and
- he multiplied two 100-digit numbers in 8 hours and 45 minutes.

12 Analysis

For the standard multiplication algorithm, the number of computational steps grows with the numbers size n as n^2. Thus, for the standard multiplication algorithm, the computation time for performing the computation also grows as n^2:

$$t(n) = C \cdot n^2. \tag{10}$$

So, for this algorithm, two different number sizes $n_1 < n_2$, we would have $t(n_1) = C \cdot n_1^2$ and $t(n_2) = C \cdot n_2^2$ and thus,

$$\frac{t(n_2)}{t(n_1)} = \frac{n_2^2}{n_1^2} = \left(\frac{n_2}{n_1}\right)^2. \tag{11}$$

For a faster algorithm, e.g., for the algorithm that requires $O(n \cdot \log(n))$ running time, the corresponding ratio will be smaller:

$$\frac{t(n_2)}{t(n_1)} = \frac{n_2 \cdot \log(n_2)}{n_1 \cdot \log(n_1))} = \frac{n_2}{n_1} \cdot \frac{\log(n_2)}{\log(n_1)}. \tag{12}$$

Thus, to check whether a human calculators uses the standard algorithm or a faster one it is sufficient to compare the the corresponding time ratio $\frac{t(n_2)}{t(n_1)}$ with the square $\left(\frac{n_2}{n_1}\right)^2$:

- If the time ratio is smaller than the square, this means that the human calculator used an algorithm which is much faster than the standard one.
- On the other hand, if the time ratio is approximately the same as the square, thus means that the human calculators most probably used the standard algorithm – or at least some modification of it that does not drastically speed up the computations.
- If it turns out that the time ratio is larger than the square, this would mean, in effect, that a human calculator used an algorithm which is asymptotically even slower than the standard one – this can happen, e.g., if there is an overhead needed to store values etc.

According to the above data, we have:

- $t(8) = 0.9$ minutes,
- $t(20) = 6$ minutes,
- $t(40) = 40$ minutes, and
- $t(100) = 8 \cdot 60 + 45 = 525$ minutes.

Here, for $n_1 = 8$ and $n_1 = 20$, we have:

$$\frac{t(n_2)}{t(n_1)} = \frac{t(20)}{t(8)} = \frac{6}{0.9} \approx 6.7 > \left(\frac{n_2}{n_1}\right)^2 =$$

$$\left(\frac{20}{8}\right)^2 = 2.5^2 = 6.25. \tag{13}$$

For $n_1 = 8$ and $n_2 = 40$, we have

$$\frac{t(n_2)}{t(n_1)} = \frac{t(40)}{t(8)} = \frac{40}{0.9} \approx 44 > \left(\frac{n_2}{n_1}\right)^2 =$$

ISBN: 978-989-95079-6-8

$$\left(\frac{40}{8}\right)^2 = 5^2 = 25. \quad (14)$$

Finally, for $n_1 = 8$ and $n_2 = 100$, we have

$$\frac{t(n_2)}{t(n_1)} = \frac{t(100)}{t(8)} = \frac{525}{0.9} \approx 583 > \left(\frac{n_2}{n_1}\right)^2 = \left(\frac{100}{8}\right)^2 = 12.5^2 \approx 156. \quad (15)$$

In all these cases, the computation time of a human calculator grows faster then n^2 corresponding to the standard algorithm.

The same conclusion can be made if instead of comparing each value n with the smallest value n_1, we compare these values with each other. For $n_1 = 20$ and $n_2 = 40$, we have

$$\frac{t(n_2)}{t(n_1)} = \frac{t(40)}{t(20)} = \frac{40}{6} \approx 7 > \left(\frac{n_2}{n_1}\right)^2 = \left(\frac{40}{20}\right)^2 = 2^2 = 4. \quad (16)$$

For $n_1 = 20$ and $n_2 = 100$, we have

$$\frac{t(n_2)}{t(n_1)} = \frac{t(100)}{t(20)} = \frac{525}{6} \approx 88 > \left(\frac{n_2}{n_1}\right)^2 = \left(\frac{100}{20}\right)^2 = 5^2 = 25. \quad (17)$$

Finally, for $n_1 = 40$ and $n_2 = 100$, we have

$$\frac{t(n_2)}{t(n_1)} = \frac{t(100)}{t(40)} = \frac{525}{40} \approx 13 > \left(\frac{n_2}{n_1}\right)^2 = \left(\frac{100}{40}\right)^2 = 2.5^2 = 6.25. \quad (18)$$

Thus, it is reasonable to conclude that the fast human calculators did not use any algorithm which is faster than the standard one.

13 Possible Future Work

People who perform computations fast appear once in a while. It may be a good idea to record and analyze their computation time – and maybe record their fuzzy explanations and try to make sense of them.

Acknowledgment

The author is thankful to the anonymous referees for useful suggestions.

References

[1] P. Beckmann, *A History of* π, Barnes & Noble, New York, 1991.

[2] Th. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, MA, 2001.

[3] D. R. Hofstadter, *Godel, Escher, Bach: an Eternal Golden Braid*, Basic Books, 1999.

[4] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*, Prentice Hall, New Jersey, 1995.

[5] A. Schönhage and V. Strassen, "Schnelle Multiplikation grosser Zahlen", *Computing*, 1971, Vol. 7, pp. 281–292.

ISBN: 978-989-95079-6-8

A parametric method to solve quadratic programming problems with fuzzy costs

Ricardo C. Silva[1] Carlos Cruz[2] Akebo Yamakami[1]

1.Department of Telematics, School of Electrical and Computer Engineering,
University of Campinas, P.O.Box 6101, 13083-970, Campinas-SP, Brazil.
Email: {rcoelhos,akebo}@dt.fee.unicamp.br
2.Department of Computer Science and Artificial Intelligence,
University of Granada, E-18071, Granada, Spain
Email: carloscruz@decsai.ugr.es

Abstract— *This work describes a novel fuzzy-sets-based method to solve a particular class of quadratic programming problems which have vagueness coefficients in the objective function. Quadratic programming problems are of utmost importance in an increasing variety of practical fields. In addition, as the ambiguity and vagueness are natural and ever-present in real-life situations requiring solutions, it makes perfect sense to attempt to address them using fuzzy quadratic programming problems. Also, two other methods to solve this kind of problems are briefly described. The proposal uses two phases to solve fuzzy quadratic programming problems. In the first, phase we parametrize the fuzzy problem in several classical alpha-problems with different cutting levels. In the second, phase each of these alpha-problems is solved by using conventional solving techniques. The final fuzzy solution to the problem is obtained by integrating all of these particular alpha-solutions. The results obtained using these two methods are compared with the two-phased proposal outlined above.*

Keywords— Fuzzy sets, decision making, fuzzy mathematical programming, quadratic optimization.

1 Introduction

The conception of Artificial Intelligence (AI) in the early fifties was inspired by the wide variety of physical and mental tasks performed by humans without any measurements and any computations. Thus, this forward thinking described the definition of AI as it is widely known: "the study and design of intelligent agents [14]", where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success[15]. The capability to compute and reason with perception-based information can be applied to real-world problems in which decision-relevant information is a mixture of measurements and perceptions [20], where in general, measurements are crisp, although perceptions are fuzzy.

The literature shows that the best way of modeling these types of problems is using Soft Computing methodologies. According to Verdegay et al.[17], Soft Computing is a family of problem-resolution methods headed by approximate reasoning, functional and optimization approximation methods, which now include search methods. More specifically, the search methods use thorough evaluations to find regions with approximate points of optimal solutions that are obtained by using optimization methods.

These methods, which are algorithms based on mathematical knowledge, are used to guarantee the convergence to the optimal solution. They are covered by the area of mathematical programming. This area has several sub-fields where the quadratic programming is a special type that optimizes a quadratic objective function of several variables subject to linear constraints. A precise definition of the constraints and the objective function to be optimized is necessary to obtain exact optimal solutions of quadratic programming problems.

Nevertheless, in most real practical applications (portfolio, game theory, engineering modeling, design and control, logistics, etc.) one lacks this kind of exact knowledge, and only approximate, vague and imprecise values are known. Moreover, these imprecise values can be dealt with fuzzy logic. In this case, the concept of fuzzy mathematical programming emerges when it is used.

With this in mind, the objective of this paper is to review some methods that are related to quadratic topic and to outline a soft computing model used to solve novel fuzzy quadratic programming problems.

The paper is organized as follows: Section 2 briefly introduces the different approaches to solving the quadratic programming problem with fuzzy costs. In this section, a novel approach is developed to solve quadratic programming problems with fuzzy costs where these are transformed into parametrical quadratic multiobjective programming problems. To clarify the above developments, three numerical examples are analyzed in section 3. Finally, conclusions are presented in Section 4.

2 Quadratic optimization with fuzzy costs

An optimization problem that is described with a quadratic objective function subject to linear constraints is called a "Quadratic Programming" problem. QP can be viewed both as a special case of the nonlinear programming and a generalization of the linear programming. Several applications and methods can be found in [3, 8, 9, 16, 19]. A quadratic programming problem can be formulated as

$$\begin{array}{ll} \min & \mathbf{c^t x} + \frac{1}{2}\mathbf{x^t Q x} \\ \text{s.t.} & \mathbf{A x} \leq \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array} \tag{1}$$

where $\mathbf{x}$ and $\mathbf{c}$ are n vectores of real numbers, $\mathbf{Q}$ is an $n \times n$ symmetric matrix of real numbers, $\mathbf{A}$ is an $m \times n$ matrix of real numbers and $\mathbf{b}$ is an m vector of real numbers.

ISBN: 978-989-95079-6-8

In this work, quadratic programming problems with fuzzy costs are considered because there are cases of real world problems whose parameters are seldom known exactly and have to be estimated by the decision maker. Hence, the $n \times n$ symmetric matrix $\mathbf{Q}$ and the n vector $\mathbf{c}$ have fuzzy numbers in these components. The application of fuzzy logic is also a way to describe, mathematically, this vagueness as described in [2, 10, 13, 18]. The uncertainties of costs of the objective function can be dealt with by fuzzy numbers.

Thus, this set of problems with uncertain costs can be formalized in the following form:

$$\begin{array}{ll} \min & \tilde{\mathbf{c}}^{\mathbf{t}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{t}}\tilde{\mathbf{Q}}\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array} \tag{2}$$

where the fuzzy numbers are characterized by membership functions that are defined by decision makers. The membership functions can be defined as

$$\mu_j, \mu_{ij} : \mathbb{R} \to [0,1], \;\; i,j \in \mathbb{J} = \{1,2,\ldots,n\}$$

In particular these membership functions will be supposed as:

$$\mu_j(y) = \begin{cases} 0 & \text{if } c_j^U \leq y \text{ or } y \leq c_j^L \\ h_j(y) & \text{if } c_j^L \leq y \leq c_j \\ g_j(y) & \text{if } c_j \leq y \leq c_j^U \end{cases} \qquad j \in \mathbb{J} \tag{3}$$

and

$$\mu_{ij}(y) = \begin{cases} 0 & \text{if } q_{ij}^U \leq y \text{ or } y \leq q_{ij}^L \\ h_{ij}(y) & \text{if } q_{ij}^L \leq y \leq q_{ij} \\ g_{ij}(y) & \text{if } q_{ij} \leq y \leq q_{ij}^U \end{cases} \qquad i,j \in \mathbb{J} \tag{4}$$

where $\mathbf{h}(\cdot)$ and $\mathbf{g}(\cdot)$ are assumed to be strictly increasing and decreasing continuous functions, respectively, $h_j(c_j) = g_j(c_j) = 1,\ j \in \mathbb{J}$ and $h_{ij}(q_{ij}) = g_{ij}(q_{ij}) = 1,\ i,j \in \mathbb{J}$.

Bellman and Zadeh describe the *fuzzy decision* in [4] as the intersection of goals and constraints which is formalized in the following definition:

Definition 1 *Assume that we are given a fuzzy goal μ_G and a fuzzy constraint μ_C in a space of alternative X. Then, μ_G and μ_C combine to form a* fuzzy decision*, μ_D, which is a fuzzy set resulting from intersection of μ_G and μ_C. In symbols*

$$\sup_{\mathbf{x} \in X} \mu_D(\mathbf{x}) = \sup_{\mathbf{x} \in X} [\mu_G(\mathbf{x}) \wedge \mu_C(\mathbf{x})]$$

Note that in Definition 1, the goals and the constraints enter into the expression for μ_D in the same way. Thus, it is able to find a maximizing decision to an extremum problem for a scalar function, as described in [13]. Let ϕ be denoted by $\phi : [0,1] \to [0,1]$ that implies $\phi(\alpha) = \sup_{\mathbf{x} \in \mathbf{X}(\alpha)} \mu_G(\mathbf{x})$ where $X(\alpha) = \{\mathbf{x} \in \mathbb{R}^n \mid \mu_X(\mathbf{x}) \geq \alpha\}$. If ϕ is continuous in [0,1] then ϕ has a *fix point* $\overline{\alpha}$ and, therefore,

$$\sup_{\mathbf{x} \in X} \mu_D(\mathbf{x}) = \sup_{\mathbf{x} \in X(\overline{\alpha})} \mu_G(\mathbf{x}) = \overline{\alpha}$$

.

A novel parametric approach to solve quadratic programming problems with fuzzy costs will be presented in the next sub-section. This approach transforms the original problem into a parametric multi-objective quadratic programming problem.

2.1 Multi-objective approach

A multi-objective approach to solve a linear programming problem with imprecise costs is described in [6, 7]. As the linear problem is a particular case of quadratic problem, this approach can be extended to solve quadratic programming problems with fuzzy costs.

The quadratic objective function can be divided into two parts, where the first one is a linear term and the second one is a quadratic term. According to this, the fuzzy costs can only be in the first part or only the second part or in both. In this section, these three ways will be presented separately.

The linear problem considered in [6] used trapezoid membership functions for the costs but here, for the sake of simplicity, they will be supposed to be like (3) and (4).

Then, by considering the $(1-\alpha)$-cut of every cost, $\alpha \in [0,1]$,

$$\begin{array}{l} \forall\, x \in \mathbb{R},\;\; \mu_j(x) \geq 1-\alpha \Leftrightarrow h_j^{-1}(1-\alpha) \leq x \leq g_j^{-1}(1-\alpha), \\ \forall\, x \in \mathbb{R},\;\; \mu_{ij}(x) \geq 1-\alpha \Leftrightarrow h_{ij}^{-1}(1-\alpha) \leq x \leq g_{ij}^{-1}(1-\alpha), \end{array}$$

where $i,j \in \mathbb{J} = \{1,2,\ldots,n\}$.

2.1.1 Quadratic programming problem with fuzzy costs vector of linear term

The approaches described in this section solve a quadratic programming problem with a vector $\tilde{\mathbf{c}}$ of fuzzy numbers. This vector describes the costs of the objective function and Problem 2 can be rewritten as follows:

$$\begin{array}{ll} \min & \tilde{\mathbf{c}}^{\mathbf{t}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{t}}\mathbf{Q}\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{array} \tag{5}$$

Thus, according to some results by Bitran [5], it has been shown that a fuzzy solution to (5) may be found from the parametric solution of the multi-objective parametric quadratic programming problem

$$\begin{array}{ll} \min & \left[(\mathbf{c^1})^{\mathbf{t}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{t}}\mathbf{Q}\mathbf{x}, (\mathbf{c^2})^{\mathbf{t}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{t}}\mathbf{Q}\mathbf{x}, \ldots, (\mathbf{c^{2^n}})^{\mathbf{t}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{t}}\mathbf{Q}\mathbf{x}\right] \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b},\;\; \mathbf{x} \geq \mathbf{0}, \\ & \mathbf{c^k} \in \mathbb{E}(1-\alpha),\;\; \alpha \in [0,1],\;\; k = 1,2,\ldots,2^n, \end{array} \tag{6}$$

where $\mathbb{E}(1-\alpha)$, for each $\alpha \in [0,1]$, is the set of vectors in $\mathbb{R}^n$ such that each of its components is either in the lower bound, $h_j^{-1}(1-\alpha)$, or in the upper bound, $g_j^{-1}(1-\alpha)$, of the respective $(1-\alpha)$-cut, that is, $\forall\, k = 1,2,\ldots,2^n$, and $\forall\, j \in \mathbb{J}$

$$c^k = (c_1^k, c_2^k, \ldots, c_n^k) \in \mathbb{E}(1-\alpha) \Leftrightarrow c_j^k = \begin{cases} h_j^{-1}(1-\alpha) & \text{or} \\ g_j^{-1}(1-\alpha). \end{cases}$$

Now observe that to find a parametric solution to (6), from which one can obtain a fuzzy solution to (5), one may use any classical multiobjective quadratic programming approach.

2.1.2 Quadratic programming problem with fuzzy costs matrix of quadratic term

The approach described in this section solves a quadratic programming problem with matrix $\tilde{\mathbf{Q}}$ of fuzzy numbers. This

ISBN: 978-989-95079-6-8

vector describes the costs of the objective function and Problem 2 can be rewritten as follows

$$\begin{aligned} \min \quad & \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t \tilde{Q} x} \\ \text{s.t.} \quad & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{7}$$

As it was shown above, a fuzzy solution to (7) may be found from the parametric solution of the multiobjective parametric quadratic programming problem

$$\begin{aligned} \min \quad & \left[\mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q^1 x}, \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q^2 x}, \ldots, \mathbf{c^t x} + \tfrac{1}{2}\mathbf{x^t Q^{2^{n^2}} x}\right] \\ \text{s.t.} \quad & \mathbf{Ax} \leq \mathbf{b}, \ \ \mathbf{x} \geq \mathbf{0}, \\ & \mathbf{Q^k} \in \mathbb{E}(1-\alpha), \ \ \alpha \in [0,1], \ \ k = 1,2,\ldots,2^{n^2}, \end{aligned} \tag{8}$$

where $\mathbb{E}(1-\alpha)$, for each $\alpha \in [0,1]$, is the set of vectors in $\mathbb{R}^n$ such that each of its components is either in the lower bound, $h_{ij}^{-1}(1-\alpha)$, or in the upper bound, $g_{ij}^{-1}(1-\alpha)$, of the respective $(1-\alpha)$-cut, that is, $\forall\, k = 1,2,\ldots,2^{n^2}$, and $\forall\, i,j \in \mathbb{J}$

$$\mathbf{Q^k} = (q_{11}^k, \ldots, q_{1n}^k, \ldots, q_{nn}^k) \in \mathbb{E}(1-\alpha) \Leftrightarrow$$
$$\Leftrightarrow q_{ij}^k = \begin{cases} h_{ij}^{-1}(1-\alpha) & \text{or} \\ g_{ij}^{-1}(1-\alpha). \end{cases}$$

Now observe that to find a parametric solution to (8), from which one can obtain a fuzzy solution to (7), one may use any classical multiobjective quadratic programming approach.

2.1.3 *Quadratic programming problem with all fuzzy costs*

Thus, according to the parametric transformations shown above, a fuzzy solution to (2) may be found from the parametric solution of the multiobjective parametric quadratic programming problem

$$\begin{aligned} \min \quad & \left[(\mathbf{c^1})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^1 x}, (\mathbf{c^2})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^1 x}, \ldots,\right. \\ & \ldots, (\mathbf{c^{2^n}})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^1 x}, (\mathbf{c^1})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^2 x}, \ldots, \\ & \left. \ldots, (\mathbf{c^{2^n}})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^2 x}, \ldots, (\mathbf{c^{2^n}})^\mathbf{t}\mathbf{x} + \tfrac{1}{2}\mathbf{x^t Q^{2^{n^2}} x}\right] \\ \text{s.t.} \quad & \mathbf{Ax} \leq \mathbf{b}, \ \ \mathbf{x} \geq \mathbf{0}, \\ & \mathbf{c^k}, \mathbf{Q^p} \in \mathbb{E}(1-\alpha), \ \ \alpha \in [0,1], \\ & k = 1,2,\ldots,2^n \text{ and } p = 1,2,\ldots,2^{n^2}, \end{aligned} \tag{9}$$

where $\mathbb{E}(1-\alpha)$, for each $\alpha \in [0,1]$, is the set of vectors in $\mathbb{R}^n$ such that each of its components is either in the lower bound, $h_j^{-1}(1-\alpha)$, or in the upper bound, $g_j^{-1}(1-\alpha)$, of the respective $(1-\alpha)$-cut, that is, $\forall\, k = 1,2,\ldots,2^n$, and $\forall\, i,j \in \mathbb{J}$

$$c^k = (c_1^k, c_2^k, \ldots, c_n^k) \in \mathbb{E}(1-\alpha) \Leftrightarrow c_j^k = \begin{cases} h_j^{-1}(1-\alpha) & \text{or} \\ g_j^{-1}(1-\alpha). \end{cases}$$

and

$$\mathbf{Q^k} = (q_{11}^k, \ldots, q_{1n}^k, \ldots, q_{nn}^k) \in \mathbb{E}(1-\alpha) \Leftrightarrow$$
$$\Leftrightarrow q_{ij}^k = \begin{cases} h_{ij}^{-1}(1-\alpha) & \text{or} \\ g_{ij}^{-1}(1-\alpha). \end{cases}$$

The obtained parametrical solutions for any of multiobjective models above, to the different α values, generate a set of solutions and then we use the Representation Theorem to integrate all of these particular alpha-solutions.

2.2 *Liu's approach*

Transforming a fuzzy quadratic programming problem into a two-level mathematical programming problem for finding the bounds of the fuzzy values in the objective function and in the set of constraints is discussed in [11, 12]. These papers describe how to transform the two-level mathematical program into the conventional one-level quadratic program. However, such a method is also valid when only the objective is fuzzy, that is, when only the vector of costs are fuzzy, and for this reason it shall be considered here. According to the goal of this work, the coefficient $\mathbf{c}$ of Problem (5) is fuzzy.

The authors derive the membership function of the goal, and then they apply Zadeh's extension principle to transform the fuzzy quadratic problem into family of classical quadratic problems that can be solved by conventional optimization techniques. Thus, the membership function of the objective function can be defined as

$$\mu_{\tilde{Z}}(z) = \sup_{\mathbf{c}} \min\{\mu_{\tilde{c}_j}(c_j), \ \forall j | z = Z(\mathbf{c})\} \tag{10}$$

where $Z(\mathbf{c})$ is the goal of the conventional quadratic problem. Membership function $\mu_{\tilde{Z}}$ can be computed by finding the functions that describe the shape of the left and right sides of the fuzzy numbers. Then, it is possible to obtain the upper bound of the objective value Z_α^U and your lower bound Z_α^L to each value α. Thus, Z_α^U is the maximum and Z_α^L is the minimum of $Z(\mathbf{c})$, respectively, that can be described as

$$Z_\alpha^U = \max_{c_j} \left\{ \begin{array}{ll} \min_{\mathbf{x}} & \sum_{j=1}^n c_j x_j + \frac{1}{2}\sum_{j=1}^n \sum_{l=1}^n q_{jl} x_j x_l \\ \text{s.a} & \sum_{j=1}^n a_{ij} x_j \leq b_i, \ \ i = 1,\ldots,m \\ & x_j \geq 0, \ \ j = 1,\ldots,n \end{array} \right. \tag{11}$$

$$Z_\alpha^L = \min_{c_j} \left\{ \begin{array}{ll} \min_{\mathbf{x}} & \sum_{j=1}^n c_j x_j + \frac{1}{2}\sum_{j=1}^n \sum_{k=1}^n q_{jk} x_j x_k \\ \text{s.a} & \sum_{j=1}^n a_{ij} x_j \leq b_i, \ \ i = 1,\ldots,m \\ & x_j \geq 0, \ \ j = 1,\ldots,n \end{array} \right. \tag{12}$$

where $c_j \in [(c_j)_\alpha^L, (c_j)_\alpha^U]$, for any $\alpha \in [0,1]$ chosen by decision maker.

These two formulations above are solved by outer-level and inner-level programs, respectively. The outer-level program obtains the values c_j that are used with parameters by inner-level program. The inner-level program solves a classical quadratic programming problem with the data obtained by outer-level program. The authors state that the formulation of two-level quadratic problems is a generalization of the conventional parametrical quadratic programming problem. Thus, the first two-level mathematical program can be transformed into the following quadratic problem by dual formulation:

$$\begin{aligned} \max \quad & \sum_{i=1}^m b_i \lambda_i - \tfrac{1}{2}\sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j \\ \text{s.t.} \quad & \sum_{i=1}^m a_{ij}\lambda_i + \delta_j - \sum_{i=1}^n q_{ij} x_i = c_j, \ \ j = 1,2,\ldots,n \\ & (c_j)_\alpha^L \leq c_j \leq (c_j)_\alpha^U, \ \ j = 1,2,\ldots,n \\ & \lambda_i, \delta_j \geq 0, \ \ i = 1,2,\ldots,m \text{ and } j = 1,2,\ldots,n. \end{aligned} \tag{13}$$

Since both the inner-level program and the outer-level of the second program have the same minimization operation, they can be combined into a conventional one-level program with the constraints of the two programs considered simultaneously. Consequently, some points must be analyzed which

ISBN: 978-989-95079-6-8

are shown in [11, 12]. The second program can be described as

$$\begin{array}{ll}\max & \sum_{j=1}^{n}(c_j)_{\alpha}^{L}+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}q_{ij}x_ix_j \\ \text{s.t.} & \sum_{i=1}^{n}a_{ij}x_i\leq b_i,\ \ i=1,2,\ldots,m \\ & x_j\geq 0,\ \ j=1,2,\ldots,n\end{array} \tag{14}$$

2.3 Ammar's approach

A method to solve a quadratic programming problem with fuzzy coefficients in the objective function and in the set of constraint is described in [1]. However, such a method is also valid when only the objective is fuzzy, that is, when only the matrix of costs are fuzzy, and for this reason it shall be considered here. This approach can be formulated in the following way:

$$\begin{array}{ll}\min & \mathbf{x^t}\tilde{\mathbf{Q}}\mathbf{x} \\ \text{s.t.} & \mathbf{Ax}\leq\mathbf{b} \\ & \mathbf{x}\geq\mathbf{0}.\end{array} \tag{15}$$

where all decision variables are non negatives and the linear term in objective function is not used. However, in this work only costs of the objective functions are imprecise.

Thus, Problem (15) can be reformulated by using α-cutting levels as

$$\begin{array}{lll}(P_\alpha): & \min & \mathbf{x^t}[\mathbf{Q}_\alpha^-,\mathbf{Q}_\alpha^+]\mathbf{x} \\ & \text{s.a} & \mathbf{Ax}\leq\mathbf{b} \\ & & \mathbf{x}\geq\mathbf{0}.\end{array} \tag{16}$$

where $\alpha\in(0,1]$. $\mathbf{Q}_\alpha^-$ represents a matrix with the lower values of the interval and $\mathbf{Q}_\alpha^+$ represents a matrix with the upper values of the interval.

[1] describes how Problem (16) can be written in two quadratic problems. Then, these problems are formulated as follows

$$\begin{array}{lll}(P_\alpha^-): & \min & \mathbf{x^t}\mathbf{Q}_\alpha^-\mathbf{x} \\ & \text{s.a} & \mathbf{Ax}\leq\mathbf{b} \\ & & \mathbf{x}\geq\mathbf{0}.\end{array} \tag{17}$$

$$\begin{array}{lll}(P_\alpha^+): & \min & \mathbf{x^t}\mathbf{Q}_\alpha^+\mathbf{x} \\ & \text{s.t.} & \mathbf{Ax}\leq\mathbf{b} \\ & & \mathbf{x}\geq\mathbf{0}.\end{array} \tag{18}$$

where the first problem uses the lower bound of the interval of the α-cutting level and the second problem uses the upper bound.

The two quadratic problems (17) and (18) can be solved by using the conditions of Karush-Kuhn-Tucker's optimality to each α value. The optimal solution of the original problem (16), (P_α), is inside the interval formed by the optimal solutions of each one problems (P_α^-) and (P_α^+).

3 Numerical example

In this section, we illustrate a quadratic programming problem that will be solved with three different way. Each way is represented by different fuzzy environment in each problem, i.e., the first problem has fuzzy costs in the linear term, while the second has fuzzy costs in the quadratic term, and the third has fuzzy costs in all coefficients in the objective function.

Example 1 *Consider the following quadratic programming problem with fuzzy coefficients in the linear term:*

$$\begin{array}{ll}\min & (-6,-5,-4)x_1+(1,1.5,2)x_2+2x_1^2-2x_1x_2+x_2^2 \\ \text{s.t.} & x_1+x_2\leq 2 \\ & 2x_1-x_2\leq 4 \\ & x_1,x_2\geq 0.\end{array} \tag{19}$$

The membership functions are:

$$\mu_{-5}(c_1)=\left\{\begin{array}{ll}c_1+6 & -6\leq c_1\leq -5\\ -c_1-4 & -5\leq c_1\leq -4\end{array}\right.$$

$$\mu_{1.5}(c_2)=\left\{\begin{array}{ll}2(c_2-1) & 1\leq c_2\leq 1.5\\ 2(2-c_2) & 1.5\leq c_2\leq 2\end{array}\right.$$

From the membership function for this costs an interval representation according to (3) can be given:

$$\begin{array}{ll}c_1^{0.2}=[-5.8,-4.2], & c_1^{0.4}=[-5.6,-4.4],\\ c_1^{0.6}=[-5.4,-4.6], & c_1^{0.8}=[-5.2,-4.8]\end{array}$$

$$\begin{array}{ll}c_2^{0.2}=[1.1,1.9], & c_2^{0.4}=[1.2,1.8],\\ c_2^{0.6}=[1.3,1.7], & c_2^{0.8}=[1.4,1.6]\end{array}$$

with $c_1^0=[-6,-4]$, $c_1^1=[-5,-5]$ *and* $c_2^0=[1,2]$, $c_2^1=[1.5,1.5]$, *and now* $\mathbb{M}=\{0,0.2,0.4,0.6,0.8,1\}$.

Thus, Problem (19) can be re-formulated as:

$$\begin{array}{ll}\min & (\alpha-6)x_1+\frac{1}{2}(\alpha+2)x_2+2x_1^2-2x_1x_2+x_2^2\\ \min & (\alpha-6)x_1+\frac{1}{2}(4-\alpha)x_2+2x_1^2-2x_1x_2+x_2^2\\ \min & (-\alpha-4)x_1+\frac{1}{2}(\alpha+2)x_2+2x_1^2-2x_1x_2+x_2^2\\ \min & (-\alpha-4)x_1+\frac{1}{2}(4-\alpha)x_2+2x_1^2-2x_1x_2+x_2^2\\ \text{s.t.} & x_1+x_2\leq 2\\ & 2x_1-x_2\leq 4\\ & x_1,x_2\geq 0.\end{array} \tag{20}$$

Problem (20) is solved by weighting objectives, with $\alpha=0.8$*, where an auxiliary problem can be obtained in the following way:*

$$\begin{array}{ll}\min & \omega_1[-5.2x_1+1.4x_2]+\omega_2[-5.2x_1+1.6x_2]+\\ & +\omega_3[-4.8x_1+1.4x_2]+\omega_4[-4.8x_1+1.6x_2]+\\ & +2x_1^2-2x_1x_2+x_2^2\\ s.t. & x_1+x_2\leq 2,\ \ 2x_1-x_2\leq 4,\ \ x_1,x_2\geq 0.\end{array} \tag{21}$$

Different optimal solutions to each of the weighting vectors is obtained by choosing the canonical base of $\mathbb{R}^4$ *as weighting vectors, as it is presented in Table 1.*

Table 1: Result of Problem 21 to each vector ω.

ω	x_1	x_2	FunObj
(1,0,0,0)	1.4600	0.5400	-3.8580
(0,1,0,0)	1.4800	0.5200	-3.7520
(0,0,1,0)	1.4200	0.5800	-3.2820
(0,0,0,1)	1.4400	0.5600	-3.1680

By taking $\omega_1=\omega_2=\omega_3=\omega_4=\frac{1}{4}$ *and solving (21) the optimal solution is* $x^*=[1.45,0.55]$ *and value of objective function is -3.5125.*

On the other hand, according to the method described in 2.2, Problem (19), which is the outer-level program, can be

ISBN: 978-989-95079-6-8

rewritten as:

$$\begin{array}{ll}\max & 2\lambda_1+4\lambda_2-\frac{1}{2}(4x_1^2-4x_1x_2+2x_2^2)\\ \text{s.t.} & \lambda_1+2\lambda_2-\delta_1-4x_1+2x_2=c_1\\ & \lambda_1-1\lambda_2-\delta_2+2x_1-2x_2=c_2\\ & -5.2\le c_1\le -4.8\\ & 1.4\le c_2\le 1.6\\ & x_1,x_2,\lambda_1,\lambda_2,\delta_1,\delta_2\ge 0.\end{array} \tag{22}$$

and the inner-level program can be written as:

$$\begin{array}{ll}\min & -5.2x_1+1.4x_2+\frac{1}{2}(4x_1^2-4x_1x_2+2x_2^2)\\ \text{s.t.} & x_1+x_2\le 2\\ & 2x_1-x_2\le 4\\ & x_1,x_2\ge 0.\end{array} \tag{23}$$

This model is a traditional quadratic program that can represent the two models described in 2.2. Hence, by solving the outer-level program the optimal solution is $x^*=[1.3,0]$ and value of objective function is -3.38. On the other hand, by solving the inner-level program the optimal solution is $x^*=[1.45,0.55]$ and value of objective function is -3.8575.

Example 2 *Consider the following quadratic programming problem with fuzzy coefficients in the quadratic term:*

$$\begin{array}{ll}\min & -5x_1+1.5x_2+(1,2,3)x_1^2+(1,2,3)x_1x_2+(0,1,2)x_2^2\\ \text{s.t.} & x_1+x_2\ge 2\\ & 2x_1-x_2\ge 4\\ & x_1,x_2\ge 0.\end{array} \tag{24}$$

The membership functions are

$$\mu_2(c_1)=\begin{cases} c_1-1 & 1\le c_1\le 2\\ 2-c_1 & 2\le c_1\le 3\end{cases}$$

$$\mu_1(c_2)=\begin{cases} c_2 & 0\le c_2\le 1\\ 2-c_2 & 1\le c_2\le 2\end{cases}$$

From the membership function for these costs an interval representation according to (4) can be given:

$$c_1^{0.2}=[1.2,2.8],\quad c_1^{0.4}=[1.4,2.6],$$
$$c_1^{0.6}=[1.6,2.4],\quad c_1^{0.8}=[1.8,2.2]$$

$$c_2^{0.2}=[0.2,1.8],\quad c_2^{0.4}=[0.4,1.6],$$
$$c_2^{0.6}=[0.6,1.4],\quad c_2^{0.8}=[0.8,1.2]$$

with $c_1^0=[1,3]$, $c_1^1=[2,2]$ and $c_2^0=[0,2]$, $c_2^1=[1,1]$, and now $\mathbb{M}=\{0,0.2,0.4,0.6,0.8,1\}$.

Thus, Problem (24) can be re-formulated as

$$\begin{array}{ll}\min & -5x_1+1.5x_2+(\alpha+1)x_1^2-(\alpha+1)x_1x_2+\alpha x_2^2\\ \min & -5x_1+1.5x_2+(\alpha+1)x_1^2-(\alpha+1)x_1x_2+(2-\alpha)x_2^2\\ \min & -5x_1+1.5x_2+(\alpha+1)x_1^2-(3-\alpha)x_1x_2+\alpha x_2^2\\ \min & -5x_1+1.5x_2+(\alpha+1)x_1^2-(3-\alpha)x_1x_2+(2-\alpha)x_2^2\\ \min & -5x_1+1.5x_2+(3-\alpha)x_1^2-(\alpha+1)x_1x_2+\alpha x_2^2\\ \min & -5x_1+1.5x_2+(3-\alpha)x_1^2-(\alpha+1)x_1x_2+(2-\alpha)x_2^2\\ \min & -5x_1+1.5x_2+(3-\alpha)x_1^2-(3-\alpha)x_1x_2+\alpha x_2^2\\ \min & -5x_1+1.5x_2+(3-\alpha)x_1^2-(3-\alpha)x_1x_2+(2-\alpha)x_2^2\\ \text{s.t.} & x_1+x_2\ge 2\\ & 2x_1-x_2\ge 4\\ & x_1,x_2\ge 0.\end{array} \tag{25}$$

Problem (25) is solved by weighting objectives, with $\alpha=0.8$, where an auxiliary problem like (21) can be obtained;

$$\begin{array}{ll}\min & -5x_1+1.5x_2+\omega 1(1.8x_1^2-1.8x_1x_2+0.8x_2^2)\\ & +\omega_2(1.8x_1^2-1.8x_1x_2+1.2x_2^2)+\omega_3(1.8x_1^2-2.2x_1x_2+0.8x_2^2)\\ & +\omega_4(1.8x_1^2-2.2x_1x_2+1.2x_2^2)+\omega_5(2.2x_1^2-1.8x_1x_2+0.8x_2^2)\\ & +\omega_6(2.2x_1^2-1.8x_1x_2+1.2x_2^2)+\omega_7(2.2x_1^2-2.2x_1x_2+0.8x_2^2)\\ & +\omega_8(2.2x_1^2-2.2x_1x_2+1.2x_2^2)\\ \text{s.t.} & x_1+x_2\ge 2,\ 2x_1-x_2\ge 4,\ x_1,x_2\ge 0.\end{array} \tag{26}$$

Different optimal solutions to each of the weighting vectors is obtained by choosing the canonical base of $\mathbb{R}^8$ as weighting vectors, as it is presented in Table 2.

Table 2: Result of Problem 21 to each vector ω.

ω	x_1	x_2	FunObj
(1,0,0,0,0,0,0,0)	1.5114	0.4886	-3.8506
(0,1,0,0,0,0,0,0)	1.5521	0.4479	-3.7630
(0,0,1,0,0,0,0,0)	1.4688	0.5312	-4.1547
(0,0,0,1,0,0,0,0)	1.5096	0.4904	-4.0505
(0,0,0,0,1,0,0,0)	1.3854	0.6146	-3.0130
(0,0,0,0,0,1,0,0)	1.2705	0.3279	-2.9303
(0,0,0,0,0,0,1,0)	1.3558	0.6442	-3.3582
(0,0,0,0,0,0,0,1)	1.4018	0.5982	-3.2040

On the other hand, the approach described in 2.3 can be obtained by using the first and last weighting vector of Table 2. Hence, by taking $\omega_L=(1,0,0,0,0,0,0,0)$ and solving (26) the optimal solution is $x^*=[1.5114,0.4886]$ and value of objective function is -3.8506. Besides, by taking $\omega_U=(0,0,0,0,0,0,0,1)$ and solving (26) the optimal solution is $x^*=[1.4018,0.5982]$ and value of objective function is -3.2040.

Example 3 *Consider the following quadratic programming problem with all fuzzy coefficients:*

$$\begin{array}{ll}\min & (-6,-5,-4)x_1+(1,1.5,2)x_2+(1,2,3)x_1^2+\\ & +(1,2,3)x_1x_2+(0,1,2)x_2^2\\ \text{s.t.} & x_1+x_2\ge 2\\ & 2x_1-x_2\ge 4\\ & x_1,x_2\ge 0.\end{array} \tag{27}$$

The membership functions are

$$\mu_{-5}(c_1)=\begin{cases} c_1+6 & -6\le c_1\le -5\\ -c_1-4 & -5\le c_1\le -4\end{cases}$$

$$\mu_{1.5}(c_2)=\begin{cases} 2(c_2-1) & 1\le c_2\le 1.5\\ 2(2-c_2) & 1.5\le c_2\le 2\end{cases}$$

$$\mu_2(c_1)=\begin{cases} c_1-1 & 1\le c_1\le 2\\ 2-c_1 & 2\le c_1\le 3\end{cases}$$

$$\mu_1(c_2)=\begin{cases} c_2 & 0\le c_2\le 1\\ 2-c_2 & 1\le c_2\le 2\end{cases}$$

From the membership function for these costs an interval

ISBN: 978-989-95079-6-8

representation according to (3) and (4) can be given:

$$c_1^{0.2} = [-5.8, -4.2], \quad c_1^{0.4} = [-5.6, -4.4],$$
$$c_1^{0.6} = [-5.4, -4.6], \quad c_1^{0.8} = [-5.2, -4.8]$$

$$c_2^{0.2} = [1.1, 1.9], \quad c_2^{0.4} = [1.2, 1.8],$$
$$c_2^{0.6} = [1.3, 1.7], \quad c_2^{0.8} = [1.4, 1.6]$$

$$c_3^{0.2} = [1.2, 2.8], \quad c_3^{0.4} = [1.4, 2.6],$$
$$c_3^{0.6} = [1.6, 2.4], \quad c_3^{0.8} = [1.8, 2.2]$$

$$c_4^{0.2} = [0.2, 1.8], \quad c_4^{0.4} = [0.4, 1.6],$$
$$c_4^{0.6} = [0.6, 1.4], \quad c_4^{0.8} = [0.8, 1.2]$$

with $c_1^0 = [-6, -4]$, $c_1^1 = [-5, -5]$, $c_2^0 = [1, 2]$, $c_2^1 = [1.5, 1.5]$, $c_3^0 = [1, 3]$, $c_1^3 = [2, 2]$ *and* $c_4^0 = [0, 2]$, $c_4^1 = [1, 1]$, *and now* $\mathbb{M} = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

Different optimal solutions to each of the weighting vectors can be showed by the multi-objective method described choosing the canonical base of $\mathbb{R}^{32}$ *as weighting vectors. According to the previous examples, it can obtain the minimum and maximum values that form the optimal interval of this problem. The minimum value is -4.5021 where the optimal solution is* $[1.4792, 0.5208]$ *while the maximum value of this optimal interval is -2.6579 where the optimal solution is* $[1.1803, 0.2186]$.

4 Conclusions

This paper presented three approaches for solving fuzzy quadratic programming problems. A novel approach that transforms a quadratic programming problem with imprecise costs in the objective function into a parametric multiobjective quadratic programming problem. The other two approaches are methods published that are used in this work for comparison. Quadratic Programming is very important in theoretical and practical areas. When real-world applications are considered, vagueness appears in a natural way, and hence it makes perfect sense to think of fuzzy quadratic programming problems. In contrast to what happens in fuzzy linear programming problems, unfortunately until now not much research has been conducted to this important class of problems.

It was shown that the parametric multiobjective programming problem can obtain the optimal solutions of the other approaches because these solutions are a quadratic or linear combination of the optimal solutions obtained for each of weighting vectors. Hence, this shows that the parametric multiobjective approach contains the optimal solutions obtained by the others. Therefore, parametric approaches can be used as a general method to solve quadratic programming problems with uncertain costs in the objective function.

The authors aim to extend the line of investigation involving Fuzzy Quadratic Programming problems in order to try to solve practical real-life problems by facilitating the building of Decision Support Systems.

Acknowledgment

The authors want to thank the support provided by the Brazilian agency CAPES and the Spanish projects TIN2005-08404-C04-01 (70% of which are FEDER funds) and MINAS (TIC-00129-JA) for the support provided.

References

[1] E. Ammar and H. A. Khalifa. Fuzzy portfolio optimization a quadratic programming approach. *Chaos, Solitons and Fractals*, 18:1045–1054, 2003.

[2] S.S. Appadoo, S.K. Bhatt, and C.R. Bector. Application of possibility theory to investment decisions. *Fuzzy Optimization and Decision Making,*, 7:35–57, 2008.

[3] E. M. L. Beale. On quadratic programming. *Naval Research Logistics Quarterly*, 6:227–244, 1959.

[4] R. E. Bellman and L. A. Zadeh. Decision-marking in a fuzzy environment. *Management Science*, 17(4):B141–B164, 1970.

[5] G.R. Bitran. Linear multiple objective problems with interval coefficients. *Management Sciences*, 26(7):694–706, 1980.

[6] M. Delgado, J.L. Verdegay, and M.A. Vila. Imprecise costs in mathematical programming problems. *Control and Cybernetics*, 16(2):113–121, 1987.

[7] M. Delgado, J.L. Verdegay, and M.A. Vila. Relating different approaches to solve linear programming problems with imprecise costs. *Fuzzy Sets and Systems*, 37:33–42, 1990.

[8] C. A. Floudas, P.M. Pardalos, C. Adjiman, W. R. Esposito, Z. H. Gümüs, S. T. Harding, J. L. Klepeis, C. A. Meyer, and C. A. Sshweiger. *Handbook of test problems in local and global optimization*, volume 33 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 1999.

[9] W. Hock and K. Schittkowski. *Test examples for nonlinear programming codes*, volume 187 of *Lecture Notes in Economics and Mathematical Systems*. Spring-Verlag, 1981.

[10] Y. J. Lai and C. L. Hwang. *Fuzzy mathematical programming : methods and applications*, volume 394 of *Lecture notes in economics and mathematical systems*. Springer, Berlin, 1992.

[11] S. T. Liu. Quadratic programming with fuzzy parameters: A membership function approach. Chaos, Solitons and Fractals. Article in press. doi:10.1016/j.chaos.2007.07.054, 2007.

[12] S. T. Liu. Solving quadratic programming with fuzzy parameters based on extension principle. In *FUZZ-IEEE 2007*, pages 1–5, London, UK, 2007. IEEE International Conference Fuzzy Systems.

[13] C. V. Negoita and D. A. Ralescu. *Applications of fuzzy sets to systems analysis*. Birkhauser Verlag, Stuttgard, 1975.

[14] D. Poole, A. Mackworth, and R. Goebel. *Computational intelligence: A logical approach*. Oxford University Press, 1998.

[15] S. J. Russel and P. Norvig. *Artificial intelligence: A modern approach*. Prentice Hall, Upper Saddle River, NJ, second edition, 2003.

[16] K. Schittkowski. *More test examples for nonlinear programming codes*, volume 282 of *Lecture Notes in Economics and Mathematical Systems*. Spring-Verlag, 1987.

[17] J. L. Verdegay, R. R. Yager, and P. P. Bonissone. On heuristics as a fundamental constituent of soft computing. *Fuzzy Sets and Systems*, 159(7):846–855, 2008.

[18] S. Wang and S. Zhu. On fuzzy portfolio selection problems. *Fuzzy Optimization and Decision Making,*, 1:361–377, 2002.

[19] P. Wolfe. The simplex method for quadratic programming. *Econometrica*, 27:382–398, 1959.

[20] L. A. Zadeh. A new direction in AI: Toward a computational theory of perceptions. *AI Magazine*, 22(1):73–84, 2001.

ISBN: 978-989-95079-6-8

Comparison of Multicriteria Methods for Land-use Suitability Assessment

Jozo Dujmović[1] Guy De Tré[2] Suzana Dragićević[3]

1.Department of Computer Science, San Francisco State University,
1600 Holloway Ave, San Francisco, CA 94132, U.S.A.
2.Department of Telecommunications and Information Processing Ghent University,
St.-Pietersnieuwstraat 41, B-9000 Ghent, Belgium
3.Department of Geography, Simon Fraser University,
8888 University Drive, Burnaby, BC V5A 1S6, Canada
Email: jozo@sfsu.edu, Guy.DeTre@UGent.be, suzanad@sfu.ca

Abstract— *In this paper we investigate properties of multicriteria methods that are used for building land-use suitability assessment criteria. We identify and describe fundamental properties that are of interest in the land-use suitability analysis and the design of suitability maps. The existing multicriteria methods can be evaluated from the standpoint of their ability to support the desirable properties that affect the expressive power of evaluation methods. In this paper we investigate simple additive scoring, MAVT, MAUT, AHP, OWA, outranking methods and LSP.*

Keywords— GIS, land-use, MCDM.

1 Introduction

This paper has two main goals. The first goal is to identify properties that affect the expressive power of multicriteria decision methods (MCDM) that can be used to create suitability maps of a specific geographic region. The second goal is to investigate land-use suitability assessment methods that are based on simple additive scoring (SAS), multiattribute value technique (MAVT), multiattribute utility technique (MAUT), analytic hierarchy process (AHP), ordered weighted average (OWA), outranking methods (ELECTRE, PROMETHEE), and logic scoring of preference (LSP) in view of satisfaction of the identified properties.

We assume that the analyzed geographic region is divided into cells and that each cell with coordinates x, y is characterized by an array of suitability attribute values $a_1(x,y), \ldots, a_n(x,y)$, $n \geq 1$ [5]. The attributes are defined as quantitative parameters that affect the suitability of a cell for some specific land-use (e.g. housing, recreation, agriculture, industrial development, etc.). The set of attributes must be complete, i.e., it must include *all* relevant components. Generally, the attributes must be justifiable and not redundant with each other.

The purpose of MCDM is to provide a criterion function $\sigma : \mathbb{R}^n \rightarrow [0,1]$ for computing an overall degree of suitability $S(x,y) = \sigma(a_1(x,y), \ldots, a_n(x,y))$ that reflects the suitability of location x, y for specific land-use. The overall suitability is a matter of degree: $0 \leq S(x,y) \leq 1$. As in all soft computing models, here 0 denotes a completely unsuitable location, and 1 denotes the highest level of suitability. A suitability map is defined as a distribution of the overall suitability $S(x,y)$ for a specific geographic region [7].

Nowadays suitability maps are assumed to be dynamically created using data from GIS databases in a way illustrated in Fig. 1. The multicriteria decision model must be interfaced both with the user and with the GIS database. An attribute ETL interface is necessary to Extract, Transform, and Load the set of cell attribute values from the GIS database. The multicriteria decision model is used to implement the criterion function σ and to generate the overall suitability $S(x,y)$. The user input includes a specification of desired suitability attributes and parameters of the decision model. A suitability criterion interface is necessary for accepting user input and for rendering the resulting suitability map.

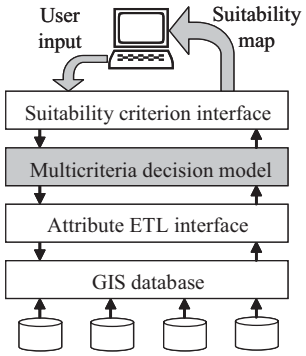


Figure 1: Dynamic generation of suitability maps using GIS.

Suitability maps introduced in GIS literature use a variety of decision models [16, 9, 11, 5, 20]. The emphasis of such efforts is primarily (and naturally) on the selection of attributes and the use of suitability maps. So far the GIS literature avoided problems of evaluating the credibility of MCDM used for the development of suitability maps.

Except for interfacing with GIS components, the land-use suitability assessment problems do not differ from evaluation problems in other areas. All multicriteria decision models are ultimately models of human decision making in the area of evaluation of complex alternatives. Therefore, the credibility of MCDM used in GIS context depends on their ability to express observable properties of human evaluation logic [4]. The primary purpose of this paper is to evaluate the expressive power of GIS-related MCDM based on the level of their ability to support the concepts of human evaluation logic.

The remainder of this paper is structured as follows. In Section 2, the logic properties of GIS-related MCDM are studied. In Section 3, the logic properties of the selected decision meth-

ISBN: 978-989-95079-6-8

ods are investigated. The major contributions of the paper are summarized and some conclusions are presented in Section 4.

2 Ten fundamental properties of MCDM

The properties of human evaluation logic are easily observable, and they are also observable in the context of land-use evaluation. Our goal is to first identify properties that are frequently encountered in all worth assessment problems. The most important properties include the following:

1. Ability to combine any number of attributes.
2. Ability to combine objective and subjective inputs.
3. Ability to combine absolute and relative criteria.
4. Flexible adjustment of relative importance of attributes.
5. Modelling of simultaneity requirements.
   (a) Modelling of soft simultaneity.
   (b) Modelling of hard simultaneity.
6. Modelling of replaceability requirements.
   (a) Modelling of soft replaceability.
   (b) Modelling of hard replaceability.
7. Modelling of balanced simultaneity/replaceability.
8. Modelling of mandatory, desired, and optional requirements.
9. Modelling of sufficient, desired, and optional requirements.
10. Ability to express suitability as an aggregate of usefulness and inexpensiveness (separation of the usefulness analysis and the cost analysis).

2.1 Ability to combine any number of attributes

We assume that the analyzed geographic region is divided into cells and that each cell is characterized by attributes $a_1, \ldots, a_n$, $n \geq 1$. In special cases $n = 1$. For example, the acoustic pollution (noise) maps presented in [11] uses in each cell a single scalar value, the level of noise. However, in all cases of more complex suitability maps we have $n > 1$. For example, the housing suitability maps discussed in [5] use $n = 11$. If we want to develop maps of complex suitability indicators that depend on variety of inputs it is reasonable to expect that the number of inputs can be large. So, in a general case, MCDM used to compute the land-use suitability must be able to support any number of inputs.

It should be noted that this requirement indirectly transforms into a requirement to use nonlinear MCDM models. Indeed, in the case of linear weighted aggregation models, the overall suitability $S(x, y)$ is a dot product of a vector of n suitability components $\mathbf{s} = [s_1(x, y), \ldots, s_n(x, y)]$ (one component for each attribute) and a corresponding vector of relative weights $\mathbf{W} = [W_1, \ldots, W_n]$: $S(x, y) = \sum_{i=1}^{n} W_i s_i(x, y)$ where $0 \leq s_i(x, y) \leq 1$, $0 < W_i < 1$, $i = 1, \ldots, n$, and $W_1 + \cdots + W_n = 1$.

Obviously, a large value of n yields a low significance of individual components. For example, if $n = 100$ the average significance (the ability to change the overall suitability S) of each input is only 1%, and in the case of unequal weights some inputs will have the significance considerably below 1% which means that such inputs may (and should) be neglected. This is not acceptable since credible decision models must be equally applicable for any number of inputs.

2.2 Ability to combine objective and subjective inputs

Some input attributes are objectively measurable values (e.g., distances, slopes, altitudes, temperatures, etc). Other inputs are subjective and must be assessed by experts (e.g., aesthetic quality of an area, the quality of educational and/or medical institutions, quality of public transportation, development costs, etc.). MCDM for GIS applications must be able to combine and aggregate objective and subjective inputs.

2.3 Ability to combine absolute and relative criteria

Each suitability attribute must satisfy specific requirements. The requirements can be defined as elementary criterion functions $g_i : \mathbb{R} \rightarrow [0, 1]$, $i = 1, 2, \ldots, n$. By definition, the attribute suitability $s_i(x, y) = g_i(a_i(x, y))$ is the degree of satisfaction of the attribute requirements. The suitability of an attribute is a component of the overall suitability of the evaluated location x, y.

Suppose that a decision maker (DM) needs a suitability map for housing in a rural area where it is possible to buy land and build a house. Let the attribute $a_i(x, y) = t$ denote the traveling time between location $L(x, y)$ and the closest elementary school (or hospital, or airport, or any other point of interest). An absolute criterion is a criterion that evaluates location L regardless of other competitive locations. E.g., a DM may specify the absolute requirements as threshold times t_{min} and t_{max} so that all times $t \geq t_{max}$ are considered unacceptable and all times $t \leq t_{min}$ are considered perfectly acceptable. Then the elementary criterion function might be defined as follows: $g_i(t) = \max(0, \min(1, (t_{max} - t)/(t_{max} - t_{min})))$. So, $g_i(t)$ specifies the attribute suitability for any value of t. If DM wants to compare locations L_1 and L_2, with respect to the access to school then the comparison can be based on $g_i(t_1)$ and $g_i(t_2)$. If t_{min} and t_{max} are well justifiable values, then $g_i(t_1)$ and $g_i(t_2)$ are very credible results.

Another approach is to define a relative criterion. In the case of k locations an example of such a criterion could be $g_i(t) = t_{min}/t$, $t_{min} = \min(t_1, \ldots, t_k)$. The relative criterion evaluates relationships between competitors regardless of DM's actual needs: the closest location is considered perfectly suitable ($g_i(t_{min}) = 1$) and a location where $t = 2t_{min}$ gets $g_i(t) = 0.5$. Obviously, the credibility of relative criteria is much less under DM's control than the credibility of corresponding absolute criteria. E.g., if only two locations are available, then the location that is 2 minutes from the closest school is likely to be equally attractive as the location that is 1 minute from the school, and locations that are 1 hour and 2 hours from the closest school might be equally unacceptable; in both cases the suitability levels of 1 and 0.5 are inappropriate. For some other values, e.g., $t_{min} = 7$ minutes, the presented relative criterion might be appropriate.

It is highly unlikely that DM knows what is better, but does not know what is good. In other words, if DM can specify a justifiable relative criterion, it is very likely that DM can also specify a justifiable absolute criterion. Absolute criteria can be used to evaluate a single alternative (e.g., a single location), while relative criteria can be applied only if we have multiple alternatives. Relative criteria are appropriate only in situations where expected values of attributes are unknown. Such situations are not frequent, but exist, and MCDM must be able to combine and aggregate absolute and relative attribute criteria.

ISBN: 978-989-95079-6-8

2.4 Flexible adjustment of the relative importance of attributes

In a general case, multiple attributes are not equally important. The relative importance of an attribute is usually expressed using multiplicative or implicative weights [3]. The relative importance has two roles in suitability criteria: it defines the level of contribution of an attribute to the overall suitability, and it defines compensatory properties between attributes. Assuming that the overall suitability is a function of attribute suitability degrees, $S = \lambda(s_1, \ldots, s_j, \ldots, s_k, \ldots, s_n)$, we expect that if the suitability s_j is more important than the suitability s_k, then $\partial S/\partial s_j \geq \partial S/\partial s_k$. Under the same assumptions, if some compensation is possible, then $\lambda(s_1, \ldots, s_j, \ldots, s_k, \ldots, s_n) = \lambda(s_1, \ldots, s_j - p, \ldots, s_k + q, \ldots, s_n)$ and $p < q$. In other words, a suitability decrement p of an attribute can be compensated by the suitability increment q of a less important attribute, but the increment q must be greater than the decrement p. These properties are essential in human reasoning and must be supported by MCDM.

2.5 Modelling of simultaneity requirements

Function $\lambda : [0,1]^n \to [0,1]$ that aggregates all attribute suitability degrees and computes the overall suitability degree is essentially a logic function. In human decision making the aggregation of suitability degrees is usually a stepwise process where small groups of related suitability degrees are aggregated and replaced by an aggregated suitability degree. The process of stepwise aggregation terminates when the suitability degrees of subsystems at the highest level are aggregated yielding the overall suitability degree S as a compound function of attribute suitability degrees.

Let us investigate an aggregation step where DM aggregates m suitability degrees using an aggregation function $\mu : [0,1]^m \to [0,1]$. In human decision making $\mu(s_1, \ldots, s_m)$ is most frequently a model of simultaneity. E.g., a homebuyer regularly prefers locations that are both close to schools for children *and* to jobs for parents. Such a simultaneity requirement can be modelled using some form of partial conjunction [3], i.e., μ is expected to have an adjustable degree of similarity with conjunction $s_1 \wedge \cdots \wedge s_m$. A normalized degree of proximity between $\mu(s_1, \ldots, s_m)$ and $s_1 \wedge \cdots \wedge s_m$ is called andness and it can be defined by: $\alpha = \frac{(s_1 \vee \cdots \vee s_m) - \mu(s_1, \ldots, s_m)}{(s_1 \vee \cdots \vee s_m) - (s_1 \wedge \cdots \wedge s_m)}$. Since $(s_1 \wedge \cdots \wedge s_m) \leq \mu(s_1, \ldots, s_m) \leq (s_1 \vee \cdots \vee s_m)$ it follows that $0 \leq \alpha \leq 1$ and $\mu(s_1, \ldots, s_m)$ is a model of simultaneity if $0.5 < \alpha \leq 1$ and $(s_1 \wedge \cdots \wedge s_m) \leq \mu(s_1, \ldots, s_m) < (s_1 + \cdots + s_m)/m$. This form of andness depends on input degrees of suitability and humans usually think globally using an average andness, e.g., $\overline{\alpha} = \int_{[0,1]^m} \alpha(s_1, \ldots, s_m) ds_1 \ldots ds_m$, $0 \leq \overline{\alpha} \leq 1$. If DM needs a high simultaneity (e.g., $\overline{\alpha} > 0.7$), that usually means that all inputs must be at least partially satisfied. In other words, $\mu(s_1, \ldots, s_m) = 0$, $s_i = 0$, $i \in \{1, \ldots, m\}$ and it is mandatory to satisfy all inputs. This kind of simultaneity is called a *hard simultaneity*. In other cases DM may need a *soft simultaneity*, where the average andness is slightly above 0.5 and $\mu(s_1, \ldots, s_m) > 0$, $s_i > 0$, $i \in \{1, \ldots, m\}$. Supporting models of simultaneity is an important requirement that MCDM must satisfy.

2.6 Modelling of replaceability requirements

Replaceability requirements are symmetrical and complementary to simultaneity requirements. Replaceability means that a high suitability in a group of attributes can be achieved using any one of the attributes (i.e., they can replace each other). E.g., a home location can be considered suitable for recreational activities if it is close to lake, *or* close to ski terrains. The intensity of replaceability can be determined using the orness indicator that is a complement of andness: $\overline{\omega} = 1 - \overline{\alpha}$. Replaceability aggregators satisfy conditions $0.5 < \overline{\omega} \leq 1$ and $(s_1 + \cdots + s_m)/m < \mu(s_1, \ldots, s_m) \leq (s_1 \vee \cdots \vee s_m)$. High orness means low andness and vice versa. Similarly to the case of simultaneity, a high level of replaceability (e.g., $\overline{\omega} > 0.7$) may be combined with the requirement for hard replaceability where $\mu(s_1, \ldots, s_m) = 1$, $s_i = 1$, $i \in \{1, \ldots, m\}$. Soft replaceability is any form of replaceability that does not satisfy the hard replaceability requirements. In the case of soft replaceability we have $\mu(s_1, \ldots, s_m) < 1$, $s_i < 1$, $i \in \{1, \ldots, m\}$.

2.7 Modelling of balanced simultaneity/replaceability

If the simultaneity and replaceability are balanced then $\alpha = \omega = 0.5$. In the case of two variables, from definition $\alpha = \frac{(s_1 \vee s_2) - \mu(s_1, s_2)}{(s_1 \vee s_2) - (s_1 \wedge s_2)} = 0.5$ it follows that $\mu(s_1, s_2) = ((s_1 \vee s_2) + (s_1 \wedge s_2))/2 = (s_1 + s_2)/2$. This result indicates that the arithmetic mean is a soft computing logic function that combines simultaneity and replaceability requirements in a balanced way: DM would like that all attributes are simultaneously satisfied, but at the same time s/he accepts that any attribute can compensate any other attribute. It is important to understand that the arithmetic mean represents a model of this specific logic condition and nothing more. MCDM that use the arithmetic mean are acceptable only in cases where DM can justify the use of this specific logic condition.

According to Malczewski [16], "GIS implementations of the weighted summation procedures are often used without full understanding of the assumptions underlying this approach."

2.8 Modelling of mandatory, desired, and optional requirements

Using models of simultaneity, replaceability and negation ($x \mapsto 1 - x$) it is possible to create a variety of compound soft computing logic functions that precisely reflect the needs of DM. A compound aggregator that is most frequent in human reasoning is used to combine mandatory and nonmandatory attributes. Most frequently there are one mandatory attribute and one or two nonmandatory attributes. E.g., a DM may reject home locations that do not have good ground transportation, but accept locations that are far from an international airport. In such cases the ground transportation is a mandatory requirement, and the vicinity of an international airport is desired, but not mandatory. Optional attributes are also nonmandatory and have lower significance than desired attributes. To model such requirements we need aggregators $\mu(s_{man}, s_{des}, s_{op})$ that satisfy the conditions $\mu(s_{man}, s_{des}, s_{op}) = 0$, $s_{man} = 0$, $s_{des} > 0$, $s_{op} > 0$ and if the compensation between s_{des} and s_{op} is possible, then $\mu(s_{man}, s_{des}, s_{op}) = \mu(s_{man}, s_{des} - p, s_{op} + q)$, $s_{man} > 0$, $p < q$. Optional attributes can be omitted, and in such cases

ISBN: 978-989-95079-6-8

we use only mandatory and desired inputs. An aggregator with these properties is a partial absorption function introduced in [2] and expanded in [19].

2.9 Modelling of sufficient, desired, and optional requirements

Sufficient, desired and optional requirements are a disjunctive counterpart of mandatory, desired and optional requirements. If the sufficient input is completely satisfied then the desired and optional inputs have no effect. If the sufficient attribute has low or even zero suitability degree, this can be partially compensated by the desired and optional attributes. The corresponding aggregator $\mu(s_{suf}, s_{des}, s_{op})$ satisfies conditions $\mu(s_{suf}, s_{des}, s_{op}) = 1, s_{suf} = 1, s_{des} < 1, s_{op} < 1$, and if the compensation between s_{des} and s_{op} is possible, then $\mu(s_{suf}, s_{des}, s_{op}) = \mu(s_{suf}, s_{des} - p, s_{op} + q)$, $s_{suf} < 1$, $p < q$.

2.10 Ability to express suitability as an aggregate of usefulness and inexpensiveness

Land-use is regularly related to a variety of costs (e.g., the cost of land, the cost of building infrastructure and objects, the cost of financing, etc.). The overall suitability depends on two simultaneous requirements: finding locations that are very useful for specific purpose and at the same time inexpensive. A convenient way to solve that problem is to define *usefulness* as a non-financial part of the overall suitability, and *inexpensiveness* as an overall result of cost analysis (an aggregate of cost components only). Then, the overall suitability can be conveniently expressed as an aggregate of usefulness and inexpensiveness. Separation of cost and usefulness attributes reflects human reasoning where the overall cost is compared with the corresponding overall satisfaction. Of course, there are cases where the overall suitability does not depend on cost. In such cases the overall suitability reduces to usefulness.

The presented list of properties is not proved to be necessary and sufficient in all cases. However, the presented conditions are relevant for many land use decision problems and show that logic aggregation of attribute suitabilities is a frequently needed property. For detailed analysis of mathematical conditions see [4], and for sample applications see [4, 5].

The necessary properties of MCDM's do not change if attributes of a cell are functions of time, or functions of the values of attributes in other cells.

3 Properties of GIS related MCDM

Several GIS related MCDM approaches have been presented in literature [15, 16]. Among the decision methods used in these approaches we selected a representative set that consists of the following techniques: *simple additive scoring* (SAS) [8, 6], *multiattribute value technique* (MAVT) [22, 13, 14], *multiattribute utility technique* (MAUT) [13, 10], *analytic hierarchy process* (AHP) [18, 1], *ordered weighted average* (OWA) [21, 17], *outranking methods* [12] and *logic scoring of preference* (LSP) [4, 5].

The ten fundamental features presented in the previous section can be used to investigate the selected decision methods in view of their appropriateness for land-use evaluation and suitability map construction.

3.1 Simple additive scoring

The SAS technique [8, 6] is based on the concept of a weighted average in which weights are used to denote relative importance of suitability attributes. A DM directly assigns a weight w_i to each suitability attribute a_i, $i = 1, \ldots, n$. These assigned weights are rescaled to normalized weights W_i, $i = 1, \ldots, n$, such that $\sum_{i=1}^{n} W_i = 1$. The overall score or overall degree of suitability $S(x, y)$ of each cell x, y is then computed by:

- Determining the n suitability components $s_1(x, y), \ldots, s_n(x, y)$ that are obtained from the evaluation of the n attributes for the cell.
- Multiplying each suitability component $s_i(x, y)$ with the normalized weight W_i of its corresponding attribute.
- Summing the products over all attributes, i.e., $S(x, y) = \sum_{i=1}^{n} W_i s_i(x, y)$.

Therefore, SAS uses a simple linear weighted aggregation model.

3.2 Multiattribute value technique

In MAVT [22, 13, 14] suitability attributes a_i, $i = 1, \ldots, n$ are evaluated using value functions that aim to mathematically represent human judgements. A single-attribute value function translates the performance of the alternative attribute values into a value score which represents the degree to which a decision objective is achieved. As such, value function v_i associates a number (or 'value') $v_i(a)$ with each alternative value a of attribute a_i in such a way that a preference order on the alternatives consistent with DM value judgements is obtained.

For aggregation, more complex value functions are used. The most commonly used function is the simple additive weighting function $S(x, y) = \sum_{i=1}^{n} W_i v_i(a_i(x, y))$ where W_i is the weight of suitability attribute a_i and $v_i(a_i(x, y))$ is the value of suitability attribute value $a_i(x, y)$ (of cell x, y). This approach is valid if suitability attributes are preferentially independent.

The weights W_i are scaling constants that have to be derived with reference to the attribute ranges. These need to be elicited through questions which capture acceptability of trade-offs (e.g., 'how many units of one suitability attribute are worth how many of units of another suitability attribute?'). Weights must sum up to 1, i.e., $\sum_{i=1}^{n} W_i = 1$.

As such, the MAVT approach is similar to the 'scoring method', except that the scores $s_i(x, y)$ are replaced by values $v_i(a_i(x, y))$ that are obtained with the value function v_i.

3.3 Multiattribute utility technique

MAUT [13, 10] is used and treated separately from MAVT when 'risks' or 'uncertainties' have a significant role in the definition and assessment of alternatives. The attitude of the DM toward risk is incorporated into the assessment of a single-attribute utility function u_i, which is obtained through utility analysis and translates the values of suitability attribute a_i into 'utility units'. A 'utility unit' is a relative value between 0 and 1 (where 0 and 1 resp. denote the worst and best values). The concept of a utility function is inherently probabilistic in nature.

ISBN: 978-989-95079-6-8

Aggregation is done by an overall utility function. A simple additive weighting function is most commonly used. In such a case $S(x,y) = \sum_{i=1}^{n} W_i u_i(a_i(x,y))$. Suitability attributes must be preferentially independent. The weights W_i, $i = 1, \ldots, n$ have to sum up to 1, i.e., $\sum_{i=1}^{n} W_i = 1$.

From aggregation point of view, the MAUT approach is similar to the MAVT and 'scoring' approaches.

3.4 Analytic hierarchy process

AHP [18, 1] uses a different approach. Cognitive psychology has found that people are poor at assimilating large quantities of information on problems. The subsequent steps in AHP can be summarized as follows:

- Model the problem as a hierarchy containing the decision goal, the alternatives for reaching it, and the suitability attributes for evaluating the alternatives.
- Establish priorities (normalized weights) among the elements of the hierarchy by making a series of judgments based on pairwise comparisons of the elements.
- Synthesize these judgments to yield a set of overall weights for the hierarchy. This is done by means of a sequence of multiplications of the matrices of relative weights at each level of the hierarchy.
- Check the consistency of the judgments.
- Come to a final decision based on the results of this process.

Suitability attributes are subdivided in subattributes and hierarchically structured. Overall suitability $s(x,y)$ for an attribute a is computed by an additive weighting function $s(x,y) = \sum_{j=1}^{m} W_{s_j(x,y)} s_j(x,y)$ where $s_j(x,y)$, $j = 1, \ldots, m$ denote the suitability components of the subattributes of a. All weights $W_{s_j(x,y)}$ must to sum up to 1, i.e., $\sum_{i=j}^{n} W_{s_j(x,y)} = 1$.

3.5 Ordered weighted averaging

In an OWA approach [17], the DM specifies the decision-relevant suitability attributes to be used as evaluation criteria; identifies preferred criteria values on a qualitative scale; and defines the relative importance of each criterion by assigning weights. The weighted criterion values are then combined using an OWA aggregation operator [21], resulting in an evaluation score for each cell. OWA allows the DM to specify a decision strategy that reflects his decision-related preferences.

The OWA operators [21], provide a parameterized class of mean type aggregation operators. Many notable mean operators such as the max, arithmetic average, median and min, are members of this class. OWA operators allow to model linguistically expressed aggregation instructions.

An OWA operator of dimension n is a mapping function $F : [0,1]^n \rightarrow [0,1]$ that has an associated collection of weights $\mathbf{W} = [W_1, \ldots, W_n] \in [0,1]^n$, for which it holds that $\sum_{i=1}^{n} W_i = 1$, and with $F(s_1, \ldots, s_n) = \sum_{i=1}^{n} W_i s'_i$ where s'_i is the ith largest value of the s_i.

By choosing different $\mathbf{W}$, different aggregation operators can be implemented. The OWA operator is a non-linear operator as a result of the process of determining the values s'_i.

3.6 Outranking methods

Outranking methods, such as variants of ELECTRE and PROMETHEE are used in the areas of GIS and environmental planning [12]. The basic idea of these methods is to use strictly relative criteria that express a range from indifference to strong preference of one alternative over another alternative separately for all individual attributes. The attribute preferences are then averaged using the arithmetic mean to generate the credibility of the outranking relation of two alternatives (PROMETHEE). In such a case no input is mandatory. The overall degree of outranking of two alternatives can also be computed using a product (ELECTRE III), making all attributes mandatory.

3.7 Logic scoring of preference

With the LSP method [4, 5] suitability maps are created with the following main steps:

1. Creation of an attribute tree. This tree contains and structures all parameters that affect the overall suitability and is build by the DM.
2. Definition of elementary criteria. The DM has to provide an elementary criterion for each attribute involved in the decision process. These criteria will be evaluated during suitability map construction. For each analyzed cell x, y, the evaluation of each attribute a_i, $i = 1, \ldots, n$ will result in an elementary satisfaction degree $s_i(x,y)$.
3. Creation of the aggregation structure. For each analyzed cell, all associated elementary satisfaction degrees must be aggregated. Therefore, the DM has to create an aggregation structure, which adequately reflects his domain knowledge and reasoning.
4. Computation of the overall suitability degree. Once the attribute tree, the elementary criteria and the aggregation structure are available, the suitability map construction can start. The elementary criteria are evaluated and their resulting elementary satisfaction degrees can be aggregated in order to compute the overall satisfaction degree of each analyzed cell.

Aggregation in LSP is done via the aggregation structure, build by the DM. The basic building blocks of the aggregation structure are the simple and compound LSP aggregators. The DM can use them to construct an easily understandable aggregation schema which is consistent with observable properties of human reasoning in the area of evaluation.

The simple LSP aggregators are all graded preference logic functions and based on a superposition of the fundamental *Generalized Conjunction/Disjunction* (GCD) function. Most frequently, GCD is implemented by the weighted power mean function $GCD(s_1, \ldots, s_n) = (W_1 s_1^r + \cdots + W_n s_n^r)^{1/r}$ where $r \in [-\infty, +\infty]$ and $0 < W_i < 1$ such that $\sum_{i=1}^{n} W_i = 1$ [5]. The parameter r determines the logical behavior of the function. As such a continuous variety of logical functions ranging from full conjunction to full disjunction can be modelled. The simple LSP aggregators can be used to construct more complex, compound operators like the conjunctive partial absorption which can be used to aggregate a mandatory and a desired criterion.

For suitability map construction, the overall suitability degree of each analyzed cell x, y must be computed. This is

ISBN: 978-989-95079-6-8

done in two steps: First, the elementary satisfaction degrees $s_1(x,y),\ldots,s_n(x,y)$ are aggregated using the logic aggregation structure. This results in the overall satisfaction degree $s(x,y)$ of the cell. Second, cost (if applicable) is taken into account. Cost is dealt with separately. This better reflects human reasoning and allows for more efficient cost/satisfaction studies. Cost is considered to be a function c of the analyzed cells. For each cell x,y the cost function returns the associated cost $c(x,y)$ of the cell. If the importance of high satisfaction of criteria is the same as the importance of low cost, then the overall suitability degree $S(x,y)$ of the cell can be computed by $S(x,y)=s(x,y)/c(x,y)$. Alternative definitions of $S(x,y)$ are possible.

4 Conclusions

All multicriteria decision methods should be models of human decision making and the only way to prove their credibility is to show their proximity to human evaluation logic. Our analysis shows that the majority of existing MCDM are not derived with an explicit goal to model observable properties of human reasoning. Indeed, human reasoning is neither restricted to the use of arithmetic mean nor restricted to relative criteria only. Humans use a spectrum of absolute and relative elementary criteria and a spectrum of soft computing logic aggregators, such as soft and hard simultaneity and replaceability with nonmandatory, mandatory and sufficient attributes, conjunctive and disjunctive partial absorption, as well as other compound aggregators.

Decision methods that are used in land-use evaluation problems cannot be randomly selected, without appropriate justification. The justification for using a specific evaluation method in a GIS environment should be based on investigating the capability of the method to support features that are proved to characterize human decision making. Many oversimplifications that are frequent in GIS literature [16] (particularly those based on simple additive models) are based on the fact that they are "easy-to-understand and intuitively appealing." Unfortunately, that is not enough. Before using mathematical models it is first necessary to prove that they are appropriate.

LSP is a method developed with the goal to support logic operators observed in human reasoning. Consequently, it is realistic to expect that the LSP method can provide highly accurate and justifiable models for GIS applications, such as land-use evaluation, suitability maps, and natural resources planning.

References

[1] R. Banai. Fuzziness in geographic information systems: contributions from the analytic hierarchy process. *International Journal of Geographical Information Systems*, Vol. 7, No. 4, pp. 315–329, 1993.

[2] J. Dujmović. Partial Absorption Function. *Journal of the University of Belgrade, EE Dept., Series Mathematics and Physics*, No. 659, pp. 156–163, 1979.

[3] J. Dujmović and H.L. Larsen. Generalized Conjunction / Disjunction. *International Journal of Approximate Reasoning*, Vol. 46, pp. 423–446, 2007.

[4] J. Dujmović. Preference Logic for System Evaluation. *IEEE Transactions on Fuzzy Systems*, Vol. 15, No. 6, pp. 1082–1099, 2007.

[5] J. Dujmović, G. De Tré and N. Van de Weghe. Suitability Maps Based on the LSP Method. In *Proc. of the 5th MDAI conference (Modeling Decisions for Artificial Intelligence)*, Sabadell, Spain, 2008. Lecture Notes in Artificial Intelligence, vol. 5285, pp. 15–25, Springer-Verlag, Heidelberg, Germany, 2008.

[6] I. Heywood, J. Oliver and S. Tomlinson. Building an exploratory multi-criteria modelling environment for spatial decision support. In P. Fisher (Ed.) *Innovations in GIS 2*, pp. 127–136, Taylor & Francis, London, UK, 1995.

[7] L.D. Hopkins. Methods for Generating Land Suitability Maps: A Comparative Evaluation. *JAPA*, Vol. 43, No. 4, 1977.

[8] P. Jankowski and L. Richard. Integration of GIS-based suitability analysis and multicriteria evaluation in a spatial decision support system for route selection. *Environment and Planning B*, Vol. 21, No. 3, pp. 326–339, 1994.

[9] P. Jankowski, N. Andrienko and G. Andrienko. Map-centered exploratory approach to multiple criteria spatial decision making. *Int. J. Geographical Information Science*, Vol. 15, No. 2, pp. 101–127, 2001.

[10] R. Janssen and P. Rietveld. Multicriteria analysis and geographical information systems: an application to agricultural land use in the Netherlands. In H.J. Scholten and J.C.H. Stillwell (Eds.) *Geographical information systems for urban and regional planning*, pp. 129–139, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[11] F. Joerin, M. Thériault and A. Musy. Using GIS and outranking multicriteria analysis for land-use suitability assessment. *Int. J. Geographical Information Science*, Vol. 15, No. 2, pp. 153–174, 2001.

[12] A. Kangas, J. Kangas and J. Pykäkäinen. Outranking Methods As Tools in Strategic Natural Resources Planning. *Silva Fennica*, Vol. 35, No. 2, pp. 215–227, 2001.

[13] R.L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs (2nd Edition)*. Cambridge University Press, Cambridge, 1993.

[14] R.L. Keeney. Value-focused thinking: identifying decision opportunities and creating alternatives. *European Journal of Operational Research*, Vol. 92, pp. 537–549, 1996.

[15] J. Malczewski. *GIS and multicriteria decision analysis*. John Wiley & Sons Inc., New York, USA, 1999.

[16] J. Malczewski. GIS-based multicriteria analysis: a survey of the literature. *International Journal of Geographic Information Science*, Vol. 20, No. 7, pp. 703–726, 2006.

[17] C. Rinner and M. Raubal. Personalized Multi-Criteria Decision Strategies in Location-Based Decision Support. *Journal of Geographic Information Sciences*, Vol. 10, No. 2, pp. 149–156, 2004.

[18] T.L. Saaty. *The Analytic Hierarchy Process*. McGraw Hill, New York, USA, 1980.

[19] S.Y.W. Su, J. Dujmović, D.S. Batory, S.B. Navathe and R. Elnicki. A Cost-Benefit Decision Model: Analysis, Comparison, and Selection of Data Management Systems. *ACM Transactions on Database Systems*, Vol. 12, No. 3, pp. 472–520, 1987.

[20] L.J. Wood and S. Dragićević. GIS-based multicriteria evaluation and fuzzy sets to identify priority sites for marine protection. *Biodivers. Conserv.*, Vol. 16, pp. 2539–2558, 2007.

[21] R.R. Yager. On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 18, No. 1, pp. 183–190, 1988.

[22] M. Zeleny. *Multiple Criteria Decision Making*. McGraw Hill, New York, USA, 1982.

ISBN: 978-989-95079-6-8

Computing with Actions: The case of driving a car in a simulated car race

Sergio Guadarrama *
Fundamentals for Soft Computing Unit
European Centre for Soft Computing
Mieres, Spain
sergio.guadarrama@softcomputing.es

Abstract— *Describing and modeling a complex system is very hard and time consuming, so, we propose to use a linguistic approach introduced by Zadeh. By taking a linguistic approach we can use linguistic variables and approximate rules to characterize approximately phenomenons which are too complex or too ill-defined to be done in numerical terms.*

Starting from the results obtained in our previous works, in this paper we propose to decompose complex behaviors in simple behaviors, and context dependent rules for the behaviors coordination, which are described by set of fuzzy rules. The use of fuzzy rules for behavior coordination and behavior description has allowed us to have a shared framework that makes easy to develop and to compute with linguistic descriptions of actions.

We propose to use a mixture of competitive and cooperative co-evolutionary algorithms to manage the big search space and tackle a complex and decomposable problem.

We use the Fuzzy-IEEE 2007 Car Racing Competition as an example of a problem that allows to show the benefits of our approach. In the example presented the results obtained are better and the flexibility is increased with respect to previous works.

Keywords— Computing with Words, Computing with Actions, Fuzzy Linguistic Descriptions, Fuzzy Controllers, Co-evolutionary Algorithms, Behavior coordination.

1 Introduction

Real-life problems with solutions that are complex and unknown in advance require to build and test several prototypes before being able to build a good enough solution. Prototypes allow designers to devise what an appropriate solution might look like while they work at intermediate stages of development. In this way, designers require to test potential solutions, following a design-test-update cycle before reaching a good enough solution (as suggested in the Agile software development [1]). Computers are useful tools to develop and test these prototypes, which in turn serve designers to explore, build and test different ideas.

Describing and modeling a complex system is very hard and time consuming. So, we propose to use a linguistic approach introduced by Zadeh in [2, 3, 4, 5, 6] to deal with complex systems and complex concepts. By taking a linguistic approach we can use linguistic variables and approximate rules to characterize approximately phenomenons which are too complex or too ill-defined to be done in numerical terms [7, 8]. We use linguistic descriptions to describe the desired solution, problem decomposition techniques and fuzzy controllers to build the solution, and co-evolutionary algorithms to tune and improve the solutions.

*His work has been supported by the Spanish Department of Science and Innovation (MICINN) under program Juan de la Cierva JCI-2008-3531, and the European Social Fund.

Indeed, in this work we are introducing a tentative proposal to compute with actions by computing with linguistic descriptions of actions, paralleling Zadeh's Computational Theory of Perceptions [9], in which he proposed to compute with perceptions by means of computing with linguistic descriptions of perceptions.

To cope with these problems, we combine our previous ideas, improvements and results in the areas of computing with words [10], fuzzy controllers [11] and co-evolutionary algorithms [12].

The rest of the paper is organized as follows. In section 2, we provide details of the problem tackled, the FUZZ-IEEE 2007 Car Racing Competition. In section 3, the solution based on linguistic description of the behaviors is presented. The coordination between behaviors is presented in section 5. The mixture of co-evolutionary algorithms is presented in section 6. In section 7 we present the results and finally, we summarize our work and provide conclusions in section 8.

2 The problem: FUZZ-IEEE 2007 Car Racing Competition

Car Racing Competition is a well-known sport that gets attention from many people in a number of different scenarios, ranging from popular real-life Fomula One to virtual computer games. Particularly, we tackle the challenge of building a car pilot for the FUZZ-IEEE 2007 Car Racing Competition.

Driving is a complex problem and how to find a general solution is unknown. Everyone need a lot practice before becoming a good driver. Actually, people become driving experts by practicing in many different scenarios. So we will use the simulator provided by the FUZZ-IEEE 2007 Car Competition to simulate different races that will allow us to learn and test solutions in many different scenarios.

In this problem, the challenge was to develop the best fuzzy car driver controller for a basic form of car race. A competition in which a pair of simulated cars race through a series of waypoints within a 2d plane, and compete between them to reach the maximum number of waypoints. Basic aspects of the car race are the following:

1. A car only sees the next three waypoints at a given time.
2. Waypoints have to be visited in order and the next one to be visited is highlighted.
3. Waypoints are randomly generated on a 2d plane.
4. The car that first reaches the next waypoint scores.
5. Cars cannot collide with each other, so two cars can follow the same path at the same time.

ISBN: 978-989-95079-6-8

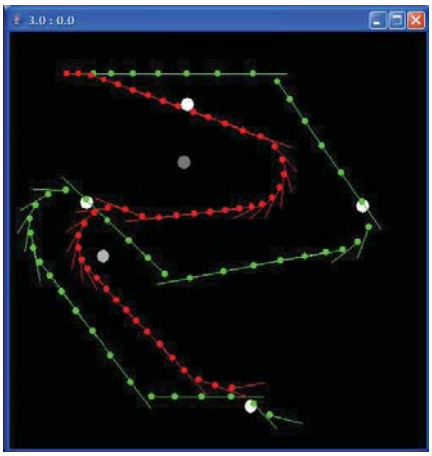


Figure 1: Car Racing Simulation on a 2d Plane.

6. Cars have imperfect sensors and actuators: gaussian noise is used to simulate the lack of perfect observations on position and velocity.

7. General driving ability rather than specialization to a particular track it is tested.

3 The solution: Behaviors described by linguistic actions

We describe a desired behavior by describing linguistically a set of actions modeled by fuzzy rules. So each behavior will be described by a set of fuzzy rules, and a solution will be composed by one or more behaviors coordinated appropriately.

Fuzzy rules are linguistically interpretable units that can serve our purposes of designing and testing prototypes easily and quickly, since they are easily human-interpretable, as well as, adjustable.

We have used fuzzy controllers [13, 14, 15] to implement individual behavior units. Fuzzy controllers incorporate heuristic control knowledge in the form of if-then rules, whose membership functions are later tuned by the co-evolutionary algorithm, similarly as it was proposed by Karr in [16].

We have used Xfuzzy 3.0, a fuzzy system development environment, to design our fuzzy inference system [17]. This system contains a set of tools that allows not only to easily describe a fuzzy inference model but also to implement it, and export it to programming languages such as JAVA, C and C++.

In this section we present the three different behaviors that we introduced in our previous paper [12]. The first behavior "Go straight to next target" was a local solution in the sense that the other targets were not considered (the other two way-points). The second behavior "Follow a spline curve" takes into account the next three way-points, thus following a more global approach. The third behavior "Avoiding ellipses around targets" cope with unwanted situation of going around a target indefinitely.

3.1 Behavior 1: Go straight to the target

Basically, this fuzzy controller takes as input variables: velocity, distance to next target and angle to next target; and has as output variables: acceleration and turn angle. The final controller is composed of two rule-bases:

- Acceleration rules:
  If Distance is Large then Acceleration is Very Positive.
  If Distance is Medium then Acceleration is Positive.
  If Distance is Small and Velocity is Very Positive then Acceleration is Very Negative.
  If Distance is Small and Velocity is Positive then Acceleration is Negative.
  If Distance is not Small and Velocity is Zero then Accelerations is Positive.

- Turn rules:
  If Angle is Zero then Turn Zero
  If Angle is Positive then Turn Positive
  If Angle is Negative then Turn Negative
  If Angle is not Zero and Distance is Very Small then Turn Zero

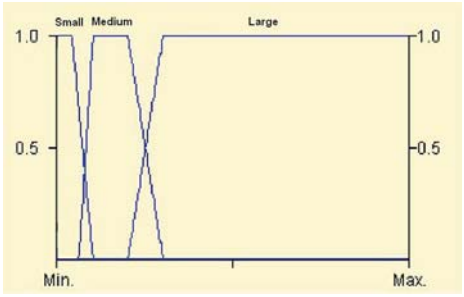


Figure 2: Distance Linguistic Variable

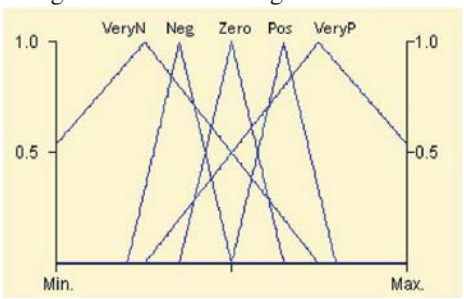


Figure 3: Acceleration Linguistic Variable

We impose the following ordering constrains [18] between the linguistic labels of the linguistic variables to maintain the interpretability of the rules, although these don't impose that they should form a partition:

- Distance = { Small $\prec$ Medium $\prec$ Large }
- Velocity = { Very Negative $\prec$ Negative $\prec$ Zero $\prec$ Positive $\prec$ Very Positive }
- Acceleration = { Very Negative $\prec$ Negative $\prec$ Zero $\prec$ Positive $\prec$ Very Positive }

After tuning parameters of the fuzzy membership functions using a genetic algorithm (see figures 2 and 3) of the fuzzy rules this behavior was able to defeat the heuristic controller in all races (see table 2 for comparison). So from now on we took it as the new baseline for comparison with the rest of the solutions.

ISBN: 978-989-95079-6-8

3.2 Behavior 2: Follow a spline curve

We observed that the previous behavior was able to align with the first target (next target) quite well, but it did not consider the positions of second and third targets. In other words, direction of the car was perfect for the next target, but it was not exploiting the knowledge about the rest other targets.

To do that, we use a spline curve instead of following straight lines, in order to follow a global strategy and gain advantage over the opponent, as soon as possible. We built a spline curve that goes through four points on a 2d plane: the position of our car, and the next three targets. We use the spline curve to calculate/approximate intermediate points between our car and the next target, and then follow them using these fuzzy rules:

- Acceleration rules:
  If Distance is Large then Acceleration is Very Positive.
  If Distance is Medium then Acceleration is Positive.
  If Distance is not Small and Velocity is Zero then Accelerations is Positive.
- Turn rules:
  If Angle is Zero then Turn Zero
  If Angle is Positive then Turn Positive
  If Angle is Negative then Turn Negative

On the negative side of this approach we can say that the distance that the car is covering is higher than in our previous behavior, that is, without using a spline curve, so when we set up a competitive co-evolutionary strategy this behavior (B2) was always loosing, unable to defeat the previous one (B1) (see table 3 for comparison).

3.3 Behavior 3: Avoid ellipses around targets

We also noticed that some races were ended when both cars were unable to reach a target. This situation occurred when a new target appeared and both cars were: 1, too close to the new target; and 2, with such a high angle to the target that it could not be corrected enough, given the short distance and the velocity of the cars. As a result, cars spent their last moments turning around the target.

To avoid that, we add this behavior (B3), that detects the unwanted situation and takes a different action. In order to detect that unwanted situation, we keep track of what happened in last n iterations. Particularly, we consider three facts during the last n iterations: 1, the target remained the same; 2, the angle to target was around 80 degrees; 3, the distance to target was small.

To address the described scenario, when the unwanted situation is detected then this new behavior is activated during a number m of iterations. This new behavior was built following the same rules that *"Go straight to the target"* with one sole difference: it did not go towards the target but in the opposite direction (it has the same rules as behavior 1 but with opposite sign).

4 Tuning Fuzzy Membership Functions by Co-evolution

Developing a good prototype from scratch is very time consuming and difficult since it involves a good understanding of the problem to do a good design, and a good understanding of the solution to adjust all its parameters. We did this initially manually and then used genetic algorithms, and co-evolutionary algorithms.

Table 1: Summary of previous results

	HC	GA	$Comp_1$	$Coop_\alpha$
HC	-	15	0	0
GA	83	-	1	5
$Comp_1$	100	99	-	47
$Coop_\alpha$	100	81	47	-

To avoid complexity and exploit the features of co-evolutionary algorithms we defined different species with different attributes to optimize. Actually, we divided our population into three species. Species number 1 consisted of individuals of behavior 1 *"Go straight to target"*, species number 2 was formed by individuals belonging to behavior 2 *"Follow a spline curve"*, and species number 3 was formed by individuals of behavior 3 *"Avoid ellipses"*.

So, we built a population with three different species with 20 individuals each, which evolve during 1000 generations. These species only met when racing (to calculate the fitness functions of individuals), while for the rest (cross, mutation, selection of individuals) were kept apart.

In [12] we compared how the use of a competitive or a cooperative co-evolutionary algorithm affects to the final solution. Obtaining that both approaches can find very good solutions, once the appropriate crossover operator is chosen. Table 6 recall the summary of results.

1. *HC*, hand coded model;
2. *GA*, complete model including three pilots, automatically tuned using a genetic algorithm;
3. $Comp_1$, complete model including three pilots, automatically tuned using a competitive coevolutionary approach, with 1-point cross operator;
4. $Coop_\alpha$, complete model including three pilots, automatically tuned using a cooperative coevolutionary approach, with BLX-alpha cross operator.

5 Behavior coordination

The main two problems of behavior coordination are:

1. Arbitration: deciding which behavior should be activated in each situation. This can be decided a priori or dynamically by establishing variable priorities. The simplest solution is a switching scheme that selects one behavior and ignores the rest. More flexible arbitration can be obtained using fuzzy context rules, that allow to activate concurrently several behaviors with different degrees of activation.
2. Command fusion: If only one behavior is active at a time, then there is not need to do it, but if there are more than

ISBN: 978-989-95079-6-8

one behavior active simultaneously, then it is needed to combine in some way the different commands generated by them. It is needed to fuse, or aggregate, properly these commands in order not to send a set of contradictory commands but to send a coherent fused command.

Context dependent blending is a flexible and general form of solving the problem of behavior coordination, since it solves both arbitration and command fusion by using a common scheme, fuzzy context rules and fuzzy aggregation. It was initially proposed by Ruspini in [19] and later expanded by Saffioti and himself in [20, 21].

5.1 Context dependent behavior coordination

In our previous work we defined a fix arbitration scheme to switch between behaviors. And compared competitive and cooperative co-evolutionary algorithms to tune the membership functions of the linguistic variables used in the fuzzy rules.

In this work we use context dependent blending, using fuzzy rules for both behavior coordination and behavior description, since that allowed us to have a shared framework that made easier to develop complex behaviors.

To address the coordination problem, we defined context rules for each behavior, so they could be activated in different situations, and serve to merge their outputs when are activated at the same time.

- Context Rule 1:
  If Distance is Small then "Go Straight to the target".
- Context Rule 2:
  If Distance is Large then "Follow a spline curve".
- Context Rule 3:
  If Distance is Small and Angle is approx. 80^o then "Avoid ellipses".

This allows to follow a local strategy when the car is near to the next target, a global strategy when the car is far to the next target, and avoiding ellipses around targets

We set up a mixture of competitive and cooperative co-evolutionary strategy, the individual behaviors cooperate to form the complex behavior (B123) while competing with the best solution obtained previously (B1). (see table 5 for comparison).

In figure 4 we can see a diagram of how the coordination of the different behaviors works. Using a weighted fusion between the commands of behaviors 1 and 2 and a switching scheme with behavior 3.

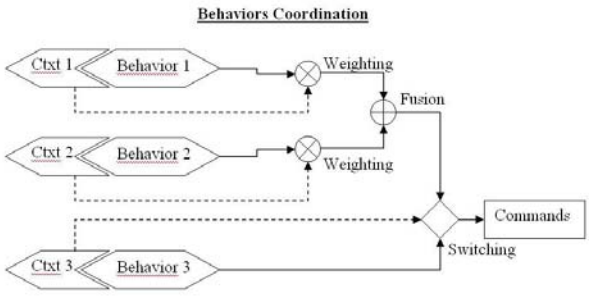


Figure 4: Behaviors Coordination

6 Mixture of Co-evolutionary algorithms

In some way, approaches pointing in the same direction that ours have already been proposed. In fact, many papers on fuzzy systems tuned with genetic algorithms have been published (for a good review see [22]).

Genetic algorithms, of single-population algorithms, are not well suited to problems which have solutions that are long, complex and interdependent [23]. In the last years, co-evolutionary algorithms have shown their ability to manage big search spaces and to tackle complex problems. Specially when the solution to the problem may be decomposable, or when there exist different competitive solutions [24, 25].

Co-evolutionary algorithms can be seen as an extension of traditional evolutionary algorithms, where there are, at least, two species that are simultaneously evolved and their fitness functions are coupled [24, 25]. Therefore, each species has its own coding scheme and reproduction scheme, but each individual is evaluated considering individuals from other species. Depending on the type of interaction between species, we can distinguish between competitive or cooperative species.

6.1 Competitive Co-evolution

In this context, each individual represents a candidate solution to the problem and competes with individuals of other species to find better solutions. Thus, if one individual increases its fitness, then fitness functions from the rest individuals decrease [24]. This may be seen as a kind of struggle between species, which presses individuals to compete between them for finding better solutions.

6.2 Cooperative Co-evolution

In this case, each individual represents a component of the solution and requires the rest components to build a complete solution. Therefore, an individual needs to cooperate with others to find a better solutions [25]. Thus, an individual fitness is linked with fitness values of its cooperative individuals. This is often seen as a kind of symbiosis between species to preserve existence, leading individuals to cooperate for finding better solutions.

6.3 Mixture of Competitive and Cooperative Co-evolution

We have different solutions that compete or cooperate to obtain better solutions. Following the two co-evolutionary approaches we have defined a mixed environment, where alternative solutions were competing to obtain better solutions, and where complementary solutions were cooperating to obtain better solutions. At the same time that the three different behaviors evolve in a cooperative way, they are competing with the best previous solutions.

The main features of the co-evolutionary algorithm used to tune the parameters of the membership functions are the following:

generations = 1000.
population size = 20.
crossover probability = 0.9.
mutation probability = 0.1.
crossover operator = BLX-α.
$\alpha = 0.7$.
species = 3.
co-evolutionary strategy = mixed.

ISBN: 978-989-95079-6-8

Each specie have its own coding as follows:

- Go Straight to Target (51 genes) + Context 1 (2 genes)
- Following a Spline Curve (39 genes) + Context 2 (2 genes)
- Avoiding Ellipses (51 genes) + Context 3 (5 genes)

7 Results

In this section it is shown the sum of the results we obtained during the development of the system and a comparative evaluation of the obtained solutions (for previous results see [12]).

Table 2:
Heuristic vs Behavior 1

Race	HC	B1
1	18	37
2	15	35
3	12	42
4	16	44
5	11	43
6	19	41
7	13	46
8	12	42
9	20	44
10	17	39
Total	153	413

Table 3:
Behavior 1 vs Behavior 2

Race	B1	B2
1	47	4
2	44	2
3	44	5
4	43	7
5	42	6
6	44	3
7	43	9
8	42	8
9	42	5
10	44	5
Total	435	54

Table 2 shows the number of way-points (targets) got in 10 races by the hand coded (HC) against the behavior 1 (B1) "Go straight to the next target", in which we can see that the behavior 1 is much better than the hand coded controller, and it is wining all the races by a big difference.

Table 3 shows the number of way-points (targets) got in 10 races by the behavior 1 (B1) "Go straight to the next target" against the behavior 2 (B2) "Follow the spline curve", in which we can see that behavior 2 is performing very bad against behavior 1, loosing all the races by a huge difference of points, and performing even worse than the hand coded controller.

These results indicate that the best individual behavior is "Go straight to the next target", and that to improve these results we need to combine it with other behaviors, as we did in the next experiments.

Table 4 shows the number of way-points (targets) got in 10 races by the behavior 1 (B1) "Go straight to the next target" against the combined behavior 1 and 2 (B12) "Go straight to the next target; Follow the spline curve", in which we can see that the combined behavior 12 is much better that the behavior 1 alone, and it is wining all the races by some difference. Table 5 shows the number of way-points (targets) got in 10 races by the behavior 1 (B1) "Go straight to the next target" against the combined behavior 1, 2 and 3 (B123) "Go straight to the next target; Follow the spline curve; Avoid ellipses", in which

Table 4:
Behavior 1 vs Behavior 12

Race	B1	B12
1	26	35
2	27	37
3	25	32
4	31	32
5	24	38
6	32	31
7	28	36
8	26	36
9	28	35
10	30	40
Total	277	352

Table 5:
Behavior 1 vs Behavior 123

Race	B1	B123
1	19	35
2	18	45
3	13	43
4	22	33
5	18	32
6	20	34
7	16	37
8	26	28
9	24	32
10	20	36
Total	196	355

we can see that behavior 123 is performing very well against behavior 1 wining all the races by a big difference of points, and performing even better than the combined behavior 1 and 2 (B12).

These results indicate that the best behavior is the combined behavior 123. But after these comparisons and to evaluate the impact of the context blending behavior coordination, we compared it with the best solutions from our previous work $Comp_1$ and $Coop_\alpha$.

We set up a league competition among all of them, by playing 100 races between each pair of solutions. Table 6 shows the results of the five solutions considered: *Hand-coded*, $Comp_1$, $Coop_\alpha$ and $B123$, and their associated wins for each one in the 100 races between each pair of solutions.

Table 6: Wins Against Each Other

	HC	$Comp_1$	$Coop_\alpha$	$B123$	Total
HC	-	0	0	0	0
$Comp_1$	100	-	47	20	167
$Coop_\alpha$	100	51	-	31	182
$B123$	**100**	**74**	**63**	–	**237**

Table 6 shows the associated wins of each solution. Both behaviors, $Comp_1$ and $Coop_\alpha$, obtained good performance but the new solution $B123$ outperformed them with a greater number of wins than the other approaches, which indicates that: 1, the context dependent blending provides a great advantage compared to fixed arbitration coordination; 2, automatically tuning of fuzzy membership functions yield better results than using hand-coded solutions.

8 Conclusions

In the particular context of the FUZZ-IEEE 2007 Car Racing Competition we have introduced a tentative proposal to compute with actions by computing with linguistic descriptions of behaviors. In this context we have identified and tackled the

ISBN: 978-989-95079-6-8

following key aspects: how to define different behaviors by describing them linguistically and how to coordinate several behaviors using context dependent rules.

We have set up a mixture of competitive and cooperative co-evolutionary algorithms to model the membership functions of the linguistic variables used. On one hand, competitive co-evolution allows to find alternative solutions that compete among them to find better solutions and to avoid being trapped in local minima. On the other hand, cooperative co-evolution allows to find complementary solutions that cooperate among them to find complex and better solutions.

By following an incremental approach, we were able to decompose the solution to this problem in simple behaviors with a lower degree of complexity. The use of fuzzy rules for behavior coordination and behavior description has allowed us to have a shared framework that makes easy to develop complex behaviors and to compute with linguistic descriptions of actions. The use of co-evolutionary algorithms has allowed us to test, compare, combine and tune the different behaviors appropriately and obtain a better solution.

In general, it is not possible to predict a priori all the scenarios, problems, or solutions. So a "learning by doing" approach is a need in these cases. The know-how obtained by testing and playing with different solutions is irreplaceable in order to build systems that compute with actions described by words. This approach allows to mix designer's knowledge and knowledge learned from the data, maintaining a good interpretability of the solution. This represents a very important step-stone that would allow to change and improve the system in front of new environments and situations.

Acknowledgment

The author want to thank to Claudio Moraga, David Perez and to the anonymous reviewers for their comments and fruitful discussion about the paper.

References

[1] J. Highsmith and A. Cockburn. Agile software development: The business of innovation. *Computer*, 34(9):120–127, 2001.

[2] L.A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transaction on Systems, Man and Cybernetics*, 3:28–44, 1973.

[3] L.A. Zadeh. On the analysis of large-scale systems. In *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers*, pages 195–209. World Scientific Pub Co Inc, 1996.

[4] L.A. Zadeh. A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. *Int. J. Man-Machine Studies*, 8:249–291, 1976.

[5] L.A. Zadeh. Precisiated natural language (PNL). *AI MAGAZINE*, 25(3):74–91, 2004.

[6] L.A. Zadeh. A New Frontier in Computation? Computation with Information Described in Natural Language. In *Proceedings Symposium on Fuzzy Systems in Computer Science FSCS*, pages 1–2, 2006.

[7] L. A. Zadeh. Fuzzy logic = computing with words. *IEEE Trans. Fuzzy Systems*, 4:103–111, 1996.

[8] L.A. Zadeh. From computing with numbers to computing with words.from manipulation of measurements to manipulation of perceptions. In P.P. Wang, editor, *Computing with words*, pages 35–68. John Wiley & Sons, 2001.

[9] L.A. Zadeh. A New Direction in AI: Toward a Computational Theory of Perceptions. *AI Magazine*, 22(1):73, 2001.

[10] S. Guadarrama. *A Contribution to Computing with Words / Perceptions*. PhD thesis, Universidad Politécnica de Madrid, Madrid, April 2007. http://oa.upm.es/448/.

[11] C. Moraga, E. Trillas, and S. Guadarrama. Multiple-valued logic and artificial intelligence: Fundamentals of fuzzy control revisited. *Artif. Intell. Rev.*, 20(3-4):169–197, 2003.

[12] S. Guadarrama and R. Vazquez. Tuning a fuzzy racing car by coevolution. In *Genetic and Evolving Systems, 2008. GEFS 2008. 3rd International Workshop on*, pages 59–64, 2008.

[13] L.A. Zadeh. A rationale for fuzzy control. *Jr. of Dynamic Systems, Measurements and Control*, 34:3–4, 1972.

[14] S. Assilian and E.H. Mamdani. Learning control algorithms in real dynamic systems. In *Int. IFAC/IFIP Conf. on Digital Computer Appl. to Process Control*, ZÜrich, 1974. IFAC/IFIP.

[15] E.H. Mamdani. Advances in the linguistic synthesis of fuzzy controllers. *Int. Jr. Man-Machine Studies*, 8:669–678, 1976.

[16] C. Karr. Genetic algorithms for fuzzy controllers. *AI Expert*, 6(2):2633, 1991.

[17] F.J. Moreno-Velo, I. Baturone, S. Sanchez-Solano, and A. Barriga. XFUZZY 3.0: A Development Environment for Fuzzy Systems. In *Proc. International Conference in Fuzzy Logic and Technology*, pages 93–96, 2000.

[18] JJ Saade and H. Schwarzlander. Ordering fuzzy sets over the real line: an approach based on decisio making under uncertainty. *Fuzzy sets and systems*, 50(3):237–246, 1992.

[19] E.H. Ruspini. Fuzzy logic in the flakey robot. In *Procs. of the Int. Conf on Fuzzy Logic and Neural Networks*, pages 767–770, Iizuka, Japan, 1990.

[20] A. Saffiotti, K. Konolige, and E.H. Ruspini. A multivalued logic approach to integrating planning and control. *Artificial Intelligence*, 76(1-2):481–526, 1995.

[21] A. Saffiotti, E.H. Ruspini, and K. Konolige. Blending reactivity and goal-directedness in a fuzzy controller. In *Fuzzy Systems, 1993., Second IEEE International Conference on*, pages 134–139, 1993.

[22] O. Cordon, F. Herrera, F. Gomide, F. Hoffmann, and L. Magdalena. Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy Sets and Systems*, 141(1):5–31, 2004.

[23] C.A. Pena-Reyes and M. Sipper. Fuzzy CoCo: a cooperative-coevolutionary approach to fuzzy modeling. *Fuzzy Systems, IEEE Transactions on*, 9(5):727–737, 2001.

[24] C.D. Rosin and R.K. Belew. New Methods for Competitive Coevolution. *Evolutionary Computation*, 5(1):1–29, 1997.

[25] M.A. Potter and K.A.D. Jong. Cooperative Coevolution: An Architecture for Evolving Coadapted Subcomponents. *Evolutionary Computation*, 8(1):1–29, 2000.

ISBN: 978-989-95079-6-8

Analysis of Similarity Measures for Atanassov's Intuitionistic Fuzzy Sets

Eulalia Szmidt and Janusz Kacprzyk

Systems Research Institute, Polish Academy of Sciences,
ul. Newelska 6, 01–447 Warsaw, Poland
Email: {szmidt, kacprzyk}@ibspan.waw.pl

Abstract— *We consider some existing similarity measures for Atanassov's intuitionistic fuzzy sets (A-IFSs, for short). We show that neither similarity measures treating an A-IF as a simple interval values fuzzy set, nor straightforward generalizations of the similarity measures well-known for the classic fuzzy sets work under reasonable circumstances. Next, expanding upon our previous works, we consider a family of similarity measures constructed by taking into account both all the three functions (the membership, non-membership and hesitation) describing an A-IF, and the complements of the elements we compare to each other. That is, we use all kinds and fine shades of information available. We point out their proper behavior and an intuitive appeal.*

Keywords— Atanassov's intuitionistic fuzzy sets, similarity measures.

1 Introduction

Atanassov's intuitionistic fuzzy sets (Atanassov [1], [2], [3]) – to be called A-IFSs, for short – can be viewed as a tool that may better model and process imperfect information. The use of positive and (independently) negative information that is the core of A-IFSs is natural in real life and is also well-known, advocated and studied in psychology and other social and behavioral sciences [e.g., [21], [14]]. It also attracted much attention and research interest in soft computing. It would be difficult to deal with machine learning (making use of examples and counter-examples), modeling of preferences or voting without taking into account positive and (independently) negative testimonies or opinions. Although from a mathematical point of view A-IFSs are equipotent, under some assumptions, to interval-valued fuzzy sets (cf. Atanassov and Gargov in 1989 [4]), from the point of view of solving problems (starting from the stage of collecting data), they are different as A-IFSs force the user to explicitly consider positive and negative information independently. On the other hand, while employing the interval-valued fuzzy sets, the user's attention is focused on positive information (in an interval form) only. Notably, Dubois [9] noticed recently that A-IFSs correspond to an intuition that differs from that behind the interval valued fuzzy sets.

The fact that people tend to notice and take into account only most obvious aspects (e.g. advantages only) when making decision is well known in psychology (cf. Kahneman [14]), Sutherland [21]) and may often lead to improper decisions. In this context, A-IFSs ("forcing" a decision maker to take into account both negative and positive aspects of the decisions) may be seen as belonging to modern and promising means of knowledge representation and processing.

We consider here some major similarity measures for A-IFSs. First we present a whole array of similarity measures (known from the literature) for A-IFSs viewed in terms of single intervals. Second, we consider measures that are straightforward generalization of the similarity measures for the conventional fuzzy sets. Unfortunately, both do not meet our expectations, and we provide some counter-intuitive examples. It seems to be an indirect hint that A-IFSs are not functionally equivalent to the interval valued fuzzy sets.

Next, we reconsider our (cf. Szmidt and Kacprzyk [41]) concept of a similarity measure between A-IFSs taking into account all three functions (the membership, non-membership and hesitation), and explicitly add to the above three functions the complements of the elements we compare to each other.

2 A brief introduction to A-IFSs

One of the possible generalizations of a fuzzy set in X (Zadeh [48]), given by

$$A^{'} = \{< x, \mu_{A^{'}}(x) > | x \in X\} \tag{1}$$

where $\mu_{A^{'}}(x) \in [0,1]$ is the membership function of the fuzzy set $A^{'}$, is Atanassov's intuitionistic fuzzy set (Atanassov [1], [2], [3]), A-IF, A given by

$$A = \{< x, \mu_A(x), \nu_A(x) > | x \in X\} \tag{2}$$

where: $\mu_A : X \rightarrow [0,1]$ and $\nu_A : X \rightarrow [0,1]$ such that

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1 \tag{3}$$

and $\mu_A(x)$, $\nu_A(x) \in [0,1]$ denote the degree of membership and a degree of non-membership of $x \in A$, respectively.

Each fuzzy set may be represented by the following A-IF

$$A = \{< x, \mu_{A^{'}}(x), 1 - \mu_{A^{'}}(x) > | x \in X\} \tag{4}$$

For each intuitionistic fuzzy set in X, we will call

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \tag{5}$$

an *intuitionistic fuzzy index* (or a *hesitation margin*) of $x \in A$ and, it expresses a lack of knowledge of whether x belongs to A or not (cf. Atanassov [3]). It is obvious that $0 \leq \pi_A(x) \leq 1$, for each $x \in X$.

The hesitation margin turns out to be important while considering the distances (Szmidt and Kacprzyk [26], [29], [39], entropy (Szmidt and Kacprzyk [31], [40]), similarity (Szmidt and Kacprzyk [41]) for the A-IFSs, etc. i.e., the measures that play a crucial role in virtually all information processing tasks.

In our further considerations we will use the following distances between fuzzy sets A, B in $X = \{x_{1,}, \ldots, x_n\}$ Szmidt

ISBN: 978-989-95079-6-8

and Baldwin [23], [24], Szmidt and Kacprzyk [29], [39]: the normalized Hamming distance

$$l_{IFS}(A,B) = \frac{1}{2n}\sum_{i=1}^{n}(|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|) \quad (6)$$

and the normalized Euclidean distance:

$$q_{IFS}(A,B) = (\frac{1}{2n}\sum_{i=1}^{n}(\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2)^{\frac{1}{2}} \quad (7)$$

For distances (6), and (7) we have $0 \leq l_{IFS}(A,B) \leq 1$, and $0 \leq q_{IFS}(A,B) \leq 1$. Clearly these distances satisfy the conditions of the metric.

In our further considerations we will use the notion of the complement elements, which definition is a simple consequence of a complement set A^C

$$A^C = \{< x, \nu_A(x), \mu_A(x) > | x \in X\} \quad (8)$$

The use of A-IFSs instead of fuzzy sets implies the introduction of another degree of freedom (non-memberships) into the set description. Such a generalization of fuzzy sets gives us an additional possibility to represent imperfect knowledge which leads to describing many real problems in a more adequate way. Applications of intuitionistic fuzzy sets to group decision making, negotiations, voting and other situations are presented in Szmidt and Kacprzyk [25], [27], [28], [30], [32], [34], [33], [35], [38], Szmidt and Kukier [42], [43].

3 Some counter-intuitive results given by the traditional similarity measures

In the literature there is a multitude of similarity measures both for A-IFSs (Atanassov [1, 2, 3], and vague sets (Gau and Buehrer [10]) which have also been proved to be equivalent to A-IFSs (Bustince and Burillo [5]). Here we adopt the notation for A-IFSs but we will consider the measures originally introduced for vague sets, too.

Chen [6, 7] considered similarity measures between two A-IFSs A and B as

$$S_C(A,B) = 1 - \frac{\sum_{i=1}^{n}|S_A(x_i) - S_B(x_i)|}{2n} \quad (9)$$

where $S_A(x_i) = \mu_A(x_i) - \nu_A(x_i)$, $S_A(x_i) \in [-1,1]$ and $S_B(x_i) = \mu_B(x_i) - \nu_B(x_i)$, $S_B(x_i) \in [-1,1]$.

But, as Hong and Kim [11] noticed

$$\mu_A(x_i) - \nu_A(x_i) = \mu_B(x_i) - \nu_B(x_i) \Rightarrow S_C(A,B) = 1 \quad (10)$$

which is counterintuitive as, e.g., for $A = (x,0,0)$ and $B = (x,0.5,0.5)$, we have $S_C(A,B) = 1$.

To overcome the problem of S_C (9), Hong and Kim [11] proposed the similarity measures S_H and S_L

$$S_H(A,B) = 1 - (\sum_{i=1}^{n}|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)|)/2n \quad (11)$$

$$S_L(A,B) = 1 - \frac{1}{4n}((\sum_{i=1}^{n} S_A(x_i) - S_B(x_i)) - (\sum_{i=1}^{n}|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)|)) \quad (12)$$

Since $S_H(A,B)$ takes into account the absolute values, it does not distinguish the positive from negative differences, e.g., for $A = \{(x,0.3,0.3)\}$, $B = \{(x,0.4,0.4)\}$, $C = \{(x,0.3,0.4)\}$, and $D = \{(x,0.4,0.3)\}$, we obtain from (11) that $S_H(A,B) = S_H(C,D) = 0.9$ which seems counter-intuitive.

$S_L(A,B)$ also gives counter-intuitive results, e.g. for $A = \{(x,0.4,0.2)\}$, $B = \{(x,0.5,0.3)\}$, $C = \{(x,0.5,0.2)\}$, we obtain from (12) $S_L(A,B) = S_L(A,C) = 0.95$ which seems counter-intuitive.

The same problem like with S_H occurs with the similarity measure (cf. Li et al. [16]):

$$S_O(A,B) = 1 - (1/2n)^{0.5}(\sum_{i=1}^{n}(\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2)^{0.5} \quad (13)$$

Dengfeng and Chuntian [15] considered the similarity measure:

$$S_{DC}(A,B) = 1 - (1/n^{1/p})(\sum_{i=1}^{n}(|m_A(x_i) - m_B(x_i)|)^p)^{1/p} \quad (14)$$

where $m_A(x_i) = (\mu_A(x_i) + 1 - \nu_A(x_i))/2$, $m_B(x_i) = (\mu_B(x_i) + 1 - \nu_B(x_i))/2$, $1 \leq p < \infty$. Unfortunately, as for (14), the medians of two intervals are compared only, it is rather easy to point out the counter-intuitive examples, e.g., $A = (x,0.4,0.2)$, $B = (x,0.5,0.3)$, then $S_{DC}(A,B) = 1$, for each p.

Mitchell [18] modified Dengfeng and Chuntian's measure S_{DC} (14) using a statistical approach by interpreting A-IFSs as families of ordered fuzzy sets. Let $\rho_\mu(A,B)$ and $\rho_\nu(A,B)$ denote a similarity measure between the "low" membership functions μ_A and μ_B, and between the "high" membership functions $1 - \nu_A$ and $1 - \nu_B$, respectively, as:

$$\rho_\mu(A,B) = S_{DC}(\mu_A, \mu_B) = 1 - (1/n^{1/p})(\sum_{i=1}^{n}(|\mu_A(x_i) - \mu_B(x_i)|)^p)^{1/p}$$

$$\rho_\nu(A,B) = S_{DC}(1 - \nu_A, 1 - \nu_B) = 1 - (1/n^{1/p})(\sum_{i=1}^{n}(|\nu_A(x_i) - \nu_B(x_i)|)^p)^{1/p}$$

Then the modified similarity measure between A and B is

$$S_{HB}(A,B) = (\rho_\mu(A,B) + \rho_\nu(A,B))/2 \quad (15)$$

Unfortunately, S_{HB} gives the same counter-intuitive results as S_H, for $p = 1$ and for one-element sets.

To overcome the drawbacks of S_{DC}, Liang and Shi [17] proposed $S_e^p(A,B)$, $S_s^p(A,B)$, $S_h^p(A,B)$ as:

ISBN: 978-989-95079-6-8

$$S_e^p(A,B) = 1 - (1/n^{1/p})(\sum_{i=1}^{n}(\phi_{\mu AB}(x_i) - \phi_{\nu AB}(x_i)|)^p)^{1/p} \quad (16)$$

where $\phi_{\mu AB}(x_i) = |\mu_A(x_i) - \mu_B(x_i)|/2$, $\phi_{\nu AB}(x_i) = |(1 - \nu_A(x_i))/2 - (1 - \nu_B(x_i))|/2$. But, for $p = 1$ and for one-element sets, $S_e^p(A,B)$=S_{HB}=S_H, which are again the same counter-intuitive results.

$$S_s^p(A,B) = 1 - (1/n^{1/p})(\sum_{i=1}^{n}(\varphi_{s1}(x_i) - \varphi_{s2}(x_i))^p)^{1/p} \quad (17)$$

where: $\varphi_{s1}(x_i) = |m_{A1}(x_i) - m_{B1}(x_i)|/2$,
$\varphi_{s2}(x_i) = |m_{A2}(x_i) - m_{B2}(x_i)|/2$,
$m_{A1}(x_i) = (\mu_A(x_i) + m_A(x_i))/2$,
$m_{A2}(x_i) = (m_A(x_i) + 1 - \nu_A(x_i))/2$,
$m_{B1}(x_i) = (\mu_B(x_i) + m_B(x_i))/2$,
$m_{B2}(x_i) = (m_B(x_i) + 1 - \nu_B(x_i))/2$,
$m_A(x_i) = (\mu_A(x_i) + 1 - \nu_A(x_i))/2$,
$m_B(x_i) = (\mu_B(x_i) + 1 - \nu_B(x_i))/2$.

S_s^p (17) avoids the problematic results obtained from S_{DC} (14) (for the intervals with equal medians) but, again, a problem of counter-intuitive results remains. For example, for $A = \{(x, 0.4, 0.2)\}$, $B = \{(x, 0.5, 0.3)\}$, $C = \{(x, 0.5, 0.2)\}$, we obtain $S_s^p(A,B) = S_s^p(A,C) = 0.95$ which seems difficult to accept.

S_h^p is given as

$$S_h^p(A,B) = 1 - (1/(3n)^{1/p})(\sum_{i=1}^{n}(\eta_1(i) + \eta_2(i) + \eta_3(i))^p)^{1/p} \quad (18)$$

where $\eta_1(i) = \phi_\mu(x_i) + \phi_\nu(x_i)$ (the same as for S_e^p),
$\eta_2(i) = m_A(x_i) - m_B(x_i))$ (the same as for S_{DC}),
$\eta_3(i) = \max(l_A(i), l_B(i)) - \min(l_A(i), l_B(i))$,
$l_A(i) = (1 - \nu_A(x_i) - \mu_A(x_i))/2$,
$l_B(i) = (1 - \nu_B(x_i) - \mu_B(x_i))/2$. But, again, there are counter-intuitive cases for this measure. For $A = (x, 0.3, 0.4)$, and $B = (x, 0.4, 0.3)$, i.e., for an element and its complement, we obtain $S_h^p(A,B) = 0.933$ (which seems to be rather too big a similarity for an element and its complement).

Hung and Yang [12] proposed the similarity measures S_{HY}^1, S_{HY}^2, S_{HY}^3 in which Hausdorff distances were employed:

$$S_{HY}^1(A,B) = 1 - d_H(A,B) \quad (19)$$

$$S_{HY}^2(A,B) = 1 - (e^{d_H(A,B)} - e^{-1})/(1 - e^{-1}) \quad (20)$$

$$S_{HY}^3(A,B) = (1 - d_H(A,B))/(1 + d_H(A,B)) \quad (21)$$

where
$d_H(A,B))$=$\sum_{i=1}^{n} \max(|\mu_A(x_i)$–$\mu_B(x_i)|, |\nu_A(x_i)$–$\nu_B(x_i)|)$
Unfortunately, (19)–(21) give counter-intuitive results (implied by the calculation of $d_H(A,B)$ - cf. Szmidt [22]). For example if $A = (x, 0.4, 0.5)$, $B = (x, 0.5, 0.4)$, $C = (x, 0.5, 0.3)$, $D = (x, 0.6, 0.4)$, $E = (x, 0.6, 0.3)$, $F = (x, 0.4, 0.3)$ then $S_{HY}^1(A,B) = 0.9$ (a counter-intuitive large similarity for A and its complement as $B = A^C$), and also $S_{HY}^1(C,D) = S_{HY}^1(C,E) = S_{HY}^1(C,F) = 0.9$. Next, $S_{HY}^2(A,B) = S_{HY}^2(C,D) = S_{HY}^2(C,E) = S_{HY}^2(C,F) = 0.85$, and also $S_{HY}^3(A,B) = S_{HY}^3(C,D) = S_{HY}^3(C,E) = S_{HY}^3(C,F) = 0.85$.

A straightforward attempt to calculate the similarity between A-IFSs just by adding the non-memberships values to the existing similarity measures for fuzzy sets is due to Hung and Yang [13]. Their measures (22) and (23) are the extension of Wang's measures [46]:

$$S_{w1}(A,B) = (1/n) \sum_{i=1}^{n} \frac{\min(\mu_A(x_i), \mu_B(x_i)) + \min(\nu_A(x_i), \nu_B(x_i))}{\max(\mu_A(x_i), \mu_B(x_i)) + \max(\nu_A(x_i), \nu_B(x_i))} \quad (22)$$

But, it is easy to give counter-examples again. For example, for $A = \{(x, 0, 0.5)\}$, $B = \{(x, 0.1, 0.5)\}$, $C = \{(x, 0, 0.6)\}$, we obtain $S_{w1}(A,B) = S_{w1}(A,C) = 0.8(3)$ (for different B and C we obtain the same result), for $A = \{(x, 0, 0.5)\}$, $B = \{(x, 0.18, 0.5)\}$, $C = \{(x, 0, 0.68)\}$, we obtain $S_{w1}(A,B) = S_{w1}(A,C) = 0.735$ (again - for different B and C the same result) etc., which seems to be difficult to accept (S_{w1} is not bijective).

$$S_{w2}(A,B) = (1/n) \sum_{i=1}^{n}(1 - 0.5(|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)|)) \quad (23)$$

But for $A = \{(x, 0, 0.5)\}$, $B = \{(x, 0, 0.4)\}$, $C = \{(x, 0, 0.6)\}$, we obtain $S_{w2}(A,B) = S_{w2}(A,C) = 0.95$ (again - for different B and C the same similarity), which seems to be difficult to accept (S_{w2} is not bijective).

Another straightforward extensions of fuzzy similarity measures proposed by Hung and Yang [13] for A-IFSS are measures (24), (25) and (26) (they are extensions of Pappis and Karacapilidis' [20] measures for fuzzy sets).

$$S_{pk1}(A,B) = \frac{\sum_{i=1}^{n}(\min(\mu_A(x_i), \mu_B(x_i)) + \min(\nu_A(x_i), \nu_B(x_i)))}{\sum_{i=1}^{n}(\max(\mu_A(x_i), \mu_B(x_i)) + \max(\nu_A(x_i), \nu_B(x_i)))} \quad (24)$$

For (24) the counter-intuitive examples are like for (22).

$$S_{pk2}(A,B) = 1 - 0.5(\max_i(|\mu_A(x_i) - \mu_B(x_i)|) + \max_i(|\nu_A(x_i) - \nu_B(x_i)|)) \quad (25)$$

For (25) it is easy to give counter-examples again - especially for one-element sets. For example, for $A = \{(x, 0, 0.5)\}$, $B = \{(x, 0.1, 0.5)\}$, $C = \{(x, 0, 0.6)\}$, we obtain $S_{pk2}(A,B) = S_{pk2}(A,C) = 0.95$ (for different B and C just the same result).

$$S_{pk3}(A,B) = 1 - \frac{\sum_{i=1}^{n}(|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)|)}{\sum_{i=1}^{n}(|\mu_A(x_i) + \mu_B(x_i)| + |\nu_A(x_i) + \nu_B(x_i)|)} \quad (26)$$

But for $A = \{(x, 0, 0.5)\}$, $B = \{(x, 0, 0.26)\}$, $C = \{(x, 0, 0.965)\}$, we obtain $S_{pks}(A,B) = S_{pk3}(A,C) = 0.68$ (again, for different B and C the same similarity), which seems to be difficult to accept.

ISBN: 978-989-95079-6-8

3.1 Why the measures presented may yield counter-intuitive results?

It is worth noticing that all the above measures were constructed to satisfy the following conditions:

$$S(A,B) \in [0,1] \quad (27)$$

$$S(A,B) = 1 \Longleftrightarrow A = B \quad (28)$$

$$S(A,B) = S(B,A) \quad (29)$$

$$If A \subseteq B \subseteq C, \quad then \quad S(A,C) \leq S(A,B) \; and \; S(A,C) \leq S(B,C) \quad (30)$$

Conditions (27)–(29) are obvious. The problem lies in (30) as this condition is meant as:

$$A \subset B \text{ iff } \forall x \in X, \; \mu_A(x) \leq \mu_B(x) \; and \; \nu_A(x) \geq \nu_B(x) \quad (31)$$

Unfortunately, (31) is not constructive and operational for A-IFSs as for many cases it can not be used. For example, for the elements $(x : (\mu, \nu, \pi))$: x_1: $(0.12, 0.4, 0.48)$ and x_2: $(0.1, 0.3, 0.6)$ we can not come to a conclusion. Moreover, element x: $(0, 0, 1)$ seems to be always beyond consideration in the sense of (31) which is very specific, and mostly practically irrelevant.

Furthermore, most of the similarity measures shown above are in fact similarity measures comparing just two intervals (each interval representing one of the A-IFSs under comparison). But we should bear in mind that elements of A-IF are described via the membership and non-membership function and the hesitation margin. In other words, in the terms of intervals, we have both the membership in an interval, and the non-membership in an interval so that we should represent an A-IF via two (not one) intervals.

Considering the representation of A-IFSs as single intervals implies some problems while calculating distances. Distances used in the (counter-intuitive) similarity measures mentioned in the previous section are calculated without taking into account the hesitation margins as the membership and non-membership functions only are taken into account. The counter-intuitive results obtained in such a case are in Szmidt and Kacprzyk [29], [39], Szmidt [22].

4 Some examples of intuitively justified and operational similarity measures

First we recall the measure of similarity between A-IFSs presented by Szmidt and Kacprzyk [37], [36]).

In the simplest situations we calculate the similarity of any two elements X and F belonging to an A-IF (A-IFSs). The proposed measures indicate if X is more similar to F or to F^C, where F^C is the complement of F. In other words, the proposed measures answer the question: is X more similar or more dissimilar to F?

Definition 1

$$Sim_{rule}(X,F) = \frac{l_{IF}(X,F)}{l_{IF}(X,F^C)} \quad (32)$$

where: $l_{IFS}(X,F)$ *is a distance from* $X(\mu_X, \nu_X, \pi_X)$ *to* $F(\mu_F, \nu_F, \pi_F)$*,*
$l_{IFS}(X,F^C)$ *is a distance from* $X(\mu_X, \nu_X, \pi_X)$ *to* $F^C(\nu_F, \mu_F, \pi_F)$*,*
F^C *is a complement of* F*, distances* $l_{IFS}(X,F)$ *and* $l_{IFS}(X,F^C)$ *are calculated from (6).*

For (32) we have

$$0 \leq Sim_{rule}(X,F) \leq \infty \quad (33)$$

$$Sim_{rule}(X,F) = Sim_{rule}(F,X)$$

The similarity has typically been assumed to be symmetric. Tversky [44], however, has provided some empirical evidence that the similarity should not always be treated as a symmetric relation. We stress this to show that a similarity measure (32) may have some features which can be useful in some situations but are not welcome in others (see Cross and Sudkamp [8], Wang et al. [47], Veltkamp [45]).

It is obvious (cf. Szmidt and Kacprzyk [36]) that the formula (32) can also be stated as

$$Sim_{rule}(X,F) = \frac{l_{IFS}(X,F)}{l_{IFS}(X,F^C)} = \frac{l_{IFS}(X^C,F^C)}{l_{IFS}(X,F^C)} = \frac{l_{IFS}(X,F)}{l_{IFS}(X^C,F)} = \frac{l_{IFS}(X^C,F^C)}{l_{IFS}(X^C,F)} \quad (34)$$

It is worth noticing that
– $Sim_{rule}(X,F) = 0$ means the identity of X and F.
– $Sim_{rule}(X,F) = 1$ means that X is to the same extent similar to F and F^C (i.e., values bigger than 1 mean a closer similarity of X and F^C to X and F).
– When $X = F^C$ (or $X^C = F$), i.e. $l_{IFS}(X,F^C) = l_{IFS}(X^C,F) = 0$ means the complete dissimilarity of X and F (or in other words, the identity of X and F^C), and then $Sim_{rule}(X,F) \to \infty$.
– When $X = F = F^C$ means the highest possible entropy (see [31]) for both elements F and X i.e. the highest "fuzziness" – not too constructive a case when looking for compatibility (both similarity and dissimilarity).
In other words, while applying the measure (32) to analyze the similarity of two objects, one should be interested in the values $0 \leq Sim_{rule}(X,F) < 1$.

The proposed measure (32) was constructed for selecting objects which are more similar than dissimilar [and well-defined in the sense of possessing (or not) attributes we are interested in]. In Szmidt and Kacprzyk [37] it was shown that a measure of similarity defined as above, (32), between $X(\mu_X, \nu_X, \pi_X)$ and $F(\mu_F, \nu_F, \pi_F)$ is more powerful than a simple distance between them. The following conclusion were drawn:

– when a distance between two (or more) objects/elements, or sets, is large, then it means for sure that the similarity does not occur.
– when a distance is small, we can say nothing for sure about similarity just on the basis of a distance between two objects [when not taking into account complements of the objects as in (32)]. The distance between objects can be small and the compared objects can be more dissimilar than similar.

We have shown on a simple example (cf. Szmidt and Kacprzyk [37]) that the measure (32) gives reasonable results when applied to assessing agreement in a group of experts. The only disadvantage of the proposed measure is that it does

ISBN: 978-989-95079-6-8

Table 1: Example results obtained from the similarity measures (35)–(38)

1	2	3	4	5
$X = (\mu, \nu, \pi)$	$(0.3, 0.4, 0.3)$	$(0.4, 0.2, 0.4)$	$(0.4, 0.2, 0.4)$	$(0, 0, 0)$
$F = (\mu, \nu, \pi)$	$(0.4, 0.3, 0.3)$	$(0.5, 0.3, 0.2)$	$(0.5, 0.2, 0.3)$	$(0.5, 0.5, 0)$
Sim_1	0	0.6	0.75	0.5
Sim_2	0	0.43	0.6	0.33
Sim_3	0	0.72	0.88	0.6
Sim_4	0	0.48	0.65	0.38

not follow the range of the usually assumed values for the similarity measures. But it is possible to construct a whole array of similarity measures following the philosophy, and preserving the advantages of the measure (32), and whose numerical values are consistent with the common scientific tradition (i.e. belonging to $[0, 1]$). For example:

$$Sim_1(X, F) = Sim_1(l_{IFS}(X, F), l_{IFS}(X, F^C)) = 1 - \frac{l_{IFS}(X, F)}{l_{IFS}(X, F) + l_{IFS}(X, F^C)} \quad (35)$$

$$Sim_2(X, F) = Sim_2(l_{IFS}(X, F), l_{IFS}(X, F^C)) = \frac{1 - f(l_{IFS}(X, F), l_{IFS}(X, F^C))}{1 + f(l_{IFS}(X, F), l_{IFS}(X, F^C))} \quad (36)$$

$$Sim_3(X, F) = Sim_3(l_{IFS}(X, F), l_{IFS}(X, F^C)) = \frac{1 - f(l_{IFS}(X, F), l_{IFS}(X, F^C))^2}{1 + f(l_{IFS}(X, F), l_{IFS}(X, F^C))^2} \quad (37)$$

$$Sim_4(X, F) = Sim_4(l_{IFS}(X, F), l_{IFS}(X, F^C)) = \frac{e^{-f(l_{IFS}(X,F), l_{IFS}(X,F^C))} - e^{-1}}{1 - e^{-1}} \quad (38)$$

where

$$f(l_{IFS}(X, F), l_{IFS}(X, F^C)) = = \frac{l_{IFS}(X, F)}{l_{IFS}(X, F) + l_{IFS}(X, F^C)} \quad (39)$$

and $0 \leq f(l_{IFS}(X, F), l_{IFS}(X, F^C)) \leq 1$.

The measures (35) – (38) give intuitive results. Some examples, being troublesome for other measures, are presented in Table 1. It is worth noticing that each measure assigns similarity equal 0 for an element $(0.3, 0.4, 0.3)$ and its complement $(0.4, 0.3, 0.3)$. In general, similarity measures (35) – (38) satisfy the following properties:

$$Sim_i(X, F) \in [0, 1], \quad (40)$$

$$Sim_i(X, X) = 1, \quad (41)$$

$$Sim_i(X, X^C) = 0, \quad (42)$$

$$Sim_i(X, F) = Sim_i(F, X), \quad (43)$$

for $i = 1, \ldots, 4$.

The similarity measures introduced in this section assess similarity of any two elements (X and F) belonging to an intuitionistic fuzzy set (or sets). The counterpart similarity measures for A-IFSs A and B containing n elements each, are:

$$Sim_k(A, B) = \frac{1}{n} \sum_{i=1}^{n} Sim_k(l_{IFS}(X_i, F_i), l_{IFS}(X_i, F_i^C)) \quad (44)$$

for $k = 1, \ldots, 4$.

5 Conclusions

We considered two groups of similarity measures between A-IFSs. First, we dealt with similarity measures constructed as if an A-IFSs was equal to a simple interval valued fuzzy set, or similarity measures being straightforward generalizations of those well known for the fuzzy sets. Unfortunately, in some situations, both approaches give counter-intuitive results. Second, we considered similarity measures accounting for all three functions describing an A-IF (the membership, non-membership, and hesitation margin) which is different from viewing an A-IF as a single interval. Next, we also took into account the complements of the elements compared. That is, we employed all kinds and fine shades of information available. It seems that these last measures are the most promising because, first of all, they help avoid some strong counter-intuitive results. This is crucial for both theory and applications.

References

[1] Atanassov K. (1983), Intuitionistic Fuzzy Sets. VII ITKR Session. Sofia (Centr. Sci.-Techn. Libr. of Bulg. Acad. of Sci., 1697/84) (in Bulgarian).

[2] Atanassov K. (1986) Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems, 20, 87–96.

[3] Atanassov K. (1999), Intuitionistic Fuzzy Sets: Theory and Applications. Springer-Verlag.

[4] Atanassov K. and Gargov G. (1989), Interval-valued intuitionistic fuzzy sets. Fuzzy sets and Systems, 31 (3), 343–349.

[5] Bustince H., Burillo P. (1996) Vague sets are intuitionistic fuzzy sets. Fuzzy Sets and Systems 67, 403–405.

[6] Chen S. M. (1995) Measures of similarity between vague sets. Fuzzy Sets and Systems 74(2), 217–223.

[7] Chen S. M. (1997) Similarity measures between vague sets and between elements. IEEE Trans. Syst. Mn Cybernet. 27(1), 153–158.

[8] Cross V. and Sudkamp T. (2002) Similarity and Compatibility in Fuzzy Set Theory. Physica-Verlag.

[9] Dubois D. On degrees of truth, partial ignorance and contradiction. Magdalena L., Ojeda-Aciego M., Verdegay J.M. (eds): Proc. IPMU08, pp. 31–38.

[10] Gau W.L., Buehrer D.J. (1993) Vague sets. IEEE Trans. Systems Man Cybernet 23, 610–614.

[11] Hong D.H. and Kim C. (1999) A note on similarity measures between vague sets and between elements. Inform Science 115, 83–96.

ISBN: 978-989-95079-6-8

[12] Hung W-L. and Yang M-S. (2004) Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance. Pattern recognition Letters 25, 1603–1611.

[13] Hung W.L and Yang M.S. (2008) On similarity measures between intuitionistic fuzzy sets. International Journal of Intelligent Systems 23 (3): 364–383.

[14] Kahneman D. (2002) Maps of bounded rationality: a perspective on intuitive judgment and choice. Nobel Prize Lecture.

[15] Li D.F, Cheng C.T. (2002) New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions. Pattern Recognition Letters 23, 221–225.

[16] Li Y., Zhongxian C., Degin Y. (2002) Similarity measures between vague sets and vague entropy. J. Computer Sci. 29(12), 129–132.

[17] Liang Z. and Shi P. (2003) Similarity measures on intuitionistic fuzzy sets. Pattern Recognition Lett. 24, 2687–2693.

[18] Mitchell H.B. (2003) On the Dengfeng-Chuntian similarity measure and its application to pattern recognition. Pattern Recognition Lett. 24, 3101–3104.

[19] Montero J., Gómez D. and Bustince H. (2007): *On the relevance of some families of fuzzy sets*. Fuzzy Sets and Systems 158, 2429–2442.

[20] Pappis C. P. and Karacapilidis N. (1993) A comparative assessment of measures of similarity of fuzzy values. Fuzzy Sets and Systems 56, 171–174.

[21] Sutherland S. (1994) Irrationality. The Enemy Within. Penguin Books.

[22] Szmidt E. (2006) On a problem with the distances based on the Hausdorff metric. Report RB/52/2006, IBS PAN.

[23] Szmidt E. and Baldwin J. (2003) New Similarity Measure for Intuitionistic Fuzzy Set Theory and Mass Assignment Theory. Notes on IFSs, 9 (3), 60–76.

[24] Szmidt E. and Baldwin J. (2004) Entropy for Intuitionistic Fuzzy Set Theory and Mass Assignment Theory. Notes on IFSs, 10(3), 15-28.

[25] Szmidt E. and Kacprzyk J. (1996c) Remarks on some applications of intuitionistic fuzzy sets in decision making, Notes on IF, 2(3), 22–31.

[26] Szmidt E. and Kacprzyk J. (1997) On measuring distances between intuitionistic fuzzy sets, Notes on IF, 3(4), 1–13.

[27] Szmidt E. and Kacprzyk J. (1998a) Group Decision Making under Intuitionistic Fuzzy Preference Relations. IPMU'98, 172–178.

[28] Szmidt E. and Kacprzyk J. (1998b) Applications of Intuitionistic Fuzzy Sets in Decision Making. EUSFLAT'99, 150–158.

[29] Szmidt E. and Kacprzyk J. (2000) Distances between intuitionistic fuzzy sets, Fuzzy Sets and Systems, 114(3), 505–518.

[30] Szmidt E. and Kacprzyk J. (2000) On Measures on Consensus Under Intuitionistic Fuzzy Relations. IPMU 2000, 1454–1461.

[31] Szmidt E., Kacprzyk J. (2001) Entropy for intuitionistic fuzzy sets. Fuzzy Sets and Systems, 118 (3), 467–477.

[32] Szmidt E. and Kacprzyk J. (2001) Analysis of Consensus under Intuitionistic Fuzzy Preferences. Proc. Int. Conf. in Fuzzy Logic and Technology. Leicester, UK, 79–82.

[33] Szmidt E. and Kacprzyk J. (2002) An Intuitionistic Fuzzy Set Based Approach to Intelligent Data Analysis (an application to medical diagnosis). In A. Abraham, L.Jain, Kacprzyk (Eds.) Recent Advances in Intelligent Paradigms and Applications. Springer-Verlag, 57-70.

[34] Szmidt E. and Kacprzyk J. (2002) Analysis of Agreement in a Group of Experts via Distances Between Intuitionistic Fuzzy Preferences. IPMU 2002, Annecy, France, 1859–1865.

[35] Szmidt E. and J. Kacprzyk J. (2002b) An Intuitionistic Fuzzy Set Based Approach to Intelligent Data Analysis (an application to medical diagnosis). In A. Abraham, L. Jain, J. Kacprzyk (Eds.): Recent Advances in Intelligent Paradigms and Applications. Springer-Verlag, 57–70.

[36] Szmidt E. and Kacprzyk J. (2004) Similarity of intuitionistic fuzzy sets and the Jaccard coefficient. IPMU 2004, 1405–1412.

[37] Szmidt E., Kacprzyk J. (2004) A Concept of Similarity for Intuitionistic Fuzzy Sets and its use in Group Decision Making. 2004 IEEE Conf. on Fuzzy Systems, Budapest, 1129–1134.

[38] Szmidt E. and Kacprzyk J. (2006) An Application of Intuitionistic Fuzzy Set Similarity Measures to a Multi-criteria Decision Making Problem. ICAISC 2006, LNAI 4029, Springer-Verlag, 314–323.

[39] Szmidt E. and Kacprzyk J. (2006) Distances Between Intuitionistic Fuzzy Sets: Straightforward Approaches may not work. 3rd Int. IEEE Conf. Intelligent Systems IS06, 716–721.

[40] Szmidt E. and Kacprzyk J. (2007). Some problems with entropy measures for the Atanassov intuitionistic fuzzy sets. Applications of Fuzzy Sets Theory. LNAI 4578, 291–297. Springer-Verlag.

[41] Szmidt E. and Kacprzyk J. (2007a). A New Similarity Measure for Intuitionistic Fuzzy Sets: Straightforward Approaches may not work. 2007 IEEE Conf. on Fuzzy Systems, 481–486.

[42] Szmidt E. and Kukier M. (2006). Classification of Imbalanced and Overlapping Classes using Intuitionistic Fuzzy Sets. 3rd International IEEE Conference on Intelligent Systems IS'06, London, 2006, 722-727.

[43] Szmidt E. and Kukier M. (2008) A New Approach to Classification of Imbalanced Classes via Atanassov's Intuitionistic Fuzzy Sets. In: Hsiao-Fan Wang (Ed.): Intelligent Data Analysis : Developing New Methodologies Through Pattern Discovery and Recovery. Idea Group, 85–101.

[44] Tversky A. (1977). Features of similarity. Psychol. Rev., 84, 327–352.

[45] Veltkamp R.C. (2001) Shape matching: similarity measures and algorithms. Proc. Shape Modelling International, Genova, Italy, IEEE Press, 188–197.

[46] Wang W.J. (1997) New similarity measures on fuzzy sets and on elements. Fuzzy Sets and Systems, 85, 305–309.

[47] Wang X., De Baets B., and Kerre E. (1995). A comparative study of similarity measures. Fuzzy Sets and Systems, 73 (2), 259–268.

[48] Zadeh L.A. (1965). Fuzzy sets. Information and Control, 8, 338–353.

ISBN: 978-989-95079-6-8

Non-commutative EQ-logics and their extensions

Vilém Novák, Martin Dyba

Institute for Research and Applications of Fuzzy Modelling, University of Ostrava
Ostrava, Czech Republic
Email: Vilem.Novak,martin.dyba@osu.cz

Abstract— *We discuss a formal many-valued logic called EQ-logic which is based on a recently introduced special class of algebras called EQ-algebras. The latter have three basic binary operations (meet, multiplication, fuzzy equality) and a top element and, in a certain sense, generalize residuated lattices. The goal of EQ-logics is to present a possible direction in the development of mathematical logics in which axioms are formed as identities. In this paper we propose a basic EQ-logic and three extensions which end up with a logic equivalent to the MTL-logic.*

Keywords— EQ-algebra, fuzzy equality, residuated lattice, MTL-logic, fuzzy logic.

1 Introduction

One of possible directions of the development of mathematical logic which can be tracked down to Leibnitz, Wittgenstein, and Ramsey (cf. [1]) is to develop it on the basis of identity (equality) as the principal connective. This direction has been crowned in a noble way by Henkin in [2] who developed the type theory (a higher order logic) using identity as a sole primitive constant. This is nice because it enables to treat equality between elements on the basis of the same principles not depending on their character. Thus, equality between truth values is, in principle, a relation of the same kind as an equality between other kinds of elements. Hence, we will speak about equality instead of equivalence even when speaking about truth values.

Note that there are works developing classical boolean logic on the basis of equivalence as the main (not sole) connective (a recent book presenting logic in this way is [3]). Moreover, the logic is developed there in the "equational style", i.e. proofs proceed as sequences of equations (in fact, equivalences).

This gives rise to an idea whether also fuzzy (many-valued) logic could be developed on the basis of fuzzy equality as the main connective where by fuzzy equality we mean, in fact, a fuzzy equivalence, i.e. a generalization of the classical logical equivalence. We will prefer the term "fuzzy equality", though. Recall that this idea is supported by the mentioned Henkin successful type theory (for a detailed development of this theory, see also [4]) as well as by the development of fuzzy type theory (see [5, 6]). On the other hand, unlike classical type theory, it seems impossible to have fuzzy equality a sole connective. At least conjunction is also necessary as follows from the considerations concerning the structure of the algebra of truth values. Namely, two approaches are apparent.

In the first approach, the structure of truth values is assumed to form a residuated lattice where the basic operations are, besides the lattice ones, the multiplication "$\otimes$" and its residuum "$\rightarrow$". The latter is in fuzzy logic a natural interpretation of implication which is a primary operation while the equivalence is interpreted by a biresiduation $a \leftrightarrow b = (a \rightarrow b) \wedge (b \rightarrow a)$ which is a derived operation. Since the basic connective, in our case, should be the fuzzy equality, it seems unnatural to interpret it by a derived operation.

The second approach follows an idea presented in [7, 8] where a specific kind of algebra called EQ-algebra is developed. Unlike the residuated lattices, the basic operation in EQ-algebras is a fuzzy equality "$\sim$" while implication is derived from it. Of course, a crucial question remains, which properties such an equality should have. At first sight it seems clear that these properties should be reflexivity, symmetry, and transitivity. We are convinced, however, that the fuzzy equality should be characterized by a deeper property reflecting some basic structure so that both symmetry as well as especially transitivity would follow. The most essential is, in our opinion, the "substitution principle" stating that if we replace an object by another one equal to the former then the result is not changed; in the case of fuzzy equality, the replacement should not make the structure in concern "worse" in some reasonable sense.

What basic structure should we consider in fuzzy logic? Clearly, this is *ordering* since the basic idea standing behind fuzziness is introducing degrees to enable measuring imprecision stemming from vagueness. Since ordering is in no way included in the equality, it must be introduced explicitly. Therefore, we will consider the basic structure to be a $\wedge$-semilattice with the greatest element $\mathbf{1}$. At the same time, we must "fuse" in some way statements about ordering and equality. The fusion is in our algebra realized by a third operation of multiplication "$\otimes$" whose properties, however, should be as weak as possible because we think that it should not participate on the ordering but enable only to "fuse things together". This is the main motivation behind axiom (E4) below. The symmetry and transitivity of $\sim$ are then consequences of it. The fusion $\otimes$ needs not be commutative and, perhaps, nor even associative.

Step by step we arrive at 8 axioms which determine the EQ-algebra. In its "crude form", there is no need to introduce the smallest element $\mathbf{0}$. Morever, the axioms make it possible to have two different truth values equivalent in the degree $\mathbf{1}$ and so, the fuzzy equality generalizes classical equivalence in an essential way. From the algebraic point of view, EQ-algebras generalize, in a certain sense, residuated lattices and so, they look as interesting class of algebraic structures.

When attempting to develop a formal logical calculus it came out, surprisingly, that the crude structure though reflecting the basic principles of fuzzy equality, is not strong enough. Namely, the logic enforced adding an axiom of "goodness" (E10) (or, as a logical formula (EQ1)) below which says that

ISBN: 978-989-95079-6-8

an element a is equal to $\mathbf{1}$ always in a degree a. This axiom implies that the algebra is separated, i.e. that two elements equal in the degree $\mathbf{1}$ must be identical in the classical sense. We conclude that the logic allows only a slight generalization of the classical equivalence. Of course, the goodness axiom has many important algebraic consequences (see [9, 8]). Among them is also the fact that each good EQ-algebra gives rise to a BCK-algebra. Putting all together we expect that EQ-algebras may put a different light on residuated lattices and on their general role in algebra and logic.

This paper is an attempt to introduce a formal logic called EQ-logic*) whose truth values are taken from a good EQ-algebra. Several basic properties of this logic demonstrating that it behaves reasonably are proved. In comparison with residuated fuzzy logics, however, our logic seems to be more limited. For example, the deduction theorem even in its weaker form probably does not hold and we guess that its validity is equivalent with residuation. Still, it is interesting to learn how far can we go with the development of EQ-logic. Besides others, it enables us to separate properties, which stem from the fuzzy equality itself, from the properties for which residuation is necessary. It may also shed light on classical logic, especially on its equational variant ([3]).

2 Definition and fundamental properties of EQ-algebras

Definition 1
A semicopula-based EQ-algebra $\mathscr{E}$ is an algebra of type (2, 2, 2, 0), i.e.

$$\mathscr{E} = \langle E, \wedge, \otimes, \sim, \mathbf{1} \rangle\,, \tag{1}$$

where for all $a, b, c, d \in E$:

(E1) $\langle E, \wedge, \mathbf{1}\rangle$ *is a commutative idempotent monoid (i.e. $\wedge$-semilattice with top element $\mathbf{1}$) where the ordering $a \le b$ is defined in standard way as $a \wedge b = a$,*

(E2) $\otimes$ *is a binary multiplication operation isotone w.r.t. $\le$, i.e. $a \le b$ implies both $a \otimes c \le b \otimes c$ as well as $c \otimes a \le c \otimes b$ and $\mathbf{1}$ is a neutral element, i.e. $a \otimes \mathbf{1} = \mathbf{1} \otimes a = a$.*

(E3) $a \sim a = \mathbf{1}$,

(E4) $((a \wedge b) \sim c) \otimes (d \sim a) \le c \sim (d \wedge b)$,

(E5) $(a \sim b) \otimes (c \sim d) \le (a \sim c) \sim (b \sim d)$,

(E6) $(a \wedge b \wedge c) \sim a \le (a \wedge b) \sim a$,

(E7) $(a \wedge b) \sim a \le (a \wedge b \wedge c) \sim (a \wedge c)$,

(E8) $a \otimes b \le a \sim b$.

The operation $\sim$ is called a fuzzy equality.

For all $a, b \in E$ we put

$$a \to b = (a \wedge b) \sim a. \tag{2}$$

This is a derived operation called *implication*†).

Note that axioms (E6) and (E7) can be rewritten as

$$a \to (b \wedge c) \le a \to b\,, \tag{E6'}$$
$$a \to b \le (a \wedge c) \to b\,. \tag{E7'}$$

The following theorem confirms that $\sim$ is indeed a fuzzy equality.

Theorem 1
Let $\mathscr{E}$ be an EQ-algebra. The following holds for all $a, b, c \in E$:

(a) Symmetry: $a \sim b = b \sim a$,

(b) Transitivity: $(a \sim b) \otimes (b \sim c) \le a \sim c$,

(c) Transitivity of implication: $(a \to b) \otimes (b \to c) \le a \to c$.

We say that the multiplication $\otimes$ is *$\to$-isotone* if

$$a \to b = \mathbf{1} \quad \text{implies} \quad a \otimes c \to b \otimes c = \mathbf{1} \tag{3}$$

for all $a, b, c \in E$. Let us remark that the $\to$-isotonicity of $\otimes$ is usually fulfilled. There exist EQ-algebras, however, which have not this property.

A semicopula-based EQ-algebra has the operation $\otimes$, in general neither commutative nor associative. If the latter is only non-commutative but associative then we will speak about a *non-commutative* EQ-algebra.

Theorem 2
Let $\mathscr{E} = \langle E, \wedge, \otimes, \sim, \mathbf{1}\rangle$ is a semicopula-based EQ-algebra and put $b \bar{\otimes} a = a \otimes b$. Then $\mathscr{E}' = \langle E, \wedge, \bar{\otimes}, \sim, \mathbf{1}\rangle$ is also a semicopula-based EQ-algebra.

The following lemmas summarizes some of the properties of EQ-algebras useful below. Its proof can be found in [9, 8].

Lemma 1
Let $\mathscr{E}$ be an EQ-algebra. For all $a, b \in E$ such that $a \le b$ it holds that

(a) $a \to b = \mathbf{1}$,

(b) $a \sim b = b \to a$,

(c) $c \to a \le c \to b$ *and* $b \to c \le a \to c$.

Lemma 2
Let $\mathscr{E}$ be an EQ-algebra. For all $a, b, c, d, a', b', c', d' \in E$ it holds that

(a) $a \otimes b \le a, a \otimes b \le a \wedge b, c \otimes (a \wedge b) \le (c \otimes a) \wedge (c \otimes b)$,

(b) $a \sim b \le a \to b, a \to a = \mathbf{1}$,

(c) $(a \to b) \otimes (b \to a) \le a \sim b$,

(d) $a = b$ *iff* $a \sim b = \mathbf{1}$,

(e) $a = \mathbf{1} \to a$ *and* $a \to \mathbf{1} = \mathbf{1}$,

(f) $a \otimes (a \sim b) \le b$,

(g) $b \le a \to b$,

(h) $((a \wedge b) \sim (c \wedge d)) \otimes (a \sim a') \otimes (b \sim b') \otimes (c \sim c') \otimes (d \sim d') \le (a' \wedge b') \sim (c' \wedge d')$.

*) This logic has been first introduced in Barcelona, October 2008 at the occasion of Prof. F. Esteva 65th birthday.

†) Note that this operation has been, in fact, introduced already by G. W. Leibnitz (cf. [10])

ISBN: 978-989-95079-6-8

Note that if $\mathscr{E}$ is separated, then $a \to b = \mathbf{1}$ implies $a \leq b$. Indeed, $a \to b = (a \wedge b) \sim a = \mathbf{1}$ implies $a \wedge b = a$, i.e. $a \leq b$.

Let $\mathscr{E}$ contain also the bottom element $\mathbf{0}$. Then we put

$$\neg a = a \sim \mathbf{0}, \qquad a \in E \tag{4}$$

and call $\neg a$ a *negation* of $a \in E$.

Definition 2
Let $\mathscr{E}$ be an EQ-algebra and $a, b, c, d \in L$. We say that $\mathscr{E}$ is:

(i) separated *if for all $a \in E$,*

(E9) $a \sim b = \mathbf{1}$ *implies* $a = b$.

(ii) good *if*

(E10) $a \sim \mathbf{1} = a$.

(iii) residuated *if for all $a, b, c \in E$,*

(E10) $(a \otimes b) \wedge c = a \otimes b$ *iff* $a \wedge ((b \wedge c) \sim b) = a$.

(iv) involutive *(IEQ-algebra) if for all $a \in E$,*

(E11) $\neg\neg a = a$.

(v) prelinear *if for all $a, b \in E$,*

(E12) $\sup\{a \to b, b \to a\} = \mathbf{1}$.

(vi) lattice EQ-algebra *(ℓEQ-algebra) if it is a lattice and*

(E13) $((a \vee b) \sim c) \otimes (d \sim a) \leq (d \vee b) \sim c$.

(vii) complete *if it's $\wedge$-semilattice reduct is complete.*

(viii) EQ(R)-algebra *if for all $a, b \in E$ there exist*

$$a/b = \max\{c \in E \mid a \otimes c \leq b\}$$
$$a\backslash b = \max\{c \in E \mid c \otimes a \leq b\}$$

(the latter equalities will be referred to as (R)-condition).

It is easy to show that good EQ-algebras are *separated*. Obviously, $\otimes$ in good EQ-algebras is $\to$-isotone. As already discussed, the "goodness property" appeared to be indispensable for logic and so, from now on, *by EQ-algebra we will always mean a good one*.

Remark 1
Let $\mathscr{E}$ be a semicopula-based EQ-algebra. A set $F \subseteq E$ is a filter if $\mathbf{1} \in F$; $a, b \in F$ implies $a \otimes b \in F$; $a, a \to b \in F$ implies $b \in F$; $a \to b \in F$ implies $(a \otimes c) \to (b \otimes c) \in F$ and $(c \otimes a) \to (c \otimes b) \in F$. Then it can be proved that a factor algebra $\mathscr{E}|F$ is a separated semicopula-based EQ-algebra. Surprizingly, there exist EQ-algebras with no proper filter and so, we cannot embed each EQ-algebra into a separated one.

The concept of EQ(R)-algebra is due to El Zekey [11]. Obviously, it inserts residuation inside EQ-algebra. In case that $E = [0, 1]$, it simply requires the operation $\otimes$ to be left-continuous.

Lemma 3
Let $\mathscr{E}$ be an IEQ-algebra. Then

(a) $\mathscr{E}$ is good ℓEQ-algebra with join defined by

$$a \vee b = \neg(\neg a \wedge \neg b).$$

(b) $a \leq b$ *iff* $\neg b \leq \neg a$.

(c) $a \sim b = \neg a \sim \neg b$.

(d) $a \otimes \neg a = \mathbf{0}$.

Theorem 3 ([11])
Let $\mathscr{E}$ be a good non-commutative EQ(R)-algebra. Then the following is equivalent:

(a) $\mathscr{E}$ is representable (i.e. subdirectly embeddable into a product of linearly ordered good EQ(R)-algebras[‡]).

(b) $\mathscr{E}$ satisfies the formula

$$(a \to b) \vee (d \to (d \otimes (c \to (b \to a) \otimes c))) = \mathbf{1}$$

for all $a, b, c, d \in E$.

3 Basic EQ-logic

In this section we introduce a propositional EQ-logic which we will call basic. This logic seems to be the simplest logic definable on the basis of EQ-algebras. Of course, we obtain more sophisticated logics when adding further axioms. From this point of view, all core (residuated) fuzzy logic in the sense of [12]) are extensions of the basic EQ-logic.

Definition 3
The language of EQ-logic consists of propositional variables $p_1, p_2, \ldots$, binary connectives $\wedge, \&, \equiv$ and a truth (logical) constant $\top$. Formulas are defined in the obvious way: each propositional variable is a formula, $\top$ is a formula and if A, B are formulas, then $A \wedge B, A \& B, A \equiv B$ are formulas. Implication is defined as a short

$$A \Rightarrow B := (A \wedge B) \equiv A. \tag{5}$$

Let J be a language of EQ-logic, F_J a set of all formulas in the language J and $\mathscr{E} = \langle E, \wedge, \otimes, \sim, \mathbf{1}\rangle$ be a good EQ-algebra. A truth evaluation $e : F_J \longrightarrow E$ is defined as usual: if $p \in J$ is a propositional variable then $e(p) \in E$. Furthermore, we put

$$e(\top) = \mathbf{1},$$
$$e(A \wedge B) = e(A) \wedge e(B)$$
$$e(A \& B) = (e(A) \otimes e(B)),$$
$$e(A \equiv B) = (e(A) \sim e(B))$$

for all formulas $A, B \in F_J$.

A formula $A \in F_J$ is a tautology if $e(A) = \mathbf{1}$ for each truth evaluation $e : F_J \longrightarrow E$.

Note that our logic has a non-commutative fusion connective $\&$ but only one implication because it is derived from equivalence (i.e. its interpretation is not residual operation adjoint with multiplication). We are convinced that this is an advantage.

[‡] A recent result of El-Zekey (personal communication) implies that the (R)-condition is unnecessary for the representability. This may lead to simplification of the structure of EQ-logics discussed below.

ISBN: 978-989-95079-6-8

3.1 Logical axioms and inference rules

The following formulas are axioms of the EQ-logic:

(EQ1) $(A \equiv \top) \equiv A$

(EQ2) $A \wedge B \equiv B \wedge A$

(EQ3) $(A \wedge B) \wedge C \equiv A \wedge (B \wedge C)$

(EQ4) $A \wedge A \equiv A$

(EQ5) $A \wedge \top \equiv A$

(EQ6) $A \,\&\, \top \equiv A$

(EQ7) $\top \,\&\, A \equiv A$

(EQ8a) $((A \wedge B) \,\&\, C) \Rightarrow (B \,\&\, C)$

(EQ8b) $(C \,\&\, (A \wedge B)) \Rightarrow (C \,\&\, B)$

(EQ9) $((A \wedge B) \equiv C) \,\&\, (D \equiv A) \Rightarrow (C \equiv (D \wedge B))$

(EQ10) $(A \equiv B) \,\&\, (C \equiv D) \Rightarrow (A \equiv C) \equiv (D \equiv B)$

(EQ11) $(A \Rightarrow (B \wedge C)) \Rightarrow (A \Rightarrow B)$

(EQ12) $(A \Rightarrow B) \Rightarrow (A \wedge C) \Rightarrow B$

Remark 2
The above axioms have been written using $\Rightarrow$ for better readability because formulas with this connective are more usual than those with $\equiv$. However, technically we apply equational-style proofs in the sense of [3], which are sequences of formulas of the form $A_1 \equiv A_2, \ldots, A_{n-1} \equiv A_n$ such that each of the individual theorems $A_i \equiv A_{i+1}$ has an independent individual proof. Therefore, a fussy presentation would require to rewrite all the axioms using $\equiv$ and the definition (5) only.

By $A[p := B]$ we denote a formula resulting from A by replacing all occurrences of a propositional variable p in A by the formula B. Then we introduce the following inference rules of EQ-logic:

$$(EA)\quad \frac{A, A \equiv B}{B} \qquad (L)\quad \frac{B \equiv C}{A[p := B] \equiv A[p := C]}$$

The rule (EA) is the *equanimity rule* and (L) is the *Leibnitz rule* for formulas (cf. [3] and elsewhere). The notion of provability is classical. A formal theory is any subset $T \subseteq F_J$. As usual, we suppose that T is defined by a set of special axioms.

3.2 Main properties

The proof of the following three lemmas is technical. They demonstrate several reasonable properties of the basic EQ-logic.

Lemma 4
The following are special derived rules:

(a) $A \equiv \top \vdash A$. *(rule (T1))*

(b) $A \vdash A \equiv \top$. *(rule (T2))*

(c) $A \wedge D \equiv C, A \equiv B \vdash B \wedge D \equiv C$. *(rule (C))*

(d) $(A \equiv D) \equiv C, A \equiv B \vdash (B \equiv D) \equiv C$. *(rule (E))*

(e) $A \,\&\, D \equiv C, A \equiv B \vdash B \,\&\, D \equiv C$. *(rule (F1))*

(f) $D \,\&\, A \equiv C, A \equiv B \vdash D \,\&\, B \equiv C$. *(rule (F2))*

Lemma 5
(a) $\vdash A \equiv A$,

(b) $A \equiv B \vdash B \equiv A$,

(c) $A \equiv B, B \equiv C \vdash A \equiv C$,

(d) $A, A \Rightarrow B \vdash B$, *(Modus Ponens)*

(e) $\vdash (\top \Rightarrow A) \equiv A$,

(f) $A \Rightarrow B, B \Rightarrow C \vdash A \Rightarrow C$,

(g) $A, B \vdash A \,\&\, B$,

(h) $\vdash (A \equiv B) \equiv (B \equiv A)$,

(i) $A \Rightarrow (B \equiv C), B \vdash A \Rightarrow C$,

(j) $\vdash (A \equiv D) \Rightarrow ((A \equiv B) \equiv (D \equiv B))$,

(k) $A \Rightarrow (B \equiv C), B \equiv D \vdash A \Rightarrow (D \equiv C)$,

(l) $A \Rightarrow (B \equiv C), C \equiv D \vdash A \Rightarrow (B \equiv D)$,

(m) $\vdash (A \equiv B) \Rightarrow (A \Rightarrow B)$,

(n) $A \equiv B, C \equiv D \vdash A \,\&\, C \equiv B \,\&\, D$,

(o) $A \Rightarrow B, C \Rightarrow D \vdash A \,\&\, C \Rightarrow B \,\&\, D$,

(p) $\vdash ((A \equiv B) \equiv C) \,\&\, (A \equiv D) \Rightarrow ((D \equiv B) \equiv C)$,

(q) $\vdash (A \equiv B) \,\&\, (C \equiv D) \Rightarrow (A \equiv C) \equiv (B \equiv D)$,

(r) $\vdash (A \Rightarrow B) \,\&\, (B \Rightarrow A) \Rightarrow (A \equiv B)$,

(s) $\vdash B \Rightarrow (A \Rightarrow B)$.

Lemma 6
(a) $\vdash (A \equiv B) \,\&\, (B \equiv C) \Rightarrow (A \equiv C)$,

(b) $\vdash (A \,\&\, (A \equiv B)) \Rightarrow B$

(c) $\vdash (A \Rightarrow B) \,\&\, (B \Rightarrow C) \Rightarrow (A \Rightarrow C)$,

(d) $\vdash (A \,\&\, (A \Rightarrow B)) \Rightarrow B$

(e) $\vdash B \Rightarrow (B \equiv \top)$,

(f) $A \Rightarrow (B \Rightarrow C) \vdash (A \,\&\, B) \Rightarrow C$,

(g) $\vdash A \,\&\, B \Rightarrow A \equiv B$,

(h) $\vdash (A \wedge B) \Rightarrow A$,

(i) $\vdash (C \Rightarrow A) \,\&\, (C \Rightarrow B) \Rightarrow (C \Rightarrow (A \wedge B))$.

By the straightforward verification we can prove the following two lemmas.

Lemma 7
All axioms of EQ-logic are tautologies.

Lemma 8
The deductive rules of EQ-logic are sound in the following sense. Let $e : F_J \longrightarrow E$ be a truth evaluation:

ISBN: 978-989-95079-6-8

(a) *If* $e(A) = \mathbf{1}$ *and* $e(A \equiv B) = \mathbf{1}$ *then* $e(B) = \mathbf{1}$.

(b) *If* $e(B \equiv C) = \mathbf{1}$ *then* $e(A[p := B] \equiv A[p := C]) = \mathbf{1}$ *for any formula* A.

The following is a standard Lindenbaum-Tarski technique. Put

$$A \approx B \quad \text{iff} \quad \vdash A \equiv B, \quad A, B \in F_J.$$

It follows from Lemmas 5(a), (h) and 6(a) that $\approx$ is an equivalence on F_J. Let us denote by $[A]$ an equivalence class of A and put $\bar{E} = \{[A] \mid A \in F_J\}$. If we define

$$\begin{aligned} \mathbf{1} &= [\top], \\ [A] \wedge [B] &= [A \wedge B], \\ [A] \otimes [B] &= [A \,\&\, B], \\ [A] \sim [B] &= [A \equiv B], \end{aligned}$$

then we obtain an algebra $\bar{\mathscr{E}} = \langle \bar{E}, \wedge, \otimes, \sim, \mathbf{1} \rangle$.

Lemma 9
The algebra $\bar{\mathscr{E}}$ *is a good semicopula-based EQ-algebra.*

Theorem 4 (Soundness)
The basic EQ-fuzzy logic is sound.

PROOF: This is a consequence of Lemmas 7 and 8. □

Theorem 5 (Completeness)
The following is equivalent for every formula A*:*

(a) $\vdash A$,

(b) $e(A) = \mathbf{1}$ *for every good semicopula-based EQ-algebra* $\mathscr{E}$ *and a truth evaluation* $e : F_J \longrightarrow E$.

PROOF: The implication (a) to (b) is soundness.

(b) to (a): By Lemma 9 the algebra $\bar{\mathscr{E}}$ of equivalence classes of formulas is a good semicopula-based EQ-algebra. Thus, if (b) holds then it holds also for $e : F_J \longrightarrow \bar{E}$. If $e(A) = \mathbf{1}$ then it means that $[A] = [\top]$, i.e. $\vdash A \equiv \top$ and so, $\vdash A$ by rule (T2). □

4 Extensions of the basic EQ-logic

In this section we will briefly discuss some extensions of the basic EQ-logic.

4.1 Involutive EQ-logic

This logic (we will speak about IEQ-logic) contains falsity and is characterized by the property of double negation which leads to some simplifications.

The language J is the same as that of basic EQ-logic with the exception that $\top$ is replaced by $\bot$. Furthermore, we introduce the following shorts of formulas:

$$\top := \bot \equiv \bot, \tag{6}$$

$$\neg A := A \equiv \bot, \tag{7}$$

$$A \vee B := \neg(\neg A \wedge \neg B). \tag{8}$$

(clearly, (7) is a *negation* and (8) a disjunction.

Axioms of IEQ-logic are (EQ2)–(EQ12) plus the following ones:

(EQ13) $(A \,\&\, B) \,\&\, C \equiv A \,\&\, (B \,\&\, C)$,

(EQ14) $(A \wedge \bot) \equiv \bot$,

(EQ15) $\neg\neg A \equiv A$.

Axiom (EQ14) characterizes the basic property of $\bot$ and it can be written as $\bot \Rightarrow A$ (ex falso quodlibet). Axiom (EQ1) is in this logic provable.

A contradiction is a formula $A \,\&\, \neg A$. Then we say that a theory T is *contradictory* if $T \vdash A \,\&\, \neg A$ for some formula $A \in F_J$.

A straightforward proof will give the following (classical) theorem:

Theorem 6
A theory T *is contradictory iff* $T \vdash A$ *for all* $A \in F_{J(T)}$.

Semantics of this logic is formed by IEQ-algebras.

Theorem 7 (Completeness)
The following is equivalent for every formula A*:*

(a) $\vdash A$,

(b) $e(A) = \mathbf{1}$ *for every IEQ-algebra* $\mathscr{E}$ *and a truth evaluation* $e : F_J \longrightarrow E$.

4.2 EQ(R)-fuzzy logic

This logic seems to be closest to the residuated fuzzy logics. In fact, it already introduces some kind of residuated structure which enables to prove a stronger variant of the completeness theorem.

The language of this logic is that of IEQ-logic extended by two binary connectives: "\" and "/" and the shorts (6) and

$$A \vee B := ((A \Rightarrow B) \Rightarrow B) \wedge ((B \Rightarrow A) \Rightarrow A).$$

Axioms are (EQ1)–(EQ14) plus the following:

(EQ16) $(A \Rightarrow B) \vee (D \Rightarrow (D \,\&\, (C \Rightarrow (B \Rightarrow A) \,\&\, C)))$

(EQ17) $C/(A \otimes B) \equiv (C/(B/A))$

(EQ18) $(A \otimes B)\backslash C \equiv ((A\backslash B)\backslash C)$

Semantics of this logic is formed by good non-commutative prelinear EQ(R)-algebras.

Using Theorem 3, the following can be proved:

Theorem 8 (Completeness)
For every formula $A \in F_J$ *the following is equivalent:*

(a) $\vdash A$.

(b) $e(A) = \mathbf{1}$ *for every truth evaluation* $e : F_J \longrightarrow E$ *and every linearly ordered good non-commutative EQ(R)-algebra* $\mathscr{E}$.

(c) $e(A) = \mathbf{1}$ *for every truth evaluation* $e : F_J \longrightarrow E$ *and every good non-commutative prelinear EQ(R)-algebra* $\mathscr{E}$.

ISBN: 978-989-95079-6-8

4.3 Residuated EQ-logic

This is stronger logic close to MTL-logic because of introducing the residuation property.

The language J is the same as that of basic EQ-logic. Its axioms are (EQ1)–(EQ12) plus the following:

(EQ19) $((A \,\&\, B) \equiv C) \,\&\, (D \equiv A) \Rightarrow (C \equiv (D \,\&\, B))$

Lemma 10
The following is provable residuated EQ-logic:

(a) $\vdash A \,\&\, B \equiv B \,\&\, A$,

(b) $(A \,\&\, B) \,\&\, C \equiv A \,\&\, (B \,\&\, C)$,

(c) $\vdash ((A \,\&\, B) \Rightarrow C) \equiv (A \Rightarrow (B \Rightarrow C))$.

By A^n we mean n-times repetition of A w.r.t. $\&$ (as usual).

Theorem 9 (Deduction)
For every theory T and formulas $A, B \in F_J$:

$T \cup \{A\} \vdash B$ *iff* *there is* $n \geq 1$ *such that* $T \vdash A^n \Rightarrow B$

Semantics for this logic is formed by residuated EQ-algebras. If we replace the constant $\top$ in the language J by $\bot$, take the definition (6) and add axiom (EQ14) and the axiom

(EQ20) $((A \Rightarrow B) \Rightarrow C) \Rightarrow (((B \Rightarrow A) \Rightarrow C) \Rightarrow C)$

then the resulting logic is equivalent with the MTL-logic.

5 Conclusion

In this paper we introduced a class of logics based on a new algebra of truth values called EQ-algebra. This algebra is specific by introducing fuzzy equality (equivalence) as the basic operation. Accordingly, the basic connective of EQ-logic is equivalence.

It is easy to show that BCK-axioms are provable in basic EQ-logic and so, basic EQ-logic is also a BCK-one (cf. [13]). However, since modus-ponens is a derived rule in the former, translation of properties of BCK-logic to EQ-logic is not straightforward. It is an open question whether there is a formal system equivalent with basic EQ-logic with modus ponens as its sole rule.

Acknowledgment

The research was supported by the project MSM 6198898701 of the MŠMT ČR.

References

[1] F.P. Ramsey. The foundations of mathematics. *Proceedings of the London Mathematical Society*, 25:338–384, 1926.

[2] L. Henkin. A theory of propositional types. *Fundamenta Math.*, 52:323–344, 1963.

[3] G. Tourlakis. *Mathematical Logic*. New York, J. Wiley & Sons, 2008.

[4] P. Andrews. *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof*. Kluwer, Dordrecht, 2002.

[5] V. Novák. On fuzzy type theory. *Fuzzy Sets and Systems*, 149:235–273, 2005.

[6] V. Novák. EQ-algebra-based fuzzy type theory and its extensions. *Logic Journal of the IGPL*, page (submitted).

[7] V. Novák. EQ-algebras in progress. In O. Castillo, editor, *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing*, pages 876–884. Springer, Berlin, 2007.

[8] V. Novák and B. de Baets. EQ-algebras. *Fuzzy Sets and Systems*, page (to appear), 2009.

[9] M. El-Zekey, V. Novák, and R. Mesiar. Semicopula-based EQ-algebras. *Fuzzy Sets and Systems*, page (submitted).

[10] H.N. Castañeda. Leibniz's syllogistico-propositional calculus. *Notre Dame Journal of Formal Logic*, XVII(4):338–384, 1976.

[11] M. El-Zekey. Representable EQ(R)-algebras. page (submitted).

[12] P. Hájek and P. Cintula. On theories and models in fuzzy predicate logics. *Journal of Symbolic Logic*, 71(3):863—-880, 2006.

[13] P. Cintula, P. Hájek, and R. Horčík. Formal systems of fuzzy logic and their fragments. *Annals of Pure and Applied Logic*, 150:40–65, 2007.

ISBN: 978-989-95079-6-8

Discriminant Analysis for fuzzy random variables based on nonparametric regression

Ana Colubi, M.A. Gil, [1] Gil González-Rodríguez, Wolfgang Trutschnig [2]

1.Department of Statistics and OR, University of Oviedo
Oviedo, Spain
2.Unit of Intelligent Data Analysis and Graphical Models, European Centre for Soft Computing
Mieres, Spain
Email: {colubi, magil}@uniovi.es, {gil.gonzalez, wolfgang.trutschnig}@softcomputing.es

Abstract*— *This communication is concerned with the problem of supervised classification of fuzzy data obtained from a random experiment. The data generation process is modelled through fuzzy random variables which, from a formal point of view, can be identified with a kind of functional random element. We propose to adapt one of the most versatile discriminant approaches in the context of functional data analysis to the specific case we handle. The discriminant analysis is based on the kernel estimation of the nonparametric regression. The results are applied to an experiment concerning fuzzy perceptions and linguistic labels

***Keywords*—** fuzzy data, random experiments, supervised classification, kernel estimation, nonparametric regression.

1 Introduction

In many random experiments, as for instance sociological surveys, ecological studies, etc., some characteristics of interest can be assessed in a more meaningful scale if the respondents or the experts are allowed to indicate the degree of precision/imprecision of their answers/judgment/perceptions by means of fuzzy sets (see, for instance, [3] and Section 4 for more details). The results of those experiments can be soundly modelled by means of fuzzy random variables in Puri & Ralescu's sense [16]. It should be remarked that fuzzy random variables in this sense are used to model fuzzy-set-valued random elements and that our goal regards the fuzzy random variables 'per-se'. However, there are other experiments in which the quantity of interest is an underlying real-valued random variables that cannot be precisely observed. In this cases, different approaches may be considered (see, for instance, [5]).

From a formal point of view fuzzy random variables can be identified with a special case of functional random variables, although with some particular features concerning the natural arithmetic and metric structure (see [7] for a deeper discussion). Functional Data Analysis has become an important area of research in recent years (see, for instance, [12, 17, 8, 19]), and as it was suggested in [7], it is possible to take advantage of some of the results developed for those elements to analyze fuzzy data.

As a first step in classification problems concerning fuzzy random variables, some unsupervised approaches have been considered in the literature (see, for instance, [10]). In this communication we deal with the supervised classification problem. That is, given a set of possible groups and a training random sample of fuzzy data of each group, the goal is to predict the group membership of a new fuzzy datum.

As it is indicated in [6], there are different approaches to this problem in the functional context. Most of them are based on modifications of the linear discriminant analysis, however to avoid the inconveniences that frequently arises from the presence of nonlinear class boundaries, we propose a nonparametric method inspired by [6].

The rest of the paper is organized as follows. In Section 2 we introduce the notation and the basic concepts. In Section 3 the classification approach is discussed. Section 4 is devoted to the empirical results, and finally in Section 5 we conclude with some remarks and open problems.

2 Preliminaries

We will denote by $\mathcal{F}_c(\mathbb{R}^p)$ the *class of fuzzy sets* $U : \mathbb{R}^p \to [0,1]$ whose α-levels U_α are nonempty compact convex subsets of $\mathbb{R}^p$ for all $\alpha \in [0,1]$, where $U_\alpha = \{x \in \mathbb{R}^p \,|\, U(x) \geq \alpha\}$ for all $\alpha \in (0,1]$, and $U_0 = \mathrm{cl}(\{x \in \mathbb{R}^p \,|\, U(x) > 0\})$.

Recently, a new class of metrics based on the generalization of the mid-point and the spread of an interval has been defined. This class is very intuitive and versatile and has good properties for the statistical analysis (see [18]).

The generalized mid-point and spread of a fuzzy set A is a way of identifying levelwise the center (location) and the extent (imprecision) by considering each direction in the multidimensional case through the unit sphere $\mathbb{S}^{p-1}$.

Formally, let $\alpha \in [0,1]$ and $u \in \mathbb{S}^{p-1}$, and calculate the lengths $\pi_u(A_\alpha)$ of all orthogonal projections of A_α on this direction, i.e.

$$\pi_u(A_\alpha) = \big[\underline{\pi}_u(A_\alpha)\,,\, \overline{\pi}_u(A_\alpha)\big] = \big[-s_{A_\alpha}(-u)\,,\, s_{A_\alpha}(u)\big]$$

where s stands for the support function of a nonempty convex compact set, that is, $s_{A_\alpha}(u) = \sup_{a \in A_\alpha} \langle u, a\rangle$ ($\langle\cdot,\cdot\rangle$ denoting the usual inner product in $\mathbb{R}^p$). Thus, the generalized mid-point and spread of A are defined as the functions $\mathrm{mid}_A, \mathrm{spr}_A : \mathbb{S}^{p-1} \times [0,1] \to \mathbb{R}$ so that

$$\mathrm{mid}_A(u,\alpha) = \mathrm{mid}_{A_\alpha}(u) = \frac{1}{2}\big(s_{A_\alpha}(u) - s_{A_\alpha}(-u)\big),$$

$$\mathrm{spr}_A(u,\alpha) = \mathrm{spr}_{A_\alpha}(u) = \frac{1}{2}\big(s_{A_\alpha}(u) + s_{A_\alpha}(-u)\big).$$

Note that in the interval case, $\mathbb{S}^{p-1} = \{-1,1\}$ and $\mathrm{mid}_A(-1,\alpha) = -\mathrm{mid}\, A$, $\mathrm{mid}_A(1,\alpha) = \mathrm{mid}\, A$, $\mathrm{spr}_A(-1,\alpha) = \mathrm{spr}_A(1,\alpha) = \mathrm{spr}\, A$ holds for all $\alpha \in [0,1]$.

ISBN: 978-989-95079-6-8

The generalized mid-point and spread are not defined as numbers summarizing the 'central tendency' and the 'imprecision', but as functions identifying the 'central points' and the 'imprecision' in the different directions of the Euclidean space. In this way, we get a meaningful characterization of the fuzzy sets alternative to the classical support functions.

The class of distances in [18] is defined from the distances between the level sets as a meaningful generalization of the Bertoluzza et al. metric [1], by considering L_2-type distances between the mid-points and the spreads. Specifically, for each level set $\alpha \in [0,1]$, we define

$$d_\theta^2\big(A_\alpha, B_\alpha\big) = \|\mathrm{mid}\, A_\alpha - \mathrm{mid}\, B_\alpha\|^2 + \theta\|\mathrm{spr}\, A_\alpha - \mathrm{spr}\, B_\alpha\|^2$$

where $\|\cdot\|$ is the usual L_2-norm in the space of the square-integrable functions $L^2(\mathbb{S}^{p-1})$, and $0 < \theta \leq 1$ determines the relative importance of the squared distance between the spreads in relationship with the squared distance between the mids.

The metric between fuzzy sets D_θ^φ is defined as a weighting mean (w.r.t. a probability measure φ with support $[0,1]$) of the distances between the level sets by

$$D_\theta^\varphi(A,B) = \left(\int_{[0,1]} d_\theta^2(A_\alpha, B_\alpha) d\varphi(\alpha)\right)^{1/2}$$

The weight measure φ reflects the intuitive (or subjective) interpretation of fuzzy sets - for instance one may treat every α-level as equally important (and therefore use the Lebesgue measure as weight measure φ), or give more mass to α-levels close to 1 or to α-levels close to 0.

If we denote by $\|\cdot\|_{\tilde{2}}$ the usual L_2-norm on the Hilbert space $\mathcal{H} = L^2(\mathbb{S}^{p-1} \times [0,1])$, we have the following alternative intuitive expression for the metric:

$$\big(D_\theta^\varphi(A,B)\big)^2 = \big\|\mathrm{mid}_{\,A} - \mathrm{mid}_{\,B}\big\|_{\tilde{2}}^2 + \theta\big\|\mathrm{spr}_{\,A} - \mathrm{spr}_{\,B}\big\|_{\tilde{2}}^2.$$

Let $(\Omega, \mathcal{A}, P)$ be a probability space. A *Fuzzy Random Variable* (FRV) can be identified with a Borel measurable mapping $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R}^p)$ (see [2, 16]). The concepts of induced probability distribution and independence are the same as for general metric-space valued random elements.

3 Nonparametric discriminant approach for fuzzy data

Assume that we have a population $(\Omega, \mathcal{A}, P)$, and for each individual we observe a fuzzy datum. In addition, each individual may belong to one of k different categories $g_1, \ldots, g_k$, and as learning sampling we have the fuzzy data and the group of n independent individuals. The goal is to find a rule that allows us to classify a new individual in one of the k groups from the fuzzy datum. For this purpose we suggest to use a nonparametric approach.

Formally, let $(\mathcal{X}, G) : \Omega \to \mathcal{F}_c(\mathbb{R}^p) \times \{g_1, \ldots, g_k\}$ be a random element in such a way that $\mathcal{X}(\omega)$ is a fuzzy datum and $G(\omega)$ is the membership group (g_1,... or g_k) of each individual $\omega \in \Omega$. Assume that we have n independent copies of $(\mathcal{X}, G)$ as training sample, that is, we have a random sample $\{\mathcal{X}_i, G_i\}_{i=1}^n$. As in a more general functional context (see [6]), we propose to estimate nonparametrically

$$P(G = g_j | \mathcal{X} = \widetilde{x})$$

for $j = 1, \ldots, k$, $\widetilde{x} \in \mathcal{F}_c(\mathbb{R}^p)$, and then to assign the new data to the class of higher estimated probability.

In order to find a reasonable estimator of the preceding probability consider first $\delta > 0$ and a closed ball $B(\widetilde{x}; \delta)$ defined in terms of the metric introduced in Section 2. From Bayes Theorem we have that

$$P(G = g_j | \mathcal{X} \in B(\widetilde{x}; \delta))$$

$$= \frac{P(X \in B(\widetilde{x}; \delta) | G = g_j) P(G = g_j)}{\sum_{l=1}^k P(X \in B(\widetilde{x}; \delta) | G = g_l) P(G = g_l)}$$

$$= \frac{P(D_\theta^\varphi(\mathcal{X}, \widetilde{x}) \leq \delta | G = g_j)) P(G = g_j)}{\sum_{l=1}^k P(D_\theta^\varphi(\mathcal{X}, \widetilde{x}) \leq \delta | G = g_l) P(G = g_l)}$$

For each $\widetilde{x}$, obviously $D_\theta^\varphi(\mathcal{X}, \widetilde{x})$ is a (real-valued) random variable. Let $F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}$ denote the distribution function of this variable in the group g_j. Consequently,

$$P(G = g_j | \mathcal{X} \in B(\widetilde{x}; \delta))$$

$$= \frac{F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(\delta) P(G = g_j)}{\sum_{l=1}^k F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_l}(\delta) P(G = g_l)}$$

If we assume (as it is natural in the nonparametric setting) that $F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}$ is uniformly continuous for all $j = 1, \ldots k$, then we have that there exists an underlying density function $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}$ and

$$\lim_{\delta \to 0} \frac{F_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(\delta)}{\delta} = f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)$$

As a result,

$$P(G = g_j | \mathcal{X} = \widetilde{x})$$

$$= \frac{f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0) P(G = g_j)}{\sum_{l=1}^k f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_l}(0) P(G = g_l)}.$$

Note that the denominator is just $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$, and for this reason according to [6] we could estimate $P(G = g_j | \mathcal{X} = \widetilde{x})$ by means of

$$\widehat{P}(G = g_j | \mathcal{X} = \widetilde{x})$$

$$\frac{\sum_{i=1}^n I_{G=g_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}{\sum_{i=1}^n K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}$$

$$= \frac{(nh)^{-1} \sum_{i=1}^n I_{G=g_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}{(nh)^{-1} \sum_{i=1}^n K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}$$

ISBN: 978-989-95079-6-8

$$= \frac{\left[(n_j h)^{-1} \sum_{i \in N_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))\right] \cdot (n_j n^{-1})}{(nh)^{-1} \sum_{i=1}^{n} K(h^{-1} D_\theta^\varphi(\mathcal{X}_i, \widetilde{x}))}$$

where $N_j = \{i \in \{1, \ldots, n\} | G_i = g_j\}$, n_j is the cardinality of N_j, K is a kernel and h the bandwidth.

The last expression shows clearly that the estimator depends on an estimate of $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)$ (the densities within each group at the point 0), the empirical estimate of $P(G = g_j)$ and an estimate of $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$ (the overall density at the point 0). In this approach the bandwidth is the same for all the estimates, and it is determined by employing the whole sample. In this way, it is verified that

$$\sum_{j=1}^{k} \widehat{P}(G = g_j | \mathcal{X} = \widetilde{x}) = 1.$$

However, since that $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$ is composed by a mixture of the variables $\{D_\theta^\varphi(\mathcal{X}, \widetilde{x}) | G = g_j\}_{j=1}^k$ whose distributions are expected to be different (recall that the aim is to discriminate among them), it seems more convenient to estimate separately each distribution (each one of them with its own bandwidth) and then to combine them to estimate $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})}(0)$. Thus, we propose to estimate $f_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)$ by means of

$$\widehat{f}_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0) = (n_j h_j)^{-1} \sum_{i \in N_j} K(h^{-1} D_\theta^\varphi(\mathcal{X}, \widetilde{x}))$$

for all $j = 1, \ldots, k$. In this way,

$$\widehat{P}(G = g_j | \mathcal{X} = \widetilde{x}) = \frac{n_j n^{-1} \widehat{f}_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_j}(0)}{\sum_{l=1}^{k} (n_j n^{-1}) \widehat{f}_{D_\theta^\varphi(\mathcal{X},\widetilde{x})|G=g_l}(0)},$$

and it is also verified that

$$\sum_{j=1}^{k} \widehat{P}(G = g_j | \mathcal{X} = \widetilde{x}) = 1.$$

The considered estimators are statistically consistent under the unique assumption that the corresponding densities exist for each $\widetilde{x} \in \mathcal{F}_c(\mathbb{R}^p)$ (and the usual regularity conditions are satisfied), however there is no assumption about the changes in the distributions as $\widetilde{x}$ varies.

4 Empirical results

In an experiment about the visual perception of the length of a rule relatively to a maximum, we have shown images like those in Figure 1 to different people. The aim has been to know the relative measure of the dark line (below)in contrast with the light line (above) for these people. They have been allowed to express the lack of precision that they may have by using a fuzzy scale. After a trial with the software, people have seen different rules with random length and they have indicated their perception about the length in the fuzzy scale and the linguistic label (very small, small, medium, large, very large) that they consider suitable for such a length.

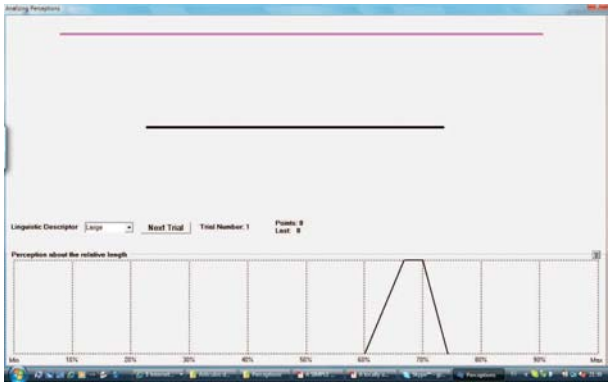


Figure 1: Software to evaluate the visual perception of a rule

The way of using fuzzy scale as explained in the software is as follows:

"*This experiment regards your perception about the relative length of different lines.*

On the top of the screen, we have plotted in light color the longest line that we could show to you. This line will remain visible in the current position during all the experiment, so that you can always have a reference of the maximum length.

At each trial of the experiment we will show you a dark line and you will be asked about its relative length (in comparison with the length of the reference light line):

- *Firstly you will be asked for a linguistic descriptor of the relative length. We have consider five descriptors (Very Small; Small; Medium; Large; Very Large). The aim is to select one of these descriptors at first sign (you can change it later if you want to).*
- *Secondly you will be asked for your own estimate or perception (without physically measuring it) of the relative length (in percentage) by means of a Fuzzy Set (the information about the design and interpretation of the Fuzzy Set will be shown to you at this time).*
- *Finally, in case your initial perception had been changed during the process you can readjust again the linguistic descriptor of its relative length.*"

Concerning the design and interpretation of the fuzzy set, when the image of Figure 1 is shown, the respondents have been asked to draw a fuzzy set representing their perceptions. In accordance with the usual interpretation, the respondents have to choose the 0-level (set of all those points with a positive degree of membership) as the set of all values that they consider compatible with the length to a greater or lesser extent. In the same way, the 1-level (set of all those points with total degree of membership) has to be fixed as the set of values that they consider completely compatible with their perception about the measure of the line. Although it is possible to change the shape of the resulting fuzzy sets, by default the trapezoidal fuzzy set formed by the interpolation of both intervals is fixed (as shown in Figure 1).

In Table 1 we show some of the data of a person who made 220 trials.

The goal here is to predict the category (very small, small, medium, large or very large) that this person would consider

ISBN: 978-989-95079-6-8

Table 1: Perceptions about the relative length of the rule.

Trial	$\inf P_0$	$\inf P_1$	$\sup P_1$	$\sup P_0$	Ling. descrip.
1	78.27	80.94	84.41	87.40	large
2	54.93	58.00	62.20	65.67	large
3	47.25	49.43	50.89	53.31	medium
4	92.65	95.72	97.58	99.11	very large
5	12.92	15.51	17.77	20.03	very small
6	32.55	36.03	39.90	42.89	small
7	2.50	4.44	6.22	9.21	very small
8	24.80	28.19	30.45	33.28	small
9	55.17	58.40	61.79	65.75	large
10	2.26	3.63	5.57	8.08	very small

suitable from the fuzzy perception that he/she has about the length of the line. It should be remarked that the categories are treated here simply as different classes, which may be also labelled as 1, 2, 3, 4 and 5, irrespectively of the fuzzy representation that they may have. The consideration of fuzzy labels would lead to a different approach.

The line showed at each trial has been chosen at random, although to obtain also a validation sample we have proceeded as follows:

- 166 lengths were generated by means of random numbers between 0 and 100. They will be used as training sample.
- The 9 lengths in the equally spaced discrete set $\{100/27 + (i/8)100(1 - 2/27)\}_{i=0,\ldots,8}$ have been repeated 6 times. Thus, we have 54 lengths that are representative of the different situations that may arise, and that can be used as validation sample.
- The 166 random lengths of the training sample and the 54 lengths of the validation sample are interspersed and shown at random.

As it was explained in Section 3, we need to estimate the density of the random variable $D_\theta^\varphi(\mathcal{X}, \widetilde{x})$ at the point 0 for any fuzzy response $\widetilde{x}$ we obtain. The problem of estimating a density in the boundary of its support is not easy. Several approaches can be found in the literature (see for instance, [15], [13] and [11]), however none of them have solved all the inconveniences that may arise. We have considered different options to evaluate its performance in this case study:

M1: Firsty a kernel density estimation with Quartic Kernel

$$K_1(u) = \frac{15}{16}(1-u^2)^2 I_{(|u|\leq 1)}.$$

and bandwidth chosen by cross-validation were considered. It is well-known that this estimator is non consistent in the boundaries. In particular the estimation at 0 in this case is not good in general (see for instance, [15]). As we have mentioned, several boundary corrections to overcome this problem are available. Many of these corrections (basically based on Jacknife) lead to consistent estimators at the boundaries, although taking negative values in many practical situations (which makes them useless in our case). There are some exceptions in the literature that yield to both consistent and non-negative estimates.

M2: In particular, as a second scenario we have considered the proposal in [6] of using an asymmetric kernel. To be precise, we consider the absolute value of the Quartic kernel K_1, that is,

$$K_2(u) = \frac{15}{8}(1-u^2)^2 I_{(u\in[0,1])}.$$

This approach leads to consistent estimates at the boundary (but with a low convergence speed ratio). There are no studies about the optimal way of choosing the bandwidth for this asymmetric kernel, so we have used the common one provided by cross-validation techniques applied to this particular case.

M3: Finally, as a third strategy we have considered the estimator proposed by [14] which is a kernel density estimation with boundary correction that provides with non-negative estimates, that is,

$$f_P(0) = \overline{f}(0)\exp\left\{\frac{\hat{f}(0)}{\overline{f}(0)} - 1\right\}$$

where $\hat{f}(0)$ denotes the basic kernel density estimator divided by $a_0(0) = \int_{-1}^{0} K(u)du$ and $\overline{f}$ is the boundary corrected kernel density estimator based on generalised jackknifing by combining $\hat{f}$ with the analogous estimator based on the kernel $L(u) = 0.5K(2u)$. In this case and optimal (local) bandwidth is the solution of a complex variational problem (see [15] for a similar situation), that in general is not possible to obtain. For this reason we have considered the global cross-validation bandwidth for the kernel $K = K_1$.

In Table 2 we show the percentage of right classification corresponding to each one of the considered methods. We observe a better behaviour for **M2** than for **M3**. In addition **M1** is quite similar to **M3** in spite of being theoretically inconsistent. This seems to indicate that the selection of a right bandwidth is more important than the theoretical accuracy of the considered estimator (**M1** and **M2** are worse from a theoretical point of view than **M3**, provided that the optimal bandwidth is considered).

Table 2: Percentage of right classification with different methods.

	M1	**M2**	**M3**
% of right classification	75.92	83.33	77.77

5 Concluding remarks

This works is a preliminary study concerning the problem of supervised classification of fuzzy random variables. We propose to use a nonparametric approach to better capture nonlinear boundaries. However, other interesting viewpoints may be used (either by extending those in functional data analysis, as the penalized or flexible discriminant analyses, or by being developed ad-hoc for this case). Further theoretical and empirical comparative studies should be developed.

As it is mentioned in Section 4, the nonparametric estimation problem to be solved here is not easy, and there is no

ISBN: 978-989-95079-6-8

general optimal method to do it. Our empirical results seem to indicate the importance of the bandwidth selection. As it is well-known, the optimal bandwidth for the estimation of a density function is not appropriate in general for the estimation of its curvature. In the same way, the optimal bandwidth for the estimation of a density function is non necessarily appropriated for the classification problem. As an open problem, we propose to optimize the bandwidth selection for the classification problem by choosing, for instance, the values maximizing the proportion of right classifications.

On the other hand, it seems also very interesting to consider the case in which the group membership of the training data is imprecise, as it is done in [4] for real learning data.

Acknowledgment

The research in this communication has been partially supported by the Grant from the Spanish Ministry of Education and Science MTM2006-07501. This financial support is gratefully acknowledged.

References

[1] Bertoluzza, C., Corral, N., Salas A., 1995. On a new class of distances between fuzzy numbers. Mathware & Soft Computing 2, 71–84.

[2] Colubi, A., Domínguez-Menchero, J.S., López-Díaz M., Ralescu, D. A. On the formalization of fuzzy random variables, Inform. Sci. 133 (2001) 3–6.

[3] Colubi, A. 2009. Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data. Fuzzy Sets and Systems. 160 (3), 344-356.

[4] Côme, E., Oukhellou, L., Denœux, T., Aknin, P. 2009 Learning from partially supervised data using mixture models and belief functions. Pattern Recognition 42 (3), pp. 334-348

[5] Denœux, T., Masson, M.-H., Hébert, P.-A. 2005 Nonparametric rank-based statistics and significance tests for fuzzy data Fuzzy Sets and Systems 153 (1), 1-28.

[6] Ferraty, F., Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Comput. Statist. Data Anal.* **44**, 161–173.

[7] Flórez, J.L., Colubi, A., González-Rodríguez, G., Gil, M.A. 2006. Nonparametric regression with fuzzy random variables through the support functions. In Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty (IPMU'2006, Paris), pp. 724-730.

[8] González-Manteiga, W.G., Vieu, P. Statistics for Functional Data. Computational Statistics and Data Analysis 51 (10), pp. 4788-4792. 2007.

[9] González-Rodríguez, G., Colubi, A., Trutschnig, W. 2009. Simulation of fuzzy random variables. Information Sciences. 179 (5), 642–653.

[10] González-Rodríguez, G., Colubi, A., D'Urso, P., Montenegro, M. 2009. Multi-sample test-based clustering for fuzzy random variables. International Journal of Approximate Reasoning.

[11] Hall, P., Park, B.U. 2002. New methods for bias correction at endpoints and boundaries. The Annals of Statistics 30 (5), 14601479.

[12] Hall, P., Müller, H.-G., Wang, J.-L. 2006. Properties of principal component methods for functional and longitudinal data analysis. Annals of Statistics 34 (3), pp. 1493-1517

[13] Jones, M. C., Foster, P. J. 1993. Generalized jackknifing and higher order kernels. J.Nonparametric Statist. 3, 81-94.

[14] Jones, M. C., Foster, P. J. 1996. A simple nonnegative boundary correction method for kernel density estimation. Statistica Sinica 6(1996), 1005-1013.

[15] Müller, H.-G. 1991. Smooth optimum kernel estimators near endpoints. Biometrika 78, 521-530.

[16] Puri, M. L., Ralescu, D. A. 1986. Fuzzy random variables. J. Math. Anal. Appl. 114, 409–422.

[17] Ramsay, J. O., Silverman, B. W. (2002). *Applied functional data analysis. Methods and case studies.* Springer Series in Statistics. Springer-Verlag, New York.

[18] Trutschnig, W. González-Rodríguez, G., Colubi, A., Gil, M.A. 2009. A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. Submitted.

[19] Zhang, J.-T., Chen, J. Statistical inferences for functional data. Annals of Statistics 35 (3), pp. 1052-1079. 2007

ISBN: 978-989-95079-6-8

Confidence regions for the mean of a fuzzy random variable

Gil González-Rodríguez, Wolfgang Trutschnig[1], Ana Colubi[2]

1. European Centre for Soft Computing
Mieres, Spain
2. Department of Statistics and OR, University of Oviedo
Oviedo, Spain
Email: {gil.gonzalez,wolfgang.trutschnig}@softcomputing.es, colubi@uniovi.es

Abstract— *The aim of this paper is to extend the classical problem of confidence interval estimation for the mean of a random variable to the case of a fuzzy random variable. The key idea consists in considering a confidence region defined as a ball w.r.t. a given metric, which is centered in the sample mean and whose radius is determined via bootstrapping. The developed approach is illustrated both by means of simulated examples using C and R and by means of a practical example consisting of courses evaluations.*

Keywords— Fuzzy random variables, confidence interval, confidence region, expectation, bootstrap.

1 Introduction

Inferential procedures for handling fuzzy information have gained in importance in the literature during the last years. Hypothesis testing problems concerning fuzzy random variables (hereafter FRV for short) in different situations (see, for instance, [9, 13, 14, 5, 6, 3]) as well as consistency properties of the considered estimators, i.e. the associated point-wise estimation problems, have been studied extensively (see, for instance, [12, 10, 16, 7]). Nevertheless, the problem of confidence interval estimation for FRVs has not received too much attention. In this communication we will focus on the the problem of estimating by a confidence interval the expected value of an FRV in Puri & Ralescu's sense [15].

Initially, there are two main difficulties in connection with this estimation problem: the well-known lack of ordering and the lack of linearity of the space of fuzzy sets. The lack of ordering is not crucial - indeed this difficulty has been faced previously in some classical situations (as for instance in the determination of confidence bands for the linear regression). On the other hand the lack of linearity makes it senseless to attempt to express the confidence interval in the standard way (sample mean plus/minus a given quantity depending on the variability and the sample size). Consequently, we will speak about confidence regions instead of confidence intervals in the following, keeping in mind that these regions cannot be specified simply by means of an upper and lower bound.

It is well-known that if X is a real-valued random variable with mean μ and finite variance, then, based on a random sample $X_1, \ldots, X_n$ of n independent random variables having the same distribution as X, a confidence interval with confidence level $100(1-\beta)\%$ for μ, can be determined by

$$IC_n = [\overline{X} - \delta, \overline{X} + \delta],$$

whereby $\overline{X}$ is the sample mean and $\delta = \delta(X_1, \ldots, X_n)$ is so that $P(\mu \in IC_n) = 1 - \beta$. The keys are, firstly, that this confidence interval can be seen as a ball with respect to the Euclidian distance which is centered in the sample mean $\overline{X}$ and has radius δ and, secondly, that in practice δ can be computed via bootstrapping. In other words: classical confidence intervals for the mean μ can equivalently be seen as balls, having the sample mean as center and a suitable radius.

Having this is in mind, our goal is to consider a *good* metric (that is, both easy to handle and easy to interpret) on the space of all fuzzy sets and, on the basis of this metric, to define confidence regions as balls centered in the sample mean and with a given radius which is empirically determined by applying bootstrap procedures. By the comments made before this technique surely is a natural extension of the classical procedure for real random variables.

The rest of the manuscript is organized as follows. In Section 2 we will recall some preliminary concepts concerning FRVs. At the beginning of Section 3 we propose an initial algorithm to determine a confidence ball for the mean of an FRV by using bootstrap. After checking its capability by means of various simulations made in C and R we finally present an improvement of the algorithm. In Section 4 we will apply the developed procedure to a concrete fuzzy sample consisting of courses evaluations. Finally, some concluding remarks, next developments and open problems are gathered in Section 5.

2 Preliminaries

Fuzzy random variables in Puri and Ralescu's sense [15] are an extension of both random variables and random (convex compact) sets. They allow us to model situations in which an imprecise (fuzzy) value is associated with the outcome of a stochastic experiment. In order to give an exact definition of FRVs we need some preliminary notions.

$\mathcal{K}_c(\mathbb{R})$ will denote the family of all non-empty compact intervals. $\mathcal{F}_c(\mathbb{R})$ denotes the class of fuzzy numbers, i.e. functions $U : \mathbb{R} \to [0,1]$, whose α-levels U_α fulfill $U_\alpha \in \mathcal{K}_c(\mathbb{R})$ for all $\alpha \in [0,1]$ whereby, as usual, $U_\alpha := [\underline{U}_\alpha, \overline{U}_\alpha] := \{x \in \mathbb{R} | U(x) \geq \alpha\}$ for every $\alpha \in (0,1]$ and $U_0 := [\underline{U}_0, \overline{U}_0]$ is defined as the topological closure of $\cup_{\alpha \in (0,1]} U_\alpha$.

The class $\mathcal{F}_c(\mathbb{R})$ is usually endowed with a semilinear structure that is compatible with Zadeh's Extension Principle (see [17]) - i.e. with a *sum* (extending the Minkowsky addition of intervals), defined by $(U+V)_\alpha = U_\alpha + V_\alpha$ for all $\alpha \in [0,1]$, and with a *product by a scalar*, $(aU)_\alpha = aU_\alpha$ for all $\alpha \in [0,1]$ and $a \in \mathbb{R}$. For every $U \in \mathcal{F}_c(\mathbb{R})$ mid_U and spr_U are defined to be real-valued functions on $[0,1]$ such that $mid_U(\alpha) := (\underline{U}_\alpha + \overline{U}_\alpha)/2$ and $spr_U(\alpha) := (\overline{U}_\alpha - \underline{U}_\alpha)/2$

ISBN: 978-989-95079-6-8

for every $\alpha \in [0,1]$.
A very useful metric on $\mathcal{F}_c(\mathbb{R})$ from the statistical point of view is the one introduced by Bertoluzza *et al* [1]. It depends on two weighting probability measures W and φ, defined on $([0,1], \mathcal{B}_{[0,1]})$ ($\mathcal{B}_{[0,1]}$ being the Borel σ-field on $[0,1]$). For any two fuzzy numbers $U, V \in \mathcal{F}_c(\mathbb{R})$ the D_W^φ-distance is defined as:

$$\sqrt{\int_{[0,1]^2} \left[t(\underline{U}_\alpha - \underline{V}_\alpha) + (1+t)(\overline{U}_\alpha - \overline{V}_\alpha)\right]^2 \, dW(t)d\varphi(\alpha)}$$

Thereby φ is assumed to have a strictly increasing distribution function on $[0,1]$ (and therefore has the whole interval $[0,1]$ as support) and W is non-degenerate (i.e. it is not a Dirac measure). A especially useful subclass of choices for W, on that we will concentrate in the sequel is that for which $\int_0^1 t dW(t) = 1/2$ (which in fact is equivalent to assuming D_W^φ to be invariant to rigid motions on $\mathbb{R}$). In this case it is easy to see that $(D_W^\varphi(U,V))^2$ can be rewritten as follows:

$$(D_W^\varphi(U,V))^2 = \int_{[0,1]} (\mathrm{mid}_U(\alpha) - \mathrm{mid}_V(\alpha))^2 d\varphi(\alpha)$$

$$+\theta_W \int_{[0,1]} (\mathrm{spr}_U(\alpha) - \mathrm{spr}_V(\alpha)^2 d\varphi(\alpha)$$

That is, the squared D_W^φ-distance of $U, V \in \mathcal{F}_c(\mathbb{R})$ is the sum of the squared φ-weighted L_2-distances between the mids and between the spreads, where the parameter $\theta_W \in (0,1]$ allows us to weight the relative importance of the distance between mids w.r.t. the distance between spreads (in particular, when W is chosen as the Lebesgue measure λ then $\theta_W = 1/3$).

Having this, we can now return to the definition of an FRV mentioned before. Let $(\Omega, \mathcal{A}, P)$ be an arbitrary probability space. an FRV $\mathcal{X}$ is a mapping $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R})$, which is Borel-measurable with respect to the D_W^φ-metric. This definition is different from the original one by Puri and Ralescu, however in many cases and in particular in the cases we will consider in the sequel coincides with the original one (see [15, 11, 2]).
The FRV $\mathcal{X}$ is said to be integrably bounded whenever $\max\{|\min X_0|, |\max X_0|\} \in L^1(\Omega, \mathcal{A}, P)$. If $\mathcal{X}$ is an integrably bounded FRV then its *expected value (or mean)* is defined as the unique fuzzy number $\mathbb{E}(\mathcal{X}) \in \mathcal{F}_c(\mathbb{R})$ such that $\big(\mathbb{E}(\mathcal{X})\big)_\alpha$ is the Aumman integral of the random set $\mathcal{X}_\alpha$ for all $\alpha \in [0,1]$ (see [15]). In particular,

$$\big(\mathbb{E}(\mathcal{X})\big)_\alpha = [\mathbb{E}(\min \mathcal{X}_\alpha), \mathbb{E}(\max \mathcal{X}_\alpha)]$$

holds for all $\alpha \in [0,1]$ if $\mathcal{X}$ is integrably bounded.
If, in addition, $\max\{|\min \mathcal{X}_0|, |\max \mathcal{X}_0|\} \in L^2(\Omega, \mathcal{A}, P)$, then the (W,φ)-*variance* of $\mathcal{X}$ (see [12, 10]) is defined as

$$\mathrm{Var}(\mathcal{X}) = \mathbb{E}([D_W^\varphi(\mathcal{X}, \mathbb{E}(\mathcal{X}))]^2). \quad (1)$$

Remark: The main reasons for using the above mentioned D_W^φ-metric are that (1) it has more good intuitive properties (see [1]) than the usual L_p metrics d_p (see [4]) do not have, and (2) it can be expressed as an inner product on the Hilbert space of all square integrable functions on $\mathcal{H} = L^2([0,1]\times\{-1,1\})$ (via using the concept of support function).

3 Confidence regions for the mean of an FRV

Let $(\Omega, \mathcal{A}, P)$ be an arbitrary probability space, and consider an FRV $\mathcal{X} : \Omega \to \mathcal{F}_c(\mathbb{R})$ verifying $\max\{|\min \mathcal{X}_0|, |\max \mathcal{X}_0|\} \in L^2(\Omega, \mathcal{A}, P)$. As mentioned in the introduction, our aim is to calculate a confidence ball for the expectation $\mu = \mathbb{E}(\mathcal{X}) \in \mathcal{F}_c(\mathbb{R})$ on the basis of n independent observations $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$ of $\mathcal{X}$.
We will denote by $\overline{\mathcal{X}}$ and $\widehat{S}$ the associated sample fuzzy mean and sample (W,φ)-deviation respectively, that is,

$$\overline{\mathcal{X}} = \frac{1}{n}\sum_{i=1}^n \overline{\mathcal{X}}_i, \quad \widehat{S} = \sqrt{\frac{1}{n}\sum_{i=1}^n [D_W^\varphi(\mathcal{X}_i, \overline{\mathcal{X}})]^2}. \quad (2)$$

Following the approach mentioned in the introduction, for a given significance level $\beta \in (0,1)$ we define the confidence ball CR_β with respect to the D_W^φ-metric as

$$CR_\beta = B(\overline{\mathcal{X}}, \delta) = \{U \in \mathcal{F}_c(\mathbb{R}) | D_W^\varphi(U, \overline{\mathcal{X}}) \le \delta\}, \quad (3)$$

whereby the radius δ has to verify the coverage condition

$$P(\mu \in B(\overline{\mathcal{X}}, \delta)) = P(D_W^\varphi(\mu, \overline{\mathcal{X}}) \le \delta) = 1 - \beta. \quad (4)$$

Due to the non-existence of sufficiently general parametric models for FRVs in general it is not possible to find δ fulfilling the coverage condition. Nevertheless we may choose δ as the $(1-\beta)$-quantile of the distribution of $D_W^\varphi(\mu, \overline{\mathcal{X}})$, which in turn can be approximated by the corresponding bootstrap $(1-\beta)$-quantile. In fact, given a sample $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$ of $\mathcal{X}$ we propose to proceed as follows:

Algorithm 3.1

Step 1. *Fix the significance level $\beta \in (0,1)$ and the number B of bootstrap replications.*

Step 2. *Obtain B bootstrap samples $(\mathcal{X}_1^{b\star}, \ldots, \mathcal{X}_n^{b\star})$ $(b = 1, \ldots, B)$ from $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$, and for each one compute its corresponding sample mean $\overline{\mathcal{X}}^{b\star}$.*

Step 3. *Compute the distance between the sample mean and each bootstrap sample mean, $d_b^\star = D_W^\varphi(\overline{\mathcal{X}}, \overline{\mathcal{X}}^{b\star})$, for each $b = 1, \ldots, B$.*

Step 4. *Choose δ as one of the $(1-\beta)$-quantiles of the sample $(d_1^\star, \ldots, d_B^\star)$ (that is, choose δ so that at least the $100(1-\beta)\%$ of the computed distances are smaller or equal than δ and at least the $100\beta\%$ of the computed distances are greater or equal than δ).*

In order to analyze the quality and accuracy of the confidence regions obtained by means of this algorithm we have made simulations with different sample sizes n. We have chosen φ as the Lebesgue measure on $[0,1]$ and used two different weight distributions W (in fact $\theta_W = 1/3$, which corresponds to the Lebesgue measure, and $\theta_W = 9/10$ that gives more weight to the spreads, were used). Each simulation corresponds to 10000 iterations with 1000 bootstrap replications and a significance level β of 0.05.
For the simulation of general FRVs the approach introduced in [8] has been applied and $n_0 = 101$ equally spaced alpha-levels $(\alpha_i = (i-1)/(n_0-1),\ i = 1, \ldots, n_0)$ were considered. In short, this approach consists essentially of the following three steps (for more details see [8]):

ISBN: 978-989-95079-6-8

SIM1 **Decomposition:**
Fix the (future) expectation $V \in \mathcal{F}_c(\mathbb{R})$, an index $n_0 \in \mathbb{N}$ and α-levels $0 = \alpha_1 < \alpha_2 < \cdots < \alpha_{n_0} = 1$. Let V^c be the mid-point of the 1-level of V, $V^l := V^c - \min(V_0)$ the total left spread and $V^r := \max(V_0) - V^c$ the total right spread. Furthermore define

$$L_\alpha = \begin{cases} \{0\} & \text{if } V^l = 0, \\ [(\min(V_\alpha) - V^c)/V^l, 0] & \text{if } V^l \neq 0 \end{cases}$$

$$R_\alpha = \begin{cases} \{0\} & \text{if } V^r = 0, \\ [0, (\max(V_\alpha) - V^c)/V^r] & \text{if } V^r \neq 0 \end{cases}$$

for every $\alpha \in [0,1]$ and set $F^l(t) := \max(-L_{1-t})$ and $F^r(t) := \max(R_{1-t})$ for every $t \in [0,1]$ (in fact F^l and F^r are distribution function but we do not use their stochastic interpretation).

SIM 2 **Discretization:**
Define $p_1^l = F^l(\alpha_1)$ and $p_i^l = F^l(\alpha_i) - F^l(\alpha_{i-1})$ for all $i = 2, \ldots, n_0$, and $p_1^r = F^r(\alpha_1)$ and $p_i^r = F^r(\alpha_i) - F^r(\alpha_{i-1})$ for all $i = 2, \ldots, n_0$ respectively. Let $B_x \in \mathcal{F}_c(\mathbb{R})$ denote the fuzzy set fulfilling $(B_x)_\alpha = [0, \mathbf{1}_{[\alpha,1]}(x)]$ for every $\alpha \in [0,1]$ and set

$$V_{n_0} = V^c + \sum_{i=1}^{n_0} -B_{1-\alpha_i} V^l p_i^l + \sum_{i=1}^{n_0} B_{1-\alpha_i} V^r p_i^r$$

SIM 3 **Stochastic Perturbation**
Consider a $(2n_0+1)$-dimensional random vector

$$\mathcal{Y} = (C_0, C_1^l, \ldots, C_{n_0}^l, C_1^r, \ldots, C_{n_0}^r) : \Omega \to \mathbb{R} \times [0,\infty)^{2n_0}$$

of coefficients for the 'approximating' simulated FRV $\mathcal{X}^\star$ as random perturbation of $(V^c, V^l p_1^l, \ldots, V^l p_{n_0}^l, V^r p_1^r, \ldots, V^r p_{n_0}^r)$ in such a way that

$$\mathbb{E}(C_0, C_1^l, \ldots, C_{n_0}^l, C_1^r, \ldots, C_{n_0}^r) = \\ = (V^c, V^l p_1^l, \ldots, V^l p_{n_0}^l, V^r p_1^r, \ldots, V^r p_{n_0}^r)$$

holds and set

$$\mathcal{X}^\star := C^0 + \sum_{i=1}^{n_0} -B_{1-\alpha_i} C_i^l + \sum_{i=1}^{n_0} B_{1-\alpha_i} C_i^r \qquad (5)$$

This procedure generates a sample of size 1 of the FRV $\mathcal{X}$ - for a sample of size n SIM 3 simply has to be repeated n times.

Remark: To be precise, in the simulations we have made we have used perturbations from the same distributions, i.e. $C_1^l, \ldots, C_{n_0}^l$ are independent and distributed as a fixed D^l, and $C_1^r, \ldots, C_{n_0}^r$ are independent and distributed as a fixed D^r (these distributions are also mentioned in the tables below).

In particular we made simulations for three FRVs $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ with different expectations - the above-mentioned parameters defining these FRVs are listed in Table 1. Moreover for the FRV $\mathcal{Z}$ (1) the mean, (2) examples of the simulated samples of $\mathcal{Z}$, (3) the distance between the mean V and one sample mean

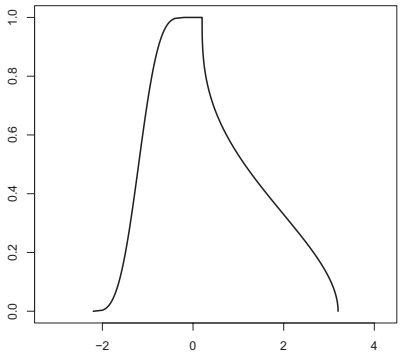


Figure 1: Expectation V of the FRV $\mathcal{Z}$

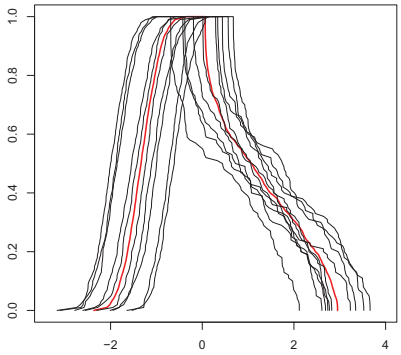


Figure 2: Simulated sample and sample mean (in red)

Table 1: Parameters defining the FRVs for the simulation

	$V = [V^c, V^l, V^r, F^l, F^r]$	$C = [C^0, D^l, D^r]$
$\mathcal{X}$	[0,1,1,Beta(1,1),Beta(1,1)]	$[U_{(-\frac{1}{2},\frac{1}{2})}, U_{(0,2)}, U_{(0,2)}]$
$\mathcal{Y}$	[4,2,2,Beta(0.5,1),Beta(3,2)]	$[U(0,8), \chi_1, \chi_2/2]$
$\mathcal{Z}$	[0,2.2,3.2,MixL,MixR)]	$[U(-1,1), U(0,2), Ex(1)]$

and (4) the empirical distribution function of the quantity $d_b^\star$ are depicted in Figure 1 - Figure 4.

Thereby MixL denotes a mixture distribution consisting of a point mass in 0 with weight 0.083 and a Beta(0.3,0.3)-distribution with weight 0.9167, and MixR denotes a mixture distribution consisting of a point mass in 0 with weight 0.067 and a Beta(3,2)-distribution with weight 0.933.

In Table 2 and Table 3 the percentage of confidence regions (among the 10,000 simulated ones) containing the populational mean in each situation is collected.

Table 2: Percentage of accurate confidence regions using Algorithm 3.1 - Part 1

	$n = 10$			$n = 30$		
θ_W	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$
1/3	89.86	90.03	90.15	93.58	93.25	93.80
9/10	89.89	89.69	0.00	93.07	93.64	93.64

ISBN: 978-989-95079-6-8

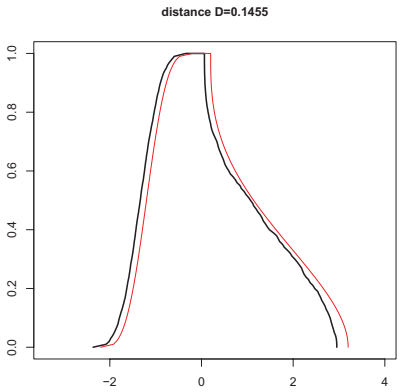


Figure 3: D_W^{φ}-distance of mean and sample mean ($\theta_W = 1/3$)

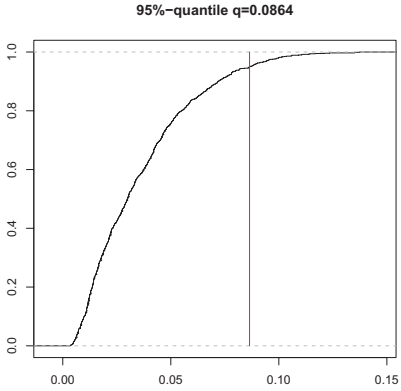


Figure 4: Empirical distribution function of $d_b^{\star}$ and 95%-quantile

Table 3: Percentage of accurate confidence regions using Algorithm 3.1 - Part 2

	$n = 100$		
θ_W	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$
$1/3$	94.51	94.83	0.00
$9/10$	94.54	94.65	0.00

Taking a look at the results in Table 2 and Table 3 it seems that (irrespective of the underlying distribution and the selected value for the parameter θ_W), the accuracy of the confidence regions generated by means of Algorithm 3.1 is sufficient for either moderate or large sample sizes (30-100 observations), whereas the results are not satisfactory for small sample sizes (10 observations). As usual, the bigger the sample size, the closer the empirical accuracy to the theoretical one.

In order to improve the results, especially for small sample sizes, we are going to modify the proposed method. The modification we make is the usual one in classical statistics - we define the radius of the confidence ball to be directly proportional to the (estimated) population variability (standardization). For this purpose, define a new confidence ball w.r.t. the D_W^{φ}-metric as follows

$$CR_\beta = B(\overline{\mathcal{X}}, \widehat{S}\delta) = \{U \in \mathcal{F}_c(\mathbb{R}) | D_W^{\varphi}(U, \overline{\mathcal{X}})/\widehat{S} \leq \delta\}, \quad (6)$$

whereby the radius δ has to verify the coverage condition

$$P(\mu \in B(\overline{\mathcal{X}}, \widehat{S}\delta)) = P(D_W^{\varphi}(\mu, \overline{\mathcal{X}})/\widehat{S} \leq \delta) = 1 - \beta. \quad (7)$$

For the same reasons mentioned in the previous case we again use bootstrapping. In this case the variability of the population, from which the bootstrap sample is taken, is completely known (denoted by $\widehat{S}$). Consequently, one possibility consists in approximating the distribution of $D_W^{\varphi}(\mu, \overline{\mathcal{X}})/\widehat{S}$ by means of $D_W^{\varphi}(\overline{\mathcal{X}}, \overline{\mathcal{X}^{\star}})/\widehat{S}$, which, however, leads exactly to the procedure proposed in Algorithm 3.1. Another possibility is to reestimate the variability of the bootstrap population instead of using the exact known value $\widehat{S}$. We will see that this alternative approach leads to better results.
In fact, the proposed procedure works as follows:

Algorithm 3.2

Step 1. *Fix the significance level $\beta \in (0, 1)$ and the number B of bootstrap replications.*

Step 2. *Obtain B bootstrap samples $(\mathcal{X}_1^{b\star}, \ldots, \mathcal{X}_n^{b\star})$ $(b = 1, \ldots, B)$ from $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$, and for each one compute its corresponding sample mean $\overline{\mathcal{X}}^{b\star}$ and its sample deviation $\widehat{S}^{b\star}$.*

Step 3. *Compute the distance between the sample mean and each bootstrap sample mean, and calculate $d_b^{\star} = D_W^{\varphi}(\overline{\mathcal{X}}, \overline{\mathcal{X}}^{b\star})/\widehat{S}^{b\star}$ for each $b = 1, \ldots, B$.*

Step 4. *Choose δ as one of the $(1 - \beta)$ quantiles of the sample $(d_1^{\star}, \ldots, d_B^{\star})$.*

In Table 4 the percentage of accurate confidence regions using Algorithm 3.2 in the same situations as in the preceding simulating study is collected. As it was expected the accuracy of the confidence regions obtained by using the second approach is much better than the one of the first approach. Indeed, the procedure can be applied also for small sample sizes ($n = 10$) and seems to be a little bit conservative (with an empirical confidence level greater than the theoretical one).

Table 4: Percentage of accurate confidence regions using Algorithm 3.2 - Part 1

	$n = 10$			$n = 30$		
θ_W	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$
$1/3$	96.31	95.18	95.70	95.38	95.17	95.42
$9/10$	96.17	95.77	95.57	94.89	95.26	95.27

Table 5: Percentage of accurate confidence regions using Algorithm 3.2 - Part 2

	$n = 100$		
θ_W	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{Z}$
$1/3$	94.93	95.06	0.00
$9/10$	94.94	94.92	0.00

ISBN: 978-989-95079-6-8

4 Case Study

In this section we will apply the proposed methodology in order to obtain a confidence ball for the mean overall rating of the II Summer Course organized by the European Centre for Soft Computing in July 2008. A survey regarding different aspects of this summer course was presented to the students at the end of the course. Among other questions, the students where asked about the overall rating of the course. The answers where collected directly as fuzzy sets. In fact, a scale ranging from 0 (minimum rating) to 100 (maximum rating) was presented (see Figure 5). The students were asked to draw a trapezoidal fuzzy set R representing their rating by fixing two intervals, the 0-level and the 1-level. Firstly the 0-level was chosen in such a way that their perception about the overall rating was surely not outside of this interval. Finally the 1-level was chosen as an interval in which they thought their overall rating would be contained.

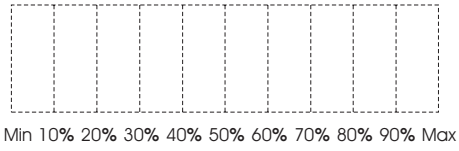


Figure 5: Scale for the made questionary

The overall rating of 33 students were collected, edited in Table 6 and the sample mean was calculated.

Applying Algorithm 3.1 and Algorithm 3.2 with $B = 10000$ bootstrap iterations and a significance level $\beta = 0.05$ we obtained the following results for the confidence balls listed in Table 7. The centre of the balls in each case is the sample mean $\overline{\mathcal{X}}$ (see Figure 6).

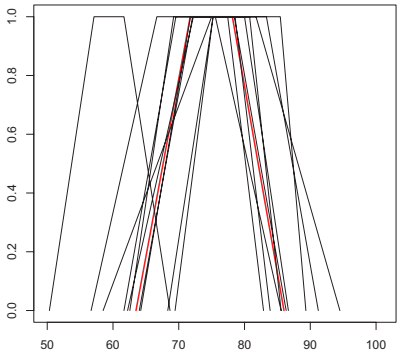


Figure 6: Thn elements of the fuzzy sample and the sample mean (red)

Figures 7 and Figures 8 depict the empirical distribution functions for the quantities $d_b^\star$ in Algorithm 3.1 and Algorithm 3.2 as well as the corresponding 95%-quantiles (for the case $\theta_W = 1/3$).

Remark: Having calculated the 95%-confidence balls for the mean overall rating, one question that naturally arises is, how this confidence balls really look like. Going back to the definition of the D_W^φ-metric it is clear that $U \in B(V, \delta)$ (ball with respect to D_W^φ), for instance, does not imply any restrictions on the support of U - in fact, the support can be arbitrarily big as long as (loosely speaking) not too many[1] α-levels are arbitrarily big. The reason for this behaviour lies in the fact that the D_W^φ-metric (because of the integration involved) averages over the distances of the α-levels (the same holds for most metrics on $\mathcal{F}_c(\mathbb{R})$). Nevertheless the condition $U \in B(V, \delta)$ (ball with respect to D_W^φ) surely implies strong restrictions on U - if for instance the spread of V is small for all α, then the spread of U can not be too big for many α's. As an example, the trapezoidal fuzzy number $(73, 73, 78, 78)$ (representation as in Table 6) is not contained in the calculated confidence Ball $B(\overline{\mathcal{X}}, 2.8239)$ since the mean spread is

Table 6: Overall ratings of the II Summer Course.

Student	$\min R_0$	$\min R_1$	$\max R_1$	$\max R_0$
1	58,50	75,00	81,75	94,50
2	64,22	71,78	78,44	86,33
3	69,44	75,22	80,78	85,56
4	64,00	72,11	85,44	89,33
5	61,67	69,56	83,33	91,22
6	56,67	66,67	75,56	85,56
7	62,22	72,22	80,00	86,67
8	50,33	57,11	61,67	68,67
9	68,33	75,22	77,44	82,89
10	62,56	69,22	78,56	83,89
11	82,22	87,33	88,11	93,44
12	50,33	60,00	64,44	74,44
13	45,56	56,67	70,00	81,11
14	69,13	75,00	89,00	94,00
15	58,89	68,89	73,11	83,11
16	51,11	62,22	63,33	77,78
17	65,00	72,50	78,50	86,25
18	51,89	61,22	68,22	86,56
19	45,89	51,67	59,44	66,33
20	74,11	80,56	88,78	95,11
21	67,89	76,22	78,33	85,56
22	74,17	85,00	88,33	97,33
23	72,22	75,00	80,44	85,33
24	51,11	60,00	71,11	78,89
25	55,56	65,56	68,89	77,22
26	76,11	80,89	90,11	93,44
27	69,89	75,11	78,00	83,22
28	80,00	90,00	90,00	100
29	71,29	79,14	82,86	89,57
30	63,78	70,89	79,44	86,67
31	63,33	74,44	85,56	95,56
32	60,00	72,11	72,11	79,89
33	78,00	83,00	87,80	92,60
Mean	63,50	71,74	78,15	86,00

Table 7: Results of the case study - calculated radius

		Radius	
θ_W	$\hat{S}$	Algorithm 3.1	Algorithm 3.2
$1/3$	8.41	2.8239	0.3587
$9/10$	8.61	2.8224	0.3476

[1] "many" in this context has to be understood as "for a set of big Lebesgue measure"

ISBN: 978-989-95079-6-8

too small.

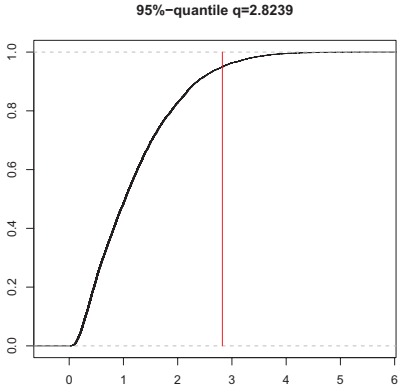


Figure 7: Empirical distribution function of $d_b^\star$ and 95%-quantile (Algoritm 3.1)

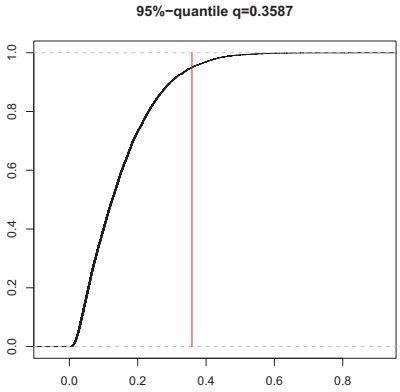


Figure 8: Empirical distribution function of $d_b^\star$ and 95%-quantile (Algoritm 3.1)

5 Conclusions and open points

In this paper we have proposed two algorithms for calculating confidence balls for the mean of an FRV - we used the sample mean as centre of the ball and calculated the radius of the ball via bootstrapping. Since we already have implemented the developed procedures (and others) in C and R our aim is to write a general R-package for statistics with fuzzy data, which could then be freely used by everybody working with R.

The simulations indicate that the second method is quite accurate, in the sense of achieving the nominal significance level for moderate and even small samples. Nevertheless, it is essential to analyze theoretically this approach in order to soundly justify its usefulness.

We have focused on a kind of percentile bootstrapping for the estimations, however it can be interesting to consider other kinds of bootstrap approaches.

Finally, for a future work we propose to analyze a possible modification of the metrics to make the estimation as reasonable as possible, by avoiding to include in the confidence region fuzzy sets which are close to the sample mean, but are not too realistic.

Acknowledgment

The research in this communication has been partially supported by the Grant from the Spanish Ministry of Education and Science MTM2006-07501. This financial support is gratefully acknowledged.

References

[1] Bertoluzza, C., Corral, N., Salas A., On a new class of distances between fuzzy numbers, Mathware & Soft Computing 2, 71–84 (1995)

[2] Colubi, A., Domínguez-Menchero, J.S., López-Díaz M., Ralescu, D. A., On the formalization of fuzzy random variables, Inform. Sci. 133, 3–6 (2001)

[3] Colubi, A., Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data, Fuzzy Sets and Systems, 160 (3), 344-356 (2009) (doi:10.1016/j.fss.2007.12.019)

[4] Diamond, P., Kloeden, P.: Metric spaces of fuzzy sets, World Scientific, Singapore, 1994

[5] Gil, M.A., Montenegro, M., González-Rodríguez, G., Colubi, A., Casals, M.R., Bootstrap approach to the classic Oneway ANOVA test with imprecise data, Comp. Stat. Data Anal, (submitted).

[6] G. González-Rodríguez, M. Montenegro, A. Colubi, M.A. Gil, Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data, Fuzzy Sets and Systems 157, 2608-2613 (2006)

[7] González-Rodríguez, G., Blanco, A., Colubi, A., Lubiano, M.A., Estimation of a simple linear regression model for fuzzy random variables, Fuzzy Sets and Systems 160 (3), 357–370 (doi:10.1016/j.fss.2008.07.007)

[8] González-Rodríguez, G., Colubi, A., Trutschnig, W., Simulation of fuzzy random variables. Information Sciences. 179 (5), 642–653 (2009) (doi:10.1016/j.ins.2008.10.018)

[9] Körner, R., An asymptotic α-test for the expectation of random fuzzy variables, J. Stat. Plann. Inference 83, 331–346 (2000)

[10] Körner, R., Näther, W., On the variance of random fuzzy variables, In: Bertoluzza, C., Gil, M.A., Ralescu, D.A. (Eds.) Statistical Modeling, Analysis and Management of Fuzzy Data, Physica-Verlag, Heidelberg, pp. 22–39 (2002).

[11] H. Kwakernaak, Fuzzy random variables, definition and theorems, Inform. Sci. 15 (1989) 1–29.

[12] M. A. Lubiano, M. A. Gil, M. López-Díaz, M. López-García, The $\overrightarrow{\lambda}$-mean squared dispersion associated with a fuzzy random variable, Fuzzy Sets and Systems 111 (2000) 307–317.

[13] Montenegro, M., Casals, M. R., Lubiano, M. A., Gil, M. A., Two-sample hypothesis tests of means of a fuzzy random variable, Inform. Sci 133, 89–100 (2001).

[14] Montenegro, M., Colubi, A., Casals, M. R., Gil, M. A., Asymptotic and Bootstrap techniques for testing the expected value of a fuzzy random variable, Metrika 59, 31–49 (2004)

[15] Puri, M. L., Ralescu, D. A., Fuzzy random variables, J. Math. Anal. Appl. 114, 409–422 (1983)

[16] Ramos-Guajardo, A.B., Colubi, A., González-Rodríguez, G., Gil, M.A., One sample tests for a generalized Fréchet variance of a fuzzy random variable, Metrika, Article in Press (doi:10.1007/s00184-008-0225-0)

[17] L.A. Zadeh, L.A., The concept of a linguistic variable and its application to approximate reasoning, Part 1, Inform. Sci. 8 (1975) 199–249; Part 2, Inform. Sci. 8 (1975) 301–353; Part 3, Inform. Sci. 9 (1975) 43–80.

ISBN: 978-989-95079-6-8

A Fuzzy Multi-Criteria Approach to WEEE Treatment Strategy Selection*

İlke Bereketli[1] Tuncay Gürbüz[2] Müjde Erol Genevois[3]

1, 2, 3 Department of Industrial Engineering, Galatasaray University
İstanbul, Turkey
Email: ibereketli@gsu.edu.tr, tgurbuz@gsu.edu.tr, merol@gsu.edu.tr

Abstract— *Today's one of the biggest environmental problem is the alarming increase of the wastes electrical and electronic equipments (WEEE). This type of waste threats the environment and the human health by contaminating air, soil and water, because of toxic materials including in the electrical and electronic equipments (EEE). That's why we should look for exploring the environmental friendly ways to dispose of these wastes and the producers should select an appropriate strategy. Waste treatment strategies contribute also to either local or global economies by creating a new sector and employment, and by reducing use of scarce resources. In this paper, we use a fuzzy hierarchical TOPSIS method for solving multi attribute group decision making (MAGDM) problems with preference information on alternatives in fuzzy environment. Our aim is to develop a fuzzy hierarchical TOPSIS model to evaluate and to select of a waste treatment strategy for electrical and electronic equipments (EEE). In this direction, three treatment strategies alternatives and six criteria are determined. The best strategy is selected and the key criterion is found.*

Keywords— Fuzzy AHP, Fuzzy Hierarchical TOPSIS, Fuzzy TOPSIS, WEEE treatment strategies.

1 Introduction

Both technological innovation and market expansion continue to accelerate the replacement of equipment leading to a significant increase of waste electric and electronic equipment (WEEE). Hence, electrical and electronic equipments, which are a subset of technological equipments, have already begun to accumulate at the garbage dumps.

Many countries have drafted legislation to improve the reuse, recycling and other forms of recovery of such wastes so as to reduce disposal [1]. The European Commission identified the need for legislation to address the escalating problem of WEEE at the Community level, and this has taken the form of the WEEE Directive. The EU's Directive 2002/96/EC about WEEE has extended producer responsibility as its main principal and aims to increase reuse and recycling ratios [2]. This Directive obligates electronic and electrical equipment producers to take back old equipment from customers free of charge and to dispose of these wastes environmentally [3]. Turkey is also on the way to prepare a new regulation about WEEE.

In this paper, a Fuzzy multi-criteria decision making (MCDM) approach is proposed for multiple attribute group decision making (MAGDM) problems with preference information on alternatives. This approach is used to evaluate and select a waste treatment strategy for electrical and electronic equipments (EEE). In this study, Fuzzy Hierarchical Technique for Order Preference by Similarity to Ideal Solution (FH-TOPSIS) is used to reflect the decision maker's subjective preference information and to determine the weight vector of criteria, and to get the rankings of different alternatives.

In the remaining part of the paper, waste treatment strategies and the case study are mentioned. Section 2 includes some definitions, alternatives and criteria required for the evaluation of the strategies. The proposed methodology is represented in Section 3. Section 4 contains the scope of the application and the results of the study. Finally in Section 5, conclusions driven from the research are provided.

2 Selection of Waste Treatment Strategies and a Case Study

WEEE, or e-waste, can be described in general as all electrical and electronic equipments that have terminated their useful life. Widmer *et al.* defines WEEE as a generic term embracing various forms of electric and electronic equipment that have ceased to be of any value to their owners, or a waste type consisting of any broken or unwanted electrical or electronic appliance [4]. In respect of EU's Directive 2002/96/EC, WEEE's definition is as follows: WEEE means electrical or electronic equipment including all components, subassemblies and consumables which are part of the product at the time of discarding [5]. So many official and non-official foundations give its own definition, therefore there is, as yet, no standard definition.

There are generally three different ways of treating WEEE: reuse, recycling and disposal (such as incineration, landfill). The EEE producers prefer at least one of these treatment strategies in order to discard their e-wastes. In this paper, it is examined that which one should be selected considering related criteria.

A case study was already realized by Bereketli *et al.*, in a telecommunication enterprise which produces and sells cell phones in Turkey [6]. A linear programming technique for multidimensional analysis of preference (LINMAP) method for solving MAGDM problems with preference information on alternatives in fuzzy environment was proposed. The aim was to develop a fuzzy LINMAP model to evaluate and to select of a waste treatment strategy for electrical and electronic equipments (EEE). At the end of the study, The

* This work is financially supported by Galatasaray University Research Fund.

ISBN: 978-989-95079-6-8

Fuzzy LINMAP method gave the ranking of the alternatives under the problems criteria as $A1 \phi\ A3 \phi\ A2$.

The objective of this study is to evaluate the alternatives considering all criteria with a different approach, Fuzzy Hierarchical TOPSIS method, and to observe the differences between the results. As we need the independence between the problem's criteria with the proposed method, some of the criteria between which there may be dependences are grouped under one criterion.

There are three alternatives: treating WEEE by reuse and recycling methods (2R), by insourced disposal methods, and by outsourced disposal methods. In the first alternative, the enterprise reuses or recycles its own WEEE within the facility. In the second alternative, the enterprise carries its own WEEE from its own waste collection store to an area which is purchased in order to dispose of the e-wastes by landfill or incineration. In the third and last alternative, the enterprise outsources its disposal process of its own WEEE to a specialized company. This company goes to the waste collection center periodically, gets there e-waste of the telecommunication enterprise, and brings them to its own facility. In its facility, it realizes disassembly of the e-waste and dissects some valuable integrated circuits. The telecommunication enterprise takes back these circuits periodically and reuses them in its assembly line. The outsourced company environment friendly disposes of the other parts of the e-waste.

It has been determined 6 criteria, 3 of them are quantitative (Type-1: T1) and 3 of them are qualitative (Type-2: T2). The criteria based on three alternatives are as follows:

- C1 - Waste release period: It is the period in which e-waste returns to the production line by passing reuse or recycling processes, or ends its life cycle by being disposed of. Small number of periods is preferred. Its unit is period time. (T1)
- C2 - Cost: This criterion includes the first investment, stock and process costs. It is the cost for creating new lines, implementing hardware, buying new machines, and new trucks for transporting e-waste, or purchasing an area where the telecommunication enterprise disposes of its e-waste, or contracting with the outsourced enterprise, stocking e-waste transported from waste collection center, every operation for each process during the period of e-waste treatment. Its unit is thousands of Turkish Liras (TL). (T1)
- C3 - Ratio of resource conservation: It is the ratio of the returned products (e-waste) quantity over total electrical and electronic equipment produced. Its unit is percentage. (T1)
- C4 - Capacity need: It is the need of labor-hour. (T2)
- C5 - Risk of damage for the nature and the human health: It is the potential of damaging the nature and the human health for each alternative when e-wastes are treated by their own way. (T2)
- C6 - Convenience to the possessed technology: It is the criteria about the degree of sufficiency of the technology that the telecommunication enterprise have. (T2)

3 Methodology

3.1 Basic concepts of Fuzzy Analytic Hierarchy Process (FAHP)

In this paper the FAHP approach is introduced, with the use of Triangular Fuzzy Numbers for pairwise comparison scale of FAHP according to the method of Chang's [7] fuzzy extent analysis by applying correct normalization formula given later by Wang *et al.* [8].

3.1.1 The construction of fuzzy judgment

In the conventional AHP, the pairwise comparisons for each level with respect to the goal are conducted using a nine-point scale proposed by Saaty [9]. According to Zadeh [10], it is very difficult for conventional quantification to define the complex situations, so the notion of a linguistic variable, whose values are words or sentences, is necessary. To assess the relative importance of the criteria and to evaluate the alternatives respecting the problems criteria an assumed weighting set has been developed.

The decision makers can use the linguistic rating set. The triangle fuzzy conversion scale of the linguistic values in the weighting set is shown in Table 1.

Table 1: The triangular fuzzy conversion scale

Very Low (VL)	(0, 3/6,1)
Low (L)	(2/3, 7/6, 5/3)
Medium (M)	(4/3, 11/6, 7/3)
High (H)	(2, 15/6, 3)
Very High (VH)	(8/3, 19/6, 11/3)

3.1.2 Value of fuzzy synthetic extent

The steps of Chang's extent analysis can be given as in the following [7]:

- First, by fuzzy arithmetic operations, take the sum of each row of the fuzzy comparison matrix.

$$RS_i = \sum_{j=1}^{n} \widetilde{a}_{ij} = \left(\sum_{j=1}^{n} l_{ij},\ \sum_{j=1}^{n} m_{ij},\ \sum_{j=1}^{n} u_{ij} \right),\quad i = 1,\mathrm{K}\ ,n \tag{1}$$

- Using fuzzy synthetic extent analysis, the value of fuzzy synthetic extent with respect to the ith object x_i, $i = 1,2,\mathrm{K}\ ,n$ that represents the overall performance of the object across all goals can be determined by the following normalization formula given by Wang et al. [8]:

$$\widetilde{S}_i = \frac{RS_i}{\sum_{j=1}^{n} RS_j} = \left(\frac{\sum_{j=1}^{n} l_{ij}}{\sum_{j=1}^{n} l_{ij} + \sum_{\substack{k=1 \\ k \neq i}}^{n} \sum_{j=1}^{n} u_{kj}},\ \frac{\sum_{j=1}^{n} m_{ij}}{\sum_{k=1}^{n} \sum_{j=1}^{n} m_{ij}},\ \frac{\sum_{j=1}^{n} u_{ij}}{\sum_{j=1}^{n} u_{ij} + \sum_{\substack{k=1 \\ k \neq i}}^{n} \sum_{j=1}^{n} l_{kj}} \right),\quad i = 1,\mathrm{K}\ ,n \tag{2}$$

ISBN: 978-989-95079-6-8

- Compute the degree of possibility of $\widetilde{S}_i \geq \widetilde{S}_j$ by the following equation where $\widetilde{S}_i = (l_i, m_i, u_i)$ and $\widetilde{S}_j = (l_j, m_j, u_j)$:

$$V(\widetilde{S}_i \geq \widetilde{S}_j) = \begin{cases} 1, & \text{if } m_i \geq m_j, \\ \dfrac{u_i - l_j}{(u_i - m_i) + (m_j - l_j)}, & \text{if } l_j \leq u_i, \quad i, j = 1, \ldots, n, i \neq j \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

- To compare $\widetilde{S}_i = (l_i, m_i, u_i)$ and $\widetilde{S}_j = (l_j, m_j, u_j)$, we need both the values of $V(\widetilde{S}_i \geq \widetilde{S}_j)$ and $V(\widetilde{S}_j \geq \widetilde{S}_i)$. The degree of possibility for a convex fuzzy number $\widetilde{S}$ to be greater than k convex fuzzy numbers $\widetilde{S}_i$, i=1,…,k can be defined by:

$$V(M \geq M_1, \ldots, M_k) = V[(M \geq M_1) \wedge \ldots \wedge (M \geq M_k)] = \min V(M \geq M_i) \quad (4)$$
$$\text{for } i = 1, 2, \ldots, k$$

- Assume that for the alternative Ai,

$$d'(A_i) = \min V(S_i \geq S_j), \text{ for } j = 1, 2, \ldots, n; j \neq i \quad (5)$$

- Then the weight vector is given by,

$$W' = (d'(A_1), d'(A_2), \ldots, d'(A_n))^T \quad (6)$$

- Via normalization, the normalized weight vectors are,

$$W = (d(A_1), d(A_2), \ldots, d(A_n))^T \quad (7)$$

3.2 Fuzzy TOPSIS

3.2.1 Basic concepts of Fuzzy TOPSIS

TOPSIS, developed by Yoon [11] and Hwang and Yoon [12] uses the following intuitive principle: The chosen alternative has to be the one having the smallest distance to the ideal solution and the biggest distance to the negative-ideal solution. But, the alternative which has the smallest distance to the ideal solution doesn't have necessarily the biggest distance to the nadir solution and vice versa. Hence, TOPSIS considers both distances to the ideal and nadir solutions simultaneously by measuring the relative closeness, RC, of the alternatives to the alternative called the « ideal solution ».

We can convert the decision matrix into a fuzzy decision matrix and construct a weighted normalized fuzzy decision matrix once the decision makers' fuzzy ratings have been collected. According to the TOPSIS model, we define the fuzzy ideal solution and fuzzy negative-ideal solution. And then a vertex method, proposed by Chen [13] is used in this paper to calculate the distance between two triangular fuzzy ratings.

Finally, the relative closeness of each alternative is defined to determine the final ranking of the alternatives.

3.2.2 Selection Procedure with Fuzzy TOPSIS

- Construct the Decision Matrix where all the data for the alternatives according the problems criteria is collected.
- Normalize the DM: To avoid the complicated normalization formula used in classical TOPSIS, the linear scale transformation is used here to transform the various criteria scales into a comparable scale.

$$\widetilde{x}_{ij} = \begin{cases} \left(\dfrac{a_{ij} - a_j^-}{c_j^* - a_j^-}, \dfrac{b_{ij} - a_j^-}{c_j^* - a_j^-}, \dfrac{c_{ij} - a_j^-}{c_j^* - a_j^-} \right) & \text{for j of a criteria of output nature} \\ \left(\dfrac{c_j^* - c_{ij}}{c_j^* - a_j^-}, \dfrac{c_j^* - b_{ij}}{c_j^* - a_j^-}, \dfrac{c_j^* - a_{ij}}{c_j^* - a_j^-}, \right) & \text{for j of a criteria of input nature} \end{cases} \quad (8)$$

- Find the weighted DM and define the ideal and the nadir alternatives namely A^+ and A^-. Note that after this normalization, the ideal (resp. nadir) alternative will be found by taking the maximum (resp. minimum) value for each criterion.
- Define distances of A_i from A^+ and A^- as follows:

$$d_i^+ = \sum_{j=1}^{n} d(\widetilde{x}_{ij}, \widetilde{x}_j^+) \text{ where } i = 1, 2, \ldots, M \quad (9)$$

$$d_i^- = \sum_{j=1}^{n} d(\widetilde{x}_{ij}, \widetilde{x}_j^-) \text{ where } i = 1, 2, \ldots, M \quad (10)$$

with

$$d(\widetilde{x}_i, \widetilde{x}_j) = \sqrt{\frac{1}{3}\left[(l_i - l_j)^2 + (m_i - m_j)^2 + (u_i - u_j)^2\right]} \quad (11)$$

- Rank the preference order of m alternatives by their relative closeness (RC) to the ideal alternative (the smaller the value of RC for an alternative, the more that alternative is preferred) which is given for the ith alternative as:

$$RC_i = \frac{d_i^+}{(d_i^+ + d_i^-)} \quad (12)$$

3.2.3 Proposed Method

In this study, the proposed method works as follows:

- Relative priorities for the criteria and the alternatives for the subjective criteria are obtained from the pairwise comparison matrices of the FAHP method using the extension principle.
- Subjective criteria ratings for the alternatives obtained in the previous step and objective criteria values are collected in a decision matrix.
- Finally Fuzzy TOPSIS is applied in order to find the ranking between the alternatives.

ISBN: 978-989-95079-6-8

4 Application

For the application of the proposed method, we asked their opinions to three decision makers. You can see in Table 2 one of the comparison matrices.

Table 2: Comparison matrices

C6	*A1*	*A2*	*A3*
A1	*	MH, H, H	VH, VH, VH
A2		*	MH, H, MH
A3			*

According to the ratings they have given, the pairwise comparison matrices are formed in order to weigh the problem's criteria and to obtain the relative priorities of the subjective criteria.

From these comparison matrices, using Fuzzy AHP method, the following results presented in Table 3 and Table 4 are obtained:

Table 3: Criteria weights

Criteria	C1	C2	C3	C4	C5	C6
Weights	.125	.198	.20	.11	.213	.15

Table 4: Relative priorities of the alternatives for the subjective criteria

	C4	C5	C6
A1	0.49	0.047	0.273
A2	0.415	0.735	0.281
A3	0.095	0.218	0.446

For the objective criteria, the values obtained by the alternatives are provided by the experts. For C1, the values were given by crisp values. On the other hand, for C2 and C3 they were agreed on giving by using the fuzzy values presented in Table 5.

As we defuzzified the fuzzy linguistic values given by the experts on the subjective criteria using Fuzzy AHP, while we apply Fuzzy TOPSIS in order to rank our alternatives, for these criteria, we will use the relative priorities given in Table 3 as crisp values. For example, the value obtained by A1 for the criteria C4, will be noted as (.49,.49,.49).

Therefore, we constructed the final decision matrix and normalized decision matrix as follows:

Table 5: Final decision matrix

	C1	C2	C3
A1	(4,4,4)	(887.5, 1095, 1207)	(30, 50, 80)
A2	(1,1,1)	(354.5, 460, 516)	(0,0,0)
A3	(2,2,2)	(174.5, 205, 231)	(3, 6, 10)
	C4	**C5**	**C6**
A1	(.49,.49,.49)	(.087,.087,.087)	(.273,.273,.273)
A2	(.415,.415,.415)	(.689,.689,.689)	(.281,.281,.281)
A3	(.095,.095,.095)	(.224,.224,.224)	(.446,.446,.446)

Table 6: Normalized decision matrix

	C1	C2	C3
A1	(0,0,0)	(0,.11,.31)	(.38,.63,1)
A2	(1,1,1)	(.67,.72,.83)	(0,0,0)
A3	(.67,.67,.67)	(.95,.97,1)	(.04,.08,.13)
	C4	**C5**	**C6**
A1	(0,0,0)	(1,1,1)	(0,0,0)
A2	(.19,.19,.19)	(0,0,0)	(.05,.05,.05)
A3	(1,1,1)	(.75,.75,.75)	(1,1,1)

From this matrix, using the weights given in Table 3, the weighted normalized decision matrix is calculated and ideal and nadir alternatives are deducted as follows:

$$A^{+} = [(.13,.13,.13)\ \ (.19,.19,.2)\ \ (.08,.,16,.,2)\ \ (.11,.11,.11)\ \ (.21,.21,.21)\ \ (.15,.15,.15)]$$

$$A^{+} = [(0,0,0)\ \ (0,.02,.06)\ \ (0,0,0)\ \ (0,0,0)\ \ (0,0,0)\ \ (0,0,0)]$$

Using the distance calculation formula given in (3.2.2) we calculated the distances of the alternatives from ideal and nadir alternatives and finally found the RC values as shown in the following table:

Table 6: Relative closeness values

	d^{-}	d^{+}	RC
A1	0.357	0.553	0.607
A2	0.272	0.639	0.701
A3	0.689	0.221	0.243

Hence the final ranking:

$$A3 \phi\ A1 \phi\ A2$$

5 Conclusions

Nowadays, the depletion of resources and the pollution are two serious problems in the world. WEEE, or e-waste for short, is one of the most influential factors on these problems. This type of waste threats the environment and the human health by contaminating air, soil and water, because of toxic materials including in the EEE. While the hazardous substances in e-waste pollute the environment, their shorter life cycle causes the fast consumption and the continuous overproduction; hence the depletion of resources increases. Therefore, a strong interest is arising in the field improving to recovery of e-waste. To cope with these problems, governments make laws about e-waste treatment and the producers begin to take the responsibility of their e-wastes. There are generally three different ways of treating WEEE: reuse, recycling and disposal.

In this paper, a Fuzzy MCDM approach is proposed for MAGDM problems with preference information on alternatives. This approach is used to evaluate and select a waste treatment strategy for EEE.

In this direction, three treatment alternatives are determined and six criteria based on these alternatives are also stated. The Fuzzy Hierarchical TOPSIS method gives the ranking

ISBN: 978-989-95079-6-8

of the alternatives under the problems criteria as $A3 \phi A1 \phi A2$. So, the best alternative is treating e-wastes by outsourcing its disposal process of its own WEEE to a specialized company. And the worst alternative is to carry its own WEEE from its own waste collection store to an area which is purchased in order to dispose of the e-wastes by landfill or incineration.

The difference of ranking between the LINMAP method and the method used in this paper may be due to the following reasons:

- Major difference between the first and third alternatives' cost. Relatively small difference for C5 and the difference in C3 for the advantage of the first alternative are not enough to compensate this huge difference of cost as well as capacity need and the convenience to the possessed technology.
- Assumption of independence between criteria caused us to unite some of the problem's criteria and as the LINMAP model doesn't require this assumption, this situation possibly caused a different ranking of the alternatives.
- Not enough consistency in expert judgments.

References

[1] J. Cui, and E. Forssberg, Mechanical recycling of waste electric and electronic equipment: a review, *Journal of Hazardous Materials*, B99, 243–263, 2003.

[2] L. Darby, and L. Obara, Household recycling behavior and attitudes towards the disposal of small electrical and electronic equipment, *Resources, Conservation and Recycling*, Vol.No.44, 17–35, 2005.

[3] M. Schmidt, A production-theory-based framework for analyzing recycling systems in the e-waste sector, *Environmental Impact Assessment Review*, Vol. No. 25, 505-524, 2005.

[4] R. Widmer, H. Oswald-Krapf, D. Sinha-Khetriwal, M. Schnellmann, and H. Böni, Global perspectives on e-waste, *Environmental Impact Assessment Review*, 25, 436-458, 2005.

[5] Directive 2002/96/Ec of the European Parliament and of the Council of 27 January 2003 on waste electrical and electronic equipment (WEEE), *Official Journal of the European Union*, L 37/24-38, 13.2.2003.

[6] İ. Bereketli, M. Erol Genevois, Y. E. Albayrak, M. Özyol, WEEE Treatment Strategies' Evaluation Using Fuzzy LINMAP Method, *Proceedings of 8th International FLINS Conference*, 647-652, 2008.

[7] D. Y. Chang, Applications of the extent analysis method on fuzzy AHP, *European Journal of Operational Research* 95, 649–655, 1996.

[8] Y.-M. Wang, Y. Luo, Z. Hua, On the extent analysis method for fuzzy AHP and its applications, *European Journal of Operational Research* 186, 735 – 747, 2008.

[9] T. L. Saaty, *The Analytic hierarchy process*, Mc. Graw-Hill, New York, 1980.

[10] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338-353, 1965.

[11] K. Yoon, *Systems selection by multiple attribute decision making*. Ph.D. dissertation. KS: Kansas State University Press, Manhattan, 1980.

[12] CL Hwang, K. Yoon, *Multiple attribute decision making: methods and applications, a state-of-the-art survey*. Berlin: Springer-Verlag, 1981.

[13] C.-T. Chen, Extensions of the TOPSIS for group decision-making under fuzzy environment, *Fuzzy Sets And Systems*, 114, 1-9, 2000.

ISBN: 978-989-95079-6-8

IFSA-EUSFLAT 2009

HIERARCHIC APPROACH IN THE ANALYSIS OF TOMOGRAPHIC EYE IMAGE

Robert Koprowski, Zygmunt Wróbel

University of Silesia, Faculty of Computer Science and Materials Science
Institute of Computer Science, Department of Biomedical Computer Systems
ul. Będzińska 39, 41-200 Sosnowiec
e-mail: koprow@us.edu.pl, wrobel@us.edu.pl.

Abstract *- The paper presents an algorithm designed to detect layers of eye's retina using an analysis in a hierarchic approach. This type approach has been implemented and tested in images obtained by means of Copernicus OCT (Optical Coherence Tomography). The algorithm created is an original approach to detect contours, layers and their thicknesses. The approach presented is an expansion of approaches described in [1], [2] and [4] and enables identification and recognition of external limiting membranes, retina and others in a very short time. The algorithm has been implemented in Matlab and C environment.*

Keywords - tomographic eye image, algorithm, recognition

1. Introduction

Images originating from a Copernicus tomograph due to its specific nature of operation are obtained in sequences of a few, a few dozen 2D images within approx. 1s, which provide the basis for 3D reconstruction. Because of their number, the analysis of a single 2D image should proceed within a time not exceeding 10 ms, so that the time of operator's waiting for the result would not be onerous (as it is easy to calculate for the above value, for a few dozen images of resolution usually MxN = 740x820 in a sequence, this time will be shorter than 1 s).

At the stage of image preprocessing (like in [1], [2] and [3]) the input image L_{GRAY} is initially subject to filtration using a median filter of (M_h x N_h) size of mask h equal to M_h x N_h=3x3 (in final version of the software this mask may be set also at M_h x N_h=5 x 5 so as to obtain better precision of algorithm operation for a certain specified group of images). Image L_M obtained this way is subject consecutively to decomposition to an image of lower resolution and analysed in terms of layers detection.

2. Image decomposition

It is assumed that the algorithm described should give satisfactory results considering mainly the criterion of operating speed. Although methods (algorithms) described in [1], [2] and [3] feature high precision of computations, however, they are not fast enough (it is difficult to obtain the speed of single 2D image analysis on a PII 1.33 GHz processor in a time not exceeding 10 ms). Therefore a reduction of image L_M resolution by approx. a half was proposed to such value of pixels number in lines and columns, which is a power of '2', i.e.: MxN=256x512 (L_{M2}) applying further on its decompositions to image L_{D16} (where symbol 'D' means decompositions, while '16' the size of block, for which it was obtained). Each pixel of the input image after decomposition has a value equal to a median of the area (block) of 16x16 size of the input image, acc. to Figure. 1.

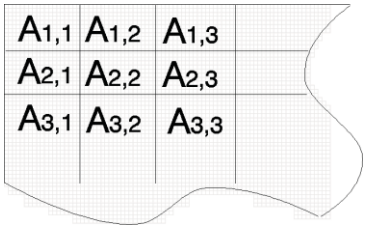


Figure. 1. Blocks arrangement in L_M image

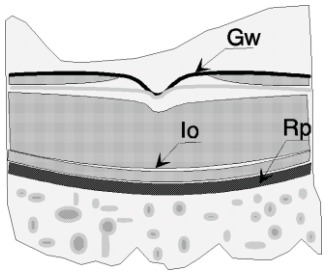


Figure. 2. Pictorial diagram of layers sought arrangement in a tomographic image

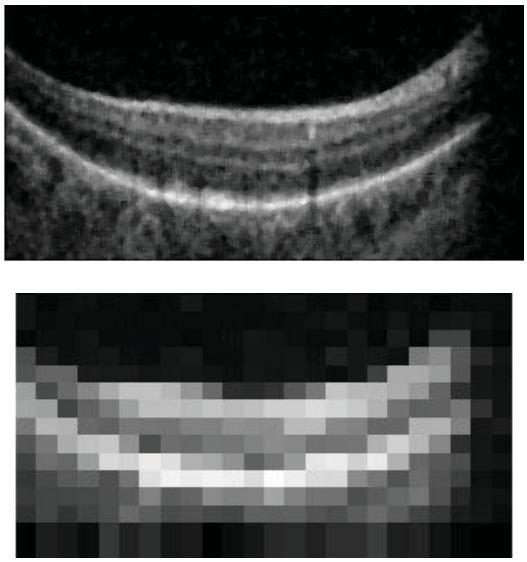


Figure. 3. Image before and after decomposition – L_{M2} and L_{D16}, respectively

ISBN: 978-989-95079-6-8

An example of L_{D16} result and the input image L_{M2} is shown in Figureure. 3. Image L_{D16} is then subject to determination of pixels position of maximum value for each column, i.e.:

$$L_{DM16}(m,n)=\begin{cases}1 & \text{if} \quad L_{D16}(m,n)=\max\limits_{m}(L_{D16}(m,n))\\ 0 & \text{others}\end{cases} \tag{1}$$

where:

m – means a line numbered from one;

n – means a column numbered from one.

Using the described method of threshold setting for the maximum value in lines, in 99 percent of cases only one maximum value in a column is obtained (Figureure. 4).

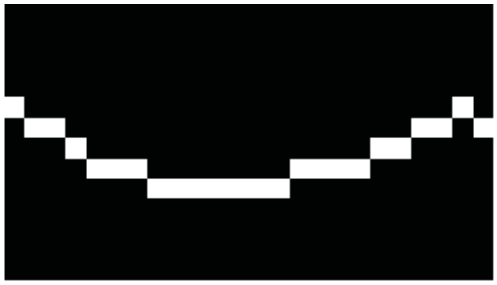


Figure. 4. Example of L_{DM16} image

Figure. 5. Example of L_{DB16} image

To determine precise position of Gw and Rp limits (Figure. 2) it turned out necessary to use one more L_{DB16} image, i.e.:

$$L_{DB16}(m,n)=\begin{cases}1 & \text{if} \quad |L_{D16}(m,n)-L_{D16}(m+1,n)|>p_r\\ 0 & \text{others}\end{cases} \tag{2}$$

for $m\in(1,M-1)$, $n\in(1,N)$,

where:

pr – threshold assumed within the range (0, 0.2).

As a result, coordinates of Gw(n) and Rp(n) limits position points are obtained as such positions of values '1' in L_{DB16} image, for which $Gw(n) \le Rp(n)$ and Rp(n) is obtained from L_{DB16} image in the same way.

This method for pr threshold selection at the level of 0.01 gives satisfactory results in around 70 percent of cases of not composed images (i.e. such, which are not images with a visible pathology). Unfortunately for the other 30 percent cases the selection of pr threshold in the adopted limits does not reduce the originated errors (Figure. 5).

The correction on this level of erroneous recognitions of Gw(n) and Rp(n) layers is that important, that for this approach these errors will not be duplicated (in the hierarchic approach presented below) for the subsequent more precise approximations.

3. Correction of erroneous recognitions

In L_{DB16} image (Figure. 5) white pixels are visible in an excess number for most columns. Two largest objects arranged along 'maxima' in columns entirely coincide with Gw and Rp limits position. Based on that and having carried out the above analysis for a few hundred images, the following limitations were adopted:

- for coordinates Rp(n) found in L_{DM16} image there must be at the same time $L_{DM16}(m,n)=1$ in other cases this point is considered as disturbance or as a point of Gw(n) layer,
- if only one pixel of value '1' occurs in image L_{DM16} and L_{DB16} for the same position, i.e. for the analysed n there is $L_{DM16}(m,n) = L_{DB16}(m,n)$ the history is analysed for $n>1$ and it is checked, whether $|Gw(n-1) - Gw(n)| > |Rp(n-1) - Rp(n)|$, i.e.:

$$Rp(n)=\begin{cases}m & \text{if} \quad L_{DB16}(m,n)=L_{DM16}(m,n)=1 \wedge \\ & \wedge |Gw(n-1)-Gw(n)|>|Rp(n-1)-Rp(n)|\\ 0 & \text{others}\end{cases} \tag{3}$$

for $m\in(1,M-1)$, $n\in(2,N)$,

- if $|Gw(n-1) - Gw(n)| \le |Rp(n-1) - Rp(n)|$, the condition $Gw(n-1) - Gw(n) = \pm 1$ is checked (giving thereby up fluctuations against history n-1 within the range ± 1 of area A (Figure. 1). If so, then this point is the next Gw(n) point. In the other cases the point is considered as a disturbance. It is assumed that lines coincide Gw(n) = Rp(n) if $Rp(n-1) - Rp(n)=\pm 1$ and only one pixel occurs of value '1' in L_{DM16} image.
- in the case of occurrence in specific column of larger number of pixels than 2, i.e. if $\operatorname{sum}_{m}(L_{DB16}(m,n))>2$ a pair is matched (if occurs) Gw(n-1), Rp(n-1) so that $|Gw(n-1) - Gw(n-1)| - |Rp(n) - Rp(n)| = \pm 1$ would occur. In this case it may happen that lines Gw(n) and Rp(n) will coincide. However, in the case of finding more than one solution, that one is adopted, for which $L_{D16}(Gw(n),n) + L_{D16}(Rp(n),n)$ assumes the maximum value (the maximum sum of weights in L_{D16} occurs).

The presented correction gives for the above class of images the effectiveness of around 99% of cases. Despite adopted limitations the method gives erroneous results for the initial value n=1, unfortunately these errors continue to be duplicated.

ISBN: 978-989-95079-6-8

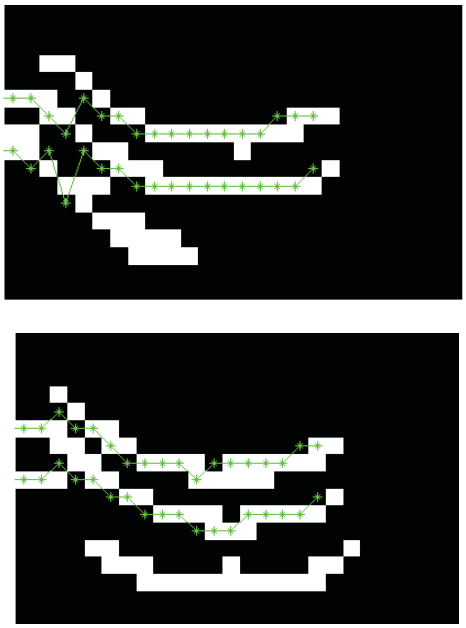


Figure. 6. Examples of L_{DB16} images for p_r=0.01 with incorrectly marked Gw(n), Rp(n) points (layers)

Unfortunately, the adopted relatively rigid conditions of acceptable difference |Gw(n-1) - Gw(n-1)| or |Rp(n)-Rp(n)| cause origination of large errors for another class of tomographic images, in which a pathology occurs in any form (Figure. 7).

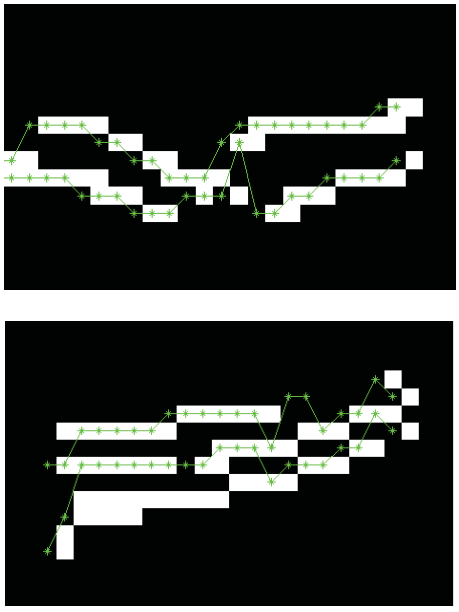


Figure. 7. Examples of L_{DB16} images for p_r=0.01 with incorrectly marked Gw(n), Rp(n) points (layers)

As it may be seen in Figure. 6 and Figure. 7 problems occur not only for the initial n values, but also for the remaining points. The reason of erroneous recognitions of layers positions consists of difficulty in distinguishing proper layers in the case of discovering three 'lines', three points in a specific column, which position changes in acceptable range for individual n.

These errors cannot be eliminated at this stage of decomposition into 16x16 pixels areas (or 16x32 image resolution). They will be the subject of further considerations in the next sections.

4. Reducing the decomposition area

The increasing of accuracy and thereby reducing the $A_{m,n}$ area size (Figure. 1) – block in L_M image – is a relatively simple stage of tomographic image processing with particular focus on the operating speed. It has been assumed that $A_{m,n}$ areas will be sequentially reducing by half in each iteration – down to 1x1 size. The reduction of $A_{m,n}$ area is equivalent to performance of the next stage of lines Gw and Rp position approximation.

The increasing of accuracy (precision) of Gw and Rp lines position determined in the previous iteration is connected with two stages:

- concentration of (m,n) coordinates in the sense of determining intermediate ((m,n) points situated exactly in the centre) values by means of linear interpolation method;
- change of concentrated points position so that they would better approximate the limits sought.

If the first part is intuitive and results only in resampling, the second requires more precise clarifications. The second stage consists in matching individual points to the layer sought. As in the ox axis the image by definition is already decomposed and pixel's brightness in the image analysed corresponds to the median value of the original image in window A (Figure. 1), the modification of points Rp and Gw position occurs only in the vertical axis. The analysis of individual Rp and Gw points is independent in the sense of dependence on n-1 point position, as was the case in the previous section.

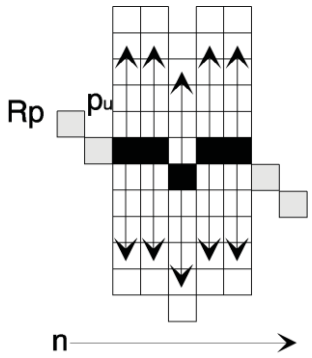


Figure. 8. Pictorial diagram of the process of Rp course matching to the edge of the layer sought. Individual pixels independent of each other may change the position within the $\pm p_u$ range

Each of Rp points, left from the previous iteration, and newly created from interpolation, in the consecutive algorithm stages is matched with increasingly high precision to the RPE layer. Point's Rp(n) position changes within the range of $\pm p_u$ (Figure. 8), where the variation range does not depend on the scale of considerations (size of A area) and strictly results from the distance between Gw and Rp (Figure. 2). For blocks A of size from 16 x 16 to 1 x 1 p_u is constant and equal 2. This value has been taken based on typical average, for analysed a few hundred L_{GRAY} images, distance between

ISBN: 978-989-95079-6-8

Gw and Rp equal to around 32 pixels, what means that after decomposition into blocks A of size 16x16 these are two pixels, i.e. p_u=2. In this ±2 range a maximum is sought in L_{DM} image and a new position of point Rp or Gw assumed for it. Thus the course of Rp or Gw is closer to the actual course of the layer analysed.

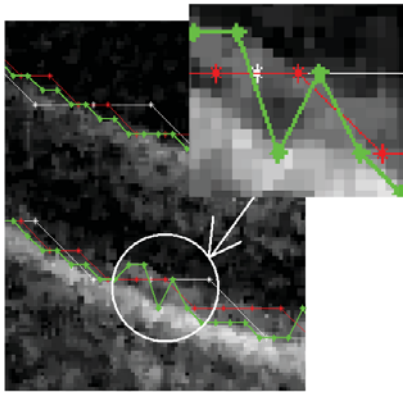


Figure. 9. Results of matching for two iterations White colour marks input Rp points and red and green – consecutive approximations

The obtained results of matching are presented in Figure. 9. White colour shows input Rp values as input data for this stage of algorithm and decomposition into A blocks of size 16 x 16 (L_{DM16} and L_{DB16} images), red colour – results of matching for A blocks of size 8 x 8 (L_{DM8} and L_{DB8} images), and green colour – results of matching for A blocks of size 4 x 4 (L_{DM4} and L_{DB4} images). As may be seen from Figure. 9 the next decompositions into consecutive smaller and smaller A areas and thus image of higher resolution, a higher precision is obtained at the cost of time (because the number of analysed Rp(n), Gw(n) points and their neighbourhoods ±p_u increases).

This method for A of 16 x 16 size has that high properties of global approach to pixels brightness that there is no need to introduce at this stage additional actions aimed at distinguishing layers situated close to each other (which have not been visible so far due to image resolution). While at A areas of 4x4 size other layers are already visible, which should be further properly analysed. At increased precision, Io layer is visible, situated close to Rp layer (Figure. 9). Thereby in the area marked with a circle there is a high position fluctuation within the oy axis of Rp layer. Because of that the next step of algorithm has been developed, taking into account separation into Rp and Io layers for appropriately high resolution.

5. Analysis of Rp and Io layers

The analysis of layers consists in separating line Io from line Rp originating from previously executed stages of the algorithm. The case is facilitated by the fact that on average around 80, 90% of pixels in each tomographic image has a maximum value in each column exactly in point Rp (this property has been already used in the first section). So the only problem is to detect the position of Io line. One of possible approaches consists of an attempt to detect the contour of the layer sought in L_{IR} image. This image originated from L_M image thanks to widening of Rp(n) layer range within oy axis within the range of ±p_I=20 pixels. L_{IR} image has been obtained with the number of columns consistent with the number of L_M image columns and with the number of lines $2 \cdot p_I+1$.

Figure. 10. Image $L_{IR} = L_M(m\text{-}R_p(n),n)$

Figure. 10 shows image $L_{IR} = L_M(m\text{-}Rp(n),n)$ originating from L_M image from Figure. 9.

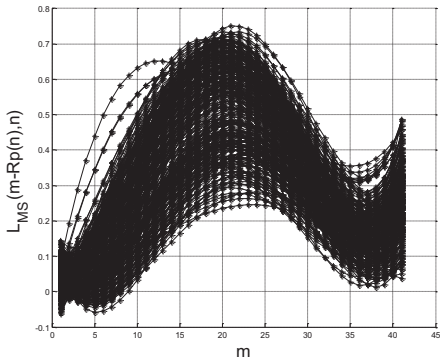


Figure. 11. Courses $L_{IRS} = L_{MS}(m\text{-}Rp(n),n)$ versus m

The upper layer visible in Figure. 10 as a pretty sharp contour is the sought course of Io. Unfortunately, because of a pretty high individual variation within the Io layer position relative to Rp the selected p_I range in further stages of the algorithm may be increased even twice (that will be described later). To determine consecutive points of Io layer position interpolations with 4th order polynomial of grey level degree for individual columns of L_{IR} image obtaining this way L_{IRS}, which changes of grey levels in individual columns are shown in Figure. 11. The position of point Io(n) occurs in the place of the highest gradient occurring within the range (Rp(n)- p_I) ÷ Rp(n) relative to L_{MS} image or 1÷ p_I relative to L_{IRS} image.

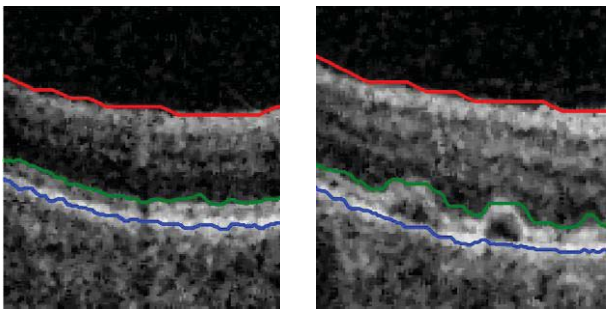


Figure. 12. Parts of L_M images with marked courses Gw – red, Rp – blue, and Io – green

As may be seen in Figure. 12, the method presented perfectly copes with detecting Gw, Rp and Io layers marked in red, blue and green, respectively.

ISBN: 978-989-95079-6-8

6. Layers thickness map and 3D reconstruction

The analysis of L_M images sequence and precisely the acquiring of Gw, Rp and Io layers allows performing 3D reconstruction and layers thickness measurement. A designation for an image sequence with an upper index (i) has been adopted, where i = {1,2,3,...,k-1,k) i.e. $L_M^{(1)}$, $L_M^{(2)}$, $L_M^{(3)}$,.., $L_M^{(k-1)}$, $L_M^{(k)}$. For a sequence of 50 images the position of Gw layers (Figure. 13), Rp (Figure. 14) and Io (Figure. 15) was measured as well as Io - Rp layer thickness (Figure. 16).

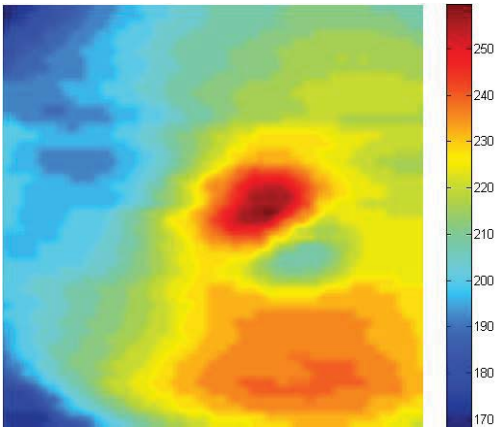


Figure. 13. Spatial position of Gw

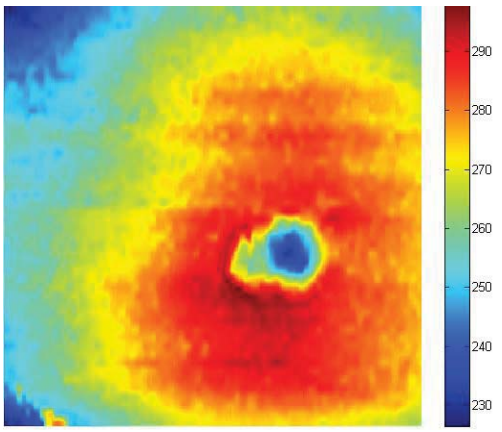


Figure. 14. Spatial position of Rp

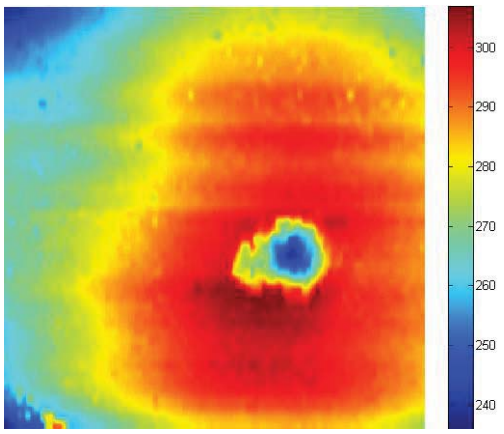


Figure. 15. Spatial position of Io

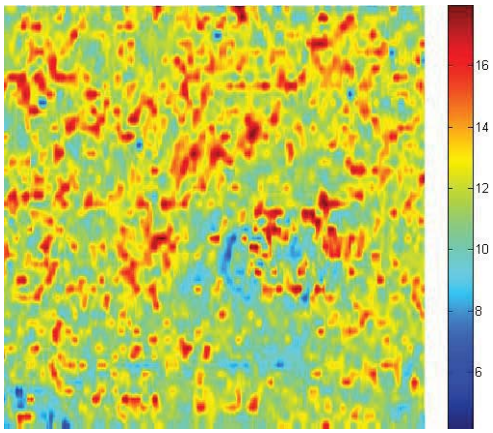


Figure. 16. Thickness of Io - Rp layer

3D reconstruction performed based on $L_M^{(i)}$ images sequence is the key element crowning the results obtained from the algorithm suggested. The sequence of images, and more precisely the sequence of $Gw^{(i)}(n)$, $Rp^{(i)}(n)$ and $Io^{(i)}(n)$ layers position provides the basis for 3D reconstruction of a tomographic image. For an example sequence of 50 images and one $L_M^{(i)}$ image resolution of MxN= 256 x 512 a 3D image is obtained, composed of three Gw, Rp and Io layers of 50 x 512 size.

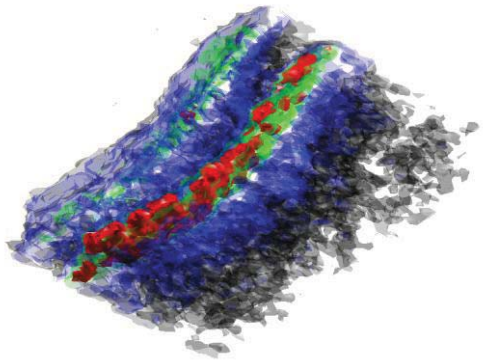


Figure. 17. Example 3D reconstruction of layers: Gw – blue, Rp – red, and Io - green

In an obvious way a possibility of automatic determination of the thickest or the thinnest places between any points results from layers presented in Figure. 18.

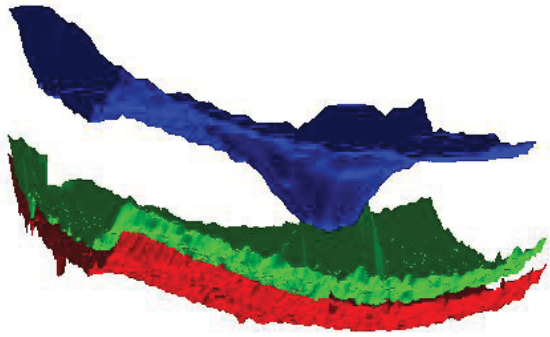


Figure. 18. Example 3D reconstruction of layers: Gw – blue, Rp – red, and Io - green

ISBN: 978-989-95079-6-8

Results are shown in Figure. 17 for an example reconstruction of original images (without processing described above) based on pixels brightness and in Figure. 18 – the reconstruction performed using the algorithm described above was carried out based on $Gw^{(i)}(n)$, $Rp^{(i)}(n)$ and $Io^{(i)}(n)$ information.

7. Summary

The algorithm presented detects Gw, Rp and Io layers within up to 50 ms time on a PC with a 2.5 GHz Intel Core 2 Quad processor. The time was measured as a mean value of 700 images analysis dividing individual images into A blocks (Figure. 1) consecutively of sizes 16 x 16, 8 x 8, 4 x 4, 2 x 2. This time may be reduced modifying the number of approximation blocks and at the same times increasing the layer position identification error – results are presented in the table below.

Table 1. Execution time of algorithm to find Gw, Rp and Io layers and 3D reconstruction

Processing stage	Total time since processing start [ms]	Time of individual stages computations [ms]
Preprocessing	10.10	10.10
Preliminary breakdown into Gw and Rp+Io	13.20	3.20
Gw and Rp+Io approximation for A – 16x16	15.90	2.74
Gw and Rp+Io approximation for A – 8x8	23.70	7.63
Precise Rp and Io breakdown	44.60	20.85

The specification of individual algorithm stages' analysis times presented in the table above clearly shows the longest execution of the first stage of image preprocessing, where filtration with a median filter is of prevailing importance (in terms of execution time) as well as of the last stage of precise determination of Rp and Io layers position. Because precise Rp and Io breakdown is related to the analysis and mainly to the correction of Rp and Io points position in all columns of the image for the most precise approximation (because of a small distance between Rp and Io it is not possible to perform this breakdown in earlier approximations). So the reduction of computation times may occur only at increasing the error of layers thickness measurement. And so for example for the analysis in the first approximation for A of 32 x 32 size and then for 16 x 16 gross errors are obtained generated in the first stage and duplicated in the next ones. For approximations for A of 16 x 16 and then 8 x 8, 4 x 4, 2 x 2 and 1 x 1 sizes the highest accuracy is obtained, however the computation time increases approximately twice.

References

1 Koprowski R., Wróbel Z.: Layer Analysis in a Tomographic Eye Image (in Polish), Proceedings Conference Decision Support Systems, Zakopane Poland (2007).

2 Koprowski R., Wróbel Z.: Identification of layers in a tomographic image of an eye based on the canny edge detection., Advances in Soft Computing 47, 2008, pp. 232-239

3 Koprowski R., Wróbel Z.: Layers Recognition in Tomographic Eye Image Based on Random Contour Analysis – sent to conference 2009 IFSA World Congress - 2009 EUSFLAT Conference

4 Koprowski R., Wróbel Z.: Determining the contour of cylindrical biological objects using the directional field. Proceedings 5th International Conference on Computer Recognition Systems CORES-07, Advances in Soft Computing, (2007)

5 Wróbel Z,. Koprowski R.: Automatic methods of microtubules orientation analysis (In Polish), Wydawnictwo UŚ, (2007).

ISBN: 978-989-95079-6-8

A Two Step Solution Procedure to a Fuzzy Medical Waste Disposal Facility Location Problem

H. Ziya Ulukan Yeşim Kop

Industrial Engineering Department, Galatasaray University

Ciragan Cad. No:36 Ortakoy, Istanbul, Turkey
Email: zulukan@gsu.edu.tr, ykop@gsu.edu.tr

Abstract— *This paper handles the medical waste disposal facility location problem in Istanbul by using fuzzy TOPSIS (FETOPSIS) to select the adequate place between some candidate points that are obtained from the Ishii's undesirable facility location algorithm. Medical waste must be disposed without damaging environment and human health, complying with new regulations. The aim is to provide an alternative method to the medical waste disposal facility location problem in Istanbul, able to handle the fuzziness of the real world. People attempt to minimize the undesirable effects introduced by the new facility by maximizing its minimum Euclidean distance with respect to all demand points and also to minimize the total transportation costs. This study proposes a modified model of Ishii's algorithm to obtain a list of candidate points. Then a selection is made between these candidate points considering other criteria as earthquake risk, distance to natural water sources and level of air pollution.*

Keywords— Undesirable Facility Location, Maximin, FETOPSIS, Medical Waste, Euclidean Distances.

1 Introduction

The infectious, pathological, cutter-driller wastes emanating from medical institutions are known as medical wastes. Recently, the controlled collection and safe disposal of medical waste has become an important field of environment protection in Turkey as in whole world. The 50% of environmental pollution has occurred in the last 35 years. As a result, the scientists work on eliminating all harmful factors on air, earth, water and human life. The safe disposal and the regain of waste are some of the important fields that they work on. Even if the medical waste consist a small part of waste amount, when we consider the threat they constitute on human life, their safe disposal is vital.

The present position is that medical wastes are stored at special hazardous waste storage areas or medical waste disposal areas if they cannot be disposed by incineration. The technical criteria concerning storage are included in the By-Law on the Control of Medical Wastes (2005).

Throwing away without control or common disposal with household waste poses a serious threat for all livings. That is why medical waste must be disposed without damaging the human mental and physical health, the animal health, the flora, the water and the welfare of the society, complying with new regulations.

The aim of this study is to provide an alternative method to solve the medical facility location problem in Istanbul, able to handle the fuzziness of transportation costs and people preferences. Nuclear plant, oil refining plant, waste disposal plant must not be close to residential area. People attempt to minimize the undesirable effects introduced by the new facility by maximizing its minimum Euclidean distance with respect to all demand points and also to minimize the total transportation costs. These are semi-obnoxious facilities and to locate them, facility planners determined two objectives [1]. The first objective aims to maximize the minimum distance from the new facility to the demand points; this is the maximin problem. The second aim is to minimize the total distance from the facility to the demand points in order to minimize the transportation costs. In this study, we propose the following two-step solution procedure where we obtain all efficient solutions at the first step and then we choose the best solution at the second step. Considering the fuzzy nature of the people attitude towards the location of this kind of facilities we will try to find the site of the facility which maximizes the minimal satisfaction degree among all demand points and maximizes the preference of the site by using a method based on the Ishii's where the attitudes of people are expressed by a trapezoidal membership function [2]. The function represents the satisfaction degree of demand points with respect to the distance from these points to the facility site. We reformulate the Ishii's model to obtain a list of candidate points. Then, we make a selection between these candidate points by solving a fuzzy TOPSIS (FETOPSIS). The use of this method allows us to consider other criteria as earthquake risk, distance to natural water sources and level of air pollution. These criteria will be evaluated by experts with the first two criteria: distance to districts and transportation costs.

2 Disposal of medical waste

There are many proven techniques for the safe disposal of medical waste all around the world. Nowadays, new contagious diseases that appeared in diverse countries induced all medical authorities to take severe precautions. In 2003, with the outbreak of severe acute respiratory syndrome (SARS) the authorities take more serious steps in managing medical waste. The procedures for handling, treatment and disposal of this waste were required to comply with the most stringent standards [3, 4]. The health authority in Taiwan had to handle the current status of waste

ISBN: 978-989-95079-6-8

production for further management planning. The quantity of waste generated by hospitals varies by changes in local legislation according to the studies [5, 6, 7, 8].

In Turkey, the medical waste production increases due to the economic and social changes and the population increase. The amount of waste is too huge to dispose in the dumping ground and requires an integrated management concept involving collection, transportation and disposal. By 2005 the number of hospital in Turkey has become 1198 and due to development of medical technologies and hygienic precautions, total waste amount has been increasing. According to MEF's researches, total waste amount in Turkey is 238.26 tons per day and 86 968 tons per year. 23 000 tons of this amount is collected in Istanbul by ISTAC (sub institution of Istanbul Metropolitan Municipality) and district municipalities.

The mortal contagious diseases that appeared recently in Turkey as Avian Influenza, as well-known name bird flu, Crimean-Congo hemorrhagic fever (CCHF) and the holocausts related to these diseases, furthermore the increase of other contagious diseases as aids, hepatitis and tuberculosis accentuate the vitality of the collection and the safe disposal of medical waste.

The regulations of medical waste control inure on 20.07.2005 in Turkey. The regulations are based on the 8th article of the environment code where it is denoted that inappropriate collection, transportation and disposal are forbidden. In the 35th article of regulations there is a legal decision on the medical waste disposal facility location: The distance of disposal facility can't be closer than 1000 meters to the residential area. As it is mentioned clearly in the regulations these following criteria must be considered in the decision making process:

1. Distance between disposal facility and districts. (C_1)
2. Transportation costs. (C_2)
3. Earthquake risk. (C_3)
4. Distance to the water sources. (C_4)
5. Existing air quality. (C_5)

The first two quantitative criteria will be handled by using a fuzzy facility location method so as to obtain worthwhile candidate points. Then these points will be handled by a fuzzy TOPSIS method considering the five criteria above all together to select a single point.

3 Methodology

3.1 Fuzzy Facility Location with Preference of Candidate Sites

Facilities as power plants, chemical plants, dumping grounds, airports are undesired close to residential area but they must also be at a reasonable distance easy to reach in order to minimize transportation costs. By using different decision making methods one must compromise two objectives. People don't want to live near a dumping ground but want also to get rid of the waste they produce as fast as possible. So the facility can not be placed to an inaccessible distance.

The first step of this study is based on the [2]'s fuzzy facility location problem, where the attitudes of the demand points (people, clients) towards the location of new facility are expressed by a trapezoidal membership function. Ishii's problem is reformulated to obtain a list of candidate points. Reference 2 categorizes people attitudes in three categories. In this study we have two categories of attitude: 1. People don't want the facility near. 2. People don't want high costs of transportation. As it is mentioned in the medical waste regulations, the burning and storage facility can not be close more than 1000 meters. Anyway people prefer the facility to be more than 5000 meters far from their residential area.

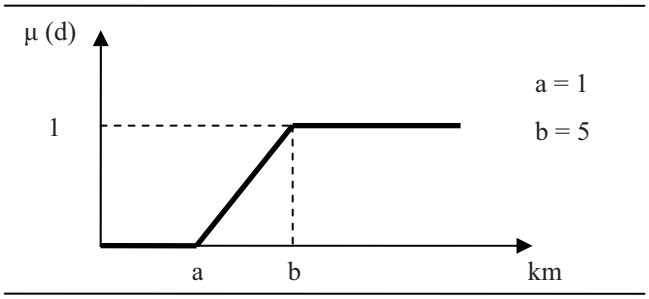


Figure 1: Membership function related to distances.

$$d_i(X) = \begin{cases} E(x-p_i)+N(y-q_i) & (x \ge p_i, y \ge q_i) \\ W(x-p_i)+N(y-q_i) & (x \le p_i, y \ge q_i) \\ E(x-p_i)+S(y-q_i) & (x \ge p_i, y \le q_i) \\ W(x-p_i)+S(y-q_i) & (x \le p_i, y \le q_i) \end{cases} \quad (1)$$

$$\mu(d_i) = \begin{cases} 0 & (d_i \le a) \\ \dfrac{d_i - a}{b-a} & (a \le d_i \le b) \\ 1 & (b \le d_i) \end{cases} \quad (2)$$

N: North = 1; *S*: South = –1; *E*: East = 1; *W*: West = –1 are constants used to provide nonnegative differences between coordinates. Let be n demand points in the rectangular area. These points are represented as (p_i, q_i), $i = 1, 2, \ldots, n$. The facility that we will build is represented as $\chi = (x, y)$ and $X = \{(x, y) | p_L \le x \le p_U,\ q_L \le y \le q_U\}$ [2]. Site $\chi \in X$.

Define α: α = $\mu(d_i)$. Then, we model the problem as [2]:

Maximize α

such that

$$\alpha \le \frac{d_i(x) - a}{b - a}$$
$$0 < \alpha \le 1$$
$$p_L \le x \le p_U$$
$$q_L \le y \le q_U \quad (3)$$

But the way that we will pursue to solve this problem is to partition the rectangular area to small rectangles by tracing the horizontal and vertical lines passing by each demand

ISBN: 978-989-95079-6-8

points. For each line intersection point called x_{jm} we will solve the sub-problem (4). The optimums of each sub problem can be handled by another decision making tool in order to select the best location for the facility.

$$\begin{aligned} Q:&\ \text{Maximize}\ \alpha \\ &\text{such that}\ \ \alpha(b-a)+a \le E(x-p_i)+N(y-q_i) \\ &\alpha(b-a)+a \le W(x-p_i)+N(y-q_i) \\ &\alpha(b-a)+a \le E(x-p_i)+S(y-q_i) \\ &\alpha(b-a)+a \le W(x-p_i)+S(y-q_i) \\ &0<\alpha\le 1 \qquad x\in X \qquad i=1,2,...,\ n \end{aligned} \tag{4}$$

3.2 Selection of the Best Candidate Site with Fuzzy TOPSIS

Method TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) was developed by Hwang and Yoon (1981) as an alternative to the ELECTRE method. The basic concept of the method is that the selected alternative should have the shortest distance to the ideal solution and the farthest distance to the negative-ideal solution. The Euclidian distance approach was proposed to evaluate the relative closeness of the alternatives to the ideal solution [4]. It solves the dilemma of the choice between ideal and anti-ideal by using an idea that Dasarathy (1976) applies to the data analysis. The TOPSIS method evaluates the decision matrix which refers to m alternatives which are evaluated in terms of n criteria [9].

In real-world situation, because of incomplete or non-obtainable information, (for example, human judgments including preferences are often vague and cannot estimate his preference with an exact numerical data), the data are not deterministic; therefore they usually are fuzzy / imprecise, so, TOPSIS for fuzzy data is used [10]. The main advantage of fuzzy formulation compared to the crisp formulation is that the decision maker is not obligated to give a precise formulation, for the sake of mathematical reasons, even though he or she might be able or willing to describe the problem in fuzzy terms [11].

The extension of TOPSIS to fuzzy TOPSIS provides a new multi-criteria decision making method compatible with the real world decisions. There are diverse applications of this method in the literature as the evaluation of airline service quality of [12], selection of expatriate host country of [13], bridge risk assessment of [14], new product introduction of [15], industrial robotic system selection of [16] etc.

Let $A_1, A_2,..., A_m$ be m alternatives among which we will make the selection, $C_1, C_2,..., C_n$ be n the criteria that are under consideration during the decision making process. $\tilde{x}_{ij}$ is the fuzzy rating of alternative A_i according to the criterion C_j. Fuzzy data used here is triangular fuzzy number. We can express this fuzzy multi criteria decision making problem in matrix format [10]:

$$\begin{array}{cccccc} & C_1 & C_2 & \cdots & C_n \\ A_1 & \tilde{x}_{11} & \tilde{x}_{12} & \cdots & \tilde{x}_{1n} \\ A_2 & \tilde{x}_{21} & \tilde{x}_{22} & \cdots & \tilde{x}_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_m & \tilde{x}_{m1} & \tilde{x}_{m2} & \ldots & \tilde{x}_{mn} \end{array} \tag{5}$$

The purpose of linear scales transformed with normalization function used in this study is to preserve the property that the ranges of normalized triangular fuzzy numbers to be included in (0,1). [17]:

$$\tilde{r}_{ij}=\left(\frac{a_{ij}}{c_j^*};\frac{b_{ij}}{c_j^*};\frac{c_{ij}}{c_j^*}\right) \tag{6}$$

Here, $c_j^*=\max_i c_{ij}$

Thereby, the normalized fuzzy decision matrix $\tilde{R}=\left[\tilde{r}_{ij}\right]_{m\times n}$ is generated. In decision making process each criterion can have a different importance for the decision maker who will assign different weights to each criterion. But in our study we assume that all criteria are equally important.

We define the positive ideal solution A^* and the negative ideal solution A^- [17]:

$$A^*=\left(\tilde{r}_1^*,\tilde{r}_2^*,\ldots,\tilde{r}_n^*\right) \tag{7}$$

$$A^-=\left(\tilde{r}_1^-,\tilde{r}_2^-,\ldots,\tilde{r}_n^-\right) \tag{8}$$

Then, we have to calculate the distance of each alternative to A^* and A^-, by using the formula of distance between two fuzzy numbers.

$$d_i^*=\sum_{j=1}^{n} d\left(\tilde{r}_{ij},\tilde{r}_j^*\right), i=1,\ldots,m \tag{9}$$

$$d_i^-=\sum_{j=1}^{n} d\left(\tilde{r}_{ij},\tilde{r}_j^-\right), i=1,\ldots,m \tag{10}$$

The distance between two fuzzy numbers is calculated as:

$$\tilde{x}=\left(a_1,b_1,c_1\right),\ \ \tilde{y}=\left(a_2,b_2,c_2\right) \tag{11}$$

$$d\left(\tilde{x},\tilde{y}\right)=\sqrt{\frac{1}{3}\left[\left(a_1-a_2\right)^2+\left(b_1-b_2\right)^2+\left(c_1-c_2\right)^2\right]} \tag{12}$$

A relative closeness index between d_i^* and d_i^- is calculated to determine the ranking order of alternatives [17,10]:

$$R_i=\frac{d_i^-}{d_i^*+d_i^-},\ \ i=1,\ldots,m \tag{13}$$

ISBN: 978-989-95079-6-8

4 Application

This two-step model is applied to the medical waste facility selection for the most populated region of Istanbul including the districts Şişli, Kağıthane, Beşiktaş and Sarıyer. The figure 2 is the Google Earth map showing the application region. The distances and coordinates are provided by Google Earth.

Figure 2: Selection of a medical waste disposal facility in Istanbul.

In rectangular area there are four demand points and all lines passing through form 25 points which means that we will solve the sub problem (4) 25 times by using LINDO software. The optimum membership values and optimum candidate facility sites are in the Table 1. As our aim is to maximize membership value α, we can select the biggest α value as new facility site but there are many α values closer to each other. So, the most reasonable attitude is to handle some candidate points having big α value, considering other criteria. In this study, we handle seven points as alternatives having α value bigger than 0.600 in terms of an observation made on the map.

Table 1: Optimum α values and optimum candidate facility sites

SUB PROBLEMS	OPTIMUM POINT	OPTIMUM α VALUE
1	-	-
2	(1.790; 3.450)	0.110
3	(6.350; 0.000)	0.140
4	(5.330; 7.110)	0.167
5	(5.080; 2.160)	0.217
6	(5.080; 3.645)	0.261
7	(5.330; 3.645)	0.324
8	(4.445; 2.160)	0.376
9	(5.080; 2.795)	0.376
10	(5.080; 2.795)	0.376
11	(3.230; 0.000)	0.410
12	(6.350; 2.160)	0.410
13	(2.675; 0.540)	0.413
14	(6.350; 7.110)	0.422
15	(2.675; 0.000)	0.549
16	(0.000; 0.565)	0.565
17	(3.230; 7.110)	0.567
18	(6.350; 3.645)	0.579
19	(1.640; 5.645)	0.621
20	(6.350; 3.045)	0.631
21	(0.000; 3.100)	0.645
22	(0.000; 0.000)	0,700
23	(1.640; 7.110)	0.965
24	(0.000; 5.520)	1.000
25	(0.080; 5.690)	1.000

Seven candidate waste disposal facilities are highlighted in the Table 1. As there are several candidate points, we need a second decision process where we will introduce new decision criteria that are mentioned in the regulations to find out the best place for a medical waste disposal facility. The five criteria mentioned in the second section are evaluated for each candidate site by experts. The fuzzy TOPSIS method of [17] was used to handle these evaluations.

For the evaluation of the first criterion, the distance between the facility and the districts, experts used the α values of seven candidate points. Bigger α values improve our model.

To figure out the transportation costs, the distances between districts and candidate points are measured and given in the Table 2. The waste amount generated by each district is in the Table 3. These amounts help us to determine the weights of each district that are used in the calculation of transportation costs.

Table 2: Distances of candidate sites to each demand point

SITE	SISLI	BESIKTAS	KAGITHANE	SARIYER
1	6.680	5.870	3.800	10.770
2	3.580	7.890	2.580	7.670
3	3.525	3.525	5.595	3.915
4	5.300	10.310	5.000	5.250
5	3.785	8.795	3.485	3.485
6	5.390	10.400	5.090	5.000
7	5.250	10.260	4.950	4.860

Table 3: Waste amounts of each district and weights (Istanbul Metropolitan Municipality).

DISTRICTS	MEDİCAL WASTE AMOUNT (KG/YEAR)	WEIGHT
SISLI	87 600	0.362
BESİKTAS	64 800	0.268
KAGITHANE	64 800	0.268
SARIYER	24 600	0.102

To evaluate the earthquake risk criteria, the experts used the earthquake map of Istanbul published by earthquake observatory of Istanbul. There are five different risk levels in the map. They evaluated each candidate point according to their location on the map.

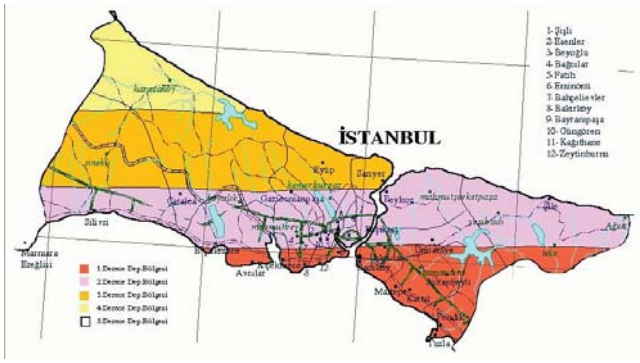


Figure 3: Earthquake Risk Map of Istanbul.

Distances of each candidate point to the water sources were measured using Google Earth. The results were evaluated by experts.

For the evaluation of air quality criterion, the experts used the data of Istanbul Technical University pollution reports.

ISBN: 978-989-95079-6-8

The figure 4 is the regional pollution map. The amount of SO_2 differs for each color in the map. It is denoted in the regulations that the waste disposal facility must be located in a region where it will not increase the pollution to high levels. So it is convenient to locate the facility in a less polluted area.

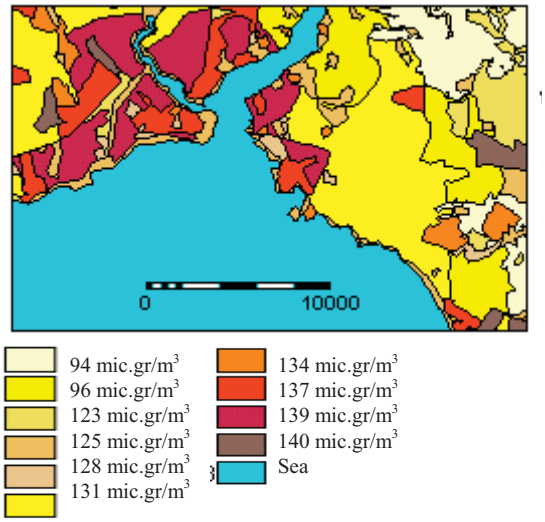


Figure 4: SO_2 amount for different places in Istanbul

The linguistic variables used by experts are in the Table 4. In our study we used a scale of [17]. The evaluation matrix and the normalized decision matrix of fuzzy TOPSIS decision process of this study are given in the Table 5 and Table 6.

Table 4: Linguistic Variables

Very poor (VP)	(0; 1; 3)
Poor (P)	(1; 3; 5)
Fair (F)	(3; 5; 7)
Good (G)	(5; 7; 9)
Very Good (VG)	(7; 9; 10)

Table 5: Expert Evaluation

	C1	C2	C3	C4	C5
A1	VP	G	F	VP	VG
A2	VP	VG	P	F	VP
A3	P	G	P	G	F
A4	F	P	VP	VG	P
A5	VG	VP	F	P	VG
A6	VG	VP	F	P	VG
A7	VG	VP	F	P	VG

Table 6: Normalized Decision Matrix

	C_1	C_2	C_3	C_4	C_5
	0.00	0.50	0.43	0.00	0.70
A_1	0.10	0.70	0.71	0.10	0.90
	0.30	0.90	1.00	0.30	1.00
	0.00	0.70	0.14	0.30	0.00
A_2	0.10	0.90	0.43	0.50	0.10
	0.30	1.00	0.71	0.70	0.30
	0.10	0.50	0.14	0.50	0.30
A_3	0.30	0.70	0.43	0.70	0.50
	0.50	0.90	0.71	0.90	0.70
	0.30	0.10	0.00	0.70	0.10
A_4	0.50	0.30	0.14	0.90	0.30
	0.70	0.50	0.43	1.00	0.50
	0.70	0.00	0.43	0.10	0.70
A_5	0.90	0.10	0.71	0.30	0.90
	1.00	0.30	1.00	0.50	1.00
	0.70	0.00	0.43	0.10	0.70
A_6	0.90	0.10	0.71	0.30	0.90
	1.00	0.30	1.00	0.50	1.00
	0.70	0.00	0.43	0.10	0.70
A_7	0.90	0.10	0.71	0.30	0.90
	1.00	0.30	1.00	0.50	1.00

The results are given in the Table 7. According to these results, three areas, A_5, A_6, and A_7 are equally adequate for the medical waste disposal facility. Selected point is sufficiently far from residential area and not too far to increase the transportation costs. So one can conclude that this study is meaningful for this kind of problems. But this model does not take into account other logistical constraints.

Table 7: Range of alternatives

	d_i^*	Range	d_i^-	Range	R_i	Range
A_1	1.64	3	1.83	3	0.53	3
A_2	2.13	5	1.35	5	0.39	5
A_3	1.57	2	1.93	2	0.55	2
A_4	2.04	4	1.45	4	0.42	4
A_5	1.30	1	2.17	1	0.62	1
A_6	1.30	1	2.17	1	0.62	1
A_7	1.30	1	2.17	1	0.62	1

5 Conclusions

This paper deals with the problem of determining the undesirable facility location site in Istanbul. We tried first to find the site of the facility which maximizes the minimal satisfaction degree among all demand points and maximizes the preference of the site by using a method based on the Ishii's.

There are four drawbacks in this kind of solution procedure. The first is that this solution procedure suggests that residential areas are discrete points. This problem can be handled by revising the intervals of membership function as in this study. Second is that more than one sub-problem can have the same optimum solution. To handle this problem we use the optimum of each sub-problem having nearly same α values, bigger than a fixed value, in a new decision making process. The inaptitude to introduce qualitative criteria to the decision process is another drawback. A fuzzy TOPSIS

ISBN: 978-989-95079-6-8

method is introduced to the decision making process as a second step to improve the model. The last drawback is not handled in this study but can be subject to a further study. When the number of districts increases, the number of sub problems increases hugely and it complicates the decision making process. For a limited number of districts, the apparent advantage of our solution procedure is that all problems are converted to linear problems and we solve them easily using LINDO software.

Acknowledgment

This research has been financially supported by Galatasaray University Research Fund.

References

[1] E. Melachrinoudis, Z. Xanthopulos, A Semi-obnoxious single location in Euclidean space, *Computers&Operations Research*, 30, 2191-2209, 2002.

[2] H. Ishii, Y.L. Lee, K.Y. Yeh, Fuzzy facility location problem with preference of candidate sites, *Fuzzy Sets and Systems*, 158, 1922-1930, 2007.

[3] C.F. Chiang, F.C. Sung, F.H. Chang, C.T. Tsai, Hospital waste generation during the SARS outbreak in Taiwan, *Infection Control and Hospital Epidemiology*, 27, 519-522,2006.

[4] Y.W. Cheng, F.C. Sung, Y. Yang, Y.H. Lo, Y.T. Chung, K.-C. Li, Medical waste production at hospitals and associated factors, *Waste Management*, 29, 440-444, 2009.

[5] M. Askarian, M. Vakili, G. Kabir, Results of a hospital waste survey in private hospitals in Fars Province, Iran, *Waste Management*, 24, 347-352, 2004.

[6] S. Fisher, Healthcare waste management in the UK: the challenges facing healthcare waste producers in light of changes in legislation and increased pressures to manage waste more efficiently, *Waste Management*, 25, 572-574, 2005.

[7] R. Mohee, Medical wastes characterization in healthcare institutions in Mauritius, *Waste Management*, 25, 575-581, 2005.

[8] A. Bdour, B. Altrabsheh, N. Hadadin, M. Al-Shareif, Assessment of medical wastes management practice: a case study of Northern Part of Jordan, *Waste Management*, 27, 746-759, 2007.

[9] J-C. Pomero, S. Barba-Romero, *Choix multicritère dans l'entreprise, principe et pratique*, Editions Hermes, Paris, (1993).

[10] G.R. Jhanshahloo, F. H. Lotfi, M. Izadikhah, Extension of the TOPSIS metod for decision-making problems with fuzzy data, *Applied Mathematics and Computation*, 181(2), 1544-1551, 2006.

[11] H.-J. Zimmermann, Fuzzy set theory and mathematical programming in: A. Jones et al., *Fuzzy Set Theory and Applications*, 99-114, Reidel, Dordrecht, 1986.

[12] S-H. Tsaur, T-Y. Chang and C-H. Yen, The evaluation of airline service quality by fuzzy MCDM, *Tourism Management*, 23(2), 107-115, 2002.

[13] M-F. Chen and G-H. Tzeng, Combining grey relation and TOPSIS concepts for selecting an expatriate host country, *Mathematical and Computer Modelling*, 40(13), 1473-1490, 2004.

[14] Y-M. Wang and T. M.S. Elhag, Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment, *Expert Systems with Applications*, 31(2), 309-319, 2006.

[15] C. Kahraman, G. Büyüközkan and N. Y. Ates, A two phase multi-attribute decision-making approach for new product introduction, *Information Sciences*, 177(7), 1567-1582, 2007.

[16] C. Kahraman, S. Çevik, N. Y. Ates and M. Gülbay, Fuzzy multi-criteria evaluation of industrial robotic systems, *Computers & Industrial Engineering*, 52(4), 414-433, 2007.

[17] T-C. Wang, T-H. Chang, Application of TOPSIS in evaluating initial training aircraft under a fuzzy environment, *Expert Systems with Applications*, 33, 870-880, 2007.

ISBN: 978-989-95079-6-8

Online Classification of Machine Operation Modes Based on Information Compression and Fuzzy Similarity Analysis

Gancho Vachkov

Faculty of Engineering, Kagawa University
Takamatsu City, Kagawa 761-0396, Japan
Email: vachkov@eng.kagawa-u.ac.jp

Abstract—*This paper proposes a computational scheme for online classification of process data sets that represent different operation modes of machines and other systems by use of fuzzy similarity analysis. The classification procedure starts with a preliminary given small number of known operation modes (data sets) which constitute the initial size of the Knowledge Base (KB). During the online classification, the dissimilarity (difference) degree between the newly submitted operation mode and each of the current modes stored in the KB is computed. As a result, depending on the preliminary given threshold for classification, the newly submitted mode is classified as belonging to a certain class (operation mode) from the KB or is considered as a new (unknown and different) mode. In the latter case, this mode is added as a new entry of the KB and used for the further online classification.*

An unsupervised learning algorithm (a modification of the Neural-Gas leaning algorithm) is used in the paper for compression of the original "raw" data from each operation mode into Compressed Information Model (CIM) with a smaller number of neurons. Then the similarity analysis is performed as a two-input fuzzy inference procedure that uses the Center-of-Gravity Distance and the Weighted Average Size Difference between two CIMs. The membership functions and the singletons of the fuzzy inference procedure are tuned in the paper by using a given set of test operation modes. The whole computational scheme is illustrated and discussed on a real example for classification of 5 main operation modes of a hydraulic excavator.

Keywords— Fuzzy Similarity, Information Compression, Online Classification, Operation Modes, Unsupervised Learning

1 Introduction

Classification of different available process data taken from the real operation of machines and other systems is a problem of utmost importance for the proper monitoring and performance evaluation of the machines. Generally speaking, the classification and pattern recognition problems, as well as various methods for their solution are very well presented and discussed in the literature [1-6]. However in the most often cases the problem is viewed as off-line classification of preliminary given set of data (patterns) with fixed size. Then the task is to classify every single data (pattern) from the given data set as belonging to one or another class.

When dealing with industrial systems and machines, the problem is sometimes different and more complicated, since one single data (i.e one measurement from a group of sensors at a given instant) is not sufficient to reveal clearly the current machine status (known as machine *operation mode*). We need a data set containing multiple measurements during a limited time of machine operation, in order to classify the current operation mode as *belonging* (or *not belonging*) to a certain operation mode.

During long time operation of the machine, *large* (or even *endless*) sequence of limited time operation data sets is obtained. Therefore a flexible *online* and *incremental classification scheme* is needed that should be able to *1)* properly classify the data sets into respective operation modes and *2)* to discover *new* operation modes that are stored in the KB in order to be used in the further online classification.

In this paper we present a special *two-stage* computation scheme for solving the problem of online classification, The *first stage* is information compression of the "raw operation data" representing a certain operation mode into a respective compressed information model (CIM), which consists of a small number of neurons. As a tool for information compression, a modified version of the off-line Neural-Gas unsupervised learning algorithm [3,4] is used in the paper. In the *second stage* a special fuzzy inference procedure for similarity analysis is proposed that uses two parameters, extracted from the CIM, namely the C*enter-of-Gravity* and the *Weighted Average Size* of the CIM. The differences between the respective parameters for a given pair of operation modes are used as two input *features* in the fuzzy inference procedure for similarity analysis.

The proposed classification scheme has the ability to perform *incremental classification*, which is very useful in the typical scenario of an unlimited sequence of operation data sets that represent different (n*ew*, unknown) *operation modes*. The flexibility and applicability of this scheme is shown in the paper on a real example for classification of five main operation modes of a hydraulic excavator.

2 The proposed unsupervised classification scheme

The Block-Diagram of the proposed scheme is shown in Fig. 1. This scheme is a further development of our previous idea for similarity analysis, discussed in [7].

The procedure starts with a small number of preliminary known *core operation modes* (core CIMs), which build the initial size of the Knowledge Base. During classification, the *difference (dissimilarity) degree* between the newly submitted (unknown) operation mode (as CIM) and all the core CIMs in the KB is computed. As a result, depending on a preliminary given *threshold Th* for classification, the new operation mode could be classified as *belonging* to a certain class (mode) from the KB or is regarded as a *quite different* (new) mode thus creating a *new class* in the KB. In such way, the proposed general concept for classification is incremental one, allowing the Knowledge Base to gradually grow, when a new operation mode with a very high dissimilarity degree to any of the core modes in the KB is discovered.

ISBN: 978-989-95079-6-8

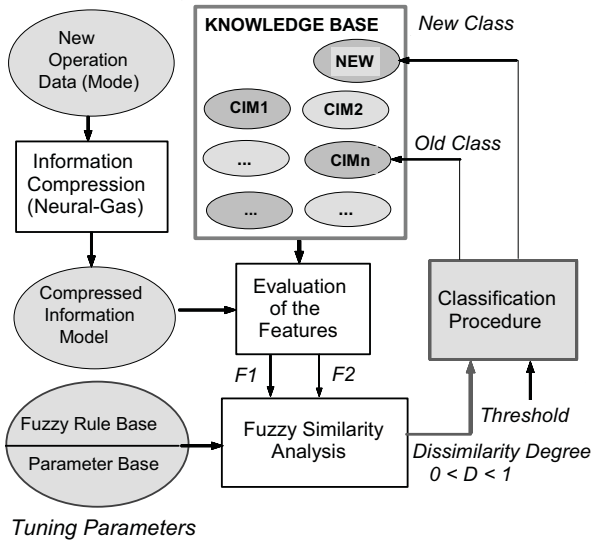


Figure 1: Block diagram of the proposed Online Classification Scheme based on fuzzy similarity analysis.

3 Unsupervised learning algorithm for information compression

The first step before the actual similarity analysis and classification of the operation modes is to find a way to reduce the large amount of the "raw data" information contained in the data set. Further on we refer to this computation step as *Information Compression*. From a computational viewpoint the information compression can be viewed as a kind of *transformation* of the original large data set: $\mathbf{x}_i = [x_{i1}, x_{i2}, ..., x_{iK}], i = 1,2,..., M$, consisting of *M* data in the *K*-dimensional input space into a respective *Neural Model* consisting of *N neurons* in the same space. Here $N << M$. and *CR = M/N* is the so called *Compression Ratio.*

The *information compression* of the original large data set (e.g. pixels or process data from sensors) can be perform by using different unsupervised competitive learning algorithms, such as clustering algorithms [1,2], Neural-Gas and its modifications [3,4,7], Support Vector Machines and Self-Organizing Maps [5,6] etc. All these algorithms try to find the most appropriate positions of the preliminary fixed number of *N* neurons (clusters) in the *K*-dimensional data space so that the resulting group of neurons resembles as much as possible the density distribution of the original data in the same space.

The essential part of any unsupervised learning algorithm is the so called *updating rule* for the neuron *centers* $\mathbf{c}_i, i = 1,2,...,N$ in the *K*-dimensional space. The algorithm updates the neuron centers iteratively with preliminary fixed number of iterations *T* (*t = 0,1,2,...,T*), as follows:

$$\mathbf{c}_i(t) = \mathbf{c}_i(t-1) + \Delta\mathbf{c}_i(t),\ i = 1,2,...,N. \qquad (1)$$

Here the computation of the update $\Delta\mathbf{c}_i(t)$ varies depending on the type of the unsupervised algorithm.

In this paper we use a modified version of the original *Neural-Gas* unsupervised learning algorithm, first presented in [3]. At every iteration the update is computed as:

$$\Delta\mathbf{c}_i(t) = R(t) H_s(t, r_i)\left[\mathbf{x}_s - \mathbf{c}_i(t-1)\right], \quad i = 1,2,...,N;\ s = 1,2,...,M \qquad (2)$$

Here $R(t),\ 0 \le R(t) \le 1,\ t = 0,1,2,...,T$ is a monotonically decreasing *Learning Rate,* which guarantees the convergence and stability of the learning process:

$$R(t) = R_0 \exp\left(-t/T_C\right),\ t = 0,1,...,T \qquad (3)$$

The so called *Neighborhood Function* in (2) $0 \le H_s(t, r_i) \le 1$ also decreases exponentially with the iterations. It computes the dynamically changing (decreasing) *activity area* for each neuron during the iterations, as follows:

$$H_s(t, r_i) = \exp\left[-(r_i - 1)/B(t)\right], \quad t = 0,1,...,T;\ s = 1,2,...,M;\ i = 1,2,...,N \qquad (4)$$

$$\text{where} \quad B(t) = \exp\left(-t/T_W\right),\ t = 0,1,...,T \qquad (5)$$

Here $r_i \in [1,2,...,N]$ is an integer number representing the so called *ranking position* of the *i*-th neuron (*i = 1,2,...,N*) to the *s*-th data point (*s = 1,2,...M*). This position is determined by the distance between the *i*-th neuron and the *s*-th data point. The closest neuron (in a sense of a minimal *Euclidean* distance) is called "winning neuron" and gets ranking position *r = 1*. The second closest neuron gets *r = 2* and so on. Example for information compression of a given data set by using the above algorithm is presented in the next Section.

4 Example of machine operation data for online classification and similarity analysis

The hydraulic excavator (HE) is a typical example of a complex machine. It is equipped with a turbo-diesel engine, which powers a special hydraulic system for performing different kinds of movements and working operations. HE normally works in a dynamical sequence of several repetitive *operation modes.* The following *five operation mode*s are typical for a normal long time operation of the excavator, when it does not move on the ground, as follows:

Mode 1. Loading the bucket with the raw material (sand, soil, stones etc);

Mode 2. Transporting the load in the bucket to the nearby truck by moving the arm (arrow) with the full bucket;

Mode 3. Unloading the bucket material into the truck;

Mode 4. Returning the arm with the empty bucket to the initial position for the next loading;

Mode 5. Short repetitive movements of the arm (*up* and *down*) for *pressing* the raw material in the truck;

All these modes differ from one to another in terms of the required engine power and the amount of the load in the bucket. Then the real practical problem is to properly *recognize* all these five operation *modes* and to *discover* possibly *new* ones (if any), by analyzing the available *online sequence* of data sets, each of them corresponding to *one operation mode*.

Such recognition and classification of the modes is important for detecting a slow trend of *deterioratio*n in the performance of the HE that needs appropriate maintenance or repair.

From an engineering viewpoint, the following six parameters are considered as important for the proper mode recognition, namely: *P1*- Engine Speed [*rpm*]; *P2* – Engine Boost Pressure; *P3* - Engine Oil Pressure; *P4* - Fuel Consumption; *P5* – Left Hydraulic Pump Pressure and *P6* – Right Hydraulic

ISBN: 978-989-95079-6-8

Pump Pressure. During the experiments, online measurements (at every second) were performed and the following Fig. 2 depicts the normalized data collected separately for each of the five typical operation modes.

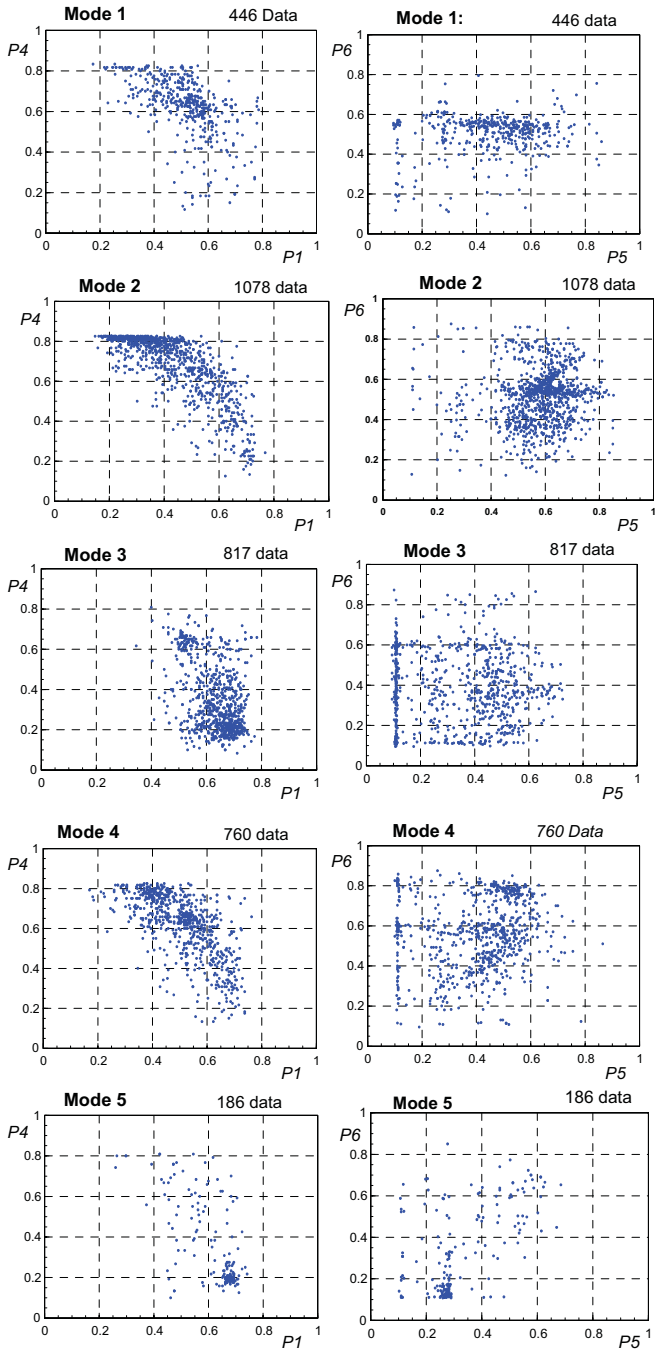


Figure 2: Plot of the normalized "raw data" representing all *5* operating modes of a hydraulic excavator.

As seen above, *two* different 2-dimensional plots are presented in Fig. 2, *P1-P4* and *P5-P6* respectively, for a better understanding of the complexity of the original *6*-dimensional problem.

Fig. 3. serves as illustration of the information compression of the "raw data" set, corresponding to operation *Mode 2*, by the Neural-Gas algorithm explained in Section 3. The following settings of the learning parameters for information compression have been used for all data sets from Fig. 2, as follows: $T = 500; R_0 = 0.16$ and $T_C = T_W = T/5$. The number of the neurons has been fixed to $N = 50$ in all simulations.

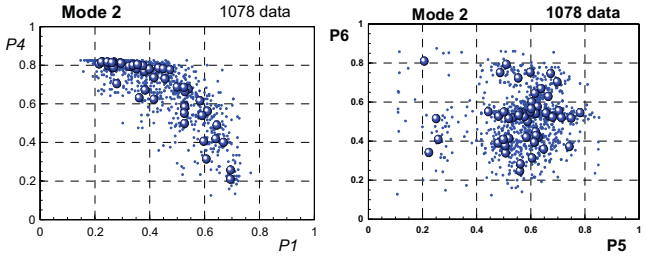


Figure 3: Results from the information compression of the "raw data" for operation *Mode 2* with $N = 50$ neurons.

5 Feature selection and computation for the fuzzy similarity analysis

As seen from Fig. 1 in Section 2, in order to evaluate the similarity between a given pair of operation modes, based on their compressed models (CIMs), we have to evaluate two important *features F1* and *F2* that characterize in an easy-to-understand numerical way the relation (similarity) between each pair of modes. For this purpose, we propose here to extract the following two distinct *parameters P1* and *P2* that characterize the *location* and the *size* of each operation mode in the *K*-dimensional input space. They are further called: *P1 - Center-of-Gravity* **CG** and *P2 - Weighted Average Size WAS* of the given operation mode.

1) The *Center-of-Gravity* $\mathbf{CG} = [CG_1, CG_2, ..., CG_K]$ of a *K*-dimensional operation mode is a *vector* that is computed directly from the respective CIM as follows:

$$CG_j = \sum_{i=1}^{N} c_{ij} g_i \Big/ \sum_{i=1}^{N} g_i, \; j = 1,2,...,K \qquad (6)$$

Here $c_{ij}, j = 1,2,...,K$ denotes the *center* (coordinates) of the *i*-th neuron in the *K*-dimensional input space and $0 < g_i \le 1, i = 1,2,...,N$ are the *normalized weights* of the neurons:

$$g_i = m_i / M; \; i = 1,2,...,N \qquad (7)$$

$m_i \le M, i = 1,2,...,N$ is the number of all data points: $\mathbf{x}_s, s = 1,2,...,m_i$, for which the *i*-th neuron is a *winning neuron* (i.e. the neuron with the shortest *Euclidean* distance to all of these data points). Obviously, the following equation holds: $\sum_{i=1}^{N} m_i = M$ and therefore $\sum_{i=1}^{N} g_i = 1$.

2) The *Weighted Average Size WAS* of the operation mode (and its respective CIM) is a *scalar* value, which takes into account the normalized weights of all neurons and the Euclidean distance ED_{pq} between all pairs of neurons, $\{p,q\}, p = 1,2,...,N; q = 1,2,...,N$, as shown in the next two equations (8) and (9):

$$WAS = \sum_{p=1}^{N-1} \sum_{q=p+1}^{N} ED_{pq} w_{pq} \Big/ \sum_{p=1}^{N-1} \sum_{q=p+1}^{N} w_{pq} \; . \qquad (8)$$

where $w_{pq} = g_p \times g_q, \quad p = 1,2,...,N; q = 1,2,...,N \qquad (9)$

Fig. 4 shows the locations of the centers-of-gravity **CG** for all *5* operation modes, computed by (6) and (7). It is seen from this figures that **CG** of some modes (such as: *Mode 1* and *Mode 4* and also *Mode 3* and *Mode 5*) are quite close to each

ISBN: 978-989-95079-6-8

other in the *K*-dimensional input space, which could result in a wrong classification.

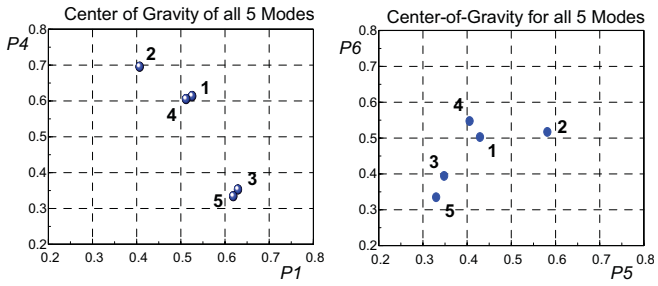


Figure 4:Center-of-Gravities **CG** for all *5* operation modes.

The following Table 1. shows the sizes *WAS* of all *5* Operation Modes, computed by (8) and (9). It also can be noticed here that some sizes are quite similar (such as *WAS* for *Mode 3* and *Mode 4*), which could also lead to wrong classification.

Table 1: Model Size *WAS* for all *5* operation modes.

Mode	Mode**1**	Mode**2**	Mode**3**	Mode**4**	Mode**5**
Model Size	*0.39538*	*0.38647*	*0.42592*	*0.42463*	*0.44486*

The above two parameters *P1* = **CG** and *P2* = *WAS* carry important information that can be used for selection of the two *features F1* and *F2* used as inputs of the Fuzzy Inference procedure for similarity analysis from Fig. 1.

We propose here an easy way to extract the features *F1* and *F2* as follows:

- The feature *F1* is a *scalar* value, computed as the *distance CGD* between the centers-of-gravities **CG** of a given pair $\{A,B\}$ of operation modes:

$$F1 = CGD_{AB} = \sqrt{\sum_{j=1}^{K} [CG_j^A - CG_j^B]^2} \qquad (10)$$

- Similarly the feature *F2* is a *scalar* value computed as the *difference WSD* between the weighted average sizes *WAS* of the same pair $\{A,B\}$ of operation modes, namely:

$$F2 = WSD_{AB} = |WAS_A - WAS_B| \qquad (11)$$

6 Structure of the block for fuzzy rule based similarity analysis

According to the block diagram shown in Fig. 1., we use the features *F1* and *F2*, computed by (10) and (11) as the inputs of the *Fuzzy Rule Based Inference Procedure* for similarity analysis of a given pair $\{A,B\}$ of operation modes.

Thus the Fuzzy Rule Based Procedure becomes a two-input / one output fuzzy system, as follows: $D = \mathbf{F}(F1,F2)$. Here $0.0 \le D \le 1.0$ is the *Difference Degree* (or *Dissimilarity Degree*). A difference degree $D = 0$ means that the operation modes *A* and *B* are *identical* (equal) and difference degree $D = 1$ means that *A* and *B* are *completely different modes*.

As well known [2], the fuzzy decision procedure consists of the following three main computation steps, as follows:

1) *Fuzzyfication* (with triangular Membership Functions);
2) *Fuzzy Inference* (with Product Operation) and
3) *Defiuzzification* (Weighted Mean Average).

For the next simulations in the paper, we assume *five triangular membership functions* that characterize linguistically the two inputs (features), namely *F1* and *F2*. They are used in the *fuzzification step* and have the following linguistic meaning: *VS = Very Small*; *SM = Small*; *MD = Medium*; *BG = Big* and *VB = Very Big*.

The Fuzzy Rule Base for the Fuzzy Inference procedure is shown in Fig. 5. It consists of *25* fuzzy rules, each of them with individual output as one of the *9 crisp* numerical values (*Singletons*): $U_1,U_2,...,U_9$, as seen in the figure. Note that the following inequality is required for achieving meaningful (plausible) results from the fuzzy similarity analysis:

$$U1 < U2 < ... < U9 \qquad (12)$$

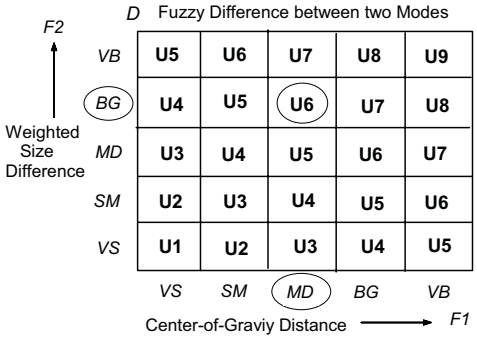


Figure 5: The Fuzzy Rule Base used for fuzzy similarity analysis.

The structure of the fuzzy rule base in Fig. 5 has been generated by using general *human logic and experience* from comparison and evaluation of data sets and other objects.

The well common weighted average method is used for the *Defuzzification* step, as follows:

$$D = \sum_{i=1}^{L} u_i v_i \Big/ \sum_{i=1}^{L} v_i \qquad (13)$$

Here $0 \le v_i \le 1, i=1,2,...,L$ is the *Firing* (*Activation*) *Degree* of the *i*-th fuzzy rule and $L = 25$ is the total number of the fuzzy rules. All rules have their individual crisp values (*Singletons*): $u_i \in [U_1,U_2,...,U_9], i=1,2,...,L$, according to the notations of the Fuzzy Rule Base in Fig. 5. For example, the *crisp* output of the Fuzzy Rule *No. 14* ($i = 14$), marked by circles in Fig. 5 is, as follows:

IF(*P1 is* MD AND *P2 is* BG) THEN $u_{14} = U6$

7 Optimization of the procedure for fuzzy similarity analysis

Once the structure of the Fuzzy Rule base is fixed, according to Fig. 5., then all the remaining parameters in the fuzzy inference procedure should be appropriately tuned. These are the parameters (*locations*) of the triangle membership functions and the *singletons* (consequents of the fuzzy rules). If successful optimization of these parameters is performed (according to a given optimization criterion), then we can expect correct (plausible) classification results.

There are two problems in connection with the optimization, namely: 1) construction of the *optimization criterion* with possible constraints and 2) selection of the *optimization method* (algorithm) to be used. As for the best selection of the

ISBN: 978-989-95079-6-8

optimization method, this is not a topic of interest in this paper, so we concentrate further on in the paper on the first problem, namely the selection of the optimization criterion.

We have collected *three groups* of data sets for all *5* operation modes of the hydraulic excavator, as well as *two* additional data sets representing two *different* (unknown) modes. One data set actually represents an *idling mode* of HE, while the other mode is not quite typical for the excavator (*moving* on the ground).

All the collected data sets have been first compressed into respective compressed information models (CIMs), according to the concept of the block diagram in Fig. 1. Then the respective CIMs have been split into the following *three* groups of models:

1) Five *training* models for each of the preliminary known operation modes, named as Mode**1**, Mode**2**,…,Mode**5**.

2) Five *test* models for the same group of five operation modes, named as: Test**1**, Test**2**, …, Test**5**;

3) A sequence of *16* operation models, used for *validation* of the classification procedure. These validation models are named as: Val**1**, Val**2**, Val**3**, …, Val**16** (or shortly: **1,2,…,16**). Then the sequence **1,2,…,16** is used to simulate the *Online Classification* process.

Table 2 shows the results from the similarity analysis between the *test modes* and the *training modes* for the initially assumed parameters (before optimization). The columns named as *Rank No.1* and *Rank No.2* refer to the *first* and *second* choice for classification (i.e the *least* difference and the *second least* difference). It is easy to notice the *contradiction* in this table arising at operation *Mode 5*. It is wrongly classified as *Mode 3* in this table.

Table 2: Similarity results before optimization.

	Training Modes: Mode **1** – Mode **5**			
Test	Rank **No.1**		Rank **No.2**	
Modes:	*No.*	*D*	*No.*	*D*
Test **1**	**1**	0.092	2	0.146
Test **2**	**2**	0.050	1	0.156
Test **3**	**3**	0.012	5	0.106
Test **4**	**4**	0.069	2	0.174
Test **5**	**3**	0.138	5	0.144

In order to correct this contradiction, we have to tune the parameters of the fuzzy inference procedure. It is done in this paper by assuming a special type of *optimization criterion,* which counts the *discrepancy DIS* between the human decision (or *human preference*) and the result from the computer-based classification. Table 3 shows example of one (plausible) human preference for classification of all *5* operation modes.

Table 3: Construction of the Optimization Criterion: *Preferred Human Decision*

Rank No. 1			**Rank No. 2**		
Test Mode	Train Mode	*D*	Test Mode	Train Mode	*D*
1	**1**	*0.10*	**1**	**4**	*0.20*
2	**2**	*0.10*	**2**	**1**	*0.20*
3	**3**	*0.10*	**3**	**5**	*0.20*
4	**4**	*0.10*	**4**	**1**	*0.20*
5	**5**	*0.10*	**5**	**3**	*0.20*

Now, by summing up the discrepancies between all *10* human preferences listed in Table 3 and the respective computer results from Table 2, we can compute the initial value of the optimization criterion as: *DIS = 0.5052.*

Further on we use a relatively simple (random search) optimization algorithm (further details omitted here) for a *two-stage* tuning of the fuzzy inference parameters:

- *Stage 1* optimizes the three intermediate locations (*SM, MD, BG*) of the membership functions for both inputs *F1* and *F2*, which means *3 + 3 = 6* optimization parameters;
- *Stage 2* optimizes the singleton values: $U_1,U_2,...,U_9$, taking into account the constraints (12).

The two stages were performed once in a consequence: *Stage1* → *Stage 2* with *30000* iteration steps for each stage. As a result the optimization criterion *DIS* was decrease to *0.3011* (after *Stage1*) and to *0.2983* (after *Stage2*). The final classification results are shown in Table 4. It is seen that all the classification results have become *correct* and with much closer degrees to the human preferences from Table 3.

Table 4: Similarity results after optimization.

	Training Modes Mode **1** – Mode **5**			
Test	Rank **No. 1**		Rank **No. 2**	
Modes:	*No.*	*D*	*No.*	*D*
Test **1**	**1**	0.123	2	0.169
Test **2**	**2**	0.096	1	0.192
Test **3**	**3**	0.016	5	0.168
Test **4**	**4**	0.114	2	0.208
Test **5**	**5**	0.141	3	0.189

The next three figures: Fig. 6., Fig. 7. and Fig. 8. are self explanatory. They show how the optimization has changed the parameters of the membership functions, the singletons and the Fuzzy Rule base response surface.

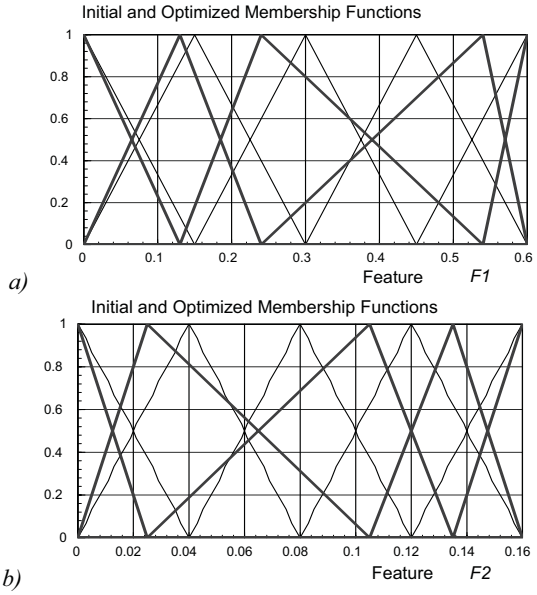


Figure 6: Results from the tuning of the membership functions. Bold lines denote the optimal positions.

8 Online classification results

After the fuzzy inference procedure has been tuned appropriately, then the whole *online classification* procedure has been tested on the sequence **1,2,…,16** of *16 validation* models Val**1**,Val**2**,…, Val**16**, according to the Block Diagram in Fig. 1. The results from the classification are shown in Fig. 9 and Fig. 10.

ISBN: 978-989-95079-6-8

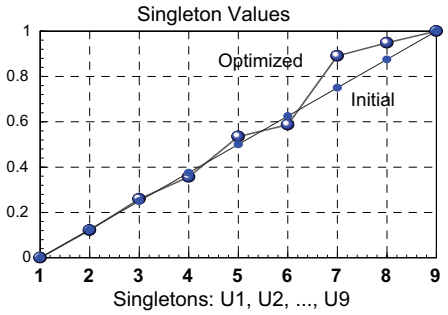


Figure 7: Results from the tuning of the singleton values.

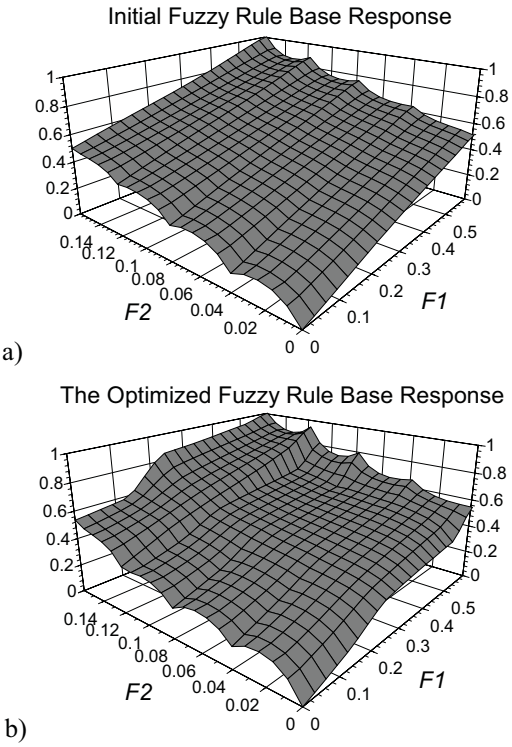


Figure 8: Response Surface of the fuzzy rule base: *a)* before optimization; *b)* after optimization.

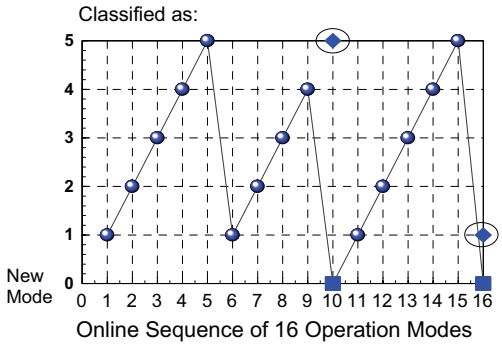


Figure 9: Classification Results for all *16* validation modes 1,2,…,16, presented as online sequence.

The "ball" symbols in Fig. 9. correspond to the correct classified modes. The two "diamond" symbols for operations *10* and *16* represent the computer classification results, while the respective two "square" symbols represent the actual "new" (unknown) operation modes.

It is seen from the similarity results in Fig. 9 that the operation sequences *10* and *16* were classified as *Mode 5* and *Mode 1* respectively. However, this decision was later *rejected,* as shown in Fig. 10, because their dissimilarity degrees exceed the predetermined threshold: *Th = 0.25*.

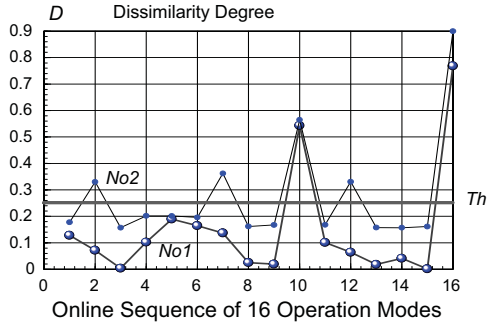


Figure 10: Confirmation results from the online classification with the predetermined threshold *Th = 0.25*.

Therefore finally the two operations *10* and *16* are assumed to correspond to *new* (*unknown*) operation modes. This result reflects the real situation.

9 Conclusions

The *Onine* classification of operation modes in continuously working machines and systems proposed in this paper uses fuzzy inference for similarity analysis. It is also suitable for possible applications in other areas of interest, such as search engines, classification and similarity analysis of images etc.

Essential feature of the whole proposed computation scheme is its *incremental* ability, in a sense that the newly detected modes could be added as new members of the current Knowledge Base and further on used for online classification.

The main originality of the proposed classification scheme is that it is a *human-assisted* classification, in which human experience and preference is taken into account and included into the optimization criterion. This criterion is further on used for tuning the parameters of the membership functions and the singletons in the fuzzy inference procedure so that to achieve the best possible matching between the computer results and human preferences.

There are several directions to improve the current status of this research. One is to investigate other optimization criterions and to select good, effective optimization method. Another direction is to solve the problem of *ever-growing* Knowledge Base during long time operation, by appropriate *pruning* and *merging* the modes in KB. In such way the whole classification system would become more flexible, *evolving classification* system.

References

[1] J.C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, 1981.

[2] Ch.M. Bishop, *Neural networks for pattern recognition.* Oxford University Press, Oxford, 2003.

[3] T. Martinetz, S. Berkovich and K. Schulten, Neural-Gas network for vector quantization and its application to time-series prediction, *IEEE Trans. Neural Networks*, vol. 4, No. 4, 558 – 569, 1993.

[4] Ya-Jun Zhang and Zhi-Quang Liu, Self-Splitting competitive learning: A new On-Line clustering paradigm, *IEEE Trans. on Neural Networks,* vol. 13, No. 2, 369 – 380, 2002.

[5] Xueju Fu, C. Ong, S. Keerthi, G. Huang and L. Goh, Extracting the knowledge embedded in support vector machines, Proc. of the *IEEE Int. Joint Conf. on Neural Networks*, IJCNN 2004, vol. 1, 291-296, 2004.

[6] N. Kwak and Ch-Ho Choi, Input feature selection for classification problems, *IEEE Trans. on Neural Networks,* vol. 13, No. 2, 369 – 380, 2002.

[7] G. Vachkov, Classification of images based on information compression and fuzzy rule based similarity analysis, CD-ROM Proc. *World Congress on Computational Intelligence* WCCI 2008, Hong Kong, June, 2326 – 2332, June, 2008.

ISBN: 978-989-95079-6-8

Introducing Relation Compactness for Generating a Flexible Size of Search Results in Fuzzy Queries

Xiaohui Tang Guoqing Chen Qiang Wei

School of Economics and Management, Tsinghua University
Beijing 100084, China
Email: weiq@sem.tsinghua.edu.cn

Abstract-*Generating relevant and manageable search results is of great interest in database queries and information retrieval nowadays. This paper proposes an approach to measuring relation compactness and to generating a compact set of query outcomes under a fuzzy relational data model. The approach has desirable properties that (1) the resultant compact set is unique and contains mutually distinct tuples at the specified degree, and (2) the resultant compact set is information-equivalent to the collection of original query outcomes. Moreover, it is deemed appealing that the approach provides a way to obtain query outcomes of different size in a flexible manner according to the specified degree of compactness or the preferred number of tuples by users.*

Keywords—Tuple closeness, Fuzzy databases and queries, Relation compactness, Information equivalence, Web search

1 Introduction

The commonly used database model nowadays is the relational database (*RDB*) model initiated by Codd [1] which is generally designed under the assumption that the data/information is precise and queries are crisp. However, decision makers often face a growing need for representing and processing imprecise and uncertain information. Particularly in a web environment, approximate match is often preferred in generating search outcomes when exact match is unable to produce satisfactory results from targeted databases.

Since 1980's, a number of fuzzy relational database (*FRDB*) models and fuzzy query techniques [2-6] have been proposed to deal with imprecision and uncertainty of fuzziness type by means of fuzzy logic [7]. For instance, with the data stored in classical *RDB*, Kacprzyk and Zadrozny [8] proposed a system called *FQUERY* that could execute fuzzy queries such as "find (all) records such that most of the (important) clauses are satisfied (to a degree in [0, 1])" for crisp data stored in the Microsoft Access DBMS. In [9], a method to transform query conditions described by fuzzy numbers into the classical *SQL* queries was presented. In [10], the authors extended the *SQL* queries into *SQLf* to allow the expression of imprecise queries. Furthermore, a method was developed to transform the *FRDB* into weighted *RDB* in that a nesting mechanism was introduced to support the expression of some query results such as projection-selection-join operations [11]. In [12-13], we extended basic algebraic operators to deal with fuzzy data modeling and queries in forms of database design as well as equivalent definitions and transformation rules of the algebraic operators.

Notably, when querying databases with data of high volume, the size of the outcomes can easily be very large, even massive in a web database search context. For instance, a fuzzy query such as "*Select* Customer Name *From R Where* Age is about 25 AND Location is near B" requires an approximate match between query conditions and values of corresponding attributes. Another example is a web search, say via Google. Just click and see how many pages of outcomes to appear when keying in *"Population of Beijing"* for an exact match and *Population of Beijing* (notably without quotation marks) for an approximate match. Efforts have been made to deal with queries and web search with imprecise information and approximate match measures, so as to enrich the representation semantics and strengthen the power of search engines [5, 14-16]. Generally speaking, approximate match provides flexible queries, which is considered meaningful and useful in many cases, whereas the size of its outcomes is usually larger than that of the exact match.

Hence, the size of query outcomes becomes an issue of concern. One notable approach was to provide the users with the top-*k* results according to a ranking based on a certain evaluation function *F* between query conditions and all data records [17-18]. While sometimes the top-*k* results are sufficient for fuzzy/soft queries with approximate match, it is desirable and meaningful if a query/search could also guarantee that the selected results have no information loss as far as all the query results are concerned. The focus of this paper is then to provide the users with a compact set of query outcomes, in that the compact set (a) is smaller than and information-equivalent to the set of all query outcomes, and (b) at the same time has its size flexibly specified by users upon their need and preference.

2 Fuzzy Relational Database Model

Fuzzy logic incepted by Zadeh [7] attempts at quantifying and reasoning with imprecision and uncertainty that is common in the real world. Possibility distributions provide a graded semantics to natural language statements such as "the customer is young", which are often used in our daily communications. For example, the linguistic term "Young" can be defined as

$$\pi_{Young}(age) = \begin{cases} 1 & 0 \le age \le 30 \\ (50 - age)/20 & 30 < age \le 50 \\ 0 & age > 50 \end{cases} \quad (1)$$

ISBN: 978-989-95079-6-8

Two specific fuzzy relations are of particular interest, namely, closeness relation and similarity relation. A closeness relation c is a fuzzy relation on $X\times X$ such that $\forall x, y\in X$, $c(x, x) = 1$ (reflexive) and $c(x, y) = c(y, x)$ (symmetric). A similarity relation s is a fuzzy relation on $X\times X$ such that $\forall x, y, z\in X$, $s(x, x) =1$ (reflexive), $s(x, y) = s(y, x)$ (symmetric), and $s(x, z) \geq \sup_{y\in X} \min(s(x, y), s(y, z))$ (sup-min transitive). Apparently, similarity relation is a special case of closeness relation.

In this paper, the extended-possibility-based model is considered to be the underlying fuzzy database model since it facilitates handling both imprecision in attribute values (e.g., Age is young) and fuzziness in domain elements (e.g., between classic and gold for the domain of (customer) Class) in terms of possibility distributions (including fuzzy sets and linguistic terms) and closeness relations, respectively [4, 19]. In light of data representation, this model is deemed to be a general setting compared with two known models, namely, the so-called possibility-based model [4] where attribute values can be possibility distributions, and the so-called similarity-based model [3] where domain elements can be associated by similarity relations (reflective, symmetric and sup-min transitive). Specifically, in the extended-possibility-based model, a relation R is a subset of $\Pi(D_1) \times \Pi(D_2) \times \ldots \times \Pi(D_g)$, where $\Pi(D_i) =\{\pi_{Ai} \mid \pi_{Ai}$ is a possibility distribution of attribute A_i on domain $D_i\}$, and a closeness relation c_i is associated with domain D_i ($1 \leq i \leq g$). In addition, an g-tuple t of R is of the form: $t(\pi_{A1}, \pi_{A2}, \ldots, \pi_{Ag})$. An example tuple recording a customer's Name, Age and Class can be (Tony, {0.7/28, 1.0/33, 0.8/36}, diamond). The closeness relation c_{Class} on domain D_{Class}={classic, silver, gold, diamond} can be pre-defined by the managers as shown in Table 1.

Table 1: A closeness relation c_{Class}

c_{Class}	classic	silver	gold	diamond
classic	1.0	0.5	0.0	0.0
silver		1.0	0.75	0.25
gold			1.0	0.75
diamond				1.0

Given two g-tuples $t_p(\pi_{A1}, \pi_{A2}, \ldots, \pi_{Ag})$, $t_q(\pi'_{A1}, \pi'_{A2}, \ldots, \pi'_{Ag})$, where π_{Ai} and π'_{Ai} are normalized ($1 \leq i \leq g$), an example measure for the closeness of two values is [22]:

$$E_c(\pi_{A_i}, \pi'_{A_i}) = \sup_{\substack{x,y\in D_i \\ c_i(x,y)\geq\alpha_i}} \min(\pi_{A_i}(x), \pi'_{A_i}(y)) \qquad (2)$$

Here c_i is a closeness relation on domain D_i, $\alpha_i \in [0, 1]$ is a threshold specified by experts or the database managers. Other measures can be found in [19-21]. Then, the closeness of two tuples t_p and t_q can be considered as:

$$F_c(t_p, t_q) = \min(E_c(\pi_{A_1}, \pi'_{A_1}), E_c(\pi_{A_2}, \pi'_{A_2}), \ldots, E_c(\pi_{A_g}, \pi'_{A_g})) \qquad (3)$$

3 Tuple Extraction

Regarding tuple extraction in fuzzy relational databases, there are two issues to consider. One is to evaluate tuple closeness, the other is to extract representative tuples.

3.1 The evaluation of tuple closeness

For the fuzzy database, given an attribute A with domain D, let $\pi_1, \pi_2, \ldots, \pi_n$ be n possibility distributions as values of A, then a closeness relation on $\Pi(D)\times\Pi(D)$ is a mapping from $\Pi(D)\times\Pi(D)$ to [0,1], which can be represented by an $n\times n$ fuzzy matrix M where $e_{ij} = E_c(\pi_i, \pi_j)$, $1\leq i, j \leq n$, as follows:

$$M = \begin{pmatrix} e_{11} & e_{12} & \ldots & e_{1n} \\ e_{21} & e_{22} & \ldots & e_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ e_{l1} & e_{l2} & \ldots & e_{nn} \end{pmatrix} \qquad (4)$$

Further, for $0 \leq \lambda \leq 1$, where λ is a threshold which is usually determined by users or domain experts, its λ-cut (M_λ) is a classical relation defined by:

$$(e_{ij})_\lambda = \begin{cases} 1 & e_{ij} \geq \lambda \\ 0 & e_{ij} < \lambda \end{cases} \qquad (5)$$

Moreover, it is known that if M is a closeness relation, then its transitive closure M^+ (i.e., M^+ is a series of max-min compositions onto M itself with $M^+ = M^p = M^{p+1}$, $p\geq 1$) is a similarity relation, and M^+ converges within n-1 compositions [23-25]:

$$M^+ = M^p, \ p \leq l - 1 \qquad (6)$$

Furthermore, the equivalence classes of $(M^+)_\lambda$ constitute a partition of the domain concerned [26].

Definition 1: Given a relation $R = \{t_1, t_2, \ldots, t_n\} \subseteq \Pi(D_1) \times \Pi(D_2) \times\ldots\times\Pi(D_g)$, where tuple t_i is $t_i(\pi_{i1}, \pi_{i2}, \ldots, \pi_{ig})$, and threshold $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_g)$ with equivalence classes of $(M^+)_{\lambda i}$, $i = 1, 2, \ldots, n$, then two tuples $t_i(\pi_{i1}, \pi_{i2}, \ldots, \pi_{ig})$ and $t_j(\pi_{j1}, \pi_{j2}, \ldots, \pi_{jg})$ are called $\boldsymbol{\lambda}$-close if for all $k \in \{1, 2, \ldots, g\}$, π_{ik} and π_{jk} are in the same equivalence class. □

Example 1: Given threshold $\boldsymbol{\lambda} = (0.7, 0.8)$ and a relation $S \subseteq \Pi(D_1) \times \Pi(D_2)$ as shown in Table 2. For the sake of simplicity, assume that a_1, a_2, a_3, and a_4 are mutually distinct; so are b_1, b_2, b_3 and b_4. Then the $\boldsymbol{\lambda}$-close tuples can be generated as follows.

Table 2: A relation S with fuzziness

	A_1	A_2
t_1	$0.3/a_1, 0.7/a_2, 0.7/a_3, 0.7/a_4$	$0.53/b_1, 0.59/b_2, 0.62/b_3, 0.85/b_4$
t_2	$0.6/a_1, 0.4/a_2, 0.1/a_3, 0.8/a_4$	$0.92/b_1, 0.67/b_2, 0.09/b_3, 0.06/b_4$
t_3	$0.2/a_1, 0.6/a_2, 0.1/a_3, 1.0/a_4$	$0.45/b_1, 0.75/b_2, 0.06/b_3, 0.75/b_4$
t_4	$0.5/a_1, 0.1/a_2, 1.0/a_3, 0.9/a_4$	$0.61/b_1, 0.66/b_2, 0.35/b_3, 0.69/b_4$
t_5	$0.9/a_1, 0.6/a_2, 0.2/a_3, 0.2/a_4$	$0.56/b_1, 0.32/b_2, 0.98/b_3, 0.30/b_4$
t_6	$0.4/a_1, 0.2/a_2, 0.2/a_3, 0.4/a_4$	$0.81/b_1, 0.30/b_2, 0.93/b_3, 0.99/b_4$

First, for A_1 we can have:

$$M = \begin{pmatrix} 1 & 0.7 & 0.7 & 0.7 & 0.6 & 0.4 \\ 0.7 & 1 & 0.8 & 0.8 & 0.6 & 0.4 \\ 0.7 & 0.8 & 1 & 0.9 & 0.6 & 0.4 \\ 0.7 & 0.8 & 0.9 & 1 & 0.5 & 0.4 \\ 0.6 & 0.6 & 0.6 & 0.5 & 1 & 0.4 \\ 0.4 & 0.4 & 0.4 & 0.4 & 0.4 & 1 \end{pmatrix},$$

where each element on (i, j) in the matrix represents the closeness of t_i and t_j on A_1. Then, with $(M^+=M^3=M^2)_{0.7}$, three equivalence classes are obtained: $\{\pi_{11}, \pi_{21}, \pi_{31}, \pi_{41}\}$, $\{\pi_{51}\}$ and $\{\pi_{61}\}$. Consider attribute A_2, in the similar manner, there are three equivalence classes: $\{\pi_{12}, \pi_{22}, \pi_{52}, \pi_{62}\}$, $\{\pi_{32}\}$ and $\{\pi_{42}\}$.

ISBN: 978-989-95079-6-8

That means that values of t_1 and t_2 are in the same class for A_1 (i.e., π_{11}, π_{21}) as well as for A_2 (i.e., π_{12}, π_{22}), indicating that t_1 and t_2 are λ-close.

Finally, it is worthwhile to mention that, in case there are closeness relations on domains, the calculation of M can be readily made in consideration of c_i and α_i in E_c. For instance, suppose we have $c_1 = c_{\text{Class}}$ for A_1 with a_1 = classic, a_2 = silver, a_3 = gold and a_4 = diamond as shown in Table 1. Given α_1 = 0.7, since $c_1(a_2, a_3) = c_1(a_3, a_4) = 0.75$, we have $e'_{45} = 0.6 \neq e_{45} = 0.5$. In general, $M \subseteq M'$. □

3.2 Relation compactness

In addition to closeness of any two tuples, it is of interest to investigate the compactness of a given relation R composed of n tuples in the fuzzy relational database. Here the compactness of a relation refers to the degree of non-redundancy. For a given relation, the more redundant tuples it contains, the lower the degree of compactness. Straightforward ways of evaluating the compactness are to use the minimum closeness of any two tuples of R (*e.g.,* $1 - \min_{t,t' \in R} F_c(t, t')$) or the average closeness of any two tuples of R(*e.g.,* $1 - \sum_{t,t' \in R} F_c(t, t')/n(n-1)$) to reflect the compactness of R in whole. But it shall be noted that neither $1 - \min_{t,t' \in R} F_c(t, t')$ nor $1 - \sum_{t,t' \in R} Fc(t, t')/n(n-1)$ could describe the compactness of R in a satisfactory fashion. For example, suppose that all the n tuples in R are classical tuples for the sake of simplicity, i.e., for any $t, t' \in R$, $F_c(t, t') = 1$ or $F_c(t, t') = 0$. If there are $(n-1)$ tuples which are identical, only one tuple is not identical to them, then $1 - \min_{t,t' \in R} F_c(t, t') = 1$. But we know that there are a lot of tuples whose closeness degrees to each other are 1 in such a situation. On the other hand, if half of the tuples are identical to each other, and the other half of the tuples are also identical to each other (but different from the first half), then, $1 - \sum_{t, t' \in R} Fc(t, t')/n(n-1) = 2 \times (n/2) \times (n/2 - 1)/n(n-1) = 0.5 + 1/(2n - 2) > 0.5$. But in this situation, if n is large enough, e.g., $n = 100$, we will surely think that the compactness degree of R will be very close to 0, as 98% of the tuples can be deleted. Thus, a new measure, relation compactness, is proposed to evaluate a given relation R being compact. This measure not only works well in both of the above situations, but also possesses some good properties. Let us first consider it in classical relations, and then extend it into the relations in fuzzy databases.

In information theory [27], consider a single discrete information source, it may produce different kinds of symbol sets $A = \{a_1, a_2, \ldots, a_n\}$. For each possible symbol set A there will be a set of probabilities *pi* of producing the various possible symbols a_i ($\sum_{i=1}^{n} p_i = 1$), where these symbols are assumed successive and independent. Thus there is an entropy *Hi* for each a_i. The entropy of this given piece of information will be defined as the average of these H_i weighted in accordance with the probability of occurrence of the symbols in question,

$$H = H(p_1, p_2, \ldots, p_n) = -\sum_{i}^{n} p_i H_i = -\sum_{i}^{n} p_i \log p_i \qquad (7)$$

where the default log base is 2.

Similarly to [28], if there is a relation S with a set of classical tuples, which can be divided into m classes $C_1, C_2, \ldots, C_m$, the probability of a random tuple belonging to class C_i is n_i/n, where n_i is the number of tuples in C_i, and n is the number of tuples in S. Here, n_i/n is also called the probability of class C_i in S. The expected information for classifying this given relation S is:

$$H(n_1, n_2, \ldots, n_m) = -\sum_{i=1}^{m} p_i H_i = -\sum_{i=`}^{m} \frac{n_i}{n} \log \frac{n_i}{n} \qquad (8)$$

Definition 2: Let $R = \{t_1, t_2, \ldots, t_n\}$ be a classical relation with n tuples, R be divided into m classes $C_1, C_2, \ldots, C_m$ according to tuple identity (i.e., every tuple in C_i is identical to each other), and n_i be the number of tuples in C_i. Then the relation compactness of R is defined as

$$RC'(n_1, n_2, \ldots, n_m) = H(n_1, n_2, \ldots, n_m)/\log n$$
$$= -(\sum_{i=1}^{m} p_i H_i)/\log n = -(\sum_{i=1}^{m} \frac{n_i}{n} \log \frac{n_i}{n})/\log n \qquad (9)$$

For a relation R, its degree of compactness (RC') reflects the extent to which the tuples of R are not redundant, measured by the "amount of information" in R. In other words, the higher RC' is, the less redundant the tuples in R, meaning that the more information R contains.

Next, let us consider the fuzzy extension. Suppose there are n tuples with fuzziness involved in $R = \{t_1, t_2, \ldots, t_n\}$, where these tuples can be close to each other. To describe the degree at which a given relation with fuzziness is compact, the concept of relation compactness in the classical context could be extended to cope with the closeness of these tuples. A tuple t_k may not totally belong to class C_i, but belong to class C_i at a certain degree, e.g., $O_k^i \in [0, 1]$. Thus, we will use $\sum_{k=1}^{n_i} O_k^i$ instead of n_i in the extended relation compactness. Concretely, suppose we have n_i λ-close tuples in class $C_i = \{t_1, t_2, \ldots, t_{ni}\}$ of relation R. For each tuple t_k, $1 \le k \le n_i$, the degree of t_k belonging to C_i (or the closeness of t_k to C_i) can be defined as:

$$O_k^i = \sum_{j=1}^{n_i} F_c(t_k, t_j)/n_i \qquad (10)$$

Definition 3: Given a relation with fuzziness $R = \{t_1, t_2, \ldots, t_n\}$, where R can be divided into m classes $C_1, C_2, \ldots, C_m$ according to λ-close (i.e., every tuple in C_i is λ-close to each other), and n_i is the number of tuples in C_i. The relation compactness in R is,

$$RC(n'_1, n'_2, \ldots, n'_m) = -(\sum_{i=1}^{m} \frac{n'_i}{n} \log \frac{n'_i}{n})/\log n \qquad (11)$$

where $C_i = \{t_1, t_2, \ldots, t_{ni}\}$, $1 \le i \le m$, $1 < n_i \le n$, O_k^i is the degree of t_k belonging to C_i, and $n'_i = \sum_{k=1}^{n_i} O_k^i$ is the Σcount operation for the "effective number" of tuples in class C_i. □

As the classical relation is a special case of a relation with fuzziness, if all the tuples in $C_i = \{t_1, t_2, \ldots, t_{ni}\}$ are classical tuples, then they are equal to each other, i.e., $F_c(t_k, t_j) = 1$, $O_k^i = 1$, and $n'_i = n_i$ ($1 \le j, k \le n_i$). So $RC'(n_1, n_2, \ldots, n_m)$ is a special case of $RC(n'_1, n'_2, \ldots, n'_m)$. Note that one may also use some

ISBN: 978-989-95079-6-8

other measures (e.g., [19]) instead of O_k^i for the degree at which t_k belongs to C_i in (10). For *RC* in the fuzzy database context, the following proposition holds.

Proposition 1: Give a relation $R = \{t_1, t_2, ..., t_n\} \subseteq \Pi(D_1) \times \Pi(D_2) \times \ldots \times \Pi(D_g)$, where it is divided into m classes $C_1, C_2, ..., C_m$ according to $\boldsymbol{\lambda}$-close, $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, ..., \lambda_g\}$, and n_i is the number of tuples in C_i $(1 \le i \le m)$, then

(1) R is not $\boldsymbol{\lambda}$-close, i.e., $m = n$, $n'_1 = n'_2 = ... = n'_m = 1$, if and only if $RC(n'_1, n'_2, ..., n'_m) = 1$.

(2) Suppose $\min_{i=1}^{g} \lambda_i \ge e^{-1} \approx 0.37$, if R is totally $\boldsymbol{\lambda}$-close, i.e., $m = 1$, $n'_1 = n$, then $RC(n'_1)$ decreases when any two of the tuples's closeness $F_c(t_k, t_j)$ increases $(1 \le k, j \le n)$. Especially, if $\boldsymbol{\lambda} = \mathbf{1}$, then $RC(n'_1) = RC'(n_1) = 0$.

(3) $0 \le RC(n'_1, n'_2, ..., n'_m) \le 1$.

(4) Suppose $(n'_p + n'_q)/n \le e^{-1} \approx 0.37$, if n is fixed, $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, ..., \lambda_g\}$, $\lambda_i = \lambda$, $i = 1, 2, ..., g$, then *RC* decreases when any two of the original classes (e.g., C_p, C_q) are merged into one class C'_p according to $\boldsymbol{\lambda}$-close.

For different n and m, the value of *RC* is different. If $m = 2$, then $RC = -((n'_1/n)\log(n'_1/n) + ((n-n'_1)/n)\log((n-n'_1)/n))/\log n$, with different n, *RC* is shown in Figure 1. If $m = 3$, then $RC = -((n'_1/n)\log(n'_1/n) + (n'_2/n)\log(n'_2/n) + ((n-n'_1-n'_2)/n)\log((n-n'_1-n'_2)/n))/\log n$, with $n = 100$, *RC* is shown in Figure 2. Figures 1 and 2 reflect that RC has a characteristic of convexity.

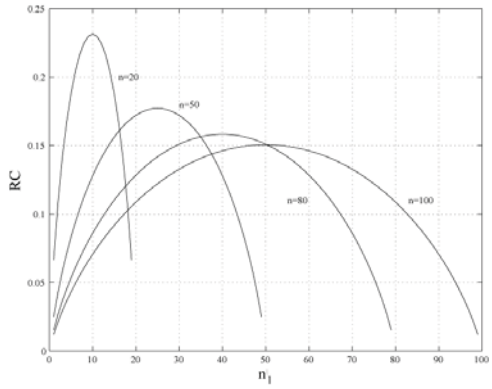


Figure 1: The value of *RC* ($m = 2$)

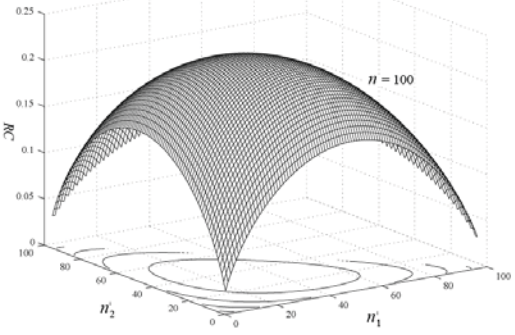


Figure 2: The value of *RC* ($m = 3$; $n = 100$)

3.3 Extraction of representative tuples

When it is known which tuples in a relation with fuzziness are $\boldsymbol{\lambda}$-close, then how to extract the representative tuples is the next problem of concern. It is considered ideal to have fewer tuples if they carry the same amount of original information. Usually, due to existence of tuple closeness, obtaining a smaller set of query outcomes becomes an effort of extracting representative tuples in light of "information equivalence".

Here we consider a center-based method for extracting the representative tuples. Conceptually, all the tuples in a same equivalence class are regarded to express approximately the same information and therefore one of them could be extracted to represent the class. Suppose we have n_i $\boldsymbol{\lambda}$-close tuples in class $C_i = \{t_1, t_2, ..., t_{ni}\}$ of a relation R. Then, we will keep the tuple whose closeness to C_i is the highest. This means, tuple t_p will be kept if $O_p^i = \max_{k=1}^{n_i} O_k^i$, $1 \le p \le n_i$.

Theorem 1: Suppose there are h $\boldsymbol{\lambda}$-close tuples in class $C = \{t_1, t_2, ..., t_h\}$ of a relation R, $C_i = C - \{t_i\}$, $1 \le i \le h$, and the relation compactness of C and C_i are *RC* and RC_i respectively. Then, tuple t_p $(1 \le p \le h)$ will be retained if and only if $RC_p = \max_{i=1}^{h} RC_i$.

Theorem 1 indicates that we can extract the representative tuples in light of relation compactness. It will keep the tuple that, without this tuple, the relation compactness of the other tuples will be the highest.

Example 2: Suppose we have had the $\boldsymbol{\lambda}$-close tuples of a fuzzy query about customers' age and salary as shown in Table 3.

Table 3: Customer's age and salary (in part)

	Age	Salary($)
t_1	0.6/25, 0.9/35, 0.6/40	0.8/4000, 0.9/5000, 0.8/6000
t_2	0.7/28, 0.8/35, 0.7/42	0.7/4500, 0.8/5000, 0.7/6000
t_3	0.6/25, 0.9/35, 0.8/42	0.9/4000, 0.7/5000, 0.6/6000

According to (2): $e_{12}(\pi_{11}, \pi_{21}) = 0.8$, $e_{13}(\pi_{11}, \pi_{31}) = 0.9$, $e_{23}(\pi_{21}, \pi_{31}) = 0.8$; $e_{12}(\pi_{12}, \pi_{22}) = 0.8$, $e_{13}(\pi_{12}, \pi_{32}) = 0.8$, $e_{23}(\pi_{22}, \pi_{32}) = 0.7$. Then, according to (3) and (10): $F_c(t_1, t_2) = 0.8$, $F_c(t_1, t_3) = 0.8$, $F_c(t_2, t_3) = 0.7$, $F_c(t_1, t_1) = F_c(t_2, t_2) = F_c(t_3, t_3) = 1$. $O_1 = 0.87$, $O_2 = 0.83$, $O_3 = 0.83$. Thus, t_2 and t_3 will be eliminated, t_1 will be kept.

Also, from the viewpoint of relation compactness, Table 3 can be considered as a class C, so, $C_1 = \{t_2, t_3\}$, $C_2 = \{t_1, t_3\}$, $C_3 = \{t_1, t_2\}$. For RC_1, $n'_1 = (F_c(t_2, t_2) + F_c(t_2, t_3))=2 + (F_c(t_3, t_3) + F_c(t_2, t_3))/2 = 1.7$. $RC_1 = -(1.7/2)\log(1.7/2)/\log 2 = 0.20$. In the same way, $RC_2 = RC_3 = 0.14$, thus, t_1 will be kept.

As far as the complexity of the tuple extraction is concerned, while introducing possibility distributions enriches the semantics that the data model represents, approximate match may lead to an increase in computational complexity. Further analysis shows that the complexity is polynomial and generally manageable. Suppose there are n tuples in relation R, the matrix M has $n \times n$ entries. For every entry, suppose there are, on average, p values in the two imprecise attribute values to compare, the cost is $O(p^2)$. The total cost to get M is then $O(p^2n^2)$. A reasonable estimate of p could be within 10. Therefore, the total cost to get M could be simplified to be $O(ln^2)$ where $l < 100$. For the transitive closure of M, the total cost is at $O(n^2)$ level for large n [24-25].

Furthermore, in consideration of generating matrix M, though detailed treatments go beyond the scope of this paper, there

ISBN: 978-989-95079-6-8

have been effective approaches in database queries, informational retrieval and web search, using a variety of measures, in respective fields, for keyword similarity, text match, semantic proximity, etc., in order to compare pair-wise closeness of data/information of interest (e.g., [30-34]).

4 Further Discussions

First, whether the extraction result is unique and representative tuples are not redundant is an interesting and important issue. Importantly, it can be proven that the result of the tuple extraction treatment in Section 3 is unique, and the resulting new relation is not $\boldsymbol{\lambda}$-close.

Second, we can also have the number of tuples in the outcome to be controlled by the setting of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_g)$. When λ_i increases, there will be more classes, thus, after the extraction, more tuples will be kept. Especially, with $\lambda_i = 1$, the treatment turns into the classical situation, in that no tuples will be eliminated.

Third, two representations are information equivalent if the transformation from one to the other entails no loss of information, i.e., if each can be constructed from the other [35]. In other words, the information in one is also inferable from the other, and vice versa [36-37]. Here, "information equivalence" in our discussions could be described as follows. Suppose there is an original query result R_1. The classified relation is R_2 (i.e., the $\boldsymbol{\lambda}$-close tuples are in the same classes), and the resultant relation (compact set) containing representative tuples is R_3, which is not $\boldsymbol{\lambda}$-close.

Compared with original relation R_1 of the query result, R_2 and R_3 contain the classification information. We can say that, in terms of $\boldsymbol{\lambda}$-close, R_2 and R_3 are information equivalent. If we have had R_2, then R_3 could be derived (i.e., via extracting) with respect to tuple closeness/relation compactness. If we have had R_3, suppose $R_3 = \{t_u, t_w, t_x, ...\}$, then every tuple of R_3 is not $\boldsymbol{\lambda}$-close, and can be considered as a class labeled as $C_u, C_w, C_x, ...$. For every tuple t_i of R, if t_i satisfies the fuzzy query conditions, then it could be compared with the tuples in the labeled classes $C_u, C_w, C_x, ...$, if there is a tuple t in one labeled class, e.g., C_u such that t_i and t are $\boldsymbol{\lambda}$-close, then add t_i to C_u. Consequently, we will get relation R_2, as the tuples in R_3 and the equivalence classes in R_2 are bijective.

Finally, let us discuss the approach from a more applied perspective. While searching via a search engine (e.g., Google or the like) or querying available databases for some information, it would be desirable if the database/web search management systems are capable of generating outcomes that are representative and of a manageable size. Even better is to control the query/search outcomes with different compactness degrees or with different sizes. A way to view different results is to specify the degree of relation compactness. If R_C is higher, there will be fewer outcomes, otherwise, more outcomes.

Example 3: Someone uses a search engine to browse for the literature about "fuzzy queries". To help discuss the procedure and treatment details of how the search results can be shortened or extended, we hereby only chose 10 resultant items as R_1, merely for the sake of illustrative simplicity:

a_1: A Fuzzy Database Model for Supporting a Concept-Based Query ...
a_2: Fuzzy Database Query Languages and Their Relational Completeness ...
a_3: FSQL (Fuzzy SQL), a Fuzzy Query Language
a_4: Fuzzy Database Modeling with XML
a_5: Fuzzy database systems - Fuzzy Systems, 1995. International Joint ...
a_6: Amazon.com: Fuzzy Database Modeling of Imprecise and Uncertain ...
a_7: Amazon.com: Fuzzy Database Modeling with XML
a_8: Database Schema with Fuzzy Classification and Classification Query ...
a_9: Type-2 Fuzzy Logic - Publications Database
a_{10}: Fuzzy Clustering for Content-based Indexing in Multimedia Database

Suppose the threshold set by the user is $\lambda = 0.7$ and the closeness of these articles generated with a keywords match technique (e.g., [30-34]) is shown in Table 4.

Table 4: The closeness degrees among $a_1, a_2, \ldots, a_{10}$

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
a_1	1	0.75	0.9	0.6	0.5	0.5	0.5	0.6	0.6	0.6
a_2		1	0.82	0.6	0.6	0.6	0.6	0.5	0.6	0.6
a_3			1	0.6	0.5	0.5	0.6	0.5	0.6	0.6
a_4				1	0.85	0.9	0.9	0.6	0.6	0.6
a_5					1	0.7	0.75	0.6	0.5	0.5
a_6						1	0.78	0.6	0.6	0.6
a_7							1	0.6	0.5	0.6
a_8								1	0.9	0.95
a_9									1	0.7
a_{10}										1

Now, let us consider how to obtain the outcomes with $RC|_{0.7}$ being 1 and 0.6 respectively. In doing so, the database management systems or search processing functions could execute the extracting process with $\lambda = 0.7$, $\lambda_1 = 0.8$, $\lambda_2 = 0.85$, $\lambda 3 = 0.9$ and $\lambda 4 = 1$ at the back-end as follows.

For $\lambda = 0.7$, R_1 can be classified as $\{a_1, a_2, a_3\}$, $\{a_4, a_5, a_6, a_7\}$ and $\{a_8, a_9, a_{10}\}$, extracting the representative tuples as described in Section 3.3, the final outcome is $S = \{a_3, a_4, a_8\}$, $RC(S)|_{0.7} = 1$.

For $\lambda_1 = 0.8$, R_1 can be classified as $\{a_1, a_2, a_3\}$, $\{a_4, a_5, a_6, a_7\}$ and $\{a_8, a_9, a_{10}\}$, $S_1 = \{a_3, a_4, a_8\}$, $RC(S_1)|_{0.8} = RC(S_1)|_{0.7} = 1$.

For $\lambda_2 = 0.85$, R_1 can be classified as $\{a_1\}$, $\{a_2\}$, $\{a_3\}$, $\{a_4, a_5, a_6, a_7\}$ and $\{a_8, a_9, a_{10}\}$, $S_2 = \{a_1, a_2, a_3, a_4, a_8\}$, $RC(S_2)|_{0.85} = 1$, $RC(S_2)|_{0.7} = 0.6$.

In the same manner, for $\lambda_3 = 0.9$, $S_3 = \{a_1, a_2, a_3, a_4, a_5, a_8\}$, $RC(S_3)|_{0.9} = 1$, $RC(S_3)|_{0.7} = 0.57$.

For $\lambda_4 = 1$, $S_4 = R_1 = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}\}$, $RC(S_4)|_1 = 1$, $RC(R_1)|_{0.7} = RC(S_4)|_{0.7} = 0.47$.

Thus, if the threshold $\lambda = 0.7$, the original result of the query is compact at a degree of $RC(R_1)|_{0.7} = 0.47$. The back-end can give the user a totally compact outcome $S = \{a_3, a_4, a_8\}$ (with $RC(S)|_{0.7} = 1$) by default, which is composed of the three items:

"FSQL (Fuzzy SQL), a Fuzzy Query Language"
"Fuzzy Database Modeling with XML"

ISBN: 978-989-95079-6-8

"Database Schema with Fuzzy Classification and Classification Query ..."

If user wants to review more outcomes, e.g., $RC|_{0.7} = 0.6$, then the DBMS can generate $S_2 = \{a_1, a_2, a_3, a_4, a_8\}$, with $RC(S_2)|_{0.7} = 0.6$. Two more specific "fuzzy query" related items are listed: "A Fuzzy Database Model for Supporting a Concept-Based Query ..." and "Fuzzy Database Query Languages and Their Relational Completeness ...". In this way, the list of the literature could be shortened or extended depending upon the need and preference of the user. □

5 Conclusion

This paper has introduced a measure, namely relation compactness, to express the degree of a given relation being compact. Then an approach to evaluating and extracting representative tuples has been proposed. It has been proven that the resultant set of query outcomes is unique, smaller in size, and information equivalent to the original result. Moreover, the approach enables the query/search users to obtain various sizes of results upon their need and preference in light of compactness.

Ongoing research is centering on an application with real data on tuple extraction in both database query and web search contexts.

Acknowledgements

The work was partly supported by the National Natural Science Foundation of China (70890083/70621061) and Tsinghua University's Research Center for Contemporary Management.

References

[1] E.F. Codd, "A relational model for large shared data banks", *Communications of the ACM*, vol. 13, no. 6, pp. 377-387, 1970.

[2] J.F. Baldwin, S.Q. Zhou, "A fuzzy relational inference language", *Fuzzy Sets and Systems*, vol. 14, pp. 155-174, 1984.

[3] B.P. Buckles, F.E. Petry, "A fuzzy representation of data for relational databases", *Fuzzy Sets and Systems*, vol. 7, pp. 213-226, 1982.

[4] H. Prade, C. Testemale, "Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries", *Information Sciences*, vol. 34, pp. 115-143, 1984.

[5] V. Owei, "An Intelligent Approach To Handling Imperfect Information In Concept-Based Natural Language Queries", *ACM Transactions on Information Systems*, vol. 20, no. 3, pp. 291-328, 2002.

[6] Z.M. Ma, W.J. Zhang, W.Y. Ma, "Extending object-oriented databases for fuzzy information modeling", *Information Systems*, vol. 29, no. 5, pp. 421-435, 2004.

[7] L.A. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 3-28, 1978.

[8] J. Kacprzyk, S. Zadrozny, "Computing with words in intelligent database querying: standalone and internet-based applications", *Information Sciences*, vol. 134, pp. 71-109, 2001.

[9] S.M. Chen, W.T. Jong, "Fuzzy Query Translation For Relational Database Systems", *IEEE Transactions on System, Man, and Cybernetics-Part B*, vol. 27, no. 4, pp. 714-721, 1997.

[10] P. Bosc, O. Pivert, "SQLf: a relational database language for fuzzy querying", *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 1, pp.1-17, 1995.

[11] P. Bosc, O. Pivert, "About project-selection-join queries addressed to possibilistic relational databases", *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 1, pp. 124-139, 2005.

[12] G.Q. Chen, E.E. Kerre, J. Vandenbulcke, "Normalization based on fuzzy functional dependency in a fuzzy relational data model", *Information Systems*, vol. 21, no. 3, pp. 299-310, 1996.

[13] X.H. Tang, G.Q. Chen, "Equivalence and Transformation of Extended Algebraic Operators in Fuzzy Relational Databases", *Fuzzy Sets and Systems*, vol. 157, no. 12, pp. 1581-1596, 2006.

[14] P. Buche, C. Dervin, O. Haemmerle, R. Thomopoulos, "Fuzzy querying of incomplete, imprecise, and heterogeneously structured data in the relational model using ontologies and rules", *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 3, pp. 373-383, 2005.

[15] D.Y. Choi, "Enhancing the power of Web search engines by means of fuzzy query", *Decision Support Systems*, vol. 35, pp. 31-44, 2003.

[16] S. Fox, K. Karnawat, M. Mydland, S. Dumais, T. White, "Evaluating Implicit Measures to Improve Web Search", *ACM Transactions on Information Systems*, vol. 23, no. 2, pp. 147-168, 2005.

[17] C. Li, K. C. Chang, I. F. Ilyas, S. Song, "RankSQL: Query Algebra and Optimization for Relational Top-k Queries", in *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, 2005, pp. 131-142.

[18] H. Yu, S. Hwang, K. C. Chang, "Enabling soft queries for data retrieval", *Information Systems*, vol. 32, pp. 560-574, 2007.

[19] G.Q. Chen, J. Vandenbulcke, E.E. Kerre, "A general treatment of data redundancy in a fuzzy relational data model", *Journal of The American Society for Information Science*, vol. 43, no. 4, pp. 304-311, 1992.

[20] J.C. Cubero, M.A. Vila, "A new definition of fuzzy functional dependency in fuzzy relational databases", *Intelligent Systems*, vol. 9, no. 5, pp. 441-448, 1994.

[21] K.V.S.V.N. Raju, A.K. Majumdar, "Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems", *ACM Transactions on Database Systems*, vol. 13, no. 2, pp. 129-166, 1988.

[22] G.Q. Chen, *Fuzzy logic in data modeling: semantics, constraints, and database design*, Kluwer Academic Publishers, Boston, 1998.

[23] S. Tamura, S. Higuchi, K. Tanaka, "Pattern classification based on fuzzy relations", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 1, no. 1, pp. 61-66, 1971.

[24] H.S. Lee, "An optimal algorithm for computing the max-min transitive closure of a fuzzy similarity matrix", *Fuzzy Sets and Systems*, vol. 123, pp. 129-136, 2001.

[25] H.L. Larsen, R.R. Yager, "Efficient computation of transitive closures", *Fuzzy Sets and Systems*, vol. 38, no. 1, pp. 81-90, 1990.

[26] S. Shenoi, A. Melton, Fan, L.T. "An equivalence classes model of fuzzy relational databases", *Fuzzy Sets and Systems*, vol. 38, pp. 153-170, 1990.

[27] C.E. Shannon, "A Mathematical Theory of Communication", *The Bell System Technical Journal*, vol. 27, pp. 379-423, pp. 623-656, 1948.

[28] J. Han, M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann publishers, 2001.

[29] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley, 1991.

[30] Y.Z. Cao, M.S. Ying, G.Q. Chen, "Retraction and Generalized Extension of Computing With Words", *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 6, pp. 1238-1250, 2007.

[31] S. Medasani, R. Krishnapuram, and Y. Choi, "Graph matching by relaxation of fuzzy assignments", *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 1, pp. 173-182, 2001.

[32] A. Koeller, V. Keelara, "Approximate matching of textual domain attributes for information source integration", in *Proc. of the 2nd international workshop on Information quality in information systems*, 2005, pp. 77-86.

[33] M. Mitra, B.B. Chaudhuri, "Information Retrieval from Documents: A Survey", *Information Retrieval*, vol. 2, pp. 141-163, 2000.

[34] E. Rahm, P.A. Bernstein, "A survey of approaches to automatic schema matching", *The International Journal on Very Large Data Bases*, vol. 10, no. 4, pp. 334-350, 2001.

[35] H.A. Simon, "On the forms of mental representation" in *Minnesota Studies in the Philosophy of Science: Perception and cognition: issues in the foundations of psychology*, C. W. Savage Eds. University of Minnesota Press, 1978.

[36] J.H. Larkin, H.A. Simon, "Why a diagram is (sometimes) worth ten thousand words", *Cognitive Science*, vol. 11, pp. 65-99, 1987.

[37] K. Siau, "Informational and Computational Equivalence in Comparing Information Modeling Methods", *Journal of Database Management*, vol. 15, no. 1, pp. 73-86, 2004.

Robot Cooperation without Explicit Communication by Fuzzy Signatures and Decision Trees

Áron Ballagi[1] László T. Kóczy[2,3] Tamás D. Gedeon[4]

1 Department of Automation,
Széchenyi István University,
H-9026, Győr, Egyetem tér 1., Hungary

2 Inst. of Mechanical, Electrical Engineering and Information Technology,
Széchenyi István University,
H-9026, Győr, Egyetem tér 1., Hungary

3 Department of Telecommunications and Media Informatics,
Budapest University of Technology and Economics,
H-1117, Budapest, Magyar tudósok krt. 2., Hungary

4Department of Computer Science,
Australian National University,
Acton, ACT (Canberra), Australia

Email: ballagi@sze.hu, koczy@tmit.bme.hu, tom.gedeon@anu.edu.au

Abstract—*This paper presents a novel action selection method for multi robot task sharing problem. Two autonomous mobile robots try to cooperate for push a box to a goal position. Both robots equipped with object and goal sensing, but do not have explicit communication ability. We explore the use of fuzzy signatures and decision making system to intention guessing and efficient action selection. Virtual reality simulation is used to build and test our proposed algorithm.*

Keywords— Mobile robots, cooperation, fuzzy signatures, action selection, intention guessing.

1 Introduction

In this paper we focus on manipulation tasks with sufficiently challenging dynamics to require the careful cooperation of two or more robots. Sensing and actuation is noisy and uncertain in mobile robot domains, resulting in partial knowledge about the world. We explore the use of fuzzy signatures to efficient action selection and intention guessing in this environment. The intention guessing is the base of a meta-communication method between the robots. In this setup is not any explicit communication line.

We chose box-pushing as the problem domain because it has both theoretical interest and practical applications as it is an instance in large class of practical object manipulation tasks that appear to require tight feedback and control of real-world physics and dynamics [1-9]. From a theoretical standpoint, box-pushing is a variant on canonical object manipulation problem that draws on issue in fine motion as well as high level planning and control. Box-pushing is related to the well-known "piano-movers problem" [10], in that it requires the achievement of top level goal of delivering the box to a particular location, as well as the maintenance of low-level requirements including obstacle avoidance, maintaining contact with the box, and maintaining forward motion. From a practical point of view, box-pushing is a prototypical problem for studying various tasks requiring cooperation of number of smaller robots moving larger objects [6].

We use simulation in our experiments where two robots push a box to a goal position. Each robot has the own behaviour based control system thus they are fully autonomous. An action selection mechanism works in our behaviour based control in which the decision about selection is done by fuzzy signature based state describing algorithm.

2 Experimental task and environment

The actual stage of our research we use simulation of our real differential driven autonomous micro-robots (Fig. 1). The physical simulation is exact model of our robots in the case of scale, weight, mechanical systems and sensors.

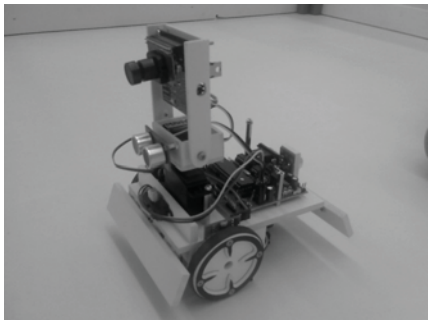


Figure 1. The real box-pushing robot

The two robots operate in a 2 x 2 m square arena. The box to be pushed is a 20 cm high and 40 cm wide and long. The

ISBN: 978-989-95079-6-8

goal region, located in one corner of the arena (Fig. 2), indicated by light sources detectable by the robot's sensor. In the experiments described here the goal location is fixed but it can be moved before and during experiments.

Each robot pushes the box with two "whiskers" which have a pair of force sensors. The left and right whiskers each provides an analog force signal which is combined to give information the relative position to the box and via the control loop keep it contact on both sides and thus perpendicular to the box.

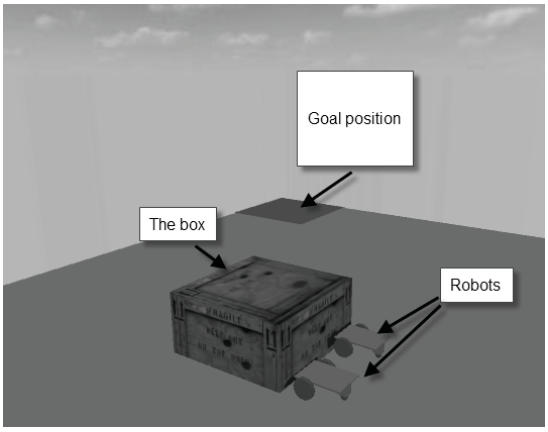


Figure 2. The simulation environment

In the description two direction sign systems are used, the absolute direction with letters N, E, S, W as in the usual sense for North, East, South and West. The second direction sign system is a box relative system where the sides of the box are N_B, E_B, S_B and W_B respectively (Fig. 3). The position of the objects (boxes and robots in this case) always can be described by the absolute course, latitude and longitude of the object. One object relative position to a box is described by the box relative system, i.e. which side of the box is touched by that object. For simplicity we assume that the sides of the box are always parallel with the N-S and E-W axes, so there is not necessary any rotation.

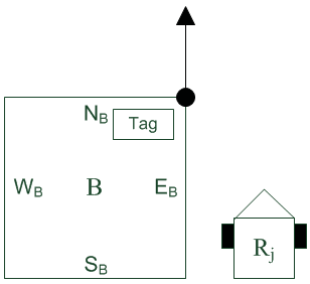


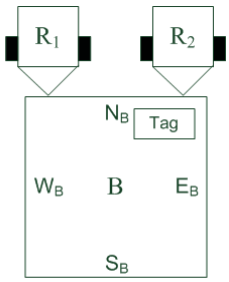
Figure 3. Symbols of boxes and robots

There are just a few essentially different robot positions allowed. Because two robots are needed for pushing the box, at each side of the boxes, two spaces are available for the robots manipulating them: the "counterclockwise position" and the "clockwise position" (see Fig. 4). The position is described by $P_r = [S,T]$ where r is the number of the robot, S is the side of the box where the robot touch it (N_B, E_B, S_B and W_B respectively) and the T is the turning position that means "counterclockwise position" or "clockwise position" (CC or CW).

The cooperating combination of robots is denoted by $C_{i,j}$ where i,j is the number of the robots. $C_{1,2} = P$ is the "pushing or shifting combination", when two robots (R_1 and R_2) are side by side at the same side of the box as Fig. 5 shows. In this case R_1 and R_2 are in the relative North (N_B) position. Of course, all the other three directions are similarly allowed. Any other combination of two robots is illegal, except see the next paragraph ("stopping combination").

$P_j = [N_b, CC]$ $\quad$ $P_j = [N_b, CW]$

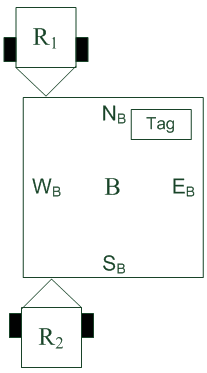
Figure 4. Robot positions at the N_B side of the box



$$\left.\begin{aligned} P_1 &= [N_B, CC] \\ P_2 &= [N_B, CW] \end{aligned}\right\} C_{1,2} = P$$

Figure 5. Allowed combinations of two robots for moving the box

Eventually, in Fig. 6, the combinations are shown where one robot intends to do a move operation, and another robot that has recognized the goal box configuration positions itself to prevent a certain move. This is an exception where a two robot combination other than the ones listed in Fig. 5 is legal as a temporary combination, clearly signalizing "stop this attempt as it is in contrary to the goal ".



$$\left.\begin{aligned} P_1 &= [N_B, CC] \\ P_2 &= [S_B, CW] \end{aligned}\right\} C_{1,2} = ST$$

Figure 6. The stopping combination

ISBN: 978-989-95079-6-8

3 Build up the codebook

After having overviewed the possible positions and combinations, let us build the part of the codebook that enables robots to recognize a situation and take action accordingly. In initial position both robots can see the goal light and know the actual position of the box. There are four good position and two pushing combination that the robots can take up.

How can take the starting position? The possible initial actions are:

1. The both robots move to reach the nearest possible good position.
2. Which reach the position first wins the temporal leader role if the other knows or guesses that this position is well for forming a useful combination. In this case the second robot turns and moves to the free position for this combination as the Fig. 7 shows. If the second robot rejects this combination then they try to take up a new form.
3. After the forming this starting combination the next is the push or shift task. We assume there are only two shifting axis, the N-S and E-S. The robots push the box a given force.

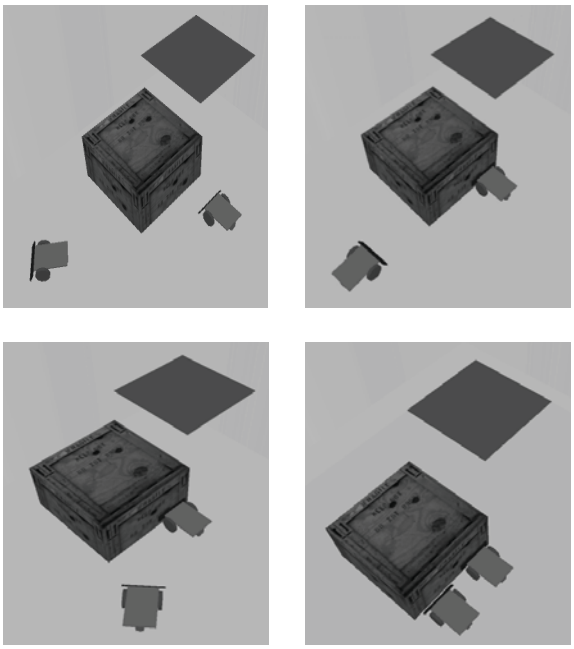


Figure 7. The robots take the starting combination

From this point a lot of scenarios are possible. Let us see some examples.

The R2 robot loses the goal sign, so it does not know the direction exactly. The R2 will slow down. The R1 robot detects the rotation of the box via its force sensors and guesses the R2 is slacken. The R1 has to make a decision, in other words it selects an action or reaction. There are two ways:

1. The R1 knows the right direction and forces to shift toward. It speeds up slightly, the R2 senses this and has to decide its reaction. The R2 makes a second self-examination and if it has not any other problem but the lost of goal sign then it switches to the *blind push* action or behavior. It means R2 pushes the box lean on R1 as follower. If R2 finds any other reason of slacken, e.g.: external obstacle or any internal error (e.g.: low battery), it keeps its own speed or decreases it. In this case the R1 gives a new reaction and so on, as long as they reach a right deal or give up the task.
2. The R1 does not sure the right direction, slow down to the speed of R2 and both robots search the goal light. If one of they find the goal takes the leader role and forces the other to push the box. If not, after a given time they stop pushing the box and move to take a new combination. After the some new unsuccessful attempt they give up.

Based on the above example and considerations it is possible to build up some elements of the action selection algorithm as a codebook. It will take the form of a decision tree, where the inputs are the direct observation, the first level outputs are intention guesses and the second level outputs the concrete actions of the corresponding robot.

Now let us see a relevant part of the codebook contains a decision tree with fuzzy elements. It is a part of the above presented example.

In R1 control system:
- Does R2 {slow down}?
  - If no then *No Action*
  - else
  - Do I know the {goal position}?
    - If no then *Goal Searching Action*
    - else
    - *Force Move Action*
  - Does R2 {accelerate}?
    - If no then *Slow Down Action*
    - else
    - *No Action*

In R2 control system:
- Can I see the {goal light}?
  - If no then *Slow Down Action*
  - .
  - .
  - .
- Does R1 {force the move}?
  - If no then *No Action* (or *Goal Searching Action*)
  - else
  - Do I have any internal or external {obstacle}?
- If no then *Blind Push Action*
- else
- *Slow Down Action* (or *Keep Move Action*)

Note that in the condition parts there are fuzzy notions which are between curly brackets. It is usually hard to judge that the R2 slows down really or the box hits an obstacle as the information come from some simple sensors that might provide only approximate results.

This simple example illustrates clearly that the meta-communication among intelligent robots by intention guessing and fuzzy evaluation of the situation might lead to effective cooperation and the achievement of task that cannot be done without collaboration and communication.

ISBN: 978-989-95079-6-8

4 Fuzzy signatures and decisions

4.1 Fuzzy signatures

In 1999 Vámos, *et al.* introduced the concept of Fuzzy Signatures [11]. Some further advanced versions of the concept and it possible use for describing complex data were later proposed in [12,13,14].

The original definition of fuzzy sets was $A : X \to [0,1]$, and was soon extended to *L-fuzzy sets* by Goguen [15]

$$A_s : X \to [a_i]_{i=1}^{k}, a_i = \begin{cases} [0,1] \\ [a_{ij}]_{j=1}^{k_i} \end{cases}, a_{ij} = \begin{cases} [0,1] \\ [a_{ijl}]_{l=1}^{k_{ij}} \end{cases} \tag{1}$$

$A_L : X \to L$, L being an arbitrary algebraic lattice. A practical special case, *Vector Valued Fuzzy Sets* was introduced by Kóczy [16], where $A_{V\,k} : X \to [0,1]^k$, and the range of membership values was the lattice of k-dimensional vectors with components in the unit interval. A further generalization of this concept is the introduction of fuzzy signature and signature sets, where each vector component is possibly another nested vector (right).

Fuzzy signature can be considered as special multidimensional fuzzy data. Some of the dimensions are interrelated in the sense that they form sub-groups of variables, which jointly determine some feature on higher level. Let us consider an example. Fig. 8 shows a fuzzy signature structure.

The fuzzy signature structure shown in Fig. 8 can be represented in vector form as follow:

$$x = \begin{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} \\ \begin{bmatrix} x_{21} \\ \begin{bmatrix} x_{221} \\ x_{222} \\ x_{223} \end{bmatrix} \\ x_{23} \end{bmatrix} \\ \begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix} \end{bmatrix}^T \tag{2}$$

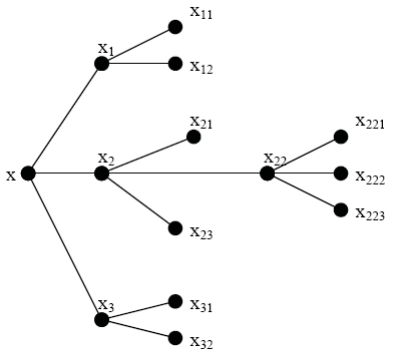


Figure 8. A Fuzzy Signature Structure

Here $[x_{11}\ x_{12}]$ from a sub-group that corresponds to a higher level compound variable of x_1. $[x_{221}\ x_{222}\ x_{223}]$ will then combine together to form x_{22} and $[x_{21}\ [x_{221}\ x_{222}\ x_{223}]\ x_{23}]$ is equivalent on higher level with $[x_{21}\ x_{22}\ x_{23}] = x_2$. Finally, the fuzzy signature structure will become $x = [x_{221}\ x_{222}\ x_{223}]$ in the example.

The relationship between higher and lower level is govern by the set of fuzzy aggregations. The results of the parent signature at each level are computed from their branches with appropriate aggregation of their child signature. Let a_1 be the aggregating associating x_{11} and x_{12} used to derive x_1, thus $x_1 = x_{11}a_1x_{12}$. By referring to Fig. 8, the aggregations for the whole signature structure would be a_1, a_2, a_{22} and a_3. The aggregations a_1, a_2, a_{22} and a_3 are not necessarily identical or different. The simplest case for a_{22} might be the min operation, the most well known t-norm. Let all aggregation be min except a_{22} be the averaging aggregation. We will show the operation based on the following fuzzy signature values for the structure in the example.

Each of these signatures contains information relevant to the particular data point x_0; by going higher in the signature structure, less information will be kept. In some operations it is necessary to reduce and aggregate information obtained from another source (some detail variables missing or simply being locally omitted). Such is when interpolation within a fuzzy signature rule base is done, where the fuzzy signature flanking an observation are not exactly of the same structure. In this case the maximal common sub-tree must be determined and all signatures must be reduced to that level in order to be able to interpolate between the corresponding branches or roots in some cases [17].

$$x = \begin{bmatrix} \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix} \\ \begin{bmatrix} 0.2 \\ \begin{bmatrix} 0.6 \\ 0.8 \\ 0.1 \end{bmatrix} \\ 0.9 \end{bmatrix} \\ \begin{bmatrix} 0.1 \\ 0.7 \end{bmatrix} \end{bmatrix}^T \tag{3}$$

After the aggregation operation is perform to the lowest branch of the structure, it will be described on higher level as:

$$x = \begin{bmatrix} 0.3 \\ \begin{bmatrix} 0.2 \\ 0.5 \\ 0.9 \end{bmatrix} \\ 0.1 \end{bmatrix}^T \tag{4}$$

Finally, the fuzzy signature structure will be:

ISBN: 978-989-95079-6-8

$$x = \begin{bmatrix} 0.3 \\ 0.2 \\ 0.1 \end{bmatrix}^T \quad (5)$$

4.2 Fuzzy signatures for box-pushing robot cooperation

Now, let us construct the fuzzy signatures for robot cooperation. The fuzzy signatures presented here are implemented in R1 robot action selection system as a part of the meta-communication codebook. Of course here is described only a slice of the whole signatures.

Fig. 9 presents the signatures which describe R2 behaviour. The information communicated to R1 is partly the observation about the last move of R2 and partly the evaluation of the situation with the box, according to those membership degrees will be attached to each leave of the actual signature.

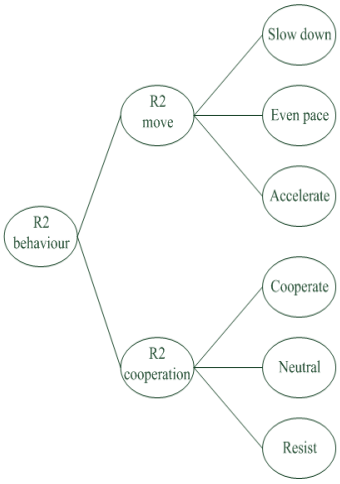


Figure 9. R2 robot's behavior fuzzy signatures

Figure 11 presents a portion of decision tree applied by R1 when it observes some action or reaction of R2 described by above mentioned behavior fuzzy signatures.

After all action selection the R1 control system reevaluates the values of the membership degrees of observed R2 behavior, environmental properties and the own state as the lines partially show in Fig. 11. After the evaluation the robot makes a decision which action is taken.

Parallel, in the R2 control system runs a similar task which reacts to the R1 decision and action. Thus, the two robots produce a circle of action and reaction (Fig. 10), where an initial action triggers the circulation of reactions and composes a meta-communication between the robots.

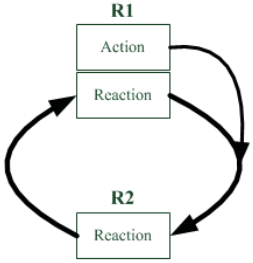
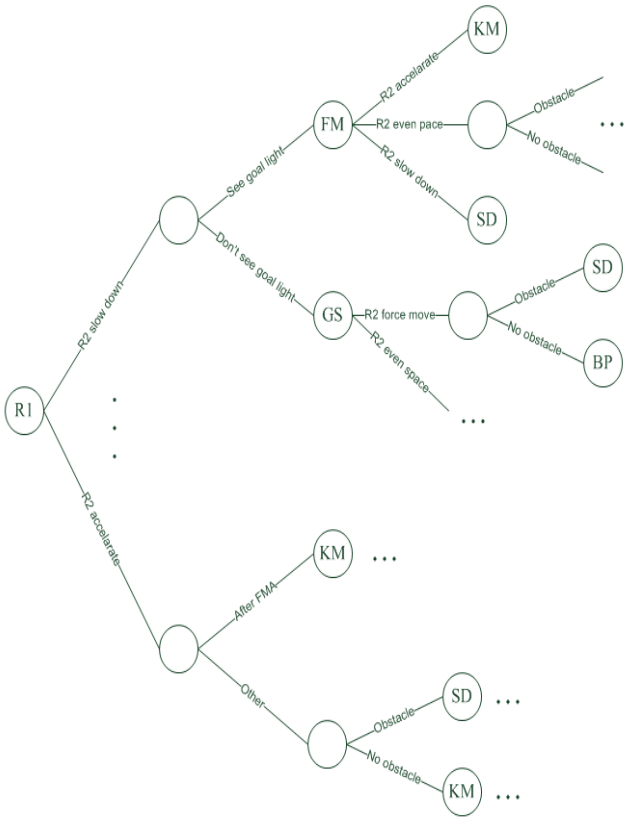


Figure 10. Action – reaction circle

This is a type of context dependent or fuzzy communication [18, 19], which means every robot has the own codebook and communicate in a vague, compressed or quasi channel. Then the robots build up and interpret this imprecise information with their codebooks.



Where the actions are

- BP – Blind push
- FM – Force move
- GS – Goal Searching
- KM – Keep move
- SD – Slow down

Figure 11. The decision tree of R1 robot

5 Conclusions

Fuzzy communication contains vague or imprecise components and it might lack abundant information. If two robots are communicating by a fuzzy channel, it is necessary that both ends possess an identical part within the codebook. The codebook might partly consist of common knowledge but it usually requires a context dependent part that is learned by communicating. Possibly it is continuously adapting to the input information. If such a codebook is not available or it contains too imprecise information, the information to be transmitted might be too much distorted and might lead to misunderstanding, misinterpretation and serious damage. If however the quality of the available codebook is satisfactory, the communication will be efficient i.e. the original contents of the message can be reconstructed. At the same time it is cost effective, as fuzzy communication is compressed as compared to traditional communication. This advantage can be deployed in many areas of engineering, especially where the use of the communication channels is expensive in some sense, or where there is no proper communication channel available at all.

ISBN: 978-989-95079-6-8

Here we illustrate clearly that the communication among intelligent robots by intention guessing and fuzzy evaluation of the situation might lead to effective cooperation and the achievement of tasks that cannot be done without collaboration and communication.
We simulated many scenarios and almost got acceptable results, but sometimes the robots made deadlock combination and gave up the work. In future we want to work out new algorithms to solve these situations.

Acknowledgment

The research was supported by HUNOROB project (HU0045), a grant from Iceland, Liechtenstein and Norway through the EEA Financial Mechanism and the Hungarian National Development Agency, the Széchenyi University Main Research Direction Grant 2009, and a National Scientific Research Fund Grant OTKA T048832 and K75711.

References

[1] P. Caloud, W. Choi, J. C. Latombe, C. L. Pape, and M. Yim, Indoor automation with many mobile robots, *in IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 67-72, 1990.

[2] B. Donald, J. Jennings, and D. Rus, Analyzing teams of cooperating mobile robots, *in IEEE International Conference on Robotics and Automation,* pp. 1896-1903, 1994.

[3] C. R. Kube and H. Zhang, Collective robotic intelligence, *in Second International Confernece on Simulationof Adaptive Behavior: From Animals to Animats*, pp. 460-468, MIT Press, (Cambridge, MA), 1992.

[4] C. R. Kube and H. Zhang, Task modelling in collective robotics, *Autonomous Robots* 4(1), pp. 53-72, 1997.

[5] C. R. Kube and E. Bonabeau, Cooperative transport by ants and robots, *Tech. Rep.* 99{01-008, Santa Fe Institute, Santa Fe, NM, 1999.

[6] M. J. Matari_c, M. Nilsson, and K. T. Simsarian, Cooperative multi-robot box-pushing, *in IEEE International Conference on Intelligent Robots and Systems*, vol. 3, pp. 556-561, 1995.

[7] F. R. Noreils, An architecture for cooperative and autonomous mobile robots, *in IEEE International Conference on Robotics and Automation*, pp. 2703-2710, 1992.

[8] L. E. Parker, ALLIANCE: An architecture for fault tolerant, cooperative control of heterogeneous mobile robots, *in IEEE/RSJ/GI International Conference on Intelligent Robots and Systems*, pp. 776-783, 1994.

[9] D. J. Stilwell and J. S. Bay, Toward the development of a material transport system using swarms of ant-like robots, *in IEEE International Conference on Robotics and Automation*, pp. 766-771, 1993.

[10] J. Schwartz and M. Sharir, On the Piano Mover's problem. II. General techniques for computing topological properties of real algebraic manifolds', *Advances in Applied Math*, 1983.

[11] T. Vámos, Gy. Bíró and L. T. Kóczy, Fuzzy signatures, *in SIC-EUROFUSE'99,* Budapest, pp. 210-217, 1999.

[12] B. S. U. Mendis, T. D. Gedeon and L. T. Kóczy, Investigation of Aggregation in Fuzzy Signatures, *in 3rd International Conference on Computational Intelligence, Robotics and Autonomous Systems*, (CD proc), Singapore, 2005.

[13] K. W. Wong, T. D. Gedeon and L. T. Kóczy, Construction of fuzzy signature from data, *in FUZZ-IEEE 2004*, Budapest, pp. 1649-1654, 2004.

[14] K. W. Wong, A Chong, T. D. Gedeon, L. T. Kóczy, and T. Vámos, Hierarchical fuzzy signature structure for complex structured data, *in International Symposium on Computational Intelligence and Intelligent Informatics (ISCIII),* Nabeul, pp. 105-109, 2003.

[15] J. A. Goguen, L-fuzzy Sets, *J. Math. Anal. Appl.* 18 pp. 145-174, 1967.

[16] L. T. Kóczy, Vectorial I-fuzzy Sets, *in M. M. Gupta & E. Sanchez (eds.): Approximate Reasoning in Decision Analysis*, North Holland, Amsterdam, pp. 151-156, 1982.

[17] T. Vámos, L.T. Kóczy and Gy. Biró, Fuzzy Signatures in Data Mining, *in IFSA World Congress and 20th NAFIPS International Conference*, Vancouver, pp. 2842-2846, 2001.

[18] L.T. Kóczy and T.D. Gedeon, Context Dependent Reconstructive Communication, *3rd International Symposium on Computational Intelligence and Intelligent Informatics*, ISCIII 2007, Agadir, Morocco, March 28-30, pp. 13-19, 2007.

[19] A. Ballagi, L.T. Kóczy and T.D. Gedeon, Local Codebook Construction for Fuzzy Communication in Cooperation of Mobile Robots, *in First Győr Symposium on Computational Intelligence*, Győr, 2008.

ISBN: 978-989-95079-6-8

A Fuzzy Filter for the Removal of Gaussian Noise in Colour Image Sequences

Tom Mélange, Mike Nachtegael, Etienne E. Kerre

Ghent University, Dept. of Applied Math. and Computer Science
Fuzziness and Uncertainty Modelling Research Group
Krijgslaan 281-S9, 9000 Ghent, Belgium
Email: {tom.melange,mike.nachtegael,etienne.kerre}@ugent.be

Abstract— *In this paper a new fuzzy filter for the removal of additive Gaussian noise in colour image sequences is presented. The proposed method, which is a colour extension of our previous work on greyscale images, consists of two subfilters. The first subfilter averages the noise in each of the colour bands separately. In this averaging, the weights of the neighbouring pixels are determined by fuzzy rules. Even though the filtering is performed in each colour band separately, information from the other colour bands is also used in the fuzzy rules. In this way, we can be more confident whether a neighbouring pixel is similar to the pixel being filtered. In order to further improve the results, a second subfilter is applied too. This colour restorating filter is based on the simplified assumption that for similar pixels the pixel value difference should be approximately the same in all colour bands. A pixel is estimated from its neighbours by estimating the differences in each colour band equal to the average over all colour bands. Experimental results show that, in terms of average PSNR and average NCD, the proposed colour extension of our previous work outperforms the usual colour extension of a greyscale filter, in which the Y component in the YUV transform is filtered with the original greyscale method and the chrominance bands U and V are averaged.*

Keywords— Video, noise filter, colour, Gaussian noise, fuzzy rule.

1 Introduction

Images sequences are among the most important information carriers in today's world. They are used in numerous applications such as broadcasting, video-phone, traffic observations, surveillance systems and autonomous navigation to name a few. The used sequences are however often affected by noise due to bad acquisition, transmission or recording. In this paper we will concentrate on image sequences corrupted with additive white Gaussian noise of zero mean and variance σ^2:

$$I_{n,i} = I_{o,i} + \epsilon_i, \qquad i = 1, \ldots, P \tag{1}$$

where $I_{n,i}$ and $I_{o,i}$ denote the i-th pixel from the noisy and the original frame respectively, $\epsilon_i \sim N(0, \sigma^2)$ and P is the number of pixels per frame.

The goal of noise filtering is not only visual improvement but also an improvement of the further analysis or coding of the sequences. During this filtering process a compromise needs to found between the removal of noise and the preservation of fine image details.

Most video filters that exist in literature are designed for greyscale sequences. Some examples are e.g. [2, 3, 4, 5]. These greyscale methods can nevertheless be extended to the RGB colour space [1] in a straightforward way by filtering each of the colour bands R, G and B separately. This might however result in the introduction of colour artefacts since the correlation between the different colour bands is neglected. Therefore, the commonly used alternative is to filter only the luminance component Y of the YUV-transform with the given greyscale method, possibly with an additional averaging of the chrominance components U and V.

The proposed filtering framework can be seen as a colour extension of our previous work presented in [4], which was inspired by the multiple class averaging filter from [5]. The presented filter consists of two separate subfilters. In the first subfilter, we add colour information to the fuzzy logic framework from our work in [4]. In this first fuzzy subfilter, each of the colour bands is denoised separately by averaging the noise. The weights assigned to the pixels considered in the averaging are determined by fuzzy rules. Even though the filtering is performed in each colour band separately, the fuzzy rules also require information from the other colour bands. Due to this increase in information, we can expect a more reliable estimation of the degree to which a neighbouring pixel is similar to the pixel that is filtered. However, especially around edges in the image, some colour artefacts might appear because sometimes not enough similar neighbours can be found to completely average the noise and it might also happen that a neighbouring pixel is wrongly considered as belonging to the same object (similar). To cope with this problem, the first subfilter is combined with an additional second subfilter, which is an extension of the second subfilter in [6]. This subfilter is based on the simplified assumption that for similar pixels the pixel value differences in the three different colour components should all three be approximately the same. The pixel being filtered is estimated from a neighbour by estimating the differences in each band equal to the average over the different colour bands.

The experimental results show that the proposed colour video denoising framework performs better than the YUV-colour extension in terms of average PSNR and NCD.

The paper is structured as follows: Section 2 gives some preliminaries about fuzzy set theory. Next, the proposed colour video denoising algorithm is explained in Section 3. Finally, experimental results are presented in Section 4.

2 Preliminaries

In this section, we will introduce some basic notions concerning fuzzy set theory and fuzzy if-then rules.

2.1 Fuzzy Sets

In classical set theory, an element x in a universe X belongs or does not belong to a certain set C, defined over the uni-

ISBN: 978-989-95079-6-8

verse X. This can be modelled by membership degrees belonging to $\{0,1\}$. In fuzzy set theory [7], there is a more gradual transition between 'belonging to' and 'not belonging to'. A fuzzy set F in the universe X is characterized by a $X \rightarrow [0,1]$ mapping μ_F, which assigns a degree of membership $\mu_F(x) \in [0,1]$ of x in the fuzzy set F, with every element x in X. This membership degree may also lie between 0 and 1, which makes fuzzy sets very useful and more natural than crisp sets to work with complex systems and human knowledge, where linguistic variables are used. In this paper, we will use a linguistic variable "large" for several parameters that will be introduced further. Using fuzzy set theory, a parameter is not necessarily large or not large, but can also be large to some degree.

2.2 Fuzzy Rules and Fuzzy Logical Operators

The general form of a fuzzy rule is "IF A THEN B", where the premise A and the consequent B are (collections of) propositions containing linguistic variables. These propositions can contain AND, OR and NOT operators, which correspond to respectively the intersection, the union and the complement of fuzzy sets.

The membership degree of an element x in the intersection (respectively union) of two fuzzy sets A and B in X is obtained through a mapping T, (respectively S) that maps the membership degrees of an element in the fuzzy sets A and B onto its membership degree in the fuzzy set $A \cap B$ (respectively $A \cup B$): $\mu_{(A \cap B)}(x) = T(\mu_A(x), \mu_B(x))$, $\forall x \in X$ (respectively $\mu_{(A \cup B)}(x) = S(\mu_A(x), \mu_B(x))$,$\forall x \in X$). In fuzzy logic for the mappings T and S, respectively a triangular norm [8] and a triangular conorm [8] are used. For the results in this paper, we have used the algebraic product and the probabilistic sum as triangular norm and conorm respectively. This norm and conorm led to the best results for our filter and are the most simple from a computational point of view.

To specify the complement of a fuzzy set A in X, we use a mapping N to derive the membership degree of an element x in the complement of the fuzzy set A from its membership degree in the fuzzy set A: $\mu_{coA}(x) = N(\mu_A(x))$, $\forall x \in X$. In fuzzy logic for the mapping N an involutive negator [8] is used. In this paper we have chosen for the well-known standard negator $N_s(x) = 1 - x, \forall x \in [0,1]$.

Let's take for example the following fuzzy rule:

Fuzzy Rule 1. `IF` (u *is* `U` `AND` v *is* V) `OR` w *is* `NOT` `W` `THEN` z *is* `Z`.

The membership degree $\mu_Z(z)$ of z in Z, which corresponds to the activation degree of the rule, i.e. the degree of thruthfulness, is then calculated as:

$$\mu_Z(z) = (\mu_U(u) \cdot \mu_V(v)) + (1 - \mu_W(w)) - (\mu_U(u) \cdot \mu_V(v)) \cdot (1 - \mu_W(w)). \quad (2)$$

3 The proposed algorithm

In this section, we will outline the proposed filtering framework. The method is a superposition of two subfilters, presented in respectively Subsection 3.1 and Subsection 3.2.

In the first subfilter a $3 \times 3 \times 2$ sliding window is used, which is moved through the frame from top left to bottom right, each time filtering the central position in the window. This window consists of 3×3 pixels in the current frame and 3×3 pixels in the previous frame as shown in Fig. 1. The central position in the window is denoted by $(\mathbf{r}, t)$, where $\mathbf{r} = (x, y)$ and t respectively stand for the spatial and temporal position in the image sequence. An arbitrary pixel position in the sliding window (which may also be the central position) is denoted by $(\mathbf{r'}, t')$, with $\mathbf{r'} = (x + k, y + l)$, $(-1 \leq k, l \leq 1)$ and $t' = t$ or $t' = t - 1$. Further, the second subfilter uses a 3×3 window in the current frame for which similar notations will be used as for the $3 \times 3 \times 2$ window.

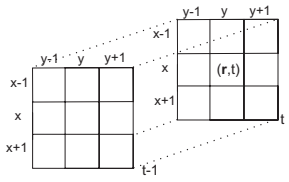


Figure 1: The $3 \times 3 \times 2$ filtering window consisting of 3×3 pixels in the current frame and 3×3 pixels in the previous frame.

3.1 First Subfilter

The subfilter explained in this subsection, can be seen as a colour extension of the filtering framework of our previous work, presented in [4]. This method was inspired by the multiple class averaging filter [5], from which we adopted the following ideas: (i) temporal blur should be avoided by taking into account only pixels from the current frame when motion is detected; (ii) image details should be preserved by filtering less when large spatial activity (details) is detected. More noise will be left, but the eye is less sensitive for the high spatial frequencies to which large spatial activity corresponds [9]. In homogeneous areas on the other hand, as much noise as possible should be removed by strong filtering.

The filtering is based on averaging the noise using the pixel values in the neighbourhood that are similar to the given pixel value and probably belong to the same object. Each colour band is filtered separately, but in the filtering of each colour band, the information from the other colour bands is used to confirm that a neighbouring pixel does indeed belong to the same object.

In the following the output of the first fuzzy subfilter is denoted by I_f, while the noisy input sequence is denoted by I_n. The output of the first subfilter for the central pixel in the window is determined as a weighted mean of the pixel values in the $3 \times 3 \times 2$ window ($i = 1, 2, 3$):

$$I_f(\mathbf{r}, t, i) = \frac{\sum_{\mathbf{r'}} \sum_{t'=t-1}^{t} W(\mathbf{r'}, t', \mathbf{r}, t, i) I_n(\mathbf{r'}, t', i)}{\sum_{\mathbf{r'}} \sum_{t'=t-1}^{t} W(\mathbf{r'}, t', \mathbf{r}, t, i)}, \quad (3)$$

The weights $W(\mathbf{r'}, t', \mathbf{r}, t, i)$ in the above weighted means are determined using fuzzy logic. The weight of a pixel is chosen equal to its membership degree in the fuzzy set "large weight". This membership degree is calculated as the activation degree of a fuzzy rule that corresponds to the ideas given at the beginning of this subsection. If large detail is detected, i.e., if a calculated detail value is large, then we should filter less by averaging only over pixels that are quite similar, i.e.,

ISBN: 978-989-95079-6-8

for which there is no large difference in the considered colour component and also in at least one of the other components. On the other hand, if there is not much detail detected, i.e., in the case that the calculated detail value is not large, strong filtering should be performed, i.e., we don't put a condition on the difference between the considered and the filtered pixel in the considered colour band (the check in the other colour bands remains for the case that the calculated detail value was not completely reliable). Further, if the pixel for which the weight is calculated belongs to the previous frame, we only want to give it a large weight if there is no motion detected in the filtering window, i.e., if a calculated motion value is not large. In the filtering of the red colour band, this results in the following two fuzzy rules, depending on whether the pixel lies in the current or the previous frame:

Fuzzy Rule 2. *Assigning the membership degree in the fuzzy set "large weight" of the weight $W(\mathbf{r'}, t', \mathbf{r}, t, 1)$ for the red value at position $\mathbf{r'}$ in the current frame ($t' = t$) of the window with central pixel position $(\mathbf{r}, t)$:*

IF $\Big($*(the detail value $d(\mathbf{r},t)$ is* LARGE AND $\Delta_1(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE AND $(\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE OR $\Delta_3(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE$)\Big)$

OR $\Big($*(the detail value $d(\mathbf{r},t)$ is* NOT LARGE$)$ AND $(\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE OR $\Delta_3(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE$)\Big)\Big)$

THEN *the red value at position $\mathbf{r'}$ has a* LARGE *weight $W(\mathbf{r'}, t', \mathbf{r}, t, 1)$ in (3).*

Fuzzy Rule 3. *Assigning the membership degree in the fuzzy set "large weight" of the weight $W(\mathbf{r'}, t', \mathbf{r}, t, 1)$ for the red value at position $\mathbf{r'}$ in the previous frame ($t' = t - 1$) of the window with central pixel position $(\mathbf{r}, t)$:*

IF $\Big($*(the detail value $d(\mathbf{r},t)$ is* LARGE AND $\Delta_1(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE AND $(\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE OR $\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE$)\Big)$

OR $\Big($*(the detail value $d(\mathbf{r},t)$ is* NOT LARGE$)$ AND $(\Delta_3(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE OR $\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ *is* NOT LARGE$)\Big)\Big)$

AND *the motion value $m(\mathbf{r},t)$ is* NOT LARGE

THEN *the red value at position $\mathbf{r'}$ has a* LARGE *weight $W(\mathbf{r'}, t', \mathbf{r}, t, 1)$ in (3).*

Similar fuzzy rules, switching the role of the red colour band and the colour band that needs to be filtered, are used to determine the weights $W(\mathbf{r'}, t', \mathbf{r}, t, 2)$ and $W(\mathbf{r'}, t', \mathbf{r}, t, 3)$ in expression (3) to filter the green and blue colour band respectively. In these fuzzy rules, a detail value $d(\mathbf{r}, t)$, three difference values $\Delta_1(\mathbf{r'}, t', \mathbf{r}, t)$, $\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ and $\Delta_3(\mathbf{r'}, t', \mathbf{r}, t)$ (one for each colour band) and a motion value $m(\mathbf{r}, t)$ are used, which we will now discuss. In our proposed method, only one detail value $d(\mathbf{r}, t)$ is used for all three colour bands. This detail value depends however on three detail values computed on each colour band separately. These three detail values are equal to the standard deviation calculated in the respective colour bands on the 3×3 pixels of the filtering window belonging to the current frame ($i = 1, 2, 3$):

$$I_{av}(\mathbf{r}, t, i) = \frac{1}{9} \sum_{\mathbf{r'}} I_n(\mathbf{r'}, t, i),$$

$$d_i(\mathbf{r}, t) = \Big(\frac{1}{9} \sum_{\mathbf{r'}} \big(I_n(\mathbf{r'}, t, i) - I_{av}(\mathbf{r}, t, i) \big)^2 \Big)^{\frac{1}{2}}.$$

We don't need to know the exact value of $d(\mathbf{r}, t)$. Only the membership degree $\mu_d\big(d(\mathbf{r}, t)\big)$ of $d(\mathbf{r}, t)$ in the fuzzy set "large detail value" is needed to calculate the activation degree of Fuzzy Rules 2 and 3 that determine the weights in the expression (3). This membership degree corresponds to the activation degree of the following fuzzy rule:

Fuzzy Rule 4. *Assigning the membership degree in the fuzzy set "large detail value" of the detail value $d(\mathbf{r}, t)$ for the pixel at the central position $(\mathbf{r}, t)$ in the filtering window of the current step:*

IF $d_1(\mathbf{r}, t)$ *is* LARGE AND $d_2(\mathbf{r}, t)$ *is* LARGE AND $d_3(\mathbf{r}, t)$ *is* LARGE

THEN $d(\mathbf{r}, t)$ *is* LARGE.

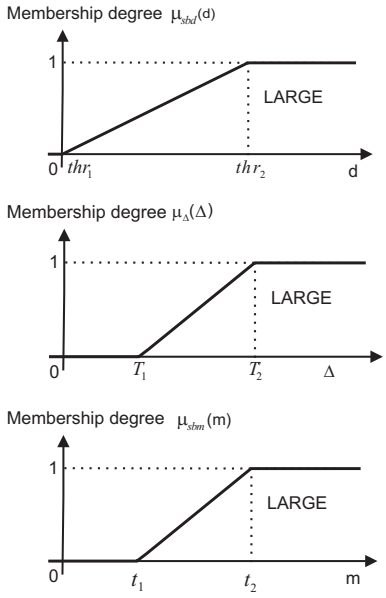


Figure 2: The respective membership functions μ_{sbd}, μ_Δ and μ_{sbm} of the respective fuzzy sets "large single band detail value", "large difference" and "large single band motion value".

The membership function μ_{sbd} of the fuzzy set "large single band detail value" is given in Fig. 2. The membership degree $\mu_d\big(d(\mathbf{r}, t)\big)$ is thus given by:

$$\mu_d\big(d(\mathbf{r}, t)\big) = \mu_{sbd}\big(d_1(\mathbf{r}, t)\big) \cdot \mu_{sbd}\big(d_2(\mathbf{r}, t)\big) \cdot \mu_{sbd}\big(d_3(\mathbf{r}, t)\big). \quad (4)$$

The three difference values $\Delta_1(\mathbf{r'}, t', \mathbf{r}, t)$, $\Delta_2(\mathbf{r'}, t', \mathbf{r}, t)$ and $\Delta_3(\mathbf{r'}, t', \mathbf{r}, t)$ that are used to determine the weights in (3) are given by ($i = 1, 2, 3$):

$$\Delta_i(\mathbf{r'}, t', \mathbf{r}, t) = |I_n(\mathbf{r'}, t', i) - I_n(\mathbf{r}, t, i)|. \quad (5)$$

The membership function μ_Δ of the fuzzy set 'large difference' is given in Fig.2.

ISBN: 978-989-95079-6-8

Analogously to the detail value $d(\mathbf{r},t)$ also only one motion value $m(\mathbf{r},t)$ is used for the filtering of all three colour bands. This value depends however again on three values computed for each of the colour bands separately. These three single band motion values are ($i=1,2,3$):

$$\begin{aligned} m_i(\mathbf{r},t) &= \Big|I_{av}(\mathbf{r},t,i) - I_{av}(\mathbf{r},t-1,i)\Big| \\ &= \Big|\frac{1}{9}\sum_{\mathbf{r'}} I_n(\mathbf{r'},t,i) - \frac{1}{9}\sum_{\mathbf{r'}} I_v(\mathbf{r'},t-1,i)\Big|. \end{aligned}$$

Just as it was the case for the detail value $d(\mathbf{r},t)$, we don't need to know the exact value of $m(\mathbf{r},t)$ either. Only its membership degree $\mu_m\big(m(\mathbf{r},t)\big)$ in the fuzzy set "large motion value" is needed for the Fuzzy Rules 2 and 3 (or thus the calculation of the weights in (3)). This membership degree is obtained from the following fuzzy rule:

Fuzzy Rule 5. *Assigning the membership degree in the fuzzy set "large motion value" of the motion value $m(\boldsymbol{r},t)$ for the pixel at the central position $(\boldsymbol{r},t)$ in the filtering window of the current step:*

IF ($m_1(\boldsymbol{r},t)$ *is* LARGE AND $m_2(\boldsymbol{r},t)$ *is* LARGE) OR
($m_1(\boldsymbol{r},t)$ *is* LARGE AND $m_3(\boldsymbol{r},t)$ *is* LARGE) OR
($m_2(\boldsymbol{r},t)$ *is* LARGE AND $m_3(\boldsymbol{r},t)$ *is* LARGE)
THEN $m(\boldsymbol{r},t)$ *is* LARGE.

The membership function μ_{sbm} of the fuzzy set "large single band motion value" is given in Fig 2. The membership degree $\mu_m\big(m(\mathbf{r},t)\big)$ is thus given by:

$$\mu_m\big(m(\mathbf{r},t)\big) = \alpha + (\beta+\gamma-\beta\cdot\gamma) - \alpha\cdot(\beta+\gamma-\beta\cdot\gamma), \quad (6)$$

with

$$\begin{aligned} \alpha &= \Big(\mu_{sbm}\big(m_1(\mathbf{r},t)\big)\cdot\mu_{sbm}\big(m_2(\mathbf{r},t)\big)\Big), \\ \beta &= \Big(\mu_{sbm}\big(m_1(\mathbf{r},t)\big)\cdot\mu_{sbm}\big(m_3(\mathbf{r},t)\big)\Big), \\ \gamma &= \Big(\mu_{sbm}\big(m_2(\mathbf{r},t)\big)\cdot\mu_{sbm}\big(m_3(\mathbf{r},t)\big)\Big). \end{aligned} \quad (7)$$

Summarized, the membership degree of the pixel at position $(\mathbf{r'},t')$ in the fuzzy set "large weight", which corresponds to the weight $W(\mathbf{r'},t',\mathbf{r},t,1)$ in (3) is thus given by

$$W(\mathbf{r'},t',\mathbf{r},t,1) = \omega\cdot\theta\cdot\phi + (1-\omega)\cdot\phi - (\omega\cdot\theta\cdot\phi)\cdot((1-\omega)\cdot\phi), \quad (8)$$

where

$$\begin{aligned} &\omega = \mu_d(d(\mathbf{r},t)) \\ &\theta = (1-\mu_\Delta(\Delta_1(\mathbf{r'},t',\mathbf{r},t))) \\ &\phi = (1-\mu_\Delta(\Delta_2(\mathbf{r'},t',\mathbf{r},t))) + (1-\mu_\Delta(\Delta_3(\mathbf{r'},t',\mathbf{r},t))) \\ &\quad - (1-\mu_\Delta(\Delta_2(\mathbf{r'},t',\mathbf{r},t)))\cdot(1-\mu_\Delta(\Delta_3(\mathbf{r'},t',\mathbf{r},t))) \end{aligned}$$

For pixel positions in the window belonging to the current frame. For pixel positions in the window belonging to the previous frame, an extra factor $1-\mu_m(m(\mathbf{r},t))$ is needed.

3.2 Second Subfilter

Because sometimes not enough similar neighbours can be found to completely average the noise in the first subfilter and because some pixels might have been wrongly considered similar in the first subfilter, some colour artefacts might still be present after applying the first subfilter. To further improve the result, the first subfilter is combined with an additional second subfilter, which is an extension of the second subfilter in [6]. Based on the simplified assumption that the difference between similar pixels is approximately the same in all three colour bands, a pixel is estimated from a neighbour by estimating a difference in a given colour component equal to the average over all three colour bands. So a difference that is larger than the average is made smaller and vice versa. The final output is a weighted average over the estimations obtained from the different neighbours, where the weight is the degree to which we believe that the neighbour belongs to the same object. The weights are introduced because for neighbours not belonging to the same object, the simplified assumption does not hold.

3.2.1 Local Differences and Correction Terms

As mentioned before, for this second subfilter, a 3×3 sliding window is used. In each step the central pixel in this window, at position $(\mathbf{r},t)$ in the image sequence, is filtered. For each pixel in the sliding window, local differences (gradients) in the three colour bands (each separately) are calculated. The differences in the red, green and blue neighbourhoods are respectively denoted by LD_1, LD_2 and LD_3 and they are calculated based on the output of the first subfilter ($i=1,2,3$):

$$LD_i(\mathbf{r'},t,\mathbf{r},t) = I_f(\mathbf{r'},t,i) \;-\; I_f(\mathbf{r},t,i). \quad (9)$$

Next, for each position in the window one correction term is determined using the calculated local differences. This correction term is defined as the average of the local difference in the red, green and blue component at the given position:

$$\epsilon(\mathbf{r'},t,\mathbf{r},t) = \frac{1}{3}\sum_{i=1}^{3} LD_i(\mathbf{r'},t,\mathbf{r},t). \quad (10)$$

3.2.2 Output of the second subfilter

In [6], the output for each component of the central pixel is an average of the corresponding components of the neighbourhood pixels, corrected with the corresponding correction term ($i=1,2,3$):

$$Out(\mathbf{r},t,i) = \frac{\sum_{\mathbf{r'}}\Big(I_f(\mathbf{r'},t,i) - \epsilon(\mathbf{r'},t,\mathbf{r},t)\Big)}{9}. \quad (11)$$

However, pixels that belong to another object and that have another colour, have a negative influence on the output. In homogeneous areas, neighbouring pixels are expected to be almost the same, and the local differences to be almost 0. So the method further averages the remaining differences caused by the noise. For a pixel belonging to another object however, the assumption that the local differences are expected to be equal in all components does not always hold. Therefore we assign weights $WT(\mathbf{r'},t,\mathbf{r},t)$ to the neighbouring pixels, based on whether they are expected to belong to the same object or not. To make this decision, we use the Euclidian distance between the central pixel and the considered neighbourhood pixel, given by

$$\delta(\mathbf{r'},t,\mathbf{r},t) = \big(\sum_{i=1}^{3} LD_i(\mathbf{r'},t,\mathbf{r},t)^2\big)^{\frac{1}{2}} \quad (12)$$

ISBN: 978-989-95079-6-8

The weights themselves are then calculated using the following fuzzy rule that expresses that the value $\delta(\mathbf{r'}, t, \mathbf{r}, t)$ should not be large. Otherwise, the considered pixel is expected to belong to another object.

Fuzzy Rule 6. *Assigning the weight in the second subfilter for the pixel at position* $(\boldsymbol{r'}, t)$ *in the filtering window:*

IF $\delta(\boldsymbol{r'}, t, \boldsymbol{r}, t)$ *is* NOT LARGE

THEN *the pixel at position* $(\boldsymbol{r'}, t)$ *has a* LARGE WEIGHT $WT(\boldsymbol{r'}, t, \boldsymbol{r}, t)$ *in the second subfilter.*

The membership function μ_δ that determines the fuzzy set "large Euclidian distance" is depicted in Fig 3. The weights in

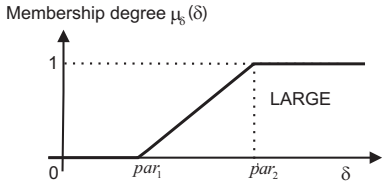


Figure 3: The membership function μ_δ of the fuzzy set "large Euclidian distance".

the filtering are again chosen equal to their membership degree in the fuzzy set "large weight", i.e., $WT(\mathbf{r'}, t, \mathbf{r}, t) = 1 - \mu_\delta(\delta(\mathbf{r'}, t, \mathbf{r}, t))$.

Finally, if not $WT(\mathbf{r'}, t', \mathbf{r}, t) = 0$ for all neighbouring pixels in the 3×3 window, the output of the second subfilter for the central pixel in the window is determined as follows ($i = 1, 2, 3$):

$$Out(\mathbf{r}, t, i) = \frac{\sum_{\mathbf{r'}} WT(\mathbf{r'}, t, \mathbf{r}, t)\Big(I_f(\mathbf{r'}, t, i) - \epsilon(\mathbf{r'}, t, \mathbf{r}, t)\Big)}{\sum_{\mathbf{r'}} WT(\mathbf{r'}, t', \mathbf{r}, t)} \tag{13}$$

where $\epsilon(\mathbf{r'}, t, \mathbf{r}, t)$ is the correction term for the components of the neighbouring pixel at position $(\mathbf{r'}, t')$. If the central pixel is so corrupt that all neighbouring pixels get a weight equal to zero, the output is calculated by giving all neigbhouring pixels in the window a weight equal to 1 and the corrupt central pixel the weight 0 ($i = 1, 2, 3$):

$$Out(\mathbf{r}, t, i) = \frac{\sum_{\mathbf{r'} \neq \mathbf{r}} \Big(I_f(\mathbf{r'}, t, i) - \epsilon(\mathbf{r'}, t, \mathbf{r}, t)\Big)}{8} \tag{14}$$

4 Experimental Results

In this section we present some experimental results obtained from the test sequences "Salesman", "Tennis", "Chair" and "Flowers", corrupted with additive Gaussian noise of zero mean and standard deviation $\sigma_n = 5, 10, 15, 20, 25$.

To draw conlusions about the proposed filtering framework, we have used the peak signal to noise ratio (PSNR) and the normalized colour difference (NCD) as measures of objective similarity and dissimilarity between the original and the filtered frames [6]. The higher the PSNR and the lower the NCD, the more similar the original and the filtered frame.

4.1 Parameter Selection

The parameters, that determine the membership functions in the above described filtering framework, have been set as follows. For the respective noise levels $\sigma_n = 5, 10, 15, 20$, the optimal parameters in terms of the mean PSNR values averaged over the sequences "Salesman", "Tennis", "Flowers" and "Chair" have been determined by letting them vary over a range of possible values. As illustrated in Fig. 4 for the parameter thr_2, this led to a linear relationship between these optimal values and the noise level. Hence, the parameters are set as the best fitting line through the observations, as shown in Fig. 4. The equations of those straight lines are given in Table 1, where σ_n stands for the standard deviation of the Gaussian noise. If this standard deviation is not known, it can be estimated by e.g. the method from [10].

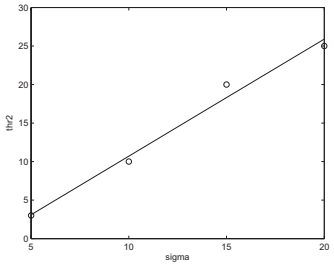


Figure 4: Optimal parameter values in terms of the PSNR.

Table 1: The used parameter values.

Parameter	Value	Parameter	Value
thr_1	0	t_1	$0.72\sigma_n - 4.0$
thr_2	$1.52\sigma_n - 4.5$	t_2	$2.22\sigma_n - 4.5$
T_1	0	par_1	$1.1\sigma_n - 7.5$
T_2	$3.14\sigma_n - 1.0$	par_2	$6\sigma_n + 35$

4.2 Experiments

In the experiments, we have compared the proposed filter (denoted by FMDAF-CR) to the filtering scheme in which a wavelet extension of the original pixel domain greyscale method [4], which outperforms both the pixel domain filter and other state-of-the-art greyscale methods of a similar complexity (as shown in [4]), is applied on the Y component of the YUV transform, followed by an averaging (3×3 window) of the chrominance components U and V (denoted by FMDAF-YUV). The average PSNR and NCD values found for the test sequences corrupted with different noise levels and processed by these two above approaches are given in Tables 2 and 3 respectively. From these tables, we can conclude that the proposed colour extension performs better in terms of average PSNR and average NCD than the commonly used YUV-filtering. For a visual comparison, we have made the original and noisy ($\sigma_n = 15$) "Salesman" sequence and the results after applying respectively the FMDAF-CR and the FMDAF-YUV filter available on `http://users.ugent.be/~tmelange/colourrule`. If

ISBN: 978-989-95079-6-8

Table 2: Average PSNR value for the processed sequences.

Sequence	σ_n	FMDAF-CR	WRFMDAF-YUV
"Salesman"	5	38.34	37.37
	10	34.36	33.64
	15	32.07	31.20
	20	30.45	29.35
	25	29.18	27.88
"Tennis"	5	36.51	33.40
	10	32.75	29.83
	15	30.35	27.92
	20	28.68	26.68
	25	27.39	25.59
"Chair"	5	39.27	39.92
	10	36.17	35.67
	15	33.94	32.99
	20	32.17	30.89
	25	30.67	29.16
"Flower garden"	5	33.32	29.03
	10	30.07	26.39
	15	27.65	24.98
	20	25.87	23.67
	25	24.47	22.44

Table 3: Average NCD value for the processed sequences.

Sequence	σ_n	FMDAF-CR	WRFMDAF-YUV
"Salesman"	5	0.0446	0.0531
	10	0.0676	0.0829
	15	0.0861	0.1121
	20	0.1027	0.1397
	25	0.1183	0.1652
"Tennis"	5	0.0269	0.0339
	10	0.0376	0.0475
	15	0.0470	0.0601
	20	0.0555	0.0724
	25	0.0638	0.0848
"Chair"	5	0.0132	0.0122
	10	0.0182	0.0215
	15	0.0238	0.0309
	20	0.0296	0.0404
	25	0.0353	0.0498
"Flower garden"	5	0.0543	0.0792
	10	0.0743	0.0945
	15	0.0907	0.1064
	20	0.1041	0.1190
	25	0.1157	0.1319

we concentrate on the side of the phone, we see that for the FMDAF-YUV method more colour artefacts (green and red spots) are visible than for the FMDAF-CR filter, which might be an explanation for the better PSNR and NCD values. Further, we also see that the wavelet domain method FMDAF-YUV has removed more noise and produces a smoother result. This smoother result can however be attributed to the use of a wavelet domain filter. If the original pixel domain greyscale method would have been used, we would also have had a little more noise remaining. Remark also that the smoother result also has as a result that the details have been smoothed a little more. This trade-off between noise removal and detail preservation is one of the main challenges in the development of a noise filter.

Acknowledgment

This work was financially supported by the GOA-project B/04138/01 IV 1 of Ghent University.

References

[1] S.J. Sangwine and R.E.N. Horne. *The colour image processing handbook*, Chapman&Hall, London, 1998.

[2] E.J. Balster, Y.F. Zheng and R.L. Ewing. Combined spatial and temporal domain wavelet shrinkage algorithm for video denoising. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(2):220–230, Feb. 2006.

[3] S.M. Mahbubur Rahman, M. Omair Ahmad and M.N.S. Swamy. Video denosing based on inter-frame statistical modeling of wavelet coefficients. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(2):187–198, Feb. 2007.

[4] T. Mélange et al. A new fuzzy motion and detail adaptive video filter. *Proc. of ACIVS 2007, LNCS*, 4678:640–651, 2007.

[5] V. Zlokolica, A. Pizurica and W. Philips. Video denoising using multiple class averaging with multiresolution. *International Workshop VLBV03, LNCS,* 2849:172–179, 2003.

[6] S. Schulte et al. A New Fuzzy Additive Noise Reduction Method. *Proc. of ICIAR 2007, LNCS* 4633:12–23, 2007.

[7] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353,1965.

[8] S. Weber. A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms. *Fuzzy Sets and Systems*, 11(2):115–134, 1983.

[9] E.B. Bellers and G. De Haan. *De-interlacing: A Key Technology for Scan Rate Conversion*, Elsevier Science B.V., Sara Burgerhartstraat, Amsterdam, 2000.

[10] D. Donoho and I. Johnstone. Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, 8:425–455, 1994.

ISBN: 978-989-95079-6-8

Visualising Rough Time Intervals in a Two-Dimensional Space

Yi Qiang[1], Katrin Asmussen[1], Matthias Delafontaine[1], Birger Stichelbaut[2], Guy De Tré[3], Philippe De Maeyer[1], Nico Van de Weghe[1]

1 CartoGIS Cluster, Dept. of Geography, Ghent University, Belgium
2 Dept. of Archaeology and Ancient History of Europe, Ghent University, Belgium
3 Dept. of Telecommunications and Information Processing, Ghent University, Belgium

E-mail: {Yi.Qiang, Katrin.Asmussen, Matthias.Delafontaine, Birger.Stichelbaut, Guy.DeTre, Philippe.DeMaeyer, Nico.VandeWeghe}@UGent.be

Abstract *— A lot of disciplines (e.g. archaeology) have to process imprecise temporal information. There are different possibilities to handle this kind of information, amongst them e.g. fuzzy set theory and rough set theory. In this paper, due to its capability in the context of many data acquisition applications, the focus has been set on rough set theory. To illustrate temporal information, an interval is often visualised by means of a one-dimensional segment in a one-dimensional space. An alternative representation of time intervals is called the Triangular Model (TM) by which a time interval is represented by a point in a two-dimensional space. In this paper, rough set theory is applied into TM, which gets extended to the Rough Triangular Model (RTM). In RTM, Rough Time Intervals (RTI) and their mutual relations can be visualised diagrammatically, which offers opportunities to visualise and analyse imprecise temporal information. Aerial photos, taken during World War I, containing imprecise temporal information with archaeological background, are used to illustrate the potentials of the model in processing RTI.*

Keywords — Imprecise temporal information, rough set theory, temporal reasoning, temporal relation, time interval, triangular model.

1 Introduction

For time intervals, the most widely adopted representation is the linear model where time intervals are modelled as finite linear segments in a one-dimensional space. Much research has been carried out on representing and reasoning about time intervals, most of which is based on this linear concept and simple temporal relations [1] [3] [6] [7] [10]. Though Rough Time Intervals (RTI), i.e. intervals starting and/or ending at uncertain time stamp are also frequently used in many disciplines, there is still a shortage of methods and tools to visualise them. An alternative temporal model, the Triangular Model (TM), has been proposed in [11] [12] [2]. This model is based on the W-diagram introduced in [4] [5]. Up till now, TM has remained mainly a theoretical concept. However, it seems to offer a promising design for several applications.

A lot of disciplines (e.g. archaeology, geography, psychology, and philosophy) are faced by the problem of having imprecise temporal information. To handle this information it can be reverted to different approaches, for instance to fuzzy set theory [13]. This theory aims to formalise inherently fuzzy concepts by permitting the gradual assessment of the membership of an element in relation to a set. Another way to deal with imprecise temporal information can be found in rough set theory [8] [9]. This theory introduces a concept of lower and upper approximation and a boundary region, describing a set where elements can or can not be decisively classified into a set X. In contrast to fuzzy set theory, rough set theory is particularly useful in the context of many data acquisition applications. Therefore, in this paper the rough set theory has been chosen to be applied into the TM.

This paper extends TM to RTM in order to visualise and analyse RTIs. First, the basics of TM are introduced in section two. The remainder of this section describes RTIs and their visualisation in the RTM. In section three, we illustrate how to visualise rough time relations of RTIs by means of the RTM. This section is followed by a description of an application of the RTM to incomplete temporal data which is deriving from an archaeological background. Finally, conclusions are drawn and future work is pointed out.

2 The Triangular Model

2.1 Representing time intervals with TM

Time is usually conceptualised as a linear, one-dimensional time line (Fig. 1). In this classical concept, a temporal interval I is visualised by a segment that is bounded by a begin point I^- and end point I^+. In the linear model, the vertical dimension is only used to differentiate multiple overlapping intervals, if used at all. An interval without duration is visualised as a point (zero dimension). The basic concept of TM is the construction of two lines through the extremes of a linear time interval (Fig. 2). For each time interval I, two straight lines (L_1 and L_2) are constructed, with L_1 going through I^-; L_2 going through I^+, and where the angle $\alpha_1 = \alpha_2$. The intersection of L_1 and L_2 is called the interval point I. The position of I in the two-dimensional space completely determines both, the beginning and end point of the interval (Fig. 3).

ISBN: 978-989-95079-6-8

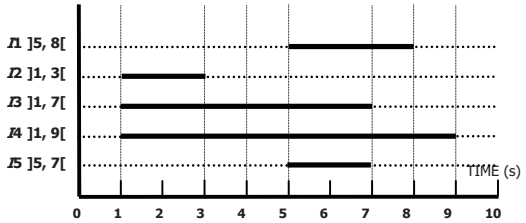


Figure 1: Linear representation of time intervals.

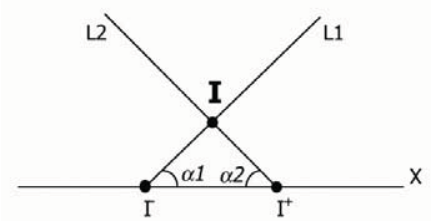


Figure 2: Construction of a simple interval point.

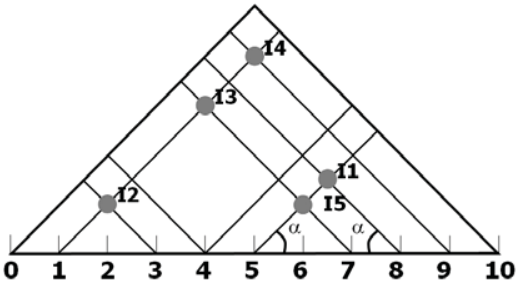


Figure 3: Time intervals in TM.

2.2 Representing RTIs with TM

It is often difficult for scientists to obtain precise temporal information about events or processes. Geological periods, for example, are always associated with uncertain starting and ending times. In history and archaeology as well, many events are lacking precise temporal demarcation. For instance, consider the impreciseness coming along with radiocarbon dating.

Scientifically underpinned treatments for imprecise information handling are already existing. For instance, fuzzy set theory [13] aims to formalize inherently fuzzy concepts by permitting the gradual assessment of the membership of an element in relation to a set. It extends conventional (crisp) set theory and handles the concept of partial truth, i.e. truth values between 0 (complete false) and 1 (complete true).

Another way to deal with imprecise information can be found in rough set theory [8] [9]. The main purpose of rough set theory is the induction of approximations of concepts, which are represented by the upper approximation $\overline{B}$ and the lower approximation $\underline{B}$. Within $\underline{B}$, elements can be decisively classified into a set X; outside $\overline{B}$, elements are not members of X. The difference between $\overline{B}$ and $\underline{B}$ forms a boundary region. If the boundary region is nonempty, a set is said to be rough; otherwise the set is crisp. In the boundary region, elements can not be decisively classified as members or not members of X.

Rough set theory is particularly useful in the context of temporal information, due to the nature of data acquisition in many scientific applications. Remote sensing, for instance, is a world-wide applied tool that relies on images taken at discrete time stamps. One can determine the state of a feature on these snapshots, whereas this state is uncertain in between two time stamps. However, we rely on the very natural assumption that this state does not change in between two snapshots which show similar states, i.e. the uncertain parts only remain in between two snapshots showing different states. Hence, we might consider a period of snapshots showing similar states (closed time interval) as a lower approximation for this state and its neighbouring periods of uncertainty (two open time intervals) as boundary region; both then cumulate into the state's upper approximation (open time interval) (Fig. 4). Thus, a feature's state can be considered as a rough set or more specifically a RTI. Next to remote sensing, this idea applies to numerous other fields, such as, soil core sampling, socio-economical census and surveys, and opinion polls.

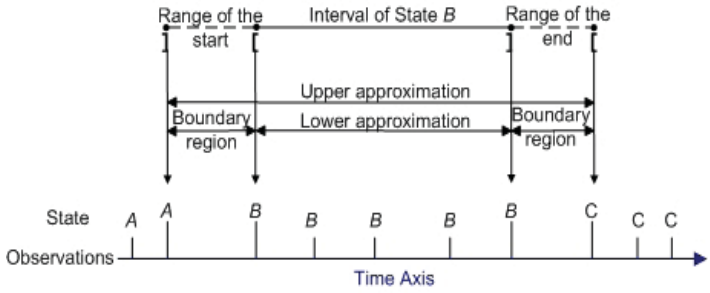


Figure 4: Rough set interval in remote sensing.

In the classical linear model, RTIs are represented as linear segments with an uncertain beginning range and an uncertain end range (Fig. 5). If there is a huge amount of rough intervals, it is quite difficult for humans to abstract information from this representation. Therefore the analyse capacities of the linear model are quite limited. For an advanced temporal analysis based on visualisation an alternative is needed.

In TM, a RTI is represented by a polygon. Four lines are constructed respectively from the earliest/latest beginning and the earliest/ latest end, forming a diamond (Fig. 6). This polygon indicates a zone within which the uncertain interval can be found. We call this representation the Rough Triangular Model (RTM).

In some cases, the beginning and end range have intersections. Take I_5 in Fig. 6 for example, we only know that it starts between 3 and 5, and ends between 3 and 5. In this case, the interval is represented as a triangle incumbent on the x-axis in RTM. In some other cases, intervals are partially rough: they have a precise beginning/end, but an imprecise end/beginning. These semi-rough intervals can be represented by lines (e.g. I_2 in Fig. 6). There are still other possibilities such as semi-open intervals, two-sided open intervals and totally indefinable intervals. In this paper, we are focussing on the

ISBN: 978-989-95079-6-8

first type of RTIs which are having a twofold, non-overlapping boundary region.

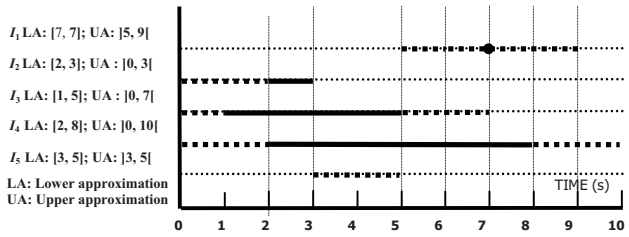


Figure 5: Linear representation of RTIs[1].

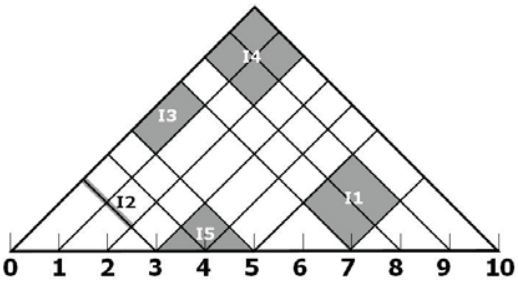


Figure 6: Visualising RTIs in RTM.

3 Rough temporal relations

When time intervals are rough, the relations between them also become rough. RTM provides the possibility of visualising and analysing temporal relations in a two-dimensional space

3.1 Representing fine temporal relations in TM

Based on the basic work of Allen [1], an interval I is represented as a pair (I^- ; I^+) with real numbers I^- and I^+, denoting respectively the beginning and end points of the interval. This means that Allen only deals with simple intervals having a specified duration. Let the beginning I^- and end point I^+ of two simple intervals have the following three possible relations: smaller than (<), equal (=) and larger than (>). Then, thirteen possible fine relationships between two intervals can be defined (see Tab. 1).

Table 1: Thirteen Allen relations [1].

Allen's temporal relations			
I_1 *equal* I_2	if $I_1^- = I_2^-$	^ $I_1^+ = I_2^+$	
I_1 *starts* I_2	if $I_1^- = I_2^-$	^ $I_1^+ < I_2^+$	
I_1 *started-by* I_2	if $I_1^- = I_2^-$	^ $I_2^+ < I_1^+$	
I_1 *finishes* I_2	if $I_1^+ = I_2^+$	^ $I_1^- > I_2^-$	
I_1 *finished-by* I_2	if $I_1^+ = I_2^+$	^ $I_2^- > I_1^-$	
I_1 *meets* I_2	if $I_1^+ = I_2^-$		
I_1 *met-by* I_2	if $I_2^+ = I_1^-$		
I_1 *overlaps* I_2	if $I_2^- > I_1^-$	^ $I_1^+ < I_2^+$	^ $I_1^+ > I_2^-$
I_1 *overlapped-by* I_2	if $I_1^- > I_2^-$	^ $I_1^- < I_2^+$	^ $I_2^+ < I_1^+$
I_1 *during* I_2	if $I_1^- > I_2^-$	^ $I_1^+ < I_2^+$	
I_1 *contains* I_2	if $I_2^- > I_1^-$	^ $I_2^+ < I_1^+$	
I_1 *before* I_2	if $I_1^+ < I_2^-$		
I_1 *after* I_2	if $I_2^+ < I_1^-$		

[1] In this paper, we use ISO standard notation to distinguish open intervals and close intervals. In this notation,]a, b[denotes open interval, [a, b] denotes close interval and]a, b] denotes left-open but right-close intervals. Since Allen's intervals are open, we use]a,b[here to denotes Allen's intervals.

Using TM [12], these relations can be visualised. Each relation thereby corresponds to a specific Fine Relation Zone (FRZ) within TM. Given a study period beginning at 0 and ending at 100, all examined intervals are located within the isosceles triangular of I]0, 100[. To obtain the best visualisation, the reference interval I_2]33,66[is chosen to be located in the centre of TM. As shown in Fig. 7b several intervals (I_{1a}, I_{1b}, I_{1c}, I_2) may exist *before* interval I_2]33,66[. All possible intervals for which $I_1^+ < I_2^-$ applies are generalised, with respect to interval I_2, into the FRZ *before,* displayed by the black triangle in Fig. 7c. Note that as Allen worked with open intervals, also the interval zone corresponding to I]0,33[is open. The right boundary of FRZ *before* represents all intervals for which applies $I_1^+ = I_2^-$. Therefore, the intervals have their end point at 33, resulting in the *meets* relationship.

Comparing the visualisations of TM with the linear model, both visualise the same fine interval relations, as displayed in Fig. 7a and Fig. 7c. The benefit of TM in visualising time relations gets spontaneously definite. The time intervals in Fig. 7c is faster to capture than the one in Fig. 7a. An overview of positions and names of the thirteen FRZs is given in Fig. 8.

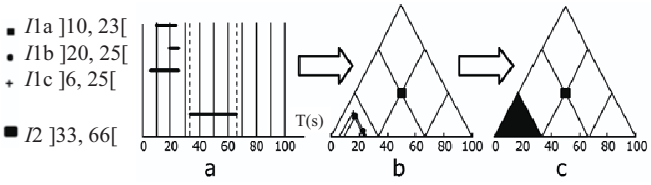


Figure 7: Visualisation of fine intervals by means of the linear model a) and TM b). FRZ *before* within TM c).

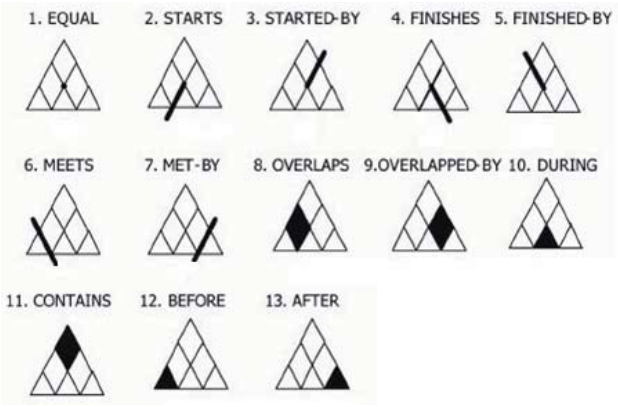


Figure 8: Thirteen FRZs in TM.

3.2 Relation zones in RTM

While TM represents fine intervals and fine interval relations, RTM represents rough intervals and its relation zones are Rough Relation Zones (RRZs).

To transform FRZs into RRZs, all point and line zones of TM (*equal, starts, started-by, finish, finished-by, meets,* and *met-*

ISBN: 978-989-95079-6-8

by) are expanded according to the duration of the boundary regions of the reference interval I_2. Hence, two polygon zones arise from the intersections of the expanded line zones *starts* and *meets* as well as from the intersections of the expanded line zones *finishes* and *met-by* (Fig. 9). Note that the absolute position of these zones may change according to the used data, while the relative position always remains.

Transforming the TM to RTM, fifteen relation zones are generated; nine of these are expanded zones (Fig. 9 and Tab. 2). The names of the expanded zones are based on the names of FRZs, but preceded by 'maybe'. Therefore these expanded relation zones are named *maybe* zones.

In TM, only one specific fine relation is possible within a FRZ (Tab. 2). Different from that, in RTM *maybe* zones represent zones where several fine relations are possible (Tab. 2). This is caused by the fact that a RRZ consists of parts of corresponding neighbouring FRZs in TM. The dashed lines in Fig. 9 represent the borders of original FRZs. In RTM, the positions of these dashed lines are uncertain and can be anywhere within the corresponding *maybe* zone. This is different from fuzzy set interval where the dashed lines have gradually-changing probability of appearance. Thus, we only use flat colour (white) to mark the interior of *maybe* zones. Note that each maybe zone contains the relations of its neighbouring zones in RTM (Fig. 9).

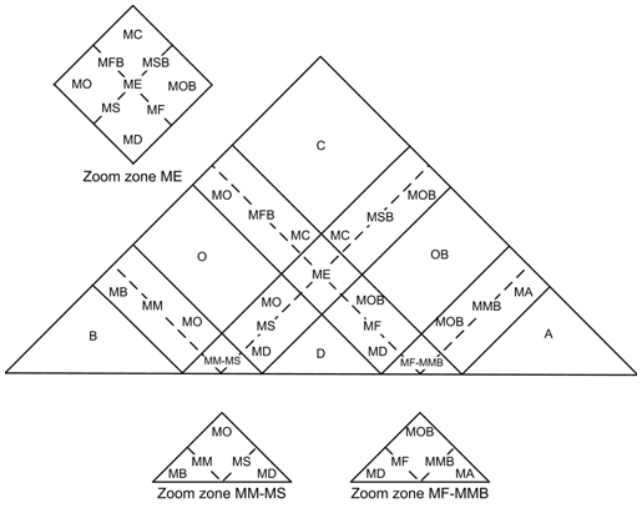


Figure 9: RRZs in RTM.

Table 2: Thirteen FRZs in TM and fifteen RRZs in RTM

FRZ		RRZ		
Name	**Abbr.**	**Name**	**possible relations**	**Abbr.**
equal	E	*maybe equal*	*contains, started-by, overlapped-by, finishes, during, starts, overlaps, finished-by, equal*	ME
starts	S	*maybe starts*	*starts, overlaps, during*	MS
started-by	SB	*maybe started-by*	*started-by, overlapped-by, contains*	MSB
finishes	F	*maybe finishes*	*finishes, during, overlapped-by*	MF
finished-by	FB	*maybe finished-by*	*overlapps, finished-by, contains*	MFB
meets	M	*maybe meets*	*meets, before, overlaps*	MM
met-by	MB	*maybe met-by*	*met-by, overlapped-by, after*	MMB
overlaps	O	*overlaps*	*overlaps*	O
overlapped-by	OB	*overlapped-by*	*overlapped-by*	OB
during	D	*during*	*during*	D
contains	C	*contains*	*contains*	C
before	B	*before*	*before*	B
after	A	*after*	*after*	A
/	/	*maybe meets or maybe starts*	*before, meet, overlaps, during, starts*	MM-MS
/	/	*maybe finishes or maybe met-by*	*during, finishes, overlapped-by, met-by, after*	MF-MMB

Unlike with FRZs, the border lines of RRZs do not make up separate relation zones but belong to one of their neighbouring RRZs. Due to their definition, the boundary regions in the beginning and the end of RTIs are open intervals (Fig. 4). The corresponding assignments are illustrated in Fig. 10, where the arrows are pointing into the zone to which the respective border line is assigned.

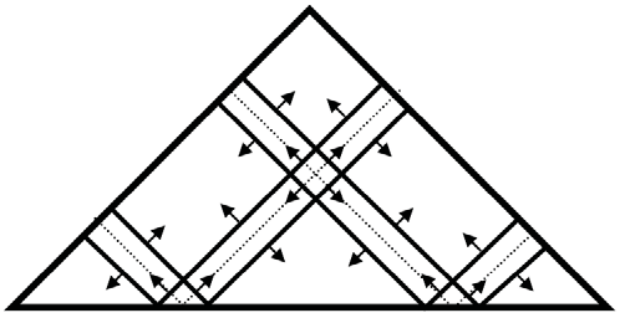


Figure 10: Assignment of the borders of RRZs.

3.3 Applying RTM

In order to evaluate whether RTM is helpful for the analysis of imprecise temporal information; we tested RTM with data deriving from an archaeological context. During World War I, aerial photos covering the Belgian-German front line in West-Flanders (Belgium) have been taken. From these aerial photos, we can observe whether a feature such as a fire trench, a gun position or a barrack, was not yet present, was present or was destroyed. For each feature, several photos with according date of acquisition exist. Hence, if we can find (1)

ISBN: 978-989-95079-6-8

the last photo where the feature is not yet found, (2) the first photo where the feature is found, (3) the last photo where the feature is found, and (4) the first photo where the feature is destroyed, we can derive a RTI representing the feature's uncertain lifetime (Fig. 4).

In RTM, fifteen RRZs can be created for one feature (Fig. 9). Lots of similar intervals will yield overlapped RRZs. If different grey shades are added to denote the number of overlaps in the concerning area, there will appear some patterns which reflect characteristics of the underlying data.

Fig. 11 and Fig. 12 visualise the overlaps of *maybe equal* zones of gun positions and breastworks respectively. More overlapped areas are marked with dark grey, and vice versa. Comparing these two visualisations, we can have a direct overview about how the two types of features are temporally distributed. As in the model most polygons of gun positions are distributed right of the breastworks zones, comparing Fig. 11 with Fig. 12, we may observe that gun positions exist relatively later than breastworks. It also can be observed that breastworks generally exist longer than gun positions because most of the breastworks zones are higher on the *y*-axis than the zones of the gun positions.

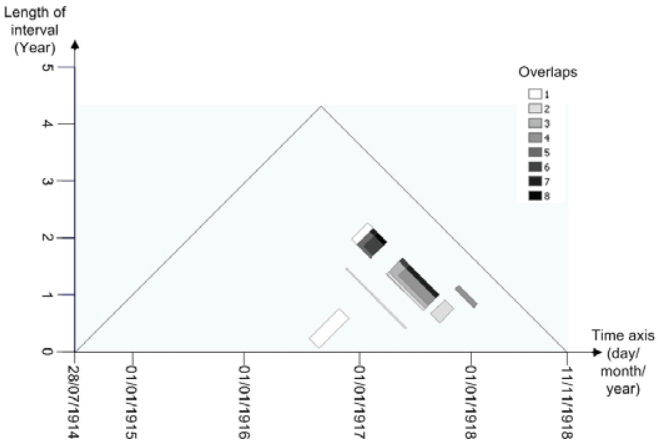


Figure 11: Overlaps of maybe equal zones of gun positions.

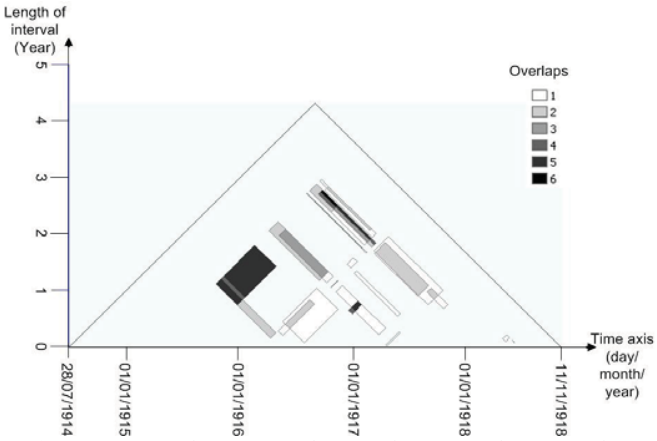


Figure 12: Overlaps of maybe equal zones of breastworks.

If we select *contain* interval zones of gun positions (Fig. 13), the display of the model can be divided into several zones according to the colour grade. If an interval point is in the dark zone, it contains intervals of most features. In other words, in the dark zone, all intervals contain most features lifetime intervals. In natural language, we could say: if an interval is in a dark zone, it witnesses the lifetime of most features. This approach can be applied to other RRZs and even to combinations of relation zones.

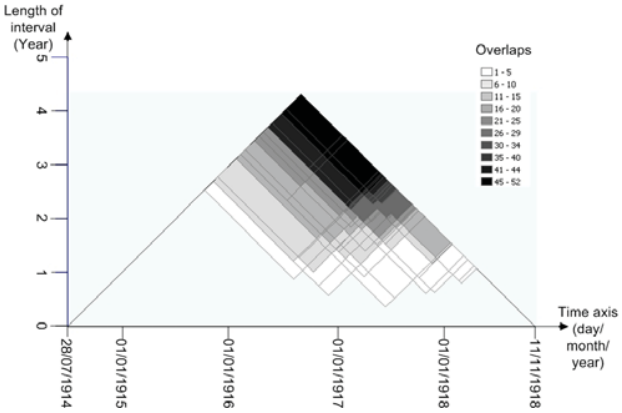


Figure 13: The overlaps of the *contain* zone of gun positions.

4 Conclusions and future work

Since a lot of disciplines (e.g. archaeology, geography, psychology, and philosophy) are faced by the problem of having imprecise temporal information we extended TM into RTM in order to visualise RTIs and temporal relations between them. We do not intend to create a new or extend an existing temporal calculus or temporal logic. Our approach takes special care for the intuitive and visualization aspects. In RTM, RTIs are represented by simple geometries (e.g. lines and polygons) in a two-dimensional space. Compared to the classical linear model, which has limited analytical capacities, RTM gives people a direct overview of the temporal distribution of RTIs. When handling a huge amount of rough intervals, RTM provides a compact visualisation pattern, which helps in further exploratory analysis. Furthermore, the temporal relations between RTIs can be visualised as zones, i.e. RRZs, in the two-dimensional space, which gives potentials in visual queries of rough temporal relations.

In this paper, we described the basics of RTM and RRZs. However, further research needs to be done.

First, as explained in 2.2; also other types of RTIs should be considered, e.g. partially rough intervals, one-sided open intervals and intervals with overlapped begin and end boundary regions. This would imply a wide modification of the RRZs. This major change in the division of the RRZ consequently has to be followed by an adaptation of visualisation and interpretation. Whether these visualisation and analyses are delivering feasible results has to be tested in further research.

ISBN: 978-989-95079-6-8

Second, the idea of conceptual neighbourhood has proved its importance in qualitative reasoning about time and space. When temporal relations become rough, there are more possibilities that one relation continuously changes to another relation without passing through other relations. Thus, the conceptual neighbourhood diagram of rough temporal relations will change accordingly. Also, the practical use of conceptual neighbourhood of temporal relations of rough intervals has to be studied in detail.

Third, so far we dedicated the borders of the RRZs of the RTM to one of the RRZs. But following the example of the TM, these lines could be relation zones by themselves. That would mean that additionally to the fifteen RRZ we would have nineteen line zones and probably six point zones. In total RTM would be build up out of forty zones. The additional line zones would include, for example, line zones of TM like *starts*, *started-by*, *finishes*, *finished-by, meets* and *met-by*. But also new line zones would be created like the line zones which are located between the zones *maybe starts or maybe meets* and *maybe meets*. Within this line zone a wide range of relations are possible. Whether this great number of RRZ provides useful visualisations and if the interpretation could be aided by a computer still needs to be determined. Also whether this interpretation would deliver helpful information still needs to be investigated.

Finally, this model can be more useful if it is implemented as an interactive tool. More flexible and interactive visualisation tools can help to better analyse complex temporal data.

Acknowledgment

The research work of Yi Qiang, Matthias Delafontaine and Birger Stichelbaut is funded by the Research Foundation–Flanders. The research work of Katrin Asmussen is co-funded by the Belgian Federal Science Policy.

References

[1] J.F. Allen. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM,* 26:832-843, 1983.

[2] G. De Tré, et al. Towards a flexible Visualisation Tool for dealing with Temporal Data. *In proc. of "FQAS" LN AI*, 4027: 109-120, 2006.

[3] C. Freksa. Temporal Reasoning based on Semi-Intervals. *Artificial Intelligence*, 54:199-227, 1992.

[4] Z. Kulpa. Diagrammatic Representation for a Space of Intervals. *Machine Graphics & Vision* 6: 5-24, 1997.

[5] Z. Kulpa. Diagrammatic Representation of Interval Space in Proving Theorems about Interval Relations. *Reliable Computing* 3:209-217, 1997.

[6] V. Kumar, R. et al. Allen. Metadata Visualization for Digital Libraries: Interactive Timeline Editing and Review. *In proc. of "ACM Digital Libraries"*, Pittsburgh, USA, 1998.

[7] J.D. Mackinlay. The Perspective Wall: Detail and Context Smoothly Integrated. *In proc. "ACM CHI'91"*, New York, 1991.

[8] Z. Pawlak. Rough Sets. *Int. J. of Information and Computer Science*, 11:341-356, 1982.

[9] Z. Pawlak. Rough Sets: Theoretical Aspects of Reasoning About Data. *Kluwer Academic Publishers,* Dordrecht, The Netherlands, 1991.

[10] S. Schockaert, M.D. Cock, and E.E. Kerre. Fuzzifying Allen's Temporal Interval Relations. *IEEE Transactions on Fuzzy Systems*, 16: 517-533, 2008.

[11] N. Van de Weghe, et al. The Triangular Model as an Instrument for Visualising and Analysing Residuality. *J. of Archaeological Science* 34: 649-655, 2007.

[12] N. Van de Weghe. Development of a Conceptual Data Model for digital spatio temporal Geographical Information. *In proc. "ER", LNCS*, Tampere, Finland, 2002.

[13] L.A. Zadeh. Fuzzy Sets. *Information and Control* 8: 338-353, 1965.

ISBN: 978-989-95079-6-8

Repairing infeasibility in fuzzy goal programming

Mariano Jiménez[1], Amelia Bilbao[2], Mar Arenas[2], Mª Victoria Rodríguez-Uría[2]

1 Dpto. de Economía Aplicada I, Universidad del País Vasco,
Plaza Oñati 1. 20018 San Sebastián. Spain
2 Dpto. de Economía Cuantitativa, Universidad de Oviedo,
Avda. del Cristo s/n. 33006 Oviedo. Spain
E-mail: mariano.jimenez@ehu.es , ameliab@uniovi.es , mariamar@uniovi.es , vrodri@uniovi.es

Abstract*—The problem of solving a multi-objective programming problem, by assuming that the decision maker has fuzzy goals for each of the objectives, is addressed. In this case a fuzzy goal programming problem has to be solved. But when we have several objectives it can be a difficult task for the decision maker to determine coherent aspiration levels, so it is possible that the tolerance thresholds be inconsistent with each other, therefore obtaining an infeasible problem. We present a procedure in order to repair the infeasibility. We rebuild the unattainable fuzzy goals keeping their relaxation levels as small as possible while maintaining, at the same time, a good balance vis-à-vis the degree of achievement of the others goals.*

***Keywords*—**, Fuzzy goal programming, Goal programming, Infeasibility, Multi-objective programming.

1 Introduction

Let a multi-objective programming (MOP) problem with k objective functions $z_i(x) = c_i x$, $i = 1,...,k$ be

$$\text{(MOP)} \quad \text{Min} \quad z(x) = (z_1(x), z_2(x), ..., z_k(x))$$
$$\text{S. t.} \quad x \in X$$

Where $X = \{x \in \mathbb{R}^n \mid Ax \geq b, x \geq 0\}$, $c_i = (c_{i1}, ..., c_{in}) \in \mathbb{R}^n$, $i = 1,...,k$; $b = (b_1, ..., b_m) \in \mathbb{R}^m$ and A is a $\mathbb{R}^{m \times n}$ matrix.

For the sake of simplicity we suppose that all the objectives should be minimized. However, the procedure demonstrated is easily extendable were the case to include some maximizing objectives.

In (MOP) it is unlikely that all objectives will simultaneously achieve their optimal value. Therefore in practice the decision maker (DM) chooses a satisfying solution, according to the aspiration level fixed for each objective.

Assuming that the DM proposes imprecise aspiration levels such as, "$z_i(x)$ should be essentially less than or equal to some value g_i", model (MOP) can be written as

$$\text{(FGP)} \quad \text{Find } x$$
$$\text{Such that} \quad z_i(x) \lesssim g_i \quad i = 1,2,...,k$$
$$x \in \text{X}$$

Each expression $c_i x \lesssim g_i$ is represented by a fuzzy set called fuzzy goal, whose membership function, $\mu_i(z_i)$, $\mu_i : \mathbb{R} \to [0,1]$, provides the satisfaction degree λ_i to which the i^{th} fuzzy inequality is satisfied. In order to define the membership function $\mu_i(z_i)$ the DM has to provide the tolerance margins g_i+t_i that she/he is willing to accept. So $\mu_i(z_i)$ should be equal to 1 if $z_i \leq g_i$, strictly monotone decreasing from 1 to 0 over the interval $(g_i, g_i + t_i)$ and equal to 0 if $z_i \geq g_i + t_i$.

Models like this are named fuzzy goal programming (FGP) problems. (FGP) should be considered as an auxiliary model in order to solve (MOP).

As it has been widely seen, a FGP problem, using the fuzzy decision max-min of Bellman and Zadeh [1] and introducing the auxiliary variable λ, adopts the following formulation [2]

$$\text{(B-Z)} \quad \text{Max } \lambda$$
$$\text{S.t.} \quad 0 \leq \lambda \leq \mu_i(z_i(x)) \quad i = 1,\ldots,k$$
$$x \in X, \ \lambda \leq 1$$

But when there are several objective functions it can be a difficult task for the DM to determine coherent aspiration levels. Therefore it is possible that some tolerance thresholds will be inconsistent with each other or with the crisp constraints, some of them being unattainable (membership degree=0). In this case (B-Z) is infeasible. Chinneck [3] summarizes the state of the art in algorithms related to infeasibility in optimization. Most published research has been focused on the diagnosis of infeasibility. Little investigation has been made in infeasibility resolution and very limited literature can be found in infeasibility resolution using fuzzy approach. León and Liern [5] use a fuzzy sets approach to relax constraints to repair infeasibility in crisp mathematical programming problems. Gupta et al. [6] take a similar fuzzy approach to finding a best approximate solution to an infeasible mathematical programming problem. In this paper we propose a procedure in order to solve a FGP should the above described drawback occur. We demonstrate the

ISBN: 978-989-95079-6-8

operativeness of our approach by means of a numerical example.

2 Repairing infeasibility

Let us assume that (B-Z) is infeasible because the tolerance thresholds of some goals are unattainable. The first step (M.I) is to identify one minimum-cardinality subset of {1, 2,…, k}, that will be denoted *CS*, so that the goals {1, 2,…, k}−*CS* provide a feasible (B-Z) problem. This procedure is related to the minimum cardinality irreducible infeasible subsystem (IIS) set-covering problem (MIN IIS COVER) (see [3, 4]). Hence the Chinneck's algorithm [4] will be applied to the following sets of linear constraints

$$z_i(x) \le g_i + t_i \qquad i = 1,2,...,k$$
$$x \in \mathrm{X}$$

where only the k -first constraints (the goals) are elasticized. So we will work with the model

$$\text{(M.I)} \quad \text{Min} \sum_{i=1}^{k} p_i$$
$$\text{S.t.} \quad z_i - p_i \le g_i + t_i \quad i = 1,\ldots,k$$
$$x \in X$$

Next, in order to repair infeasibility, we have to shift the unattainable goals, i.e. the *CS*-goals, as low as possible without penalizing excessively the achievement degree of other goals. Taking this into account we consider the following two positions: 1) we allow the feasible goals to deteriorate to their corresponding tolerance threshold 2) we aim to attain a satisfactory solution for the feasible goals. Obviously the relaxation of the unattainable goals will be smaller in the first position than in the second. Therefore we assign a satisfaction degree equal to 1 to the relaxed values obtained in the first position and a satisfaction degree equal to 0 to that obtained in the second position, decreasing monotonously between these values. This way, we obtain the membership functions $\mu_i^*\left(z_i(x)\right)$ that rebuild the new fuzzy goals corresponding to the, previously, unattainable goals.

Assuming, for the sake of simplicity, that the unattainable goals are the first *r*. Consider the aforementioned position 1, that is to say, we allow the feasible goals to deteriorate until their tolerance threshold and, being an inclusive condition, we look for the lowest relaxation levels for the unattainable goals by conventional goal programming approach [7]. Thus we propose to solve the following model

$$\text{(M.II)} \quad \text{Min} \sum_{i=1}^{r} w_i p_i$$
$$\text{S.t.} \quad z_i(x) - p_i \le g_i + t_i, \quad i = 1,...,r$$
$$z_i(x) \le g_i + t_i, \qquad i = r+1,...,k$$
$$x \in X, p_i \ge 0$$

where $\sum_{i=1}^{r} w_i = 1$ and $w_i \ge 0$ for all i=1,…,r. The variables p_i have the same significance as the positive deviation in goal programming (GP). (M.II) searches for a solution with the smallest weighted average relaxation levels in order to achieve feasibility, considering that the other goals could reach their corresponding tolerance threshold. Weight w_i represents the priority or importance of the relaxation levels of tolerance threshold g_i+t_i. It is well known that when $w_i>0$ for all i=1,…,r, any optimal solution of (M.II) is efficient [8], which give us a large set of possible goal shifts that provide a feasible solution.

Let x^{II} be an optimal solution of (M.II). In line with what we have said before, the values $z_i^{\mathrm{II}} = z_i\left(x^{\mathrm{II}}\right), i = 1,...,r$, have membership degree equal to 1 in the new (repaired) fuzzy goals that we are constructing.

Now let us consider the aforementioned position 2, that is to say we are looking for a good degree of satisfaction of the goals that are attainable in (M.I). In order to achieve this we solve the following max-min model in which only the feasible goals are included

$$\text{(M.III)} \quad \text{Max } \lambda$$
$$\text{S. t. } z_i(x) \ge z_i^{\mathrm{II}}, \ i = 1,...,r$$
$$0 \le \lambda \le \mu_i\left(z_i(x)\right) \quad i = r+1,\ldots,k$$
$$x \in X \text{ , } \lambda \le 1$$

Let x^{III} be an optimal solution of (M.III) and $z_i^{\mathrm{III}} = z_i\left(x^{\mathrm{III}}\right)$, i=r+1,…,k. We can consider the values $\max\left\{z_i^{\mathrm{III}}, g_i\right\}$, i=r+1,…,k as sufficiently satisfactory. In the process of searching for the lowest relaxation levels of the infeasible goals, maintaining the values of the feasible goals sufficiently satisfactory, we solve the following GP model

$$\text{(M.IV)} \quad \text{Min} \sum_{i=1}^{r} w_i p_i$$
$$\text{S.t.} \quad z_i(x) - p_i \le g_i + t_i, \quad i = 1,...,r$$
$$z_i(x) \ge z_i^{\mathrm{II}}, \qquad i = 1,...,r$$
$$z_i(x) \le \max\left\{z_i^{\mathrm{III}}, g_i\right\}, \qquad i = r+1,...,k$$
$$x \in X, p_i \ge 0$$

Let x^{IV} be an optimal solution of (M.IV). The values $z_i^{\mathrm{IV}} = z_i\left(x^{\mathrm{IV}}\right), i = 1,...r$, achieved by the unattainable goals, have membership degree equal to 0 in the new (repaired) fuzzy goals that we are constructing.

We write $z_i^{\mathrm{II}} \equiv g_i^*$ and $z_i^{\mathrm{IV}} \equiv g_i^* + t_i^*$.

This way we have obtained the new fuzzy aspiration levels $\mu_i^*\left(z_i(x)\right)$ which restore feasibility with the lowest relaxation levels (see figure 1).

Finally we solve the following max-min model, by replacing the unattainable aspiration levels substituted with the new aspiration levels, which restores feasibility.

$$\text{(B-Z*)} \quad \text{Max } \lambda$$
$$\text{S.t.} \quad \mu_i^*\left(z_i(x)\right) \ge \lambda \ge 0, i = 1,...,r$$
$$\mu_i\left(z_i(x)\right) \ge \lambda \ge 0, i = r+1,...,k$$
$$x \in X, \quad \lambda \le 1$$

ISBN: 978-989-95079-6-8

The optimal solutions of (B-Z*) can be considered as solutions of (B-Z) well balanced between attainable fuzzy goals an unattainable fuzzy goals. Moreover, if the solution of this problem is not fuzzy-efficient of (FGP) and/or not Pareto optimal of (MOP), the solution can be improved following a procedure showed in [9].

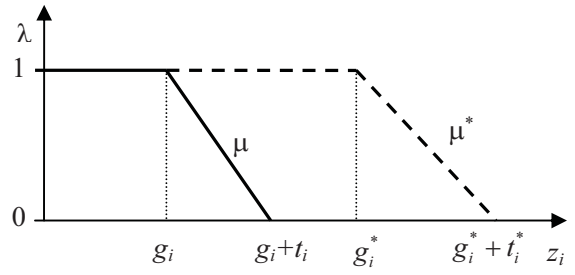


Figure1: Relaxation of an unattainable fuzzy aspiration level.

3 Example

To check the usefulness of our approach we consider the following numerical example.
Let a MOLP problem be

$$\begin{aligned}
&\text{Min } z_1=3x_1+3x_2+3x_3+2x_4+3x_5\\
&\text{Min } z_2=2x_1+x_2+2x_3+x_4+3x_5\\
&\text{Min } z_3=4x_1+4x_2+2x_3+3x_4+x_5\\
&\text{Max } z_4=x_1+6x_2+5x_3+7x_4+6x_5\\
&\text{Min } z_5= 5x_1+5x_2+4x_3+7x_4+5x_5\\
&\text{S.t. } 4x_1+2x_2+4x_3+2x_4+3x_5 \geq 18\\
&\quad x_1 \geq 1,\ 0 \leq x_3 \leq 3\\
&\quad x_2 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0
\end{aligned}$$

Assume the DM proposes the following imprecise aspiration levels: "z_1 should be essentially less than or equal to 18", "z_2 should be essentially less than or equal to 6", "z_3 should be essentially less than or equal to 13", "z_4 should be essentially greater than or equal to 24" and "z_5 should be essentially less than or equal to 17", the respective tolerance threshold being 21, 8, 15, 21 and 20.

Step 1. We solve the (B-Z) problem

$$\begin{aligned}
&\text{Max } \lambda\\
&\text{S.t.}\\
&\lambda \leq \frac{1}{3}\left(21-\left(3x_1+3x_2+3x_3+2x_4+3x_5\right)\right)\\
&\lambda \leq \frac{1}{2}\left(8-\left(2x_1+x_2+2x_3+x_4+3x_5\right)\right)\\
&\lambda \leq \frac{1}{2}\left(15-\left(4x_1+4x_2+2x_3+3x_4+x_5\right)\right)\\
&\lambda \leq \frac{1}{3}\left(\left(x_1+6x_2+5x_3+7x_4+6x_5\right)-21\right)\\
&\lambda \leq \frac{1}{3}\left(20-\left(5x_1+5x_2+4x_3+7x_4+5x_5\right)\right)\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \geq 18\\
&x_1 \geq 1,\ 0 \leq x_3 \leq 3\\
&x_2 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0,\ 0 \leq \lambda \leq 1
\end{aligned}$$

This problem is infeasible because the tolerance thresholds of some goals are unattainable.

Step 2. We will identify one minimum-cardinality irreducible infeasible subsystem of goals by applying the Chinneck's algorithm to the following system of linear constraints where only the five first constraints (the goals) are elasticized:

$$\begin{aligned}
&3x_1+3x_2+3x_3+2x_4+3x_5 \leq 21\\
&2x_1+x_2+2x_3+x_4+3x_5 \leq 8\\
&4x_1+4x_2+2x_3+3x_4+x_5 \leq 15\\
&x_1+6x_2+5x_3+7x_4+6x_5 \geq 21\\
&5x_1+5x_2+4x_3+7x_4+5x_5 \leq 20\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \geq 18\\
&x_1 \geq 1,\ 0 \leq x_3 \leq 3\\
&x_2 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0
\end{aligned}$$

We obtain $CS=\{2, 5\}$.

Step 3. In order to repair infeasibility we propose to solve the following model (see M.II)

$$\begin{aligned}
&\text{Min } p_2 + p_5\\
&\text{S.t.}\\
&2x_1+x_2+2x_3+x_4+3x_5-p_2 \leq 8\\
&5x_1+5x_2+4x_3+7x_4+5x_5-p_5 \leq 20\\
&3x_1+3x_2+3x_3+2x_4+3x_5 \leq 21\\
&4x_1+4x_2+2x_3+3x_4+x_5 \leq 15\\
&x_1+6x_2+5x_3+7x_4+6x_5 \geq 21\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \geq 18\\
&x_1 \geq 1,\ 0 \leq x_3 \leq 3\\
&x_2 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0\\
&p_2 \geq 0,\ p_5 \geq 0
\end{aligned}$$

We get $z_2^{II}=9$ and $z_5^{II}=21.5$. These values have membership degree equal to 1 in the new (repaired) fuzzy goals that we are constructing.
In order to look for the balanced relaxation levels of the infeasible goal we need to find the values sufficiently satisfactory for the feasible goals.

Step 4. We solve the following model (M. III)

$$\begin{aligned}
&\text{Max } \lambda\\
&\text{S.t.}\\
&\lambda \leq \frac{1}{3}\left(21-\left(3x_1+3x_2+3x_3+2x_4+3x_5\right)\right)\\
&\lambda \leq \frac{1}{2}\left(15-\left(4x_1+4x_2+2x_3+3x_4+x_5\right)\right)\\
&\lambda \leq \frac{1}{3}\left(\left(x_1+6x_2+5x_3+7x_4+6x_5\right)-21\right)\\
&2x_1+x_2+2x_3+x_4+3x_5 \geq 9\\
&5x_1+5x_2+4x_3+7x_4+5x_5 \geq 21.5\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \geq 18\\
&x_1 \geq 1,\ 0 \leq x_3 \leq 3\\
&x_2 \geq 0,\ x_4 \geq 0,\ x_5 \geq 0\\
&0 \leq \lambda \leq 1
\end{aligned}$$

As $z_i^{III}=15.48$, $z_3^{III}=11.32$, $z_4^{III}=26.52$ then in the process of searching for the lowest relaxation levels of the infeasible

ISBN: 978-989-95079-6-8

goals, maintaining the values of the feasible goals sufficiently satisfactory, we solve the following GP model:

Step 5. (see M.IV)

$$\begin{aligned}
&\text{Min } p_2 + p_5\\
&\text{S.t.}\\
&2x_1+x_2+2x_3+x_4+3x_5 - p_2 \le 8\\
&5x_1+5x_2+4x_3+7x_4+5x_5 - p_5 \le 20\\
&3x_1+3x_2+3x_3+2x_4+3x_5 \le 18\\
&2x_1+x_2+2x_3+x_4+3x_5 \ge 9\\
&4x_1+4x_2+2x_3+3x_4+x_5 \le 13\\
&x_1+6x_2+5x_3+7x_4+6x_5 \ge 24\\
&5x_1+5x_2+4x_3+7x_4+5x_5 \ge 21.5\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \ge 18\\
&x_1 \ge 1,\ 0 \le x_3 \le 3\\
&x_2 \ge 0,\ x_4 \ge 0,\ x_5 \ge 0\\
&p_2 \ge 0,\ p_5 \ge 0
\end{aligned}$$

We get $z_2^{IV} = 9.5,\ z_5^{IV} = 24.96$. These values achieved by the unattainable goals, have membership degree equal to 0 in the new (repaired) fuzzy goals that we are constructing.

This way we have obtained the new fuzzy aspiration levels which restore the feasibility with the lowest relaxation levels: "z_2 should be essentially greater than 9" and "z_5 should be essentially greater than 21.5" with tolerance threshold equal to 9.5 and 25, respectively. Finally we solve the following (B-Z*) model

$$\begin{aligned}
&\text{Max } \lambda\\
&\text{s.t.}\\
&\lambda \le \frac{1}{3}\left(21-\left(3x_1+3x_2+3x_3+2x_4+3x_5\right)\right)\\
&\lambda \le \frac{1}{0.5}\left(9.5-\left(2x_1+x_2+2x_3+x_4+3x_5\right)\right)\\
&\lambda \le \frac{1}{2}\left(15-\left(4x_1+4x_2+2x_3+3x_4+x_5\right)\right)\\
&\lambda \le \frac{1}{3}\left(\left(x_1+6x_2+5x_3+7x_4+6x_5\right)-21\right)\\
&\lambda \le \frac{1}{3.5}\left(25-\left(5x_1+5x_2+4x_3+7x_4+5x_5\right)\right)\\
&4x_1+2x_2+4x_3+2x_4+3x_5 \ge 18\\
&x_1 \ge 1,\ 0 \le x_3 \le 3\\
&x_2 \ge 0,\ x_4 \ge 0,\ x_5 \ge 0\\
&0 \le \lambda \le 1
\end{aligned}$$

We get $x_1^* = 1,\ x_2^* = 0.79,\ x_3^* = 2.86,\ x_4^* = 0.25,\ x_5^* = 0.13$; $z_1\left(x^*\right) = 14.92,\ z_2\left(x^*\right) = 9.2,\ z_3\left(x^*\right) = 13.83,\ z_4\left(x^*\right) = 22.75$ and $z_5\left(x^*\right) = 22.95$. This can be considered as a solution of model (B-Z) well balanced between attainable fuzzy goals and unattainable fuzzy goals.

Acknowledgment

This work was supported by the *Ministerio Español de Educación y Ciencia* under grant no. MTM2007-67634. and by the *Universidad del País Vasco* under grant no. EHU08/28.

4 Conclusions

This paper describes a method to deal with infeasibility in fuzzy goal programming problems in case of the DM proposes unattainable tolerance thresholds for some goals. Considering the inherent flexibility of fuzzy approach we restore the feasibility by shifting the unattainable goals as low as possible but, at the same time, taking care not to penalize excessively the degree of achievement of the others goals. This way we rebuilt the unattainable goals obtaining a new feasible fuzzy goal programming problem whose solution is well balanced between attainable and unattainable fuzzy goals.

References

[1] R.E. Bellman and L.A. Zadeh, Decision making in a fuzzy environment, *Management Science,* 17: 141-164, 1970.

[2] H.J. Zimmermann, Fuzzy programming and linear programming with several objective functions, *Fuzzy Sets and Systems*, 1 (1): 45-55, 1978.

[3] J.W. Chinnek, *Feasibility and Infeasibility in Optimization. Algorithms and Computational Methods.* New York: Springer, 2008.

[4] J. W. Chinnek, Fast Heuristics for the Maximum Feasible subsystem Problem, *INFORMS Journal on Computing,* Vol. 13, Nº 3, Suumer: 210-223, 2001.

[5] T. León and V. Liern, A fuzzy method to repair infeasibility in linearly constrained problem, *Fuzzy Sets and Systems,* 122: 237-243, 2001.

[6] P. Gupta, M. Vlach and D. Bhatia, Fuzzy approximation to an infesible generalized linear complementary problem, *Fuzzy Sets and Systems,* 146: 221-233, 2004.

[7] J. Jang, Infeasibility resolution based on goal programming, Computers & Operations Research, 35:1483-1493, 2008.

[8] Y. Sagarawi, H. Nakayama, T. Tanino. Theory of Multiobjective Optimization. London: Academia Press, 1985.

[9] M. Jiménez and A. Bilbao, Pareto-optimal solutions in fuzzy multi-objective linear programming, *Fuzzy Set and Systems*, 2008, doi: 10.1016/j.fss.2008.12.005.

Sorted kernel matrices as cluster validity indexes

Francisco A. A. Queiroz[1] Antonio P. Braga[1] Witold Pedrycz[2]

1 Department of Electronics Engineering, Federal University of Minas Gerais
Belo Horizonte, Brazil
2 University of Alberta, Department of Electrical and Computing Engineering
Edmonton, Canada
Email: faaq@ufmg.br, apbraga@ufmg.br, pedrycz@ece.ualberta.ca

Abstract—*Two basic issues for data analysis and kernel-machines design are approached in this paper: determining the number of partitions of a clustering task and the parameters of kernels. A distance metric is presented to determine the similarity between kernels and FCM proximity matrices. It is shown that this measure is maximized, as a function of kernel and FCM parameters, when there is coherence with embedded structural information. We show that the alignment function can be maximized according FCM and kernel parameters. The results presented shed some light on the general problem of setting up the number of partitions in a clustering task and in the proper setting of kernel parameters according to structural information.*

Keywords— Affinity matrix, Clustering, Fuzzy C-Means (FCM), Kernel matrix, Reordering, Sorting.

1 Introduction

The importance of kernels has been much emphasized in the literature as the basic construct of learning machines like SVMs [6]. Kernel parameters play, therefore, a major role in achieving good performance of these classifiers. Setting up kernel parameters is usually accomplished by an exhaustive search carried out in the space of parameters. Quite often this search is completed without a full understanding of how kernels represent relations in the input space. The proper set-up of kernels is only expected to capture the input to feature space mapping that results on an acceptable overall machine performance, without any direct constraints imposed on the internal representation of relationships between patterns.

However, like Fuzzy C-Means (FCM) [7] proximity matrices (FPM) [8], kernels may incorporate important structural information about the data. In fact, data clustering can be accomplished by using information already contained in kernels, by finding the proper order of columns and rows of the kernel matrix [2, 3, 4, 5]. A similar affinity matrix [1] representation can also be obtained for FPMs, which may suggest that kernels and FPMs embody similar information about the data if their parameters are set accordingly. This observation also suggests that supervised and unsupervised learning may exhibit a closer relationship by the exchange of information between kernels and FPMs.

The problem of approximating kernels and FPMs relies, therefore, on setting their parameters so that they yield matrices that are close to each other or, in other words, that represent the same structural information. The parameters to be set are basically the width *r* of the Gaussian functions, for RBF kernels, and the number of partitions c, the basic parameter of FPMs. In this paper, we investigate how these two parameters are related to the alignment between RBF kernels and FPMs.

Experimental results on synthetic data yielded maximum alignment on values of *c* that coincide with the number of clusters of the data generator functions. This suggests that, for a given kernel matrix, the proper number of clusters can be induced by maximizing the similarity of FPM and the kernel matrix. In addition, it is also shown that, for the same problems, SVMs [6] can also be designed by setting the kernel parameter to the corresponding value of *r* that maximizes the similarity.

The paper is organized as follows. We start with a general overview of the main concepts including kernels and FPM. The concept of affinity matrix representation of kernels and FPMs is then described. Another important link between these two representations of data is presented in section 5 where kernels are shown to incorporate the same structural information the FPMs. The metric for similarity measure is then presented, which is followed by the description of the optimization algorithm, the main results, discussions and conclusions.

2 Mercer Kernels

The main characteristic of (Mercer) kernels [9] in the supervised learning context is that they allow implicit mapping of input data into feature space without actually computing the mapping itself. Instead, the mapping is accomplished by computing the internal product located in the input space. The kernel matrix $\mathbf{K}=[k(x_i,x_j]$ is regarded as a Mercer kernel if it is symmetric and positive semi-definite. In general, the kernel matrix can be considered as a matrix capturing similarity between all pairs of points of a data set. Kernels can be implemented by polynomials, linear and sigmoidal functions as well as Gaussian functions. A Gaussian, or RBF, kernel is described as follows

$$k(x_i, x_j) = e^{\frac{-(x_i - x_j)^2}{2r^2}} \quad (1)$$

where *r* is the radius of the Gaussian function.

ISBN: 978-989-95079-6-8

As an example, let us consider Fig. 1, where data were sampled from two Gaussian distributions. The corresponding RBF kernel where *r=0.25* is shown in Fig. 2. As patterns are not ordered according to the generator functions, no structural information can be directly visualized in the kernel matrix. In order to observe pattern relation properties directly from the kernel matrix, it is necessary to order the patterns accordingly, as it will be described in the next sections.

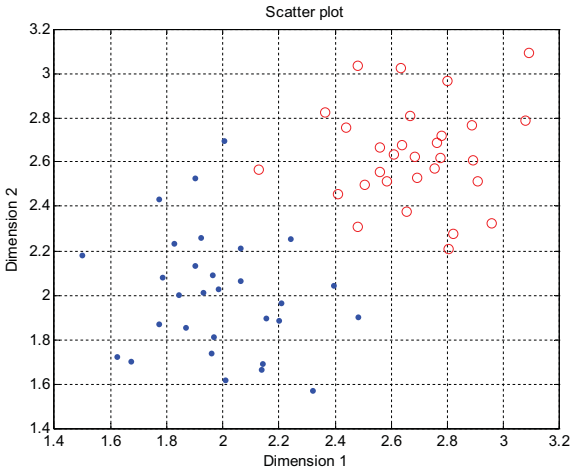


Figure 1: Data sample from two Gaussian distributions.

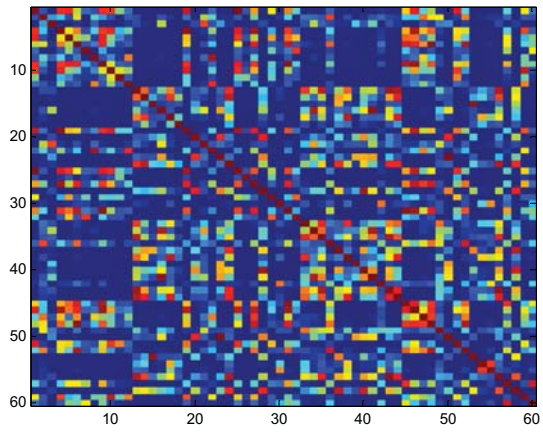


Figure 2: Unordered RBF kernel matrix (r=0.25) for the data of Fig. 1.

3 FCM Proximity Matrix

Fuzzy clustering realized by the well-known FCM algorithm [7] is formulated as a constrained optimization problem, where the partition matrix $\mathbf{U}=[u_{ik}]$ satisfies the constraint $\sum_{i=1}^{c} u_{ik} = 1$, while the objective function to be minimized reads as follows

$$J = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{m} \| \mathbf{x}_k - \mathbf{v}_i \|^2 \qquad (2)$$

where c is the number of partitions (clusters) in the data set, N is the data set size, m is the fuzziness coefficient (fuzziness factor) and $\mathbf{v}_i$ is the prototype of cluster i.

The obtained partition matrix $\mathbf{U}=[u_{ij}]$ can be used to generate the *NxN* proximity matrix $\mathbf{P}=[p_{kl}]$, according to the known relationship available in the literature [8] that is presented in (3). Likewise kernels, the proximity matrix **P** embodies relationships between patterns according to the clusters represented in the partition matrix **U**.

$$p_{kl} = \sum_{i=1}^{c} \min(u_{ik}, u_{il}) \qquad (3)$$

4 Affinity matrices

Given a data set $\Gamma_u = \{x_k\}_{k=1}^{N}$, where N is the number of patterns, the elements s_{ij} of the Affinity Matrix $\mathbf{S}=[s_{ij}]$ contain a measurement or estimation of the affinity of the pair of patterns $(\mathbf{x}_i, \mathbf{x}_j)$, where affinity is defined as a likeness based on relationship or causal connection [1]. For reflexive affinities, **S** is symmetrical, what implies on $s_{ij}=s_{ji}$. In general, the Affinity Matrix can be represented as a block diagonal symmetrical matrix as the one of (4).

$$S = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ S_{21} & S_{22} & \cdots & S_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{k1} & S_{k2} & \cdots & S_{kk} \end{bmatrix} \qquad (4)$$

where $\boldsymbol{S}_{ij}=\boldsymbol{S}_{ji}$ are sub-matrices of **S** and k is the total number of sub-groups of **S**.

For ordered data sets, kernels and proximity matrices can be represented in the form of (4), where sub-matrix $\mathbf{S}_{ii}$ represents within-cluster affinities and $\mathbf{S}_{ij}$ represents inter-cluster affinities, for $i \neq j$. Consider, for instance, the data set of Fig. 1 and the corresponding kernel presented in Fig. 2, which is now ordered according to the generator distributions and presented in Fig. 3. The block-diagonal affinity matrix form of the kernel presented in Fig. 3 shows clearly the structure of the data set. The block-diagonal matrices $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ represent the two clusters of Fig. 1 with 30 elements each, whereas the matrices $\mathbf{S}_{21}$ and $\mathbf{S}_{12}$ represent the relationships between data of matrices $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$. Matrices $\mathbf{S}_{21}$ and $\mathbf{S}_{12}$ have in fact very small values, indicating that the two clusters are not quite related to each other, as can be confirmed in Fig. 1. A similar result to the one presented in Fig. 3 would have been obtained if the proximity matrix **P** were obtained and ordered according to the partition matrix **U** for c=2.

ISBN: 978-989-95079-6-8

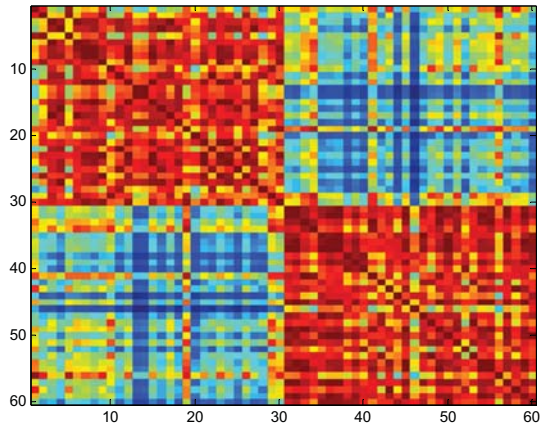


Figure 3: Ordered RBF kernel matrix ($r=0.25$) for the data set of Fig. 1.

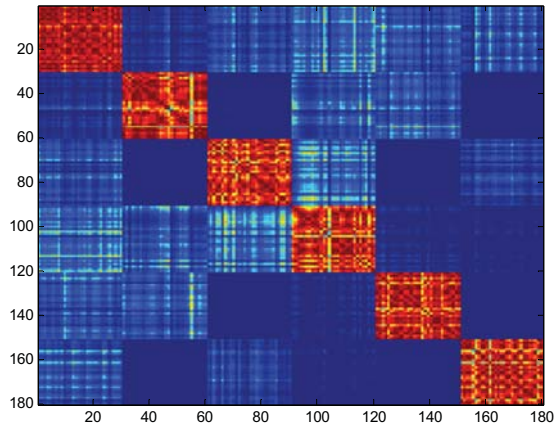


Figure 5: Affinity matrix representation of the ordered RBF kernel matrix of the data set of Fig. 4.

5 Clustering by Sorting Kernels

In the previous section we have pointed out to the close relationship between kernels and proximity matrices as represented by affinities of the data. Clustering algorithms can be described with the objective of sorting kernel matrices. There are in fact clustering algorithms in the literature that work by sorting rows and columns of the kernel matrix [2, 3, 4, 5] in order to obtain affinity matrices like the one presented in Fig. 3. Once the affinity matrix has been formed, embedded cluster information can be directly extracted from the kernels.

The eigenvector ordering method [1, 2] was applied to the data set presented in Fig. 4, which was generated from 6 Gaussian generator functions. The corresponding ordered RBF kernel is presented in Fig. 5. As it can be observed, the ordered kernel shows clearly the presence of 6 groups of data in the input space. The affinity matrix representation of kernels simply reveal the information that is already present in the kernel, since only permutation operations were accomplished in the rows and columns of the original matrix. This indicates that (properly set) kernels already contain information about data structure that is equivalent to that obtained by clustering methods.

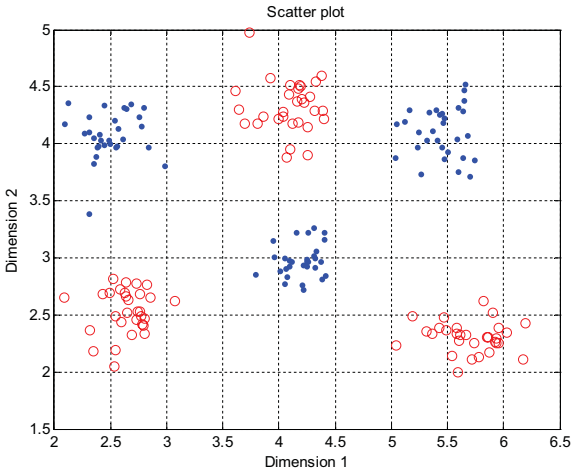


Figure 4: Data set obtained from 6 Gaussian functions.

6 Alignment of Kernels and FPMs

In order to quantify the similarity between kernels and proximity matrices, the Empirical Alignment described in [2] was adopted. The alignment quantity A is described in the form

$$A(K,P) = \frac{\langle K,P\rangle_F}{\sqrt{\langle K,K\rangle_F \langle P,P\rangle_F}} \tag{3}$$

where K and P, respectively, are the kernel and proximity matrices and $\langle .,. \rangle_F$ is the Frobenius inner product [2],

$$\langle K,P\rangle_F = \sum_{i=1}^{N}\sum_{j=1}^{N} K(i,j)P(i,j) \tag{4}$$

7 Alignment as an Optimization Problem

As discussed in the previous sections, depending on their parameters, kernels and FPM may incorporate analogous information about patterns and cluster relationships which hold in the input space. For the RBF kernels, structural information may be revealed by the proper setting of the width r, whereas FPM ability to describe pattern relations depends on the previously set number of clusters c. Nevertheless, these parameters are usually blindly configured in advance by the user. In the SVM design, the value of r can be fine-tuned according to the overall learning machine performance, while further cluster analysis can also give a hint for fine tuning of the value of c. When the setting of (c, r) results on affinity matrices that are coherent with the actual data structure, the corresponding kernel and FPM will be aligned to each other, according to the metric presented in (3). These arguments suggest that A(K,P) has a maximum on (c^*, r^*), where c^* and r^* are, respectively, the number of partitions and RBF width that best describe the original data structure. This characterizes the optimization problem presented in (5). The objective is therefore to obtain the values of the parameters c and r that maximize the alignment, which are expected to be the ones that describe better the data set for both kernel and proximity matrix.

$$\arg \max_{(c,r)} A(K,P) \tag{5}$$

ISBN: 978-989-95079-6-8

In order to demonstrate the arguments above, the values of A(K,P) were computed as a function of *c* and *r* for the data sets coming from Fig. 1 and 4, respectively. The obtained results are shown in Fig. 6 and 7. The same calculations were also completed for another data set with 4 original Gaussian generator functions. The obtained graph is presented in Fig. 8. As can be observed in Fig. 6, 7 and 8, the maximum occurs for *c=2*, *c=6* and *c=4*, respectively, which correspond to the original number of generator functions used in each data set. These results confirm the principle that the alignment function (3) has a maximum corresponding to the number of generator functions and that (5) could be optimized in order to obtain the values of *c* and *r* that result on the best representation of the data set. In the next section, the optimization of (5) will be implemented with a simple Genetic Algorithm (GA) [10] approach.

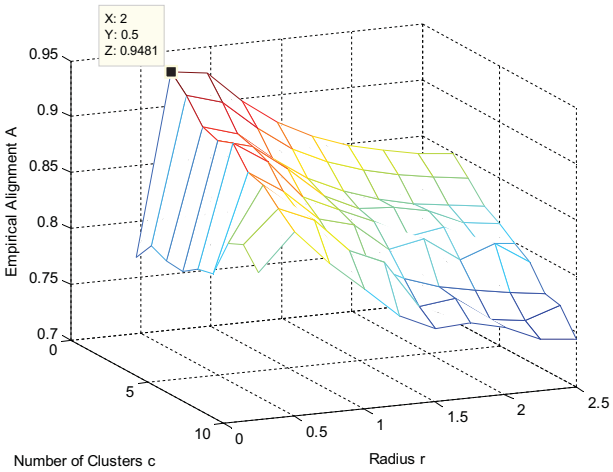


Figure 6: Alignment A(K,P) as a function of c and r for the data set of Fig. 1.

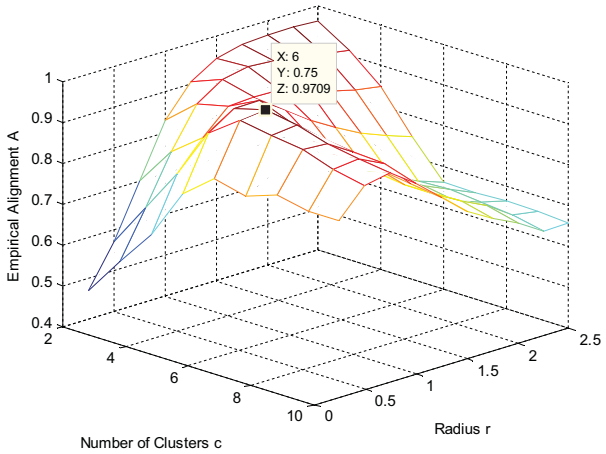


Figure 7: Alignment A(K,P) as a function of c and r for the data set of Fig. 4.

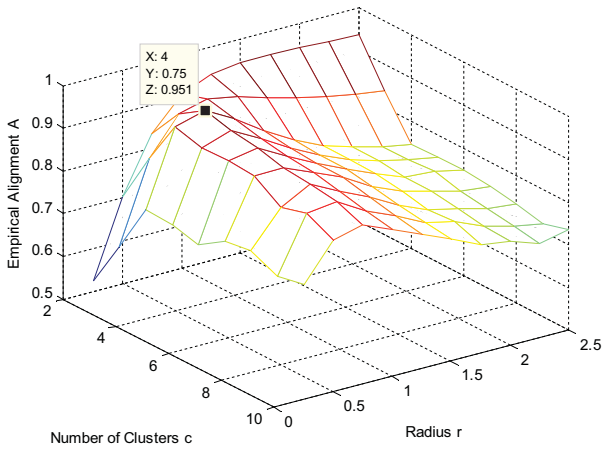


Figure 8: Alignment A(K,P) as a function of c and r for a data set with 4 Gaussian generator functions.

8 Evolutionary Optimization

A simple Genetic Algorithm (SGA) [10] was implemented in order to optimize (5) for the three problems represented by the objective functions of Fig. 6, 7, and 8. GA parameters are presented in Table 1 and the obtained results in Table 2. As can be observed, the maximum obtained for the three objective functions are quite close to the ones observed in the figures. As for the number of clusters, the optimization yielded exactly the same number of generator functions.

Table 1: Parameters of the SGA for optimization of the objective functions of Fig. 6, 7 and 8.

Number of variables in search space	2
Population size	50
Maximum number of generations	65
Crossover rate	0.6
Mutation rate	0.03
Biased linear cross-over coefficient	0.9
Linear cross-over coefficient	0.5

Table 2: SGA solutions for the objective functions of Fig. 6, 7 and 8.

Number of generator functions	***Number of iterations***	***r***	***c***	**A**
Two	52	0.5894	2	0.9540
Four	19	0.6808	4	0.9537
Six	63	0.6551	6	0.9758

The SGA was also applied to the data set of Fig. 9, represented as a binary classification problem. In order to distinguish between structural information and data set labels for classification problems, one of the classes was sampled from two Gaussian distributions, as can be observed in Fig. 9. In such a situation, there are three generator functions and two classes. This can be visualized by the ordered kernel of Fig. 10, which suggests the existence of the three clusters, in spite of the number of labels. The SGA optimization resulted on *r=0.66959*, *c=3* and *A=0.92225*,

ISBN: 978-989-95079-6-8

what is consistent with the number of clusters of the original distribution.

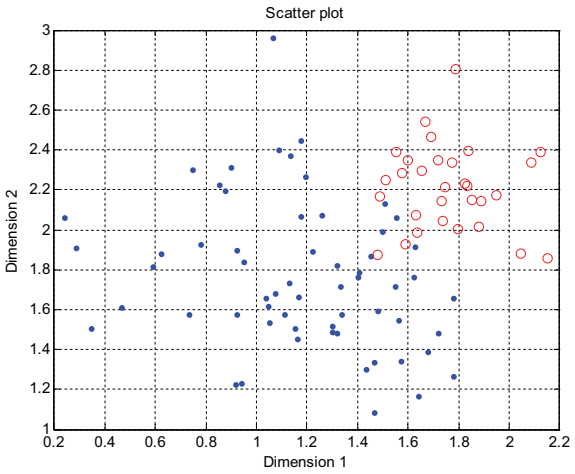


Figure 9: Binary classification problem sampled from three generator functions.

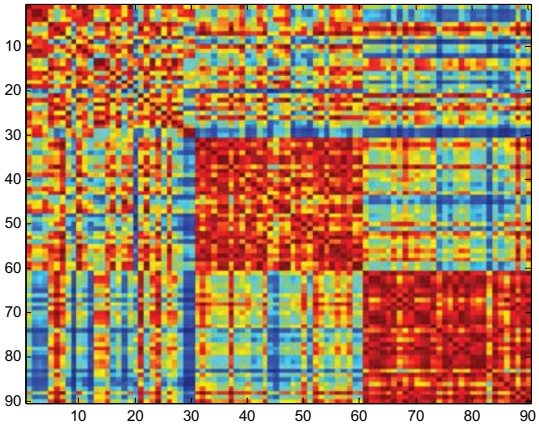


Figure 10: Affinity matrix representation of the ordered RBF kernel matrix of the data set of Fig. 9.

In order to show the consistency of the RBF width resulted from SGA optimization for classification purposes, a SVM was designed with the obtained value of $r=0.66959$. The corresponding separating surface is presented in Fig. 11 and, as can be observed, the resulting surface is coherent with the classification problem (the margin parameter was set to $C=2$). This suggests that the maximum obtained from (5) also points out to a proper tuning of kernel parameters for SVM design. Similar procedures were applied to the Iris data set [12], which resulted on $c=2$ in the maximum alignment. This result is consistent with the literature and with statistical analysis of the data set [13].

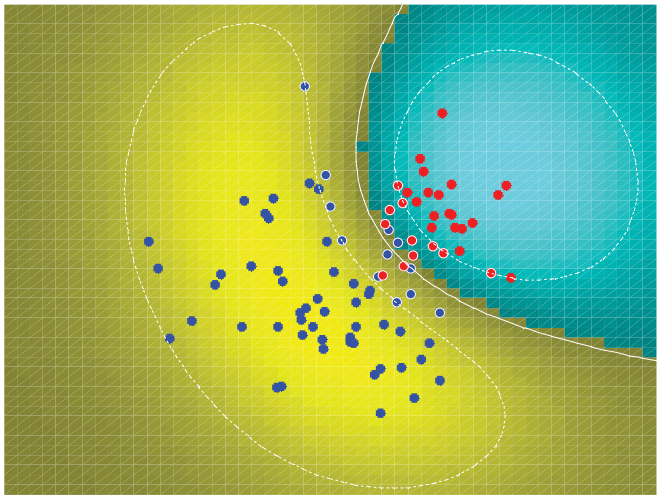


Figure 11: SVM separation surface for Fig. 9. RBF kernel was selected with and $r=0.66959$ obtained from SGA optimization.

9 Discussions and conclusions

Kernel and clustering design was presented in this paper in the perspective of function optimization. Our main argument to support this principle is that when both kernel and FPM are coherent with the embedded structural information contained in the data set, they should be aligned to each other. Alignment is measured in this paper according to the Frobenius inner product [2] between kernel and FPM, computed as a function of the parameters c and r. Therefore, we argue that maximum alignment should result on a proper setting of the parameters c and r. This approach differs from other kernel FCM algorithms [11], since the original objective functions are used.

This principle sheds some light on the discussion of determining the number of partitions of a clustering task. It has been shown throughout the paper that coherent partitions can be inferred according to the maximum alignment principle. Likewise, proper kernel design can also be obtained from the alignment. This suggests that kernel can be designed by exploring the structural data properties instead of performing an exhaustive blind search in the space of parameters with the aim of solely accomplishing the input-output mapping. According to these arguments, kernels should be more representative of the underlying data instead of serving uniquely as a mapping engine. This also suggests that kernels and FCM clustering should be co-designed and not described independently. According to the same principle, supervised and unsupervised learning may exhibit a closer relationship by collaborative interaction between kernels and FPMs.

References

[1] Y. Weiss, Segmentation using eigenvectors: a unifying view, *Proceedings of the International Conference on Computer Vision*, Volume 2, p. 975, May 20-25, 1999.

[2] N. Cristianini, J. Kandola, A. Elisseeff and J. Shaw-Taylor, On kernel target alignment, *Journal of Machine Learning Research*, 1, 2002.

[3] L. Võhandu, R. Kuusik, A. Torim, E. Aab, G. Lind, Some algorithms for data table (re) ordering using Monotone Systems, *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases: 5th WSEAS International*

ISBN: 978-989-95079-6-8

Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'06) , 417-422, Madrid, Spain, Feb 15-17, 2006.

[4] D. Tsafrir, I. Tsafrir, L. Ein-Dor, O. Zuk, D. A. Notterman and E. Domany, Sorting points into neighborhoods (SPIN): data analysis and visualization by ordering distance matrices. *Bioinformatics*, 21:2301-8, 2005.

[5] P. Arabie, L. J. Hubert, The bond energy algorithm revisited, *IEEE Transactions on Systems, Man and Cybernetics*, 20 (1): 268-274, Jan-Feb 1990.

[6] C. Cortes, V. Vapnik, Support Vector Network, *Machine Learning*, vol.20, pp.273-297.

[7] J. C. Dunn, A Fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybernetics* 3: 32-57. 1973.

[8] V. Loia, W. Pedrycz and S. Senatore, Proximity fuzzy clustering for web context analysis, *Proceedings of EUSFLAT 2003*, pp. 59-62.

[9] N. Cristianini and J. Shawe-Taylor, Support vector machines and other kernel-based learning methods. Cambridge University Press, 2000.

[10] D. E. Goldberg, Genetic algorithms in search, optimization and machine learning. Kluwer Academic Publishers, Boston, MA, 1989.

[11] W. Cai, S. Chen and D. Zhang, (2007). Robust fuzzy relational classifier incorporating the soft class labels**,** *Pattern Recognition Letters*, vol. 28(16), pp 2250-2263.

[12] A. Asuncion, D.J. Newman, UCI Machine Learning Repository [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, 2007.

[13] S. R. Gunn, Support vector machines for classification and regression (Technical Report). Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.

ISBN: 978-989-95079-6-8

Complementary source information cooperation within a decision system for crop monitoring

Mahmoud El Hajj[1] Serge Guillaume[2] Agnès Bégué[3]

1 NOVELTIS, Ramonville Saint-Agne, F-31520 France
2 Cemagref, UMR ITAP, Montpellier, F-34196 France
3 CIRAD, UMR TETIS, Montpellier, F-34093 France
Email: mahmoud.elhajj@noveltis.fr, serge.guillaume@montpellier.cemagref.fr, agnes.begue@teledetection.fr

Abstract— *Multi-source information fusion usually deals with redundant sources; the goal is to aggregate decisions, or evidence which supports one or another class. Many frameworks are then available: OWA, fuzzy integrals, etc. In this paper we are interested in managing complementary information sources. The integration is done using a fuzzy rule base. The rules are generated using a fuzzy decision tree algorithm. The application consists in the design of a crop harvest monitoring system. The three information sources are a time series of satellite images, a crop growth model and expert knowledge. The results show that the system is ready to use in an operational way.*

Keywords— Crop growth modeling, Decision, Expert knowledge, Fuzzy rules, Learning, Remote sensing.

1 Introduction

Over the past decade, time series of satellite images acquired at high spatial resolution have proven to be an important source of information for different agricultural applications. Several authors have recognized the benefits of this kind of data for monitoring agricultural lands [1], classifying land cover [2-4], mapping seasonal patterns and crop rotations [5, 6], and for many other uses (see papers collected in [7, 8]).

Nevertheless, the quantity of information extracted from time series of optical images is often restricted by several factors: acquisition gaps, atmospheric conditions, imperfect radiometric normalization, radiometric confusion, etc. Therefore, to make credible decisions, this information needs to be supplemented with data from other sources.

Let us focus on the sugarcane industry. One of its principal needs is to have continuous information about the harvest progress throughout the season. Such information helps increasing the effectiveness in fields and factory. The main method reported in the literature, to monitor the sugarcane harvest is based on a multispectral classification [9-11]. The major limitations of this method are the subjectivity and the considerable time to invest in the photo-interpretation phase and the confusion between the various states of a sugarcane field when the gap between two released images is important (more than 2 months). Since the harvest campaign extends over several months, it is difficult in some cases to distinguish between a standing crop and the regrowth in a field harvested at the beginning of the season.

In order to improve and automate the monitoring of sugarcane harvest using time series of high spatial resolution satellite images, it is necessary to develop a new method that process the time series in co-operation with other sources. This co-operation allows compensating for the possible lack of data in the time series of images by the integration of temporal and environmental constraints.

As the sources are complementary and of different nature, usual frameworks of fusion information cannot be used. In this paper we propose to make the source integration within a fuzzy inference system. Next section introduces the study site and the data used. The characteristics of the three information sources are presented in section 3. Section 4 describes the fuzzy inference system and section 5 goes into detail of the experiments. The last section deals with concluding remarks.

2 Study site and data sets

The study site consists of two sugarcane farms located in the north-east part of Reunion Island, which is a small territory of ~2,500 km² in the Indian Ocean (21°7' to 19°40' S, 55°13' to 61°13' E), where sugarcane is the main crop. The first farm is at an average altitude of 70 m and includes 33 fields with an average size of 5.4 ha. The second farm is at an altitude ranging from 400 to 700 m, and has 46 fields with an average size of 3.5 ha. As the study area is located in a tropical zone, the year is divided into two seasons: a hot rainy season from November to April, and a cool dry season from May to October.

The satellite data set used in this study consists of 18 SPOT5 images acquired over Reunion Island between January 10th, 2003 and December 7th, 2004. Both SPOT5 instruments (HRG1 and HRG2) acquire radiation in four spectral bands[1] with high spatial resolution: 10 m for the Green, Red, and Near Infra-Red (NIR) bands, and 20 m for the Short Wave Infra-Red (SWIR) band. The images belong to the Kalideos-ISLE REUNION database

[1] Green (0.50 - 0.59 µm), Red (0.61 - 0.68 µm), Near Infra-Red (0.78 - 0.89 µm), Short Wave Infra-Red (1.58 - 1.75 µm).

ISBN: 978-989-95079-6-8

set up by the CNES[2] [12, 13]. All images were ortho-rectified and co-registered to the UTM coordinate system (zone 40 South) with a root mean square error of less than 0.5 pixel per image.

The radiometry of the images was corrected so that pixel values represent the top of canopy reflectances in the four spectral bands [14, 15]. Cloud mask was available for each image.

Block field boundaries for all Reunion Island were provided by the DDAF[3] and were refined by the CIRAD[4] to define the boundaries of each field in the study site.

Daily climatic data recorded at La Mare meteorological station near the two sugarcane farms were collected for the period covered by the satellite time series. These data are daily estimations of rainfall (mm), potential evapo-transpiration (mm), global radiation (J/m^2), and minimum, maximum and mean temperature values (°C). Climatic data were required to run the crop growth model.

A ground truth database was built by using harvest dates reported by farmers for each field during the 2003 and 2004 harvest campaigns. This database indicates the status of each field (whether it was harvested or not) between each pair of consecutive satellite acquisition dates in the time series.

3 Three information sources

3.1 SPOT5 time series

The time series of SPOT5 images are the principal source of information. By using field boundaries, the temporal profiles of reflectances in the four spectral bands, Green, Red, NIR and SWIR, were extracted from the images for each field in the study site. This extraction was carried out by calculating the average value of field pixels in each spectral band. The calculation was done after discarding cloud pixels using the cloud masks. Temporal profiles of NDVI (Normalized Difference Vegetation Index) where then calculated for each field using reflectance values in the Red, ρ_{Red}, and NIR, ρ_{NIR}, bands:

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (1)$$

The temporal profile of NDVI provides useful information about the actual field status, and about its different historical stages. In general, this profile can be divided into two periods: a period in which NDVI values increase, corresponding to the vegetative development of the field crop, and another period with steady or decreasing values, corresponding to the maturation phase. Fig.1 shows an example of the temporal profile of a sugarcane field NDVI, extracted from a time series of SPOT5 images acquired with high temporal repeatability.

3.2 Sugarcane growth model

To cope with missing data or cloudy images, the sugarcane growth model MOSICAS [16] was used. It provides information on the field harvest possibility that is independent of the satellite images and based essentially on climatic data. This information is particularly interesting when the gap between two cloud-free satellite images is more than two months.

MOSICAS is a dynamic model that estimates the growth of the sugarcane on a daily time scale by using climatic and biophysical data of the environment of the sugarcane field. It allows the simulation of LAI (Leaf Area Index) profiles (Fig.1) that can be converted into NDVI profiles using a linear model [17]. An illustration of such a simulated profile is plotted in Fig.1. From the simulated NDVI temporal profiles, we build a helpful indicator for harvest detection: this indicator represents the nominal time required (in days) to reach a given threshold of NDVI starting with a given harvest date.

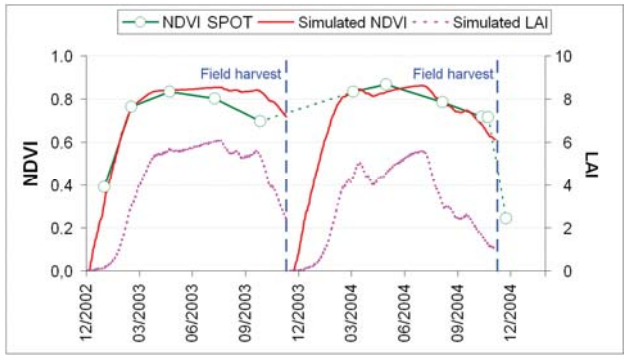


Figure 1: Example of temporal profiles of a sugarcane field NDVI extracted from SPOT5 time series, as well as NDVI and LAI temporal profiles simulated using the sugarcane growth model MOSICAS.

3.3 Expert Knowledge

The third source of information is the knowledge about phenological stages of the sugarcane, as well as its cropping system. This knowledge allows a better understanding of the relationship between the dynamics of the signals extracted from the satellite image time series and the sugarcane field status; it offers also very useful temporal constraints that particularly help in making decisions when there is a lack of satellite data. Examples of parameters defining these temporal constraints are the nominal length of the sugarcane cycle and the dates of opening and closing of the sugar factories which determine the harvest campaigns, as illustrated in Fig.2.

Various variables are built from the three information sources and used as inputs of the decision system. The inputs are divided into three main groups:

- The first group (4 inputs) is based on the NDVI and SWIR values calculated at field scale using each satellite image.
- The second group (5 inputs) relates to the dynamics in the temporal profiles of NDVI and SWIR.
- The third group (5 inputs) represents the temporal constraints resulting from the sugarcane growth model and the expert knowledge.

[2] *Centre National d'Etudes Spatiales* ; French Spatial Agency.

[3] *Direction Départementale de l'Agriculture et de la Forêt* ; Departmental Directorate of Agriculture and Forestry.

[4] *Centre de coopération Internationale en Recherche Agronomique pour le Développement* ; French Agricultural Research Centre for International Development.

ISBN: 978-989-95079-6-8

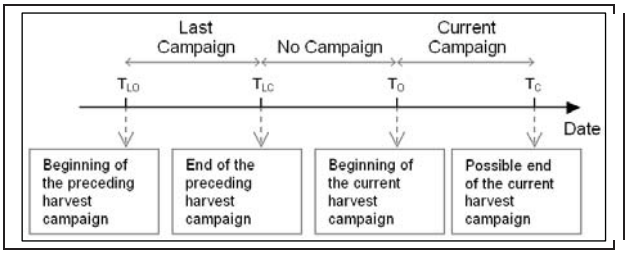


Figure 2: Temporal intervals used for the classification of image acquisition dates. T_{LO}, T_{LC}, T_O and T_C are opening and closure dates of the preceding and the current campaign respectively.

A thorough description of the inputs can be found in [18].

4 Decision system

The information coming from the three sources (SPOT5 time series, sugarcane growth model and expert knowledge) is heterogeneous and of different nature. The information extracted from the SPOT5 images is based on reflectance "measurements". The sugarcane model provides "estimates" using parameters that are not directly connected to the studied phenomenon (*i.e.* the field harvest). Expert knowledge contribution is neither "measurements" nor "estimates". Another important characteristic to highlight is that the sources are not redundant but complementary. That means that none of them is able to make a decision on its own; the cooperation is needed within the decision process. This discards usual aggregation techniques such as votes, OWA or fuzzy integrals. The general framework used to combine the different and complementary information sources is a fuzzy inference system.

The rules are of the following form: "*if X Then Y*", e.g.

$$\begin{aligned}
&\text{Rule}_1\text{: if } x_1 \text{ is } A_1^1 \text{ and } x_2 \text{ is } A_2^1 \ldots \text{ and } x_n \text{ is } A_n^1 \text{ then } y \text{ is } B^1\\
&\text{Rule}_2\text{: if } x_1 \text{ is } A_1^2 \text{ and } x_2 \text{ is } A_2^2 \ldots \text{ and } x_n \text{ is } A_n^2 \text{ then } y \text{ is } B^2\\
&\quad\vdots\\
&\text{Rule}_r\text{: if } x_1 \text{ is } A_1^r \text{ and } x_2 \text{ is } A_2^r \ldots \text{ and } x_n \text{ is } A_n^r \text{ then } y \text{ is } B^r
\end{aligned} \quad (2)$$

where A_k^r is the fuzzy set of the k^{th} input variable involved in the rule r, and B^r is the conclusion of the r^{th} rule.

The rule conclusion is a symbolic label, related to the field status. Two are possible: H (harvested) or NH (not harvested).

For each input universe a partition is defined. Some of the input partitions are crisps (see Fig.2), the other are fuzzy. In this case, we built standardized partitions according to expert knowledge. Fig.3 illustrates the fuzzy sets of NDVI and an example of NDVI profiles plotted according to thermal time for several sugarcane fields. NDVI fuzzy sets were designed according to expert knowledge about the phenology and field status of sugarcane as well as about its NDVI temporal profiles.
Linguistic labels are related to the NDVI:

- "Low" NDVI values (< 0.30) generally correspond to residues and bare soil after field harvesting.
- "Medium" NDVI values (between 0.30 and 0.75) are observed in the growth and senescence phases. The growth phase is characterized by an accumulation of biomass and an increasing length of the stems. The senescence is caused by the sugarcane ageing, the reduction in the temperature and the lack of water.
- "High" NDVI values (>0.75) are obtained at the end of the growth stage and before senescence. In this phase sugarcane fields are well developed and green covered.

The inference technique used in our system is based on Mamdani's method [19]. The weight attributed to the conclusion of an activated rule is calculated by combining the membership degrees of rule premises in a conjunctive way using the *min* operator. The aggregation of the distinct conclusions of the activated rules is done in a disjunctive way using the *max* operator.
Once system inputs and output as well as inference parameters are defined, the last step is reasoning rule description. In a previous version [18], the rules were defined according to expert knowledge. 116 rules were needed to cover all possible situations. The system proved to be accurate. A comparison between fuzzy and crisp partitions showed that the former ones increase a 15% the overall accuracy. The rules can also be generated by an automatic procedure. A fuzzy decision tree induction algorithm, which is a fuzzy extension of the well known ID3 algorithm [20], is used. The main advantage of the decision trees is to generate incomplete rules, only defined by a subset of the available input variables. The generated rules are informative for experts to the condition that the partitioning is carefully defined. The Fispro [21] implementation is used with the partitions proposed by the experts, those which proved efficient in the expert fuzzy system.

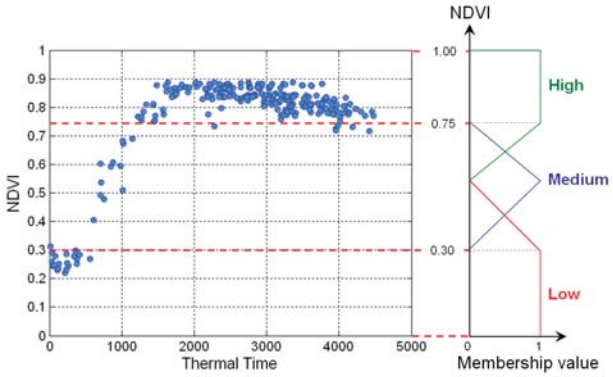


Figure 3: NDVI profiles plotted according to thermal time for several sugarcane fields. On the right are the fuzzy sets of NDVI-defined inputs.

5 Results and discussion

The learning process includes a cross validation. It has been carried out with various configurations to assess the system sensitivity to important parameters: the size of the training set and the number of satellite images per year. Finally, the contribution of each of the information sources is evaluated.
The data set has been randomly split to build training and test sets. For each proportion, the given result is the

ISBN: 978-989-95079-6-8

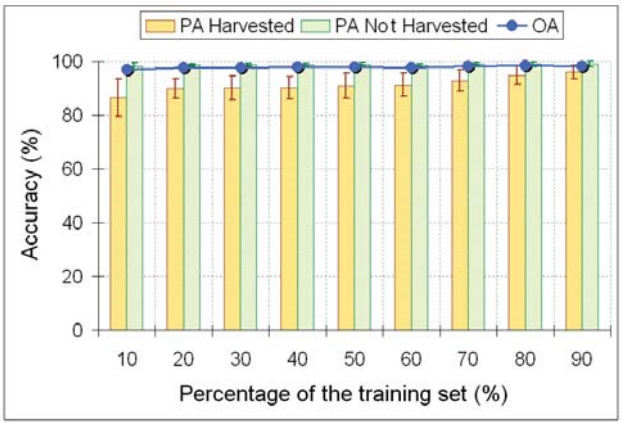


Figure 4: System performances using different percentages of training set (OA: overall accuracy; PA Harvested and PA Not Harvested: producer's accuracy for the two classes "Harvested" and "Not Harvested" respectively). These performances are obtained from confusion matrices.

average of ten runs. Nine percentages have been tested for the training set: 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% and 90%. The performance is measured over the test set.

Fig.4 shows the average values and the standard deviations (for ten iterations) of the overall accuracy (OA) of the system and the producer's accuracy for the classes "Harvested" and "Not Harvested" (PA Harvested and PA Not Harvested, respectively). Let's notice that the performance for the Not Harvested decision is quite independent from the size of the training set, while the Harvested one increases with the sample size. This is due to the respective proportions of both cases in a given image: the not harvested fields are much more numerous than the harvested ones. So, only a small number of images is needed to learn the not harvested case.

The other key parameter related to system robustness, is the number of available images. It is likely to vary from a year to another because of the lack of acquisition or cloudy weather which makes unusable the satellite images.

Fig.5 shows the selected images for each of the studied years. The selection takes into account the campaign organisation. For instance, in the case of 3 images per year, one is chosen before, the other between and the last one after the harvest campaign.

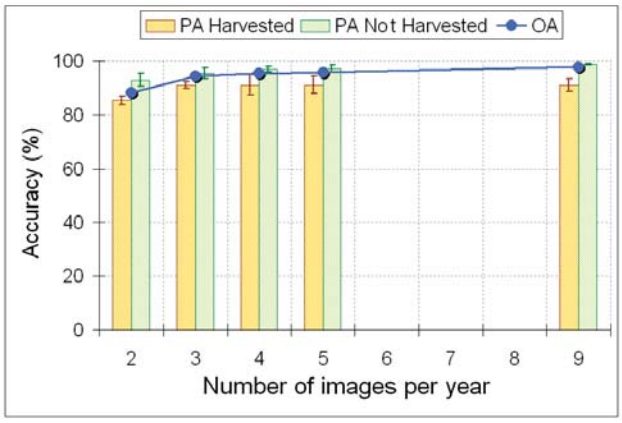


Figure 6: System performances according to the number of satellite images per year. (OA: overall accuracy; PA Harvested and PA Not Harvested: producer's accuracy for the two classes "Harvested" and "Not Harvested" respectively).

Fig.6 shows that the system performance is still acceptable even when the number of images per year drops to 3.

The number of generated rules also evolves with the number of images per year. It reaches an average of 38 for the 5 image configuration as shown in Fig.7.

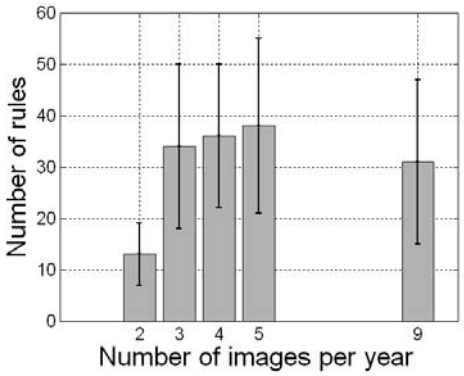


Figure 7: Number of induced rules according to the number of satellite images per year. Histogram bars represent the average values for the different training sets, and error bars the standard deviation.

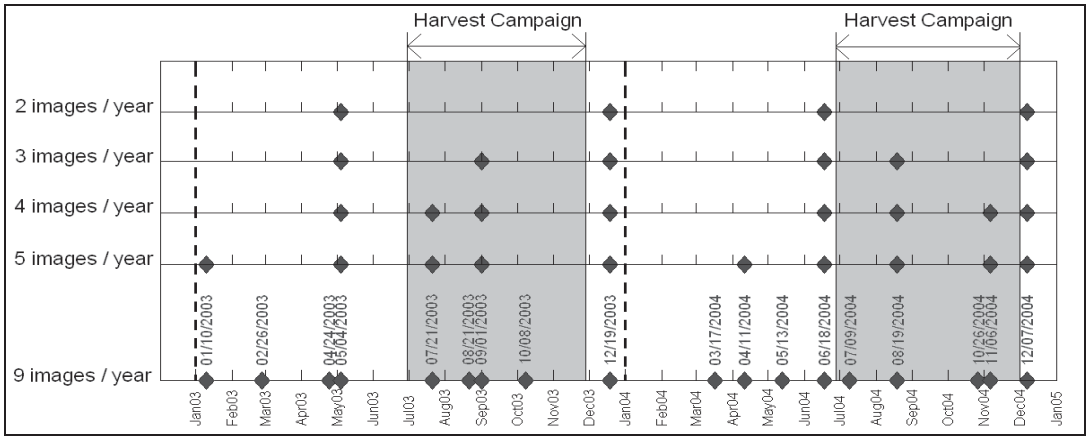


Figure 5: Acquisition dates selected at different stages of the assessment of system robustness with respect to the number of images per year.

ISBN: 978-989-95079-6-8

To assess the information source contribution three sets of input variables have been considered:

- SAT: the set of input variables that are related to the information extracted from the satellite images;
- MOD: the set of input variables built using the sugarcane growth model;
- EXP: the set of some input variables dealing with expert knowledge. Although the expert knowledge is difficult to isolate since it contributes to the design of the whole system, some variables are only related to this sources. They are encoding agronomic knowledge such as the nominal cycle length.

Four combinations have been tested and the results are reported in Fig.8. The best ones are the closest to the origin. As expected, all of them include SAT and MOD, and in most cases, the best one is SAT+MOD+EXP.

The sugarcane growth model contribution is interesting to analyze for each of the two classes "Harvested" and "Not Harvested". Fig.9 shows that it is higher in the harvested case. This result was expected: there is no need of a growth model to label a field as not harvested. The essential role of the model is to remove potential ambiguity concerning harvested fields.

6 Conclusions

This paper presented a novel approach for dealing with time series of optical satellite images used for crop monitoring. Data extracted from time series were combined with information from crop model output and expert knowledge, in order to make credible decisions. The description of the approach was done using an application example of sugarcane harvest monitoring with a SPOT5 time series. A decision system designed and implemented for automatic harvest detection was described.

Results obtained when evaluating the system were in substantial agreement with ground truth data; the overall accuracy reached 96.07%. The next step concerning the sugarcane application consists in examining the robustness of the automatically generated rules by testing the system at other sites and in other years.

The approach outlined in this paper is generic and very promising. Many models that simulate the growth of the main annual crops exist (e.g., STICS [22]), and expert knowledge about these crops could be obtained easily either from farmers or from agronomic knowledge bases [23].

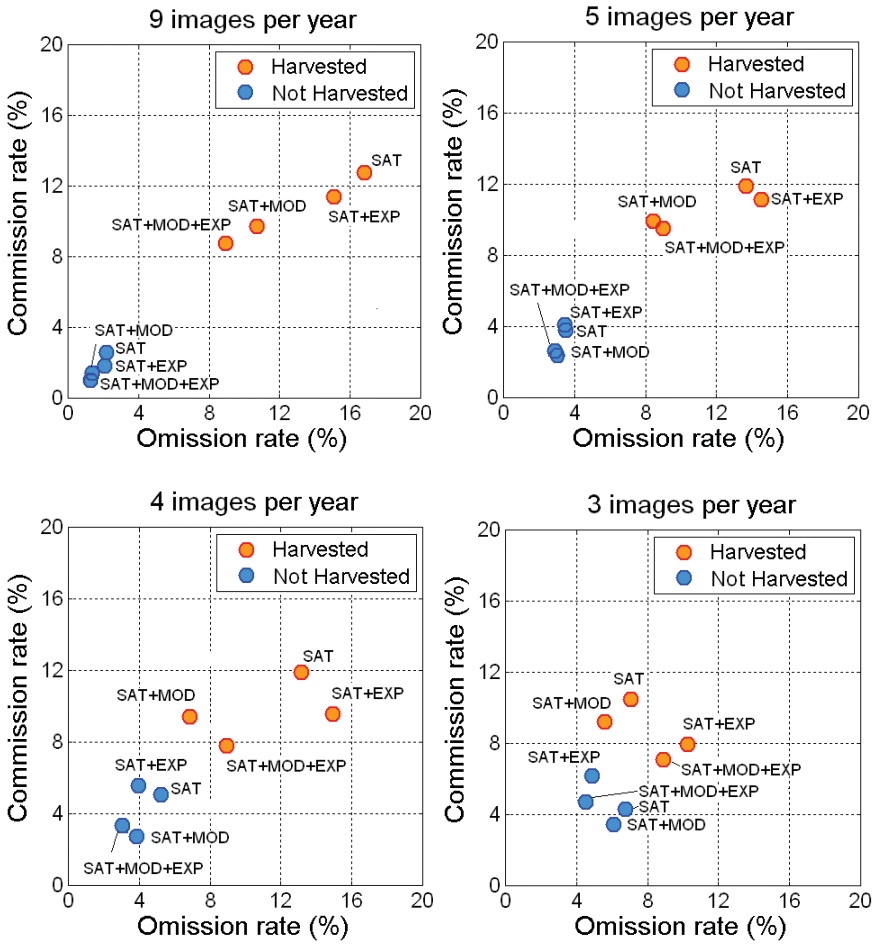


Figure 8: System performances obtained for different image frequencies by using different combinations of information sources.

ISBN: 978-989-95079-6-8

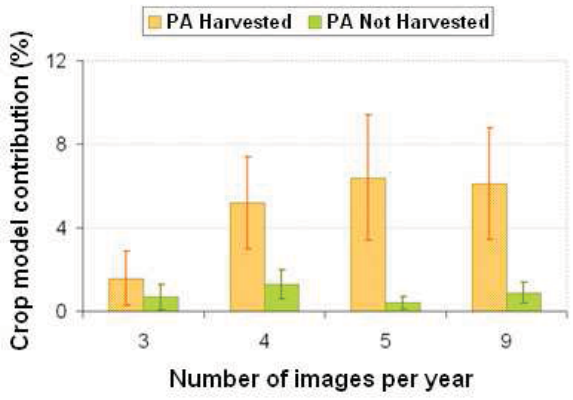


Figure 9: Crop growth model contribution in the system performances (PA Harvested and PA Not Harvested: producer's accuracy for the two classes "Harvested" and "Not Harvested" respectively). Histogram bars represent the average values for the different training sets, and error bars the standard deviation.

The combination of crop model outputs and expert knowledge with time series of high spatial-resolution satellite-images seems to be an excellent tool for crop monitoring.

Acknowledgment

We would like to thank Hélène de Boissezon and Sébastien Garrigues (CNES) for their help and support. We would also like to thank the CNES for funding the Kalideos database (http://kalideos.cnes.fr) and for generously providing the SPOT5 images.

References

[1] M. Pax-Lenney and C. E. Woodcock, "Monitoring agricultural lands in Egypt with multitemporal landsat TM imagery: How many images are needed?," *Remote Sensing of Environment*, vol. 59, pp. 522-529, 1997.

[2] G. A. Ippoliti-Ramilo, J. C. N. Epiphanio, and Y. E. Shimabukuro, "Landsat-5 Thematic Mapper data for pre-planting crop area evaluation in tropical countries," *International Journal of Remote Sensing*, vol. 24, pp. 1521-1534, 2003.

[3] C. S. Murthy, P. V. Raju, and K. V. S. Badrinath, "Classification of wheat crop with multi-temporal images: Performance of maximum likelihood and artificial neural networks," *International Journal of Remote Sensing*, vol. 24, pp. 4871-4890, 2003.

[4] M. D. Turner and R. G. Congalton, "Classification of multi-temporal SPOT-XS satellite data for mapping rice fields on a West African floodplain," *International Journal of Remote Sensing*, vol. 19, pp. 21-41, 1998.

[5] J. A. Martinez-Casasnovas, A. Martin-Montero, and M. A. Casterad, "Mapping multi-year cropping patterns in small irrigation districts from time-series analysis of Landsat TM images," *European Journal of Agronomy*, vol. 23, pp. 159-169, 2005.

[6] S. Panigrahy and S. A. Sharma, "Mapping of crop rotation using multidate Indian Remote Sensing Satellite digital data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 52, pp. 85-91, 1997.

[7] L. Bruzzone and P. Smits, "Analysis of Multi-temporal Remote Sensing Images : Multitemp 2001," in *Series in remote sensing*, vol. 2, A. P. Cracknell, Ed.: World Scientific Publishing Co. Pte. Ltd., 2002, pp. 440.

[8] P. Smits and L. Bruzzone, "Analysis of Multi-temporal Remote Sensing Images : MultiTemp 2003," in *Series in remote sensing*, vol. 3, A. P. Cracknell, Ed.: World Scientific Publishing Co. Pte. Ltd., 2004, pp. 387.

[9] C. Gers and E. Schmidt, "Using SPOT4 satellite imagery to monitor area harvested by small scale sugarcane farmers at Umfolozi," presented at 75th South African Sugar Technologists' Association (SASTA), 2001.

[10] E. Bappel, A. Bégué, P. Degenne, V. Lebourgeois, and B. Siegmund, "Real time sugarcane harvest monitoring using SPOT4&5 satellite data," presented at Proceedings of International Society of Sugar Cane Technologists (ISSCT), Guatemala, 2005.

[11] A. Bégué, P. Degenne, A. Pellegrino, P. Todoroff, and F. Baillarin, "Application of remote sensing technology to monitor sugar cane cutting and planting in Guadeloupe (French West Indies)," presented at Geomatica, Cuba, 2004.

[12] H. DeBoissezon and A. Sand, "Reference Remote Sensing Data Bases: Temporal series of calibrated and ortho-rectified satellite images for scientific use," presented at Proceedings of Recent Advances in Quantitative Remote Sensing, Valencia, Spain, 2006.

[13] http://kalideos.cnes.fr, "Images Spot: copyrigth CNES, Distribution Spot Image," 2007.

[14] M. El Hajj, A. Bégué, B. Lafrance, O. Hagolle, G. Dedieu, and M. Rumeau, "Relative radiometric normalization and atmospheric correction of a SPOT 5 time series," *Sensors*, vol.8, pp. 2774-2791, 2008.

[15] M. El Hajj, M. Rumeau, A. Bégué, O. Hagolle, and G. Dedieu, "Radiometric normalization of high spatial resolution multi-temporal imagery: A comparison between a relative method and atmospheric correction," presented at Proceedings of SPIE Europe Remote Sensing - The International Society for Optical Engineering, Florence, Italy, 2007.

[16] J. F. Martiné and P. Todoroff, "Le modèle de croissance Mosicas et sa plateforme de simulation Simulex: état des lieux et perspectives," *Revue Agricole et Sucrière de l'Ile Maurice*, vol. 81, pp. 133-147, 2002.

[17] M. El Hajj, A. Bégué, and S. Guillaume, "Multi-source information fusion: Monitoring sugarcane harvest using multi-temporal images, crop growth modelling, and expert knowledge," presented at MULTITEMP-2007, Fourth International Workshop on the Analysis of Multi-Temporal Remote Sensing Images, Provinciehuis Leuven, Belgium, 2007.

[18] M. El Hajj, A. Bégué, S. Guillaume, and J. F. Martiné, "Integrating SPOT-5 time series, crop growth modeling and expert knowledge to monitor agricultural practices - Case of sugarcane harvest on Reunion Island," *Accepted in Remote Sensing of Environment*, 2009.

[19] E. H. Mamdani and S. Assilian, "Experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, pp. 1-13, 1975.

[20] J.R. Quinlan, "Induction of decision trees," *Machine learning*, vol.1, pp. 81-106, 1986.

[21] S. Guillaume, B. Charnomordic, and J.-L. Lablée, "FisPro: An open source software for fuzzy inference systems," http://www.inra.fr/internet/Departements/MIA/M/fispro/, Ed.: INRA-Cemagref, 2002.

[22] INRA, "STICS : Simulateur mulTIdisciplinaire pour les Cultures Standard," http://www.avignon.inra.fr/stics/.

[23] G. Russell, R. I. Muetzelfeldt, K. Taylor, and J. M. Terres, "Development of a crop knowledge base for Europe," *European Journal of Agronomy*, vol. 11, pp. 187-206, 1999.

ISBN: 978-989-95079-6-8

Soft Computing Confronting Philosophical and Sociological Critiques to Classical AI

Verónica Sanz

Visiting Graduate Student
European Centre for Soft Computing
Mieres, Asturias (Spain)
E-mail: veronica.sanz@softcomputing.es

Abstract— *Artificial Intelligence has been one of the fields within Computer Science that has generated more interest and debates among philosophers. Later on, the most recent field of Science and Technology Studies (STS) also has shown some interest in AI. In both cases most of the authors have been quite critical about the promises, the practices and particularly the epistemological basis of Classical AI. The first part of the paper consists on an enumeration of the most important authors and their critiques to AI in Philosophy and STS studies. Since Soft Computing implies important changes with respect to traditional AI approaches like Symbolic AI, the second part of the paper will be devoted to confront Soft Computing with the critiques and challenges and to weight up to what extent Soft Computing could (or could not) answer differently than other AI approaches regarding the critiques received.*

Keywords— Epistemology, Philosophical critiques to AI, Science and Technology Studies (STS), Soft Computing.

1 Introduction

Since its origins Artificial Intelligence has been one of the fields within Computer Science that has generated more interest and debates among philosophers and cognitive scientists. Not in vain, Artificial Intelligence deals with some of the traditionally most debated issues in philosophy such as knowledge and reasoning. In concrete, some of the ambitious statements of the first AI researchers touch very sensitive "fibers" of philosophers. Authors like John Searle, Daniel Dennett, Hubert Dreyfus or Roger Penrose were among the first ones to cope with these issues [1, 2, 3, 4, 5].

The philosophical debate during the 1970s and 80s focused on Artificial Intelligence via Symbolic AI (the branch of artificial intelligence research that concerns itself with attempting to explicitly represent human knowledge in a declarative form -i.e. facts and rules-), since this was the approach followed by the main scientists in the field from the early days on. The main interest of philosophers has been traditionally centred in the question about whether computers can or cannot think, which generally implies doing it "in the sense humans do". There was an assumption among philosophers (though induced by claims made by some of the first AI researches) that the aim of AI is to create "an artificial mind". Despite the successes of some AI symbolic systems during the 60s, the enormous promises of early AI texts posed the debate in terms of the possible success or failure of the whole enterprise of creating an artificial mind.

Because of that the discussion was pursued in a very bi-polar manner with some philosophers (like Dennett) clearly optimistic about the question and others (like Searle and Dreyfus) very critical about it. The philosophical critique was in tune with a significant resistance in general culture, which tried to assure the specificity of human nature with regard to computational artifacts, often appealing to non-cognitive aspects of the mind (intentionality, emotion, perception, etc).

2 Philosophical Critiques

2.1 Turing

The British mathematician Alan Turing was not only the one who set the basic elements of computation, but also the one who posed the first stone of this debate about the possibility of "thinking machines", proposing a mental experiment known, since then, as the "Turing test" [6]. It consists on an interrogator asking questions (through a separating barrier) to a machine and a human, trying to guess who (which) is answering. If the interrogator cannot tell whether the responder is the human or the machine, then the intelligence of the machine would be proved.

The interest of Turing's idea lies on the practical way he is proposing the test as an "imitation game". He is not defining, at least *a priori*, what intelligence *is* but establishing a practice were the performance of human intelligence is showed. Whether the machine thinks "in the way humans do" or not is not addressed by the Turing's test. In fact, it is precisely this kind of philosophical question what he tries to avoid.

2.2 Searle

In his famous paper from 1980 [1] the American philosopher John Searle argued against the strong claim that computers can be intelligent *in the same way* humans are, which for him implies having cognitive or psychological states (i.e. conscious understanding and intentionality). He distinguishes two strands of AI. The former claim is called "Strong AI": intelligent machines would replicate not only intelligent *products,* but the intelligent *process*[1]. By contrast, "Weak AI" is the perspective that understand computers as useful tools that mimic some of the abilities and tasks of humans (included those that are considered intelligent), in part because they can simulate "smart actions"[2]. Weak AI does not make the claim that computers *actually understand* or *have* mental states.

Searle's critique is directed to Strong AI. To do that, he proposes a thought experiment known as "The Chinese room" in response to Alan Turing's suggestion to replace the question "Can machines think?" with the question of whether

ISBN: 978-989-95079-6-8

they can succeed in an "imitation game". In short, Searle's conclusion is that a native English speaker who knows no Chinese together with a book of instructions for manipulating the symbols would able to pass out Chinese symbols which are correct answers to the questions. The instructions enable him/her to pass the Turing Test but he does not understand a word of Chinese[3].

The axis of the argument relies also on the idea that *syntactic manipulation* (formal symbol manipulation or computation) does not imply understanding *semantics* (understanding of meanings). For Searle, we can not attribute intelligence to such a system since "real intelligence" is related with understanding. The background argument of the Chinese room is that one of the main features of human minds is intentionality (which is a mental state) and that is something that computers do not (and can not) have.

2.3 Dennet

Though many of the philosophers of AI have been very critical, Daniel Dennett is an exception [2]. He shared with Symbolic AI researchers what in philosophy of mind and cognitive sciences is called the "Computational (and Representational) Theory of Mind". Dennett has been positively linked to AI and he has participated in some AI projects. He disagrees with Searle in two important points:

a) The way that Searle understands intentionality is already assumed from the beginning as limited to humans. Against that, Dennett affirms that human consciousness and intentionality are not as specific as we use to think, which is showed in the way we use a "mentalist language" when describing the actions of animals, machines and so on.

b) The Chinese Room thought experiment is very ingenious but is very far from actual practices in everyday life. It involves a number of things that are quite impossible to do so it is very doubtful to be used as a proof of something[4].

2.4 Dreyfus

However, the philosopher who most energy and time has devoted to the critique of AI during the past 30 years has been the phenomenologist Hubert Dreyfus [3, 4]. Drawing upon Heidegger and existential phenomenology, his arguments are quite different from Searle's (though he also affirms that neither current machines nor future ones would be intelligent).

Dreyfus criticizes the epistemological assumption of Symbolic AI that all activity can be formalized in the form of propositional knowledge (which would allow computation in the form of "rule-based symbol manipulation") as it is done in many Symbolic AI based expert systems. For Dreyfus, many parts of human knowledge and competences cannot be reduced to an algorithmic procedure, i.e. the kind of knowledge one might employ when learning a skill like walking, riding a bike or ice skating. This learning is done through practice with our bodies: to learn how to do it one must get on and try it, it's not enough to follow written rules. This requires having a body and experimenting with it in a changing environment.

Though Dreyfus concedes that certain well-circumscribed experts systems for particular domains have succeeded, he predicts that many others will never fulfil the expectations (i.e. Lenat y Guha's "Cyc" [7]). This is not because of technical shortages (that could be overcome in the future) but because symbolic type of artificial systems are incapable of representing bodily-skills knowledge, which can not be captured in an algorithm.

3 Soft Computing confronting Philosophical Critiques

3.1 Not related to Strong AI

The problem with Searle's thought experiment is that most of practitioners in AI do not take issue with intentionality or creating an artificial mind. This debate is in fact a philosophical one (the old debate between Behaviorism and Mentalism[5]) and is far from the interests of Soft Computing (SC) practitioners, as far as I could observe. Although this debate was in part provoked by the excessively optimistic predictions of some of the founding fathers of AI, it can be said that the SC community is more focused on solving concrete problems by constructing machines that can perform "intelligent" tasks, which is more in tone with what Searle called "Weak AI". Because of that, these kind philosophical critiques do not apply to Soft Computing.

However in one of his last articles, Zadeh says that there is a possibility of achieving "human-level machine intelligence", though for that there is needed a paradigm shift [8].

3.2 Necesity of Semantics

The other axis of Searle's critique relied on the idea that *syntax* (formal symbol manipulation or computation) does not imply *semantics* (understand meanings). Searle applied it to say that, to have "real intelligence", syntax is not enough. One of the fundaments of Fuzzy Logic is precisely to deal with meaning in natural language. Fuzzy Logic agrees with Wittgenstein's definition of meaning as its use in language [9], that is, as used by concrete people in concrete contexts.

To model and operate with natural language and its non-static meanings would require tools which can deal with imprecision and uncertainty, which is precisely the main characteristic of Fuzzy Logic. The enormous importance of natural language for SC is shown in the turn Zadeh proposed in 1996 saying that Fuzzy Logic was equivalent to "Computing with Words*"* [10]. Computing with words (CW), Zadeh says, is a methodology in which words are used in place of numbers for computing and reasoning. As fuzzy sets are able to model some natural language's concepts, it would be possible to use these models to compute directly with them.

3.3 The "Bodily-skills knowledge" Problem

Dreyfus's main critique to symbolic type of artificial systems is its incapability of representing bodily-skills knowledge. Other types of AI systems have tried to avoid this problem, like Rodney Brooks' "bottom-up" strategies in *embedded* or *situated robotics* [11]. The work of Brooks and his colleagues focus on *embodiment,* the idea that intelligent behaviour is an *emergent* phenomenon resulting from the

ISBN: 978-989-95079-6-8

individual's interactions with its environment, which are more related to action than to representation.
What Soft Computing has to say about this debate? In general we could say not much. The problem of embodiment is typical in Robotics (where there are "embodied machines", not only software) but not necessarily in most of the branches of AI. Fuzzy Logic indeed deals with language and representation of meanings and not so much with actions. However, applications of Fuzzy Logic to concrete systems (fuzzy control) do actually deal with systems-in-action. One clear example is the "inverted pendulum" solved by Yamakawa [12] using a small set of simple fuzzy rules. Fuzzy control system design is based on empirical methods, basically an approach of trial-and-error, which is quite related to the way humans learn bodily-skills knowledge.

4 STS Critiques

Among the different non-technical disciplines not only philosophy has been interested in the study of AI. Within the so-called area of STS, at least some practitioners have turn around to Artificial Intelligence.
An STS methodology implies the analysis of the social and cultural contexts where all scientific and technological developments take part. This model resists classical accounts that science and technology are totally objective and neutral with respect to the context where they are made.
Regarding AI, STS takes a different stand on that philosophical critique. They don't take part in abstract discussions about the ultimate goal of AI. Their methodology consists on looking at real practices, including ethnographic observation of AI practitioners. From the STS view, the important thing is the way these engineers understand and use (or model) concepts such as knowledge and reasoning in their daily practice.

4.1 Harry Collins

Harry Collins' study of AI [13] is the most important contribution from part of STS to the study of AI. The field of Experts Systems has attracted the most attention from part of social scientists. His critique is in a way quite similar to Dreyfus' regarding the impossibility of formalizing certain types of knowledge.
Based on recent Sociology of Scientific Knowledge [14] and Laboratory Studies [15], in which is shown how experimenters need to share a culture in order to produce science, Collins maintains that skills (know-how knowledge) are principally acquired through enculturation and the acquisition of tacit capacities whose transmission is implicit (and cannot, therefore, fully be spelt out in formal algorithms). This kind of knowledge it is commonly known in STS as "tacit knowledge", which can be also understood as the "common knowledge" of a particular community (what is taken for granted within it)[6].
However, Collins understands "tacit knowledge" more widely than Dreyfus' embodied knowledge. Acquisition of skills-type knowledge in his view is related to socialization rather than embodiment. The underlying argument of Collins against Dreyfus is the importance of natural language. Collins affirms that one can learn to speak fluently about the world without being bodily involved in it. This model assumes that socialization (in a certain group) is necessary to communicate knowledge. In this way, Collins brings the discussion to the cultural arena
Another of the issues addressed by Collins is his focus on AI systems that are already being used and seem like functioning, not in possible future ones. The challenge is, then, to explain that "success". On the contrary of philosopher's understanding of computers as artificial brains, to him the role (and success) of AI is to create "social prostheses". And it is the ability of humans to interpret their actions and repairing their deficiencies which makes them successful. In a way, humans interpret machines actions as one kind of actions humans can perform (what Collins calls "machine-like action"), but this is only a little part of the whole kind of humans' actions. This human ability to interpret and repair normally remains invisible, but, for Collins, is here where the intelligence attributed to machines rests on.

4.2 Diana Forshyte

The work of Diana Forsythe [16] belongs to the field of Anthropology of Science[7] rooted in cultural anthropology's tradition. This normally implies a more or less long-term field research using the ethnographic method of participant observation. In her ethnographic study of an Expert Systems community in the US, she investigated what she called the "Culture of the Experts Systems Engineers", defined as the values and assumptions (explicit but mostly implicit) that constitute what they take for granted. Forshyte also refers to this as "common sense truths", which are those things 'everybody knows' within a given setting (this shared and tacit knowledge is indeed the main criterion to be considered a member of the community). The importance of the ethnographic method relies on the main thesis of cultural anthropology about the complex relation between beliefs and practices: usually what people do, what they think they do, and what they report they do are someway different.
Investigating the culture of the Expert Systems practitioners she found they share a particular notion of knowledge (a very restricted one), that has important implications for the products they make. Their epistemological stances are very important because they are going to be incorporated (encoded) into the expert systems produced. To prove this she focused on the first step of the process of building an expert system: 'knowledge acquisition' (also known as elicitation process), consisting in the "extraction" of the knowledge from human experts. The very use of the terms "extraction" and others such "store" (knowledge is presumed to be "stored" in the head of the experts) are metaphors which implies an understanding of knowledge as objective and susceptible of being formally transferred into a machine.
Despite that, the elicitation process is considered by knowledge engineers themselves a very persistent problem and is thought to be the cause of failures in the systems when encountering 'real-world situations' (situations that the systems builders did not anticipate). Knowledge engineers blame human experts interviewed for not being able to explicit clearly their expert knowledge, that is, in the way they can encode it (it is not casual that the kind of expert

ISBN: 978-989-95079-6-8

knowledge they prefer to work with are already those very formalized in well-defined narrow domains).
They would never think about the possibility that their implicit assumptions about knowledge can be the cause of their problems. Forshyte affirms that the knowledge encoded in experts systems is static, brittle and narrow, while in real life knowledge is continuously modified in relation to contextual factors and trough negotiations between different actors, and that is the real cause of the 'brittleness' of the systems[8].
In general terms, the kind of expert systems studied by Forshyte are included in the more general approach of Symbolic AI, since, in the end, knowledge "extracted" from human experts has to be written-down in rules-type propositional knowledge. With respect to this, the same critique from Dreyfus and Collins about the inability to model other types of knowledge (particularly social knowledge) is applicable relating Forshyte's work.

4.3 Alison Adam

One tradition within STS studies are those investigating the relations of gender factors with science and technology[9]. Notions of gender are constructed not only in relation to economy or moral values but also science and technology became in our culture more associated with the masculine than the feminine, reflecting cultural's view of what constitutes masculine or feminine. However this is not a fixed but a historical process[10].
Alison Adam is a British computer scientist and feminist who has pursued the most extended study about gender and Artificial Intelligence [17]. She accurately applies the insights of feminist epistemology (an alternative kind of epistemology which criticizes the classical conservative epistemologically) to the main traditions in AI as Symbolic AI, Expert Systems, and also more recent approaches such as Artificial Life and Situated Robotics. Adam's work is devoted to "uncover" the implicit (gendered) epistemological assumptions in AI, because on them depend how AI researchers understand some key concepts which very much influence their work such 'intelligence', 'knowledge', 'reasoning' and 'efficiency'. The idea, then, is not to look for the contributions of individual women in the history of AI, but rather to look if there can be gendered models of knowledge implicit in AI systems.
One of Adam's theses is that traditional AI systems are based on the Cartesian ideal of a disembodied mind and the over-valuation of mental (abstract) knowledge over corporeal (concrete) knowledge; the former has been historically associated with masculine realm while the latter has been largely associated with the feminine. For Adam as for other feminist epistemologists [18, 19], this is very much related to the "mathematization" of the Western Culture (started in ancient Greece and increased in the Modern Age) and the idea that the entire universe was governed by numbers. Although there was already among the Greeks a plea for other ways of knowing it, was this Platonic and then Aristotelian conception of the world which finally succeeded. Aristotle is considered "the father of Logic", and defines the correct way of reasoning in terms of a dichotomous logic (there are only true or false statements which imply only valid or invalid inferences). As we know, the rejection of this thesis was the starting point of Zadeh's Fuzzy Logic.
Following the work of the philosopher Lorrain Code [20], Adam shows that traditional AI is based on the rationalistic epistemological model of "S knows that p". 'S' refers to the knowledge subject (*who knows*), and 'p' the object of that knowledge (w*hat can be known*).
Regarding 'p', as we already said, propositional knowing (or *knowing that)* is considered to be superior knowledge while others (skill-knowledge or *knowing how*) are less important. Feminist epistemologies consider this stratification as an *epistemic discrimination* (which they affirm is not innocent of gender factors).
Regarding 'S', the ideal knower of Classical AI is an anonymous, universal and "knowing from nowhere" subject[11]. On the contrary, Adam emphasizes the "situatedness" of the knower. For example, who are the "knowledge subjects" of systems like "Cyc"? This project leaded by Douglas Lenat [7] tries to build a vast knowledge base which will include most of human common-sense knowledge. Above all, Lenat assumes that there is one *and only one* consensus reality available common to all humans. The development of STS and Anthropology is the last decades have shown that it is quite problematic to assume that hypothesis. We "see" the world according to our theories of the world, and these theories depend on our cultural location. At bottom, what CYC knowledge base would consist on is what Adam's calls "*The-World-As-The-Builders-Of-Cyc-Believe-It-To-Be",* who are a group of predominantly middle-class male US university professors, that, without realizing it or intending it, are privileging their "consensus reality" over other groups' [17].
As a way of summarizing the differences between the two kind of epistemologies we can make the following table:

Table 1: Contrasting Epistemologies

Symbolic AI Epistemology	**STS Epistemology**
Objective Knowledge *(context free, universal)*	Situated Knowledge *(contextual, local)*
Propositional knowing (*knowing that*)	Skills-type and Common-sense Knowledge
Abstract reasoning	Concrete reasoning
Epistemological Monism	Epistemological Pluralism
Lack of Reflexivity	Accountability and Responsibility

5 Soft Computing confronting STS Critiques

5.1 Contextualism

As we said before, workers in the field of Symbolic AI claim that cognition is the manipulation of internal symbols by logical rules, so, therefore, human knowledge is, to a large extent, *context free*. We have explained how this idea has been criticized by authors like Dreyfus or Collins.
Wittgenstein's pragmatist turn in *The Philosophical Investigations*, which is assumed by SC, implies that meaning can only be grasped when used by concrete people

ISBN: 978-989-95079-6-8

in concrete contexts. The philosopher, Wittgenstein says, must "look and see" the variety of uses to which the word is put in particular cases. In this way, previous empirical analysis of the context of use it is an obligatory step for SC developments.
This obligates SC to be an e*mpirically-based research,* not only a theoretical approach (this is also because the ultimate goal of SC is engineering, that is, constructing machines, which in the end has to be used by users).
In spite that the idea of contextualism may be implicit in SC practitioners, the problem about the "situatedness" of the knowledge subject regarding the difference or culture, class, nation, race or gender has not been addressed as such neither in theory nor in specific projects, as far as I know.

5.2 The "Common-sense knowledge" Problem

As we have seen, the apelation to the inability of Classical AI to deal with common-sense knowledge is an important part of STS critiques. Their argument is that this kind of knowledge is mostly tacit and took for granted, and, because of that, it can not be formalized.
On the contrary, many SC articles contain the affirmation that fuzzy logic could represent common-sense knowledge. The most explicit article from Zadeh about the topic [21] explain that common-sense knowledge is expressed in *dispositions*, as opposed to propositions. A disposition in philosophy is a kind of belief that is stored in the mind but is not currently being considered (in a way we can say is a kind of tacit knowledge). For the purpose of representation Zadeh defines a disposition as a proposition with implicit fuzzy quantifiers, so in that way fuzzy logic can be used to deal with that kind of knowledge.
Although this strategy is only in its first steps, I think is a very promising way of dealing with common-sense knowledge, at least much better than Guha and Leant kind of propositional knowledge data base. However, there are many issues that are not addressed or solved yet in SC. A very important one has to do with the very definition of "common-sense knowledge" and its interrelations with concepts as "tacit knowledge", "skill-knowledge", "situated knowledge" and so on. Is it all common-sense knowledge tacit or just "taken for granted"? The difference between common-sense and skill-knowledge is that one can be formulated in *dispositions* and the other one not? What Zadeh exactly means when he talks about "world knowledge" –defined as "the knowledge humans acquire though experience, communication and education"?

5.3 Pluralism and Hybridization

Though we have not explicitly say it before, one of the main proposals from STS and feminist epistemologies is the appeal to alternative ways of knowing and doing science. Lucy Suchman, regarding computer science, talks about **"**expanding the frame of basic modeling assumptions" [22], which means expanding the very foundational paradigms of the discipline (particularly its assumptions about knowledge) and introduce alternative paradigms. This, regarding theory of knowledge, is what is called *Epistemological Pluralism*.
I think SC can be a little bit closer to these suggestions than other approaches. On one hand the basic characteristic of SC is precisely the combination of methods. Quoting Zadeh: *"SC is not a single methodology; rather is a consortium of methodologies which are aimed at exploiting the tolerance for imprecision (...) The constituent methodologies in SC are for the most part complementary and synergistic rather than competitive* [22]*".* So we can say pluralism and hybridization are at the very core of what SC is, but… do they mean with that the same as STS scholars when they mention it?
This question leads us to the following point.

5.4 Issues of Accountability and Responsibility

Pluralism in STS and feminist epistemology is related to issues of accountability and responsibility, which means that the ultimate goal is to introduce (some) changes in science and technology to make them more inclusive and socially just. Authors as Lucy Suchman [23] and Philip Agre [24] whose work is devoted to Artificial Intelligence, appeal to epistemological and methodological pluralism as a way of accommodating diversity. Agre calls this possible way of doing technology a *critical technical practice*, which consists on identifying core metaphors of the field, noticing what (when working with these metaphors) is marginalized, and inverting dominant metaphors bringing alternative perspectives from the margins of the discipline to the center.
As far as I could observe, Soft Computing practitioners have not introduce the issues of accountability and responsibility explicitly in their practice (on the other hand like most part of computer science and technology except but some particular approaches in software engineering like participatory design or user-centred design). However my suggestion is that, because of some particular characteristics (that we already pointed out), Soft Computing is more suitable to include this issues than other more abstract approaches such, for example, Symbolic AI.

6 Conclusion Remarks

Along the article I have enumerated various critiques made to traditional approaches in Artificial Intelligence from part of Philosophy and, later, from Science and Technology Studies. Soft Computing distinguishes itself very much from traditional AI approaches. Because of that, it appears to me as a very interesting topic to question how Soft Computing could cope with these critiques and challenges.
Although this research is still in a preliminary state, I think the issues treated in this article can provide a good frame for future work in the new field of Soft Computing in Humanities and Social Sciences. In addition to that, it could be also of some use for Soft Computing itself. Soft Computing have to face many challenges in order to improve other AI methodologies and try to achieve what they weren't able to. Having a clear picture of these challenges will be very useful. Philosophical and sociological studies of science could help to identify problems and challenges, and, although far from engineering practice, can inspire and suggest new ways of facing the most durable challenges.

ISBN: 978-989-95079-6-8

Acknowledgments

The author wants to thank Sergio Guadarrama for his comments and fruitful discussions about the topic of the paper.

In addition, I want to thank the Foundation for the Advancement of Soft Computing for providing me the best place and environment to proceed with my research in this topic.

[1] Strong AI is would be, then, the materialization of the "Computational Theory of Mind" (the philosophical view that holds that the mind is *literally* a digital computer and reasoning consists on formal manipulations of symbolic representations of the world). Through this thesis the debate became not only a philosophical one but also a psychological one: it provides a working model of human intelligence.

[2] In fact this is in tune with the most quoted definition of AI: *"AI is the ability of a computer or other machine to perform those activities that are normally thought to require intelligence"* [25]

[3] For a detailed description see *Stanford Encyclopedia of Philosophy's* entry "The Chinese Room Argument".

[4] Actually Dennett elaborated this critique in general against all kind of "thought experiments".

[5] That is, whether or not there are "internal mental states" in the mind that can be studied by psychologists, or only observable behavior.

[6] Very much related with a very important concept in STS known as "epistemic cultures"

[7] The pioneers of the application of Anthropology to the study of science are the so-called laboratory studies of Latour and Woolgar [15], Karin Knorr-Cetina, and Michael Lynch.

[8] As an anthropologist whose background understanding about knowledge is precisely the opposite (in Social Sciences knowledge is a highly problematic issue, mostly tacit and unconscious and whose "locus" is the a social group or community) is very shocking to confront herself with such a different approach.

[9] We should explain here what we mean with "gender" in gender studies. It is not about the gender of the particular individuals involved in science and technology (particular men or women) not even men and women in general, but the gender systems (defined as gender structures, gender identities and gender symbolisms) in which these science and technology is developed (in our case Western gender system), and how this system influence its process, for example gendered models of knowledge, which is the main topic of the so-called Feminist Epistemology.

[10] However there is no intrinsic reason why a technology may become associated with the masculine or the feminine gender, this may change over time. It is precisely in the contingence of this process where feminism locates its hope in change towards a more egalitarian relation between women and technology.

[11] This epistemology considers logic as the paradigm of valid knowledge. Logic characterizes precisely by depersonalizing the subject and being the paradigm of context-free knowledge.

References

[1] John Searle. Minds, brains and programs, *Behavioral and Brain Sciences* 3:417-457, 1980

[2] Daniel Dennett. *The Intentional Stance*, Cambridge: MIT Press, 1987

[3] Hubert L. Dreyfus. *What Computers Can't Do: The Limits of Artificial Intelligence.* New York: Harper and Row, 1972.

[4] Hubert L. Dreyfus. *What Computers Still Can't Do: A Critique Of Artificial Reason ,* MIT Press, Cambridge, MA, 1992

[5] Roger Penrose. *The Emperor's New Mind* Oxford: OUP, 1989.

[6] Alan Turing. Computing Machinery and Intelligence, *Mind* 59:433-460, 1950

[7] Douglas B. Lenat and R.V. Guha, Building Large Knowledge-Bused Systems: *Representation and Inference in the Cyc Project.* Reading, Mass.: Addison Wesley, 1990

[8] Lotfi A. Zadeh. Toward Human Level Machine Intelligence- Is It Achievable? The Need for a Paradigm Shift. *IEEE Computational Intelligence Magazine*, August 2008:11-22

[9] Ludwig Wittgenstein. *Philosophical Investigations (PI)*, G.E.M. Anscombe and R. Rhees (eds.), Oxford: Blackwell, 1953.

[10] Lotfi A. Zadeh. Fuzzy Logic = Computing with Words. *IEEE Trans Fuzzy Systems*, 4(2):103-111, 1996

[11] Rodney A. Brooks. Intelligence without representation, Artificial Intelligence 47:139–159, 1991.

[12] Takeshi Yamakawa., Stabilization of an inverted pendulum by a high-speed fuzzy logic controller hardware system, *Fuzzy Sets and Systems* 32:161–180, 1989.

[13] Harry M. Collins, *Artificial Experts: Social Knowledge and Intelligent Machines. Cambridge*, Mass.: MIT Press, 1990.

[14] David Bloor. *Knowledge and Social Imagery*, London, Routledge, 1976.

[15] Bruno Latour and Steve Woolgar. *Laboratory Life,* Beverly Hills: Sage Publications, 1979.

[16] Diana Forshyte "Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence", *Social Studies of Science*, Vol. 23, No. 3: 445-477, 1993

[17] Alison Adam. *Artificial Knowing. Gender and the Thinking Machine*. London and New York, Routledge, 1998

[18] Sandra Harding. *The Science Question in Feminism,* Ithaca, Cornell University Press, 1986.

[19] Evelyn Fox Keller. *Reflections on Gender and Science.* New Haven: Yale University Press, 1985.

[20] Lorraine Code, *What Can She Know? Feminist Theory and the Construction of Knowledge*. Ithaca, N.Y.: Cornel1 Univ. Press, 1999

[21] Lotfi A. Zadeh. Commonsense Knowledge Representation based on Fuzzy Logic, *IEEE*: 61-65, October 1983

[22] Lotfi A. Zadeh. Soft computing and fuzzy logic. IEEE Software 11(6):48-56, 1994.

[23] Lucy Suchman. *Human-Machine Reconfigurations. Plans and Situated Action* 2nd Edition, Cambridge: Cambridge University Press, 2006.

[24] Philip Agre. *Computation and Human Experience,* Cambridge: Cambridge University Press, 1997.

[25] Enric Trillas. *La inteligencia artificial. Máquinas y personas.* Madrid, Ed. Debate, 1998

Optimizing Fuzzy Flip-Flop Based Neural Networks by Bacterial Memetic Algorithm

Rita Lovassy[1,2] László T. Kóczy[1,3] László Gál[1,4]

1 Faculty of Engineering Sciences, Széchenyi István University Győr, Hungary
2 Inst. of Microelectronics and Technology, Kandó Kálmán Faculty of Electrical Engineering, Budapest Tech, Budapest, Hungary
3 Dept. of Telecommunication and Media Informatics, Budapest University of Technology and Economics, Hungary
4 Dept. of Technology, Informatics and Economy
Szombathely University of West Hungary
Email: lovassy.rita@kvk.bmf.hu, koczy@sze.hu, gallaci@ttmk.nyme.hu

Abstract—In our previous work we proposed a Multilayer Perceptron Neural Networks (MLP NN) consisting of fuzzy flip-flops (F^3) based on various operations. We showed that such kind of fuzzy-neural network had good learning properties. In this paper we propose an evolutionary approach for optimizing fuzzy flip-flop networks (FNN). Various popular fuzzy operation and three different fuzzy flip-flop types will be compared from the point of view of the respective fuzzy-neural networks' approximation capability.

Keywords— Bacterial Memetic Algorithm, feedbacked fuzzy J-K and fuzzy D flip-flops, Multilayer Perceptron Neural Networks

1 Introduction

Fuzzy set theory, artificial neural networks, bacterial evolutionary algorithms and their hybrid combination have been the subject of intense study and application, especially in the last decades. These were developed with the aim to deal with problems which were hard to solve using traditional techniques. Fuzzy systems are transparent and interpretable, neural networks possess the property of auto-adaptability, while evolutionary, especially bacterial algorithms have been used for the approximation of the optimal structure. To approximate various test functions we use a combination of the above mentioned three main branches of Computational Intelligence. Although some paper report about Support Vector Machines (SVM) outperforming traditional Multilayer Perceptron (MLP) in classification task, but in prediction and regression problems the MLP gives smaller errors with lower network complexity [12]. The paper is organized as follows. After the Introduction Section 2 defines triangular norms and co-norms; and then highlight the five well known fuzzy operations (algebraic, Yager, Dombi, Hamacher and Frank). Section 3 presents the concept of a single fuzzy J-K and D flip-flop, using the fundamental equation as it was proposed in [13]. A comparative study of several types of fuzzy J-K flip-flops based on the five fuzzy operations is made. Comparisons of fuzzy J-K flip-flops with feedback and of two different interpretations of fuzzy D flip-flops are presented. Section 4 is devoted to the investigation of the F^3 based neurons and the MLP [11] constructed from them. We proposed the Fuzzy Flip-Flop Neural Network (FNN) architecture that can be used for learning and approximating various simple transcendental functions. In Section 5 we present Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM) [5] for FNN variables optimization. Experimental results and comparison between different types of FNNs trained with Levenberg-Marquardt (LM) method and the BMAM are discussed in Section 6, followed by a brief conclusion in the last Section.

2 Fuzzy operations

Klement, Mesiar and Pap enumerate and give the basic definitions and properties of the most general triangular norms in [8], including also graphical illustrations, comparisons. In general a triangular norm [8] (t-norm) is a binary operation T on the unit interval [0, 1], i.e., a function $T:[0,1]^2 \rightarrow [0,1]$, such that for all $a, b \in [0,1]$ the commutativity, associativity, monotonicity and boundary conditions are satisfied. Triangular co-norms (s-norm) were introduced [8] as dual operations of t-norms.

Table 1 shows various t-norms and s-norms selected by us, (as possible bases of the FNN), all five can be used as basic operators in the definition of the fundamental formulas of the fuzzy flip-flop introduced in the next Section. (Many other did not produce good enough neural network properties, cf. [10].) For simplicity we denote the t-norm with i, and the s-norm with u, where the subscripts refer to the initial of the name of the norm: e.g., in case of Yager norms: $i_Y(a,b) = a\ i_Y\ b$ and $u_Y(a,b) = a\ u_Y\ b$.

3 Fuzzy flip-flops

3.1 Fuzzy J-K Flip-Flops

The fuzzy J-K flip-flop is an extended form of the binary J-K flip-flop. In this approach the truth table for the J-K flip-flop is fuzzified, where the binary NOT, AND and OR operations are substituted by their respective fuzzy counterparts, i.e. fuzzy negation, t-norm, and co-norm respectively.

ISBN: 978-989-95079-6-8

Table 1: Some t-norms and co-norms.

Fuzzy operation	t-norm; i (a, b)	s-norm; u (a, b)
Algebraic (A)	ab	$a+b-ab$
Yager (Y)	$1-\min\left[1,((1-a)^w+(1-b)^w)^{1/w}\right]$	$\min\left[1,(a^w+b^w)^{1/w}\right]$
Dombi (D)	$\frac{1}{1+\left[(1/a-1)^\alpha+(1/b-1)^\alpha\right]^{1/\alpha}}$	$\frac{1}{1+\left[(1/a-1)^{-\alpha}+(1/b-1)^{-\alpha}\right]^{-1/\alpha}}$
Hamacher (H)	$\frac{ab}{\gamma+(1-\gamma)(a+b-ab)}$	$\frac{a+b-(2-\gamma)ab}{1-(1-\gamma)ab}$
Frank (F)	$\log_s\left[1+\frac{(s^a-1)(s^b-1)}{s-1}\right]$	$1-\log_s\left[1+\frac{(s^{1-a}-1)(s^{1-b}-1)}{s-1}\right]$

Parameters w, α, γ and s lie within the open interval $(0, \infty)$. The next state $Q(t+1)$ of a J-K flip-flop is characterized as a function of both the present state $Q(t)$ and the two present inputs $J(t)$ and $K(t)$. (For simplicity (t) will be omitted in the next.) The so called fundamental equation of the J-K type fuzzy flip-flop [13] is

$$Q(t+1) = (J \vee \neg K) \wedge (J \vee Q) \wedge (\neg K \vee \neg Q) \quad (1)$$

where $\neg, \wedge, \vee$ denote fuzzy operations (e.g. $\neg K = 1-K$). As a matter of course, it is possible to substitute the standard operations by any other reasonable fuzzy operation triplet (e.g. De-Morgan triplet), thus obtaining a multitude of various fuzzy flip-flop pairs.

In [9] we studied the behavior of fuzzy J-K flip-flops based on various fuzzy operations, and illustrated their behavior by the graphs belonging to the next states of fuzzy flip-flops for typical values of Q, J and K.

3.2 Fuzzy D Flip-Flops

Connecting the inputs of the fuzzy J-K flip-flop in a particular way, namely, by applying an inverter in the connection of the input J to K, case of $K = 1-J$, a fuzzy D flip-flop is obtained. Substituting $D = \neg K = J$ in equation (1), the fundamental equation of fuzzy D flip-flop will be

$$Q(t+1) = (D \vee D) \wedge (D \vee Q) \wedge (D \vee \neg Q) \quad (2)$$

As an alternative approach, Choi and Tipnis [4] proposed an equation which also exhibits the characteristics of a fuzzy D flip-flop, as follows

$$Q(t+1) = D \wedge (D \vee Q) \wedge (\neg Q \vee D) \quad (3)$$

We will refer to this type of fuzzy D flip-flop as Choi (type fuzzy) D flip-flop (because of the first author). Comparing the characteristical equation of the fuzzy D flip-flop (2), with expression (3), there is essential difference between the two fuzzy flip-flops.

As $D = D \wedge D$, and $D = D \vee D$ i.e. the t-norm (T) and co-norm (S) are idempotent [1], hold only in the exceptional case ($T(x, x) = x$ and $S(x, x) = x$ for all $x \in [0,1]$); when

T = min, and S = max.

For example, using the algebraic norm

$$u_A(a,a) = a + a - a \cdot a = 2 \cdot a - a^2 = a \quad (4)$$

is satisfied only in the two borderline cases, when $a = 0$, or $a = 1$. It is surprising how much the satisfaction of idempotence influences the behavior of the fuzzy D flip-flops, as it will be shown.

In the next Section we will give an overview of the different type fuzzy J-K, D and Choi D F^{3}s, based on familiar norms listed in Table 1.

3.3 Fuzzy Flip-Flops Based on Various Fuzzy Operations

The behavior of fuzzy flip-flops based on algebraic, Yager, Dombi, Hamacher and Frank t-norms, combined with the standard negation $c(a)$=1-a in every case has been analyzed and compared [9] , in order to investigate, whether and to what degree they present more or less sigmoidal (S-shaped) $J \to Q(t+1)$ transfer characteristics in the particular cases, when K=1-Q, K=1-J, with a fixed value of Q. F^3 based algebraic type t-norm presents non-sigmoidal behavior, with piecewise linear characteristics and several breakpoints, but having the advantage of the hardware implementation of F^3 [14], this type was also studied for comparison.

3.4 Fuzzy J-K, D and Choi D Flip-Flops Based on Some Classes of Fuzzy Set Unions and Intersections

Using a triplet consisting of the standard negation, and a dual pair of fuzzy operations, i.e., using Yager t-norm and s-norm, the maxterm form in the unified equation (1) can be rewritten as:

$$Q_Y(t+1) = \left(J \; u_Y \left(1-K\right)\right) \; i_Y \; \left(J \; u_Y \; Q\right) \; i_Y \; \left(\left(1-K\right) \; u_Y \left(1-Q\right)\right) \quad (5)$$

This is the fundamental equation of the Yager type fuzzy J-K flip-flop.

By substituting the above mentioned triplet into equations (2) and (3) we defined the fundamental equations of the Yager type fuzzy D (6) and Yager type Choi D (7) flip-flops.

$$Q_Y(t+1) = (D \; u_Y \; D) \; i_Y (D \; u_Y \; Q) \; i_Y \; (D \; u_Y \; (1-Q)) \quad (6)$$

$$Q_Y(t+1) = D \; i_Y (D \; u_Y \; Q) \; i_Y \; ((1-Q) \; u_Y \; D) \quad (7)$$

In a very similar way, in the cases of algebraic, Dombi, Hamacher and Frank operation triplets we defined the corresponding fundamental equations of the respective fuzzy J-K, D and Choi D flip-flops.

ISBN: 978-989-95079-6-8

The parameter values in the Yager, Dombi, Hamacher and Frank norms strongly influence the $J \to Q(t+1)$ characteristics curvature. We have compared the characteristics for various typical parameter values and choose w=2, α=2, γ=10 and s=100, where we obtained more or less S-shaped $J \to Q(t+1)$ characteristics. A change of the t-norms in the characteristical equations of fuzzy J-K, D and Choi D flip-flops leads to the modification of the slope of the transfer function, which will affect the learning rate in the implementation of neural networks.

4 Fuzzy flip-flop based neurons

In this section we propose the use of the fuzzy flip-flops discussed above as neurons in a MLP. In the next, we study the effect of applying the mentioned five t-norms and co-norms in the investigation of the F^3 based neurons and the MLPs constructed from them. An important aspect of these F^3s is that they all have a certain convergent behavior when their input J is excited repeatedly. This convergent behavior guarantees the learning property of the networks constructed this way.

4.1 Function Approximation by Fuzzy Flip-Flop Network Trained with Levenberg – Marquardt Algorithm

A fuzzy flip-flop based supervised feedfordward backpropagation network is applied in order to approximate test functions. This function is represented by a set of 100 input/output data sets. All the input and output signals are distributed in the unit interval. In general, two trainable layer networks with sigmoid transfer functions in the hidden layer and linear transfer functions in the output layer have good approximation and interpolation properties [7], analogously the neural system model proposed is based on two hidden layers constituted from fuzzy flip-flop neurons.

The nonlinear characteristics exhibited by fuzzy neurons are represented by quasi sigmoid transfer functions given by fuzzy J-K, D and Choi D flip-flops based on algebraic, Yager, Dombi, Hamacher and Frank operations. The proposed network activation function is the same at each hidden layer, from unit to unit. For simplicity we did not apply activation function or threshold to the output layer, and considered our model with only one output. The number of neurons was chosen after experimenting with different size hidden layers. Smaller neuron numbers in the hidden layer result in worse approximation properties, while increasing the neuron number results in better performance, but longer simulation time. In our approach the weighted input values are connected to input J of the fuzzy flip-flop based on a pair of t-norm and t-conorm, having quasi sigmoid transfer characteristics. The output signal is then computed as the weighted sum of the input signals, transformed by the transfer function.

During network training, the weights and thresholds are first initialized to small, random values and the network was trained with Levenberg-Marquardt algorithm with 100 maximum numbers of epochs as more or less sufficient. First we fixed the activation function, the number of layers and the number of units in each layer. The chosen target activation function was the *tansig* (hyperbolic tangent sigmoid transfer function). This function is well suited to the demands of backpropagation.

5 Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM)

The Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM) [5] is a very recent soft computing tool. It combines a special kind of Genetic Algorithm [6] - Pseudo-Bacterial Genetic Algorithm (PBGA) - and the Levenberg-Marquardt (LM) method. We use this algorithm for training fuzzy flip-flop based neural network (FNN) to improve function approximation performance. The Bacterial Memetic Algorithm (BMA) resulted in better approximation properties in fuzzy modeling problems than the Bacterial Evolutionary Algorithm (BEA) (which outperformed the traditional Genetic Algorithms [2], [3]).

The core of BMAM contains the bacterial mutation, which is inspired by the biological bacterial cell model. Its basic idea is to improve the parts of chromosomes contained in each bacterium. The bacterial mutation mechanism is used by the bacteria which can transfer genes to other bacteria. To find the optimal approximation for our network we encoded our FNN weighs, biases, Q and fuzzy operation parameter values in a bacterium (chromosome). Therefore a procedure is working on changing the variables, testing the model obtained in this way and selecting the best models.

5.1 Function Approximation by Fuzzy Flip-Flop Network Trained with BMAM

The learning of the FNN is formulated as a parameter optimization problem, using the mean square error as the fitness evaluation set-up. The basic steps followed by the algorithm embrace the bacterial mutation operation and the LM method.

In the initial population a number of individuals are randomly created and evaluated. Next, an evolutionary cycle is started by applying the bacterial mutation operation for each individual. A number of copies (clones) of the bacterium are generated. Then the same part or parts of the chromosome is choose and mutate randomly, except one single clone that remains unchanged during this mutation cycle. The LM method nested into evolutionary algorithm is applied for a few times for each individual. The selection of the best clone is made and transfers its parts to the other clones. The part choosing-mutation-LM method-selection-transfer cycle is repeated until all the parts are mutated, improved and tested. The best individual is remaining in the population, and all other clones are deleted. This process is repeated until all the individuals have gone through the modified bacterial mutation.

6 Simulation results

The combination of two sine wave forms with different period lengths as test function was,

$y = \sin(c_1 * x) * \sin(c_2 * x)/2+0.5$,

ISBN: 978-989-95079-6-8

where the input vector x generated a sinusoidal output y. The values of constants c_1 and c_2 were selected to produce a frequency proportion of the two components 1:0.35.
The function approximation goodness depended on the flexible Q and fuzzy operation parameter values, on the fuzzy flip-flop types, hidden layer neuron numbers and in addition on the fuzzy operations themselves. To check the goodness of training, we used the Mean Squared Error (MSE) as a measure of the error made by the FNN. The simulation results are summarized in the next two tests.

6.1 Test 1

In one of our earlier paper [9] we optimized the value of Q, for every combination of J-K, D and Choi D type F^3s with all five fuzzy operation pairs, the parameters of the norms chosen for suitable values, solving the problem of the selection of the value of Q with the fine tuning of FNN.
The fuzzy flip-flop based FNNs are based on algebraic, further parametric Yager, Dombi, Hamacher and Frank norms. Table 2 shows the experimental results obtained with LM optimization [9].

Table 2: Q optimums obtained with LM algorithm.

	Fuzzy neuron type		
Fuzzy operation	J-K	D	Choi D
Algebraic (A)	**0.21**	**0.15; 0.91**	**0.09; 0.81**
Yager (Y)	0.06	0.20	0.17
Dombi (D)	**0.11**	0.92	0.06
Hamacher (H)	0.32	0.57	0.38
Frank (F)	0.31	0.45	0.55

The $J \to Q(t+1)$ fuzzy D and Choi D flip-flop characteristics based on algebraic norm present about the same sigmoidal character in multiple points of the domain [7], which fact can leads to rather different optimum results, corresponding to minimal median MSE values. In Test 1 we propose a new method to find the optimal Q – fuzzy operation parameter pair for every combination of J-K, D and Choi D type F^3s by training a 1-8-8-1 FNN with bacterial memetic algorithm. The advantage of this network size lies in good learning capability. In our application a population with 29 parameters - according to the network size - was initialized. During simulations 20 generations of 5 individuals with 5 clones were chosen to obtain the best fitting variables. Table 3 shows the unique optimal Q values found by BMAM algorithm, which are approximately equal to the respective optimum values listed in Table 2. We obtained almost the same results after every run, although in short term the training is not always successful.

Table 3: Q optimums obtained with BMAM algorithm.

	J-K	D	Choi D
Algebraic	0.25	0.12	0.13

The experimental results in case of J-K F^3 type FNN based on Dombi norm further indicated that the proposed network had multiple optimal Q – α parameter pairs. The test revealed, for example, the Q = 0.12, α = 2.9 values as a result that was close to the optimum solution mentioned above.

6.2 Test 2

In the next we will compare the function approximation performance of FNN trained in two different ways. In the first one the network training function updates weight and bias values according to Levenberg-Marquardt optimization (section 4.1). In the second approach the FNN variables - weights, bias, for fixed Q and fuzzy operation parameter values - optimization was made by the bacterial memetic algorithm with the modified operation execution order (section 5.1). During the simulations we covered all 15 possible combinations of fuzzy J-K, D and Choi D type FNNs with all five fuzzy operation pairs to approximate the above mentioned test function.
By changing the number of the layers, we used a 1-4-3-1 FNN in order to have a network that approximated accurately and also fast enough the sine waves, to emphasize the difference between various FNNs.
The parameter of Dombi, Yager, Hamacher and Frank norms were fixed α=2, w=2, γ=10 and s=100, which values provided good learning and convergence properties.
In our approach we fixed the value of Q, the present state value belonging to each fuzzy flip-flop, according to the values in Table 2. Tables 4, 5 and 6 present the 30 runs average approximation goodness, by indicating the minimum, median, mean and maximum mean squared error of the training values for each of the *tansig*, algebraic, Yager, Dombi, Hamacher and Frank types of FNNs.
Figures 1 and 2 present the graphs of the simulations in case of fuzzy J-K flip-flop with feedback based neural network trained with LM and respectively BMAM algorithms. In both of cases the algebraic F^3 provides a fuzzy neuron with rather bad learning ability. Comparing the median MSE values, considering them as the most important indicators of trainability, the Yager and Dombi types FNNs performed best, they can be considered as rather good function approximators. Our simulations have shown that BMAM performs better in every case than the original LM technique.
Figures 3, 4 and 5, 6 compare the behavior of fuzzy D flip-flop and Choi type fuzzy D flip-flop based NNs. It is interesting that according to the numerical illustrations (Tables 5 and 6) the average of 30 run mean squared errors in these cases the best results after the idealistic *tansig* function are given by the Hamacher and Yager F^3, which are followed by the Dombi and finally by the algebraic one. The Hamacher and Yager types FNNs have excellent approximation properties. It is surprising how much the satisfaction of idempotence influences the behavior of the fuzzy D flip-flop based NN. Comparing the simulation results belonging to the two types of fuzzy D flip-flop with the same norms, it can be seen that, for the same value of Q, the value of the MSE differs, which fact leads to a rather different behavior in the applications.
Using BMAM algorithm in the FNN variable optimization task, we obtained better function approximation capability with lower error than in case of FNNs trained with LM method.

ISBN: 978-989-95079-6-8

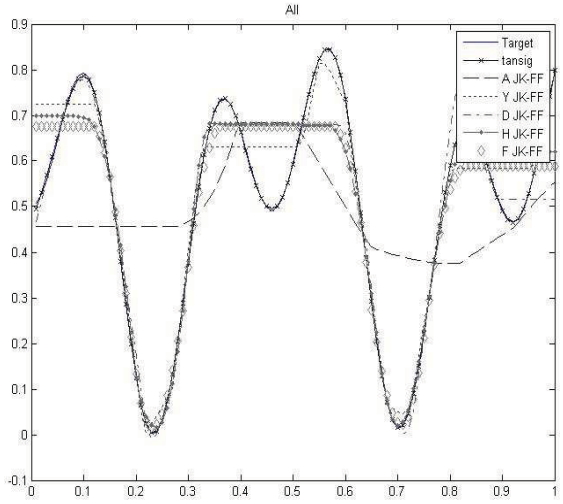


Figure 1: Simulation results of J-K FNN trained with LM algorithm.

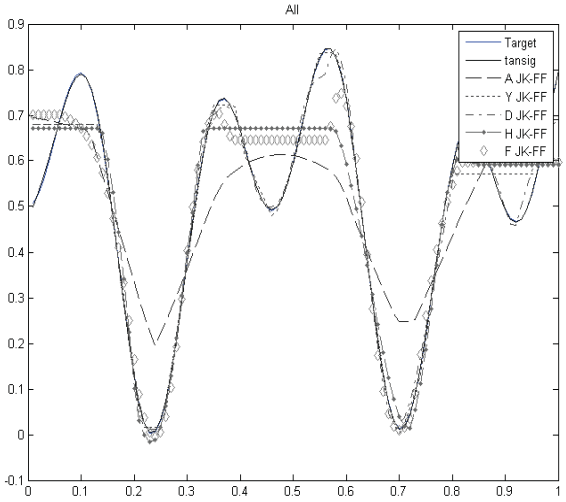


Figure 2: Simulation results of J-K FNN trained with BMAM algorithm.

Table 4: Mean squared error values case of JK FNN.

	LM				BMAM			
	min	median	mean	max	min	median	mean	max
tansig	6.6912×10^{-6}	**0.0056865**	0.0185562	0.0600578	4.8485×10^{-6}	**7.5218×10^{-5}**	0.0006223	0.0041448
A	0.04402231	**0.0548533**	0.0555179	0.0818562	0.00888158	**0.0170071**	0.0195847	0.0368526
Y	0.00286463	**0.0534354**	0.0474332	0.0615071	0.00133805	**0.0031207**	0.0310288	0.0064944
D	0.00420610	**0.0512862**	0.0439112	0.0691943	9.2216×10^{-5}	**0.0022534**	0.0024855	0.0052513
H	0.00551096	**0.0385112**	0.0400807	0.0720822	0.00351781	**0.0071140**	0.0081815	0.0256592
F	0.05552121	**0.0485408**	0.0425236	0.0660629	0.00228814	**0.0066725**	0.0076724	0.0312439

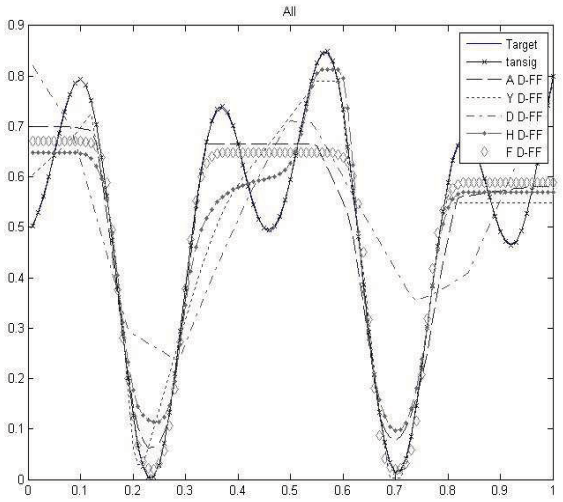


Figure 3: Simulation results of D FNN trained with LM algorithm.

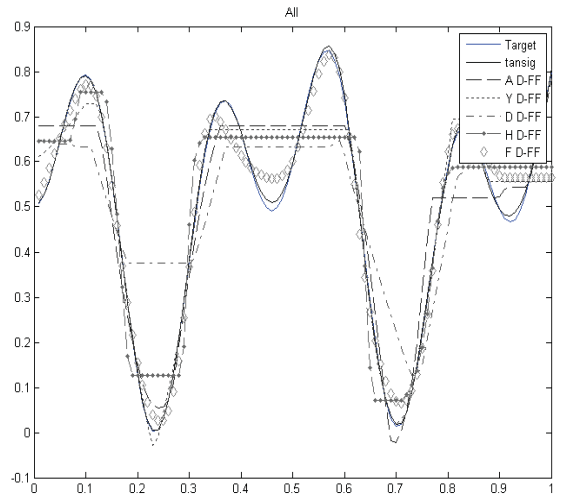


Figure 4: Simulation results of D FNN trained with BMAM algorithm.

Table 5: Mean squared error values case of D FNN.

	LM				BMAM			
	min	median	mean	max	min	median	mean	max
tansig	4.3611×10^{-6}	**0.0068415**	0.0203523	0.0565211	9.6899×10^{-7}	**4.7893×10^{-5}**	0.0006056	0.0033605
A	0.00771386	**0.0531832**	0.0468789	0.0838021	0.00490354	**0.00918654**	0.0114988	0.0330397
Y	0.00562192	**0.0515206**	0.0448304	0.0653909	0.00031811	**0.00335034**	0.0033998	0.0065352
D	0.03089988	**0.0502911**	0.0502276	0.0692478	0.01059031	**0.02699423**	0.0255351	0.0365076
H	0.00511140	**0.0485498**	0.0425834	0.0679154	0.00017240	**0.00586916**	0.0073385	0.0271431
F	0.00694345	**0.0506654**	0.0458813	0.0640876	0.00051204	**0.00681308**	0.0065404	0.0229583

ISBN: 978-989-95079-6-8

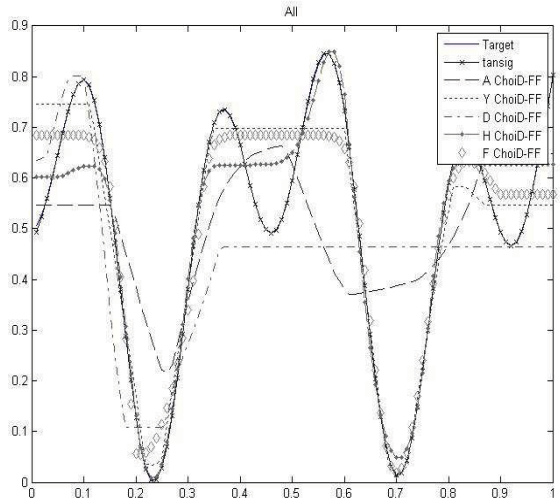


Figure 5: Simulation results of Choi D FNN trained with LM algorithm.

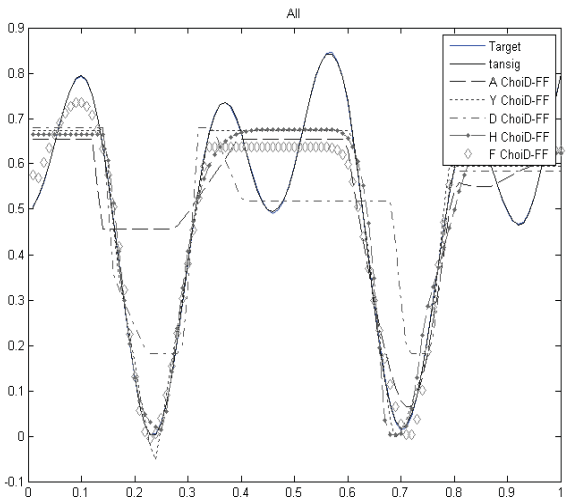


Figure 6: Simulation results of Choi D FNN trained with BMAM algorithm.

Table 6: Mean squared error values case of Choi D FNN.

	LM				BMAM			
	min	median	mean	max	min	median	mean	max
tansig	1.1454×10^{-6}	**0.0291153**	0.0251398	0.0624285	1.9737×10^{-7}	$\mathbf{2.7722 \times 10^{-5}}$	0.0005026	0.0024156
A	0.03693712	**0.0567222**	0.0557563	0.0878214	0.00777224	**0.0265908**	0.0251665	0.0363809
Y	0.00660108	**0.0508917**	0.0483636	0.0641307	0.00328527	**0.0077983**	0.0087765	0.0254508
D	0.03222194	**0.0534598**	0.0537578	0.0861916	0.01436651	**0.0235421**	0.0247559	0.0347627
H	0.00459423	**0.0497143**	0.0446552	0.0737876	0.00546122	**0.0077690**	0.0111944	0.0286523
F	0.00668745	**0.0518677**	0.0445716	0.0727711	0.00293771	**0.0092323**	0.0126012	0.0310469

7 Conclusions

In this paper we proposed the use of BMAM to optimizing fuzzy flip-flop based neural network (FNN). We compared the function approximation performance of five different types of FNNs, which depends from the choice of different fuzzy flip-flop types and from the training algorithm. The result of learning as well as its performance might differ quite significant under these two learning modes. The simulations have shown that the bacterial memetic algorithm can improve the function approximation capability by optimizing the FNN variables values.

Acknowledgment

This paper was supported by the Széchenyi University Main Research Direction Grant 2009, National Scientific Research Fund Grant OTKA T048832 and K75711, and National Office for Research and Technology.

References

[1] R. Bellman and M. Giertz, "On the analytic formalism of the theory of fuzzy sets", Information Science, vol. 5, 1973, pp. 149-156.

[2] J. Botzheim, C. Cabrita, L. T. Kóczy, A. E. Ruano, **"**Fuzzy rule extraction by bacterial memetic algorithms" International Journal of Intelligent Systems, 2009, pp. 312-339.

[3] J. Botzheim, C. Cabrita, L. T. Kóczy, A. E. Ruano, "Genetic and bacterial programming for B-spline neural networks design", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 11., No. 2, 2007, pp. 220-231.

[4] B. Choi and K. Tipnis, "New Components for Building Fuzzy Logic Circuits", Proc. of the 4th Int. Conf. on Fuzzy Systems and Knowledge Discovery, vol. 2, 2007, pp. 586-590.

[5] L. Gál, J. Botzheim and L. T. Kóczy, „Improvements to the Bacterial Memetic Algorithm used for Fuzzy Rule Base Extraction", Computational Intelligence for Measurement Systems and Applications, CIMSA 2008, Istanbul, pp.38-43.

[6] J. H. Holland, "Adaptation in Nature and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence", The MIT Press Cambridge, 1992.

[7] K. M. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators", Neural Networks, vol.2, no.5, 1989, pp. 359-366.

[8] E. P. Klement, R. Mesiar and E. Pap, "Triangular Norms", Series: Trends in Logic , vol.8, 2000.

[9] R. Lovassy, L. T. Kóczy and L. Gál, "Function Approximation Capability of a Novel Fuzzy Flip-Flop Based Neural Network", IJCNN 2009 Atlanta- accepted.

[10] R. Lovassy, L. T. Kóczy and L. Gál, "Analyzing Fuzzy Flip-Flops Based on Various Fuzzy Operations", Acta Technica Jaurinensis Series Intelligentia Computatorica vol. 1, no. 3, 2008, pp. 447-465.

[11] R. Lovassy, L. T. Kóczy and L. Gál, "Multilayer Perceptron Implemented by Fuzzy Flip-Flops", IEEE World Congress on Computational Intelligence, WCCI 2008, Hong Kong, pp. 1683-1688.

[12] S. Osowski, K. Siwek and T. Markiewicz, "MLP and SVM Networks – a Comparative Study", Proceedings of the 6th Nordic Signal Processing Symposium, Finland, 2004, pp. 37-40.

[13] K. Ozawa, K. Hirota and L. T. Kóczy, "Fuzzy flip-flop", In: M. J. Patyra, D. M. Mlynek, eds., Fuzzy Logic. Implementation and Applications, Wiley, Chichester, 1996, pp. 197-236.

[14] K. Ozawa, K. Hirota, L. T. Kóczy and K. Omori, "Algebraic fuzzy flip-flop circuits", Fuzzy Sets and Systems 39/2, North Holland, 1991, pp.215-226.

ISBN: 978-989-95079-6-8

Bi-criteria Genetic Selection of Bagging Fuzzy Rule-based Multiclassification Systems

Krzysztof Trawiński, Arnaud Quirin, Oscar Cordón

European Centre for Soft Computing. Mieres, Spain
Email: {krzysztof.trawinski, arnaud.quirin, oscar.cordon}@softcomputing.es

Abstract— *Previously we proposed a scheme to generate fuzzy rule-based multiclassification systems by means of bagging, mutual information-based feature selection, and a multicriteria genetic algorithm (GA) for static component classifier selection guided by the ensemble training error. In the current contribution we extend the latter component by the use of two bi-criteria fitness functions, combining the latter error measure with the selected ensemble likelihood. A study on four popular UCI datasets with different dimensionalities is conducted in order to analyze the accuracy-complexity trade-off obtained by the two GAs, the initial fuzzy ensemble and a single fuzzy classifier.*

Keywords— Bagging, feature selection, fuzzy rule-based multiclassification systems, genetic selection of individual classifiers, multicriteria genetic algorithm.

1 Introduction

Multiclassification systems (MCSs) are promising data mining tools dealing with complex classification problems, especially when the number of dimensions or the size of the data are really large [1]. They usually combine decision trees [2] or neural networks [3], but also more recently fuzzy classifiers [4, 5, 6].

In a previous study [7], we described a methodology in which classical MCS design approaches such as bagging [8] and random subspace [2] are used to generate fuzzy rule-based multiclassification systems (FRBMCSs) using a basic heuristic fuzzy classification rule generation method [9], as well as a classifier selection technique based on a GA driven by a multicriteria fitness function [10]. Later, we improved our methodology in [10] using a feature selection approach based on the Battiti's method [11] and the GRASP procedure [12]. Finally, in [13], we extended the genetic selection method by considering additional error measures.

We drew the conclusion that a feature and instance selection procedure combined with a simple grid partitioning FRBS is a good approach to overcome the curse of dimensionality problem in large datasets while using FRBMCS, mainly due to the fact that these kind of classifiers are instable enough. Nevertheless, once a set of classifiers has been trained, we still need to deal with the high number of rules and the correlations between individual classifiers. This is why a selection of the classifiers is so crucial. As said before, we already proposed a GA guided by several single-criteria fitness functions, based on the training error [10], the likelihood [7], or the Out-Of-Bag error [13]. This methodology, quite novel in this topic, lead us to the generation of a compact sets of rules, while still preserving its accuracy, in a single GA run, without resorting on a Pareto-based multi-objective optimization technique. However, many experimentations suggested the choice of the fitness function is very dependent of the problem being solved. For instance, when using the training error, the accuracies of two FRBMCSs can be similar or even null, making difficult for the GA to discriminate between them in order to improve the generalization ability. On the contrary, using the likelihood alone seems to give bad results on many datasets. This suggests the fact that a combination of some criteria could be a good idea to overcome this issue, producing better results than any criterion in isolation.

The aim of the current contribution is to propose a solution by exploring two new fitness functions based on a combination of the training error and the likelihood measures. By doing so, we will try to combine them using the two most simple ways: weighted average and lexicographic order (i.e. considering the optimization of a single criterion, and using the second in case of tie). Introducing such elaborated method we hope that it will allow the FRBMCS to deal with high dimensional data.

We aim to check if the new GA fitness functions will perform better in terms of accuracy than the previous ones for some datasets, while still being competitive for the others datasets. A preliminary study will be conducted on small and medium size datasets from the UCI machine learning repository to test the two new fitness functions in comparison to a single classifier, the original FRBMCS, and the GA-selected FRBMCSs using the said fitness functions. Several parameter settings for the global approach (e.g. different granularity levels as well as different feature selection methods) will be tested and compared regarding the accuracy and the size of the rule base obtained by a single classifier and the original FRBCS ensemble.

This paper is set up as follows. In the next section, existing GA-based methods to select MCSs are reviewed. Sec. 3 recalls our approach for designing FRBMCSs considering bagging and feature selection, while Sec. 4 describes the proposed multicriteria GA for component classifier selection. The experiments developed and their analysis are shown in Sec. 5. Finally, Sec. 6 collects some concluding remarks and future research lines.

2 Related work on genetic selection of MCSs

In general, the selection of a subset of classifiers is done using the *overproduce-and-choose strategy* (OCS) [14], in which a large set of classifiers is produced and then selected to extract the best performing subset. GAs are a popular technique within this strategy. In the literature, performance, complexity and diversity measures are usually considered as search criteria. Complexity measures are employed to increase the interpretability of the system whereas diversity measures are used to avoid overfitting.

ISBN: 978-989-95079-6-8

Among the different genetic OCS proposals, we can remark the following ones. In [15], a hierarchical multi-objective GA (MOGA) algorithm, performing feature selection at the first level and classifier selection at the second level, is presented which outperforms classical methods for two handwritten recognition problems. The MOGA allows both performance and diversity to be considered for MCS selection. In [16] a GA is used to select from seven diversity heuristics for k-means cluster-based ensembles and the ensemble size is also encoded in the genome. In the study of Martínez-Munoz et al. [17], a GA is compared to five other techniques for ensemble selection. Even if the performance of the GA was the worst obtained, they showed that while selecting a small subset of classifiers, the generalization error was significantly decreased. In [18], the authors developed a multidimensional GA to optimize two weight-based models, in which the weights are assigned to each classifier or to each class. They applied their system to 6 different classifiers (only linear and quadratic classifiers are explored), but on only two small datasets and without comparing to the results obtained on a single classifier. Finally, our own previous studies [10, 7] also consider a multicriteria GA for the ensemble selection in an OCS fashion, with performance (training error) and complexity as criteria to guide the GA.

In general, the performance obtained with the initial MCS is outperformed by the ensemble selected by the GA, while simplifying the system. In our current contribution, we will confirm this conclusion by the study of two improved fitness functions mixing the two most used criteria: the accuracy and the complexity of the classifiers. The fitness function will directly incorporate one or two accuracy criteria (i.e., the training error and the likelihood), while the MCS complexity will be implicitly optimized by the considered coding scheme.

3 Bagging and feature selection-based FRBMCSs

In this section we will both detail how the individual classifiers and the FRBMCSs are designed. A normalized dataset is split into two parts, a training set and a test set. The training set is submitted to an instance selection and a feature selection procedure in order to provide individual training sets (the so-called *bags*) to train simple FRBCSs (through the method described in section 3.1). The instance selection and the feature selection procedures are described in section 3.2. After the training, we got an initial FRBMCS, which is validated using the training and the test errors (*Ensemble Training Error* and *Ensemble Test Error*), as well as a measure of complexity based on the total number of rules in the FRBCSs. This ensemble is selected using a multicriteria GA (described in the next section) guided by two accuracy-based fitness functions. The final FRBMCS is validated using different accuracy (Training Error, Test Error) and complexity measures (number of classifiers, total number of rules).

3.1 Individual FRBCS composition and design method

The FRBCSs considered in the ensemble will be based on fuzzy rules R_j with a class C_j and a certainty degree CF_j in the consequent: If x_1 is A_{j1} and ... and x_n is A_{jn} then Class C_j with CF_j, $j = 1, 2, \ldots, N$, and they will take their decisions by means of the single-winner method [9]. This fuzzy reasoning method has been selected due to its high simplicity and interpretability. The use of other more advanced ones [19] is left for future works.

To derive the fuzzy knowledge bases, one of the heuristic methods proposed by Ishibuchi et al. in [9] is considered. The consequent class C_j and certainty degree CF_j are statistically computed from all the examples located in a specific fuzzy subspace $D(A_j)$. C_j is computed as the class h with maximum confidence according to the rule compatible training examples $D(A_j) = \{x_1, \ldots, x_m\}$: $c(A_j \Rightarrow Class\ h)$ = $|D(A_j) \bigcap D(Class\ h)|/|D(A_j)|$. CF_j is obtained as the difference between the confidence of the consequent class and the sum of the confidences of the remainder (called CF_j^{IV} in [9]).

This method is good for our aim of designing FRBCS ensembles since it is simple and quick. However, it carries two drawbacks: its low accuracy and the generation of large fuzzy rule bases. We aim to consider more advanced techniques in the future.

3.2 FRBMCS design approaches

The generation of the FRBMCSs is performed by means of a bagging approach combined with a feature selection method [10]. Three different feature selection methods, random subspace and two variants of Battiti's MIFS, greedy and GRASP, are considered.

As said before, *random subspace* [2] is a method in which we select randomly a set of features from the original dataset. The greedy Battiti's MIFS method [11] is based on a forward greedy search using the Mutual Information measure [20], with regard to the class. This method selects the set S of the most informative features about the output class which cannot be predicted with the already selected features. It uses a coefficient, β, to set up the penalization on the information brought by the already selected features.

The MIFS-GRASP variant is an approach where the set is generated by iteratively adding features randomly chosen from a Restricted Candidate List (RCL) composed of the best τ percent decisions according to the Battiti's quality measure. Parameter τ is used to control the amount of randomness injected in the MIFS selection. With $\tau = 0.5$, we get an average amount of randomness, while still preserving the quality-based ordering of the features.

For the bagging approach, the bags are generated with the same size as the original training set, as commonly done. In every case, all the classifiers will consider the same fixed number of features.

Finally, no weights will be considered to combine the outputs of the component classifiers to take the final FRBMCS decision, but a pure voting approach will be applied: the ensemble class prediction will directly be the most voted class in the component classifiers output set.

4 A multicriteria GA-based MCS selection method

In our previous studies, we used a multicriteria GA, which is able to obtain a list of possible MCS designs ranked by their quality *from a single chromosome* thanks to its novel coding scheme. However, the fitness function considered was based on a single criterion, either the likelihood (L) [7], the *training*

ISBN: 978-989-95079-6-8

error (TE) [10, 13], or the out-of-bag error [13]. Although the TE-based GA provided better overall results, the L-based GA outperformed it in some of the cases. That led us to the idea of combining the both measures, which moreover show complementary characteristics.

4.1 Multicriteria genetic optimization

The GA searches for an optimal sequence of the classifiers, in the way that the most significant classifiers have the lowest indexes, while those redundant members, which can be safely excluded, are in the last positions. The coding scheme is thus based on an order-based representation, a permutation $\Pi = \{j_1, j_2, \ldots, j_l\}$ of the l originally generated individual classifiers. In this way, each chromosome encodes l different solutions to the problem, based on considering a "basic" MCS comprised by a single classifier, that one stored in the first gene, then another one composed of two classifiers, those in the first and the second genes, and so on.

So, the computation of the evaluation criteria for the whole ensemble is obtained in a *cumulative* way, defined as a vector containing the measured values of the first classifier; the subset formed by the first and the second; and so on. The fitness function is thus using the values of a multicriteria vector, being composed of an array of l values, $L^i = L'_{\{j_1,j_2,\ldots,j_i\}}$, corresponding to the cumulative measure-value of the l mentioned MCS designs.

At the end of the GA, the best chromosome is that member in the population overcoming the others using the considered criterion. Then, the final design encoded in this chromosome is the MCS comprising the classifiers from the first one to the one having the the best cumulative measured value (although any other design not having the optimal accuracy but, for example, showing a lowest complexity can also be directly extracted). In this way, an implicit use of a complexity criterion is also made.

To increase its convergence rate, the GA works following a steady-state approach. The initial population is composed of randomly generated permutations. In each generation, a tournament selection of size 3 is performed, and the two winners are crossed over to obtain a single offspring that directly substitutes the loser. In this study, we have considered OX crossover and the usual exchange mutation [21].

4.2 The two used evaluation criteria

For the definition of the fitness functions, we use TE and L as the evaluation criteria.

The TE is computed as follows. Let $h_1(\mathbf{x}), \ldots, h_l(\mathbf{x})$ be the outputs of the component classifiers of the selected ensemble for an input value $\mathbf{x} = (x_1, \ldots, x_n)$. For a given sample $\{(\mathbf{x}^k, C^k)\}_{k\in\{1\ldots m\}}$, the TE of that MCS is:

$$TE = \frac{1}{m} \cdot \#\{k \mid C^k \neq \arg\max_{j\in\{1\ldots M\}} h_j(\mathbf{x}^k)\} \qquad (1)$$

Fitness evaluation using TE alone was already studied in one of our previous publications [10]. We will call it *Training Error-based Fitness Function* (TEFF).

The L is computed as follows. Let the classes $h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_l(\mathbf{x})$ be the decisions of the component classifiers of the selected ensemble S for an input value $\mathbf{x} = (x_1, \ldots, x_n)$. We will assume that the fraction of the members of S that agree on the class of $\mathbf{x}$ is an estimate of the conditional probability of that class:

$$P_S(C|\mathbf{x}) = \frac{1}{|S|} \cdot \#\{i \in S \mid h_i(\mathbf{x}) = C\}.$$

The L of the subset S, to be maximized, is:

$$L_S = \prod_k P_S(C^k|\mathbf{x}^k).$$

As the small values of L_S may produce numerical instabilities, we use instead a bounded log-likelihood:

$$L'_S = \sum_k \log(P_S(C^k|\mathbf{x}^k) + \epsilon),$$

where the value ϵ foresees that case for which none of the members of the subset has found the true class of the pattern. In [7], we endowed the fitness function with the L, as it allows us to discern differences between ensembles with the same TE (specifically, between those with null error!). A learning process using only the TEFF will automatically end up with the learning, while L will go on improving the estimations of the probability distributions for each class, thus reducing the chances of overfitting the training data.

4.3 The two new bi-criteria fi tness functions

In this contribution we propose two approaches for the fitness function combining the L with the TE measure, the *Lexicographical Order-based Fitness Function* (LOFF) and the *Weighted Combination Fitness Function* (WCFF).

Notice that, working in this way, we are introducing a second multi-criteria optimization level in our algorithm. On the one hand, a multi-criteria optimization is made by means of the considered coding scheme and the cumulative evaluation of the possible MCS designs (see Sect. 4.1). On the other hand, a higher level is introduced when evaluating the latter possible designs by means of a bi-criteria fitness function.

In the first one, the LOFF, we use the lexicographical order to deal with the multicriteria optimization. When comparing two chromosomes, one is better than the other if it takes a better (lower) minimum value of the TE. In case of tie, the L measure is considered. The ordering scheme gives priority to TE, as it provided better results in our previous study, while taking the L only in the last resort in the case of the frequent ties encountered by the system.

In the second approach, the WCFF, we propose objective function scalarization by a weighted combination of both measures:

$$WC = factor_0 * \alpha * TE + (1 - \alpha) * L \qquad (2)$$

where α is a weight in [0,1] and $factor_0 = L_0/TE_0$ is a first evaluation-based normalization using L_0 and TE_0, the L and the TE from the initial FRBMCS. The fitness function is to be minimized.

5 Experiments and analysis of results

In this section, we discuss the performance obtained by a single FRBCS, the initial FRBMCS, and three different GA-selected FRBMCSs including our two new fitness functions on four chosen datasets.

ISBN: 978-989-95079-6-8

5.1 Experimental setup

To evaluate the performance of the generated FRBMCSs, we have selected four datasets from the UCI machine learning repository (see Table 1). In order to compare the accuracy of the considered classifiers, we used Dietterichs 5×2-fold cross-validation (5×2-cv), which is considered to be superior to paired k-fold cross validation in classification problems [22].

Table 1: Data sets considered

Data set	#attr.	#examples	#classes
Pima	8	768	2
Glass	9	214	7
Vehicle	18	846	4
Sonar	60	208	2

Three different granularities, 3, 5 and 7, are tested for the single FRBCS derivation method, for feature sets of size 5 selected by means of three approaches: the greedy Battiti's MIFS filter feature selection method [11], the Battiti's method with GRASP (with τ equal to 0.5, see section 3.2), and random subspace [2]. Battiti's method has been run by considering a discretization of the real-valued attribute domains in ten parts and setting the β coefficient to 0.1.

The FRBMCSs generated are initially comprised by 50 classifiers. The GA for the component classifier selection works with a population of 50 individuals and runs during 50 generations. The mutation probability considered is 0.05. The weights of WCFF were set to 0.8 for TE and 0.2 for L as our aim was to allow a small influence of the L in the cases in which the TE gives similar values. The other values for the weights will not improve the results significantly.

All the experiments have been run in a cluster at the University of Granada on Intel quadri-core Pentium 2.4 GHz nodes with 2 GBytes of memory, under the Linux operating system.

5.2 Comparison of the three fi tness functions

The statistics (5×2-cv error, number of rules, and run time required for each run, expressed in seconds) for the genetically selected FRBCS ensembles using LOFF, WCFF and TEFF are collected in Tables 2, 3 and 4 respectively. There are three subtables for each of the feature selection method considered. The best results for a given feature selection method are shown in bold and the best values overall are outlined.

Comparing the three fitness functions, we can see how the WCFF approach is able to outperform the TEFF and LOFF considering the individual test error 9 times (+2 draw). The best individual improvement was observed on the sonar dataset (-8% regarding LOFF, -4% regarding TEFF) with greedy and 5 labels. We observed that in 4 out of 9 cases, WCFF outperforms other approaches on the sonar and pima datasets. The best overall result was obtained on the pima dataset with GRASP and 5 labels (draw with LOFF).

The FRBMCSs based on LOFF are better than TEFF and WCFF in 12 of the 36 cases (+2 draw). The best individual improvement was observed on the sonar dataset with random subspace and 5 labels (-9% regarding WCFF, -12% regarding TEFF). The best overall result was obtained on the pima dataset with greedy and 5 labels (draw with WCFF) and on the glass dataset with GRASP and 5 labels. We can observe that in 5 out of 9 cases, LOFF outperforms the other genetic approaches on the glass dataset.

Comparing the two new fitness functions, the LOFF provides better results than WCFF considering individual test error in 21 cases (+ 2 draw). We can also observe that LOFF outperforms the other approach on the glass dataset in 7 out 9 cases (+1 draw) giving a good performance for the vehicle (7 out 9 cases) datasets. However, the WCFF is better in 6 out of 9 cases on the pima dataset and in 5 out of 9 cases (+1 draw) on the sonar dataset. In general, LOFF is better than WCFF since it gives a lower influence to L, whereas the weighted combination is better on the sonar dataset due to the many ties with the TE on this dataset, making the use of the L more likely.

The TEFF-based FRBMCSs outperforms the LOFF and the WCFF considering individual test error for 13 of the 36 times. The best overall result was obtained on the sonar dataset with random subspace and 3 labels, and on the vehicle with random subspace and 7 labels. We may conclude that the LOFF and WTEL are competitive with TEFF. The LOFF got best results 12 times (+2 draw), WTEL 9 (+2 draw) and TEFF 13 times. However, in the direct comparison, the use of LOFF improves the single TEFF performance in 18 out of 36 cases (+3 draw) and the WTEL improves the single TEFF performance in 19 out of 36 cases, which indicates that L does not produce overfitting. Thus, the use of L as a secondary criterion is only useful in some of the cases.

On 36 cases, the number of classifiers is lower in 12 cases (+1 draw) with the LOFF, in 9 cases with the WCFF and in 15 cases (+1 draw) with the TEFF. The two new fitness functions generate a higher number of classifier, since they are more conservative due to the use of L. Such fitness functions could be viewed as a proper way to improve performance of the datasets having the larger size.

5.3 Genetically selected FRBMCSs vs. single FRBCS/original FRBMCSs

The results of the single FRBCSs are presented in Table 5 while those of the original FRBMCSs are presented in Table 6. In all the 36 cases, the generated FRBMCSs improve the performance of the single FRBCS.

Although the main goal of the genetic selection is to reduce the complexity of the generated FRBMCS, the accuracy results obtained from that process are also improved in most of the cases, showing the potential of the approach. In only 10 of the 36 cases (+1 draw) the original FRBMCS outperforms the best genetically designed one in terms of accuracy. Comparing the best overall TE values of genetically selected FRBMCSs with those of the original FRBMCSs, the GA improves the results on the sonar dataset (-2% regarding TEFF) and glass (-10% regarding LOFF), showing a slight increase for the other problems (+2% for pima regarding LOFF/WTEL, +2% for vehicle regarding TEFF).

5.4 Statistical signifi cance of the results

Table 7 shows the results of the statistical tests performed to check if the performance of the initial FRBMCSs and the performance of the GA selected FRBMCSs outperform significantly the performance of the single classifier. The Wilcoxon signed-rank test [23] has been used for this purpose. The best

ISBN: 978-989-95079-6-8

Table 2: Results for the FRBCS ensembles selected by the GA using the LOFF

		Bagging + Greedy				Bagging + Random Subspace				Bagging + GRASP $\tau = 0.50$			
		Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar
	5×2-cv	0.257	**0.363**	0.468	**0.246**	0.261	**0.375**	0.423	**0.219**	0.253	0.367	0.440	**0.227**
3 labels	#classifiers	3.8	10.1	9.5	8.5	4.0	13.0	14.2	19.9	4.2	9.1	11.7	12.9
5 #attr.	#rules	659.7	1275.9	1298.6	1287.2	668.4	1450.0	2336.3	3363.0	742.4	1162.3	1795.1	2110.9
	avg. #rules	176.6	126.5	138.3	151.0	167.6	112.5	165.3	169.9	175.2	125.9	152.0	163.6
	time	599.57	127.00	665.84	164.46	436.40	111.97	482.74	117.36	534.17	134.80	591.25	144.59
	5×2-cv	**0.238**	0.374	0.391	0.259	**0.257**	0.387	0.374	0.223	**0.240**	**0.358**	0.399	0.234
5 labels	#classifiers	12.9	8.5	14.0	10.1	12.9	14.2	14.2	14.1	10.9	16.0	13.4	13.0
5 #attr.	#rules	7491.1	2211.3	6711.9	5696.9	7160.0	3439.0	10374.3	9273.3	6535.1	4190.4	7610.7	8291.5
	avg. #rules	582.0	268.1	482.1	569.4	550.8	247.5	740.3	680.7	599.5	271.7	582.1	644.4
	time	434.14	121.72	487.14	117.84	599.22	167.95	665.79	162.04	370.50	102.98	405.34	100.62
	5×2-cv	0.258	0.387	**0.375**	0.258	0.270	0.393	**0.347**	0.275	0.258	0.375	**0.355**	0.250
7 labels	#classifiers	14.4	9.5	14.0	7.3	15.8	13.4	17.8	6.2	13.3	9.1	15.5	8.5
5 #attr.	#rules	18851.1	3786.4	16266.7	7609.4	19619.0	5169.0	29596.5	7690.1	17892.8	3738.7	20978.9	10210.4
	avg. #rules	1308.1	400.2	1146.7	1056.8	1240.5	392.6	1686.0	1234.8	1338.9	409.5	1386.4	1214.6
	time	369.73	103.94	405.34	100.18	366.53	103.25	407.70	99.94	498.07	140.94	552.16	135.59

Table 3: Results for the FRBCS ensembles selected by the GA using the WCFF

		Bagging + Greedy				Bagging + Random Subspace				Bagging + GRASP $\tau = 0.50$			
		Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar
	5×2-cv	0.254	**0.364**	0.488	0.256	0.258	0.384	0.430	**0.224**	0.255	**0.370**	0.459	**0.226**
3 labels	#classifiers	6.2	8.8	15.6	6.8	6.3	10.5	17.7	22.2	7.5	13.9	10.3	10.4
5 #attr.	#rules	1081.6	1139.8	2209.2	1030.0	1020.8	1190.4	2944.8	3683.5	1331.4	1722.5	1587.2	1677.3
	avg. #rules	174.3	125.4	140.2	153.2	161.2	113.1	166.6	171.0	172.4	126.3	153.3	167.6
	time	401.63	111.35	441.11	108.23	382.65	105.19	419.06	102.40	392.34	108.28	428.73	105.53
	5×2-cv	**0.243**	0.380	0.395	**0.238**	**0.256**	**0.382**	0.372	0.244	**0.238**	0.386	0.401	0.234
5 labels	#classifiers	17.4	12.9	14.7	14.9	16.3	9.2	16.0	19.1	14.6	13.0	15.4	22.4
5 #attr.	#rules	10118.3	3438.4	7258.3	8202.4	9104.2	2347.5	11739.3	12502.4	8686.7	3457.0	8947.0	13497.0
	avg. #rules	589.8	264.5	486.4	566.8	550.0	260.3	737.4	674.3	593.8	270.3	593.6	622.2
	time	383.37	106.09	420.65	102.93	401.48	111.36	441.42	108.02	371.67	102.42	411.51	100.52
	5×2-cv	0.250	0.402	**0.368**	0.256	0.272	0.393	**0.353**	0.275	0.254	0.383	**0.363**	0.243
7 labels	#classifiers	13.8	13.4	12.6	8.0	16.2	15.1	10.3	6.2	14.8	11.7	12.1	9.7
5 #attr.	#rules	17992.3	5231.7	14853.7	8248.2	20343.7	5923.8	18443.5	7690.1	19726.7	4728.6	17199.3	11271.4
	avg. #rules	1323.8	399.7	1164.6	1053.8	1258.5	411.1	1814.7	1234.8	1343.0	402.2	1435.8	1173.8
	time	372.49	102.26	407.72	100.36	368.65	103.92	408.42	99.82	386.31	107.54	425.35	104.44

Table 4: Results for the FRBCS ensembles selected by the GA using the TEFF

		Bagging+Greedy				Bagging+GRASP $\tau = 0.50$				Bagging + Random Subspace			
		Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar
	5×2-cv	0.257	**0.360**	0.461	**0.235**	0.254	0.372	0.449	**0.237**	**0.256**	**0.381**	0.428	**0.216**
3 labels	#classifiers	4.1	7.3	10.3	12.3	14.4	10.2	12.9	13.9	4.2	13.7	13.4	20.1
5 #attr.	#rules	696.5	904.3	1431.0	1842.1	763.0	1317.9	1991.6	2252.6	703.4	1546.0	2239.5	3376.7
	avg. #rules	171.5	125.4	138.3	148.3	174.3	126.0	155.9	161.7	168.1	113.1	168.9	168.3
	time	94.06	26.35	103.26	25.32	93.37	26.49	102.09	25.18	92.77	26.39	103.24	25.08
	5×2-cv	**0.242**	0.383	0.392	0.247	**0.239**	**0.363**	0.399	0.252	0.263	0.392	0.378	0.249
5 labels	#classifiers	11.5	15.9	15.5	10.4	10.9	14.7	12.0	7.8	11.9	13.7	13.0	9.4
5 #attr.	#rules	6744.9	4233.1	7338.4	5757.7	6497.4	3986.7	7227.3	4893.9	6680.0	3312.2	9455.9	6208.8
	avg. #rules	592.8	268.7	481.9	567.0	593.5	282.0	611.3	630.0	555.8	245.0	734.3	668.8
	time	93.48	26.10	103.48	25.17	92.58	26.16	103.75	24.86	91.47	26.18	104.81	24.83
	5×2-cv	0.258	0.393	**0.374**	0.258	0.256	0.395	**0.356**	0.257	0.265	0.393	**0.337**	0.267
7 labels	#classifiers	12.7	8.9	14.6	6.3	16.4	10.3	13.2	6.7	17.0	15.5	17.5	6.4
5 #attr.	#rules	16614.3	3524.3	16102.3	6427.0	21836.6	4140.6	18296.2	7767.8	21289.5	5980.6	28854.2	7655.2
	avg. #rules	1313.9	404.5	1115.7	1040.9	1346.2	401.9	1386.5	1148.7	1248.4	386.2	1680.2	1203.7
	time	92.87	26.50	102.90	24.85	92.49	26.18	102.93	25.31	92.31	26.08	103.52	25.19

Table 5: Results for the single FRBCSs with feature selection

		Greedy				GRASP $\tau = 0.50$				Random Subspace			
		Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar
3 labels	5×2-cv	0.266	0.446	0.549	**0.261**	0.267	0.447	0.546	0.316	0.265	0.457	0.512	**0.319**
5 #attr.	#rules	178.50	135.30	136.40	146.60	179.50	137.00	135.80	169.00	161.80	109.50	154.50	174.50
	time	0.08	0.04	0.12	0.08	0.09	0.04	0.12	0.09	0.07	0.03	0.12	0.08
5 labels	5×2-cv	**0.246**	**0.376**	0.430	0.287	**0.246**	**0.375**	0.425	**0.314**	**0.262**	0.435	0.460	0.329
5 #attr.	#rules	682.70	291.00	437.60	615.20	682.70	293.50	418.90	752.70	604.20	259.60	587.80	773.60
	time	0.42	0.25	0.65	0.16	0.39	0.26	0.63	0.17	0.36	0.24	0.67	0.17
7 labels	5×2-cv	0.262	0.414	**0.402**	0.291	0.266	0.423	**0.399**	0.317	0.276	**0.418**	**0.415**	0.340
5 #attr.	#rules	1600	431.20	1021	1218	1599	437.20	907.50	1470	1432	410.90	1266	1536
	time	1.75	1.32	3.27	0.52	1.71	1.34	3.25	0.55	1.66	1.32	3.37	0.63

Table 6: Results for the FRBCS ensembles

		Bagging+Greedy				Bagging+GRASP $\tau = 0.50$				Bagging + Random Subspace			
		Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar	Pima	Glass	Vehicle	Sonar
	5×2-cv	0.261	0.463	0.525	0.255	0.262	0.464	0.494	0.246	0.299	0.450	0.453	0.250
3 labels	#rules	8578	6208	6843	7282	8609	6289	7362	7951	7936	5671	8008	8174
5 #attr.	avg. #rules	171.55	124.16	136.87	145.65	172.18	125.77	147.24	159.03	158.71	113.42	160.16	163.47
	time	3.43	1.51	4.87	2.52	3.45	1.53	4.91	2.57	3.34	1.49	5.06	2.58
	5×2-cv	**0.235**	**0.396**	0.400	**0.240**	**0.234**	**0.405**	0.399	**0.220**	**0.260**	0.430	0.378	**0.221**
5 labels	#rules	29405	12877	22177	26769	29748	13302	25578	30068	27199	11998	30799	31824
5 #attr.	avg. #rules	588.11	257.54	443.55	535.37	594.95	266.04	511.56	601.36	543.97	239.96	615.97	636.47
	time	17.93	12.11	31.21	6.66	18.05	12.23	32.79	6.96	17.64	11.94	33.91	7.13
	5×2-cv	0.243	0.430	**0.375**	0.262	0.247	0.425	**0.353**	0.242	0.263	**0.402**	**0.330**	0.241
7 labels	#rules	64891	18633	48479	49587	65802	19272	54721	54684	59824	17999	67936	57298
5 #attr.	avg. #rules	1298	372.66	969.58	991.74	1316	385.45	1094	1094	1196	359.98	1359	1146
	time	84.70	67.36	166.51	24.72	85.27	68.27	170.48	25.49	82.12	66.06	174.24	25.57

Table 7: Statistical test for the comparison of the single FRBCS and the different FRBMCSs methodology. For each dataset, the best result is marked ('*') and the others are compared to it

		Best single classifier (app./labels)	Best ensemble (app./labels)	Best ens. selected (app./labels/fitness)	Best ens. TEFF (app./labels)
Pima	Approach	GRASP/5	GRASP/5	Greedy/5/LOFF	GRASP/5
	$\mu \pm \sigma$	0.246 ±0.00991	0.234 ±0.019	0.238 ±0.0167	0.240 ±0.0159
	Symbol	=	*	=	=
Glass	Approach	GRASP/5	Greedy/5	Greedy/3/LOFF	GRASP/3
	$\mu \pm \sigma$	0.375 ±0.0526	0.396 ±0.0568	0.358 ±0.0382	0.360 ±0.0507
	Symbol	=	=	*	=
Vehicle	Approach	GRASP/7	Random/7	Random/7/LOFF	GRASP/7
	$\mu \pm \sigma$	0.399 ±0.0262	0.330 ±0.0179	0.347 ±0.0196	0.337 ±0.0168
	Symbol	+	*	=	=
Sonar	Approach	Greedy/3	GRASP/5	Random/5/WCFF	GRASP/3
	$\mu \pm \sigma$	0.261 ±0.0463	0.220 ±0.0445	0.219 ±0.0264	0.216 ±0.0265
	Symbol	+	=	=	*

ISBN: 978-989-95079-6-8

result for each dataset is marked with a star '*'.

The best results (in average) are always obtained by the initial or the selected ensembles. Even if they are only significant for two datasets, they correspond to those with the largest dimension. Notice that, on sonar, the GA outperforms significantly the single classifier. The initial ensemble outperforms the best genetically selected FRBMCSs in one of the four cases (vehicle), whereas TEFF achieves the best result for the sonar dataset and WCFF for the sonar.

In the direct comparison, TEFF is the best choice two times for the vehicle and sonar dataset. The new fitness functions outperforms one of the four cases for pima with LOFF and glass dataset with WCFF.

Thus, combining bagging and the GA selection process to design FRBMCSs performs better for high dimensional problems with a large number of attributes, producing a smaller rule base while reducing the test errors in some cases, which was our original goal. When combining these two techniques with an advanced feature selection process we also get an improvement of the accuracy for datasets with higher dimensions (glass, vehicle and sonar, see Table 7).

6 Conclusions and future works

In this study, we extended our previously developed methodology in which a bagging approach together with a feature selection technique are used to train FRBMCSs, at a later stage selected by a multicriteria GA. Two new fitness functions were tested, the LOFF and the WCFF, based on one or two accuracy criteria (i.e., the training error and the likelihood). The generated selected FRBCS ensembles are performing correctly on classification problems with a significant number of features. By using abovementioned techniques, we would like to obtain FRCMCS dealing with high dimensional data.

One of the next steps we will consider in the future line is the design of a generic framework to define the multicriteria fitness function. At least two different information levels will be studied: the chromosome and the objective level. Furthermore, we would like to extend this study on larger data sets (more than 1,000 examples), to study the influence of other parameters (the GA parameters, the weighting coefficient in the WCFF, etc.), and to design more advanced genetic MCS selection techniques (e.g. the use of Pareto-based algorithms). Analysis of the different fuzzy rule genreation techniques and introduction a diversity criterion in the algorithm are another important points for future research.

References

[1] L. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley, 2004.

[2] T. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[3] D. Optiz and R. Maclin. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11:169–198, 1999.

[4] J. Canul-Reich, L. Shoemaker, and L. O. Hall. Ensembles of fuzzy classifiers. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6, London, 2007.

[5] W. Pedrycz and K.C. Kwak. Boosting of granular models. *Fuzzy Sets and Systems*, 157(22):2934–2953, 2006.

[6] S.J. Verzi, G.L. Heileman, and M. Georgiopoulos. Boosted ARTMAP: Modifications to fuzzy ARTMAP motivated by boosting theory. *Neural Networks*, 19(4):446–468, 2006.

[7] O. Cordón, A. Quirin, and L. Sánchez. A first study on bagging fuzzy rule-based classification systems with multicriteria genetic selection of the component classifiers. In *IEEE International Workshop on Genetic and Evolving Fuzzy Systems (GEFS)*, pages 11–16, Germany, 2008.

[8] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[9] H. Ishibuchi, T. Nakashima, and M. Nii. *Classification and Modeling With Linguistic Information Granules*. Springer, 2005.

[10] O. Cordón, A. Quirin, and L. Sánchez. On the use of bagging, mutual information-based feature selection and multicriteria genetic algorithms to design fuzzy rule-based classification ensembles. In *Int. Conf. on Hybrid Intelligent Systems*, pages 549–554, Barcelona, 2008.

[11] R. Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, 5(4):537–550, 1994.

[12] T. A. Feo and M. G. C. Resende. Greedy randomized adaptive search procedures. *J. of Global Optimization*, 6:109–133, 1995.

[13] O. Cordón and A. Quirin. Comparing two genetic overproduce-and-choose strategies for fuzzy rule-based multiclassification systems generated by bagging and mutual information-based feature selection. Technical Report #AFE 2008-05, European Centre for Soft Computing, Mieres, 2008.

[14] D. Partridge and W.B. Yates. Engineering multiversion neural-net systems. *Neural Computation*, 8(4):869–893, 1996.

[15] L.S. Oliveira, M. Morita, R. Sabourin, and F. Bortolozzi. Multi-objective genetic algorithms to create ensemble of classifiers. *Lecture Notes in Computer Science*, 3410:592–606, 2005.

[16] S.T. Hadjitodorov and L.I. Kuncheva. Selecting diversifying heuristics for cluster ensembles. *Lecture Notes in Computer Science*, 4472:200–209, 2007.

[17] G. Martínez-Munoz, D. Hernández-Lobato, and A. Suárez. Selection of decision stumps in bagging ensembles. *Lecture Notes in Computer Science*, 4668:319–328, 2007.

[18] B. Gabrys and D. Ruta. Genetic algorithms in classifier fusion. *Applied Soft Computing*, 6(4):337–347, 2006.

[19] O. Cordón, M.J. del Jesus, and F. Herrera. A proposal on reasoning methods in fuzzy rule-based classification systems. *International J. of Approximate Reasoning*, 20:21–45, 1999.

[20] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illlinois Press, 1949.

[21] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, 1996.

[22] T.G. Dietterich. Approximate statistical test for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1923, 1998.

[23] E. Lehmann. *Nonparametric Statistical Methods Based on Ranks*. McGraw-Hill, New-York, 1975.

ISBN: 978-989-95079-6-8

A new Fuzzy Noise-rejection Data Partitioning Algorithm with Revised Mahalanobis Distance

M.H. Fazel Zarandi, Milad Avazbeigi[1] I.B. Turksen[2]

1 Department of Industrial Engineering, Amirkabir University of Technology
Tehran, Iran
2 Department of Industrial Engineering, TOBB Economy and Technology University
Ankara, Turkey
Email: zarandi@cic.aut.ac.ir, miladavazbeigi@cic.aut.ac.ir, bturksen@etu.edu.tr

Abstract—*Fuzzy C-Means (FCM) and hard clustering are the most common tools for data partitioning. However, the presence of noisy observations in the data may cause generation of completely unreliable partitions from these clustering algorithms. Also, application of the Euclidean distance in FCM only produces spherical clusters. In this paper, a new noise-rejection clustering algorithm based on Mahalanobis distance is presented which is able to detect the noise and outlier data and also ellipsoidal clusters. Unlike the traditional FCM, the proposed clustering tool provides much efficient data partitioning capabilities in the presence of noise and outliers. For validation of the proposed model, the model is applied to different noisy data sets.*

Keywords— Cluster Validity Index (CVI), Fuzzy C-Means (FCM), Possibilistic C-means (PCM), Revised Gustafson-Kessel (GK), Revised Mahalanobis Distance.

1 Introduction

Clustering methods have been extensively used in computer vision and pattern recognition. Fuzzy clustering methods have shown spectacular ability to detect not only volume clusters, but also clusters which are actually thin shells, i.e. curves and surfaces.

Most analytic fuzzy clustering approaches are derived from the fuzzy C-means (FCM) algorithm. FCM uses the probabilistic constraint that the membership of a data point across classes sums to 1. The constraint is used to generate the memberships update equations for an iterative algorithm. The memberships resulting from FCM and its derivative however, do not always correspond to the intuitive concept of belonging or compatibility. Moreover, the algorithms have considerable trouble in noisy environments. Reference [1] summarizes the main problems of classic FCM as follows:

- In order to get the optimal partition, initial locations of the cluster centers should be assigned. The FCM algorithm always converges to a local extreme. Different choices of initial cluster centers may lead to a different extrema.
- The scientific basis for the choice of m, the weighting exponent, is still not clear.
- The optimum number of clusters in the data is assigned a priori. There should be a criterion to assign the optimal number of clusters.

To overcome these problems and drawbacks, first, [2] introduces a new method for fuzzy clustering called possibilistic fuzzy clustering. Their approach differs from the previous clustering methods in that the resulting partition of the data can be interpreted as a possibilistic partition, and the membership values may be interpreted as degrees of possibility of the points belonging to the classes, i.e., the compatibilities of the points with the class prototypes. They construct an appropriate objective function whose minimum will characterize a good possibilistic partition of the data, and derive the membership and prototype update equations from necessary conditions for minimization of the related criterion function. Next nominated work is [1]. Melek et al. in [1] show how their PCM addresses the mentioned problems of the FCM for some example cases. However, their method has some limitations which are the concentration of this paper.

The Rest of the paper is organized as follows: Section 2 discusses the drawbacks of the previous PCM's. Then, in section 3 the proposed method is presented in order to overcome these limitations. Section 4 applies the method for different data sets to verify and validate the proposed method.

2 Limitations of the traditional PCM's

Reference [1] uses Euclidean distance in all data partitioning steps. However, clearly, Euclidean distance usually fails to recognize the appropriate shape of the clusters in complex data sets. Especially, when data include ellipsoidal shapes, Euclidean distance loses sight to those data which should be considered in a cluster from data which should not be included in that cluster.

Moreover, the cluster validity index (*CVI*) applied in [1], suffers from linear behaviour of the Euclidean distance. There are many samples of data sets which the applied *CVI* is not able to recognize the number of clusters correctly.

Another limitation is about noise identification procedure. Application of Euclidean distance can mislead the noise rejection procedure. This problem is discussed in the related section of the proposed algorithm.

While Euclidean distance implies the above mentioned limitations, Mahalanobis distance can mitigate or in some cases can overcome the cited limitations completely. Another advantage of the Mahalanobis distance is that the

ISBN: 978-989-95079-6-8

Mahalanobis distance can identify both spherical and ellipsoidal clusters correctly.

3 Improvements to the traditional PCM's

As quoted in introduction, the proposed method applies the Mahalanobis distance for noise rejection problem. However, when applying mahalanobis distance, some difficulties may occur with covariance matrix.

In this section, first, the problem with calculation of the Covariance Matrix in Mahalanobis distance is discussed. An estimation of Covariance Matrix is introduced to mitigate the problem. Then, new Mahalanobis distance and revised Gustafson-Kessel Clustering are presented based on the estimation. Next, the revised Gustafson-Kessel Clustering method with the new Mahalanobis distance is integrated into the PCM which is presented in [1].

Moreover, a new *CVI* is applied which is based on proximity of two fuzzy sets. The new CVI is independent from the distance type and therefore can be integrated to the method with Mahalanobis distance. This *CVI* is used in [3].

3.1 Revised GK with revised Mahalanobis

In statistics, Mahalanobis distance is a distance measure introduced by P. C. Mahalanobis in 1936. It is based on correlations between variables by which different patterns can be identified and analyzed. It is a useful way of determining similarity of an unknown sample set to a known one. It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant, i.e. not dependent on the scale of measurements.

Mahalanobis distance can be defined as dissimilarity measure between two random vectors $\vec{x}, \vec{y}$ of the same distribution with the covariance matrix S:

$$D_M(\vec{x},\vec{\mu}) = \sqrt{(\vec{x}-\vec{\mu})^T S^{-1}(\vec{x}-\vec{\mu})}$$

The eigenvalues and eigenvectors of the covariance matrix describe the shape and orientation of the clusters in GK method of clustering which applies mahalanobis distance instead of Euclidean one. When an eigenvalues is zero or when the ration between the maximal and minimal eigenvalues, i.e., the condition number of F, is very large the matrix is nearly singular. In such a case, the inverse of covariance matrix cannot be calculated. Also the normalization to a fixed volume fails, as the determinant (the volume of the covariance matrix) becomes zero and the following formula thus cannot be applied in GK method [4]:

$$\det(F_i)^{1/n} F_i^{-1} \quad (1)$$

A straightforward way to avoid numerical problems is to constrain the ratio between the maximal and minimal eigenvalues such that it is smaller than some predefined threshold. When this threshold exceeds, the minimal eigenvalues is increased such that the ratio equals to the threshold and the covariance is reconstructed by [4]:

$$F = \Phi \Lambda \Phi^{-1} \quad (2)$$

where, Λ is the diagonal matrix containing the limited eigenvalues and Φ is a matrix whose columns are the corresponding eigenvectors.

Fig. 1 shows an example of a data set which cannot be clustered with standard GK. Using new estimation for covariance matrix, the numerical problems would be resolved and GK can appropriately find the correct number of clusters (example from [4]).

The above modification prevents the GK algorithm from running into numerical problems. However, as a result one can get clusters that are extremely long in the direction of the largest eigenvalues and have little relationship with real distribution of data. This can cause over fitting of the data and consequently one obtains a poor model [4].

This problem occurs mainly when the number of data points in a cluster becomes too low. In such a case, the computed covariance matrix is not a reliable estimate of the underlying data distribution [7]. One way to tackle this problem is to

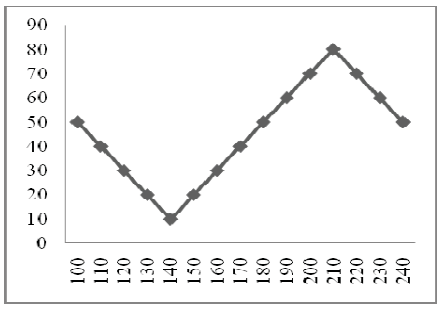


Figure 1: Linear Clusters

limit the ratio between maximal and minimal eigenvalues even further than described in the previous section. This will prevent the extreme elongation of the clusters. Another way is to add a scaled identity matrix to the covariance matrix. Reference [7] and [8] describe several different methods to improve the covariance estimation. Inspired by these methods, [4] proposes the following estimate for the GK algorithm and calculation of Mahalanobis distance:

$$F_i^{new} = (1-\gamma)F_i + \gamma \det(F_0)^{1/n} I \quad (3)$$

where $\gamma \epsilon [0,1]$ is the tuning parameter and F_0 is the covariance matrix of the whole data set. Depending on γ, the clusters are forced to have a more or less equal shape. When γ is 1, all covariance matrices are equal and have the same size, which of course limits the possibility of the algorithm to properly identify clusters.

For the complete description of the revised GK, readers can refer to [4]. The main steps of the algorithm are as follows [4]:

Repeat for *l = 1, 2, …, Ite#*

Step 1: Compute cluster prototypes (means)

$$v_i^{(l)} = \frac{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m Z_k}{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m}, 1 <= i <= K \quad (4)$$

Step 2: Compute the cluster covariance matrices

$$F_i = \frac{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m (Z_k - v_i^{(l)})(Z_k - v_i^{(l)})^T}{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m}, 1 \le i \le K \quad (5)$$

Add a scaled identity matrix:

$$F_i = (1-\gamma)F_i + \gamma \det(F_0)^{1/n} I, 1 \le i \le K \quad (6)$$

Extract eigenvalues λ_{ij} and eigenvectors ϕ_{ij} from F_i. Find $\lambda_{imax}=max_j\lambda_{ij}$ and set:

$$\lambda_{ij} = \frac{\lambda_{i\max}}{\beta} \; \forall j \; for \; which \; \frac{\lambda_{ij}}{\lambda_{i\max}} > \beta \quad (7)$$

Reconstruct F_i by

$$F_i = [\phi_{i1} \dots \phi_{in}] diag(\lambda_{i1}, \dots \lambda_{in}) [\phi_{i1} \dots \phi_{in}]^{-1}$$

Step 3: Compute the distances

$$D_{ikA_i}^2 = (Z_k - v_i^{(l)})^T [\rho_i \det(F_i)^{1/n} F_i^{-1}](Z_k - v_i^{(l)}) \quad (8)$$

where *1<=i<=k* and *1<=k<=N*.

Step 4: Update the partition matrix

$$for \; 1 \le k \le N$$

$$if \; D_{ikA_i} > 0 \; for \; 1 \le i \le K,$$

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{K} (D_{ikA_j} / D_{ikA_i})^{\frac{2}{m-1}}} \quad (9)$$

Otherwise

$$\mu_{ik}^{(l)} = 0 \; if \; D_{ikA_i} > 0, \; and \; \mu_{ik}^{(l)} \in [0,1] \quad (10)$$

with $\sum_{i=1}^{K} \mu_{ik}^{(l)} = 1 \; otherwise.$

Until $\left\| U^{(l)} - U^{(l-1)} \right\| < \varepsilon$.

Improved GK clustering does not have the problems that standard GK may face in some cases. In the next section, applied cluster validity index is described.

3.2 Cluster Validity Index (CVI)

This paper uses a cluster validity index proposed by [9].

Definition1: The relative similarity $S_{rel}(x_j{:}A_p,A_q)$ between two fuzzy sets A_p and A_q at x_j is defined as:

$$\frac{f(x_j{:}A_p \cap A_q)}{f(x_j{:}A_p \cap A_q) + f(x_j{:}A_p - A_q) + f(x_j{:}A_q - A_p)} \quad (11)$$

while $f(x_j{:}A_p \cap A_q) = u_{A_p}(x_j) \wedge u_{A_q}(x_j)$ and $\wedge$ is the minimum operator. Moreover for function of deference:

$$f(x_j{:}A_p - A_q) = Max(0, u_{A_P}(x_j) - u_{A_q}(x_j)) \quad (12)$$

Definition2: The relative similarity between two fuzzy sets A_p and A_q is defined as:

$$S_{rel}(A_p,A_q) = \sum_{j=1}^{n} S_{rel}(x_j : A_p,A_q) h(x_j) \quad (13)$$

where

$$h(x_j) = -\sum_{p=1}^{c} u_{A_p}(x_j) \log(u_{A_p}(x_j)) \quad (14)$$

Here, h is the entropy of datum x_j and $u_{A_p}(x_j)$ is the membership value of x_j to cluster A_p.

Definition3: The cluster validity index is as follows:

$$V(U,V;X) = \frac{2}{c(c-1)} \sum_{p \ne q}^{c} S_{rel}(A_p,A_q) \quad (15)$$

The optimal number of clusters is obtained by minimizing V over the range of c values (Number of clusters).

Moreover, some other popular CVI's such as "Xie and Beni" CVI [5] and "Known" CVI [6] are evaluated. Results show that the new CVI has better results and is able to find the number of clusters correctly in many complex situations.

The inputs to the CVI used in the proposed model are obtained by the revised GK.

4 The proposed method

In this section, we present our method base on the modifications presented in section 3.

For the selection of the weight exponent (m), it is suggested to be chosen far from its both extremes so as to ensure that the cluster validity index shows the optimum number of fuzzy clusters. A fuzzy total scatter matrix is defined in [10] as:

$$S_T = \sum_{k=1}^{N} (\sum_{i=1}^{K} (u_{ik})^m)(x_k - \overline{v})(x_k - \overline{v})^T \quad (16)$$

The trace of the fuzzy total scatter matrix decreases monotonically from a constant value z to zero as m varies from one to infinity. For data partitioning, a suitable value for m is that which gives a value for trace (s_T) equal to z/2 [11]. The constant value z is defined as:

$$z = trace(\sum_{k=1}^{N} [(x_k - \frac{1}{N}\sum_{k=1}^{N} x_k)(x_k - \frac{1}{N}\sum_{k=1}^{N} x_k)^T]) \quad (17)$$

Using the trace value, we determine the value of (m) which is the degree of fuzziness of the system. Next, using the introduced *CVI* and achieved (m), the appropriate number of clusters would be obtained.

After determination of m and number of clusters, we repeat the following procedure, iteratively:

We first implement the revised GK on the data set. To choose the initial cluster centers of the revised GK, we apply simple FCM for initialization of our clustering method. Other methods like [1] have applied AHC. The GK is sensitive to the initial cluster centers and initial membership function values. Our experiments show that unlike FCM, AHC is not a good method for the initialization of the revised GK.

Next, in order to find the data points that are ''too far'' from all cluster centers, [13] proposes the following index for each data point x_j:

$$W_j = \sum_{i=1}^{C} \left\| x_j - v_{hi} \right\|_A \quad (18)$$

where, *j = 1, 2, ... , N*, c is the number of clusters, and N is the number of data. The index W_j is the summation of the distance of the data point x_j to all cluster centers. This gives a measure of how far each data point is from the different cluster centers assigned in the first step of the algorithm. The noise is identified through the data points that have large

ISBN: 978-989-95079-6-8

values of W_j and therefore, a threshold X is assigned to trim these outliers from the data set. The value of the threshold depends on the range of the input data to the algorithm. While [1] applies (18) to find the W_j, we proposed the following equation in order to find the W_j:

$$W_j = \underset{i=1}{\overset{C}{Mean}}(WGHT_i \times \left\| x_j - \upsilon_{hi} \right\|_A) \quad (19)$$

where, "$WGHT_i$" is a weight assigned to the *i-th* distance after that we sorted the distances of x_j from all cluster centers (c). Reference [1] does not consider such weight. It is clear that simply the summation of distances can not provide a good measure for detection of the outliers. Instead the proposed model can assign the largest weight to the smallest distance and the smallest weight to the largest distance vice versa using (19). Weights are the input to the model. After choosing the threshold, [1] computes:

$$z = \frac{\eta_n}{N} \quad (20)$$

where, η_n is the number of noise points and N is the total number of data. The percentage of ''good'' data points, i.e., inliers can then be calculated as:

$$\hat{z} = 1 - z \quad (21)$$

After identifying the percentage of inliers in the data, we compute the corresponding chi-square data distribution value [13]. Then, we calculate the cut-off distance:

$$u_{FC}cut^2 = \upsilon_i \chi^2 \quad (22)$$

where, υ_i is a resolution parameter that depends on the number of clusters, and χ^2 is the chi-square value computed by (21). By knowing the new cut-off distance, the optimum number of clusters, the degree of fuzziness, and the initial location of the clusters centers, we calculate the membership matrix through (23):

$$u_{ij} = \frac{1}{1 + \{\frac{d^2(x_j, \upsilon_i)}{\upsilon_i}\}^{\frac{1}{m-1}}} \quad (23)$$

It should be noted that (20), (21), (22) and (23) are presented in [1]. In the proposed PCM, we use these equations with revised Mahalanobis distance.

5 Results of Experiments

In this section, the proposed model is applied for handling four different cases. The parameters of the proposed model for these cases are summarized in table 1:

- *NOC*: Number of Clusters
- *m* is the degree of fuzziness of system obtained through 16 and 17.
- *Ω* is the cut-off distance
- *Beta and Gamma* are the parameters of (3) and (7).
- *#Itr*: The number of iteration which PCM goes on.
- *W: WGHT* in (19).
- *Error:* If the changes in the value of membership functions were smaller than "*Error*", the algorithm stops.

In all cases, *Beta* is 1.00E+16, *#Itr* is 7 and *Error* is 1.00E-07. It should be noted that the data of the cases and their related configuration are very similar to [9] and [3]. However, we regenerated the data by ourselves.

Table 1: Parameters of the algorithm for different cases

Parameter	***NOC***	***m***	***Ω***	***Gamma***	***W***
Case 1	4	3	210	1.00E-07	[6 2 2 1]
Case 2	4	2	375	1.00E-09	[8 3 2 2]
Case 3	5	2	315	1.00E-01	[8 3 2 2 1]
Case 4	5	3	4590	1.00E-07	[8 4 3 2 1]
Case 5	5	3	805	1.00E-01	[8 4 3 2 1]

It should be mentioned that the experiments can't be applied for the previous noise rejection methods in the literature. This is because of the ellipsoidal forms of data which can't be handled using Euclidean distance. To make this clear, in Fig. 2, the clustering methods based on Euclidean distance apparently fail to recognize the shapes of clusters and noises correctly. That's why, only experiments are applied for the proposed method.

5.1 Case Study I

Every data cloud includes 500 data and 50 data are randomly generated as noise. Main data and noises both have Gaussian distribution.

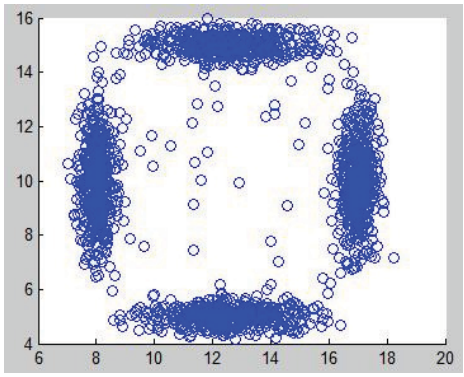


Figure 2: Data before clustering

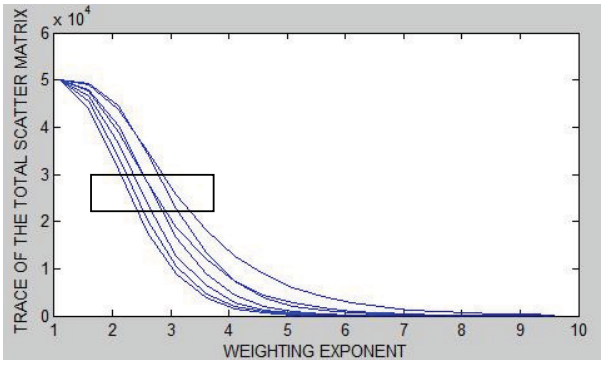


Figure 3: Finding the appropriate *m* in case study I

To determine the exponent of fuzziness (*m*), the total scatter matrix is calculated for different values of m and different

ISBN: 978-989-95079-6-8

number of clusters as is shown in (16). Then using z in (17), the appropriate m is chosen.
The trace value calculated as 5.8315e+004. We use half of trace value (2.9157e+004) to determine the optimum m (Fig. 3).

To choose the correct number of clusters, we need to calculate the CVI for different number of clusters. Fig. 4 shows the applied CVI for case study I. As the figure shows, the optimum number of clusters is equal to four.
Base on Fig. 3 and Fig. 4, we consider fuzziness exponent (m) equal to 2.5 and the number of clusters equal to 4. The result of the clustering is shown in Fig. 5. The result shows that the proposed method can identify the noises from the main data accurately. The expansion of the eclipse shape clusters can be limited by Ω as an input to the model.

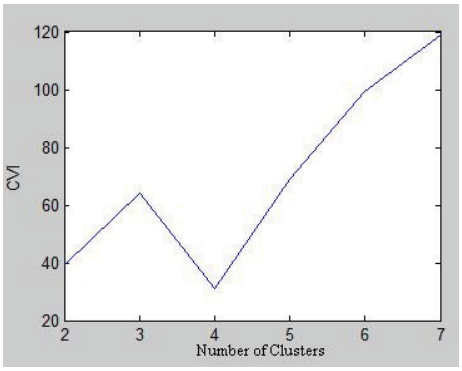


Figure 4: Identification of the optimum number of clusters

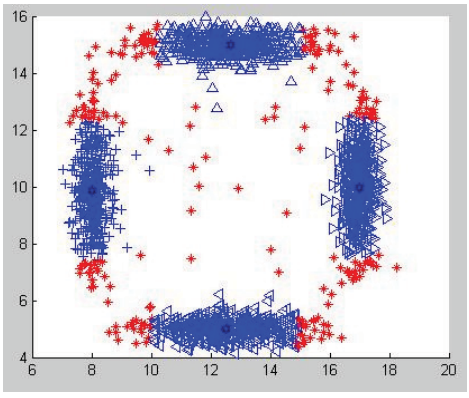


Figure 5: Data after Clustering

It is clear that, the method presented in [1] can not identify the correct shape of the clusters because of the spherical behaviour of the data. In Fig. 5, data which are shown with star "*", have a maximum membership function value lower than 0.2, hence are identified as noise.
In the other cases, each data cloud includes 500 data and 50 data are randomly generated as noise. Main data and noises both have Gaussian distribution. Also only the figures related to clustering and choosing the appropriate number of clusters are presented.

5.2 *Case Study II*

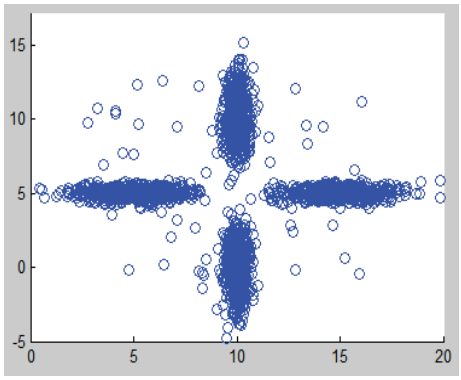


Figure 6: Data before clustering

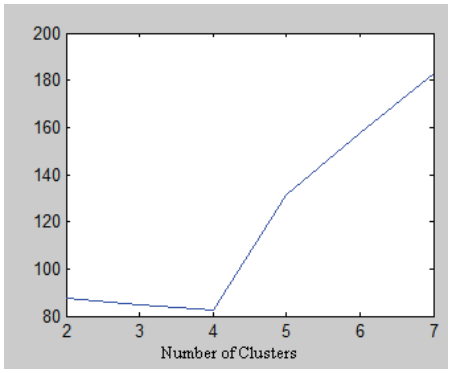


Figure 7: Identification of the optimum number of clusters

This case is presented to show the strength of the used CVI and the effect of Ω. As Figure 8 shows, the combination of these two elements in the proposed model, enable the model to identify the number of clusters, the shapes of the clusters and finally the noises accurately.

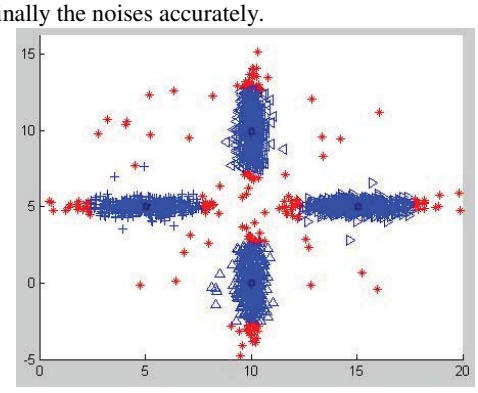


Figure 8: Data after Clustering

5.3 *Case Study III*

Fig 9 shows the data set and noises.

ISBN: 978-989-95079-6-8

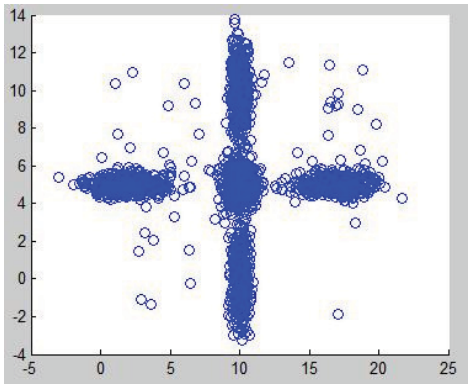


Figure 9: Data before clustering

As Fig. 10 shows, the optimum number of clusters is equal to five. This case is presented to validate that whether the method is able to find the spherical and ellipsoidal shapes simultaneously or is not. As Figure 11 shows, the method could find the circle at the center of the figure and four eclipses around the circle accurately. Also the noises are detected accurately as shown in Figure 11.

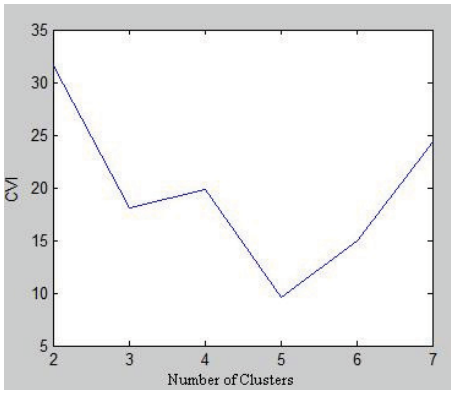


Figure 10: Identification of the optimum number of clusters

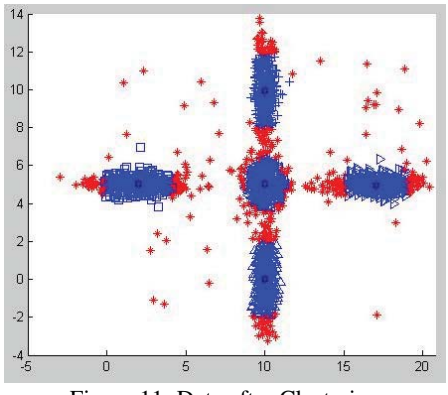


Figure 11: Data after Clustering

5.4 Case Study IV

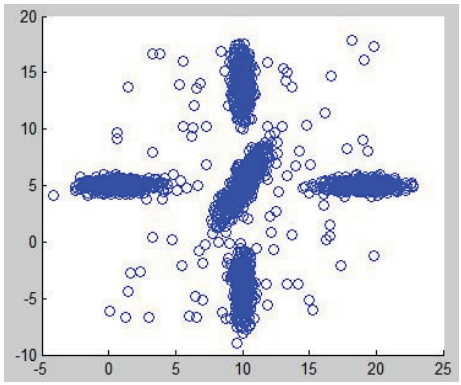


Figure 12: Data before clustering

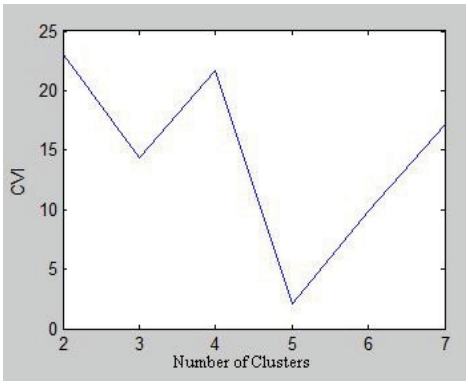


Figure 13: Identification of the optimum number of clusters

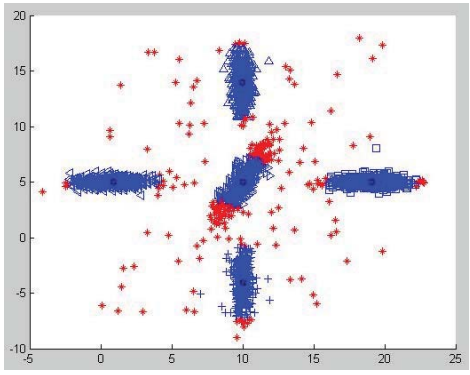


Figure 14: Data after Clustering

In this case, five eclipses are spread in different directions as shown in Figure 12. The clustering results show that the method can find the complex behaviors of data in different directions accurately as well as simple cases.

6 Conclusions

In this paper, a new noise rejection clustering method base on Mahalanobis distance and a different cluster validity index is presented. The new method can be distinguished from the existing methods in the literature from the following points:

ISBN: 978-989-95079-6-8

- The noise rejection clustering methods existing in the literature mainly use Euclidean distance in their clustering method. The proposed method, applies Mahalanobis distance which enables the method to identify the ellipsoidal behavior of data besides spherical behavior.
- In order to find the appropriate number of clusters, the method uses a well-defined cluster validity index which is independent from distance type.
- A new weighting system is attached to the noise rejection method, which helps for better detection of noises.

Method is applied to different cases and the results show that the method is capable of identification of noise and outliers within spherical and ellipsoidal data.

References

[1] W.W. Melek, A.A. Goldenberg, M.R. Emami, A fuzzy noise-rejection data partitioning algorithm, *International Journal of Approximate Reasoning*, 38: 1–17, 2005.

[2] R. Krishnapuram, J.M. Keller, A Possibilistic Approach to Clustering, *IEEE Transactions on Fuzzy Systems,* 2 (1): 98-110, 1993.

[3] M.H. Fazel Zarandi, B. Rezaee, I.B. Turksen, E. Neshat, A type-2 fuzzy rule-based expert system model for stock price analysis. *Expert Systems with Application*, 36:139-154, 2009.

[4] R. Babuska, R.J. van der Veen, U. Kaymak, Improved Covariance Estimation for Gustafson-Kessel Clustering. *IEEE*, 1081-1085, 2002.

[5] X.L. Xie, G. Beni, A validity measure for fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8): 841–847, 1991.

[6] S.H. Kwon, Cluster validity index for fuzzy clustering. *Electronics Letters*, 34(22): 2176–2177, 1998.

[7] S. Tadjudin and D.A. Landgrebe. Covariance estimation with limited training samples. *IEEE Transactions on Geosciences and Remote Sensing*, 37(4): 2113-2118, July 1999.

[8] J.F. Friedman, Regularize discriminate analysis, *J.R. Statist. Soc.*, 84:17-42, 1989.

[9] Kim, Y.I., Kim, D.W., Lee, D., Lee, K.H., A cluster validation index for GK cluster analysis based on relative degree of sharing, *Information Science*, 168:225-242, 2004.

[10] M. Sugeno, T. Yasukawa, A fuzzy-logic-based approach to qualitative modeling, *IEEE Transactions on Fuzzy Systems,* 1(1), 1993.

[11] M.R. Emami, I.B. Turksen, A.A. Goldenberg, Development of a systematic methodology of fuzzy logic modeling, *IEEE Transactions on Fuzzy Systems,* 6 (3): 346–361, 1998.

[12] J.H. Ward, Hierarchical grouping to optimize an objective function, *Journal of American Statistics Association,* 58: 236–244, 1963.

[13] W.W. Melek, Neurofuzzy Control of Modular and Reconfigurable Robots, PhD. Dissertation, University of Toronto, Canada, 2002.

ISBN: 978-989-95079-6-8

Determining OWA weights by maximizing consensus

José Luis García-Lapresta Bonifacio Llamazares Teresa Peña

PRESAD Research Group, Dept. of Applied Economics
University of Valladolid, Spain
Email: {lapresta, boni, maitepe}@eco.uva.es

Abstract— *In this paper we propose a method for generating OWA weighting vectors from the individual assessments on a set of alternatives in such a way that these weights maximize the consensus among individual assessments with respect to the outcome provided by the OWA operator.*

Keywords— OWA operators; consensus; distances; mathematical programming.

1 Introduction

In 1988 Yager [9] introduced OWA operators as a tool for aggregating numerical values in multi-criteria decision making. An OWA operator is similar to a weighted mean, but with the values of the variables previously ordered in a decreasing way. Thus, contrary to the weighted means, the weights are not associated with specific variables and, therefore, they are anonymous. Moreover, they satisfy other interesting properties, such as monotonicity, unanimity, continuity and compensativeness.

Initially, the weights of an OWA operator may be fixed taking into account the importance we want to give to the assessments. So, the outcome of an OWA operator may be the maximum, the minimum, the average or a median of the individual assessments, among a large number of possibilities.

It is important to note that the determination of the weights of OWA operators is a relevant issue since the origins of the theory of OWA operators. In this way, Yager [9] proposes to use linguistic quantifiers for generating the OWA weights; O'Hagan [6] generates the OWA weights by maximizing their entropy whenever a degree of orness has been fixed; Filev and Yager [3] consider an exponential smoothing approach for generating the OWA weights. After these seminal papers, a large variety of techniques have been proposed in the literature (see, for instance, Wang and Parkan [7] and Xu [8]).

When a group of individuals provides assessments on an alternative and these values are aggregated, it is relevant to know the degree of agreement or consensus among the individual assessments with respect to the aggregated value. In fact, it is desirable that the aggregation function used to obtain the collective value reflects the opinions of as many agents as possible. Using a specific OWA operator for aggregating individual assessments does not necessarily ensure such consensus for every opinion situation.

In our proposal, we do not fix the OWA weighting vector, but we generate an OWA operator for each profile of individual assessments, just one that maximizes the consensus (or equivalently, minimizes the disagreement) in the group with respect to the outcome provided by the OWA operator. More concretely, once the agents opinions are known, we first calculate the distances among individual assessments on the alternatives and the collective assessments generated by an arbitrary OWA operator. Secondly, we use an aggregation operator for obtaining a representative measure of disagreement from the individual assessments to the collective one. By solving a mathematical program, we obtain the weighting vector(s) that maximize(s) the consensus among individual and collective opinions.

The paper is organized as follows. Section 2 is devoted to introduce notation, basic notions and our proposal for generating an OWA operator for each profile of individual assessments. Section 3 contains some illustrative examples. Finally, Section 4 shows some open problems and further research.

2 A model for generating OWA weights

Consider a set of agents (experts or voters) $V = \{1, \dots, m\}$ ($m \geq 2$) who show their opinions on a set of alternatives $A = \{a_1, \dots, a_n\}$ ($n \geq 2$) through numbers in the interval $[0, 1]$.

A *profile* is a $m \times n$ matrix

$$P = \begin{pmatrix} a_1^1 & \cdots & a_j^1 & \cdots & a_n^1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1^i & \cdots & a_j^i & \cdots & a_n^i \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1^m & \cdots & a_j^m & \cdots & a_n^m \end{pmatrix}$$

where $a_j^i \in [0, 1]$ is the assessment that agent i assigns to alternative a_j. The set of profiles is denoted by $\mathcal{P}$.

Given a weighting vector $\boldsymbol{w} = (w_1, \dots, w_m) \in [0, 1]^m$ such that $\sum_{i=1}^m w_i = 1$, the *OWA operator associated with* $\boldsymbol{w}$ is the mapping $F_{\boldsymbol{w}} : [0, 1]^m \longrightarrow [0, 1]$ defined by

$$F_{\boldsymbol{w}}(x_1, \dots, x_m) = \sum_{i=1}^m w_i \cdot y_i$$

where y_i is the i-th greatest number of $\{x_1, \dots, x_m\}$. We use the following notation:

$$\mathcal{W} = \left\{ \boldsymbol{w} \in [0, 1]^m \;\middle|\; \sum_{i=1}^m w_i = 1 \right\}.$$

Given $\boldsymbol{w} \in \mathcal{W}$, from the opinions given by the agents on an alternative a_j, $\{a_j^1, \dots, a_j^m\}$, we generate a collective assessment on a_j:

$$v_j(\boldsymbol{w}) = F_{\boldsymbol{w}}\left(a_j^1, \dots, a_j^m\right).$$

We denote $\boldsymbol{v}(\boldsymbol{w}) = (v_1(\boldsymbol{w}), \dots, v_n(\boldsymbol{w}))$ the vector that contains the collective assessments on the alternatives of A generated by the OWA operator $F_{\boldsymbol{w}}$.

ISBN: 978-989-95079-6-8

By $\boldsymbol{a}^i = \left(a_1^i, \ldots, a_n^i\right)$ we denote the vector that contains the assessments of individual $i \in \{1, \ldots, m\}$ on the alternatives of A.

In our context, an *aggregation operator* is a continuous mapping $A : [0,1]^m \longrightarrow [0,1]$ that satisfies the following conditions:

1. *Monotonicity*, i.e., $A(x_1, \ldots, x_m) \leq A(y_1, \ldots, y_m)$ for all $(x_1, \ldots, x_m), (y_1, \ldots, y_m) \in [0,1]^m$ such that $x_i \leq y_i$ for every $i \in \{i, \ldots, m\}$.
2. *Unanimity*, i.e., $A(x, \ldots, x) = x$ for every $x \in [0,1]$.

It is easy to see that every aggregation operator is *compensative*, i.e.,

$$\min\{x_1, \ldots, x_m\} \leq A(x_1, \ldots, x_m) \leq \max\{x_1, \ldots, x_m\},$$

for every $(x_1, \ldots, x_m) \in [0,1]^m$.

On aggregation operators, see Fodor and Roubens [4], Grabisch, Orlovski and Yager [5], Calvo, Kolesárova, Komorníková and Mesiar [2] and Beliakov, Pradera and Calvo [1], among others.

2.1 The general model

In order to present our general model, it is necessary to fix two ingredients:

- A distance $d : [0,1]^n \times [0,1]^n \longrightarrow [0,1]$.
- An aggregation operator $A : [0,1]^m \longrightarrow [0,1]$.

Given a profile $P \in \mathcal{P}$, we propose to find weighting vector(s) $\boldsymbol{w} \in \mathcal{W}$ being solution(s) of the following mathematical program

$$\begin{array}{ll} \text{Min} & A\Big(d(\boldsymbol{a}^1, \boldsymbol{v}(\boldsymbol{w})), \ldots, d(\boldsymbol{a}^m, \boldsymbol{v}(\boldsymbol{w}))\Big) \\ \text{s. t. :} & \boldsymbol{w} \in \mathcal{W} \end{array} \tag{1}$$

Notice that from continuity of A and compactness of $\mathcal{W}$, the existence of solution(s) in (1) is always guaranteed.

Among the large variety of distances and aggregation operators that we may use in (1), we present with more detail those cases where Manhattan and Chebyshev distances are used, and the aggregation operators are the arithmetic mean and the maximum.

The Manhattan distance is defined by

$$d_1\big((x_1, \ldots, x_n), (y_1, \ldots, y_n)\big) = \sum_{i=1}^{n} |x_i - y_i|.$$

The Chebyshev distance is defined by

$$\begin{array}{l} d_\infty\big((x_1, \ldots, x_n), (y_1, \ldots, y_n)\big) = \\ = \max\{|x_1 - y_1|, \ldots, |x_n - y_n|\}. \end{array}$$

2.2 Using the arithmetic mean as aggregation operator

If we consider the arithmetic mean as aggregation operator, then (1) is equivalent to find the weighting vector that minimizes the sum of distances between the individual assessments and the collective assessments generated by the OWA operator associated with that weighting vector, i.e.

$$\begin{array}{ll} \text{Min} & \displaystyle\sum_{i=1}^{m} d\Big((a_1^i, \ldots, a_n^i), (v_1(\boldsymbol{w}), \ldots, v_n(\boldsymbol{w}))\Big) \\ \text{s. t. :} & \boldsymbol{w} \in \mathcal{W} \end{array} \tag{2}$$

1. If we use the Manhattan distance, then it is necessary to solve the following mathematical program:

$$\begin{array}{l} \text{Min} \displaystyle\sum_{i=1}^{m} \Big(|a_1^i - v_1(\boldsymbol{w})| + \cdots + |a_n^i - v_n(\boldsymbol{w})|\Big) \\ \text{s. t. :} \ w_1 \geq 0, \ldots, w_m \geq 0, \ w_1 + \cdots + w_m = 1 \end{array}$$

2. If we use the Chebyshev distance, then it is necessary to solve the following mathematical program:

$$\begin{array}{l} \text{Min} \displaystyle\sum_{i=1}^{m} \max\Big\{|a_1^i - v_1(\boldsymbol{w})|, \ldots, |a_n^i - v_n(\boldsymbol{w})|\Big\} \\ \text{s. t. :} \ w_1 \geq 0, \ldots, w_m \geq 0, \ w_1 + \cdots + w_m = 1 \end{array}$$

In the first case, i.e., when the Manhattan distance is used, it is possible to give the analytical solution of the problem.

Proposition 1 *The solution of Problem (2) for the Manhattan distance is the median operator.*

Proof

When the Manhattan distance is used in Problem (2), we obtain the following mathematical program:

$$\begin{array}{l} \displaystyle\min_{\boldsymbol{w} \in \mathcal{W}} \sum_{i=1}^{m} \sum_{j=1}^{n} |a_j^i - v_j(\boldsymbol{w})| = \\ \displaystyle\min_{\boldsymbol{w} \in \mathcal{W}} \sum_{j=1}^{n} \sum_{i=1}^{m} |a_j^i - v_j(\boldsymbol{w})|. \end{array} \tag{3}$$

On the other hand, it is known that given $x_1, \ldots, x_m \in \mathbb{R}$, the median operator is the solution of the following problem:

$$\min_{x \in \mathbb{R}} \frac{1}{m} \sum_{i=1}^{m} |x_i - x|.$$

Therefore, it is also the solution of

$$\min_{\boldsymbol{w} \in \mathcal{W}} \sum_{i=1}^{m} |a_j^i - v_j(\boldsymbol{w})|.$$

for all $j \in \{1, \ldots, n\}$, and, consequently it is the solution of Problem (3). □

2.3 Using the maximum as aggregation operator

We now consider the maximum as aggregation operator. Thus, with (4) we look for the weighting vector that minimizes the maximum distance between the individual assessments and the collective assessments generated by the OWA operator associated with that weighting vector, i.e.

$$\begin{array}{ll} \text{Min} & \displaystyle\max_{i=1,\ldots,m} \Big\{d\Big((a_1^i, \ldots, a_n^i), (v_1(\boldsymbol{w}), \ldots, v_n(\boldsymbol{w}))\Big)\Big\} \\ \text{s. t. :} & \boldsymbol{w} \in \mathcal{W} \end{array} \tag{4}$$

1. If we use the Manhattan distance, then it is necessary to solve the following mathematical program:

$$\begin{array}{l} \text{Min} \displaystyle\max_{i=1,\ldots,m} \Big\{|a_1^i - v_1(\boldsymbol{w})| + \cdots + |a_n^i - v_n(\boldsymbol{w})|\Big\} \\ \text{s. t. :} \ w_1 \geq 0, \ldots, w_m \geq 0, \ w_1 + \cdots + w_m = 1 \end{array}$$

ISBN: 978-989-95079-6-8

2. If we use the Chebyshev distance, then it is necessary to solve the following mathematical program:

$$\text{Min} \max_{i=1,\ldots,m} \max \Big\{ |a_1^i - v_1(\boldsymbol{w})|, \ldots, |a_n^i - v_n(\boldsymbol{w})| \Big\}$$
$$\text{s. t.: } w_1 \geq 0, \ldots, w_m \geq 0,\ w_1 + \cdots + w_m = 1$$

In this last case, i.e., when the Chebyshev distance is used, the solution of the mathematical program generates the mid-range OWA operator, as we show in the following proposition.

Proposition 2 *The solution of Problem (4) for the Chebyshev distance is the weighting vector* $\boldsymbol{w}^*$ *given by*

$$w_i^* = \begin{cases} 0.5, & \textit{if } i \in \{1, m\}, \\ 0, & \textit{otherwise.} \end{cases}$$

Proof

When the Chebyshev distance is used in Problem (4), we obtain the following mathematical program:

$$\min_{\boldsymbol{w}\in\mathcal{W}} \max_{i=1,\ldots,m} \max_{j=1,\ldots,n} \Big\{ |a_j^i - v_j(\boldsymbol{w})| \Big\} =$$
$$\min_{\boldsymbol{w}\in\mathcal{W}} \max_{j=1,\ldots,n} \max_{i=1,\ldots,m} \Big\{ |a_j^i - v_j(\boldsymbol{w})| \Big\}. \qquad (5)$$

On the other hand, it is obvious that for all $j \in \{1, \ldots, n\}$, the weighting vector given by

$$w_i^* = \begin{cases} 0.5, & \text{if } i \in \{1, m\}, \\ 0, & \text{otherwise,} \end{cases}$$

is the solution of the following problem:

$$\min_{\boldsymbol{w}\in\mathcal{W}} \max_{i=1,\ldots,m} \Big\{ |a_j^i - v_j(\boldsymbol{w})| \Big\}.$$

Therefore, $\boldsymbol{w}^*$ is also the solution of Problem (5). □

2.4 Restricted models

The general model (1) does not impose any restriction to the OWA weighting vectors that maximize the consensus among the individual assessments and the collective one. However, in some cases it could be interesting to consider some requirements for these weights by following a desired pattern, through a subclass $\mathcal{W}^* \subset \mathcal{W}$. For instance, we may restrict the search of the weighting vectors within one of the following cases.

1. Weighting vectors with a fixed *orness* or *attitudinal character* $\alpha \in (0, 1)$ (Yager [9]):

$$\mathcal{W}_\alpha^1 = \left\{ \boldsymbol{w} \in \mathcal{W} \;\middle|\; \frac{1}{m-1} \sum_{i=1}^{m} (m-i) w_i = \alpha \right\}.$$

2. Symmetric weights:

$$\mathcal{W}^2 = \left\{ \boldsymbol{w} \in \mathcal{W} \mid w_i = w_{m+1-i}\ \forall i \in \left\{ 1, \ldots, \left[\frac{m}{2} \right] \right\} \right\}.$$

3. Centered OWAs (after Yager [10]):

$$\mathcal{W}^3 = \left\{ \boldsymbol{w} \in \mathcal{W}^2 \mid w_1 \leq w_2 \leq \cdots \leq w_{\left[\frac{m+1}{2}\right]} \right\}.$$

4. Trimmed OWAs:

$$\mathcal{W}_1^4 = \{ \boldsymbol{w} \in \mathcal{W} \mid w_1 = w_m = 0 \},$$
$$\mathcal{W}_2^4 = \{ \boldsymbol{w} \in \mathcal{W} \mid w_1 = w_2 = w_{m-1} = w_m = 0 \},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

It is worth noting that $\mathcal{W}^3 \subseteq \mathcal{W}^2 \subseteq \mathcal{W}_{0.5}^1 \subseteq \mathcal{W}$.

Analogously to the general model, consider a distance $d : [0,1]^n \times [0,1]^n \longrightarrow [0,1]$ and an aggregation operator $A : [0,1]^m \longrightarrow [0,1]$. Given a profile $P \in \mathcal{P}$, we now propose to find weighting vector(s) $\boldsymbol{w} \in \mathcal{W}^*$ being solution(s) of the following mathematical program

$$\begin{array}{ll} \text{Min} & A\Big(d(\boldsymbol{a}^1, \boldsymbol{v}(\boldsymbol{w})), \ldots, d(\boldsymbol{a}^m, \boldsymbol{v}(\boldsymbol{w}))\Big) \\ \text{s. t.:} & \boldsymbol{w} \in \mathcal{W}^* \end{array} \qquad (6)$$

Notice that from continuity of A and compactness of $\mathcal{W}^*$ in the previous cases, the existence of solution(s) in (6) is always guaranteed.

3 Some illustrative examples

The following matrix shows the opinions of four experts on three alternatives.

$$P = \begin{pmatrix} 0.7 & 0.6 & 0.1 \\ 0 & 0.5 & 0.8 \\ 0.6 & 0.1 & 1 \\ 0.6 & 0.7 & 0 \end{pmatrix}$$

Case 1. If we use Model (2) with the Chebyshev distance, the following mathematical programming must be solved:

Min
$$\max\{|0.7 - v_1(\boldsymbol{w})|, |0.6 - v_2(\boldsymbol{w})|, |0.1 - v_3(\boldsymbol{w})|\} +$$
$$\max\{|0 - v_1(\boldsymbol{w})|, |0.5 - v_2(\boldsymbol{w})|, |0.8 - v_3(\boldsymbol{w})|\} +$$
$$\max\{|0.6 - v_1(\boldsymbol{w})|, |0.1 - v_2(\boldsymbol{w})|, |1 - v_3(\boldsymbol{w})|\} +$$
$$\max\{|0.6 - v_1(\boldsymbol{w})|, |0.7 - v_2(\boldsymbol{w})|, |0 - v_3(\boldsymbol{w})|\}$$

$$\text{s. t.: } w_1 \geq 0,\ w_2 \geq 0,\ w_3 \geq 0,\ w_4 \geq 0,$$
$$w_1 + w_2 + w_3 + w_4 = 1,$$

where

$$v_1(\boldsymbol{w}) = 0.7w_1 + 0.6w_2 + 0.6w_3,$$
$$v_2(\boldsymbol{w}) = 0.7w_1 + 0.6w_2 + 0.5w_3 + 0.1w_4,$$
$$v_3(\boldsymbol{w}) = 1w_1 + 0.8w_2 + 0.1w_3.$$

The previous problem is a non-smooth optimization problem. However, it can be easily replaced by an equivalent smooth linear problem by using some auxiliary variables. We have used LINGO software to solve this linear problem. The solution obtained is showed in Table 1. Likewise, that table summarizes the solutions obtained when some restrictions are imposed to the OWA weighting vectors in the problem.

ISBN: 978-989-95079-6-8

Table 1: Solutions for Case 1.

	w_1	w_2	w_3	w_4
$\mathcal{W}$	0.4706	0	0	0.5294
$\mathcal{W}^1_{0.75}$	0.75	0	0	0.25
$\mathcal{W}^1_{0.5}$	0.5	0	0	0.5
$\mathcal{W}^1_{0.25}$	0	0.375	0	0.625
$\mathcal{W}^2$	0.5	0	0	0.5
$\mathcal{W}^3$	0.25	0.25	0.25	0.25
$\mathcal{W}^4_1$	0	0.1429	0.8571	0

Case 2. If we consider Model (4) with the Manhattan distance, the mathematical programming to solve is:

$$\begin{aligned}\text{Min} \quad & \max\Big\{ \\ & |0.7 - v_1(\boldsymbol{w})| + |0.6 - v_2(\boldsymbol{w})| + |0.1 - v_3(\boldsymbol{w})|, \\ & |0 - v_1(\boldsymbol{w})| + |0.5 - v_2(\boldsymbol{w})| + |0.8 - v_3(\boldsymbol{w})|, \\ & |0.6 - v_1(\boldsymbol{w})| + |0.1 - v_2(\boldsymbol{w})| + |1 - v_3(\boldsymbol{w})|, \\ & |0.6 - v_1(\boldsymbol{w})| + |0.7 - v_2(\boldsymbol{w})| + |0 - v_3(\boldsymbol{w})|\Big\} \\ \text{s. t. :} \quad & w_1 \geq 0,\ w_2 \geq 0,\ w_3 \geq 0,\ w_4 \geq 0, \\ & w_1 + w_2 + w_3 + w_4 = 1,\end{aligned}$$

where

$$\begin{aligned} v_1(\boldsymbol{w}) &= 0.7w_1 + 0.6w_2 + 0.6w_3, \\ v_2(\boldsymbol{w}) &= 0.7w_1 + 0.6w_2 + 0.5w_3 + 0.1w_4, \\ v_3(\boldsymbol{w}) &= 1w_1 + 0.8w_2 + 0.1w_3. \end{aligned}$$

As in Case 1, this problem can be also substituted by an equivalent smooth linear problem by introducing some auxiliar variables. Table 2 shows the solution obtained by using LINGO software for different requirements on the weights.

Table 2: Solutions for Case 2.

	w_1	w_2	w_3	w_4
$\mathcal{W}$	0.667	0	0.222	0.111
$\mathcal{W}^1_{0.75}$	0.673	0	0.2307	0.0963
$\mathcal{W}^1_{0.5}$	0	0.722	0.056	0.222
$\mathcal{W}^1_{0.25}$	0	0.375	0	0.625
$\mathcal{W}^2$	0	0.5	0.5	0
$\mathcal{W}^3$	0	0.5	0.5	0
$\mathcal{W}^4_1$	0	0.834	0.166	0

4 Further research

It is worth noting that the solutions obtained in the above described models might not be unique. Even more, depending the weighting vector we choose, the outcomes provided by the corresponding OWA operator could be different. In this way, it would be necessary to provide an appropriate procedure for choosing a single weighting vector among the set of multiple solutions. This open problem constitutes part of our further research.

Acknowledgment

This work is partially supported by the Junta de Castilla y León (Consejería de Educación y Cultura, Projects VA092A08 and VA002B08), the Spanish Ministry of Education and Science (Project SEJ2006-04267/ECON), and ERDF.

References

[1] G. Beliakov, A. Pradera and T. Calvo. *Aggregation Functions: A Guide for Practicioners*. Springer, Heidelberg, 2007.

[2] T. Calvo, A. Kolesárova, M. Komorníková and R. Mesiar. Aggregation operators: Properties, classes and construction methods, in *Aggregation Operators: New Trends and Applications*, eds. T. Calvo, G. Mayor and R. Mesiar. Physica-Verlag, Heidelberg, 2002, pp. 3–104.

[3] D. Filev and R.R. Yager. On the issue of obtaining OWA operator weights. *Fuzzy Sets and Systems*, 94:157–169, 1998.

[4] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers, Dordrecht, 1994.

[5] M. Grabisch, S.A. Orlovski and R.R. Yager. Fuzzy aggregation of numerical preferences, in *Fuzzy Sets in Decision Analysis, Operations and Statistics*, ed. R. Slowinski. Kluwer Academic Publishers, Boston, 1998, pp. 31–68.

[6] M. O'Hagan. Aggregating template rule antecedents in real-time expert systems with fuzzy set logic. *Proc. 22nd Annual IEEE Asilomar Conference on Signals, Systems and Computers*. Pacific Grove, California, 1988, pp. 681–689.

[7] Y.M. Wang and C. Parkan. A minimax disparity approach for obtaining OWA operator weights. *Information Sciences*, 175:20–29, 2005.

[8] Z. Xu. An overview of methods for determining OWA weights. *International Journal of Intelligent Systems*, 20:843–865, 2005.

[9] R.R. Yager. On ordered weighted averaging operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18:183–190, 1988.

[10] R.R. Yager. Centered OWA operators. *Soft Computing*, 11:631–639, 2007.

ISBN: 978-989-95079-6-8

Weighted Trapezoidal Approximations of Fuzzy Numbers

Przemysław Grzegorzewski[1,2] Karolina Pasternak-Winiarska[2]

1.Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
2.Faculty of Mathematics and Information Science, Warsaw University of Technology
Plac Politechniki 1, 00-661 Warsaw, Poland
Email: pgrzeg@ibspan.waw.pl, karolcia.pw@gmail.com

Abstract— *Fuzzy number approximation by trapezoidal fuzzy numbers which preserves the expected interval is discussed. New algorithms for calculating the proper trapezoidal approximation of fuzzy numbers with respect to the distance based on bi-symmetrical weighted functions are proposed. It is shown that the adequate approximation operator is chosen with respect both to the global spread of a fuzzy number and the size of possible asymmetry between the spread of the left-hand and right-hand part of a fuzzy number.*

Keywords— fuzzy number, expected interval, trapezoidal approximation, width, weighted distance.

1 Introduction

Trapezoidal approximation of fuzzy numbers was considered by many authors (see, e.g. [1, 2, 3, 5, 11, 12, 9, 10, 15]). In [11] a list of criteria which trapezoidal approximation operators should possess was formulated. Then [18] considered the trapezoidal approximation under weighted distance. Some gaps in that paper were removed in [16] and [17] where so-called extended trapezoidal fuzzy numbers were introduced and applied.

The most common weighing applied to fuzzy numbers is a linearly increasing one (see, e.g. [5, 18]). However, it seems that a slightly modified weighing would be more natural and convenient in many situations. Such weighted functions, called regular bi-symmetrical weighted functions, were introduced and the nearest trapezoidal approximation operator preserving the expected interval, with respect to the distance based on such weighted functions, was suggested.

The paper is organized as follows. Firstly we recall some basic notions related to fuzzy numbers and present a few principle ideas connected with trapezoidal approximation. Then in Sec. 3 we introduce the notion of bi-symmetrical weighted functions. Finally, in Sec 4, we show trapezoidal approximation operators based on the weighted distance utilizing bi-symmetrical weighted functions and discuss some properties of the above mentioned operators.

2 Basic concepts

Let A denote a fuzzy number, i.e. such fuzzy subset A of the real line $\mathbb{R}$ with membership function $\mu_A : \mathbb{R} \to [0,1]$ which is (see [6]): normal (i.e. there exist an element x_0 such that $\mu_A(x_0) = 1$), fuzzy convex (i.e. $\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \mu_A(x_1) \wedge \mu_A(x_2)$, $\forall x_1, x_2 \in \mathbb{R}$, $\forall \lambda \in [0,1]$), μ_A is upper semicontinuous, supp$A$ is bounded, where supp$A = cl(\{x \in \mathbb{R} : \mu_A(x) > 0\})$, and cl is the closure operator. A space of all fuzzy numbers will be denoted by $\mathbb{F}(\mathbb{R})$.

Moreover, let $A_\alpha = \{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$, $\alpha \in (0,1]$, denote an α-cut of a fuzzy number A. As it is known, every α-cut of a fuzzy number is a closed interval, i.e. $A_\alpha = [A_L(\alpha), A_U(\alpha)]$, where $A_L(\alpha) = \inf\{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$ and $A_U(\alpha) = \sup\{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$.

The expected interval $EI(A)$ of a fuzzy number A is given by (see [7, 13])

$$\begin{aligned} EI(A) &= [EI_L(A), EI_U(A)] \\ &= \left[\int_0^1 A_L(\alpha)d\alpha, \int_0^1 A_U(\alpha)d\alpha\right]. \end{aligned} \tag{1}$$

The middle point of the expected interval given by

$$EV(A) = \frac{1}{2}\left(\int_0^1 A_L(\alpha)d\alpha + \int_0^1 A_U(\alpha)d\alpha\right) \tag{2}$$

is called the *expected value* of a fuzzy number and it represents the typical value of the fuzzy number A (see [7, 13]). Sometimes its generalization, called *weighted expected value*, might be interesting. It is defined as

$$EV_q(A) = (1-q)\int_0^1 A_L(\alpha)d\alpha + q\int_0^1 A_U(\alpha)d\alpha, \tag{3}$$

where $q \in [0,1]$ (see [8]).

Another useful parameter characterizing a fuzzy number is called the *width* of a fuzzy number (see [4]) and is defined by

$$w(A) = \int_{-\infty}^{\infty} \mu_A(x)dx = \int_0^1 (A_U(\alpha) - A_L(\alpha))d\alpha. \tag{4}$$

Suppose that for a certain reason or just for simplicity we want to find a suitable approximation of a fuzzy number under study. It seems that sufficiently effective simplification of the fuzzy number shape can be reached by the piecewise linear curves leading to triangle, trapezoidal or orthogonal membership functions. Since these three mentioned shapes are particular cases of the trapezoidal one, further on we will consider just the trapezoidal approximation of fuzzy numbers. It means that we want to substitute given fuzzy number A by the trapezoidal fuzzy number $T(A)$, i.e. by a fuzzy number with a following membership function

$$\mu_{T(A)}(x) = \begin{cases} 0 & \text{if} \quad x < t_1, \\ \frac{x-t_1}{t_2-t_1} & \text{if} \quad t_1 \leq x < t_2, \\ 1 & \text{if} \quad t_2 \leq x \leq t_3, \\ \frac{t_4-x}{t_4-t_3} & \text{if} \quad t_3 < x \leq t_4, \\ 0 & \text{if} \quad t_4 < x. \end{cases} \tag{5}$$

ISBN: 978-989-95079-6-8

Since we can do this in many ways we need some additional constraints which guarantee that our approximation would be reasonable. One of the most natural idea is to construct $T(A)$ that is the closest to the original fuzzy number A with respect to given distance d. Sometimes we add additional requirements which warrant that our approximation would possess some desired properties, like preservation fixed parameters or relations, continuity, etc. This problem was considered by many authors (see, e.g. [1, 2, 3, 5, 11, 12, 9, 10, 15, 16]). For example, it was suggested in [11] to consider *the nearest trapezoidal approximation operator preserving the expected interval*, i.e. the approximation operator T which produces a trapezoidal fuzzy number $T(A)$ that is the closest with respect to distance

$$d(A,T(A)) = \left(\int_0^1 [A_L(\alpha)-T(A)_L(\alpha)]^2 d\alpha + \int_0^1 [A_U(\alpha)-T(A)_U(\alpha)]^2 d\alpha\right)^{1/2} \quad (6)$$

to given original fuzzy number A among all trapezoidal fuzzy numbers having identical expected interval as the original one, i.e. satisfying a following condition

$$EI(T(A)) = EI(A). \quad (7)$$

It is worth noting the invariance of the expected interval assures many other properties (for more details we refer the reader to [11] where the broad list of desired requirements that the approximation operator should possess is also given). The research on this operator was continued in [3, 12, 9, 10].

Whatever trapezoidal approximation is considered the goal reduces to finding such real numbers $t_1 \le t_2 \le t_3 \le t_4$ that characterize $T(A) = T(t_1,t_2,t_3,t_4)$. It is so because any trapezoidal fuzzy number is completely described by four real numbers that are borders of its support and core. Let us mention that mathematical formulae for these points corresponding to the nearest trapezoidal approximation operator preserving the expected interval are given in [9].

In some applications other distances than (6) are more suitable. It is easily seen that all α-cuts in (6) are treated evenly. This feature is sometimes criticized by authors who claim that elements belonging to α_1-cut should be treated with the higher attention that those from α_2-cut if $\alpha_1 > \alpha_2$ because the membership degree for the first group is higher and so they are less uncertain. Such reasoning in trapezoidal approximation can be found in [18] devoted to weighted trapezoidal approximation. More precisely, the authors consider there a trapezoidal approximation with respect to the weighted distance

$$d_{ZL}(A,T(A)) = \left(\int_0^1 \alpha[A_L(\alpha)-T(A)_L(\alpha)]^2 d\alpha + \int_0^1 \alpha[A_U(\alpha)-T(A)_U(\alpha)]^2 d\alpha\right)^{1/2} \quad (8)$$

with increasing weighting function.

3 A bi-symmetrical weighted distance

Although such increasing weighting might be useful in some occasions, another weighted distances would be more interesting in general. This is a straightforward conclusion from the fact that the least informative α-cut is not zero but 0.5. Actually, situation $\mu_A(x) = 1$ leads to perfect information that x surely belongs to A. If $\mu_A(x)$ is close to 1 we'll say that x rather belongs to A. And conversely, $\mu_A(x) = 0$ shows that x surely does not belong to A (and belongs to $\neg A$) which is also a perfect information. Similarly, x such that $\mu_A(x)$ is close to 0 is interpreted as a point that rather does not belong to A. However, if $\mu_A(x) = 0.5$ we dot know how to classify x because it belong to A and to it's completion $\neg A$ with the same degree. The same happens if $\mu_A(x)$ is close to 0.5. Thus, to sum up, degrees of membership both high (close to 1) and low (close to 0) are much more informative than those close to 0.5. Hence, if we try to incorporate this quite obvious remark into practice we have to consider a so-called bi-symmetrical weighted distance suggested below. Before defining this distance we will introduce the notion of the bi-symmetrical weighted function.

Definition 1
A function $\lambda : [0,1] \to [0,1]$ symmetrical around $\frac{1}{2}$, i.e. $\lambda(\frac{1}{2}-\alpha) = \lambda(\frac{1}{2}+\alpha)$ for all $\alpha \in [0,\frac{1}{2}]$, which reaches its minimum in $\frac{1}{2}$, is called the bi-symmetrical weighted function. Moreover, the bi-symmetrical weighted function is called regular if

(a) $\lambda(\frac{1}{2}) = 0$,

(b) $\lambda(0) = \lambda(1) = 1$,

(c) $\int_0^1 \lambda(\alpha)d\alpha = \frac{1}{2}$.

Definition 2
For two arbitrary fuzzy numbers A and B with α-cuts $[A_L(\alpha), A_U(\alpha)]$ and $[B_L(\alpha), B_U(\alpha)]$, respectively, the quantity

$$d_\lambda(A,T(A)) = \left(\int_0^1 \lambda(\alpha)[A_L(\alpha)-B_L(\alpha)]^2 d\alpha + \int_0^1 \lambda(\alpha)[A_U(\alpha)-B_U(\alpha)]^2 d\alpha\right)^{1/2} \quad (9)$$

where $\lambda : [0,1] \to [0,1]$ is a bi-symmetrical (regular) weighted function is called the bi-symmetrical (regular) weighted distance between A and B based on λ.

One can, of course, propose many regular bi-symmetrical weighted functions and hence obtain different bi-symmetrical weighted distances. Further on we will consider mainly a following function

$$\lambda(\alpha) = \begin{cases} 1-2\alpha & \text{if} \quad \alpha \in [0,\frac{1}{2}], \\ 2\alpha-1 & \text{if} \quad \alpha \in [\frac{1}{2},1], \end{cases} \quad (10)$$

which is in some sense a bi-symmetrical counterpart of the increasing weighted function such as applied in [18].

4 Trapezoidal approximations based on bi-symmetrical weighted distances

Let us go back to the trapezoidal approximation operators $T : \mathbb{F}(\mathbb{R}) \to \mathbb{F}^T(\mathbb{R})$ which produce a trapezoidal fuzzy number $T(A)$ that is the closest to given original fuzzy number

ISBN: 978-989-95079-6-8

A among all trapezoidal fuzzy numbers having identical expected interval as the original one, i.e. satisfying (7). However, now we will look for the operators which minimize the bi-symmetrical weighted distance based on bi-symmetrical function (10).

It is easily seen that the α-cut of $T(A)$ is equal to $[t_1 + (t_2 - t_1)\,\alpha, t_4 - (t_4 - t_3)\alpha]$. Since a trapezoidal fuzzy number is completely described by four real numbers that are borders of its support and core, our goal reduces to finding such real numbers $t_1 \leq t_2 \leq t_3 \leq t_4$ that characterize $T(A) = T(t_1, t_2, t_3, t_4)$. Substituting it into (9) and (7) our problem might be expressed as follows: find t_1, t_2, t_3, t_4 which minimize

$$d(A, T(A)) = \tag{11}$$
$$= \left(\int_0^1 \lambda(\alpha) \left[A_L(\alpha) - (t_1 + (t_2 - t_1)\,\alpha) \right]^2 d\alpha \right.$$
$$\left. + \int_0^1 \lambda(\alpha) \left[A_U(\alpha) - (t_4 - (t_4 - t_3)\alpha) \right]^2 d\alpha \right)^{1/2}$$

with respect to conditions

$$\frac{t_1 + t_2}{2} = \int_0^1 A_L(\alpha) d\alpha, \tag{12}$$
$$\frac{t_3 + t_4}{2} = \int_0^1 A_U(\alpha) d\alpha \tag{13}$$
$$t_1 \leq t_2 \leq t_3 \leq t_4. \tag{14}$$

Before showing final results let us introduce some notation. Firstly let us notice that by Def. 1 the centroid of a bi-symmetrical weighted function λ is $\frac{1}{2}$. Therefore, the dispersion of the bi-symmetrical weighted function λ is given by

$$\eta = \int_0^1 (\alpha - \frac{1}{2})^2 \lambda(\alpha)\; d\alpha. \tag{15}$$

Easy computation shows that the dispersion of our bi-symmetrical weighted function (10) is $\frac{1}{16}$.

We also introduce another two parameters characterizing the dispersion of the left side and of the right side of given fuzzy number A with respect to considered bi-symmetrical weighted function λ.

Definition 3
The left (lower) spread of a fuzzy number A with respect to considered bi-symmetrical weighted function λ is a number $LSP_\lambda(A)$ given by

$$LSP_\lambda(A) = \frac{1}{2\eta} \int_0^1 (\alpha - \frac{1}{2}) \lambda(\alpha) A_L(\alpha)\; d\alpha, \tag{16}$$

while the right (upper) spread of a fuzzy number A with respect to considered bi-symmetrical weighted function λ is a following number $USP_\lambda(A)$

$$USP_\lambda(A) = \frac{1}{2\eta} \int_0^1 (\frac{1}{2} - \alpha) \lambda(\alpha) A_U(\alpha)\; d\alpha. \tag{17}$$

It can be shown that $LSP_\lambda(A) \geq 0$ and $USP_\lambda(A) \geq 0$.

Let us also denote the total spread of given fuzzy number A with respect to considered bi-symmetrical weighted function λ by $TSP_\lambda(A)$, i.e.

$$TSP_\lambda(A) = LSP_\lambda(A) + USP_\lambda(A), \tag{18}$$

while the difference between the right and the left spread of a fuzzy number will be denoted by $\Delta SP_\lambda(A)$, i.e.

$$\Delta SP_\lambda(A) = USP_\lambda(A) - LSP_\lambda(A). \tag{19}$$

It is easily seen that as $TSP_\lambda(A)$ is always nonnegative, $\Delta SP_\lambda(A)$ might be positive or negative as A is more asymmetrical to the right or to the left.

We can now formulate our main result.

Theorem 4
The nearest trapezoidal approximation operator preserving expected interval with respect to distance (11) for bi-symmetrical weighted function λ given by (10) is such operator $T : \mathbb{F}(\mathbb{R}) \to \mathbb{F}^T(\mathbb{R})$, which assigns a following trapezoidal fuzzy number $T(A) = T(t_1, t_2, t_3, t_4)$ for any fuzzy number A with α-cuts $[A_L(\alpha), A_U(\alpha)]$

(a) if $w(A) \geq TSP_\lambda(A)$ then

$$t_1 = EI_L(A) - LSP_\lambda(A), \tag{20}$$
$$t_2 = EI_L(A) + LSP_\lambda(A), \tag{21}$$
$$t_3 = EI_U(A) - USP_\lambda(A), \tag{22}$$
$$t_4 = EI_U(A) + USP_\lambda(A); \tag{23}$$

(b) if $|\Delta SP_\lambda(A)| \leq w(A) < TSP_\lambda(A)$ then

$$t_1 = EI_L(A) - \frac{1}{2} w(A) + \frac{1}{2} \Delta SP_\lambda(A), \tag{24}$$
$$t_2 = t_3 = EV(A) - \frac{1}{2} \Delta SP_\lambda(A), \tag{25}$$
$$t_4 = EI_U(A) + \frac{1}{2} w(A) + \frac{1}{2} \Delta SP_\lambda(A); \tag{26}$$

(c) if $w(A) < \Delta SP_\lambda(A)$ then

$$t_1 = t_2 = t_3 = EI_L(A), \tag{27}$$
$$t_4 = 2EI_U(A) - EI_L(A); \tag{28}$$

(d) if $w(A) < -\Delta SP_\lambda(A)$ then

$$t_1 = 2EI_L(A) - EI_U(A), \tag{29}$$
$$t_2 = t_3 = t_4 = EI_U(A). \tag{30}$$

The prove this result we apply the Karush-Kuhn-Tucker theorem for the local minimizer under given constraints.

Actually, we have received four different operators providing the nearest trapezoidal fuzzy number that preserves the expected value of the original fuzzy number, where T_1 leads to trapezoidal fuzzy number, T_2 stands for the operator that leads to triangular fuzzy number with two sides, while T_3 and T_4 produce triangular fuzzy numbers with the right side only or with the left side only, respectively.

Roughly speaking, we approximate a fuzzy number A by the trapezoidal approximation operator T_1 provided the total dispersion of given fuzzy number with respect to considered bi-symmetrical weighted function, measured by the sum of the lower and upper spread, is large enough. Otherwise, we will approximate A by a triangular number. It means that for less dispersed fuzzy numbers the solution is always a triangular fuzzy number.

ISBN: 978-989-95079-6-8

However, here we have also three possible situations: to approximate a fuzzy number A we apply operator T_2 provided the asymmetry of A is not too big (i.e.there is no big difference between the lower and upper spread). If A reveals high right asymmetry (i.e. the right spread is significantly larger than the lower spread) it would be approximated by a triangular fuzzy number with the right side only, produced by operator T_3. Otherwise, a fuzzy number with high left asymmetry would be approximated by a triangular fuzzy number with the left side only, produced by operator T_4.

It is worth noticing that we have obtained four possible solutions like in the problem with non-weighted distance considered in [3] or [9]. Moreover, operators T_3 and T_4 are identical as in [3] or [9]. It means that for very asymmetrical fuzzy numbers its nearest trapezoidal approximation preserving the expected interval remains independent whether we use weighted or non-weighted distance.

Our operators possess many desired properties, like invariance to translations and scale, monotonicity, continuity, etc. Some of those properties are guaranteed by the expected interval invariance. For the detailed list of requirements related to trapezoidal approximations we refer the reader to [11].

One may ask what happen if we use a different bi-symmetrical weighted function. Thus now, contrary to previous continuous weighted function (10) let us consider a non-continuous one

$$\lambda_1(\alpha) = \begin{cases} 1 & \text{if} \quad \alpha \in [0, \frac{1}{4}] \cup [\frac{3}{4}, 1], \\ 0 & \text{if} \quad \alpha \in (\frac{1}{4}, \frac{3}{4}), \end{cases} \tag{31}$$

which appreciates only elements with high or low degree of membership and does not take into account the other. Thus λ_1 corresponds to Pedrycz's viewpoint expressed in his shadowed sets (see [14]) where we consider only these points which rather belong to a set under study or those that rather do not belong to it. The other elements with intermediate membership degree form the so-called shadow. Easy computation shows that the dispersion of our bi-symmetrical weighted function (31) is $\frac{7}{96}$. It is interesting and worth stressing that the final solution both for λ_1 is identical with that obtained before, i.e. a following theorem holds.

Theorem 5
The nearest trapezoidal approximation operator preserving expected interval with respect to distance (11) for bi-symmetrical weighted function λ given by (31) is such operator $T : \mathbb{F}(\mathbb{R}) \to \mathbb{F}^T(\mathbb{R})$, which assigns a following trapezoidal fuzzy number $T(A) = T(t_1, t_2, t_3, t_4)$ for any fuzzy number A with α-cuts $[A_L(\alpha), A_U(\alpha)]$, has the same form as for for bi-symmetrical weighted function (10), i.e. given in Theorem 4.

5 Conclusion

In the present contribution we have considered the problem of trapezoidal approximation of fuzzy numbers for bi-symmetrical weighted functions. It was shown that the choice of adequate approximation operator depends mainly on the global spread of a fuzzy number and possible asymmetry between the spread of the left-hand and right-hand part of the original fuzzy number. It is interesting that identical operators for two different bi-symmetrical weighted functions were obtained. Thus one may ask whether the solution remains valid for other weighted function too. This would be the topic of our further research.

References

[1] S. Abbasbandy, B. Asady. The nearest approximation of a fuzzy quantity in parametric form. *Applied Mathematics and Computation*, 172:624-632, 2006.

[2] S. Abbasbandy, M. Amirfakhrian. The nearest trapezoidal form of a generalized LR fuzzy number. *International Journal of Approximate Reasoning*, 43:166-178, 2006.

[3] A. Ban. Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the expected interval. *Fuzzy Sets and Systems* 159:1327-1344, 2008.

[4] S. Chanas. On the interval approximation of a fuzzy number. *Fuzzy Sets and Systems*, 122:353–356, 2001.

[5] M. Delgado, M.A. Vila, W. Voxman. On a canonical representation of a fuzzy number. *Fuzzy Sets and Systems*, 93:125–135, 1998.

[6] D. Dubois, H. Prade. Operations on fuzzy numbers. *Int. J. Syst. Sci.*, 9:613–626, 1978.

[7] D. Dubois, H. Prade. The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24:279–300, 1987.

[8] P. Grzegorzewski. Metrics and orders in space of fuzzy numbers. *Fuzzy Sets and Systems*, 97:83–94, 1998.

[9] P. Grzegorzewski. Trapezoidal approximations of fuzzy numbers preserving the expected interval - algorithms and properties. *Fuzzy Sets and Systems*, 159:1354-1364, 2008.

[10] P. Grzegorzewski. New algorithms for trapezoidal approximation of fuzzy numbers preserving the expected interval. In: *Proceedings of the Twelfth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'08*, L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (Eds.), Spain, Torremolinos (Málaga), June 22-27, pp. 117-123, 2008.

[11] P. Grzegorzewski, E. Mrówka. Trapezoidal approximations of fuzzy numbers. *Fuzzy Sets and Systems*, 153:115–135, 2005.

[12] P. Grzegorzewski, E. Mrówka. Trapezoidal approximations of fuzzy numbers - revisited. *Fuzzy Sets and Systems*, 158:757–768, 2007.

[13] S. Heilpern. The expected value of a fuzzy number. *Fuzzy Sets and Systems*, 47:81–86, 1992.

[14] W. Pedrycz. Shadowed sets: representing and processing fuzzy sets. *IEEE Trans. Syst. Man. Cybern. B*, 2:103–109, 1998.

[15] C.T. Yeh. A note on trapezoidal approximations of fuzzy numbers. *Fuzzy Sets and Systems*, 158:747–754, 2007.

[16] C.T. Yeh. Trapezoidal and triangular approximations preserving the expected interval. *Fuzzy Sets and Systems*, 159:1345-1353, 2008.

[17] C.T. Yeh. On improving trapezoidal and triangular approximations of fuzzy numbers. *International Journal of Approximate Reasoning*, 48:297-313, 2008.

[18] W. Zeng, H. Li. Weighted triangular approximation of fuzzy numbers. *International Journal of Approximate Reasoning*, 46:137-150, 2007.

ISBN: 978-989-95079-6-8

Nonlinear-shaped Approximation of Fuzzy Numbers

Przemyslaw Grzegorzewski [1,2], Luciano Stefanini [3]

1. Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland.
2. Faculty of Mathematics and Information Sciences, Warsaw University of Technology, Warsaw, Poland.
3. Faculty of Economics, University of Urbino "Carlo Bo", Urbino, Italy.
Email: pgrzeg@ibspan.waw.pl, lucste@uniurb.it

Abstract— *In this paper, we present a general framework for the nearest approximation of fuzzy numbers simultaneously valid for the LR or the LU representations. We suggest two families of flexible monotonic functions characterized by four or five parameters that allow multiple approximation criteria and satisfy various requirements such as preservation of value, ambiguity, support, core, level set average, etc.*

Keywords— Trapezoidal Fuzzy Numbers, Approximation of Fuzzy Numbers, LR fuzzy numbers, Expected Interval, Shape Preserving Approximation

1 Introduction

The approximation of fuzzy numbers by linear or other simple shapes (e.g. polynomial, Gaussian, logistic) is an important problem in many fields and applications where extended or massive fuzzy computations are required. The recent literature has addressed both the linear case (see [3], [9], [10], [11], [12], [13], [18] and the references therein) and the reconstruction by nonlinear functions (e.g. [1], [2], [4], [15]).

Clearly, there are at least three key points to define:

(a) the basic general model to represent fuzzy numbers such as the LR (see [6], [7]) or the LU (see [8], [16], [17]) representations,

(b) the families of parameterized monotonic functions to be used which are sufficiently flexible to cover a large set of possible curves of membership functions,

(c) the criteria for the approximation, depending usually on the application at hand, such as least squares or other distance minimization, support or core or expected interval preservation, nearest ambiguity or value approximation among others (see [12]).

2 Basic representations and tools

In this paper, we present a general framework for the nearest approximation of fuzzy numbers valid simultaneously for the LR or the LU representations. To reach this goal we'll use special families of flexible monotonic curves described by sufficiently large number of parameters to allow multiple approximation criteria and satisfy various requirements.

Let us consider the following two families of nonlinear monotonic functions to use as shape generators ([16]).

1. the (2,2)-rational standardized monotonic spline:

$$p_{R2}(t;\beta_0,\beta_1) = \frac{t^2+\beta_0 t(1-t)}{1+(\beta_0+\beta_1-2)t(1-t)}, \qquad (1)$$

where $\beta_0,\beta_1 \geq 0$ and $t \in [0,1]$;

2. the mixed cubic-exponential spline

$$p_{MS}(t;\beta_0,\beta_1) = \frac{1}{a}\left[t^2(3-2t)+\beta_0 -\beta_0(1-t)^a+\beta_1 t^a\right], \qquad (2)$$

where $a = 1+\beta_0+\beta_1$, $\beta_0,\beta_1 \geq 0$ and $t \in [0,1]$.

It is easy to verify that for arbitrary nonnegative values of parameters β_0,β_1 the two families of functions are monotonic for all $t \in [0,1]$ and satisfy following conditions (the derivatives p' are made with respect to the first variable t):

$$p(0;\beta_0,\beta_1)=0, \qquad p(1;\beta_0,\beta_1)=1,$$
$$p\prime(0;\beta_0,\beta_1)=\beta_0, \qquad p\prime(1;\beta_0,\beta_1)=\beta_1.$$

Remark 1 All the proposed shape functions are monotonic over $[0,1]$. Note that it is not true for standard splines or other polynomials of degree greater than two.

Remark 2 Some special cases are of interest, i.e.

- if $\beta_0=\beta_1=1$ both p_{R2} and p_{MS} are linear;
- if $\beta_0+\beta_1=2$ both p_{R2} and p_{MS} are quadratic;
- if $\beta_0+\beta_1$ is integer (or zero), then p_{MS} is a polynomial function (increasing for all $t \in [0,1]$).

The properties given by following lemma are important.

Lemma 1 *For both p_{R2} and p_{MS} we have*

(i) $0 < \int_0^1 p(t;\beta_0,\beta_1)dt < 1 \quad \forall \beta_0,\beta_1 \geq 0$

(ii) $\lim_{\beta_1 \longrightarrow +\infty} \int_0^1 p(t;0,\beta_1)dt = 0$

(iii) $\lim_{\beta_0 \longrightarrow +\infty} \int_0^1 p(t;\beta_0,0)dt = 1.$

By Lemma 1 both p_{R2} and p_{MS} are able to "cover" the whole box $[0,1]\times[0,1]$ giving an infinite number of shape functions.

Some particular cases of these functions are of interest to meet specific requirements. For example, $p_{R2}(\frac{1}{2};\beta_0,\beta_1)=\frac{1}{2}$ if and only if $\beta_0=\beta_1=\beta$ and then $\int_0^1 p_{R2}(t;\beta,\beta)dt=\frac{1}{2}$ for all $\beta \geq 0$.

ISBN: 978-989-95079-6-8

A family of curves obtained from $p_{R2}(t;\beta_0,\beta_1)$ for different β is given in Fig. 1.

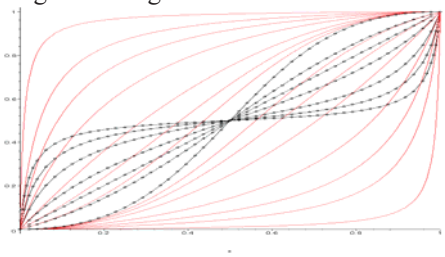


Figure 1. $p_{R2}(t;\beta_0,\beta_1)$ for different β_0,β_1; pointed lines are with $\beta_0=\beta_1$.

We can also use other single-parameter flexible forms and their combinations, like $p_{R2,1}(t,a)=p_{R2}(t;a,0)$, i.e.

$$p_{R2,1}(t,a)=\frac{t^2-at^2+at}{1+at-at^2-2t+2t^2}, \tag{3}$$

where $t\in[0,1]$, $a\geq 0$ or $p_{R2,2}(t,b)=p_{R2}(t;0,b)$, i.e.

$$p_{R2,2}(t,b)=\frac{t^2}{1+bt-bt^2-2t+2t^2}, \tag{4}$$

where $t\in[0,1]$, $b\geq 0$. Possible shapes obtained for different values of parameters a and b are shown in Fig. 2 and Fig. 3, respectively.

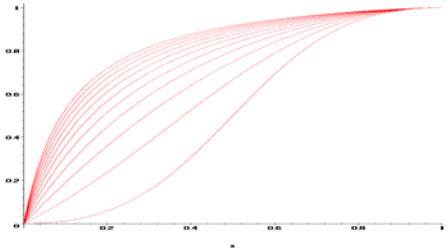


Figure 2. $p_{R2,1}(t;a)$ for different $a\geq 0$

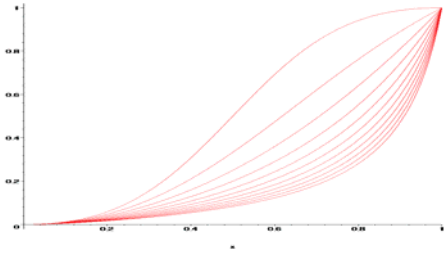


Figure 3. $p_{R2,2}(t;b)$ for different $b\geq 0$

Note that $p_{R2,1}(t;a)+p_{R2,2}(1-t;a)=1$ for all $t\in[0,1]$.

We can also combine the two using (here $\lambda\in[0,1]$)

$$P_{R2,\lambda}(t;a,b)=(1-\lambda)p_{R2,1}(t;a)+\lambda p_{R2,2}(t;b) \tag{5}$$

and have three parameters available.

Analogous constructions can be obtained by using p_{MS}, like $p_{MS,1}(t,a)=p_{MS}(t;a,0)$, i.e.

$$p_{MS,1}(t,a)=\frac{1}{1+a}[t^2(3-2t)+a-a(1-t)^{1+a}], \tag{6}$$

where $t\in[0,1]$, $a\geq 0$ or $p_{MS,2}(t,b)=p_{MS}(t;0,b)$, i.e.

$$p_{MS,2}(t,b)=\frac{1}{1+b}[t^2(3-2t)+bt^{1+b}], \tag{7}$$

where $t\in[0,1]$, $b\geq 0$, and combine the two with $\lambda\in[0,1]$:

$$P_{MS,\lambda}(t;a,b)=(1-\lambda)p_{MS,1}(t;a)+\lambda p_{MS,2}(t;b). \tag{8}$$

Examples of curves $P_{MS,0.5}(t;a,b)$ are in Fig. 4.

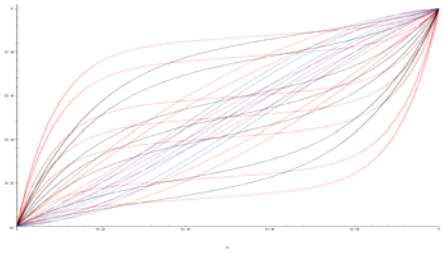


Figure 4. $P_{MS,0.5}(t;a,b)$ for different $a,b\geq 0$

The parametric functions $p_{R2}(.;\beta_0,\beta_1)$ and $p_{MS}(.;\beta_0,\beta_1)$ (or others with similar properties) can be used to construct a fuzzy number u in the LR form, by representing its membership function as

$$\mu_u(x)=\begin{cases} p(\frac{x-a}{b-a};\beta_{0,L},\beta_{1,L}) & \text{if } x\in[a,b]\\ 1 & \text{if } x\in[b,c]\\ p(\frac{d-x}{d-c};\beta_{0,R},\beta_{1,R}) & \text{if } x\in[c,d]\\ 0 & \text{otherwise}\end{cases} \tag{9}$$

or in the LU form, by representing its alpha-cuts as the compact intervals $[u_\alpha^-,u_\alpha^+]$, for $\alpha\in[0,1]$, with

$$\begin{aligned} u_\alpha^- &= a+(b-a)p(\alpha;\beta_0^-,\beta_1^-),\\ u_\alpha^+ &= d+(c-d)p(\alpha;\beta_0^+,\beta_1^+).\end{aligned} \tag{10}$$

Here, obviously, $a\leq b\leq c\leq d$ and $[a,d]$ is the support, while $[b,c]$ is the core of given fuzzy number u. If needed, we will denote a,b,c and d also by u_0^-,u_1^-,u_1^+ and u_0^+, respectively.

As it is possible to go from the LR to the LU representations by inverting the model functions, further on we will use the LU form only; however analogous formulations are possible for the LR form as well. The obvious relations between the derivatives of $\mu_u(x)$ and the derivatives of u_α^-, u_α^+ gives (provided they are not null)

$$\beta_{0,L}=\frac{1}{\beta_0^-},\beta_{1,L}=\frac{1}{\beta_1^-},\beta_{0,R}=\frac{1}{\beta_0^+},\beta_{1,R}=\frac{1}{\beta_1^+}. \tag{11}$$

Let us consider the expected interval of such fuzzy number u. It is given by $[E^-u,E^+u]$, where

$$\begin{aligned} E^-u &= u_0^-+(u_1^--u_0^-)\int_0^1 p(\alpha;\beta_0^-,\beta_1^-)d\alpha,\\ E^+u &= u_0^++(u_1^+-u_0^+)\int_0^1 p(\alpha;\beta_0^+,\beta_1^+)d\alpha\end{aligned} \tag{12}$$

and hence we have four parameters $\beta_0^-,\beta_1^-,\beta_0^+,\beta_1^+\geq 0$ free for our further purposes (or six parameters if $P_{R2,\lambda^-}(t;a^-,b^-)$ and $P_{R2,\lambda^+}(t;a^+,b^+)$ or $P_{MS,\lambda^-}(t;a^-,b^-)$ and $P_{MS,\lambda^+}(t;a^+,b^+)$ are used).

For example, if $\beta_0^-=\beta_1^-=\beta^-$ and $\beta_0^+=\beta_1^+=\beta^+$, we get $\int_0^1 p_{R2}(t;\beta^\pm,\beta^\pm)dt=\frac{1}{2}$ and the expected interval is

$$[E^-u,E^+u]=[\frac{u_0^-+u_1^-}{2},\frac{u_0^++u_1^+}{2}],$$

ISBN: 978-989-95079-6-8

for all nonnegative β^-, β^+. If a fuzzy number A with alpha cuts $[A_\alpha^-, A_\alpha^+]$ is to be approximated, we can constrain $u_0^- \leq u_1^- \leq u_1^+ \leq u_1^-$ such that

$$\begin{aligned} u_0^- + u_1^- &= 2\int_0^1 A_\alpha^- d\alpha \\ u_0^+ + u_1^+ &= 2\int_0^1 A_\alpha^+ d\alpha, \end{aligned}$$

while the two parameters $\beta^-, \beta^+ \geq 0$ still can be used for additional requirements. Thus, using our approach we obviously loose the benefits of linear shapes but the form of $p(t;\beta,\beta)$ is simple and it is easy to invert analytically by solving a quadratic equation instead of linear. It also reduces to linear if $\beta = 1$.

3 Nonlinear shape approximations

In this section, we consider the approximation problem for a fuzzy number A having membership function $\mu_A(x)$ and compact α-cuts $A_\alpha = [A_\alpha^-, A_\alpha^+]$, $\alpha \in [0,1]$.

If we like to preserve or approximate the middle set (i.e. the alpha cut for $\alpha = 0.5$) we have

$$p_{R2}(\frac{1}{2};\beta_0,\beta_1) = \frac{1+\beta_0}{\beta_0+\beta_1+2}$$

and we obtain the conditions

$$\begin{aligned} u_0^- + (u_1^- - u_0^-)\frac{1+\beta_0^-}{\beta_0^-+\beta_1^-+2} &= A_{0.5}^- \\ u_0^+ + (u_1^+ - u_0^+)\frac{1+\beta_0^+}{\beta_0^++\beta_1^++2} &= A_{0.5}^+ \end{aligned}$$

giving (if the $u_{0/1}^\pm$ are known) the following equations/constraints for the parameter β:

$$\begin{aligned} (u_1^- - A_{0.5}^-)\beta_0^- + (u_0^- - A_{0.5}^-)\beta_1^- &= 2A_{0.5}^- - u_0^- - u_1^- \\ (u_1^+ - A_{0.5}^+)\beta_0^+ + (u_0^+ - A_{0.5}^+)\beta_1^+ &= 2A_{0.5}^+ - u_0^+ - u_1^+. \end{aligned}$$

In the special case of $\beta_0 + \beta_1 = 2$ we obtain a parabolic shape function

$$p_{R2}(t;\beta_0,\beta_1) = Q(t;\beta) = t^2 + \beta t(1-t)$$

where $\beta_0 = \beta, \beta_1 = 2-\beta$ with $\beta \in [0,2]$ and

$$\int_0^1 Q(\alpha;\beta)d\alpha = \frac{1}{3} + \frac{1}{6}\beta$$

is in the range $[\frac{1}{3}, \frac{2}{3}]$. Then the equations for expected interval preservation are as follows

$$\begin{aligned} u_0^- + (u_1^- - u_0^-)(\frac{1}{3} + \frac{1}{6}\beta^-) &= \int_0^1 A_\alpha^- d\alpha, \\ u_0^+ + (u_1^+ - u_0^+)(\frac{1}{3} + \frac{1}{6}\beta^+) &= \int_0^1 A_\alpha^+ d\alpha. \end{aligned}$$

We can also consider the value and the ambiguity of fuzzy numbers or other quantities related to ranking or defuzzification. If $s : [0,1] \longrightarrow [0,1]$ is a regular reducing function (i.e. $\int_0^1 s(\alpha)d\alpha = \frac{1}{2}$) the value of the fuzzy number u relative to s is defined by (see [5])

$$Val_s(u) = \int_0^1 s(\alpha)\left(u_\alpha^- + u_\alpha^+\right)d\alpha, \tag{13}$$

while the ambiguity of u relative to s is

$$Amb_s(u) = \int_0^1 s(\alpha)\left(u_\alpha^+ - u_\alpha^-\right)d\alpha. \tag{14}$$

For our parametric forms, we obtain

$$\begin{aligned} Val_s(u) &= \frac{u_0^- + u_0^+}{2} + \\ &+(u_1^- - u_0^-)\int_0^1 s(\alpha)p(\alpha;\beta_0^-,\beta_1^-)d\alpha \\ &+(u_1^+ - u_0^+)\int_0^1 s(\alpha)p(\alpha;\beta_0^+,\beta_1^+)d\alpha \end{aligned}$$

and

$$\begin{aligned} Amb_s(u) &= \frac{u_0^+ - u_0^-}{2} + \\ &+(u_0^- - u_1^-)\int_0^1 s(\alpha)p(\alpha;\beta_0^-,\beta_1^-)d\alpha \\ &+(u_1^+ - u_0^+)\int_0^1 s(\alpha)p(\alpha;\beta_0^+,\beta_1^+)d\alpha. \end{aligned}$$

The integrals above can be computed by numerical approximations (e.g. trapezoidal or Simpson formulas), but for specific cases (e.g. $s(\alpha) = 1$ or $s(\alpha) = \alpha$) we can proceed analytically.

With reference to equations (10), define the following integrals, depending on the slope parameters:

$$I(\beta_0,\beta_1) = \int_0^1 p(\alpha;\beta_0,\beta_1)d\alpha \tag{15}$$

$$J(\beta_0,\beta_1) = \int_0^1 \alpha p(\alpha;\beta_0,\beta_1)d\alpha. \tag{16}$$

For the family of parametric functions $p_{MS}(\alpha;\beta_0,\beta_1)$ we have

$$\begin{aligned} I_{MS}(\beta_0,\beta_1) &= \frac{2+(3+2\beta_0)(\beta_0+\beta_1)}{2(1+\beta_0+\beta_1)(2+\beta_0+\beta_1)} \\ J_{MS}(\beta_0,\beta_1) &= \frac{42+(84+40\beta_0)(\beta_0+\beta_1)}{20(1+\beta_0+\beta_1)(2+\beta_0+\beta_1)(3+\beta_0+\beta_1)} \end{aligned}$$

They are nonlinear functions of β_0, β_1 but become linear if, for example, we know or we assume a given value $m =$

ISBN: 978-989-95079-6-8

$\beta_0 + \beta_1 \geq 0$. Note that fixing the value of m (we suggest m to be integer) is equivalent to fixing the "degree" of p_{MS} as a function of α. If m is integer then p_{MS} in a polynomial function of degree one (if $\beta_0 = \beta_1 = 1$), two (if $\beta_0+\beta_1 = 2$), three (if $\beta_0 = \beta_1 = 0$) or $m+1$ (if $m = \beta_0 + \beta_1 \geq 3$). In these cases we obtain

$$\begin{aligned} I_{MS}(\beta_0, m-\beta_0) &= \frac{2+(3+2\beta_0)m}{2(1+m)(2+m)} \\ J_{MS}(\beta_0, m-\beta_0) &= \frac{42+(84+40\beta_0)m}{20(1+m)(2+m)(3+m)} \end{aligned}$$

with constraint $0 \leq \beta_0 \leq m$.

For the commonly used regular reducing functions $s(\alpha) = \frac{1}{2}$ and $s(\alpha) = \alpha$, the expressions $I_{MS}(\beta_0, \beta_1)$ and $J_{MS}(\beta_0, \beta_1)$ allow to compute $Val_s(u)$ and $Amb_s(u)$ in closed form as functions of the parameters β_0^-, β_1^- and β_0^+, β_1^+ or, for fixed m^- and m^+, of β_0^-, β_0^+.

The ordering of fuzzy numbers can be approached in many ways (see [8] or [9] for recent literature and results). In most cases, the fuzzy numbers u are transformed into real numbers by a real-valued function $D(u)$. By the use of the LU-fuzzy representation, the ranking functions can be computed, either numerically or, if possible, analytically, in terms of the parameters $u_i^{\pm}$ and $\beta_i^{\pm}$ that define u.

The first example is the so called *level set average* (see [6]), also known as the expected value of a fuzzy number, and defined by

$$u_{DP}^* = \frac{1}{2}\int_0^1 \left(u_\alpha^- + u_\alpha^+\right) d\alpha. \tag{17}$$

Using (15) we obtain

$$\begin{aligned} u_{DP}^* &= \frac{1}{2}[u_0^- + (u_1^- - u_0^-)I_{MS}(\beta_0^-, \beta_1^-) + \\ &\quad + u_0^+ + (u_1^+ - u_0^+)I_{MS}(\beta_0^+, \beta_1^+)]. \end{aligned}$$

It is worth noting that the above integral is closely related to the *nearest interval approximation* of a fuzzy number (see [10]), expressed equivalently either in terms of the lower-upper functions or in terms of the membership function:

$$\begin{aligned} C(u) &= \left[\int_0^1 u_\alpha^- d\alpha\,,\ \int_0^1 u_\alpha^+ d\alpha\right] \qquad (18) \\ &= \left[u_1^- - \int_{u_0^-}^{u_1^-} \mu(x)dx\,,\ u_1^+ + \int_{u_1^+}^{u_0^+} \mu(x)dx\right]. \end{aligned}$$

Using our parametric family p_{MS} and with reference to (10) we get

$$\begin{aligned} C(u) &= \frac{1}{2}[u_0^- + (u_1^- - u_0^-)I_{MS}(\beta_0^-, \beta_1^-), \\ &\quad u_0^+ + (u_1^+ - u_0^+)I_{MS}(\beta_0^+, \beta_1^+)]. \end{aligned}$$

Other examples are the *interval valued possibilistic mean* $M(u) = [M^-(u), M^+(u)]$ and the *level-weighted average* u_{GW}^* given by

$$M^-(u) = 2\int_0^1 \alpha u_\alpha^- d\alpha, \quad M^+(u) = 2\int_0^1 \alpha u_\alpha^+ d\alpha$$

and

$$u_{GW}^* = \int_0^1 \alpha\left(u_\alpha^- + u_\alpha^+\right) d\alpha = \frac{M^-(u) + M^+(u)}{2}. \tag{19}$$

The calculations for p_{MS} family lead to

$$\begin{aligned} M^-(u) &= u_0^- + 2(u_1^- - u_0^-)J_{MS}(\beta_0^-, \beta_1^-) \\ M^+(u) &= u_0^+ - 2(u_0^+ - u_1^+)J_{MS}(\beta_0^+, \beta_1^+). \end{aligned}$$

4 Some particular approximations

We show in this section how the illustrated setting is useful to solve two specific types of approximations of a given fuzzy number A for which we know the support $[a, d]$ and the core $[b, c]$ (section 4.1) or we fix the shapes, by fixing the nonnegative parameters β_0^-, β_1^- and β_0^+, β_1^+ (section 4.2).

4.1 Approximations with known support and core

In the first case we are looking for the "best curve" under the condition that the support and the core of the approximating fuzzy number u are exactly the same of A. In other words, $u_0^- = a$, $u_0^+ = d$, $u_1^- = b$ and $u_1^+ = c$.

Now we have to estimate the shape-parameters β_0^-, β_1^- and β_0^+, β_1^+ such that a distance measure $Dist(A, u)$ is minimized, subject to the nonnegativity constraints on β_0^-, β_1^- and β_0^+, β_1^+. Here u results to be a function of the $\underline{\beta} = (\beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+)$. So our problem is

$$\begin{aligned} &Min \quad Dist(A, u(\underline{\beta})) \\ &s.t. \\ &\underline{\beta} \geq 0. \end{aligned}$$

The distance $Dist(A, u(\underline{\beta}))$ can be calculated if we have other information on A. For example, if the membership function of A is known at other points, we can approximate the distance by the least squares functional.

Suppose that $\mu_A(x_j) = \mu_j$ for given $x_j \in (a, b)$, $j = 1, 2, ..., j_L$ and for given $x_j \in (c, d)$, where $j = j_L + 1, j_L + 2, ..., j_L + j_R$ (i.e. j_L values correspond to the left arm of the fuzzy number while j_R values to the right arm). So we should minimize

$$\begin{aligned} &Dist(A, u(\underline{\beta})) = \qquad (20) \\ &= \sum_{j=1}^{j_L} \left\{x_j - \left[a + (b-a)p(\mu_j; \beta_0^-, \beta_1^-)\right]\right\}^2 \\ &+ \sum_{j=j_L+1}^{j_L+j_R} \left\{x_j - \left[d + (c-d)p(\mu_j; \beta_0^+, \beta_1^+)\right]\right\}^2. \\ &s.t. \\ &\beta_0^-, \beta_1^-, \beta_0^+, \beta_1^+ \geq 0. \end{aligned}$$

Therefore, we have obtained a nonlinear least squares problem with four variables and nonnegativity constraints which can be solved by any numerical procedure.

The minimization (20) can be split into two independent problems, one to determine β_0^-, β_1^- and the other for β_0^+, β_1^+.

A special case applies if we have available a single additional observation for the left side ($j_L = 1$), say $\mu_A(x^-) =$

ISBN: 978-989-95079-6-8

μ^- with $x^- \in (a,b)$, $\mu^- \in (0,1)$ and for the right ($j_R = 1$), say $\mu_A(x^+) = \mu^+$ with $x^+ \in (c,d)$, $\mu^+ \in (0,1)$.

An interpolating solution satisfies equations

$$\begin{aligned} a + (b-a)p(\mu^-; \beta_0^-, \beta_1^-) &= x^- \\ d + (c-d)p(\mu^+; \beta_0^+, \beta_1^+) &= x^+. \end{aligned}$$

If we use $p_{R2}(t; \beta_0, \beta_1)$ for $p(t; \beta_0, \beta_1)$, we obtain two equations and four nonnegative variables

$$(b-x^-)\beta_0^- + (a-x^-)\beta_1^- = \gamma^- \quad (21)$$
$$(c-x^+)\beta_0^+ + (d-x^+)\beta_1^+ = \gamma^+, \quad (22)$$

where

$$\begin{aligned} \gamma^- &= \frac{(a-x^-)(-2\mu^{-^2} + 2\mu^- - 1) - (b-a)\mu^{-^2}}{\mu^-(1-\mu^-)} \\ \gamma^+ &= \frac{(d-x^+)(-2\mu^{+^2} + 2\mu^+ - 1) - (c-d)\mu^{+^2}}{\mu^+(1-\mu^+)}. \end{aligned}$$

Equations (21)-(22) represent a line in the plane (β_0, β_1) having an infinite number of nonnegative solutions (note that $b - x^- > 0$, $a - x^- < 0$ and $c - x^+ < 0$, $d - x^+ > 0$). So we have many possible choices. We suggest tree interesting solutions:

1. The unique solution having least norm of (β_0, β_1) (i.e. with minimal $\beta_0^2 + \beta_1^2$), is obtained at the intersections of lines (21)-(22) with the axes. It has the following closed form:

$$\begin{aligned} &\text{if } \gamma^- = 0 \text{ then } \beta_0^- = 0, \beta_1^- = 0; \quad (23) \\ &\text{if } \gamma^- > 0 \text{ then } \beta_0^- = \frac{\gamma^-}{b-x^-}, \beta_1^- = 0; \\ &\text{if } \gamma^- < 0 \text{ then } \beta_0^- = 0, \beta_1^- = \frac{\gamma^-}{a-x^-}. \end{aligned}$$

and

$$\begin{aligned} &\text{if } \gamma^+ = 0 \text{ then } \beta_0^+ = 0, \beta_1^+ = 0; \quad (24) \\ &\text{if } \gamma^+ > 0 \text{ then } \beta_0^+ = 0, \beta_1^+ = \frac{\gamma^+}{d-x^+}; \\ &\text{if } \gamma^+ < 0 \text{ then } \beta_0^+ = \frac{\gamma^+}{c-x^+}, \beta_1^+ = 0. \end{aligned}$$

Solution (23)-(24) is illustrated in Fig. 5 with the following data: $a = 0$, $b = 2.5$, $c = 3$, $d = 5$ and the two membership functions for $\mu_A(0.5) = 0.5$, $\mu_A(3.5) = 0.9$ and $\mu_A(0.6) = 0.5$, $\mu_A(3.5) = 0.8$.

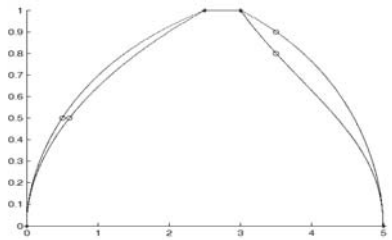


Figure 5: Solutions (23)-(24) for two examples data.

2. Recall that functions $p(t; \beta_0, \beta_1)$ represent a linear shape if and only if $\beta_0 = \beta_1 = 1$; so it is reasonable to "measure" its nonlinearity by the distance from (β_0, β_1) to $(1,1)$ and to minimize the distance $(\beta_0 - 1)^2 + (\beta_1 - 1)^2$. In this way, linear data will reproduce linear shapes. The unique solution for this criterion is obtained by the following procedure:

$$\begin{aligned} &\text{if } \widehat{\beta}_0^- \geq 0 \text{ and } \widehat{\beta}_1^- \geq 0 \text{ then } \beta_0^- = \widehat{\beta}_0^-,\ \beta_1^- = \widehat{\beta}_1^-; \quad (25) \\ &\text{if } \widehat{\beta}_0^- < 0 \text{ then } \beta_0^- = 0,\ \beta_1^- = \frac{\gamma^-}{a-x^-}; \\ &\text{if } \widehat{\beta}_1^- < 0 \text{ then } \beta_0^- = \frac{\gamma^-}{b-x^-}, \beta_1^- = 0. \end{aligned}$$

where

$$\begin{aligned} \widehat{\beta}_0^- &= \frac{\gamma^-(b-x^-) + (a-x^-)^2 - (b-x^-)(a-x^-)}{(b-x^-)^2 + (a-x^-)^2} \\ \widehat{\beta}_1^- &= 1 + \frac{(a-x^-)(\widehat{\beta}_0^- - 1)}{(b-x^-)}; \end{aligned}$$

for β_0^+, β_1^+ the procedure is analogous

$$\begin{aligned} &\text{if } \widehat{\beta}_0^+ \geq 0 \text{ and } \widehat{\beta}_1^+ \geq 0 \text{ then } \beta_0^+ = \widehat{\beta}_0^+,\ \beta_1^+ = \widehat{\beta}_1^+; \quad (26) \\ &\text{if } \widehat{\beta}_0^+ < 0 \text{ then } \beta_0^+ = 0,\ \beta_1^+ = \frac{\gamma^+}{d-x^+}; \\ &\text{if } \widehat{\beta}_1^+ < 0 \text{ then } \beta_0^+ = \frac{\gamma^+}{c-x^+}, \beta_1^+ = 0. \end{aligned}$$

where

$$\begin{aligned} \widehat{\beta}_0^+ &= \frac{\gamma^+(c-x^+) + (d-x^+)^2 - (d-x^+)(c-x^+)}{(c-x^+)^2 + (d-x^+)^2} \\ \widehat{\beta}_1^+ &= 1 + \frac{(d-x^+)(\widehat{\beta}_0^+ - 1)}{(c-x^+)}. \end{aligned}$$

Solution (25)-(26) is illustrated in Fig. 6 with the same data as before: $a = 0$, $b = 2.5$, $c = 3$, $d = 5$, $\mu_A(0.5) = 0.5$, $\mu_A(3.5) = 0.9$ and $\mu_A(0.6) = 0.5$, $\mu_A(3.5) = 0.8$.

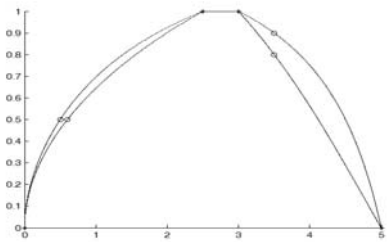


Figure 6: Solutions (25)-(26) for two examples data.

3. In cases 1. and 2. we made no assumptions on the parameters β_0^-, β_1^-, β_0^+, β_1^+, but we may know or desire to have them partially fixed. Lets consider the fitting equations in terms of the LR representation (9); if, for example, we require a differentiable membership function, this is equivalent to have $\beta_{1,L} = 0$ and $\beta_{1,R} = 0$ and we obtain the following solution

$$\beta_{0,L} = \begin{cases} \frac{\omega^-}{t^-(1-t^-)(1-\mu^-)} & \text{if } \omega^- > 0 \\ 0 & \text{if } \omega^- \leq 0, \end{cases} \quad (27)$$

where $t^- = \frac{x^- - a}{b-a}$ and $\omega^- = (1-2t^-)\mu^- - (1-2\mu^-)(t^-)^2$,

$$\beta_{0,R} = \begin{cases} \frac{\omega^+}{t^+(1-t^+)(1-\mu^+)} & \text{if } \omega^+ > 0 \\ 0 & \text{if } \omega^+ \leq 0, \end{cases} \quad (28)$$

ISBN: 978-989-95079-6-8

where $t^+ = \frac{d-x^+}{d-c}$ and $\omega^+ = (1-2t^+)\mu^+ - (1-2\mu^+)(t^+)^2$. Solution (27)-(28) is illustrated in Fig. 7 with the same data as for Fig. 5, 6. Clearly, in this case, we obtain membership functions such that $\mu(\frac{a+b}{2}) \geq \frac{1}{2}$ and $\mu(\frac{c+d}{2}) \geq \frac{1}{2}$ and interpolation is possible only for specific data.

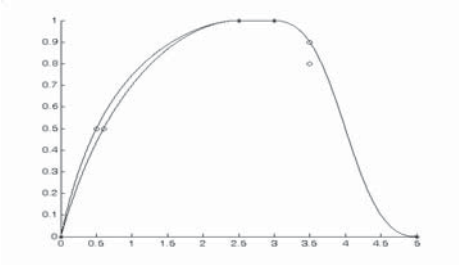


Figure 7: Solutions (27)-(28) for two examples data.

4.2 Approximations with fixed shapes

In the second case we assume that the general shape of u is fixed (by fixing the parameters $\underline{\beta}$) and we like to find the best approximation of the core and support, i.e. to find u_0^-, u_1^-, u_1^+, u_0^+ such that $d(A, u(\underline{\beta})) \rightarrow \min$.

In this case, the resulting optimization problem is easier than for Problem 1. Now, $u = u(\underline{\eta})$ is a function of $\underline{\eta} = (u_0^-, u_1^-, u_0^+, u_1^+)$ and we require $u_0^- \leq u_1^- \leq u_1^+ \leq u_0^+$.

Suppose also here that we know the membership of A at $j_L + j_R$ points, $\mu_A(x_j) = \mu_j$ for given $x_j \in [a, b]$, $j = 1, 2, ..., j_L$ and for given $x_j \in [c, d]$, $j = j_L + 1, j_L + 2, ..., j_L + j_R$.

Define $p_j = p(\mu_j; \beta_0^-, \beta_1^-)$ for $j = 1, 2, ..., j_L$ and $p_j = p(\mu_j; \beta_0^+, \beta_1^+)$ for $j = j_L + 1, j_L + 2, ..., j_L + j_R$.

Now we have to minimize

$$
\begin{aligned}
& Dist(A, u(\underline{\eta})) = \qquad (29)\\
= & \sum_{j=1}^{j_L} \left\{x_j - \left[u_0^- + (u_1^- - u_0^-)p_j\right]\right\}^2 \\
& + \sum_{j=j_L+1}^{j_L+j_R} \left\{x_j - \left[u_0^+ + (u_1^+ - u_0^+)p_j\right]\right\}^2 \\
& s.t. \\
& u_0^- - u_1^- \leq 0 \\
& u_1^- - u_1^+ \leq 0 \\
& u_1^+ - u_0^+ \leq 0.
\end{aligned}
$$

We obtain a linear least squares problem with four variables and three linear constraints and can be solved using standard well known procedures (see [14]).

5 Conclusions

In the present contribution we have suggested a general approach to fuzzy number approximation based on two families of parameterized functions. Our method enables a very rich set of approximating curves which are both easy to handle and useful for practical purposes.

References

[1] S. Abbasbandy, B. Asady, *The nearest approximation of a fuzzy quantity in parametric form*, Applied Mathematics and Computation, 172, 2006, 624-632.

[2] S. Abbasbandy, M. Amirfakhrian, *The nearest trapezoidal form of a generalized LR fuzzy number*, International Journal of Approximate Reasoning, 43, 2006, 166-178.

[3] Ban A., *Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the expected interval*, Fuzzy Sets and Systems 159, 2008, 1327-1344.

[4] Y. Chalco-Cano, H. Roman-Flores, F. Gomide, *A new type of approximation for fuzzy intervals*, Fuzzy Sets and Systems, 159, 2008, 1376-1383.

[5] M. Delgado, M.A. Vila, W. Voxman, *A fuzziness measure for fuzzy number: applications*, Fuzzy Sets and Systems 94 (1998), 205–216.

[6] D. Dubois, H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.

[7] D. Dubois, H. Prade (Ed), *Fundamentals of Fuzzy Sets*, Kluwer, Boston, The Handbooks of Fuzzy Sets Series, 2000.

[8] S. Heilpern, *Representation and application of fuzzy numbers*, *Fuzzy Sets and Systems*, 91, 1997, 259-278.

[9] P. Grzegorzewski, *Metrics and orders in space of fuzzy numbers*, Fuzzy Sets and Systems, 97, 1998, 83-94.

[10] P. Grzegorzewski, *Nearest interval approximation of a fuzzy number*, Fuzzy Sets and Systems, 130, 2002, 321-330.

[11] P. Grzegorzewski, *Trapezoidal approximations of fuzzy numbers preserving the expected value: algorithms and properties*, Fuzzy Sets and Systems, 159, 2008, 1354-1364.

[12] P. Grzegorzewski, E. Mrówka, *Trapezoidal approximations of fuzzy numbers*, Fuzzy Sets and Systems, 153, 2005, 115-135.

[13] P. Grzegorzewski, E. Mrówka, *Trapezoidal approximations of fuzzy numbers - revisited*, Fuzzy Sets and Systems 158 (2007), 757–768.

[14] C. L. Lawson, C. R. J. Hanson, *Solving least squares problems*, Prentice-Hall, 1974.

[15] E. N. Nasibov, S. Peker, *On the nearest parametric approximation of a fuzzy number*, Fuzzy Sets and Systems, 159, 2008, 1365-1375.

[16] L. Stefanini, L. Sorini, M. L. Guerra, *Parametric representation of fuzzy numbers and application to fuzzy calculus*, Fuzzy Sets and Systems, 157, 2006, 2423 – 2455.

[17] L. Stefanini, L. Sorini, M.L. Guerra, *Fuzzy Numbers and Fuzzy Arithmetic*, in W. Pedrycz et al. (Eds) "*Handbook of Granular Computing*" , Chapter 12, J. Wiley & Sons, 2008.

[18] C.T. Yeh, *Trapezoidal and triangular approximations preserving the expected interval*, Fuzzy Sets and Systems 159, 2008, 1345-1353.

ISBN: 978-989-95079-6-8

Characterizing Fuzzy Modal Semantics by fuzzy multimodal systems with crisp accessibility relations

Félix Bou[1], Francesc Esteva[1], Lluís Godo[1], Ricardo Rodríguez[2]

1. Institut d'Investigació en Intel·ligència Artificial, IIIA - CSIC,
Campus UAB, 08193 Bellaterra, Spain
2. Dpto. de Computación, Universidad Buenos Aires,
1428 Buenos Aires, Argentina
Email: {fbou,esteva,godo}@iiia.csic.es, ricardo@dc.uba.ar

***Abstract*—** *In [1] the authors considered finitely-valued modal logics with Kripke style semantics where both propositions and the accessibility relation are valued over a finite residuated lattice. Unfortunately, the necessity operator does not satisfy in general the normality axiom (K). In this paper we focus on the case of finite chains, and we consider a different approach based on introducing a multimodal logic where the previous necessity operator is replaced with a family, parametrized by truth values different from zero, of necessity operators each one semantically defined using the crisp accessibility relation given by the corresponding cut of the finitely-valued original accessibility relation. This multimodal logic is somehow more appealing than the original modal one because axiom (K) holds for each necessity operator. In this paper we axiomatize this multimodal logic and we prove that, in the case the starting residuated lattice is a finite BL chain, the modal and the multimodal languages have the same expressive power iff this algebra is an MV chain.*

***Keywords*—** many-valued modal logic, fuzzy modal logic, Łukasiewicz modal logic, fuzzy logic.

1 Introduction

Fuzzy modal logics is a subfield of mathematical fuzzy logic with growing interest. The interested reader is referred to [1] and the references therein for recent developments. Indeed, in [1] the same authors proposed a mathematical definition of what a many-valued (uni)modal logic is. According to this definition *many-valued modal logics* are sets[1] of modal formulas, denoted by $\Lambda(\mathsf{K}, \mathbf{A})$, arising naturally from the semantics given by a complete residuated lattice $\mathbf{A}$, whose support A corresponds to the intended set of truth values, and a class K of Kripke frames valuated over the set A.

In [1], due to simplicity reasons, the authors only considered *modal formulas* obtained enriching the propositional language of residuated lattices with a necessity operator $\Box$ (i.e., without a primitive possibility operator $\Diamond$). These modal formulas may include or not canonical constants to talk about the truth values. By adding *canonical constants* we mean to add one constant $\overline{a}$ for every element a in the residuated lattice $\mathbf{A}$ in such a way that each one of these constants is semantically interpreted by its canonical interpretation (i.e., the very element a). In the case that we allow canonical constants in the modal language we use the notation $\Lambda(\mathsf{K}, \mathbf{A^c})$ to stress their presence. In the present paper we will always assume that the language has canonical constants.

An assumption that we consider throughout the present paper is that $\mathbf{A}$ is a chain (to keep us inside the fuzzy realm) and finite, i.e., $\mathbf{A}$ is a finite MTL chain. We remind the reader that finiteness is crucial in the results obtained in [1], but linearity is not required there. In the next paragraphs we point out some of the results in [1] for the particular case of finite chains.

One of the main results in [1] is the presentation of a complete calculus for the many-valued modal logic $\Lambda(\mathsf{K}, \mathbf{A^c})$ when K is the class Fr of all frames valued over $\mathbf{A}$, assuming a complete calculus is already known for the non-modal logic without canonical constants (denoted by $\mathbf{\Lambda(A)}$). This axiomatization is shown in Table 1. It is worth pointing out that in the case that $\mathbf{A}$ is a finite BL chain, it is known [1, Propositions 2.5 and 2.7] that there are axiomatizations of $\mathbf{\Lambda(A)}$ based on only one rule, the Modus Ponens rule; but there are examples of finite MTL chains where it is strictly necessary to add more rules besides Modus Ponens.

Among the difficulties to find such an axiomatization is that while the meet distributivity axiom

$$(\Box\varphi \wedge \Box\psi) \leftrightarrow \Box(\varphi \wedge \psi) \qquad (\mathsf{MD})$$

is valid (as in the classical modal case), in general this is not the case in the many-valued modal setting for the normality axiom

$$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi) \qquad (\mathsf{K})$$

The normality axiom is indeed valid in $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ iff the residuated lattice is a Heyting algebra (i.e., the interpretation of the strong conjunction coincides with the meet).

The situation is much more difficult in case that there are no canonical constants in the language, and as far as the authors are aware the only known axiomatizations for many-valued modal logics of the form $\Lambda(\mathsf{Fr}, \mathbf{A})$, where $\mathbf{A}$ is not a Heyting algebra, are the ones given in [1] for the case that $\mathbf{A}$ is a finite MV chain. The main drawback of the axiomatization there given is that it is rather artificial, and hence it is not absolutely clear what are the basic principles of these many-valued modal logics.

On the other hand, if one considers K as the class CFr of *crisp Kripke frames* (i.e., those Kripke frames valuating the accessibility relation over the set[2] $\{0, 1\}$) then the problem

[1] Besides these sets in the same paper it is also considered the consequence relations $\mathbf{\Lambda}(l, \mathsf{K}, \mathbf{A})$ and $\mathbf{\Lambda}(g, \mathsf{K}, \mathbf{A})$ corresponding to the local and global many-valued modal consequence relations.

[2] It is worth pointing out that since $\mathbf{A}$ is a chain it holds that the set $\{0, 1\}$ coincides with the set $\{a \in A : 1 = a \vee \neg a\}$ of Boolean elements of $\mathbf{A}$.

ISBN: 978-989-95079-6-8

Table 1: Axiomatization of the set $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ when $\mathbf{A}$ is a finite chain

- the set of axioms is the smallest set closed under substitutions containing
  - the axiomatic basis for $\mathbf{\Lambda(A)}$,
  - the witnessing axiom $\bigvee_{a\in A}(\varphi \leftrightarrow \overline{a})$
  - the bookkeeping axioms $(\overline{a_1} * \overline{a_2}) \leftrightarrow \overline{a_1 * a_2}$ (for every $a_1, a_2 \in A$ and every $* \in \{\wedge, \vee, \odot, \rightarrow\}$),
  - $\Box 1$, $(\Box\varphi \wedge \Box\psi) \rightarrow \Box(\varphi \wedge \psi)$ and $\Box(\overline{a} \rightarrow \varphi) \leftrightarrow (\overline{a} \rightarrow \Box\varphi)$ (for every $a \in A$),
- the rules of the basis for $\mathbf{\Lambda(A)}$, the rule $\overline{k} \vee \varphi \vdash \varphi$ (where k is the coatom of $\mathbf{A}$) and the Monotonicity rule $\varphi \rightarrow \psi \vdash \Box\varphi \rightarrow \Box\psi$.

resembles much more the one in the classical modal setting since now the normality axiom is valid. It is known [1] that $\Lambda(\mathsf{CFr}, \mathbf{A^c})$ is the set of modal formulas derivable in the calculus given in Table 2. This table is, roughly speaking, saying that we only need to add to $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ the normality axiom plus the axiom $\Box(\overline{k}\vee\varphi) \rightarrow (\overline{k}\vee\Box\varphi)$ (where k is the coatom of $\mathbf{A}$) in order to capture the logic of crisp frames. We point out it is known that in general it is not enough to add the normality axiom[3] (cf. Remark 3.2); and we notice that the last axiom is a particular case of the formulas $\Box(\overline{a} \vee \varphi) \rightarrow (\overline{a} \vee \Box\varphi)$ with $a \in A$, all these formulas being valid in $\Lambda(\mathsf{CFr}, \mathbf{A^c})$. Notice that in Table 2 we have replaced the Monotonicity rule by the Necessity rule, but this is not important due to the presence of the normality axiom.

The aim of this paper is to apply to the realm of many-valued modal logics the well known [2] correspondence between

- a fuzzy binary relation over A (i.e., a function $R : W \times W \longrightarrow A$), and
- an $A \setminus \{0\}$-indexed and decreasing family of crisp binary relations (i.e., a family $\{R_a : a \in A, a \neq 0\}$ such that if $a \leqslant b$ then $R_b \subseteq R_a \subseteq W \times W$).

This correspondence is given by the following identities

$$R_a = \{(w_1, w_2) \in W \times W : R(w_1, w_2) \geqslant a\} \qquad (1)$$

$$R(w_1, w_2) = \sup\{a \in A : (w_1, w_2) \in R_a\} \qquad (2)$$

The fact that $\mathbf{A}$ is a finite chain (in particular a complete lattice) is crucial in order to see that the previous identities induce a bijective correspondence.

In other words, the previous correspondence transforms a Kripke frame into a family of crisp Kripke frames. This transformation suggests to use a multimodal language (with one modality $\Box_a$ for every element $a \in A \setminus \{0\}$) to describe properties of Kripke frames. The advantage of this method is that, since the modalities $\Box_a$'s are induced by crisp Kripke frames, the normality axioms

$$\Box_a(\varphi \rightarrow \psi) \rightarrow (\Box_a\varphi \rightarrow \Box_a\psi)$$

are valid.

[3] In [1] it is proved that if we only add the normality axiom to $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ then we get an axiomatization for the modal logic given by the class of idempotent frames (i.e., those Kripke frames valuating the accessibility relation over the set $\{a \in A : a = a \odot a\}$).

The aim of the present paper is to pursue this research line. To this purpose in Section 2 these multimodal logics are introduced, in Section 3 a complete calculus is given for them, and in Section 4 we compare the expressive power of the crisp multimodal approach with the unimodal one from [1]. Finally, in Section 5 the authors analyze what it is known about these results when there are no canonical constants in the language.

2 Defining the multimodal systems

Throughout this paper we assume that $\mathbf{A}$ is a fixed *finite residuated lattice* whose underlying order is a *chain*. We remind the reader that a residuated lattice is an algebra $\mathbf{A}$ such that:

- $\langle A, \wedge, \vee, 0, 1\rangle$ is a bounded lattice with a linear associated order $\leqslant$,
- $\langle A, \odot, 1\rangle$ is a commutative monoid with the unit 1,
- $x \odot z \leqslant y \Leftrightarrow z \leqslant x \rightarrow y$ (the law of *residuation*).

In the literature these algebras are well known under different names: residuated lattices, integral, commutative residuated monoids, $\mathsf{FL_{ew}}$-algebras, etc. [3, 4, 5]. We stress that finite MTL-algebras [6], finite BL-algebras [7] and finite MV-algebras [8] are particular cases satisfying our assumption.

We stress that the algebraic language of $\mathbf{A}$ is given by $\langle\wedge, \vee, \odot, \rightarrow, 1, 0\rangle$ (with arities $\langle 2, 2, 2, 2, 0, 0\rangle$). The *multimodal language* is the one obtained by enriching the previous one with canonical constants and a unary operator $\Box_a$ for every $a \in A \setminus \{0\}$. We remind the reader that in this paper (see p. 1) with the term *modal language* we denote the language obtained enriching $\langle\wedge, \vee, \odot, \rightarrow, 1, 0\rangle$ with canonical constants and one unary operator $\Box$. In case we are interested in the expansion having simultaneously, besides canonical constants, the operators $\Box_a$'s and $\Box$ we will use the expression *full modal language*. Of course we adopt the analogous convention to talk about *multimodal formulas*, *modal formulas* and *full modal formulas*.

A *Kripke frame* is a pair $\mathfrak{F} = \langle W, R\rangle$ where W is a non empty set (whose elements are called *worlds*) and R is a binary relation valued in A (i.e., $R : W \times W \longrightarrow A$) called *accessibility relation*. $\mathfrak{F}$ is said to be *crisp* in case that the range of R is included in $\{0, 1\}$. The classes of Kripke frames and crisp Kripke frames will be denoted, respectively, by Fr and CFr.

A *Kripke model* is a 3-tuple $\mathfrak{M} = \langle W, R, V\rangle$ where $\langle W, R\rangle$ is a Kripke frame and V is a map, called *valuation*, assigning

ISBN: 978-989-95079-6-8

Table 2: Axiomatization of the set $\Lambda(\mathsf{CFr}, \mathbf{A^c})$ when $\mathbf{A}$ is a finite chain

- the set of axioms is the smallest set closed under substitutions containing
  - the axiomatic basis for $\mathbf{\Lambda(A)}$,
  - the witnessing axiom $\bigvee_{a\in A}(\varphi \leftrightarrow \overline{a})$
  - the bookkeeping axioms $(\overline{a_1} * \overline{a_2}) \leftrightarrow \overline{a_1 * a_2}$ (for every $a_1, a_2 \in A$ and every $* \in \{\wedge, \vee, \odot, \rightarrow\}$),
  - $\Box 1$, $(\Box\varphi \wedge \Box\psi) \rightarrow \Box(\varphi \wedge \psi)$ and $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$
  - $\Box(\overline{a} \rightarrow \varphi) \leftrightarrow (\overline{a} \rightarrow \Box\varphi)$ (for every $a \in A$) and $\Box(\overline{k} \vee \varphi) \rightarrow (\overline{k} \vee \Box\varphi)$ (where k is the coatom of $\mathbf{A}$),
- the rules of the basis for $\mathbf{\Lambda(A)}$, the rule $\overline{k} \vee \varphi \vdash \varphi$ (where k is the coatom of $\mathbf{A}$) and the Necessity rule $\varphi \vdash \Box\varphi$.

to each propositional variable and each world in W an element of A (i.e., $V : \mathsf{Var} \times W \longrightarrow A$ where Var is the set of propositional variables). In such a case we say that $\mathfrak{M}$ is based on the Kripke frame $\langle W, R\rangle$.

If $\mathfrak{M} = \langle W, R, V\rangle$ is a Kripke model, then the map V can be uniquely extended to a map V' assigning to each full modal formula (in particular also multimodal and modal formulas) and each world in W an element of A (i.e., $V' : Fm \times W \longrightarrow A$) satisfying that:[4]

- V' is an algebraic homomorphism, in its first component, for the connectives $\wedge, \vee, \odot, \rightarrow, 1$ and 0,
- $V'(\overline{a}, w) = a$ for every $a \in A$.
- $V'(\Box\varphi, w) = \bigwedge\{R(w, w') \rightarrow V'(\varphi, w') : w' \in W\}$.
- $V'(\Box_a\varphi, w) = \bigwedge\{R_a(w, w') \rightarrow V'(\varphi, w') : w' \in W\} = \bigwedge\{V'(\varphi, w') : w' \in W, R(w, w') \geqslant a\}$.

Although V and V' are different mappings there is no problem, since one is an extension of the other, to use the same notation V for both. As usual we say that two formulas φ and ψ are *equivalent* (in symbols $\varphi \equiv \psi$) iff for every Kripke model $\mathfrak{M}$ it holds that $V(\varphi) = V(\psi)$.

Definition 2.1. The *local (many-valued) full modal logic* $\mathbf{F\Lambda}(l, \mathsf{Fr}, \mathbf{A^c})$ is the consequence relation obtained by defining, for all sets $\Gamma \cup \{\varphi\}$ of full modal formulas,

- $\Gamma \vdash_{\mathbf{F\Lambda}(l,\mathsf{Fr},\mathbf{A^c})} \varphi$, iff
- For every Kripke model $\langle W, R, V\rangle$ and $w \in W$, if $V(\gamma, w) = 1$ for every $\gamma \in \Gamma$, then $V(\varphi, w) = 1$.

And the *global (many-valued) full modal logic* $\mathbf{F\Lambda}(g, \mathsf{Fr}, \mathbf{A^c})$ is the one given by defining

- $\Gamma \vdash_{\mathbf{F\Lambda}(g,\mathsf{Fr},\mathbf{A^c})} \varphi$, iff
- For every Kripke model $\langle W, R, V\rangle$, if $V(\gamma, w) = 1$ for every $\gamma \in \Gamma$ and every $w \in W$, then $V(\varphi, w) = 1$ for every $w \in W$.

In the case that we restrict our attention to multimodal formulas or modal ones we will analogously use the notations $\mathbf{M\Lambda}(l, \mathsf{Fr}, \mathbf{A^c})$, $\mathbf{M\Lambda}(g, \mathsf{Fr}, \mathbf{A^c})$, $\mathbf{\Lambda}(l, \mathsf{Fr}, \mathbf{A^c})$ and $\mathbf{\Lambda}(g, \mathsf{Fr}, \mathbf{A^c})$. A formula is *valid* iff it is a theorem of the local (or the global) consequence relation. We will write $\mathsf{M\Lambda}(\mathsf{Fr}, \mathbf{A^c})$, $\mathsf{F\Lambda}(\mathsf{Fr}, \mathbf{A^c})$ and $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ to denote the set of valid formulas in the corresponding language.

From this definition it is obvious that all these local and global consequence relations are conservative expansions of the non-modal consequence relations with and without canonical constants, denoted respectively by $\mathbf{\Lambda(A^c)}$ and $\mathbf{\Lambda(A)}$.

3 Completeness of the multimodal systems

In this section we give an sketch of the proof of the following completeness theorem.

Theorem 3.1 (Completeness)**.** *Let* $\mathbf{A}$ *be a finite* MTL *chain.*

1. *The consequence relation* $\mathbf{M\Lambda}(g, \mathsf{Fr}, \mathbf{A^c})$ *is the consequence relation axiomatized by the axioms and rules given in Table 3.*
2. *The set* $\mathsf{M\Lambda}(\mathsf{Fr}, \mathbf{A^c})$ *is the set of modal formulas derivable in the calculus given in Table 3.*
3. *The consequence relation* $\mathbf{M\Lambda}(l, \mathsf{Fr}, \mathbf{A^c})$ *is axiomatized by (i)* $\mathsf{M\Lambda}(\mathsf{Fr}, \mathbf{A^c})$ *as the set of axioms, and (ii) the rules of the basis for* $\mathbf{\Lambda(A)}$ *together with the rule* $\overline{k} \vee \varphi \vdash \varphi$ *(where* k *is the coatom of* $\mathbf{A}$*).*

We notice that the behaviour of the $\Box_b$ operators given in this axiomatization (Table 3) is the same one that was given for $\Box$ in the case of crisp Kripke frames (cf. Table 2), plus the addition of some axioms telling that the $\Box_b$'s are somehow nested.

Remark 3.2. Before explaining the proof it is worth pointing out that if we delete the axioms $\Box_b(\overline{k} \vee \varphi) \rightarrow (\overline{k} \vee \Box_b\varphi)$ from Table 3 then we get an incomplete system. This fact can be checked using an standard matrix argument. Let us consider the Gödel algebra $\mathbf{G_3}$ with three elements $\{0, 0.5, 1\}$, and expand it with canonical constants (i.e., $\overline{0.5} = 0.5$) and with the connectives $\Box_{0.5}$ and $\Box_1$ being interpreted by the unary function $f : \mathbf{G_3} \longrightarrow \mathbf{G_3}$ defined by $f(x) := 0.5 \rightarrow x$. Then, this algebra enriched with the set $\{1\}$ of designated elements is a matrix that is a model of all axioms and rules given in Table 3 except for the ones of the form $\Box_b(\overline{k} \vee \varphi) \rightarrow (\overline{k} \vee \Box_b\varphi)$. It is not a model of these last axioms because for example $\Box_1(\overline{0.5} \vee 0) \rightarrow (\overline{0.5} \vee \Box_1 0) = 0.5$.

[4] The infimun of the empty set is taken, as usual, equal to 1.

ISBN: 978-989-95079-6-8

Table 3: Axiomatization of the set $\mathsf{M}\Lambda(\mathsf{Fr}, \mathbf{A^c})$ when $\mathbf{A}$ is a finite chain

- the set of axioms is the smallest set closed under substitutions containing
  - the axiomatic basis for $\mathbf{\Lambda(A)}$,
  - the witnessing axiom $\bigvee_{a\in A}(\varphi \leftrightarrow \overline{a})$
  - the bookkeeping axioms $(\overline{a_1} * \overline{a_2}) \leftrightarrow \overline{a_1 * a_2}$ (for every $a_1, a_2 \in A$ and every $* \in \{\wedge, \vee, \odot, \rightarrow\}$),
  - $\Box_b 1$, $(\Box_b\varphi \wedge \Box_b\psi) \rightarrow \Box_b(\varphi \wedge \psi)$ and $\Box_b(\overline{a} \rightarrow \varphi) \leftrightarrow (\overline{a} \rightarrow \Box_b\varphi)$ (for every $a \in A$ and $b \in A \setminus \{0\}$),
  - $\Box_b(\varphi \rightarrow \psi) \rightarrow (\Box_b\varphi \rightarrow \Box_b\psi)$ and $\Box_b(\overline{k} \vee \varphi) \rightarrow (\overline{k} \vee \Box_b\varphi)$ (for every $b \in A \setminus \{0\}$) (where k is the coatom of $\mathbf{A}$),
  - $\Box_{b_1}\varphi \rightarrow \Box_{b_2}\varphi$ (for every $b_1, b_2 \in A \setminus \{0\}$ such that $b_1 \leqslant b_2$),
- the rules of the basis for $\mathbf{\Lambda(A)}$, the rule $\overline{k} \vee \varphi \vdash \varphi$ (where k is the coatom of $\mathbf{A}$) and the Necessity rules $\varphi \vdash \Box_b\varphi$ (for every $b \in A \setminus \{0\}$).

Next we give some hints on the proof of Theorem 3.1. To this purpose for the rest of this section we will use the symbol L to denote the set of modal formulas derivable from the calculus given in Table 3, and the symbol $\vdash_{\mathsf{L}}$ to denote the consequence relation axiomatized by L as set of axioms and the rules of the basis for $\mathbf{\Lambda(A)}$ together with the rule $\overline{k} \vee \varphi \vdash \varphi$ (where k is the coatom of $\mathbf{A}$). The completeness theorem can be proved using two steps: (i) a reduction of the multimodal completeness problem to the already known strong completeness of the non-modal logic $\mathbf{\Lambda(A^c)}$, and (ii) a Truth Lemma based on a canonical Kripke model construction. Next we briefly sketch the proofs of each one of these steps in the completeness proof.

The first step is based on the fact that $\vdash_{\mathsf{L}}$ is strongly complete by definition with respect to $\mathbf{\Lambda(A)}$, and hence by [1, Corollary 2.16] we know that it is also strongly complete with respect to $\mathbf{\Lambda(A^c)}$. We point out that the rule $\overline{k} \vee \varphi \vdash \varphi$ plays a remarkable role in the proof of [1, Corollary 2.16]. Therefore, we get the following trivial consequence.

Lemma 3.3 (Non-Modal Reduction)**.** *Let $\Gamma \cup \{\varphi\}$ be a set of multimodal formulas. Then*

- *$\Gamma \vdash_{\mathsf{L}} \varphi$, iff*
- *$\Gamma \vdash_{\mathbf{\Lambda(A^c)}} \varphi$, i.e., for every homomorphism h from the algebra of multimodal formulas[5] into the algebra $\mathbf{A^c}$, if $h[\Gamma] \subseteq \{1\}$ then $h(\varphi) = 1$.*

On the other hand, the second step consists on a canonical Kripke model construction. The definition of this construction is the following one.

Definition 3.4. The *multi canonical Kripke model* $\mathfrak{M}_{\mathtt{mcan}}$ is the Kripke model $\langle W_{mcan}, R_{mcan}, V_{mcan}\rangle$ where

- the set W_{mcan} is the set of non-modal homomorphisms $v : \mathbf{Fm} \longrightarrow \mathbf{A^c}$ (we point out that the algebra $\mathbf{Fm}$ is the one given by multimodal formulas) such that $v[\mathsf{L}] = \{1\}$.
- the accessibility relation R_{mcan} is defined by[6]

$$R_{mcan}(v_1, v_2) := \bigvee\{b \in A\setminus\{0\} : \forall\varphi(v_1(\Box_b\varphi) \leqslant v_2(\varphi))\}.$$

- the valuation map is defined by $V_{mcan}(p, v) := v(p)$ for every variable p.

It is obvious that for every $b \in A \setminus \{0\}$, it holds that $b \leqslant R_{mcan}(v_1, v_2)$ iff $v_1(\Box_b\varphi) \leqslant v_2(\varphi)$ for every multimodal formula φ.

Lemma 3.5 (Truth Lemma)**.** *The multi canonical Kripke model $\mathfrak{M}_{\mathtt{mcan}}$ satisfies that $V_{mcan}(\varphi, v) = v(\varphi)$ for every multimodal formula φ and world v.*

Proof. The proof is done by induction on the multimodal formula. The only non trivial case is when this formula starts with a necessity operator $\Box_b$. By the inductive hypothesis it is clear that it is enough to prove that

$$\bigwedge\{v'(\varphi) : v' \in W_{mcan}, b \leqslant R_{mcan}(v, v')\} = v(\Box_b\varphi)$$

where v is a world. And indeed the only non trivial inequality is $\leqslant$. Hence, let us consider $a := \bigwedge\{v'(\varphi) : v' \in W_{mcan}, b \leqslant R_{mcan}(v, v')\}$ and try to prove that $a \leqslant v(\Box_b\varphi)$.

First of all we claim that

$$\mathsf{L} \cup \{\overline{d} \rightarrow \psi : \psi \in Fm, d = v(\Box_b\psi) \in A\} \vdash_{\mathbf{\Lambda(A^c)}} \overline{a} \rightarrow \varphi.$$

Why? Let us consider a homomorphism h from the algebra of multimodal formulas into the algebra $\mathbf{A^c}$ such that $h[\mathsf{L}\cup\{\overline{d} \rightarrow \psi : \psi \in Fm, d = v(\Box_b\psi) \in A\}] = \{1\}$. We have to prove that $h(\overline{a} \rightarrow \varphi) = 1$, i.e., $a \leqslant h(\varphi)$. The assumptions on h imply that $h \in W_{mcan}$, and that $v(\Box_b\psi) \leqslant h(\psi)$ for every ψ. Hence, $b \leqslant R_{mcan}(v, h)$. Using the definition of a we get that $a \leqslant h(\varphi)$. This finishes the proof of this first claim.

Hence, using that $\mathbf{\Lambda(A^c)}$ is a finitary logic (because A is finite) we get from the previous claim that there is some $m \in \omega$, some multimodal formulas $\psi_1, \ldots, \psi_m$ and some elements $d_1, \ldots, d_m \in A$ such that $d_i = v(\Box_b\psi_i)$ for every $i \in \{1, \ldots, m\}$ and

$$\mathsf{L} \cup \{\overline{d_1} \rightarrow \psi_1, \ldots, \overline{d_m} \rightarrow \psi_m\} \vdash_{\mathbf{\Lambda(A^c)}} \overline{a} \rightarrow \varphi.$$

From here it is clear[7] that $\mathsf{L} \vdash_{\mathbf{\Lambda(A^c)}} ((\overline{d_1} \rightarrow \psi_1)\wedge\ldots\wedge(\overline{d_m} \rightarrow \psi_m)) \rightarrow (\overline{k} \vee (\overline{a} \rightarrow \varphi))$. Using Lemma 3.3 we get from the

[5] When we look at the multimodal formulas as non-modal ones we are thinking that $\{\Box_a\varphi : a \in A \setminus \{0\}, \varphi$ multimodal formula$\}$ are the variables of this non-modal language.

[6] The supremum of the empty set is taken, as usual, equal to 0.

[7] The trick used here is a particular case of the following more general statement:

$$\Gamma, \gamma \vdash_{\mathbf{\Lambda(A^c)}} \varphi \quad \text{iff} \quad \Gamma \vdash_{\mathbf{\Lambda(A^c)}} \gamma \rightarrow (\overline{k} \vee \varphi).$$

This statement is a consequence of the fact that k is the coatom of $\mathbf{A}$.

ISBN: 978-989-95079-6-8

previous claim that $\mathsf{L} \vdash_{\mathsf{L}} ((\overline{d_1} \to \psi_1) \wedge \ldots \wedge (\overline{d_m} \to \psi_m)) \to (\overline{k} \vee (\overline{a} \to \varphi))$. Thus, $((\overline{d_1} \to \psi_1) \wedge \ldots \wedge (\overline{d_m} \to \psi_m)) \to (\overline{k} \vee (\overline{a} \to \varphi)) \in \mathsf{L}$. Therefore, $\Box_b((\overline{d_1} \to \psi_1) \wedge \ldots \wedge (\overline{d_m} \to \psi_m)) \to \Box_b(\overline{k} \vee (\overline{a} \to \varphi)) \in \mathsf{L}$ by the Monotonicity Rule. Thus, $((\overline{d_1} \to \Box_b\psi_1) \wedge \ldots \wedge (\overline{d_m} \to \Box_b\psi_m)) \to (\overline{k} \vee (\overline{a} \to \Box_b\varphi)) \in \mathsf{L}$ using some of the axioms in Table 3. Finally, using that $v(\overline{d_1} \to \Box_b\psi_1) = \ldots = v(\overline{d_m} \to \Box_b\psi_m) = 1$ and that $k \neq 1$, we obtain that $v(\overline{a} \to \Box_b\varphi) = 1$, i.e., $a \leqslant v(\Box_b\varphi)$. This finishes the proof. □

Remark 3.6. In the calculus given in Table 3 we have included for each one of the $\Box_b$ modalities the meet distributivity axiom, the normality axiom, the Monotonicity rule and the Necessity rule. Checking the details of the proof of Lemma 3.5 the reader can realize that among the previous four axioms/rules it is enough to include the meet distributivity axiom and the Monotonicity rule in order to get completeness. And the same is also true if we only include the normality axiom and the Necessity rule, but this time the argument is slightly different (just replace $\wedge$ with $\odot$ in the proof of Lemma 3.5).

4 Comparing Expressive Power

The aim of this section is to compare the expressive power of the multimodal language with the one of the modal language. To this purpose we introduce the following concepts.

Definition 4.1. Given two pointed[8] Kripke models $\langle \mathfrak{M}, w\rangle$ and $\langle \mathfrak{M}', w'\rangle$, we will say that they are *modally equivalent*, in symbols $\langle \mathfrak{M}, w\rangle \equiv \langle \mathfrak{M}', w'\rangle$, in the case that $V(\varphi, w) = V'(\varphi, w')$ for every modal formula φ. Analogously we will talk about *multimodally equivalent* and *full modally equivalent*, in symbols $\equiv_{\mathsf{M}}$ and $\equiv_{\mathsf{F}}$, in the case we focus, respectively, on multimodal formulas or full modal formulas.

The first remark comparing expressive powers is that in every finite residuated lattice $\mathbf{A}$ (here it is not needed the chain assumption) it holds that

$$\Box\varphi \;\equiv\; \bigwedge\{\overline{a} \to \Box_a\varphi : a \in A \setminus \{0\}\} \tag{3}$$

Therefore, the modality $\Box$ is explicitly definable using the modalities $\Box_a$'s (and these last ones have the advantage that satisfy the normality axiom). In other words, the expressive power of the modal language is smaller than the one of the multimodal one. Thus, $\Lambda(\mathsf{Fr}, \mathbf{A^c})$ can be seen as a fragment of $\mathsf{M}\Lambda(\mathsf{Fr}, \mathbf{A^c})$. It is obvious that if two pointed Kripke models are multimodally equivalent then they are also modally equivalent.

What about the converse direction in the statements from last paragraph? That is, (i) is it possible to explicitly define the modalities $\Box_a$'s in the modal language?, and (ii) are modally equivalent pointed Kripke models also multimodally equivalent? In the rest of the section we will discuss these two questions. It is worth pointing out that as far as the authors are aware there is no general result in the many-valued modal setting relating these last two questions: an study of Beth definability in this setting has not been undertaken (cf. [9, p. 277]). As a matter of fact it is necessary to distinguish between having an empty set Var of propositional variables or not.

[8] By a *pointed Kripke model* we mean a Kripke model together with a distinguished point.

Proposition 4.2 (Case $\mathsf{Var} = \emptyset$)**.** *Let* $\mathbf{A}$ *be a finite* MTL *chain. Then, if two pointed Kripke models are modally equivalent then they are also multimodally equivalent.*

Proof. First of all we point out that it is enough to prove that if $\langle \mathfrak{M}, w\rangle \equiv \langle \mathfrak{M}', w'\rangle$, $w_0 \in W$ and $R(w, w_0) \geqslant a \neq 0$, then there is some $w_0' \in W'$ such that $R'(w', w_0') \geqslant a$ and $\langle \mathfrak{M}, w_0\rangle \equiv \langle \mathfrak{M}', w_0'\rangle$. The proof finishes by realizing that the previous statement is true because for every $a \in A \setminus \{0\}$, it holds that $V(\Box\overline{\mathsf{pred}(a)}, w) \neq 1$ iff there is some w_0 such that $R(w, w_0) \geqslant a$. The notation $\mathsf{pred}(a)$ refers to the predecessor of element a. □

Therefore, it is obvious that if $\mathsf{Var} = \emptyset$ then two pointed Kripke models are modally equivalent iff they are multimodally equivalent. It is still open whether the modalities $\Box_a$'s are explicitly definable or not when there are no propositional variables.

In the case that there is some propositional variable then the situation is quite different as next proposition shows. We will see later in Theorem 4.5 that the assumption in this proposition is also a necessary condition for the case of finite BL chains.

Proposition 4.3 (Case $\mathsf{Var} \neq \emptyset$)**.** *Let* $\mathbf{A}$ *be the ordinal sum* $\mathbf{A_1} \oplus \mathbf{A_2}$ *of two finite* MTL *chains such that* $\mathbf{A_1}$ *and* $\mathbf{A_2}$ *are non trivial (i.e.,* $\min\{|A_1|, |A_2|\} \geqslant 2$*). Then, there are two pointed Kripke models that are modally equivalent but not multimodally equivalent.*

Proof. Let us define $a \in A$ as the minimum element of $\mathbf{A_2}$ (i.e., a is the idempotent element separating both components). Since $\mathbf{A_i}$'s are non trivial it is obvious that $a \notin \{0^{\mathbf{A}}, 1^{\mathbf{A}}\}$, and also that for every $b \in A$, $a \odot b = a \wedge b$. Next we consider the two Kripke models given[9] in Fig. 1. It is obvious that for every modal formula φ it holds that $V(\varphi, w_1) = V'(\varphi, w_1')$. On the other hand, by induction on the length of formulas the reader can easily prove[10] that for every modal formula φ (we remind that canonical constants are allowed), it holds both that

- $a \leqslant V'(\varphi, w_1')$ iff $a \leqslant V'(\varphi, w_2')$,
- if $a \not\leqslant V'(\varphi, w_1')$ then $V'(\varphi, w_1') = V'(\varphi, w_2')$.

These last two properties guarantees that for every modal formula φ, it holds that $V'(\varphi, w_1') \leqslant a \to V'(\varphi, w_2')$; and so $V'(\Box\varphi, w_0') = V'(\varphi, w_1')$. Once we know this last fact, it becomes easy to show by an straightforward induction that for every modal formula φ, it holds that $V(\varphi, w_0) = V'(\varphi, w_0')$. Thus, $\langle \mathfrak{M}, w_0\rangle$ and $\langle \mathfrak{M}', w_0'\rangle$ are modally equivalent. Finally, using that $V(\Box_a p, w_0) = 1 \neq a = V'(\Box_a p, w_0')$ we get that $\langle \mathfrak{M}, w_0\rangle$ and $\langle \mathfrak{M}', w_0'\rangle$ are not multimodally equivalent. □

Lemma 4.4. *Let* $\mathbf{A}$ *be the finite* MV *chain of cardinal* n*. Then, for every* $a \in A \setminus \{0\}$ *it holds that*[11]

$$\Box_a\varphi \;\equiv\; \bigwedge\left\{\left(\overline{a} \to \neg\Box\neg((\varphi \leftrightarrow \overline{b})^{n-1})\right)^{n-1} \to \overline{b} : b \in A\right\}$$

[9] The convention here adopted about presenting Kripke models as diagrams is the same one stated in [1, Convention 3.6].

[10] In this inductive proof it plays a crucial role the definition of ordinal sum.

[11] Notice that φ^{n-1} only takes crisp values (i.e., in $\{0, 1\}$). Indeed, φ^{n-1} takes value 1 when φ takes value 1, and φ^{n-1} takes value 0 elsewhere. Hence, φ^{n-1} takes the same value than $\Delta\varphi$ where Δ is the well-known operator used in fuzzy logic (see [7]).

ISBN: 978-989-95079-6-8

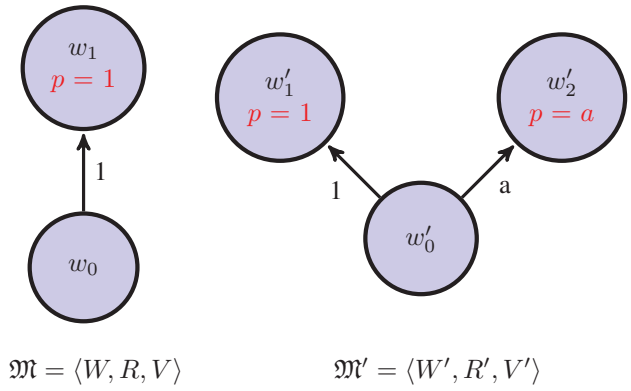


Figure 1: Two interesting Kripke models

Proof. Let us introduce the abbreviations $\Diamond\varphi := \neg\Box\neg\varphi$ and $\Delta\varphi := \varphi^{n-1}$. Then, the statement says that

$$\Box_a\varphi \;\equiv\; \bigwedge \{\Delta(\overline{a} \to \Diamond\Delta(\varphi \leftrightarrow \overline{b})) \to \overline{b} : b \in A\}$$

In this proof we will use the notation

$$f(w, \varphi, b) := \bigvee\{R(w, w') : w' \in W, V(\varphi, w') = b\}.$$

The reader can easily check the following steps.

$$V(\Diamond\Delta(\varphi \leftrightarrow \overline{b}), w) = f(w, \varphi, b)$$

$$V(\Delta(\overline{a} \to \Diamond\Delta(\varphi \leftrightarrow \overline{b})), w) = \begin{cases} 1, & \text{if } a \leqslant f(w, \varphi, b) \\ 0, & \text{if not} \end{cases}$$

$$V(\Delta(\overline{a} \to \Diamond\Delta(\varphi \leftrightarrow \overline{b})) \to \overline{b}, w) = \begin{cases} b, & \text{if } a \leqslant f(w, \varphi, b) \\ 1, & \text{if not} \end{cases}$$

Then, it easily follows that $V(\bigwedge \{\Delta(\overline{a} \to \Diamond\Delta(\varphi \leftrightarrow \overline{b})) \to \overline{b} : b \in A\}, w) = \bigwedge\{V(\varphi, w') : w' \in W, R(w, w') \geqslant a\} = V(\Box_a\varphi, w)$. This finishes the proof. □

Theorem 4.5 (Case $\mathsf{Var} \neq \emptyset$)**.** *Let* $\mathbf{A}$ *be a finite* BL *chain. Then, the following statements are equivalent.*

1. $\mathbf{A}$ *is a finite* MV *chain (i.e., the only idempotent elements of* A *are* 0 *and* 1*),*
2. *The modalities* $\Box_a$*'s are explicitly definable*[12] *in the modal language.*
3. *Two pointed Kripke models are modally equivalent iff they are multimodally equivalent,*

Proof. For the case that $|A_1| \leqslant 2$ it is trivial that these statements are equivalent. Hence, we only have to deal with the case that $|A_1| \geqslant 3$; and this case is a consequence of Proposition 4.3 and Lemma 4.4 together with the decomposition of BL chains as ordinal sums. □

5 Concluding Remarks

As it has been stressed in this paper, and also in [1], the role played by canonical constants is crucial in several of the proofs given above. For example, it is unknown to the authors whether there is a general method for converting an axiomatization of $\mathbf{\Lambda(A)}$ (and so without canonical constants) into one of $\Lambda(\mathsf{CFr}, \mathbf{A})$. One of the few statements that we know that remains true when we remove canonical constants from the language is Theorem 4.5. That is, for finite MV chains it is possible to find a modal formula (and so using $\Box$) without canonical constants that is equivalent to $\Box_a p$. The proof of this stronger version of Theorem 4.5 is completely different and is based on a combinatorial analysis of all possibilities; it is a rather involved proof and due to space limitations has not been included in this paper.

[12] That is, there is some modal formula φ such that $\varphi \equiv \Box_a p$. We remind that p refers to a propositional variable.

Acknowledgments

The collaborative work between the authors of this paper has been made possible by several research grants. The first three authors acknowlege partial support by the Spanish project MULOG2 TIN2007-68005-C04-01/04, the EU-ROCORES(LogICCC) project FP006 LoMoReVi (linked to the Spanish project FFI2008-03126-E/FILO), and the grants 2005SGR-00083 and 2005SGR-00093 of the Catalan Government. The last author has been partially financed by CyT-UBA project X484 and CONICET project PIP 112-200801-02543.

References

[1] F. Bou, F. Esteva, L. Godo, and R. Rodríguez. On the minimum many-valued modal logic over a finite residuated lattice, 2008. Submitted. Available at `http://arxiv.org/abs/0811.2107`.

[2] L. A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

[3] U. Höhle. Commutative, residuated l-monoids. In U. Höhle and E. P. Klement, editors, *Non-classical logics and their applications to fuzzy subsets (Linz, 1992)*, volume 32 of *Theory Decis. Lib. Ser. B Math. Statist. Methods*, pages 53–106. Kluwer Acad. Publ., Dordrecht, 1995.

[4] P. Jipsen and C. Tsinakis. A survey of residuated lattices. In J. Martinez, editor, *Ordered Algebraic Structures*, pages 19–56. Kluwer Academic Publishers, Dordrecht, 2002.

[5] H. Ono. Substructural logics and residuated lattices - an introduction. In V. F. Hendricks and J. Malinowski, editors, *50 Years of Studia Logica*, volume 21 of *Trends in Logic—Studia Logica Library*, pages 193–228. Dordrecht, 2003.

[6] F. Esteva and L. Godo. Monoidal t-norm based logic: towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124:271–288, 2001.

[7] P. Hájek. *Metamathematics of fuzzy logic*, volume 4 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 1998.

[8] R. Cignoli, I. M. L. D'Ottaviano, and D. Mundici. *Algebraic foundations of many-valued reasoning*, volume 7 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 2000.

[9] P. Blackburn, J. F. A. K. van Benthem, and F. Wolter, editors. *Handbook of Modal Logic, Volume 3 (Studies in Logic and Practical Reasoning)*. Elsevier Science Inc., New York, NY, USA, 2006.

ISBN: 978-989-95079-6-8

Fuzzy Optimization in Vehicle Routing Problems

J. Brito, J. A. Moreno[1] J. L. Verdegay[2]

1 Grupo de Computación Inteligente, I.U.D.R.
Universidad de La Laguna, E-38271 La Laguna, Spain
2 Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad de Granada, E-18071 Granada, Spain
Email: {jbrito,jamoreno}@ull.es, verdegay@decsai.ugr.es

Abstract—*This paper focuses on analyzing practical solution approaches for Vehicle Routing Problems (VRP) with uncertain information. Several variants of the basic problem and fuzzy optimization problem formulations are described. The fuzzy VRP is obtained when some of the elements in the formulation are fuzzy. The main versions of Fuzzy VRP that have appeared in the literature are reviewed and the standard approaches for solving the corresponding models are analyzed.*

Keywords—Vehicle Routing Problem, Fuzzy Optimization, Fuzzy Vehicle Routing Problem, VRP with Time Windows.

1 Introduction

Route planning problems are among the activities that have the highest impact in logistical planning, transport and distribution because of their effects on efficiency in resource management, service levels, and client satisfaction. Route distribution planning problems, also known as Vehicle Routing Problems (VRP), have been thoroughly studied in a variety of areas, such as Operations Research, Artificial Intelligence, etc. The standard VRP was originally introduced by Dantzig and Ramser (1959), and is NP-hard, which is a complex combinatorial optimization problem [5]. Several variants of the basic problem have been put forward and strong formulations have been proposed. Most of these problems can be modeled as linear programming problems. The most common solution techniques are exact methods that guarantee finding an optimal solution if it exists. These approaches have also been applied together with numerous heuristics solution techniques developed with enough flexibility in optimization systems and can be adapted to various practical contexts.

Given that the complex, flexible, and dynamic nature of real logistical planning produces a high degree of uncertainty related to the decision making process, is not always possible to have all of the necessary information available at the onset of the problem. Consequently the most common scenario provides incomplete or imprecise information of the parameters and variables. The use of fuzzy sets to approach these situations is very appropriate to build computing systems for the solution of solving decision and optimization problems. The modeling of these problems is complicated at various levels: not only are they difficult to define accurately and need to manage uncertainty, but there is also imprecision in the available information and stated preferences, restrictions and objectives by decision makers.

A frequent occurrence in real decision-making problems, such as those found in VRP, is the lack of precision or uncertainty in the information available in them. This characteristic complicates the process concerning the definition of their objectives and parameters. Although this type of uncertainty found in the nature of the data and their settings has traditionally been handled by means of probability theory, in most of the cases they cannot be considered random phenomena and therefore probability theory cannot be applied successfully.

The aim of this paper is to analyze how the relevance and the variety of applications found in VRP can be considered under the fuzzy context point of view. The organization of the paper follows. Section 2 introduces the broad class of optimization problems known as VRP. We also survey the concept of fuzzy optimization where the objective function and constraints have fuzzy parameters, and then we apply these concepts to the simple and classical VRP. Finally, we review the corresponding fuzzy versions, which are called Fuzzy VRP.

2 Vehicle Routing Problems

The VRP are one of the most widely studied class of problems in combinatorial optimization, and the literature provides several exact and heuristic solution techniques of general applicability. VRP are problems where several vehicles that must serve points of demands and satisfy a finite set of constraints while minimizing costs, distances or times [20].

The standard VRP (usually called capacitated VRP; CVRP) calls for the determination of a set of *m* routes whose total travel length is minimized such that: (1) each customer is visited exactly once by one route, (2) each route starts and ends at a single depot, (3) the total demand of the customers served by a route does not exceed a given vehicle capacity Q, and (4) the length of each route does not exceed a preset limit L. A constant speed is typically assumed so that to minimize distances, travel times and travel costs are considered equivalent. If each vehicle *i* is assigned to a route R_i, a feasible solution for the VRP is made up of a partition from V into *m* routes $R_1, R_2, ..., R_m$ and the corresponding permutations σ_i of $R_i \cup 0$ that specify client order along the routes.

VRP allow us to consider several types of constraints, depending on the specific characteristics of the problem and decision-making process [4]. These possibilities lead to a variety of problems, beginning with the standard or basic VRP. Their descriptions follow:

ISBN: 978-989-95079-6-8

MFVRP- *Mix Fleet VRP*. This is a VRP in which the vehicles have different capacities, in other words, fleet of vehicles with a heterogeneous capacity. Each resource must consider these capacities for each route.

MDVRP- *Multi-depot VRP*. A company can have several depots or warehouses for use when supplying client demand. If clients are grouped around the depots then the problem can be viewed as a set of independent VRP problems. However, if the clients and depots are spread apart, then this scenario is known as a MDVRP.

PVRP - *Period VRP*. In the classical VRP the planning period of the route is one day. The PVRP expands the planning period to *M* days. During this period of *M* days, each client should be visited a given number of days.

SDVRP - *Split-up Delivery VRP*. This problem is a VRP in which a client is allowed to be serviced by several vehicles if the total cost decreases. Order size is important since the total size of the client orders must exceed the capacity of one vehicle.

PDVRP - *Pickup and Delivery VRP*. This variation is a VRP since it allows clients the possibility of returning certain goods that have to be delivered to other clients. Therefore, each vehicle that is fitted must consider client pickup and delivery amounts as well as a planned order.

VRPB – *VRP with Backhauls.* This problem is VRP where clients can order or return articles. Therefore the formulation requires that returned goods from clients fit in the vehicle. In addition, all deliveries must be completed before pickups begin because vehicles are back-loaded and pickups in the vehicles are considered uneconomical or unfeasible. Orders and pickups must be known ahead of time.

VRPTW- *VRP with Time Windows*. This variation is a VRP with the additional constraint that associates a time window to each client which identifies the only interval of time when the client is willing to receive goods or services. If a vehicle arrives to the client early the vehicle has to wait. If it arrives during the interval of the time window, the vehicle makes the delivery at that moment. Finally, if it arrives late the client is not serviced.

OVRP - *Open VRP*. This problem is a special class of VRP within all open delivery routes, since vehicles are not required to return to the depot.

Finally, other versions of VRP do exist. Some versions are mixes or hybrids of the previously mentioned typologies, such as the MDVRPTW or OPDVRP. Other problem types introduce specific characteristics which require the corresponding modifications in the model. For example, one version of a particularly interesting (and increasingly complex) problem is a dynamic version of the route problem with or without time windows (DVRP, Dynamic VRP). In this version the problem characteristics vary as time passes.

3 Fuzzy optimization problems

An *optimization problem* can be described as the search for the value of specific *decision variables* so that identified *objective functions* attain their optimum values. The value of the variables is subject to stated *constraints*. In these problems the objective functions are defined on a set of solutions that we will denote by *X*. The objective function is not subject to any condition or property nor is the definition of the set *X*. Typically the number of elements of *X* is very high, essentially eliminating the possibility of a complete evaluation of all its solutions while determining the optimal solution.

Optimization problems in their most general form involve finding an optimal solution according to stated criteria. In practice, however, many situations lack the exact information that is needed in the problem, including its constraints, or in other cases, where it is unreasonable to access such specific constraints or clearly defined objective functions. In these situations it is advantageous to model and solve the problem using soft computing and fuzzy techniques.

Among all the optimization problems, the models that have received the most attention and have offered the most useful applications in different areas are Linear Programming (LP) models, which is the single objective linear case with linear constraints. The classic problem of LP is to find the maximum or minimum values of a linear function subject to constraints that are represented by linear equations or inequalities. The most general formulation of the LP problem is:

$$\begin{aligned} &\max && z = cx \\ &\text{subject to} && Ax \le b \\ & && x \ge 0 \end{aligned} \qquad (1)$$

The vector $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ represents the decision variables. The objective function is denoted by *z*, the numbers c_j are coefficients and the vector $c = (c_1, c_2, ..., c_n) \in \mathbb{R}^n$ is known as the cost vector. The matrix $A = [a_{ij}] \in \mathbb{R}^{n \times m}$ is called the constraint or technological matrix and the vector $b = (b_1, b_2, ..., b_m) \in \mathbb{R}^m$ represents the independent terms or right-hand-side of the constraints.

In many real situations not all the constraints and objective functions can be valued in a precise way. In these situations we are dealing with the general problem form of Fuzzy Linear Programming (FLP). FLP is characterized as follows: a_{ij}, b_j and c_i can be expressed as *fuzzy numbers*, x_i as variables whose states are fuzzy numbers, addition and multiplication operates with fuzzy numbers, and the inequalities are among fuzzy numbers.

Different FLP models can be considered according to the elements that contain imprecise information that is taken as a basis for the classification proposed in [22], [2]. Short descriptions of these models follow.

Models with feasible fuzzy set (fuzzy constraints)

This is the case where constraints can be satisfied, and consequently the feasible region can be defined as a fuzzy set; it should be defined by means of a membership function $\mu : \mathbb{R}^n \to [0,1]$. In such a situation, for each constraint, a desirable quantity *b* is considered, but the possibility that it is greater is accepted until a maximum *b+t* (*t* is referred to as a violation *tolerance level*). This model is represented by:

ISBN: 978-989-95079-6-8

$$\begin{aligned} &\max && z = cx \\ &\text{subject to} && Ax \le_f b \\ & && x \ge 0 \end{aligned} \quad (2)$$

where the symbol $\le_f$ indicates the imprecision of the constraints and where each fuzzy constraint $a_i x \le_f b_i$ is specified by a membership function in the form:

$$\mu_i(a_i x) = \begin{cases} 1 & \text{if } a_i x \le b_i \\ f_i(a_i x) & \text{if } b_i \le a_i x \le b_i + t_i \\ 0 & \text{if } b_i + t_i \le a_i x \end{cases} \quad (3)$$

which means that, for each constraint *i*, given the level of tolerance t_i, to each point (*n*-dimensional vector) *x is* associated a number $\mu_i(x) \in [0,1]$ known as the degree of fulfillment (or verification) of the constraint *i*. The functions f_i are assumed to be continuous and monotonous non-decreasing functions. In particular, Verdegay [21], using the representation theorem for fuzzy sets, proves that the solutions for the case of linear functions f_i can be obtained from the auxiliary model:

$$\begin{aligned} &\max && z = cx \\ &\text{subject to} && Ax \le b + t(1-\alpha) \\ & && x \ge 0, \alpha \in [0,1] \end{aligned} \quad (4)$$

where $t = (t_1, t_2, ..., t_m)$.

Models with fuzzy goals

A optimization problem with fuzzy goals allows the objective function value to be slightly below the minimum goal for a maximization problem, and similarly for a minimization problem. The corresponding linear model is expressed in the following way:

$$\begin{aligned} &\widetilde{\max} && z = cx \\ &\text{subject to} && Ax \le b \\ & && x \ge 0 \end{aligned} \quad (5)$$

If t_0 is the maximum quantity that the objective function should be inferior to the minimum goal c_0, then each vector *x* is associated to a number $\mu_0(x)$, which represents the degree that the decision maker considers to be an achieved goal. It is defined according to the following function:

$$\mu_0(x) = \begin{cases} 1 & \text{if } cx > c_0 \\ f_0(cx) & \text{if } c_0 - t_0 \le cx \le c_0 \\ 0 & \text{if } cx < c_0 - t_0 \end{cases} \quad (6)$$

where f_0 is a continuous, monotonous non-decreasing function. An operative model that provides satisfactory solutions can be found in [27].

Models with fuzzy costs as objective function coefficients

These models are those whose costs are not fully known (with imprecision). Therefore, they are represented by an *m*-dimensional fuzzy vector $c^f = (c_1^f, c_2^f, ..., c_n^f)$, and the following model:

$$\begin{aligned} &\max && z = c^f x \\ &\text{subject to} && Ax \le b \\ & && x \ge 0 \end{aligned} \quad (7)$$

Evidently, *z* is also a fuzzy number, but *x* can be a vector of fuzzy or non-fuzzy numbers, and each fuzzy cost is described by its corresponding membership function $\mu_i(x)$. Each coefficient c_i^f of the objective function is a plane fuzzy number of the L-R type with modal interval $[\underline{c}_j, \overline{c}_j]$ and membership functions g_j and h_j (which can be linear, parabolic, etc.). Delgado et al. [7] prove that the solution can be obtained with the multi-objective auxiliary model:

$$\begin{aligned} &\max && z = \left[c^1 x, ..., c^{2^n} x \right] \\ &\text{subject to} && Ax \le b \\ & && x \le 0, \alpha \in [0,1], c_j^k \in \left\{ g_j^{-1}(1-\alpha), h_j^{-1}(1-\alpha) \right\} \\ & && k = 1, ..., 2^n, \; j = 1, ..., n. \end{aligned} \quad (8)$$

Models with fuzzy coefficients in the technological matrix

Consider a problem of this type:

$$\begin{aligned} &\max && z = cx \\ &\text{subject to} && A^f x \le_f b^f \\ & && x \ge 0 \end{aligned} \quad (9)$$

where the values of the technological matrix and the coefficients are fuzzy numbers. Fuzzy constraints can also be included. Delgado et al. [6] also include imprecision in the constraints. They propose considering fuzzy solutions that are solved with the application of an ordered function *g* for the constraints. This new formulation is expressed by the auxiliary problem:

$$\begin{aligned} &\max && z = cx \\ &\text{subject to} && a_i^f x \le_g b_i^f + t_i^f (1-\alpha), \; i = 1, ...m \\ & && x \ge 0, \alpha \in [0,1] \end{aligned} \quad (10)$$

where the symbol $\le_g$ stands for a comparison relation between fuzzy numbers.

4 Fuzzy optimization Vehicle Routing Problems

Although there are different stochastic approaches to modeling and solving the VRP in the literature, this is not the case with the proposed approaches from fuzzy set theory. Furthermore, the literature offers very little in terms of modeling VRP proposals, both from the standpoint of the solutions modeled as Fuzzy Mathematical Programming, and Fuzzy VRP (FVRP). However, as we will discuss later, several proposals have been introduced in recent years, such as the VRPTW and Dynamic VRP.

Specifically, if we look at the FVRP models in the literature, the majority only assume vagueness for some of the following elements that are described in the model: a) Fuzzy demands (to be collected): customer demand is a imprecise variable and 2) Fuzzy times: service time and travel time can be imprecise variables.

The first problem deals with the demand vector of each client's ordered goods. Planning the ordered quantity by the clients is difficult to establish with sufficient notice and precise form, therefore we do not have access to a specific quantity. In other words, the information about vehicle demand at some nodes is often not precise enough. Thus,

ISBN: 978-989-95079-6-8

there is often uncertainty regarding the amount of demand at some nodes. This FVRP with fuzzy customer demand was first instructed by Teodorovic and Kikuchi [18]. In this paper, they treated the travel time and the transportation costs between two nodes in a network as fuzzy numbers. They modified the Clarke and Wright algorithm where travel times in a network are treated as fuzzy numbers. Later, Teodorović and Pavković [19] solve a VRP when demand at the nodes is uncertain and is represented by a triangular fuzzy number. The model is based on the heuristic "sweeping" algorithm, which uses fuzzy approximate reasoning procedures to decide whether or not to include a node in the route. It first uses the approximate reasoning algorithm to calculate the preference index. Once the membership function of the preference index has been determined, defuzzification must take place. In recent years, these same authors have proposed solutions to this problem [15], [16], where actual demand value is known only after the visit to the node. Their solution combines Bee and Ant systems with rules based on fuzzy logic. A new paper [14] has recently appeared which considers the VRP with uncertain demand at node. It uses the approximate reasoning algorithm to determine the preference strength to send the vehicle to next node, and the improved sweeping algorithm with vehicle coordinated strategy to determine a set of vehicle routes that minimizes costs.

The second problem, (fuzzy times in both service and travel) is characterized by other pieces of information that are increasingly imprecise, given the daily circumstances found in routing networks and traffic. In these cases service time, time windows and travel time are expressed as fuzzy numbers.

The traditional deterministic VRP is expanded to the situation so that the VRP has fuzzy travel time features. In [11] a simple description of the VRP with fuzzy traveling time, that is, a mathematical model for the problem, is built. It puts forward the concept of level effect function which can quantify the location of fuzzy number intensively and globally, and sets up the uncertain degree of measurement. In this paper the solution to the problem is based on a genetic algorithm.

In [17] the travel time based on the fuzzy mathematical model of the VRP takes the time window as a fuzzy variable. Information entropy and the path chosen by the use of random disturbance control strategy to the Ant algorithm is used to the vehicle routing problem with fuzzy travel time.

Reference [9] focuses on modeling and solution of the dynamic VRP with time-dependent and fuzzy travel time. A model of this problem is set up based on fuzzy service times of the customer, its demand and its time windows, which is regarded and ordering as a triangular fuzzy number. A hybrid genetic algorithm, which is seasoned with the model and combined with the ant colony algorithm, is presented.

In [10] the authors present a model of the real world vehicle routing and dispatching problem. The time-dependent and fuzzy travel speeds are introduced into the model. A dispatching period is divided into some time slices and each time slice is designated a triangular fuzzy speed. The method of comparing two triangular fuzzy numbers is applied to check whether or not customers' time windows are satisfied. A hybrid intelligent approach combining a genetic algorithm and an ant colony algorithm is proposed for solving the dispatching model.

Results in both types of models with uncertainty in demand or in time have been published based on fuzzy variables, fuzzy random variables, stochastic programming and Chance Constrained Programming (CCP). These concepts were introduced by Charnes and Cooper and later by Liu, and generally are applied with heuristics to find solutions to the VRP [26], [24].

In [13] the author considers the VRP with time window while assuming that the travel times cannot be precisely known, but can be regarded as fuzzy variables. Since the travel times are fuzzy variables, every customer will be visited at a fuzzy time. Credibility is introduced as a measure of confidence in the constraints so that it ensures that all customers are visited within their time windows with a confidence level, then following Chance Constraint Programming and also a hybrid intelligent algorithm by integrating fuzzy simulation and GA to solve the VRP.

To the best of our knowledge, there is little evidence in the literature on the properties of mathematical programming with random fuzzy coefficients and VRP with random fuzzy demands. The only related study appears to be a paper [8] which proposed a method to solve a class of model with random fuzzy coefficients in both the objective functions and constraint functions and applied it in the CVRP with random fuzzy demands. Based on the concept of random fuzzy variables introduced in [3], the objective is to provide workable formulations and exact algorithms for a class of uncertainty.

Reference [12] deals with a variation of uncertain VRP where the customers' demands are random fuzzy variables and the travel times between customers are random variables. The travel times between customers follow given probability distributions. The authors develop a programming model with random fuzzy and random variables for VRP which consider capacity and arrival time constraints and present a stochastic programming formulation which includes probabilistic constraints and apply a pure genetic algorithm (GA).

In [23] and [25] the authors considered a fuzzy multi-objective modeling approach for capacitated VRP with fuzzy random parameters. The first paper is based on the mixed integer linear programming model, and suggested an interactive heuristics approach using triangular fuzzy number and crisp equivalent formulae to address the demand solution. The authors in the second paper use travel time and customer demand and treat them as fuzzy random variables and chance-constrained programming is presented and converted to a crisp equivalent model under some assumptions. The authors present a hybrid multi-objective particle swarm optimization that incorporates specific heuristics to solve such a problem.

All of the above VRP can be formalized as problems of combinatorial optimization. These problems in their most general form, have integer LP formulations, based on the proposal given by Bodin *et al*. [1], and can be described, assumed that the depot is the node 0, N is the number of

ISBN: 978-989-95079-6-8

customers to be served by K vehicles and the decision variables are $x_{ij}^k \in \{0,1\}$, $i,j = 1, 2, \ldots, N$, $k = 1, 2, \ldots, K$, where $x_{ij}^k = 1$ if vehicle k travels from customer i to j and 0 otherwise.

In particular, objectives and constraints may be formulated as follows.

Objectives

a. If c_{ij}^k is the *cost of travelling* from customer i to customer j by vehicle k. The total travel cost is an objective function to be minimized and is expressed by:

$$\min \sum_{k=1}^{K} \sum_{i=0}^{N} \sum_{j=0}^{N} c_{ij}^k x_{ij}^k \tag{11}$$

In general, this cost can be expressed as total time or distance traveled by the vehicles.

b. If t_{ij}^k denotes the time needed to go from customer i to customer j and δ_i^k is the required time by vehicle k to unload the demand to the customer i. The objective function is to minimize the total travel time.

$$\min \left(\sum_{k=1}^{K} \sum_{i=0}^{N} \sum_{j=0}^{N} t_{ij}^k x_{ij}^k + \sum_{k=1}^{K} \sum_{i=0}^{N} \sum_{j=0}^{N} \delta_i^k x_{ij}^k \right) \tag{12}$$

c. Other possible objective function is to maximize the number of customers served by the vehicles.

$$\max \sum_{k=1}^{K} \sum_{i=0}^{N} \sum_{j=0}^{N} x_{ij}^k \tag{13}$$

Constraints

Constraints (14) and (15) ensure that each customer is served exactly once.

$$\sum_{k=1}^{K} \sum_{i=0}^{N} x_{ij}^k = 1, \quad j = 1,2,...,N \tag{14}$$

$$\sum_{k=1}^{K} \sum_{j=0}^{N} x_{ij}^k = 1, \quad i = 1,2,...,N \tag{15}$$

Constraint (16) ensures route continuity.

$$\sum_{i=0}^{N} x_{it}^k - \sum_{j=0}^{N} x_{tj}^k = 0, \quad t = 1,2,...,N; k = 1,2,...,K \tag{16}$$

Constraint (17) shows that the total length of each route k has a limit D_k, with d_{ij}^k the distance from i to j.

$$\sum_{k=1}^{K} \sum_{i=0}^{N} d_{ij}^k x_{ij}^k \le D_k, \quad k = 1,2,...,K \tag{17}$$

Constraint (18) ensures that the total demand of any route must not exceed the capacity Q_k of the vehicle k, with the demand of customer i, q_i.

$$\sum_{j=0}^{N} q_j \left(\sum_{i=0}^{N} x_{ij}^k \right) \le Q_k, \quad k = 1,2,...,K \tag{18}$$

Constraints (19) and (20) ensure that each vehicle is used no more than once.

$$\sum_{j=0}^{N} x_{0j}^k \le 1, \quad k = 1,2,...,K \tag{19}$$

$$\sum_{i=0}^{N} x_{i0}^k \le 1, \quad k = 1,2,...,K \tag{20}$$

This formalization assumes that the decision-maker has access to specific information on the components that define the problem; that is, on objective functions and constraints. However in real world problems it is more common that the information one has is actually imprecise or incomplete.

Real world situations reveal the difficulties of analyzing different types of models and problems that are affected by incomplete, imprecise or vague information or constraints. These problems, however, can be modeled in terms of fuzzy sets, leading to the field of fuzzy optimization. The solutions to these problems are fuzzy solutions. Discussions concerning solutions do not focus on their feasibility, nor if they are optimal solutions or not. We, in turn, have chosen to discuss the degree of feasibility and optimality of the solution.

Thus, if we suppose that any of the parameters: cost c_{ij}^k, distance d_{ij}^k, times t_{ij}^k and δ_i^k, and demand q_i can be fuzzy, the traditional model becomes a Fuzzy VRP. Intuitively, when any of these quantities are fuzzy numbers, the objective functions become fuzzy as well. If these parameters are approximately known, they can be represented by the fuzzy numbers $\tilde{c}_{ij}^k$, $\tilde{d}_{ij}^k$, $\tilde{t}_{ij}^k$, $\tilde{\delta}_i^k$ and $\tilde{q}_i$ respectively, with their corresponding membership functions. Then, for instance, the objective function can be expressed as:

$$\widetilde{\min} \sum_{k=1}^{K} \sum_{i=0}^{N} \sum_{j=0}^{N} \tilde{c}_{ij}^k x_{ij}^k \tag{21}$$

Similar changes occur with other objective functions. In the same way, constraints (17) would be expressed by:

$$\sum_{k=1}^{K} \sum_{i=0}^{N} \tilde{d}_{ij}^k x_{ij}^k \preceq \tilde{D}_k, \quad k = 1,2,...,K \tag{22}$$

with similar changes to the other constraints. However, constraints (14) to (16) are the same as those found in the crisp model.

Note that quantities Q_k and D_k also must be considered as fuzzy parameters. In addition, the summation symbol Σ in the objective functions and constraints refers to an addition of fuzzy numbers and $\preceq$ a fuzzy relation between fuzzy numbers. The meaning of min operator $\widetilde{\min}$ is also ambiguous, because it depends on the fuzzy numbers ranking index to be used. Hence there is a need to seek appropriate procedures for its solution.

Following the pattern in general Fuzzy Optimization described in a previous section, four different types of problems can be considered. Two of these problems include imprecision/uncertainty in the objective function(s), such as the case with fuzzy goals and the case with fuzzy costs. The remaining two problems consider fuzzy comparison in the constraints and in the coefficients of the technological matrix. In addition a fifth problem, the general fuzzy problem could be studied in which all of the parameters will be subject to fuzzy considerations.

ISBN: 978-989-95079-6-8

In practice, the search for optimal solutions to FVRP can be done with the following approaches. The simplest approach applies procedures for the fuzzification and defuzzification of variables. It transforms the imprecise information in fuzzy parameters and uses procedures that integrate fuzzy arithmetic to obtain fuzzy solutions. The fuzzy solution is then transformed into a crisp one using some known formulation. This approach may also be used for the introduction of sophisticated fuzzy rules in the decision-making processes to improve their quality. Linguistic variables could be used to facilitate the incorporation of "intelligent" procedures as automatic reasoning, adaptive control or automatic learning.

Another common approach consists in applying the Theory of Possibility and Chance Constraint Programming. It treats the fuzzy models, becoming the crisp equivalent, and by adding Chance Constraints changes the model. It assumes that the constraints will hold with at least a possibility α, and the chance is represented by the possibility that the constraints are satisfied.

The previous approaches combined with some metaheuristics provide robust optimization methods to obtain efficient solutions. Although in each planning problem we can establish a set of constraints, many of which can define different planning problems, this set which is defined with parameters and imprecise variables can also be considered the same if the constraints are not necessarily strictly or exactly established. These parameters or imprecise variables could include the times that cannot exceed a specified bound, or the time windows required to visit clients. In other words, it is possible to establish problems with constraints where not all of the constraints need to be satisfied and with the same degree of precision. In this formulization constraints can be called hard, which represent those constraints that must be satisfied exactly, or soft, which are not subject to the same demanding criteria.

5. Conclusions

Fuzzy Vehicle Routing Problems have been analyzed. Strong formulations are available but most algorithmic development focuses on a limited number of prototype problems. We analyze the use of sufficiently flexible and comprehensive fuzzy approaches to address the various imprecision required in practice.

Acknowledgements

The two first authors have been supported by projects TIN2008-06873-C04-01 (70% FEDER) from the Spanish Government and PI042005/044 from the Canary Islands Government. J.L. Verdegay has been supported by Project TIC-02970-JA.

References

[1] B. Bodin et al., 1983. Routing and scheduling of vehicles and crews: The state of the art. *Computers and Operations Research* 10, pp. 63–211. MathSciNet.

[2] J.M. Cadenas, J.L. Verdegay, 1999 *Optimization Models with Imprecise Data*. Ed. Universidad de Murcia. (in spanish).

[3] A. Charnes, W.W. Cooper, 1959. Chance-Constrained Programming. *Management Science*, 6:73-79.

[4] J.F. Cordeau et al., 2007. Vehicle Routing. *Handbook in OR & MS*, Vol. 14. Chapter 6. C. Barnhart and G. Laporte (Eds.), Elsevier.

[5] G.B. Dantzig, R.H. Ramser, 1959. The truck dispatching problem. *Management Science*, 6:80-91.

[6] M. Delgado, J.L. Verdegay and M.A. Vila., 1987. Imprecise costs in mathematical programming problems. *Fuzzy Sets and Systems*, 29, 21-29.

[7] M. Delgado, J.L. Verdegay and M.A. Vila, 1989. A general model for fuzzy linear programming. *Control and Cybernet*, 16 (2), 113–121.

[8] Y. He, J. Xu, 2005 A class of random fuzzy programming model and its application to vehicle routing problem. *World Journal of Modelling and Simulation* Vol. 1. N. 1, pp. 3-11

[9] L. Hong, M. Xu, 2008. A model of MDVRPTW with fuzzy travel time and time-dependent and its solution. In proceeding, *Fifth International Conference on Fuzzy Systems and Knowledge Discovery*. IEEE.

[10] L. Hong, M. Xu,, 2008. Real Vehicle Routing and Dispatching with Dynamic Fuzzy Travel Times. In proceeding, *Second International Conference on Genetic and Evolutionary Computing*. IEEE.

[11] J. Jia, et al., 2008. Genetic algorithm for fuzzy logistics distribution vehicle routing problem. In *proceeding International Conference on Service Operations and Logistics, and Informatics*. IEEE.

[12] B. Liu, 2001. Fuzzy random Chance-Constrained Programming. IEEE *Transactions on Fuzzy systems* Vol. 9, N. 5.

[13] B. Liu, 2002.*Theory and practice of uncertain programming*. Physica-berlag.

[14] C.-S. Liu, M-Y. Lai, 2009. The vehicle routing problem with uncertain demand at nodes. *Transportation Research Part E: Logistics and Transportation Review.* Available on line, 21/01/2009. doi:10.1016/j.tre.2008.11.002

[15] P. Lucić, D. Teodorović, 2003. Vehicle routing problem with uncertain demand at nodes: The Bee System and fuzzy logic Approach. *In Fuzzy Sets Based Heuristics for Optimization*, 67-82 (Editor J.L. Verdegay), Springer – Verlag.

[16] P. Lucić, D. Teodorović, 2007. The fuzzy ant system for the vehicle routing problem when demand at nodes is uncertain. *Journal on Artificial Intelligence Tools (IJAIT)*.Vol, 16 Issue: 5, 751-770.

[17] L. Tang, et al. 2007. Ant Colony Algorithm Based on Information Entropy Theory to Fuzzy Vehicle Routing Problem. *In proceedings ISKE*, series: Advances in Intelligent Systems Research.

[18] D. Teodorovic, S. Kikuchi, 1991. Application of fuzzy sets theory to the saving based vehicle routing algorithm. *Civil Engineering Systems* 8, pp. 87–93.

[19] D. Teodorovíc, G. Pavkovíc, 1996. The fuzzy set theory approach to the vehicle routing problem when demand at nodes is uncertain. *Fuzzy Sets and Systems* 82, 307-317. Elsevier.

[20] P. Toth, D.Vigo, 2002. The vehicle routing problem. *Monographs on Discrete Mathematics and Applications*, Vol. 9. SIAM

[21] J.L. Verdegay, 1982. *Fuzzy mathematical programming* In: Gupta MM, Sanchez E (eds) Fuzzy Information and Decision Processes.

[22] J.L. Verdegay, 1995. Fuzzy optimization: models, methods and perspectives. *6thIFSA-95 World Congress*. p. 39-71.

[23] B. Werners, M. Drave, 2003. Capacitated vehicle routing problem with fuzzy demand. *In Fuzzy Sets Based Heuristics for Optimization*, 317-334 (Editor J.L. Verdegay), Springer – Verlag.

[24] J. Xu, Y. He, 2004. A class of random fuzzy linear programming model and its application to vehicle routing problem. *Tech. report*, Sichuan University.

[25] Y. Yang , J. Xu, 2008. A class of multi-objective vehicle routing optimal model under fuzzy random environment and its application. *World Journal of Modelling and Simulation* Vol. 4 (2008) No. 2, pp. 112-119.

[26] Y. Zheng, and B. Liu, 2004. Fuzzy vehicle routing model with credibility measure and its hybrid intelligent algorithm, *Tech. report*, Tsinghua University.

[27] HJ. Zimmermann, 1976. Description and optimization of fuzzy system. *International Journal of general System*, 2, 209-216.

ISBN: 978-989-95079-6-8

Weighted decisions in a Fuzzy Random Forest

P.P. Bonissone[1] J.M. Cadenas[2] M.C. Garrido[2] R.A. Díaz-Valladares[3] R. Martínez[2]

1. GE Global Research. One Research Circle.
Niskayuna, NY 12309. USA
2. Dept. Ingeniería de la Información y las Comunicaciones. Universidad de Murcia.
Murcia, Spain
3. Dept. Ciencias Computacionales. Universidad de Montemorelos.
Montemorelos, México
Email: bonissone@crd.ge.com,{jcadenas,carmengarrido}@um.es,rdiaz@um.edu.mx,raquel.m.e@alu.um.es

Abstract— *A multi-classifier system - obtained by combining several individual classifiers - usually exhibits a better performance (precision) than any of the original classifiers. In this work we use a multi-classifier based on a forest of randomly generated fuzzy decision trees (Fuzzy Random Forest), and we propose a new method to combine their decisions to obtain the final decision of the forest. The proposed combination is a weighted method based on the concept of local fusion and on the data set Out Of Bag (OOB) error.*

Keywords— combination methods, fuzzy trees, local fusion, multi-classifier, random forest

1 Introduction

We have witnessed a variety of methods and algorithms addressing the classification issue. In the last few years, we have also seen an increase of multi-classifiers based approaches, which have shown to deliver better results than individual classifiers [10].

The multi-classifier Fuzzy Random Forest [2], doesn't aim to obtain the best system multi-classifier. Rather, we will focus on how to start from a multi-classifier system with performance comparable to the best classifiers and extend it to handle and manipulate imperfect information (linguistic labels, missing values, noise in nominal attributes, etc.) maintaining a proper performance. To build the multi-classifier, we follow the methodology of random forest and to incorporate the processing of imperfect data, we construct the random forest using fuzzy trees as base classifiers. Therefore, we try to use the robustness of a tree ensemble, the power of the randomness to increase the diversity of the trees in the forest, and the flexibility of fuzzy logic and fuzzy sets for imperfect data managing.

Usually, the majority vote weighted method has been the most common combination method used to combine the outputs of different classifiers. This combination method is optimal in the case of two classes and classifiers with independent outputs. In addition, although the classifiers are designed independently, it will be unlikely that they produce independent outputs [9]. In [2] we have presented various combination methods to obtain the final decision of the proposed multi-classifier.

In this work, we propose new inference implementations to obtain the decision of the multi-classifier Fuzzy Random Forest and we compare them with two of the best performing implementations in Fuzzy Random Forest: the proposed implementations are weighted implementations based on local fusion concept [3] and we compare them with weighted implementations based in the OOB cases.

In the section 2, we review the major elements that constitute a multi-classifier and some attempts for the incorporation of fuzzy logic in some techniques. In section 3, we provide a brief description of the Fuzzy Random Forest multi-classifier, showing the learning and inference phases. In section 4, we present the inference implementations that are the subject of this work. We present some comparative results of the analyzed implementations in section 5. Finally, we present our conclusions in section 6.

2 Multi-classifiers and Fuzzy Logic

When individual classifiers are combined appropriately, we usually obtain a better performance in terms of classification precision and/or speed to find a better solution. Multi-classifiers are the result of combining several individual classifiers. Multi-classifiers differ among themselves by their diverse characteristics: the number and the type of the individual classifiers; the characteristics of the subsets used by every classifiers of the set; the consideration of the decisions; and the size and the nature of the training sets for the classifiers [9].

Segrera [10] divides the methods for building multi-classifiers in two groups: ensemble and hybrid methods. The first type, such as Bagging and Boosting, induces models that merge classifiers with the same learning algorithm, while introducing modifications in the training data set. The second, type such as Stacking, creates new hybrid learning techniques from different base learning algorithms.

An ensemble uses the predictions of multiple base-classifiers, typically through majority vote or averaged prediction, to produce a final ensemble-based decision. The ensemble-based predictions typically have lower generalization error rates than the ones obtained by a single model. The difference depends on the type of base-classifiers used, ensemble size, and the diversity or correlation between classifiers [1]. Ahn [1] indicates that, over the last few years, three ensemble-voting approaches have received attention by researchers: boosting, bagging and random subspaces.

ISBN: 978-989-95079-6-8

2.1 Random Forest: A Multi-Classifier based on Decision Trees

Breiman [4] further defines a random forest as a classifiers composed by decision trees where every tree h_t has been generated from the set of data training and a vector θ_t of random numbers identically distributed and independent from the vectors θ_1, θ_2,..,θ_{t-1} used to generate the classifiers $h_1, h_2, .., h_{t-1}$. Each tree provides his unitary vote for the majority class given the entry. Examples of random forest are: randomization, Forest-RI and Forest-RC, double-bagging.

Hamza [6] concludes that: Random Forests are significantly better than Bagging, Boosting and a single tree; their error rate is smaller than the best one obtained by other methods; and they are more robust to noise than the other methods. Consequently, random forest is a very good classification method with the following characteristics: it's easy to use; it does not require models, or parameters to select except for the number of predictors to choose at random at each node.

2.2 Fuzzy Logic and Decision Trees

Decision tree techniques have proved to be interpretable, efficient and capable of treating with applications of great scale. However, they are highly unstable when small disturbances are introduced in data learning. For this reason, the fuzzy logic has been incorporated in the decision tree construction techniques.

Fuzzy logic offers an improvement in the disadvantages previously commented on the decision trees due to the elasticity of the fuzzy set's formalism. In [7, 8] we can find approaches in which fuzzy sets and their underlying approximate reasoning capabilities have been successfully combined with decision trees. This combination has preserved the advantages of both components: uncertainty management with the comprehensibility of linguistic variables, and popularity and easy application of decision trees. The resulting trees show an increased immunity to noise, an extended applicability to uncertain or vague contexts, and a support for the comprehensibility of the tree structure, which remains the principal representation of the resultant knowledge.

In the literature, we can find several proposals for building trees of fuzzy information starting with algorithms already known for building traditional trees. Fuzzy CART [7] was one of the first examples of this approach, being based on the CART algorithm. However, most authors have preferred to use the ID3 algorithm to construct trees for recursive partition of the data set of agreement to the values of the selected attribute. To use the ID3 algorithm in the construction of fuzzy trees, we need to develop attribute value space partitioning methods, branching attribute selection method, branching test method to determine the degree to which data follow the branches of a node, and leaf node labeling methods to determine classes [8].

Fuzzy decision trees are constructed in a top-down manner by recursive partitioning the training set into subsets. Some particular features of fuzzy tree learning are: the member-ship degree of examples, the selection of test attributes, the fuzzy tests (to determine the membership degree of the value of an attribute to a fuzzy set), and the stop criteria (besides the classic criteria when the measure of the information is under a specific threshold) [8].

3 Fuzzy Random Forest

Following Breiman's methodology [4], FRF is a multi-classifier that is a forest of randomly-generated fuzzy decision trees. In this section we provide a brief description of the learning and inference phases in FRF.

3.1 Fuzzy Random Forest Learning

To built a Fuzzy Random Forest (FRF) we use the algorithm 1.

FRFLearning (M,Fuzzy Random Forest)
begin
1. Take a random sample of N observations from the data set with replacement of the complete set of M observations.
2. Apply algorithm 2 (construct a fuzzy tree) to the subset of examples.
3. Repeat steps 1 and 2 up to building all fuzzy trees.
end.

Algorithm 1. FRF learning

Each tree in the forest will be a fuzzy tree generated following the guidelines of [8], adapting it where is necessary. So, the following algorithm (Algorithm 2) has been designed so that the trees could be constructed without considering all the attributes to split the nodes. We select the set of attributes as a random subset of the total set of available attributes. Then, we perform a new random selection for each split. As in the original case, some attributes (inclusive the best) might not be considered for each split, but a attribute excluded in one split might be used by other splits in the same tree.

FuzzyDecisionTree(Examples,Fuzzy Tree)
begin
1. Start with examples set of entry, having the weights of the examples (in root node) equal to 1.
2. At any node N still to be expanded, compute the number of examples of each class. The examples are distributed in part or in whole by branches. The distributed amount of each example to a branch is obtained as the product of its current weight and the membership degree to the node.
3. Compute the standard information content.
4. At each node search the set of remaining attributes to split the node.
4.1.Select with any criteria, the candidate attributes set to split the node.
4.2. Compute the standard information content to each child node obtained from from each candidate attribute.
4.3.Select the candidate attribute such that information gain is maximal.
5. Divide N in sub-nodes according to possible outputs of the attribute selected in the previous step.
6. Repeat steps 2-5 to stop criteria is satisfied in all nodes.
end.

Algorithm 2. Fuzzy decision tree learning

With Algorithm 2, we integrate the concept of fuzzy tree within the design philosophy of Breiman's random forest. In this way, we augment the capacity of diversification of random forests with the capacity of approximate reasoning of fuzzy logic.

When we built a Fuzzy Random Forest using the previous algorithm, about $1/3$ of the cases are excluded from each tree

ISBN: 978-989-95079-6-8

in the forest. These cases are called the "out of bag" (OOB); each tree will have a different set of OOB cases. The OOB cases are not used to build the tree and constitute an independent test sample for the tree. We use the set of OOB cases to obtain one of the weighted combination methods proposed in this work.

3.2 Fuzzy Random Forest Inference

In this section we will describe how the inference is carried out using Fuzzy Random Forest. First, we introduce the notation that we will use in the following sections. Then, we define two general strategies to obtain the decision of the forest for a target example. Concrete instances of these strategies are defined in the next section where we present different inference implementations in the Fuzzy Random Forest.
Notations

- T is the trees' number of the forest. We will use the index t to refer to a particular tree in the forest.
- N_t is the number of reached leaf nodes by an example, in the tree t. A characteristic inherent in fuzzy trees is that the classification of an example can derive in two or more leaves due to the overlapping of the fuzzy sets. We will use the index n to refer to a particular leaf in a tree of the forest.
- I is the number of classes that we consider. We will use the index i to refer to a particular class.
- e is an example from the dataset used to built and to infer with the Fuzzy Random Forest.
- $\chi_{t,n}(e)$ is the grade which the example e active the leaf n from t tree.
- $\overline{\omega}_{t,n}$ is a vector with I elements indicating the weights of the I possible classes in the leaf n of tree t, $\overline{\omega}_{t,n}$=$(\omega_{t,n,1}, \omega_{t,n,2}, ..., \omega_{t,n,I})$, where $\omega_{t,n,i} = \frac{E_i}{\sum_{j=1}^{I} E_j}$ and E_j is the sum of the weights of examples with class j in the leaf.

To make the inference with the Fuzzy Random Forest, we will define the fuzzy classifier module. The fuzzy classifier module operate on fuzzy trees of the forest with two possible strategies:

Strategy 1: Combining the information from the different reached leaves in each tree to obtain the decision of each individual tree and then apply the same one or another combination method to generate the global decision of the forest (Fig. 1).

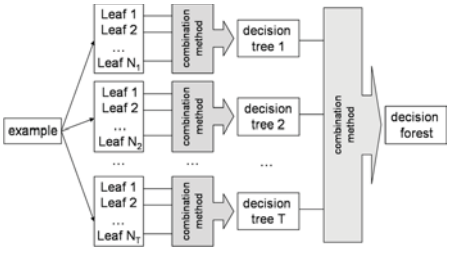


Figure 1: Inference with strategy 1

Strategy 2: Combining the information from all reached leaves from all trees to generate the global decision of the forest (Fig. 2).

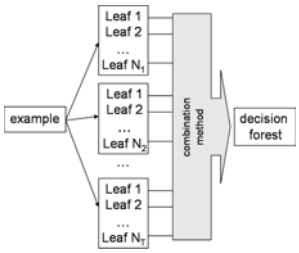


Figure 2: Inference with strategy 2

Moreover, in strategies 1 and 2 we use two functions: $K(\cdot,\cdot,\cdot)$ and $Faggre$. With $K(\cdot,\cdot,\cdot)$ we obtain information about the classes that provides a leaf reached n in a tree. This function depends of $\chi_{t,n}(e)$ and vector $\overline{\omega}_{t,n}$ and returns the weight assigned by the node n to class i. Concrete cases of this function will be defined in the next section. And, $Faggre$ is defined as a frequently used multi-classifiers combination method [9], e.g., majority vote, minimum, maximum, average, product or a weighting combination method defined in the next section.

In Algorithm 3 we implement strategy 1. In this algorithm first, $Faggre$ is used to obtain $A[t][i]$ (the weight of each tree for each class). In this case, $Faggre$ aggregates the information provided by the reached leaves in a tree. Later, the values obtained in each tree t, will be aggregating by means of the function $Faggre$ to obtain the vector $\overline{F}$ that contains the weight proposed by the Fuzzy Random Forest for the different classes. This algorithm takes a target example e and the random forest, and generates the vector $\overline{F}$ as output.

```
Begin
  TreeClasification.
  ForestClasification.
End

TreeClasification (in : e, in : Random Forest, out : A)
Begin
  For each Tree t
    For each Class i
      A[t][i] = Faggre(K(1, χ_{t,1}(e), ω_{t,1}), ...,
        K(1, χ_{t,N_t}(e), ω_{t,N_t}), ..., K(I, χ_{t,1}(e), ω_{t,1}),
        , ..., K(I, χ_{t,N_t}(e), ω_{t,N_t}))
    End For each Class
  End For each Tree
End

ForestClasification (in : A, out : F̄)
Begin
  For each Class i
    F[i] = Faggre(A[1][i], ..., A[T][i])
  End For each Class
End
```

Algorithm 3. FRF Inference (Strategy 1)

To implement strategy 2, the previous algorithm is simplified so that it does not add the information for each tree, but

ISBN: 978-989-95079-6-8

uses directly the information of all leaves reached by the example e in the different trees of the forest. The algorithm 4 implement the strategy 2 and uses as target values the example e to classify and the random forest and provides the vector $\overline{F}$ that contains the weight proposed by fuzzy random forest for the different classes.

ForestClasification ($in : e, in : Random\ Forest$,$out : \overline{F}$)
Begin
  For each Class i
    $F[i] = Faggre(K(i, \chi_{1,1}(e), \omega_{1,1}), ...,$
      $K(i, \chi_{1,N_1}(e), \omega_{1,N_1}), ..., K(i, \chi_{T,1}(e), \omega_{T,1}),$
      $..., K(i, \chi_{T,N_T}(e), \omega_{T,N_T}))$
  End For each Class
End

Algorithm 4. FRF Inference (Strategy 2)

4 Inference Implementations

In the previous section we have shown the general scheme of inference that we use to obtain the forest's final decision. In [2] specific inference instances to FRF are described to both strategies. These specific instances use the combination methods commented in previous sections to implement $Faggre$ function. To implement concrete instances of the strategies we need to define particular examples to the function K previously defined.

Given a reached leaf n in tree t, the particular case of $K(i, \chi_{t,n}(e), \omega_{t,n})$ that we use provides the weight $\chi_{t,n}(e)$ if i is the majority class in this node and 0 for all other classes. This particular case is denoted K1.

In this work we present new inference implementations to FRF based on local fusion concept. Also, we compare the results of these new implementations with the results obtained to the weighted implementations based on the OOB cases. Next we describe these implementations:

• **Weighted Implementations using the OOB cases-** Within this group we define the following implementations:

- *Majority Vote Weighted by Leaf and by Tree applied to strategy 1* (MWLT1): Uses a weight for each tree obtained by testing each individual tree with the OOB data file commented in the previous section. Lets be $\overline{p} = (p_1, p_2, ..., p_T)$ the vector with the weights assigned to each tree. Each p_i is obtained as:

$$\frac{N\_success\_OOB_i}{size\_OOB_i}$$

  where $N\_success\_OOB_i$ is the number of examples inferred correctly from the OOB file used for testing the i th tree and $size\_OOB_i$ is the total number of examples in this file.

  Once the vector $\overline{p}$ is obtained, we used it in the final decision of forest, so that the forest votes for the class c if:

$$c = \max_{i=1}^{I} \sum_{t=1}^{T} p_t \cdot A[t][i]$$

  where $A[t][i]$ is obtained using the majority vote combination method but applied to the values of reached leaves in each tree. Each reached leaf provides values to each class using the $K1$ function:

$$A[t][i] = \begin{cases} 1 & \text{if } i = \max_{j=1}^{I} \sum_{n=1}^{N_t} K1(j, \chi_{t,n}(e), \omega_{t,n}) \\ 0 & \text{other case} \end{cases}$$

- *Majority Vote Weighted by Leaf and by Tree to strategy 2* (MWLT2): Every reached leaf of the forest is considered to vote. In this case, given the example e, the forest votes for the class c if

$$c = \max_{i=1}^{I} \sum_{t=1}^{T} p_t \sum_{n=1}^{N_t} K1(i, \chi_{t,n}(e), \omega_{t,n})$$

• **Weighted Implementations based in local fusion-** Within this group we define the following implementations:

- *Local fusion applied to strategy 1* (FUS1):

  To apply this combination method, first, during the learning of the forest, by each tree generated we obtain an additional tree, that we call error tree.

  The procedure to construct the error tree associated to t-th tree in the forest is the following:

  With the training data set of the t-th tree in the forest (tds tree$_t$ in figure 3) we make a test of the tree. So, we consider the training data set as the test data set. With the results of this test we build a new dataset (tds error_tree$_t$ in the figure 3) with the same data but replacing the attribute class for the binary attribute error. The attribute error indicates whether that example has been classified correctly or not by the t-th tree in the forest (for example, that binary attribute can take the value 0 if the example has been correctly classified by the tree or 1 if it has been incorrectly classified and therefore it is a mistake made by the tree). With this new dataset is constructed a tree to learn the attribute error. This is the error tree.

  In the figure 3, tds tree$_t$ is the training dataset of the t-th tree in the forest and contains examples represented by vectors where:

  - $e_{j,t}$ is the j-th example in the training dataset of the t-th tree in the forest.
  - $class_{j,t}$ is the value of the attribute class to the j-th example. This attribute is the classification aim for the forest.

  tds error_tree$_t$ is the training dataset of the error tree associated to the t-th tree in the forest. Contains vectors represented as:

  - $e_{j,t}$ is the j-th example in the training dataset of the t-th tree in the forest.
  - $error_{j,t}$ is the attribute that acts as class in this dataset. It's a binary attribute with value:
    * 1 if $e_{j,t}$ is incorrectly classified by the t-th tree in the forest.
    * 0 if $e_{j,t}$ is correctly classified by the t-th tree in the forest.

ISBN: 978-989-95079-6-8

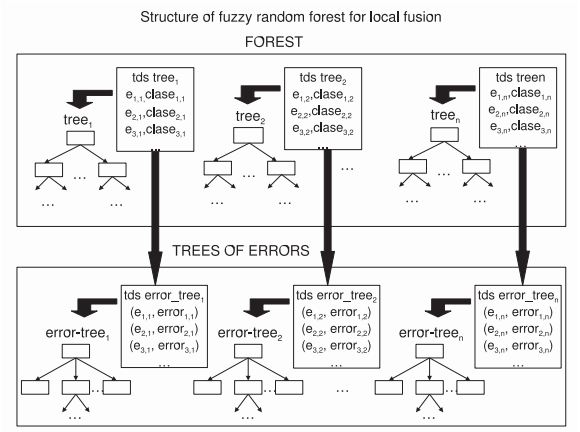


Figure 3: Inference with local fusion

Once the forest and the additional error trees are built, for each example e to classify and each tree of the forest we will obtain a vector $\overline{p}_e = (p_{e,1}, p_{e,2}, ..., p_{e,T})$ with the weights assigned to each tree to this example (local weights). Each $p_{e,t}$ is obtained as:

$$p_{e,t} = \sum_{n=1}^{N_{error_t}} \chi_{error_t,n}\omega_{error_t,n,0}, \quad \forall t = 1, .., T$$

where with the index $error_t$ we refer to the error tree of t-th tree in the forest, N_{error_t} is the number of leaves reached by the example e in the tree of errors $error_t$, $\chi_{error_t,n}$ is the weight of the leaf n of the tree of errors $error_t$ reached by the example and $\omega_{error_t,n,0}$ is the weight of the class 0 (value of binary attribute error=0) in the leaf n of the tree of errors.

The key idea that we want to capture with this implementation is that use of a local weight or a local fusion mechanism [3]. For a given probe (new example), first we evaluate the tree's performance with those examples (in the training set) which are similar to the probe we want to classify. Then, we associate a weight to the decision of that tree on the basis of its performance in the neighborhood of the probe.

Finally, the decision of the forest for the probe e is obtained by weighing the decisions of each tree with the computed weight. Therefore, the forest votes for the class c is:

$$c = \max_{i=1}^{I} \sum_{t=1}^{T} p_{e,t} \cdot A[t][i]$$

where $A[t][i]$ is obtained using the majority vote combination method but applied to the values of reached leaves in each tree. Each reached leaf provides values to each class using the $K1$ function:

$$A[t][i] = \begin{cases} 1 & \text{if } i = \max_{j=1}^{I} \sum_{n=1}^{N_t} K1(j, \chi_{t,n}(e), \omega_{t,n}) \\ 0 & \text{other case} \end{cases}$$

- *Local fusion applied to strategy 2* (FUS2):

Every reached leaf of the forest is considered to vote. In this case, given the example e, the forest votes for the class c if

$$c = \max_{i=1}^{I} \sum_{t=1}^{T} p_{e,t} \sum_{n=1}^{N_t} K1(i, \chi_{t,n}(e), \omega_{t,n})$$

5 Experiments and results

In this section we will make the FRF inference with different databases of the UCI repository and with different percentages of missing values and linguistic labels (to introduce the missing values and linguistic labels we have used the NIP 1.5 tool [5]). The percentage is divided equally between missing and labels values. In table 1 show the results of four inference implementations that we are considering in this work for each database without imperfection, with 5%, 15% and 30% of missing and linguistic labels. For each of them show the average and the standard deviation of 5 x 10 cross validation. We also show the results of the combined outputs of the trees by simple majority vote applied to strategy 1 (SM1) where each tree votes for the majority class of the reached leaves and the forest votes for the majority class among the trees and simple majority vote applied to strategy 2 (SM2) where each reached leaf in the several trees of the forest votes for its majority class and the forest votes for the majority class among the reached leaves.

Analyzing data from Table 1 we can say that the four weighted inference implementations have have similar behavior but the best inference implementation is FUS2, followed by MWLT2, FUS1 and MLWT1. It seems that the weighted inference implementations behave better with the strategy 2 that the strategy 1 and within each one of them behave better those who make use of the weighted with local fusion. We may also note that although generally not weighted strategies (SM1 and SM2) have very good performance with databases without imperfection, weighted strategies have improved the not weighted when we add imperfection in the databases.

6 Summary

In this paper we have defined implementations to combine the outputs of classifiers that composing the multi-classifier FRF. These implementations are based on the combination methods used frequently in the literature but weighting the decision of each classifier by local fusion. The results of these combination implementation is compared with other weighted implementation based on the OOB dataset that obtain a good performance in FRF and implementations based on simple majority vote.

We have presented experimental results obtained by applying these implementations to various databases. The results show that all weighted implementations obtain a similar behavior but FUS2 is the implementation that gets better results. Also the results show that the general strategy 2 provide better results with these implementations than the strategy 1. Furthermore, the weighted strategies improves the non weighted strategies when we add imperfection in the databases.

Acknowledgements

The authors thank the "Ministerio de Educación y Ciencia" of Spain and the European Fund of Regional Development

ISBN: 978-989-95079-6-8

Table 1: Testing accuracies of FRF for different levels of labels and missing

Dataset	Combination methods	original (0% missing and labels)	5% missing and labels	15% missing and labels	30% missing and labels
Ionosphere	FUS2	92.66 (2.15)	92.2 (2.16)	92.08 (1.96)	87.07 (3.29)
	FUS1	92.6 (2.11)	92.26 (2.17)	87.01 (3.36)	80.63 (2.96)
	MWLT2	92.66 (2.15)	92.2 (2.17)	91.51 (2.07)	85.82 (3.09)
	MWLT1	92.66 (2.15)	92.43 (2.17)	84.22 (3.01)	80.0 (2.77)
	SM2	93.0 (1.99)	92.14 (2.13)	85.93 (2.73)	79.83 (2.78)
	SM1	92.83 (2.05)	92.03 (2.15)	84.62 (2.82)	79.37 (2.81)
Wisconsin Breast Cancer	FUS2	96.37 (1.05)	95.94 (1.17)	96.02 (1.0)	94.79 (1.17)
	FUS1	96.34 (1.07)	95.62 (1.3)	95.54 (1.07)	94.22 (1.32)
	MWLT2	96.37 (1.05)	96.02 (1.15)	95.77 (1.04)	94.93 (1.11)
	MWLT1	96.34 (1.07)	95.65 (1.32)	95.34 (1.08)	94.05 (1.31)
	SM2	96.6 (1.03)	95.62 (1.32)	95.31 (1.07)	94.05 (1.31)
	SM1	96.6 (1.05)	95.62 (1.32)	95.31 (1.07)	94.05 (1.31)
Wine	FUS2	95.27 (2.35)	93.59 (2.62)	93.0 (2.75)	87.84 (3.85)
	FUS1	95.05 (2.49)	93.02 (2.81)	91.42 (3.14)	85.15 (4.56)
	MWLT2	95.27 (2.35)	93.47 (2.77)	92.78 (2.85)	87.61 (3.77)
	MWLT1	94.93 (2.57)	92.68 (2.91)	88.95 (3.27)	83.81 (4.43)
	SM2	95.27 (2.35)	92.01 (3.08)	88.39 (3.26)	83.58 (4.2)
	SM1	95.16 (2.45)	91.9 (3.12)	88.16 (3.36)	82.9 (4.5)
Iris Plants	FUS2	94.93 (3.13)	96.4 (2.5)	94.0 (3.2)	85.2 (5.13)
	FUS1	94.93 (3.13)	96.4 (2.5)	93.33 (3.2)	82.67 (5.67)
	MWLT2	94.93 (3.13)	96.4 (2.5)	94.0 (2.95)	85.33 (5.03)
	MWLT1	95.07 (3.16)	96.4 (2.5)	93.33 (3.2)	82.67 (5.71)
	SM2	94.93 (3.08)	96.4 (2.5)	93.33 (3.2)	82.4 (5.7)
	SM1	94.93 (3.08)	96.4 (2.5)	93.33 (3.2)	82.4 (5.7)
Pima Indian Diabetes	FUS2	76.59 (2.31)	73.36 (2.86)	72.11 (2.3)	72.42 (2.66)
	FUS1	76.51 (2.32)	73.02 (2.84)	71.72 (2.26)	72.61 (2.51)
	MWLT2	76.46 (2.34)	73.31 (2.9)	71.95 (2.31)	72.29 (2.83)
	MWLT1	76.66 (2.31)	72.89 (2.81)	71.8 (2.26)	72.16 (2.63)
	SM2	77.26 (2.23)	73.1 (2.58)	72.5 (2.38)	72.48 (2.76)
	SM1	77.0 (2.23)	73.54 (2.69)	72.4 (2.33)	72.5 (2.84)

for the support given to this work under the project TIN2008-06872-C04-03 and the "Fundación Séneca, Agencia de Ciencia y Tecnología de la Región de Murcia (Plan Regional de Ciencia y Tecnología 2007/2010)" which give support under the "Programa de Ayudas a Grupos de Excelencia de la Región de Murcia".

References

[1] H. Ahn, H. Moon, M.J. Fazzari, N. Lim, J.J. Chen and R.L. Kodell. Classification by ensembles from random partitions of high dimensional data. *Computational Statistics and Data Analysis*, 51:6166–6179, 2007.

[2] P.P. Bonissone, J.M. Cadenas, M.C. Garrido and R.A. Díaz. Combination methods in a fuzzy random forest. *Proceedings IEEE International Conference on Systems, Man and Cybernetics (SMC2008)* 1794–1799, 2008.

[3] P.P. Bonissone, F. Xue, and R. Subbur. Fast meta-models for local fusion of multiple predictive models. *Applied Soft Computing Journal*, doi:10.1016/j.asoc.2008.03.06, 2008.

[4] L. Breiman. Random forests. *Machine Learning* 45(1): 5–32, 2001.

[5] J.M. Cadenas, J.V. Carrillo, M.C. Garrido, R. Martínez and E. Muñoz. A tool to handle imperfect information in data sets *URL: http://heurimind.inf.um.es*

[6] M. Hamza and D. Larocque. An empirical comparison of ensemble methods based on classification trees. *Statistical Computati. & Simulation* 75(8):629-643, 2005.

[7] J. Jang. Structure determination in fuzzy modeling: A fuzzy CART approach. *Proceedings IEEE Conference on Fuzzy Systems*, 480–485, 1994.

[8] C.Z. Janikow. Fuzzy decision trees: issues and methods. *Transaction on Systems, Man and Cybernetics, Part B* 28(1): 1–15, 1998.

[9] L.I. Kuncheva. Fuzzy vs non-fuzzy in combining classifiers designed by boosting. *IEEE Trans. on Fuzzy Systems* 11(6): 729–741, 2003.

[10] S. Segrera, M. Moreno. Multiclassifiers: applications, methods and architectures. *Proc. of International Workshop on Practical Applications of Agents and Multiagents Systems, IWPAAMS05*, 263–271, 2005.

ISBN: 978-989-95079-6-8

Evolutionary learning of Quantified Fuzzy Rules for hierarchical grouping of laser sensor data in intelligent control

M. Mucientes, I. Rodríguez-Fdez, A. Bugarín

Intelligent Systems Group
Department of Electronics and Computer Science, University of Santiago de Compostela
Email: manuel.mucientes@usc.es, ismael.rodriguez1@rai.usc.es, alberto.bugarin.diz@usc.es

Abstract— *In complex systems it often occurs that relevant information about the system state and behavior is provided by groups of low-level variables rather than single variables. This grouping into high-level variables introduces a hierachy in the knowledge that can only be captured by means of rules involving propositions with a representation capability that is more complex than usual ones. In this paper we describe a genetic programming based approach for automated learning of Quantified Fuzzy Rules that are capable to deal with such representation capability. An application of this approach for hierarchical grouping of the distance measures provided by the laser sensors of a mobile robot (for the wall-following behaviour) is presented. Experimentation results show the control action is acceptable although no prior knowledge on the variables definition and structure was introduced in the controller.*

Keywords— Evolutionary algorithms, Genetic Programming, Fuzzy Quantification, Intelligent control.

1 Introduction

Modeling of systems using rules overcomes the interpretability issue and is a more adequate approach for scenarios when the model has to cope with noisy data or uncertainty. In this case fuzzy rules are a good choice.

Evolutionary algorithms have some characteristics that make them specially adequate for learning fuzzy rules, such as the flexibility in the representation of solutions through chromosomes, that may range from codifying a complete knowledge base, a single rule, some parameters of the fuzzy sets, ... The well-known combination of fuzzy logic with evolutionary algorithms (genetic fuzzy systems [1]) allows the designer to determine the most appropriate trade-off between accuracy and interpretability for a given system.

However, there are systems that can only be correctly modeled with rule bases that contain rules with different structures, and many times these structures are unknown. In order to learn such knowledge bases, the algorithm must have the ability to represent a set of structures. This ability is provided by genetic programming, where each chromosome of the population is represented as a tree of variable length. A complete flexibility in the structure of the rules is, generally, not desired and some restrictions have to be imposed, as not all the structures are valid. A compact representation of the valid structures of rules can be defined through a context-free grammar.

In systems with a huge number of (low level) input variables it is frequent that the values of the individual variables are not meaningful. In these cases relevant information about the system is only obtained by analyzing sets of low level variables, grouped into high level variables. This provides a hierarchical structure of knowledge that is meaningful for the experts, where the meaning of high level variables cannot be extracted taking neither an average value of the low level variables of the set, nor the maximum or minimum values. Usually this meaning can be directly captured using Quantified Fuzzy Propositions (QFPs) like *"most of the variables in the set take a high value"*.

In this paper we propose a genetic based approach to learn Quantified Fuzzy Rules (QFRs), defined as fuzzy rules of variable structure involving QFPs. The algorithm is based on the genetic programming approach and has been designed to solve regression problems in which the number of input variables is really high, and the model of the system needs the grouping of these variables in several sets (high level variables). In order to illustrate the utility of the algorithm, an application of use for hierarchical grouping of the distance measures provided by the laser sensors in a mobile robot is presented.

The paper is structured as follows: Sec. 2 introduces the motivation of the use of Quantified Fuzzy Rules (QFRs) with variable structure. Sec. 3 presents the QFRs model, while Sec. 4 describes the genetic programming algorithm that has been used to learn the QFRs. Finally, Sec. 5 points out some results, conclusions and future improvements of the algorithm.

2 Motivation for QFRs

Modeling the behavior of a mobile robot that performs a task, autonomously or remotely controlled by a human operator, is a challenging problem if no prior knowledge is provided. In previous papers [2] two successful approaches to learn fuzzy rules for the control of a robot in two different tasks were described. In both cases, the learned rules were conventional fuzzy rules and all the input variables were defined by a human expert. For the wall following task these variables were *right distance*, *left distance*, *orientation*, and *velocity*. For the moving object tracking task the variables were *distance to the object*, *deviation*, *difference in velocity with the object* and *difference in angle*.

The question that comes up is: is it possible to model the behavior of the robot without using any prior knowledge? This means that learning relies only on the input-output data pairs, i.e. on the sensors information and on the control orders that have been given for those situations. Fig. 1 shows how a robot equipped with a laser range finder senses the environment.

Laser range finders emit at the same time beams in different directions. When a beam hits an obstacle, it is reflected and registered by the scanner's receiver. With this information, the distance measured in the direction of each beam can be calculated. Fig. 1 shows in the shadowed area the space covered by the laser when the robot is placed in an typical office environment. The laser range finder provides the distances to the

ISBN: 978-989-95079-6-8

closest obstacle in each direction with a given angular resolution (number of degrees between two consecutive beams).

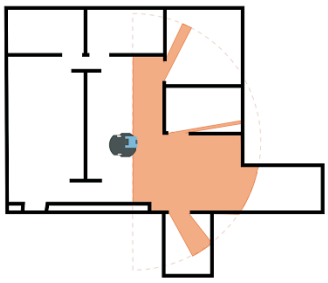


Figure 1: Sensors information for a mobile robot equipped with a laser range finder.

A robot equipped with a single laser range finder provides 361 distance measurements in each acquisition for an angular resolution of 0.5 degrees. When equipped with two laser range finders (to get information of the whole surrounding environment) it provides 722 distance input variables. As the output variables the controller must provide the linear and angular velocities. This control action cannot be decided by simply analyzing the individual distance values of each beam, since noisy measurements, gaps between objects (very frequent in cluttered environments) may occur. Usually the relevant information of the environment is modeled (or extracted) with propositions involving a more complex structure in the variables. For instance, the input variables can be arranged taking into account the importance of the information they provide to model the system.

In general, two categories can be established:

- High-level input variables: variables that provide, by themselves, information relevant and meaningful to the expert for modelling the system (e.g. the linear velocity of the robot).
- Low-level input variables: variables for which their individual values do not contribute to model the system. Their importance stems from the analysis of sets of these variables (e.g. the distance measured by a single beam in a laser range finder).

However, the values of a group of low level variables can provide valuable and meaningful information. For example, the "'frontal sector"' of a laser range finder is a high-level variable made up of a set of distances of single beams (low-level variables). Within this context Quantified Fuzzy Propositions (QFPs) such as *"some of the distances of the frontal sector are low"* are useful for representing relevant knowledge for the experts and therefore for performing intelligent control. QFPs require the definition of several elements:

- *Some*: how many distances of the frontal sector must be low?
- *frontal sector*: which beams belong to the frontal sector?
- *low*: what is the meaning of low?

This example clearly sets out the need to use propositions different from the conventional ones. In this paper QFPs (as "X is A in Q of S") are used for representing knowledge about high-level variables that are defined as the grouping of low-level variables. Conventional ("X is A") are used for other high-level variables, non related to low-level ones.

3 Quantified Fuzzy Rules (QFRs) model

An example of QFR is shown in Fig. 2, and involves both QFPs 1 and conventional ones 2 :

IF $d(h)$ *is* $HIGH$ *in most of* F^1_{sector} and (1)
...
velocity is F_{vel} (2)
THEN *linear velocity is* F_{lv} and *angular velocity is* F_{av}

Figure 2: A typical QFR to model the behavior of a mobile robot.

The general expression for QTPs in our case is:

$$d(h) \text{ is } F^i_d \text{ in } Q^i \text{ of } F^i_{sector} \quad (3)$$

where, for each i=1,...,N_d (N_d being the number of analyzed sectors of distances:

- $d(h)$ is the signal. In this particular case, it represents the distance measured by beam h.
- F^i_d is a linguistic value for variable $d(h)$.
- Q^i is a (spatial, defined in the laser beam domain) fuzzy quantifier.
- F^i_{sector} is a fuzzy set in the laser beam domain (e.g., the "frontal sector").

Evaluation of the Degree of Fulfillment (DOF) for 3 is carried out using Zadeh's quantification model for proportional quantifiers (such as "most of", "part of", ...) [3], that allows to consider non-persistence, partial persistence and total persistence situations for event $d(h)$ is F^i_d in the range of laser beams (spatial interval F^i_{sector}). This is a relevant characteristic of this model, since it allows to consider partial, single or total fulfillment of an event within the laser beams set.

Automatic learning of QFRs for this application (Fig. 2) demands an algorithm with the ability to represent rules with different structures, as the number N_d of analyzed sectors of distances, and therefore the number of QFPs per rule (Fig. 2) can change among rules. Moreover, the number of parameters involved in the definition of a proposition is higher in QFP than in conventional propositions (3 vs. 1).

4 Genetic programming algorithm

The proposed evolutionary algorithm is a genetic programming algorithm based on the genetic cooperative-competitive learning (GCCL) approach. In genetic programming, an individual is a tree of variable length. Each individual in the population can have a different structure, and the introduction of restrictions in that structure of the chromosome can be solved using, for example, a context-free grammar. As the number of beams sectors is unknown and can change among rules, the

ISBN: 978-989-95079-6-8

flexibility in the structure of the rules can be managed by the genetic programming algorithm.

In the GCCL approach [4, 5] rules evolve together but competing among them to obtain the highest fitness. In this approach it is fundamental to include a mechanism to maintain the diversity of the population (niche induction). The mechanism must warrant that there is competition among individuals of the same niche, but also has to avoid the deletion of those weak individuals that occupy a niche not covered by other individuals of the population.

We have chosen token competition [6] as the mechanism for maintaining the diversity. According to [7], this mechanism is adequate for genetic programming, as in this kind of evolutionary algorithms the structure of the individuals can be completely different and, thus, the evaluation of the similarities is hard. The advantage of token competition over other approaches, like crowding or fitness sharing, is that it is not necessary to estimate the similarities between pairs of individuals.

The learning process is based on a set of training examples. Each example e^l is represented by a tuple:

$$e^l = (d(1), \ldots, d(N_{LB}), \mathit{velocity}, \mathit{vlin}, \mathit{vang}) \quad (4)$$

where $d(h)$ is the distance measured by beam h, *velocity* is the velocity of the robot, and *vlin* and *vang* are the output variables linear and angular velocity. In token competition, each example of the training set has a token and, of all the individuals that cover this example, the token will be seized by the individual with the highest raw fitness. In this way, the individual with the highest strength in the niche will exploit it, while individuals that are weaker will reduce its strength as they cannot compete with the best individual in the niche.

4.1 Description of the context-free grammar

As we pointed out before, in genetic programming each individual is a tree of variable size. Thus, the structure of the individuals can be very different among them. In order to generate valid individuals of the population, and to produce right structures for the individuals after crossover and mutation, some restrictions have to be applied. With a context-free grammar all the valid structures of a tree (chromosome) in the population can be defined in a compact form. A context-free grammar is a quadruple (V, Σ, P, S), where V is a finite set of variables, Σ is a finite set of terminal symbols, P is a finite set of rules or productions, and S is an element of V called the start variable.

The grammar is described in Fig. 3. The first item enumerates the variables, then the terminal symbols, in third place the start variable is defined, and finally the rules for each variable are enumerated. When a variable has more than one rule, rules are separated by symbol |. Fig. 4 represents a typical chromosome generated with this context-free grammar. Terminal symbols (leafs of the tree) are represented by circles, and variables are shown as flatted circles. Terminal symbols F_{vel}, F_d, F_{lv}, F_{av}, Q correspond with the linguistic labels and the quantifier defined in Fig. 2. Moreover, F_{sector} has been codified with *initialB* and *finalB*, which represent the extreme values of a trapezium.

A description of the evolutionary algorithm is shown in Fig. 5. It is based on the GCCL approach with a population of variable size. The following sections describe each of the stages of the algorithm.

- V = { rule, antecedent, consequent, distances }
- Σ = { F_{lv}, F_{av}, F_{vel}, F_d, *initialB*, *finalB*, Q }
- S = rule
- Productions:
  - rule $\longrightarrow$ antecedent consequent
  - antecedent $\longrightarrow$ distances F_{vel} | distances
  - consequent $\longrightarrow$ F_{lv} F_{av}
  - distances $\longrightarrow$ F_d *initialB* *finalB* Q distances | F_d *initialB* *finalB* Q

Figure 3: Context-free grammar.

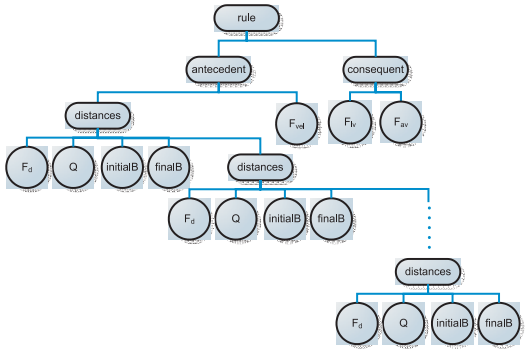


Figure 4: A chromosome representing a QFR to model the behavior of a robot.

1. Initialization
2. for iteration = 1 to *maxIterations*
   (a) Selection
   (b) Crossover and mutation
   (c) Evaluation
   (d) Resize population

Figure 5: Evolutionary algorithm.

4.2 Initialization

The first step of the algorithm is the initialization of the population: a chromosome (Fig. 4) is generated for each of the first $pop_{maxSize}$ examples in the training set. All the fuzzy sets of conventional fuzzy propositions (F_{vel}) and those of the consequent part (F_{lv} and F_{av}) are initialized with a triangular membership function centered in the corresponding example value, and with the extremes of the triangle at a distance equal to $prec_j$ from the center of the triangle. $prec_j$ represents the lowest meaningful change in the value of variable j.

The initialization of F_d, *initialB*, *finalB* and Q is quite more complex. Firstly, the number of QFPs that should be included in each rule is unknown and may be different for each example. Therefore, starting with the first beam, the average value

ISBN: 978-989-95079-6-8

of the distance is calculated incorporating consecutive beams until the variance of the distances is over a threshold. The process is repeated until all the beams have been asigned to a sector. Those sectors with a low number of beams are discarded. For each of the other sectors, *initialB* and *finalB* are directly defined using the information of the sector. F_d is generated in the same way as the fuzzy sets of the conventional propositions, but the triangular membership function is centered in the average value of the distances of the beams of the sector. Finally, Q is calculated as the percentage of beams of the sector that fulfill F_d.

4.3 Selection

The first stage of the iterative part of the algorithm is the selection of the individuals of the population. We have used a binary tournament selection. In a k-tournament selection, k individuals are randomly picked from the population with replacement, and the best of them is selected. In this case, $k = 2$ (binary tournament selection).

4.4 Evaluation

In order to estimate the fitness of an individual of the population, we first need to know the probability that an example e^i matches the output C_j:

$$P\left(C_j|e^i\right) = \exp\left(-\frac{error_j^i}{ME}\right) \tag{5}$$

where $error_j^i$ is the error due to the selection of an output different from the one codified in the example, and *ME* is a parameter that defines the meaningful error for the application. $error_j^i$ can be defined as:

$$error_j^i = \left(y_i^i - y_j^i\right)^2 \tag{6}$$

where y_i^i is the state of the system reached when the output of example e^i is applied to the state defined by e^i, and y_j^i is the state of the system reached when output C_j is applied to the state defined by e^i. As we are working with regression problems, an example can have several outputs that are different from the one codified in the example, but that make small errors, i.e., that are very similar to the desired output.

The accuracy of an individual of the population can be described as:

$$confidence = \frac{GC_{ex}}{covered_{ex}} \tag{7}$$

where GC_{ex} is the number of covered examples with $P\left(C_j|e^i\right)$ over a threshold P_{min}. The ability of generalization of a rule is calculated as:

$$support = \frac{GC_{ex}}{GC_{ex} + GU_{ex}} \tag{8}$$

where GU_{ex} is the number of uncovered examples with $P\left(C_j|e^i\right) > P_{min}$. Finally, we can define $fitness_{raw}$ as the combination of both values:

$$fitness_{raw} = \alpha_f \cdot confidence + (1 - \alpha_f) \cdot support \tag{9}$$

which represents the strength of an individual without taking into account the others. $\alpha_f \in [0,\ 1]$ is a parameter that codifies the trade-off between accuracy and generalization.

As we are using the GCCL approach, $fitness_{raw}$ will be used to decide which individual seizes each example. Thus, the fitness of an individual is defined as:

$$fitness = fitness_{raw} \cdot \frac{seized_{ex}}{covered_{ex}} \tag{10}$$

where $seized_{ex}$ is the number of examples seized by the individual, while $covered_{ex}$ is the number of examples that have been covered by it.

4.5 Crossover and mutation

Once pop_{size} individuals have been selected, each couple of them is crossed with probability p_c. The crossover operator is the parent-centric BLX (PCBLX) [8, 9]. This crossover operator is valid for genes representing real numbers. However, in the defined chromosomes (Fig. 4), there are also genes representing trapezoids. Given two trapezoids represented by tuples, $(\kappa_1\ \ldots\ \kappa_4)$ and $(\rho_1\ \ldots\ \rho_4)$ ($\kappa_j, \rho_j \in [a_j, b_j]$, $j = 1, \ldots, 4$), that are going to be crossed, the following offspring are generated:

- $(\kappa'_1\ \ldots\ \kappa'_4)$, where κ'_j is randomly selected from the interval $\left[l_j^\kappa, r_j^\kappa\right]$, with $l_j^\kappa = \max\left\{a'_j, \kappa_j - I_j\right\}$, $r_j^\kappa = \min\left\{b'_j, \kappa_j + I_j\right\}$, $I_j = |\kappa_j - \rho_j| \cdot \alpha$, $\alpha \in [0, 1]$, $a'_j = a_j$ if $j \in \{1,\ 2\}$ or $a'_j = \kappa_2$ if $j \in \{3,\ 4\}$, and $b'_j = \kappa_3$ if $j \in \{1,\ 2\}$ or $b'_j = b_j$ if $j \in \{3,\ 4\}$.
- $(\rho'_1\ \ldots\ \rho'_4)$, where ρ'_j is randomly selected from the interval $\left[l_j^\rho, r_j^\rho\right]$, with $l_j^\rho = \max\left\{a'_j, y_j - I_j\right\}$ and $r_j^\rho = \min\left\{b'_j, y_j + I_j\right\}$. Finally, $a'_j = a_j$ if $j \in \{1,\ 2\}$ or $a'_j = \rho_2$ if $j \in \{3,\ 4\}$, and $b'_j = \rho_3$ if $j \in \{1,\ 2\}$ or $b'_j = b_j$ if $j \in \{3,\ 4\}$.

When two individuals are crossed, the PCBLX operator is applied to all the genes of type terminal symbol in the chromosomes. However, as the chromosomes are trees of variable structure and/or size, for a gene of the first individual, which gene should be selected in the second of the individuals? For gene F_{vel} the choice is simple: if this gene exists in both individuals crossover is performed, otherwise not. Nevertheless, for genes of type *distances* (as there are many possible choices), the gene selected in the second individual is the one that has a higher overlap with the gene of the first individual in the definition of the sector. If there is not overlap, the gene is not crossed. For genes of type *distances*, PCBLX is applied to F_d, *initialB*, *finalB*, and Q with their corresponding counterparts in the other chromosome.

After crossover, it could happen that the nodes of type *distances* are unsorted, or that two nodes overlap (they should be merged). The algorithm to repair the chromosome has the following steps:

1. Sort all the sectors taking into account *initialB* and insert them on a list.
2. While the list is not empty.
   (a) Choose the first element of the list and check if it overlaps with other sectors.
      i. If there is overlap, then merge the sectors and sort the list.

ISBN: 978-989-95079-6-8

ii. If there is not overlap, delete the sector from the list and add it to the new tree.

When crossover is not performed, both individuals are mutated. Mutation can be implemented to generate a more general or a more specific rule. The higher the value of $seized_{ex}/covered_{ex}$, the higher the probability to generalize the rule by mutation. This occurs with rules that are very strong in their niche and that could be modified to seize examples in other niches. On the contrary, when the number of seized examples in very low, this means that the rule is weak in its niche, and in order to improve its performance some examples that are currently covered should be discarded.

In order to select the examples that are going to be covered (generalization) or uncovered (specialization), the value of $P\left(C_j|e^i\right)$ (Eq. 5) for each example is considered, where C_j is the consequent of the rule. For generalization, the best C_{ex} uncovered examples are selected while, for specialization, the worst C_{ex} covered ones are chosen. Then, from the selected set of examples, S_{ex} are randomly selected to modify the rule. The selection probability of each example is directly (generalization) or inversely (specialization) proportional to $P\left(C_j|e^i\right)$. For each selected example, each proposition of the antecedent part of the rule is analyzed to decide its mutation. For generalization, if the DOF of the proposition is null, then the proposition is mutated. The opposite occurs for specialization. The mutation of a proposition for generalization can be done with the following mutation operators:

1. Delete the proposition.
2. Merge two adjacent distance sectors. The new F_d value will be obtained applying PCBLX operator to the F_d values of the old sectors. This mutation operator can only be applied for propositions of type distances.
3. Modify the fuzzy label to cover the new example.
4. Modify the quantifier to cover the new example (only for distance propositions).

On the other hand, the mutation operators for specialization are:

1. Insert a *distances* sector (only valid for distance propositions). The insertion position is randomly selected. The values for F_d and Q are generated mutating the values of the previous or next sector. Also, *initialB* and *finalB* are randomly selected using the definitions of the contiguous sectors.
2. Divide a distances sector (only valid for distance propositions). One of the new sectors will have the same F_d and Q values, while the other sector will be generated applying non-uniform mutation to the original values.
3. Modify the fuzzy label to uncover the selected example.
4. Modify the quantifier to uncover the selected example (only for distance propositions).

4.6 Resize population

Those individuals with null fitness are removed from the population, and the best $pop_{maxSize}$ individuals are selected for the final population. If $pop_{size} < pop_{maxSize}$, and there are still uncovered examples, then new individuals are added. These individuals are chosen from the *examples population*, randomly selecting those rules that cover examples that have not been seized yet by the individuals of the population.

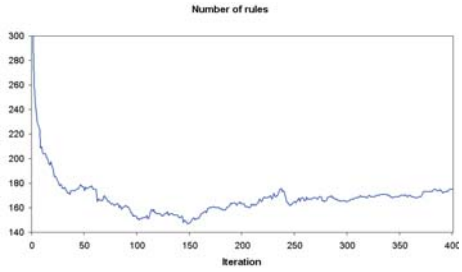
4.7 Population evaluation

The performance of the knowledge base obtained after each iteration has been measured as:

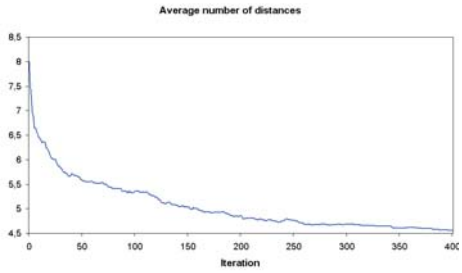
$$fitness_{pop} = \frac{n_{ex}}{\sum_l error_{pop^l}} \tag{11}$$

where n_{ex} is the total number of examples, and $error_{pop^l}$ is the error of the knowledge base when the input is example l.

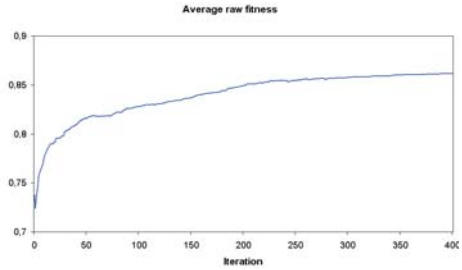
5 Results and discussion



(a) Number of rules.



(b) Average number of distances.



(c) Average $fitness_{raw}$.

Figure 6: Evolution of several variables for an execution of the algorithm.

The learning algorithm has been tested with a set of 795 examples recorded when the robot was following the wall in an environment. The values that have been used for the parameters of the evolutionary algorithm are: $maxIterations = 400$,

ISBN: 978-989-95079-6-8

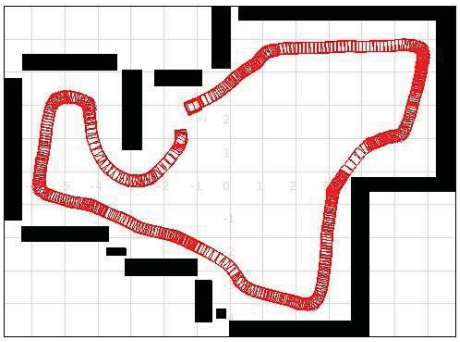


Figure 7: Path of the robot along the test environment.

$pop_{maxSize} = 300$, $p_c = 0.5$, $\alpha = 0.5$ (PCBLX crossover), and $P_{min} = 0.4$.

Fig. 6 shows the evolution of several variables for an execution of the genetic programming algorithm. These variables are the number of rules in the population after each iteration, the average (over all the individuals of the population) number of distance variables in the antecedent part of the rules, and the average $fitness_{raw}$. As can be seen from Fig. 6(a), until iteration 150 the number of rules decreases to less than 150. As the rules in the initial population where very specific (covering sometimes only one example) the algorithm finds rules with a good value of the confidence and with a higher value of the support. In the second stage, as there are rules with a high support, new rules with a higher confidence (and a lower support) seize part of the examples of the weaker rules, thus generating a higher number of rules. At the last iteration of the algorithm, the population had 175 rules. The improvement in the quality of the rules of the population is reflected by the evolution of the average raw fitness (eq. 9) of the rules (Fig. 6(c)).

Finally, the evolution in the average number of distances in the antecedent part of the rules (Fig. 6(b)) reflects the ability of the algorithm to extract valuable information about the distance sectors that are meaningful to model the system. From an initial value of more than eight distance sectors (in average), the algorithm obtains rules that, in average, use around 4.5 distance variables.

Fig. 7 shows the path of the robot along the environment that was used to test the learned controller. In this environment, the robot was following the right-hand wall. The higher the concentration of marks of the robot, the lower the linear velocity. In order to evaluate the quality of the controller we have measured four different indicators: the right distance, the linear velocity, the change in the linear velocity between two consecutive cycles —which reflects the smoothness in the control—, and the time. The average values of the indicators are calculated for each lap that the robot performs in the environment. Results presented in table 1 are the average and standard deviation values over five laps of the average values of the indicators over one lap.

Results show a distance over the expected one (0.5). However, this is in part because the distance has been measured with the information coming from only one beam (in [2, 10] the distance was measured using the input variable defined by an expert). Moreover, the velocity is low. Nevertheless, the smoothness of the control is very good.

Table 1: Results ($x \pm \sigma$) for the environment in Fig. 7

Distance (m)	0.8137 ± 0.0091
Velocity (m/s)	0.13274 ± 0.00095
Velocity change (m/s)	0.00486 ± 0.00053
Time (s)	259.26 ± 2.20

These preliminary results are promising and show an acceptable control, although no expert knowledge was introduced for the definition of the variables. Moreover, as the system has been trained with data of a unique environment, further analysis on the ability of the algorithm to generalize the extracted model of the behavior of the robot will be conducted in the future.

Acknowledgment

This work was supported by the Spanish Ministry of Science and Innovation under grant TIN2008-00040. M. Mucientes is supported by the *Ramón y Cajal* program of the Spanish Ministry of Science and Innovation.

References

[1] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena. *Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases*, volume 19 of *Advances in Fuzzy Systems - Applications and Theory*. World Scientific, 2001.

[2] M. Mucientes and J. Casillas. Quick design of behaviors in mobile robotics by fuzzy controllers with good interpretability. *IEEE Transactions on Fuzzy Systems*, 15(4):636–651, 2007.

[3] L.A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computers and Mathematics with Applications*, 9:149–184, 1983.

[4] A. Giornada and F. Neri. Search-intensive concept induction. *Evolutionary Computation*, 3(4):375–416, 1995.

[5] D.P. Greene and S.F. Smith. Competition-based induction of decision models from examples. *Machine Learning*, 3:229–257, 1993.

[6] K.S. Leung, Y. Leung, L. So, and K.F. Yam. Rule learning in expert systems using genetic algorithm: 1, concepts. In *Proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks*, pages 201–204, Iizuka (Japan), 1992.

[7] F.J. Berlanga, M.J. del Jesus, and F. Herrera. Learning compact fuzzy rule-based classification systems with genetic programming. In *Proceedings of the 4th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)*, pages 1027–1032, Barcelona (Spain), 2005.

[8] F. Herrera, M. Lozano, and A.M. Sánchez. A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study. *International Journal of Intelligent Systems*, 18:309–338, 2003.

[9] M. Lozano, F. Herrera, N. Krasnogor, and D. Molina. Real-coded memetic algorithms with crossover hill-climbing. *Evolutionary Computation*, 12:273–302, 2004.

[10] M. Mucientes, R. Alcalá, J. Alcalá-Fdez, and J. Casillas. Learning weighted linguistic rules to control an autonomous robot. *International Journal of Intelligent Systems*, 24:226–251, 2009.

ISBN: 978-989-95079-6-8

Graded Equipollence of Fuzzy Sets

Michal Holčapek[1], Matěj Turčan[2]

1. Institute for Research and Applications of Fuzzy Modeling, University of Ostrava,
30. dubna 22, 701 03 Ostrava 1, Czech Republic
2. Faculty of Economics, VŠB-Technical University of Ostrava
Sokolská 33, 701 21 Ostrava 1, Czech Republic
Email: michal.holcapek@osu.cz, matej.turcan@vsb.cz

Abstract— *The aim of this contribution is to propose an approach to comparing fuzzy sets from the cardinality point of view. The idea is based on a transformation of one fuzzy set to another one in some degree, where one-to-one mappings between suitable sets are used. To ensure the correctness of degrees in which there exist correspondences between fuzzy sets we introduce a class of sets and consider fuzzy sets and fuzzy classes with universes of discourse from this universe. We introduce a degree in which two fuzzy sets are equipollent and show that these degrees define a similarity relation on the class of all fuzzy sets. Some simplifications of the original definition of the concept of "being equipollent" are proposed and several properties are demonstrated.*

Keywords— fuzzy sets, equipollence of fuzzy sets, cardinal theory of fuzzy sets.

1 Introduction

The classical concept of cardinality of a set is a measure of the "number of elements of the set". A formal definition without knowing anything about numbers is based on the notion of one-to-one correspondence between sets. In particular, we say that two sets x and y are *equipollent* (*equipotent, equivalent, bijective*, or have *the same cardinality* etc.) and write $|x| = |y|$, if there exists a one-to-one mapping of x onto y. This definition gives rise to equivalence relation defined on the class of all sets. This equivalence is called *equipollence* (or also *equipotence, equinumerosity* etc.) and using it the concept of cardinality is introduced. For instance, if we work with the Zermelo-Fraenkel axiomatic set theory with the axiom of choice, the cardinality (or the cardinal) of a set x is defined as the least ordinal number in the class of all sets equipollent with x.

The concept of equipollent fuzzy sets is not new in fuzzy set theory. Some classical-like approaches to the concept of equipollence of fuzzy sets are presented by D. Klaua [1] or S. Gottwald [2, 3]. In the second case, a degree in which two fuzzy sets are equipollent is defined using fuzzy uniqueness of fuzzy mappings. Equivalence classes (fuzzy sets) are *fuzzy cardinals*. Note that fuzzy mappings are here defined as special fuzzy relations.

Another approach was suggested by M. Wygralak [4], where the concept of equipollence of fuzzy sets is introduced using the cardinality of α-cuts of fuzzy sets. More precisely, two fuzzy sets A and B are equipollent, if

$$\bigvee\{\alpha \in [0,1] \mid |A_\alpha| \geq i\} = \bigvee\{\alpha \in [0,1] \mid |B_\alpha| \geq i\} \quad (1)$$

for any cardinal number i. Some modifications of (1) can be found in [5, 6, 7].

In this contribution, we propose a new approach to the equipollence of fuzzy set over a universe of sets (e.g., a universe of all finite or countable sets, or all sets, or a Grothendieck universe). Analogously to the Gottwald's approach, a graded equipollence is considered where, however, the degrees of being equipollent are obtained by a simpler way than in the mentioned approach. More precisely, our approach is based on graded one-to-one correspondences between fuzzy sets where the crisp mappings between universe are considered. Such equipollence, contrary to the Gottwald's definition, can be used in practice, e.g., for a comparison of a number of well satisfied criteria for different applicants. Note that our approach is, in a certain sense, equivalent to the Wygralak's approach, if we restrict ourselves to finite fuzzy sets.

2 Preliminaries

2.1 Structures of membership degrees of fuzzy sets

In this contribution the membership degrees of fuzzy sets will be interpreted in a complete residuated lattice.

Definition 2.1. An algebra $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, \bot, \top\rangle$ is a *residuated lattice*, if $\langle L, \wedge, \vee, \bot, \top\rangle$ is a bounded lattice, where $\bot$ and $\top$ are the least and greatest elements of L, respectively, $\langle L, \otimes, \top\rangle$ is a commutative monoid and

$$\alpha \leq \beta \rightarrow \gamma \quad \text{if and only if} \quad \alpha \otimes \beta \leq \gamma, \qquad (2)$$

holds for arbitrary $\alpha, \beta, \gamma \in L$ ($\leq$ denotes the corresponding lattice ordering).

The operations $\otimes$ and $\rightarrow$ are called a *multiplication* and *residuum*, respectively. A residuated lattice is *complete* or *linearly ordered*, if $\langle L, \wedge, \vee, \bot, \top\rangle$ is a complete or linearly ordered lattice, respectively.

Example 2.1. Let T be a left continuous t-norm (see [8]). Then

$$\mathbf{L} = \langle [0,1], \min, \max, T, \rightarrow_T, 0, 1\rangle,$$

where $\alpha \rightarrow_T \beta = \bigvee\{\gamma \in [0,1] \mid T(\alpha, \gamma) \leq \beta\}$, is a complete linearly ordered residuated lattice. One of the most important residuated lattices on the unit interval is *Łukasiewicz algebra* that is determined by $T_{\mathrm{L}}(a,b) = \max(0, a + b - 1)$. The residuum is then given by $a \rightarrow_{T_{\mathrm{L}}} b = \min(1 - a + b, 1)$.

ISBN: 978-989-95079-6-8

Example 2.2. Let us consider the lattice $L = \{\bot, a, b, \top\}$ with two incomparable elements a and b. Put $x \otimes y = x \wedge y$ and $x \to y = \bigvee\{z \mid x \wedge z \leq y\}$ (see Table 1). Then

$$\mathbf{L} = \langle\{\bot, a, b, \top\}, \wedge, \vee, \to, \bot, \top\rangle$$

is a non-linearly ordered residuated lattice. Note that this is an example of a lattice called a *Heyting algebra*.

The *biresiduum* in $\mathbf{L}$ is a binary operation $\leftrightarrow$ on L defined by

$$\alpha \leftrightarrow \beta = (\alpha \to \beta) \wedge (\beta \to \alpha). \tag{3}$$

Example 2.3. One checks easily that the biresiduum for the Łukasiewicz algebra is given by $a \leftrightarrow_{\mathrm{L}} b = 1 - |a - b|$. The definition of biresiduum for the residuated lattice from Example 2.2 is given in Table 1.

$\to$	$\bot$	a	b	$\top$
$\bot$	$\top$	$\top$	$\top$	$\top$
a	b	$\top$	b	$\top$
b	a	a	$\top$	$\top$
$\top$	$\bot$	a	b	$\top$

$\leftrightarrow$	$\bot$	a	b	$\top$
$\bot$	$\top$	b	a	$\bot$
a	b	$\top$	$\bot$	a
b	a	$\bot$	$\top$	b
$\top$	$\bot$	a	b	$\top$

Table 1: Definitions of $\to$ and $\leftrightarrow$ in Example 2.2 and 2.3

2.2 Several notions from the set theory

In this contribution we work with Zermelo-Fraenkel axiomatic set theory with the axiom of choice (ZFC). We shall use small latin letters as $x, y, z, \ldots$ to denote common sets and capital latin letters as $A, B, R, \ldots$ to denote special sets (e.g., special mappings and relations). Although, the objects of ZFC are sets, we consider the informal notion of a class. This is useful for description of collections of sets that are defined by formulas of the language of set theory, i.e., if $\varphi(x, p_1, \ldots, p_n)$ is a formula, we call

$$\{x \mid \varphi(x, p_1, \ldots, p_n)\} \tag{4}$$

a *class*. We shall use calligraphic capital letters as $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \ldots$ to denote classes. Every set can be consider as a class, i.e., if y is a set, then it is sufficient to consider the class $\{x \mid x \in y\}$. A class that is not a set is a *proper class*.

We shall use $\emptyset$ to denote the empty set and $\mathrm{P}(x)$ to denote the power set. Further, we shall use the standard symbols $\cap, \cup, \sqcup, \times, \setminus$ etc. to denote the intersection, union, disjoint union, product, difference of sets etc. (and analogously for classes), respectively. The common symbols $\subseteq$ and $\subset$ denote the inclusion relations of sets (and analogously for classes). If $R \subseteq x \times y$ (and analogously for $\mathcal{R} \subseteq \mathcal{X} \times \mathcal{Y}$) is a relation, then we shall use

$$\mathrm{Dom}(R) = \{t \mid t \in x \,\&\, \exists r \in y : (t, r) \in R\},$$
$$\mathrm{Ran}(R) = \{r \mid r \in y \,\&\, \exists t \in x : (t, r) \in R\}$$

to denote the domain and range of R, respectively. If f is a mapping between sets (analogously for classes), then $\mathrm{Dom}(f)$ and $\mathrm{Ran}(f)$ denote the domain and range of f, respectively. We shall use $|x|$ to denote the cardinality of a set x.

3 Fuzzy sets and classes in universes of sets

Let us suppose that a complete residuated lattice $\mathbf{L}$ is given. The following concept of a universe of sets is motivated by the definition of Grothendieck universe (see e.g. [9]).

Definition 3.1. *A universe of sets over* $\mathbf{L}$ is a class $\mathcal{U}$ of sets having the following properties:

(U1) $x \in y$ and $y \in \mathcal{U}$, then $x \in \mathcal{U}$,

(U2) $x, y \in \mathcal{U}$, then $\{x, y\} \in \mathcal{U}$,

(U3) $x \in \mathcal{U}$, then $\mathrm{P}(x) \in \mathcal{U}$,

(U4) $x \in \mathcal{U}$ and $y_i \in \mathcal{U}$ for any $i \in x$, then $\bigcup_{i \in x} y_i \in \mathcal{U}$,

(U5) $x \in \mathcal{U}$ and $f : x \to L$, then $\mathrm{Ran}(f) \in \mathcal{U}$,

where L is the support of the complete residuated lattice $\mathbf{L}$.

Simple examples of such universes are the classes of all, finite, or countable sets. Under certain conditions (see e.g. [9]), a Grothendieck universe may be also a universes of sets over $\mathbf{L}$. One could notice that L need not be included in $\mathcal{U}$. In fact, it is sufficient to consider the class of all finite set and an infinite complete residuated lattice $\mathbf{L}$. Notice that if $\mathcal{U}$ is a proper class, then $x \subseteq \mathcal{U}$ is not necessary a set. The following theorem shows some constructions inside the universe $\mathcal{U}$.

Theorem 3.1. *Let $x, y \in \mathcal{U}$ and $y_i \in \mathcal{U}$ for any $i \in x$. Then we have*

(i) $\emptyset$ and $\{x\}$ belong to $\mathcal{U}$,

(ii) $x \times y$, $x \sqcup y$, $x \cap y$ and y^x belong to $\mathcal{U}$,

(iii) if $z \in \mathcal{U} \cup \{L\}$ and $f : x \to z$, then f and $\mathrm{Ran}(f)$ belong to $\mathcal{U}$,

(iv) if $z \subseteq \mathcal{U}$ and $|z| \leq |x|$, then z belongs to $\mathcal{U}$,

(v) $\prod_{i \in x} y_i$, $\bigsqcup_{i \in x} y_i$ and $\bigcap_{i \in x} y_i$ belong to $\mathcal{U}$.

We can see that the whole set theory (i.e., unions, intersections, products of sets, singletons, the empty set) can be formed inside a universe $\mathcal{U}$, where, moreover, we may consider mappings from sets of $\mathcal{U}$ to the support of $\mathbf{L}$. Let us put $0 = \emptyset$, $1 = 0 \cup \{0\}$, $2 = 1 \cup \{1\}$, etc. Obviously, the natural numbers $0, 1, 2, \ldots$ belong to $\mathcal{U}$ and, according to (U4), an arbitrary finite set of natural numbers is also contained in $\mathcal{U}$. Nevertheless, the set of all natural numbers need not be an element of $\mathcal{U}$ (consider the universe of all finite sets) and it is only a subset of $\mathcal{U}$, in general.

The following theorem is very important in our investigation of equipollence of fuzzy sets (see Definition 4.4) and states that a set of $\mathcal{U}$ may be extended to a suitable greater sets of $\mathcal{U}$. Note that $\mathcal{U} \notin \mathcal{U}$.

Theorem 3.2. *Let $x \in \mathcal{U}$. Then there exists $y \in \mathcal{U}$ such that $|x| \leq |y \setminus x|$.*

Now, we can define the concept of fuzzy sets in $\mathcal{U}$ as follows.

Definition 3.2. Let $\mathcal{U}$ be a universe of sets over $\mathbf{L}$. A mapping $A : x \to L$ is called a fuzzy set in $\mathcal{U}$, if $x \in \mathcal{U}$.

ISBN: 978-989-95079-6-8

A consequence of (U2), (U4) and (U5) is that each fuzzy set (perceived as a relation $A \subseteq x \times L$) belongs to $\mathcal{U}$. Hence, all fuzzy sets form a subclass of $\mathcal{U}$. It is easy to see that fuzzy sets of higher order, i.e., fuzzy sets over sets of fuzzy sets, may be established inside of $\mathcal{U}$.

Let A be a fuzzy set in $\mathcal{U}$. The set $\mathrm{Dom}(A)$ is called a *universe of discourse of* A. We shall use $\mathcal{F}(\mathcal{U})$ to denote the class of all fuzzy sets in $\mathcal{U}$. Obviously, we have $\mathcal{F}(\mathcal{U}) \subseteq \mathcal{U}$, where $\mathcal{F}(\mathcal{U})$ is a proper class, whenever $\mathcal{U}$ is a proper class. The set $\mathrm{Supp}(A) = \{x \in \mathrm{Dom}(A) \mid A(x) > \bot\}$ is called *a support of fuzzy set* A. We shall say that A *is equivalent with* B, if $\mathrm{Supp}(A) = \mathrm{Supp}(B)$ and $A(x) = B(x)$ for any $x \in \mathrm{Supp}(A)$. For example, $A = \{0.9/a, 0/b\}$ is equivalent with $B = \{0.9, /a\}$. Obviously, the relation "to be equivalent" is an equivalence on $\mathcal{F}(\mathcal{U})$. We shall use $\mathrm{cls}(A)$ to denote the class of all equivalent fuzzy sets with A.

Definition 3.3. Let $A, B \in \mathcal{F}(\mathcal{U})$, $x = \mathrm{Dom}(A) \cup \mathrm{Dom}(B)$ and $A' \in \mathrm{cls}(A)$, $B' \in \mathrm{cls}(B)$ such that $\mathrm{Dom}(A') = \mathrm{Dom}(B') = x$. Then

- *the union* of A and B is a mapping $A \cup B : x \to L$ defined by
$$(A \cup B)(a) = A'(a) \vee B'(a) \tag{5}$$
for any $a \in x$,
- *the intersection* of A and B is a mapping $A \cap B : x \to L$ defined by
$$(A \cap B)(a) = A'(a) \wedge B'(a) \tag{6}$$
for any $a \in x$.

Definition 3.4. Let $A, B \in \mathcal{F}(\mathcal{U})$, $x = \mathrm{Dom}(A) \times \mathrm{Dom}(B)$ and $y = \mathrm{Dom}(A) \sqcup \mathrm{Dom}(B)$ (the disjoint union). Then

- *the product* of A, B is a mapping $A \times B : x \to L$ defined by
$$(A \times B)(a, b) = A(a) \wedge B(b) \tag{7}$$
for any $(a, b) \in x$,
- *the disjoint union* of A, B is a mapping $A \sqcup B : y \to L$ defined by
$$(A \sqcup B)(a, i) = \begin{cases} A(a, i), & \text{if } i = 1, \\ B(a, i), & \text{if } i = 2, \end{cases} \tag{8}$$
for any $(a, i) \in y$.

Definition 3.5. Let $A \in \mathcal{F}(\mathcal{U})$ and $A : x \to L$. Then *the complement of* A is a mapping $\overline{A} : x \to L$ defined by
$$\overline{A}(a) = A(a) \to \bot \tag{9}$$
for any $a \in x$.

Let us illustrate the proposed definitions by an example.

Example 3.1. Let $\mathbf{L}$ be the Łukasiewicz algebra and $A = \{1/a, 0.4/b\}$ and $B = \{0.6/a, 0.2/c\}$. Then we have

$$A \cup B = \{1/a, 0.4/b, 0.2/c\}$$
$$A \cap B = \{0.6/a, 0/b, 0/c\}$$
$$A \times B = \{0.6/(a, a), 0.2/(a, c), 0.4/(b, a), 0.2/(b, c)\}$$
$$A \sqcup B = \{1/(a, 1), 0.4/(b, 1), 0.6/(a, 2), 0.2/(c, 2)\}$$
$$\overline{A} = \{0/a, 0.6/b\}.$$

The following theorem shows that the relation "being equivalent" is a congruence for all mentioned operations except of the complement.

Theorem 3.3. *Let* $A, B \in \mathcal{F}(\mathcal{U})$ *and* $\circledast \in \{\cap, \cup, \times, \sqcup\}$. *If* $A' \in \mathrm{cls}(A)$ *and* $B' \in \mathrm{cls}(B)$, *then* $A' \circledast B' \in \mathrm{cls}(A \circledast B)$.

Although, fuzzy set in $\mathcal{U}$ is the most important concept in our theory, it seems to be useful (analogously to the classical set theory) to introduce the concept of fuzzy class in $\mathcal{U}$.

Definition 3.6. Let $\mathcal{U}$ be a universe of sets over $\mathbf{L}$. A mapping $\mathcal{A} : \mathcal{X} \to L$ is called a *fuzzy class in* $\mathcal{U}$, if $\mathcal{X} \subseteq \mathcal{U}$.

We shall use $\mathcal{C}(\mathcal{U})$ to denote the class of all fuzzy classes in $\mathcal{U}$. Obviously, we have $\mathcal{C}(\mathcal{U}) \not\subseteq \mathcal{U}$. From (U1) of Definition 3.1, it is easy to see that $x \subseteq \mathcal{U}$, whenever $x \in \mathcal{U}$. Hence, each fuzzy set is also a fuzzy class in $\mathcal{U}$, i.e., $\mathcal{F}(\mathcal{U}) \subseteq \mathcal{C}(\mathcal{U})$. Now there is a question: When can a fuzzy class be understood as a fuzzy set? One may observe that to each fuzzy set A there is a fuzzy class $\mathcal{A}$ such that $\mathrm{Dom}(\mathcal{A}) \notin \mathcal{U}$. However, such fuzzy class could be regarded as a fuzzy set in $\mathcal{U}$. This motivates us to establish the following condition of being a fuzzy set.

Definition 3.7. We shall say that a fuzzy class $\mathcal{A}$ in $\mathcal{U}$ is a fuzzy set, if there exists a fuzzy set A such that $A \in \mathrm{cls}(\mathcal{A})$. Otherwise, we say that it is *proper*.

4 Graded equipollence of fuzzy sets

As we have mentioned in Introduction, in the classical set theory, the relation of equipollence (equipotency or to have the same cardinality) on the class of all sets is defined using a one-to-one correspondence between sets. In order to introduce an equipollence for fuzzy sets in $\mathcal{U}$, we shall follow the classical idea and use "one-to-one mappings between fuzzy sets". Nevertheless, contrary to the classical approach, it seems to be reasonable to distinguish what one-to-one mapping is better than another one. This motivates us to introduce a degree in which a one-to-one mapping between suitable sets is a one-to-one mapping between fuzzy sets as follows.

Definition 4.1. Let $A, B \in \mathcal{F}(\mathcal{U})$, $x, y \in \mathcal{U}$ and $f : x \to y$ be a one-to-one mapping of x onto y in $\mathcal{U}$. We say that f is a *one-to-one mapping of* A *onto* B *in the degree* α, if $\mathrm{Supp}(A) \subseteq x \subseteq \mathrm{Dom}(A)$ and $\mathrm{Supp}(B) \subseteq y \subseteq \mathrm{Dom}(B)$ and
$$\alpha = \bigwedge_{z \in x} (A(z) \leftrightarrow B(f(z))). \tag{10}$$

On Fig. 1 we can see an example of a one-to-one mapping

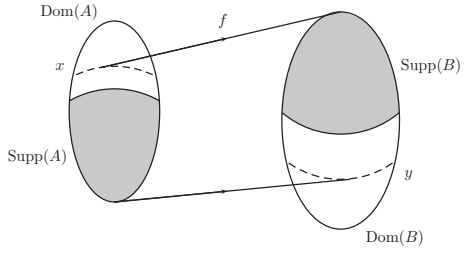


Figure 1: One-to-one mapping of A onto B

ISBN: 978-989-95079-6-8

f between fuzzy sets A and B, where not all elements of domains of A and B are respected. Obviously, $\mathrm{Dom}(A)\backslash x$ (analogously for B) contains the elements for which we have zero membership degree and thus they are unimportant elements of the fuzzy set A from this point of view. Moreover, there exists a fuzzy set A' with $\mathrm{Dom}(A') = x$ such that $A' \in \mathrm{cls}(A)$. This entitles us to omit such elements in our definition.

We shall write $[A \sim_f B] = \alpha$, if f is a one-to-one mapping of A onto B in the degree α. The following three examples demonstrate the notion of one-to-one mapping between fuzzy sets. For simplicity, we suppose that $\mathbf{L}$ is the Łukasiewicz algebra (see Example 2.2).

Example 4.1. Let $A = \{0.9/a, 0.5/b, 0/c\}$ and $B = \{0.5/1, 1/2, 0.2/3, 0/4\}$. If $f : \{a, b, c\} \to \{1, 2, 3\}$ is defined by $f(a) = 1$, $f(b) = 2$ and $f(c) = 3$, then we obtain $[A \sim_f B] = 0.5$, or by $f(a) = 2$, $f(b) = 1$ and $f(c) = 3$, then we obtain $[A \sim_f B] = 0.8$.

Example 4.2. Let A be the same as in Example 4.1 and $B' = \{0.5/1, 1/2, 0.2/3, 0.1/4\}$. Then there is no one-to-one mapping between A and B' satisfying the properties $\mathrm{Supp}(A) \subseteq x \subseteq \mathrm{Dom}(A)$ and $\mathrm{Supp}(B') \subseteq y \subseteq \mathrm{Dom}(B')$.

Example 4.3. Let us suppose that the set of all natural numbers N belongs to $\mathcal{U}$ and define $A, B : N \to [0, 1]$ by $A(n) = 1$ and

$$B(n) = \begin{cases} 1, & \text{if } n \text{ is an odd number,} \\ 0.1, & \text{otherwise.} \end{cases}$$

for any $n \in N$. Obviously, $[A \sim_f B] = 0.1$ for any one-to-one mapping between A and B. Notice that if we put 0 instead of 0.1 in the definition of B, then $[A \sim_f B] = 1$ for a suitable one-to-one mapping f.

From the latter two examples, we can see that there are some problems to derive "correct" degrees in which there exist one-to-one mappings between these fuzzy sets. More precisely, in Example 4.2, we cannot construct any one-to-one mapping between fuzzy sets, and in Example 4.3, the elements with a lower non-zero membership degree significantly influenced the degree of a one-to-one mapping between fuzzy sets. A natural idea how to solve these problems is to consider not just the given fuzzy sets, but the classes of all equivalent fuzzy sets with them. An example may be seen on Fig. 2,

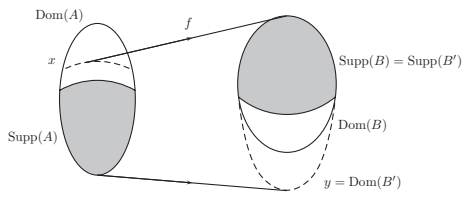


Figure 2: Extension of $\mathrm{Dom}(B)$ to $\mathrm{Dom}(B')$ for $B' \in \mathrm{cls}(B)$

where an extension of the original domain of fuzzy set B to a new domain of fuzzy set B' which is equivalent to B, i.e., $B' \in \mathrm{cls}(B)$, is given. Before we introduce the definition of being equipollent at least in the degree α, let us establish the following set of one-to-one correspondences.

Definition 4.2. Let $A, B \in \mathcal{F}(\mathcal{U})$. A mapping $f : x \to y$ belongs to the set $\mathrm{Bij}(A, B)$, if

(i) f is a one-to-one mapping of x onto y,

(ii) $\mathrm{Supp}(A) \subseteq x \subseteq \mathrm{Dom}(A)$, and

(iii) $\mathrm{Supp}(B) \subseteq y \subseteq \mathrm{Dom}(B)$.

The definition of being equipollent at least in some degree is following.

Definition 4.3. Let $A, B \in \mathcal{F}(\mathcal{U})$. We say that A *is equipollent with* B (or A has the same cardinality as B) *at least in the degree* α, if there exists $A' \in \mathrm{cls}(A)$, $B' \in \mathrm{cls}(B)$ and $f \in \mathrm{Bij}(A', B')$ such that $[A' \sim_f B'] \geq \alpha$.

Example 4.4. Let us suppose A an B' from Examples 4.1 and 4.2. Then A is equipollent with B' at least in the degree 0.8, since $[A' \sim_f B'] = 0.8$, where $A' : \{a, b, c, d\} \to [0, 1]$ with $A'(x) = A(x)$ for $x \in \{a, b, c\}$ and $A(d) = 0$, and

$$f(a) = 2,\ f(b) = 1,\ f(c) = 3,\ f(d) = 4.$$

The following result is clearly expected.

Lemma 4.1. *Let $A \in \mathcal{F}(\mathcal{U})$ and $A' \in \mathrm{cls}(A)$. Then A is equipollent with A' at least in the degree $\top$.*

If we can find a degree α in which some fuzzy sets A and B are "at least equipollent", i.e, there exists a one-to-one mapping f and $A' \in \mathrm{cls}(A)$, $B' \in \mathrm{cls}(B)$ with $[A' \sim_f B'] \geq \alpha$, a natural question arises, how to define a degree in which fuzzy sets A and B are equipollent. Obviously, one way, how to do it, is to use the supremum from all degrees of possible one-to-one mappings that can be created between all couples of fuzzy sets that are equivalent with original ones. However, here is one problem. If $\mathcal{U}$ is a proper class, then the class $\mathrm{cls}(A)$ of all equivalent fuzzy sets with A is proper. Hence, we cannot use the supremum in the mentioned sense. The following definition proposes a way, how to introduce a degree in which two fuzzy sets are equipollent and to avoid the proper classes.

Definition 4.4. Let $A, B \in \mathcal{F}(\mathcal{U})$. We shall say that A *is equipollent with* B (or A has the same cardinality as B) *in the degree* α, if there exist fuzzy sets $C \in \mathrm{cls}(A)$ and $D \in \mathrm{cls}(B)$ such that

$$\alpha = \bigvee_{f \in \mathrm{Bij}(C,D)} [C \sim_f D] \tag{11}$$

and, for each $A' \in \mathrm{cls}(A)$, $B' \in \mathrm{cls}(B)$ and $f \in \mathrm{Bij}(A', B')$, there is $[A' \sim_f B'] \leq \alpha$.

If A is equipollent with B in the degree α, then we shall write $[A \sim B] = \alpha$. One may be surprised, that it is sufficient to restrict ourselves to suitable fuzzy sets, and may ask, if this definition is really correct. That means, if, for arbitrary fuzzy sets A and B, there exists the degree in which they are equipollent. The correctness of our definition follows from the following theorem.

Theorem 4.2. *Let $A, B \in \mathcal{F}(\mathcal{U})$. Then, for any $C \in \mathrm{cls}(A)$ and $D \in \mathrm{cls}(B)$ such that $|\mathrm{Supp}(A)| \leq |\mathrm{Dom}(D)\backslash\mathrm{Supp}(B)|$ and $|\mathrm{Supp}(B)| \leq |\mathrm{Dom}(C) \setminus \mathrm{Supp}(A)|$, there is*

$$[A \sim B] = \bigvee_{f \in \mathrm{Bij}(C,D)} [C \sim_f D]. \tag{12}$$

ISBN: 978-989-95079-6-8

Example 4.5. Let A, B' be fuzzy sets from Example 4.1 and 4.2 and define $C, D : \{a, b, c, 1, 2, 3, 4\} \to [0, 1]$ such that $C \in \mathrm{cls}(A)$ and $D \in \mathrm{cls}(B')$. According to previous theorem, we can write

$$[A \sim B] = \bigvee_{f \in \mathrm{Bij}(C,D)} [C \sim_f D] = [C \sim_{f_0} D] = 0.8,$$

where $f_0 : \{a, b, c, 1\} \to \{1, 2, 3, 4\}$ is defined by $f_0(a) = 1$, $f_0(b) = 2$, $f_0(c) = 3$ and $f_0(1) = 4$ (compare with Example 4.4).

Example 4.6. Let A, B be fuzzy sets from Example 4.3. Put $M = \{0, 0.5, 1, 1.5, \dots\}$, where e.g. $0.5 = \{\{0, 1\}\}$, $1.5 = \{\{1, 2\}\}$, etc. and define $C, D : M \to [0, 1]$ such that $C \in \mathrm{cls}(A)$ and $D \in \mathrm{cls}(B)$. One checks easily that $[A \sim B] = [C \sim_f D] = 0.9$, where $f(m) = 2m + 1$ for any $m \in M$.

Lemma 4.3. *Let $f \in \mathrm{Bij}(A, B)$ and $g \in \mathrm{Bij}(B', C)$, where $B' \in \mathrm{cls}(B)$. Then there exist $C' \in \mathrm{cls}(C)$ and $h \in \mathrm{Bij}(A, C')$ such that*

$$[A \sim_f B] \otimes [B' \sim_g C] \le [A \sim_h C']. \tag{13}$$

Recall that $\mathcal{F}(\mathcal{U}) \times \mathcal{F}(\mathcal{U})$ is a subclass of $\mathcal{U}$. Hence, we can define the following fuzzy class (or also fuzzy class relation) in $\mathcal{U}$, which is analogous to the equivalence relation "to be equipollent" on the class of all sets.

Theorem 4.4. *The fuzzy class relation $\sim : \mathcal{F}(\mathcal{U}) \times \mathcal{F}(\mathcal{U}) \to L$ is a similarity relation ($\otimes$-equivalence) on the class $\mathcal{F}(\mathcal{U})$, i.e.,*

(i) $[A \sim A] = \top$,

(ii) $[A \sim B] = [B \sim A]$,

(iii) $[A \sim B] \otimes [B \sim C] \le [A \sim C]$

hold for arbitrary fuzzy sets $A, B, C \in \mathcal{F}(\mathcal{U})$.

The fuzzy class relation $\sim$ is called the *equipollence of fuzzy sets* and $[A \sim B]$ is the *degree of equipollence* of fuzzy sets A and B. The following statement is straightforward.

Corolarry 4.5. *Let $A, B \in \mathcal{F}(\mathcal{U})$ and $A' \in \mathrm{cls}(A)$. Then $[A \sim B] = [A' \sim B]$.*

In classical set theory, if $f : x \to z$ is a one-to-one mapping of x to z and $|x| = |y|$, then also $|f(x)| = |y|$, where $f(x) = \mathrm{Ran}(f)$. Analogously, we can give the following statement for fuzzy sets.

Theorem 4.6. *Let $A, B \in \mathcal{F}(\mathcal{U})$ and $f : \mathrm{Dom}(A) \to z$ be a one-to-one mapping in $\mathcal{U}$. If we put*

$$f(A)(y) = \begin{cases} A(x), & \text{if } f(x) = y, \\ \bot, & \text{otherwise,} \end{cases} \tag{14}$$

for any $y \in z$, then $[A \sim B] = [f(A) \sim B]$.

A simple consequence of Theorems 4.4 and 4.6 is the following statement.

Corolarry 4.7. *Let $A \in \mathcal{F}(\mathcal{U})$ and $f : \mathrm{Dom}(A) \to z$ be a one-to-one mapping in $\mathcal{U}$. Then $[A \sim f(A)] = \top$.*

Now there is a question, if the definition of graded equipollence, which is more complicated and practically useless, could be simplified. The remaining part of this section will give a partial positive answer.

Let $A, B \in \mathcal{F}(\mathcal{U})$ and $\mathrm{Dom}(A) = x$, $\mathrm{Dom}(B) = y$. Put

$$A^*(z, i) = \begin{cases} A(z), & \text{if } i = 1, \\ \bot, & \text{if } i = 2, \end{cases} \tag{15}$$

$$B^*(z, i) = \begin{cases} \bot, & \text{if } i = 1, \\ B(z), & \text{if } i = 2 \end{cases} \tag{16}$$

for any $(z, i) \in x \sqcup y$. According to Corollary 4.7, we have $[A \sim A^*] = \top$ and $[A \sim B] = [A^* \sim B^*]$ and, hence, it does not matter, whether we consider the original fuzzy sets A, B or their images A^*, B^* to obtain the right degree in which A and B are equipollent. A straightforward consequence of Theorem 4.2 is the following lemma.

Lemma 4.8. *Let $A, B \in \mathcal{F}(\mathcal{U})$. Then*

$$[A^* \sim B^*] = \bigvee_{f \in \mathrm{Bij}(A^*, B^*)} [A^* \sim_f B^*]. \tag{17}$$

The following lemma shows that we need not consider all one-to-one mappings to obtain the right degree of equipollence.

Lemma 4.9. *Let $A, B \in \mathcal{F}(\mathcal{U})$ and $f \in \mathrm{Bij}(A^*, B^*)$. Then for any $z \in \mathcal{U}$ with*

$$\mathrm{Dom}(f) \subseteq z \subseteq \mathrm{Dom}(A) \sqcup \mathrm{Dom}(B) \tag{18}$$

there exists $f' \in \mathrm{Bij}(A^*, B^*)$ with $\mathrm{Dom}(f') = z$ such that

$$[A^* \sim_f B^*] = [A^* \sim_{f'} B^*]. \tag{19}$$

From the previous lemma, one could notice that the set $\mathrm{Bij}(A^*, B^*)$ is too large and it is sufficient to consider the one-to-one mappings with the same domain $\mathrm{Dom}(A^*)$ or conversely with the same range $\mathrm{Dom}(B^*)$. Hence, we may ask, whether we can restrict ourselves to one-to-one correspondences between $\mathrm{Dom}(A^*)$ and $\mathrm{Dom}(B^*)$. The answer is positive as the following theorem shows and this fact is a consequence of Cantor-Bernstein theorem. We shall use $\mathrm{Perm}(A, B)$ to denote the set of all one-to-one mappings of $\mathrm{Dom}(A)$ onto $\mathrm{Dom}(B)$.

Theorem 4.10. *Let $A, B \in \mathcal{F}(\mathcal{U})$. Then*

$$[A^* \sim B^*] = \bigvee_{f \in \mathrm{Perm}(A^*, B^*)} [A^* \sim_f B^*]. \tag{20}$$

A generalization of this theorem to a wider scale of fuzzy sets is as follows (cf. Theorem 4.2).

Theorem 4.11. *Let $A, B \in \mathcal{F}(\mathcal{U})$. Then, for any $C \in \mathrm{cls}(A)$ and $D \in \mathrm{cls}(B)$ such that $|\mathrm{Supp}(A)| = |\mathrm{Dom}(D) \setminus \mathrm{Supp}(B)|$ and $|\mathrm{Supp}(B)| = |\mathrm{Dom}(C) \setminus \mathrm{Supp}(A)|$, there is*

$$[A \sim B] = \bigvee_{f \in \mathrm{Perm}(C, D)} [C \sim_f D]. \tag{21}$$

If we deal with fuzzy sets in $\mathcal{U}$ with finite supports, then it is sufficient to consider fuzzy sets which domains have the same cardinality as the following theorem shows.

ISBN: 978-989-95079-6-8

Theorem 4.12. *Let $A, B \in \mathcal{F}(\mathcal{U})$ and $C \in \mathrm{cls}(A)$, $D \in \mathrm{cls}(B)$ be such that $|\mathrm{Dom}(C)| = |\mathrm{Dom}(D)| = n$. Then*

$$[A \sim B] = \bigvee_{f \in \mathrm{Perm}(C,D)} [C \sim_f D]. \tag{22}$$

Now we may ask, when there exists just one one-to-one mapping between two fuzzy sets which determines the degree of their equipollence. If the given residuated lattice is not linearly ordered or is a complete linearly ordered residuated lattice with an infinite support, then the following examples show simple constructions of fuzzy sets, when more than one one-to-one mapping between them have to be used to find the degrees of their equipollence.

Example 4.7. Let $\mathbf{L}$ be a residuated lattice from Example 2.2, $A = \{\bot/0, b/1, \top/2\}$ and $B = \{a/0, b/1, b/2\}$. One checks easily that $[A \sim B] = \top$ but there is no permutation on x giving the degree of equipollence of A and B equal to $\top$.

Example 4.8. Let $\mathbf{L}$ be the Łukasiewicz algebra and let us suppose that the set of all natural numbers N belongs to $\mathcal{U}$. Define $A, B : N \to L$ by $A(n) = 0.5 - \frac{1}{n+2}$ for any $n \in N$ and $B(0) = 1$, $B(n) = 0$ for any $n \in N$, where $n \neq 0$. Since the presumptions of Theorem 4.11 are satisfied, we can use (21) to compute the degree of equipollence. If f is a permutation on N such that $f(n) = 0$, then one checks easily that

$$[A \sim_f B] = A(n) \leftrightarrow B(0) = 0.5 - \frac{1}{n+2}. \tag{23}$$

Note that $A(n) \leftrightarrow 0 = 0.5 + \frac{1}{n+2}$. Hence, we obtain

$$[A \sim B] = \bigvee_{f \in \mathrm{Perm}(A,B)} [A \sim_f B] = 0.5,$$

but, due to (23), no permutation on N gives such value.

Let N be the set of all natural numbers and $A \in \mathcal{F}(\mathcal{U})$ be a fuzzy set with finite support. Define a fuzzy class $p_A : N \to L$ by

$$p_A(i) = \bigvee_{\substack{z \subseteq \mathrm{Supp}(A),\\ |z| = i}} \bigwedge_{x \in z} A(x) \tag{24}$$

for any $i \in N$.

Theorem 4.13. *Let $\mathbf{L}$ be a complete linearly ordered residuated lattice and $A, B \in \mathcal{F}(\mathcal{U})$ be fuzzy sets with finite supports. Then*

$$[A \sim B] = \bigwedge_{i \in N} (p_A(i) \leftrightarrow p_B(i)).$$

Corolarry 4.14. *Let $\mathbf{L}$ be a linearly ordered residuated lattice and $A, B \in \mathcal{F}(\mathcal{U})$ be fuzzy sets with $|\mathrm{Dom}(A)| = |\mathrm{Dom}(B)| = n$. Then there exists $f \in \mathrm{Perm}(A, B)$ such that $[A \sim B] = [A \sim_f B]$.*

Example 4.9. Let $\mathbf{L}$ be the Łukasiewicz algebra and $x = \{c_1 \ldots, c_5\}$ be a set of criterion. Let us suppose that the following fuzzy sets

$$A = \{0.3/c_1, 0.8/c_2, 0.4/c_3, 0/c_4, 0/c_5\}$$
$$B = \{0/c_1, 0.6/c_2, 0.9/c_3, 0/c_4, 0.7/c_5\}$$

express the satisfaction of all criterion by two applicants and our goal is to determine, whether the numbers of satisfied criterion are more or less the same for both applicants. According to (24), we can establish

$$p_A = \{1/0, 0.8/1, 0.4/2, 0.3/3, 0/4, 0/5, \ldots\},$$
$$p_B = \{1/0, 0.9/1, 0.7/2, 0.6/3, 0/4, 0/5, \ldots\}$$

and, using Theorem 4.13, we obtain

$$[A \sim B] = \bigwedge_{i \in N} (p_A(i) \leftrightarrow p_B(i)) = 0.7.$$

Hence, we can conclude that both applicants have more or less the same number of satisfied criterion.

The following theorem demonstrates a way how we can investigate the well-known relations in cardinal set theory for fuzzy sets in $\mathcal{U}$.

Theorem 4.15. *Let $A, B \in \mathcal{F}(\mathcal{U})$. Then we have*

(i) $[A \sim B] \leq [\overline{A} \sim \overline{B}]$

(ii) $[A \sim B] \otimes [C \sim D] \leq [A \times C \sim B \times D]$,

(iii) if $\mathrm{Supp}(A \cap B) = \mathrm{Supp}(C \cap D) = \emptyset$, *then*

$$[A \sim C] \otimes [B \sim D] \leq [A \cup C \sim B \cup D].$$

5 Conclusion

Let us note that the fuzzy set cardinality is not widely developed in a comparison to the cardinal theory for sets giving many nice result. The proposed definitions of the universe, operations and equipollence of fuzzy sets are a starting point in a deeper fuzzy sets investigation from the cardinality point of view. We think that the results could be interesting for the fuzzy set theory as well as for the practical application as e.g. a determination of "number" of well satisfied criteria.

Acknowledgment

This paper has been supported by the Grant IAA108270901 of the GA AV ČR.

References

[1] D. Klaua. Zum kardinalzahlbegriff in der mehrwertigen mengenlehre. In G. Asser, J. Flashmayers, and W. Rinow, editors, *Theory of Sets and Topology*, pages 313–325. Deutsher Verlag der Wissenshaften, Berlin, 1972.

[2] S. Gottwald. Zahlbereichskonstruktionen in einer mehrwertigen mengenlehre. *Z. Math. Logic Grundl. Math.*, 17:145–188, 1971.

[3] S. Gottwald. Fuzzy uniqueness of fuzzy mappings. *Fuzzy Sets and Systems*, 3:49–74, 1980.

[4] M. Wygralak. Generalized cardinal numbers and operations on them. *Fuzzy Sets and Systems*, 53(1):49–85, 1993.

[5] M. Wygralak. An axiomatic approach to scalar cardinalities of fuzzy sets. *Fuzzy Sets and Systems*, 110(2):175–179, 2000.

[6] M. Wygralak. Fuzzy sets with triangular norms and their cardinality theory. *Fuzzy Sets and Systems*, 124(1):1–24, 2001.

[7] M. Wygralak. *Cardinalities of Fuzzy Sets*. Kluwer Academic Publisher, Berlin, 2003. isbn: 3-5400-0337-1.

[8] E.P. Klement, R. Mesiar, and E. Pap. *Triangular norms*, volume 8 of *Trends in Logic*. Kluwer Academic Publisher, Dordrecht, 2000.

[9] C. McLarty. *Elementary Categories, Elementary Toposes*. Oxford University Press, 1992.

Managing the Absence of Items in Fuzzy Association Mining

C. Molina[1] D. Sanchez[2] J.M. Serrano[1] M.A. Vila[2]

1.Department of Computer Science, University of Jaen
Jaen, Spain
2.Department of Computer Science and Artificial Intelligence, University of Granada
Granada, Spain
Email: carlosmo@ujaen.es, daniel@decsai.ugr.es, jschica@ujaen.es, vila@decsai.ugr.es

Abstract— *One of the most well-known and extended data mining techniques is that of association rule mining, a helpful tool to discover relations between items present in sets of transactions. Nevertheless, in some other scenarios, another interesting issue is that of considering not only the possible relations involving presence of items, but the absence of them. The problem gets more complex when it is necessary to represent also imprecision and/or uncertainty in the information. In this paper, we introduce a methodology to obtain fuzzy association rules involving absent items. Additionally, our proposal is based on restriction level sets, a recent representation of imprecision that extends that of fuzzy sets, and introduces some new operators, covering some misleading results obtained from usual fuzzy operators as, for example, negation. In our methodology, we define new measures of interest and accuracy for fuzzy association rules as RL-numbers, as well as we propose a new way of summarizing the resulting set of fuzzy association rules, distributed in restriction levels.*

Keywords— Absence of items, fuzzy association rules, restriction levels

1 Introduction

As the amount of information stored in databases grows continuously day after day, it is desirable the development of tools, not only to properly manage all this knowledge, but to get a profit from it. Over the past two decades, considerable efforts have been devoted to the study and development of techniques in data mining and knowledge discovery in databases. One of the most well-known and extended data mining techniques is that of association rule mining, a helpful tool to discover implicit, non-trivial and potentially useful relations between items in sets of transactions. The most known example if that of market baskets, but association rules can be also applied on relational databases. Since the original association rule definition was proposed in [1], many studies and related methodologies have been devoted to extend this tool in order to manage different scenarios and knowledge representations.

One interesting issue, approached in works like [2], and [3], is that of considering both present and absent items. In these cases, it is not only interesting to detect relations between items included in transactions, but also if the absence of given items in the transaction can be related to the presence of some others. One possible application can be the discovery of conflicts or complements between items, as well as the establishment of constraints in the sets of items involved in the data mining procedure. Some other authors have shown interest in this problem, and additional information and efficient algorithms can be found in [4],[5],[6], and [7].

On the other hand, another interesting issue is that of applying the theory of fuzzy sets [8] in the modeling of how items are included in transactions, leading to the definition of fuzzy association rules. In the literature, some interesting approaches are discussed in [9], [10], [11], [12], and [13], among others.

But, as far as we know, none of the cited works manage properly the absence of items in the fuzzy case. In particular, consider an item i that belongs to a fuzzy transaction with a membership degree of 0.5. According to the usual fuzzy extension of the complement operator, the complementary item $\neg i$ should belong to the same transaction with a membership degree of $1 - 0.5 = 0.5$, leading us to counterintuitive results. Restriction level sets [14], a recently introduced representation of fuzzy quantities that extends the representation by means of fuzzy sets, manage to solve scenarios like these.

The representation via restriction level sets can be seen as equivalent to that of fuzzy sets seen as collections of alpha cuts, and hence it offers a alternative approach to the representation and management of fuzzy quantities, differing from fuzzy sets in the sense that operations are defined on levels. However, some operations as negation result in a representation that may not correspond to the alpha cuts of a fuzzy set.

Actually, although a parallel definition of a logic model for association rules based also on restriction level sets can be found in [15], our proposal, in this paper, is a methodology for mining fuzzy association rules via restriction level sets, instead of fuzzy sets, extending the definition introduced in [16], and, in addition, considering both present and absent items.

The paper is organized as follows. Sections 2 and 3 are devoted to introduce the reader in the followed representations and previous methodologies. Next, we introduce our proposal in section 4, where we define new measures for fuzzy association rules, based on RL-numbers, and we describe some derived properties from these. Next, some aspects with respect to the implementation and our first experimentation results are discussed in section 5. Finally, we present our concluding remarks as well as propose some interesting open issues in the conclusions sections.

2 Data mining tools

2.1 Association rules

Given a set I ("set of items") and a set of transactions T (also called T-set), each transaction being a subset of I, association

ISBN: 978-989-95079-6-8

rules [1] are "implications" of the form $A \Rightarrow C$ that relate the presence of itemsets (sets of items) A and C in transactions of T, assuming $A, C \subseteq I$, $A \cap C = \emptyset$ and $A, C \neq \emptyset$.

In the case of relational databases, it is usual to consider that items are pairs $\langle attribute, value \rangle$, and transactions are tuples in a table. For example, the item $\langle X, x_0 \rangle$ is in the transaction associated to a tuple t iff $t[X] = x_0$.

The ordinary measures proposed in [1] to assess association rules are *confidence* (the conditional probability $p(C|A)$) and *support* (the joint probability $p(A \cup C)$). An alternative framework was proposed in [17]. In this framework, accuracy is measured by means of Shortliffe and Buchanan's certainty factors [18], in the following way: the certainty factor of the rule $A \Rightarrow C$ is

$$CF(A \Rightarrow C) = \frac{(Conf(A \Rightarrow C)) - S(C)}{1 - S(C)} \quad (1)$$

if $Conf(A \Rightarrow C) > S(C)$, and

$$CF(A \Rightarrow C) = \frac{(Conf(A \Rightarrow C)) - S(C)}{S(C)} \quad (2)$$

if $Conf(A \Rightarrow C) < S(C)$, and 0 otherwise.

Certainty factors take values in $[-1, 1]$, indicating the extent to which our belief that the consequent is true varies when the antecedent is also true. It ranges from 1, meaning maximum increment (i.e., when A is true then C is true) to -1, meaning maximum decrement.

2.2 Fuzzy association rules

In [16], the model for association rules is extended in order to manage fuzzy values in databases. The approach is based on the definition of fuzzy transactions as fuzzy subsets of items. Let $I = \{i_1, \ldots, i_m\}$ be a set of items and T' be a set of fuzzy transactions, where each fuzzy transaction is a fuzzy subset of I. Let $\tilde{\tau} \in T'$ be a fuzzy transaction, we note $\tilde{\tau}(i_k)$ the membership degree of i_k in $\tilde{\tau}$. A fuzzy association rule is an implication of the form $A \Rightarrow C$ such that $A, C \subset I$ and $A \cap C = \emptyset$.

It is immediate that the set of transactions where a given item appears is a fuzzy set. We call it *representation* of the item. For item i_k in T' we have the following fuzzy subset of T':

$$\tilde{\Gamma}_{i_k} = \sum_{\tilde{\tau} \in T'} \tilde{\tau}(i_k)/\tilde{\tau} \quad (3)$$

This representation can be extended to itemsets as follows: let $I_0 \subset I$ be an itemset, its representation is the following subset of T':

$$\tilde{\Gamma}_{I_0} = \bigcap_{i \in I_0} \tilde{\Gamma}_i = min_{i \in I_0} \tilde{\Gamma}_i \quad (4)$$

In order to measure the interest and accuracy of a fuzzy association rule, we must use approximate reasoning tools, because of the imprecision that affects fuzzy transactions and, consequently, the representation of itemsets. In [16], a semantic approach is proposed based on the evaluation of quantified sentences (see [19]). Let Q be a fuzzy coherent quantifier:

- The support of an itemset $\tilde{\Gamma}_{I_0}$ is equal to the result of evaluating the quantified sentence Q of T' are $\tilde{\Gamma}_{I_0}$.
- The support of the fuzzy association rule $A \Rightarrow C$ in the FT-set T', $Supp(A \Rightarrow C)$, is the evaluation of the quantified sentence Q of T are $\tilde{\Gamma}_{A \cup C}$ = Q of T are $(\tilde{\Gamma}_A \cap \tilde{\Gamma}_C)$.
- The confidence of the fuzzy association rule $A \Rightarrow C$ in the FT-set T', $Conf(A \Rightarrow C)$, is the evaluation of the quantified sentence Q of $\tilde{\Gamma}_A$ are $\tilde{\Gamma}_C$.

As seen in [16], the proposed method is a generalization of the ordinary association rule assessment framework in the crisp case.

3 Restriction-level representation

3.1 Representation

The RL-representation of an imprecise property is a collection of crisp sets, each crisp set corresponding to a crisp realization of the property under a *restriction* rule. We distinguish between *atomic* and *derived* properties. Atomic properties are those that cannot be defined in terms of other properties in our problem. Derived properties are defined by logical operations on other properties.

In [14] we consider that atomic imprecise properties are represented by fuzzy sets, and hence restrictions are of the form $degree \geq \alpha$ with $\alpha \in (0, 1]$, and restriction levels are associated to values $\alpha \in (0, 1]$. In the same case, the crisp realization of an atomic imprecise property represented by a fuzzy set A in the restriction level α corresponds to the α-cut A_α.

For every property we assume that there is a finite set of restriction levels $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying that $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$, $m \geq 1$. We call such sets *RL-sets*. The consideration that a RL-set is finite is not a practical limitation since humans are able to distinguish a limited number of restriction or precision levels and, in practice, the limit in precision and storage of computers allows us to work with a finite number of degrees (and consequently, of levels) only. In practice, the RL-set for an atomic property represented by a fuzzy set A on an universe X is

$$\Lambda_A = \{A(x) \mid x \in support(A)\} \cup \{1\} \quad (5)$$

The RL-set employed to represent a derived property is obtained as the union of the RL-sets of the atomic properties in terms of which the property is defined. Finally, the RL-representation of an imprecise property on X is defined in [14] as follows:

Definition 1 *A RL-representation is a pair (Λ, ρ) where Λ is a RL-set and ρ is a function*

$$\rho : \Lambda \rightarrow \mathcal{P}(X) \quad (6)$$

The function ρ indicates the crisp realization that represents the imprecise property for each restriction level. As an example, the RL-representation for an atomic imprecise property defined by a fuzzy set A is the pair (Λ_A, ρ_A), where Λ_A is obtained using equation (5), and $\rho_A(\alpha) = A_\alpha \; \forall \alpha \in \Lambda_A$.

Given an imprecise property P represented by (Λ_P, ρ_P), we define the set of crisp representatives of P, Ω_P, as

$$\Omega_P = \{\rho_P(\alpha) \mid \alpha \in \Lambda_P\} \quad (7)$$

ISBN: 978-989-95079-6-8

For an atomic property A, the set of crisp representatives Ω_A is the set of significant α-cuts of A, as we have seen. However, notice that in definition 1 there is no restriction about the possible crisp representatives for non-atomic properties. In particular, as a consequence of operations, they don't need to be nested, so the final RL-representation of a derived property is not always equivalent to the α-cut representation of fuzzy sets.

In order to define properties by operations, it is convenient to extend the function ρ to any RL $\alpha \in (0,1]$. Let (Λ, ρ) be a RL-representation with $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$. Let $\alpha \in (0,1]$ and $\alpha_i, \alpha_{i+1} \in \Lambda$ such that $\alpha_i \geq \alpha > \alpha_{i+1}$. Then

$$\rho(\alpha) = \rho(\alpha_i) \tag{8}$$

Finally, let us remark that a RL-representation (Λ, ρ) on X is a crisp set $A \subseteq X$ iff $\forall \alpha \in \Lambda$, $\rho(\alpha) = A$.

3.2 Interpretation in terms of evidence

Given a RL-representation (Λ_A, ρ_A) for an atomic property A, the values of Λ_A can be interpreted as values of possibility of a possibility measure defined $\forall \rho_A(\alpha_i) \in \Omega_A$ as

$$Pos(\rho_A(\alpha_i)) = \alpha_i. \tag{9}$$

Following this interpretation we define a basic probability assignment in the usual way:

Definition 2 *Let (Λ, ρ) be a RL-representation with crisp representatives Ω. The associated probability distribution $m : \Omega \to [0,1]$ is*

$$m(Y) = \sum_{\alpha_i \mid Y = \rho(\alpha_i)} \alpha_i - \alpha_{i+1}. \tag{10}$$

The basic probability assignment m_F gives us information about how representative of the property F is each crisp set in Ω_F. For each $Y \in \Omega_F$, the value $m_F(Y)$ represents the proportion to which the available evidence supports the claim that the property F is represented by Y. From this point of view, a RL-representation can be seen as a basic probability assignment in the sense of the theory of evidence, *plus a structure indicating dependencies between the possible representations of different properties*.

3.3 RL-numbers

On the basis of RL-representations and operations, RL-numbers were introduced in [20] as a representation of imprecise quantities.

Definition 3 *A RL-real number is a pair $(\Lambda, \mathcal{R})$ where Λ is a RL-set and $\mathcal{R} : (0,1] \to \mathbb{R}$.*

We shall note $\mathbb{R}_{RL}$ the set of RL-real numbers. The RL-real number R_x is the representation of a (precise) real number x iff $\forall \alpha \in \Lambda_{R_x}$, $\mathcal{R}_{R_x}(\alpha) = x$. We shall denote such RL-real number as R_x or, equivalently, x, since in the crisp case, the set Λ_{R_x} is unimportant. Operations are extended as follows:

Definition 4 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $R_1 \ldots R_n$ be RL-real numbers. Then $f(R_1, \ldots, R_n)$ is a RL-real number with*

$$\Lambda_{f(R_1,\ldots,R_n)} = \bigcup_{1 \leq i \leq n} \Lambda_{R_i} \tag{11}$$

and, $\forall \alpha \in \Lambda_{f(R_1,\ldots,R_n)}$

$$\mathcal{R}_{f(R_1,\ldots,R_n)}(\alpha) = f(\mathcal{R}_{R_1}(\alpha), \ldots, \mathcal{R}_{R_n}(\alpha)) \tag{12}$$

This approach offers two main advantages:

- RL-numbers are representations of fuzzy quantities that can be easily obtained by extending usual crisp measurements to fuzzy sets.
- Arithmetic and logical operations on RL-numbers are straightforward and unique extensions of the operations on crisp numbers, verifying the following:
  - They verify all the usual properties of crisp arithmetic and logical operations.
  - The imprecision does not necessarily increase through operations, and can even diminish. The maximum imprecision is related to the number of restriction levels employed.

4 Fuzzy association rules via restriction levels

As we noted before, restriction level sets can be viewed as sets of alpha cuts in fuzzy sets. Actually, if the restriction level set represents an atomic property, the representation is equal to that of fuzzy sets. Following this idea, we extend the definition of fuzzy association rules introduced in [16], using restriction level sets instead. Moreover, since each restriction level corresponds to a crisp set, and crisp operators hold for each level, we can define a fuzzy association rule via restriction levels as an aggregation of all the related (crisp) association rules extracted at each restriction level.

In our approach, we represent items via restriction levels as follows. Let $I = \{i_1, \ldots, i_m\}$ be a set of items and T' be a set of fuzzy transactions, where each fuzzy transaction is a restriction level set of I. Let $\tilde{\tau} \in T'$ be a fuzzy transaction, we note $\tilde{\tau}(i_k)$ the membership degree of i_k in $\tilde{\tau}$. Let $supp_\alpha(i_k)$ be the support of i_k (i.e., the probability that $\tilde{\tau}(i_k) > \alpha, \forall \tilde{\tau} \in T'$) at the restriction level α.

On the other hand, let $\neg i_k$ be the negation (absence) of item i_k, and let $supp_\alpha(\neg i_k) = 1 - supp_\alpha(i_k)$ be the support of $\neg i_k$ at the restriction level α. Note how, in the case of negated items, the subsequent restriction level set may not correspond necessarily to a fuzzy set.

Let $\bar{I} = \{\neg i_1, \ldots, \neg i_m\}$ be the set of negated items from I, and $A \subset \{I \cup \bar{I}\}, C \in \{I \cup \bar{I}\}, A \cap C = \emptyset$, the antecedent and the consequent of the fuzzy association rule $R : A \Rightarrow C$. If we see this fuzzy association rule R under its RL-representation as the pair (Λ_R, ρ_R), we define the set of crisp representations of R, Ω_R, as

$$\Omega_R = \{\rho_R(\alpha) \mid \alpha \in \Lambda_R\} \tag{13}$$

where $\rho_R)(\alpha) = (R)_\alpha$, that is, ρ represents the association rules at each restriction level. In other words, this fuzzy association rule can be viewed as the aggregation of all (crisp)

ISBN: 978-989-95079-6-8

association rules of the form $A \Rightarrow C$ that can be obtained at each restriction level α.

Let $Supp_\alpha(A \Rightarrow C)$, $Conf_\alpha(A \Rightarrow C)$, and $CF_\alpha(A \Rightarrow C)$ be, respectively, the support, confidence and certainty factor of the crisp association rule $Supp_\alpha(A \Rightarrow C)$ at restriction level α.

Following the definitions in section 2.2, we must define the measures of interest and accuracy of a fuzzy association rule in terms of evaluation of quantified sentences. Hence, we define the support, confidence and certainty factor of the fuzzy association rule $A \Rightarrow C$ as RL-numbers, according to the RL-evaluation process proposed in [21].

Definition 5 *Let m_R be the basic probability assignment for association rule $R : A \Rightarrow C$ at each restriction level. The support of the fuzzy association rule $A \Rightarrow C$, RL-Supp$(A \Rightarrow C)$, is given by*

$$RLSupp(A \Rightarrow C) = \sum_{(A \Rightarrow C)_\alpha \in \Omega_R} m_R((A \Rightarrow C)_\alpha) \cdot Supp_\alpha(A \Rightarrow C) \quad (14)$$

Definition 6 *The confidence of the fuzzy association rule $R : A \Rightarrow C$, RL-Conf$(A \Rightarrow C)$, is given by*

$$RLConf(A \Rightarrow C) = \sum_{(A \Rightarrow C)_\alpha \in \Omega_R} m_R((A \Rightarrow C)_\alpha) \cdot Conf_\alpha(A \Rightarrow C) \quad (15)$$

Definition 7 *The certainty factor of the fuzzy association rule $R : A \Rightarrow C$, RL-CF$(A \Rightarrow C)$, is given by*

$$RLCF(A \Rightarrow C) = \sum_{(A \Rightarrow C)_\alpha \in \Omega_R} m_R((A \Rightarrow C)_\alpha) \cdot CF_\alpha(A \Rightarrow C) \quad (16)$$

These measures, RL-Supp$(A \Rightarrow C)$, RL-Conf$(A \Rightarrow C)$ and RL-CF$(A \Rightarrow C)$, can be viewed as the summarization of, respectively, the support, confidence, and certainty factor of every crisp association rule $A \Rightarrow C$ obtained at each restriction level. In fact, if the considered association rule has no negated items involved, the following propositions (as generalizations of the one proposed in [21]) hold.

Proposition 1 *The RL-support of the fuzzy association rule $A \Rightarrow C$, RL-Supp$(A \Rightarrow C)$, is equal to the support of the rule as computed in [16], using the method GD proposed in [22].*

Proposition 2 *The RL-confidence of the fuzzy association rule $A \Rightarrow C$, RL-Conf$(A \Rightarrow C)$, is equal to the confidence of the rule as computed in [16], using the method GD proposed in [22].*

Let us remark that these same propositions for RL-Supp and RL-Conf are not applicable for RL-CF since, again according to [16], the certainty factor for a fuzzy association rule was still computed in the same way as for a crisp one. Thus, we are defining a new measure of accuracy for fuzzy association rules, via restriction levels. The derived properties for this new measure will be the object of study in a separate paper.

Finally, let us notice that, due to the followed RL-representation, and opposite to the fuzzy sets case, it is not possible to find itemsets containing both items i and $\neg i$, which not only is desirable but also sounds logical.

5 Experiments

5.1 Implementation aspects

In our implementation, we start from the same Apriori-based algorithm proposed in [16], with the exception that we must compute the measures of interest and accuracy in terms of evidence in restriction levels, following definitions 5, 6, and 7.

Additionally, we must take into account that now we consider also absent/negated items, which result in a larger number of itemsets to be computed. I.e., if the original Apriori algorithm [1] takes into account up to 2^n itemsets, in our proposal we must consider all the combinations of items and negated items (except those impossible cases containing both i and $\neg i$), resulting in a total number of $(n+1) \cdot 2^n \approx n \cdot 2^n$ itemsets. In each iteration, we must count the occurrences of items in the set of transactions at each restriction level, but also the absences of them. This results in that when considering combinations of items and negated items, we must correctly compute, for each restriction level, the correct support for each itemset. I.e., $supp(\{\neg i_1, \neg i_2\}) = 1 - (supp(i_1) + supp(i_2) - supp(\{i_1, i_2\}))$, $supp(\{i_1, \neg i_2\}) = supp(i_1) - supp(\{i_1, i_2\})$, and so on. One important remark is that we need to keep all the generated itemsets, in order to have access to their support values when necessary. This fact increases the amount of necessary memory, and delays the threshold pruning (by minimum support, etc.) until the second stage of the algorithm (association rule extraction).

According to this, the final efficiency of the algorithm is approximately that of classic Apriori algorithm, but multiplied by $(n+1)$ (due to negated items consideration) and by k (being k the number of restriction levels considered). One of our pending tasks will be the study and development of a more efficient algorithm. In this sense, one choice is that of parallelize the overall process. Since each one of the restriction levels is a crisp set, we can reduce the problem to that of concurrently extract crisp association rules at each level, and then aggregate the obtained measures. We will address this aspect in a future paper.

5.2 Experimentation results

In order to test the resulting algorithm, we first performed our experiments over artificially generated set of transactions. We randomly generated a set of transactions, involving 100 items (and their negations), considered $k = 10$ restriction levels, and obtained a total number of 39600 fuzzy association rules. For space saving purposes, we restrict the number of items in a rule to 2, 1 in the antecendent and 1 in the consequent. In forthcoming papers, we will afford the study of more complex rules.

Each fuzzy association rule has a related set of (crisp) association rules, present in each restriction level. Table 2 shows an example subset of the obtained rules, after establishing a threshold of minimum support and minimum certainty factor of 0.5. Note that, after establishing these thresholds, a rule

ISBN: 978-989-95079-6-8

Table 1: Example of extracted fuzzy association rules

$id\#$	Rule	Evidence	RL-Supp	RL-CF
#39283	$\neg i_{90} \Rightarrow \neg i_{92}$	0.4	0.31	0.31
#39295	$\neg i_{90} \Rightarrow \neg i_{95}$	0.4	0.30	0.25
#39347	$\neg i_{91} \Rightarrow \neg i_{92}$	0.4	0.31	0.31
#39359	$\neg i_{91} \Rightarrow \neg i_{95}$	0.1	0.05	0.05
#39463	$\neg i_{93} \Rightarrow \neg i_{95}$	0.5	0.38	0.36
#39503	$\neg i_{94} \Rightarrow \neg i_{95}$	0.5	0.38	0.36
#39519	$\neg i_{94} \Rightarrow \neg i_{99}$	0.6	0.45	0.38
#39591	$\neg i_{97} \Rightarrow \neg i_{99}$	0.5	0.38	0.37
#39599	$\neg i_{98} \Rightarrow \neg i_{99}$	0.5	0.38	0.37

may not appear in all restriction levels. For sake of simplicity and space saving matters, only the rule number is shown. See Table 1 for a more complete description of some of the rules.

Applying definitions 5 and 7, we can summarize the support and certainty factor of the rules in the restriction levels, and compute the measures for the fuzzy association rules. Table 1 shows some of the obtained fuzzy association rules. The rule numbers correspond to the referred ones in previous Table 2.

In addition, another interesting issue of our proposed methodology is that, in terms of the restriction levels evidence, we can summarize also the obtained set of fuzzy association rules. Actually, we can reduce the number of rules present at each restriction level, by discarding those with low support and/or low certainty factor. After this pruning, we can compute the associated probability distribution (see definition 10) for each rule, and then interpret the resulting set of association rules in terms of evidence. For example, the association rules shown in Table 2 could be summarize according to their evidence in the following expression,

$$\begin{aligned} Ruleset = \{\#39519\}/0.6 + \\ \{\#39463, \#39503, \#39591, \#39599\}/0.5 + \\ \{\#39283, \#39295, \#39347\}/0.4 \end{aligned} \quad (17)$$

Again, for space saving purposes, we only refer to the rules by its rule number. Let us remark that this expression is not actually a fuzzy set, but still can be very helpful in order to interpret the set of results, as we relate a relevance degree, the basic probability assignment, to each rule or set of rules.

Finally, we applied our proposed methodology on real data, obtained from a database containing soil information (see [23] for a more complete description). We reduce the set of data to those attributes modeled as fuzzy quantities ($Averagerainfall$, $Altitude$, $Depth$, PH, and attributes describing the soil chemical composition). A set of linguistic labels $\{Low, Medium, High\}$ was defined for every attribute. Table 3 shows some of the obtained fuzzy association rules, all having $Evidence = 1.0$, restraining the thresholds of minimum support to 0.5 and minimum certainty factor to 0.7, for space saving purposes.

Let us notice that considering absent items allows us to improve the semantics of the rules. I.e., considering item $\neg Avg.Rainfall = Low$ allows us to represent both $Avg.Rainfall = Medium$ and $Avg.Rainfall = High$ concepts. This aspect will be addressed in more detail in a forthcoming paper.

Table 3: Example of fuzzy association rules on real data

Rule	RL-Supp	RL-CF
$PH = High \Rightarrow \neg Avg.Rain. = Low$	0.616	0.97
$PH = Low \Rightarrow \neg Avg.Rain. = Medium$	0.616	0.97
$PH = High \Rightarrow \neg Avg.Rain. = Medium$	0.616	0.97
$PH = High \Rightarrow \neg Avg.Rain. = High$	0.616	0.97
$PH = Medium \Rightarrow \neg Avg.Rain. = Low$	0.616	0.97
$PH = Low \Rightarrow \neg Avg.Rain. = Low$	0.616	0.97
$PH = Medium \Rightarrow \neg Avg.Rain. = Medium$	0.616	0.97
$PH = Low \Rightarrow \neg Avg.Rain. = High$	0.616	0.97
$PH = Medium \Rightarrow \neg Avg.Rain. = High$	0.616	0.97
$\neg Alt. = Low \Rightarrow \neg Avg.Rain. = High$	0.810	0.83
$\neg Alt. = Medium \Rightarrow \neg Avg.Rain. = High$	0.810	0.83
$\neg Alt. = High \Rightarrow \neg Avg.Rain. = Low$	0.810	0.83
$\neg Alt. = High \Rightarrow \neg Avg.Rain. = High$	0.810	0.83
$\neg Alt. = Medium \Rightarrow \neg Avg.Rain. = Low$	0.810	0.83
$\neg Alt. = High \Rightarrow \neg Avg.Rain. = Medium$	0.810	0.83
$\neg Alt. = Low \Rightarrow \neg Avg.Rain. = Medium$	0.810	0.83
$\neg Alt. = Medium \Rightarrow \neg Avg.Rain. = Medium$	0.810	0.83
$\neg Alt. = Low \Rightarrow \neg Avg.Rain. = Low$	0.810	0.83

6 Concluding remarks

We have introduced a new methodology to represent fuzzy association rules in terms of RL-sets, as an alternative representation to that of fuzzy sets. This methodology offers some new advantages as, for example, it allows us to properly consider issues as fuzzy association rules involving absent items. As far as we know, ours is a novel methodology, combining both presence and absence of items with imprecision and uncertainty management. Considering both presence and absence of items can lead us to discern possible conflicts between items or detect when a set of items complements another.

In addition, it also offers a method of summarizing the set of obtained rules, representing them in a format similar to that of fuzzy sets, and thus allowing us to help the user to discern the relevance of a given rule or set of rules, along with the usual measures of interest and accuracy. With respect to these measures, we have seen how the resulting measures values are the same to those obtained in the fuzzy set approach. In this sense, our approach is still coherent with respect to previous ones. Nevertheless, in the general case, and specially in the case of the measure of certainty factor, new challenges appears abroad, and the derived properties for these new definitions for measuring the accuracy in rules will be approached in a future paper.

Finally, we have extended an existing Apriori-based algorithm in order to consider a fuzzy association rule as the aggregation (RL-representation) of the related association rules

ISBN: 978-989-95079-6-8

Table 2: Example of (crisp) association rules distributed through restriction levels

RL \ *Rule*	#39283	#39295	#39347	#39359	#39463	#39503	#39519	#39591	#39599
1.0	×		×		×	×	×	×	×
0.9	×	×	×		×	×	×	×	×
0.8	×	×	×		×	×	×	×	×
0.7	×	×	×		×	×	×	×	×
0.6		×			×	×	×	×	×
0.5							×		
0.4				×					

at each restriction level. As the overall performance is affected by the higher number of itemsets to be taken into account, one pending task will be the optimization of this algorithm in order to improve its efficiency.

Acknowledgment

This work was developed as part of the research project TIN2006-07262.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. Of the 1993 ACM SIGMOD Conference*, pages 207–216, 1993.

[2] Craig Silverstein, Sergey Brin, and Rajeev Motwani. Beyond market baskets: Generalizing association rules to dependence rules. *Data Min. Knowl. Discov*, 2(1):39–68, 1998.

[3] Maria-Luiza Antonie and Osmar R. Zaane. Mining positive and negative association rules: an approach for confined rules. In *In European PKDD Conf*, pages 27–38, 2004.

[4] Xindong Wu, Chengqi Zhang, and Shichao Zhang. Efficient mining of both positive and negative association rules. *ACM Transactions on Information Systems*, 22(3):381–405, July 2004.

[5] Wei-Guang Teng, Ming-Jyh Hsieh, and Ming-Syan Chen. On the mining of substitution rules for statistically dependent items. In *ICDM*, pages 442–449. IEEE Computer Society, 2002.

[6] Ashok Savasere, Edward Omiecinski, and Shamkant Navathe. Mining for strong negative associations in a large database of customer transactions. pages 494–502. IEEE Computer Society, 1998.

[7] Xiaohui Yuan, Bill P. Buckles, Zhaoshan Yuan, and Jian Zhang. Mining negative association rules. *Computers and Communications, IEEE Symposium on*, 0:623, 2002.

[8] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[9] W.H. Au and K.C.C. Chan. Mining fuzzy association rules. In *Proc. Of 6th Int. Conf. On Information and Knowledge Management. Las Vegas, NV, USA*, pages 209–215, 1997.

[10] A.W.C. Fu, M.H. Wong, S.C. Sze, W.C. Wong, W.L. Wong, and W.K. Yu. Finding fuzzy sets for the mining of fuzzy association rules for numerical attributes. In *Proc. Int. Symp. On Intelligent Data Engineering and Learning (IDEAL'98), Hong Kong*, pages 263–268, 1998.

[11] S. Ben-Yahia and A. Jaoua. A top-down approach for mining fuzzy association rules. In *Proc. 8th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'2000*, pages 952–959, 2000.

[12] Shu-Yue-J, Tsang-E, Yenng-D, and Daming-Shi. Mining fuzzy association rules with weighted items. In *Proc. IEEE Int. Conf. On Systems, Man and Cybernetics*, 2000.

[13] D. Dubois, H. Prade, and T. Sudkamp. A discussion of indices for the evaluation of fuzzy associations in relational databases. In T. Bilgic, B. De Baets, and O. Kaynak, editors, *Fuzzy Sets and Systems - Proceedings IFSA'03, LNAI 2715*, pages 111–118. Springer, 2003.

[14] D. Sánchez, M. Delgado, and M.A. Vila. A restriction level approach to the representation of imprecise properties. In *IPMU 2008*, pages 153–159, 2008.

[15] M. Delgado, M.D. Ruiz, and D. Sánchez. A restriction level approach for the representation and evaluation of fuzzy association rules. Submitted to IFSA'2009, 2009.

[16] M. Delgado, N. Marín, D. Sánchez, and M.A. Vila. fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems 11(2)*, pages 214–225, 2003.

[17] F. Berzal, I. Blanco, D. Sánchez, and M.A. Vila. Measuring the accuracy and interest of association rules: A new framework. *Intelligent Data Analysis*, 6:221–235, 2002.

[18] E. Shortliffe and B. Buchanan. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23:351–379, 1975.

[19] L.A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computing and Mathematics with Applications*, 9(1):149–184, 1983.

[20] D. Sánchez, M. Delgado, and M.A. Vila. RL-numbers: An alternative to fuzzy numbers for the representation of imprecise quantities. In *FUZZ-IEEE 2008*, pages 2058–2065, 2008.

[21] D. Sánchez, Delgado M., and M.A. Vila. Fuzzy quantification using restriction levels. Submitted to WILF'2009, 2009.

[22] M. Delgado, D. Sánchez, and M.A. Vila. Fuzzy cardinality based evaluation of quantified sentences. *International Journal of Approximate Reasoning*, 23:23–66, 2000.

[23] J. Calero, G. Delgado, M. Sánchez-Marañón, D. Sanchez, MA Vila, and J. Serrano. An experience in management of imprecise soil databases by means of fuzzy association rules and fuzzy approximate dependencies. In *Proceedings of the 6th international conference on enterprise information systems, ICEIS*, volume 4, pages 138–146. Springer, 2004.

ISBN: 978-989-95079-6-8

Towards an interpretation of the connectivity analysis in the frame of fuzzy equivalence relations for medical diagnoses

Tatiana Kiseliova

Institute of Computer Sciences
Faculty of Exact and Natural Sciences
Javakhishvili Tbilisi State University
Email: t.kiseliova@web.de

Abstract— *Connectivity analysis methodology is suitable to find representative symptoms of a disease. This methodology describes connections between symptoms in particular way and then chooses the group of symptoms that have the high level of connection, or in other words, have strong interconnections between elements of a group. In this paper we investigate the analogy between connectivity analysis and cluster analysis based on fuzzy equivalence relations. A comparison of two approaches, one of which has strong theoretical background (cluster analysis based on fuzzy equivalence relations) and more practically oriented connectivity analysis assures more convincing and accurate connectivity analysis from one side and applicability of fuzzy equivalence relations for medical diagnoses from another. Connectivity analysis, as shown in the paper, is one of the clustering methods, can be used in many applications where feature selection and extraction problem is considered, in particular, in pattern recognition and image processing. The results of the comparison are demonstrated on the examples.*

Keywords— connectivity, equivalence relations, fuzzy cluster analysis.

1 Introduction

When speaking about the physician intuition one means the ability of a doctor to establish a diagnosis for a patient. This process can be separated into different stages. One of them, considered in this paper, distinguishes an experienced doctor from a beginner: the expert uses a small amount of observations to make a conclusion about a patient's state. Calling these observations representative symptoms one highlights their high discrimination power between this and other diseases, their best characterization for a disease. "The best characterization" means: if a patient has this representative group of symptoms, his diagnosis is almost sure. Fuzzy notations such as "small amount", "almost sure" show - as was already mentioned in many scientific papers devoted to applications of fuzzy logic in medical domains - that definitions, reasoning and conclusions in medicine are usually done with uncertainty. And a task - to find the representative symptoms - is not an exception.

For reasons of simplification we use *symptoms* as a general concept for clinical parameters, different manifestations (signal, clinical), meanwhile *diseases* include disorders, different diagnostic hypotheses.

The problem of finding the representative symptoms can be solved within a computer-assisted medical diagnosis system. Ideally, an established set of representative symptoms for a particular disease allows to only partly examine a patient suspected for this disease, that, in turn, reduces costs, time, etc. Different groups of representative symptoms can be useful at the different stages of a chronic disease. Representative symptoms can help to establish an overall risk for a patient attacked from a particular disaster. For example, in life insurance medicine [16] the mortality of applicants within the period of insurance is assessed on the basis of present risk factors. Another possible application of representative symptoms is an optimization of questionaries, e.g., in screening [5].

There are at least two ways to get the representative symptoms: to ask an experienced physician or to use some methods to extract this set from the available information.

As was already mentioned, symptoms cannot normally be described in simple "black and white" terms with "yes/no" or "present/not present" answers. An expert seeking to comprehensively describe the multiple influencing factors would very quickly reach the limit of his capacity [7]. For example, if an expert should estimate a group of three symptoms, each of which has three values, he will face 27 possible states. If the number of symptoms and the number of values rise to five, the number of states increase to 3125. Moreover, some disease categories overlap. For example, no consensus exists on clinical or epidemiologic measures that can be used to classify patients based on asthma symptoms [17]. Due to the highly variable nature of asthma, classification of patients into mild, moderate and severe disease categories is necessarily imperfect. That categories may overlap means, that trying to find the representative symptoms for three categories of a disease (mild asthma, moderate asthma and severe asthma) we do not have a unique description for each of them. The USA National Asthma Education Program of guidelines suggest symptoms (wheeze, dyspnoea, cough, sputum production, allergy characteristics, etc.), the degree of airflow obstruction and frequency of use of oral glucocorticoids (never, infrequently for attacks, frequently for attacks, and daily use) to validate three asthma categories. Due to the symptom-disease-patient terminology proposed by us the representative group of symptoms could be chosen from *symptoms* (the medication use, spirometry, current symptoms) for three *diseases* (indicators of asthma severity) for this example. Therefore it is not an easy task to get representative symptoms from an expert for building a computer-assisted medical system. A specialist uses his unconscious knowledge, due to the professional intuition and acumen.

To rely on data from the patient records only (or other sources containing information about patients), using statistical methods, seems also to be not a doubtless way. Widely recognized drawbacks of statistical estimations is the demand on high amounts of data. In other situations, conclusions de-

ISBN: 978-989-95079-6-8

duced from records do not correspond to results from other sources. For example, as was reported in [17], a severity of asthma induced from oral glucocorticoids use and spirometry does not correlate with current asthma symptoms from medical records.

A combination of approaches is a matter of investigations since years [2]. Especially since the theory of fuzzy sets was proven to be useful in modeling uncertainties, in particular, in medicine. Expert knowledge is composed of the evaluations of the observed symptoms (the influencing factors) and the rules for the combinations of these evaluations. The expert formulates and describes his decision in a linguistic form in order to render it comprehensible to a layperson [16, 20]. Often, he/she demonstrates only certain key values of the influencing factors, the representative symptoms.

The complexity of a method partially depends on the initial information available. For example, if we have a collection of rules, where all symptoms and combinations of symptoms are already estimated by experts as was done in [18] for pulmonary tuberculosis with some degree of confirmation and exclusion, the task to find the representative symptoms may be easily solved. Experts are asked to establish thresholds for an exclusion and a confirmation of symptoms and this will allow to classify the symptoms to the corresponding representative groups.

The medical records is a point of departure for connectivity analysis as well, but values in these records represent some degree of expressiveness of a symptom, or the degree of compatibility between patients and symptoms. For example, for "overweight" the expert may put a number between 0 and 1 to reflect his opinion about this patient's symptom. The expert may know the exact weight of a patient or not, his estimation 0 shows that the patient has no overweight, 1 - the patient is very thick, 0.2 -the patient is almost in norm. The "ruddy complexion" is another example for the scale representation, e.g., $0.2, 0.4, 0.8$ would mean different degrees of a skin face color estimated by an expert.

The connectivity analysis has been intensively applied, in particular, in medicine, management, geology [4, 10, 11, 12, 13, 22]. In most cases its practical usefulness was proven by examples. But in general, the connectivity analysis can be considered as a part of pattern recognition problem, in particular, cluster analysis. Till now such relations were not established. Our intention in this paper is to show that there exists an analogy between the connectivity analysis and fuzzy cluster analysis based on fuzzy equivalence relations. This comparison (or, in some sense, an interpretation of connectivity analysis in terms of well-established cluster analysis) brings more insights into the connectivity analysis which is practically used but has some weaknesses in the theoretical background from one side and applicability of fuzzy equivalence relations for medical diagnoses from another.

To realize our intentions, the paper is organized as follows. In the next section we describe what we understand under connectivity analysis and how it is applied to the representative symptoms mining problem. Section 3 describes elements of a fuzzy cluster analysis based on fuzzy equivalence relations. In Section 4 the analogy between two approaches are considered and the paper is concluded by final remarks.

2 Connectivity between symptoms

Assume from the database of already diagnosed patients, *suffering from a particular disease*, the table $R_{SP} = \{r_{ij}\}$, $i = 1 \dots n$, $j = 1 \dots m$ is obtained where only presence or absence of symptoms are ticked. For example, in Table 1 information about 5 patients who have been suffering from a particular disease with 4 corresponding predefined symptoms is presented. Compared with a real practical situation, the description of e.g., *pulmonary tuberculosis (PT)* [18] consists of 30 symptoms.

Table 1: An initial table for a symptom-patient connection

	p_1	p_2	p_3	p_4	p_5
s_1	1	0	1	0	0
s_2	1	1	0	1	0
s_3	1	1	1	1	1
s_4	0	1	0	1	1

2.1 The positive level of connection

Assume, Table 1 defines the relation of connectivity between a set of patients $P = \{p_1, p_2, p_3, p_4, p_5\}$ and a set of symptoms $S = \{s_1, s_2, s_3, s_4\}$. $r_{ij} = 1$ denotes that person p_j has the symptom s_i and $r_{ik} = 0$ denotes that person p_k does not have it.

Dealing with these data, the intuitive proposal to get a group of representative symptoms is to highlight those symptoms that often meet under description of a patients' condition. For example, from Table 1 it can be seen that symptom s_3 is present by all patients, symptoms s_1 and s_2 have in common only p_1. Speaking language of connectivity, the representative symptoms can be naturally chosen based on the level of connection among symptoms that is equal to the number of common patients affected by those symptoms. For example, symptoms s_1 and s_2 have in common only p_1 and hence they are connected at level 1, meanwhile symptoms s_2 and s_3 have in common patients p_1, p_2 and p_4 , i.e., they are connected at level 3. Formula (1) describes pairwise connectivities between symptoms.

$$Q_s = R_{SP} R_{SP}^{\mathsf{T}} \tag{1}$$

In (1) ordinary multiplication of matrices is used.

The result of (1) is matrix (2). Clearly, Q_s is diagonal symmetric and can be represented in upper triangular form: the order of symptoms is not important.

$$Q_s = \begin{pmatrix} 2 & 1 & 2 & 0 \\ 1 & 3 & 3 & 2 \\ 2 & 3 & 5 & 3 \\ 0 & 2 & 3 & 3 \end{pmatrix} \tag{2}$$

Connectivity results can be represented as shown in Table 2.

0-connection is not considered. The symptom s_3 can be taken as the representative one, but the groups of symptoms are of interest from the physician point of view: one symptom, for instance, a cough, can be present due to many diseases, and cannot be distinguishable for them.

ISBN: 978-989-95079-6-8

Table 2: Pairwise connections from matrix (2)

Levels of connectivity, q	Groups of symptoms
5	$\{s_3\}$
3	$\{s_2\}\{s_2, s_3\}\{s_3, s_4\}, \{s_4\}$
2	$\{s_1\}\{s_1, s_3\}\{s_2, s_4\}$
1	$\{s_1, s_2\}$

Sets $\{s_2, s_3\}$ or $\{s_3, s_4\}$ can be considered as appropriate ones. They have, however, in common one element, s_3. A fusion of these two sets into $\{s_2, s_3, s_4\}$ would mean, that the same number of patients (i.e., 3) have in common subsets of this set.

Therefore, let us transform Table 2 into Table 3 due to the following observations:

1. if n patients have in common a symptom (or symptoms), $n-1$ patients have this symptom(s), too. The same is for $n-1$, $n-2$, etc. This property we call inheritance.
2. if the same symptoms belong to different groups at the same level of connectivity, these groups can be fused.

Table 3: After fusion and inheritance in Table 2

Levels of connectivity, q	Groups of symptoms
5	$\{s_3\}$
4	$\{s_3\}$
3	$\{s_2, s_3, s_4\}$
2	$\{s_1, s_2, s_3, s_4\}$
1	$\{s_1, s_2, s_3, s_4\}$

Information from Table 3 can be interpreted. For instance, symptoms $\{s_2, s_3, s_4\}$ are more representative for the considered disease, because they are at a high level of connectivity i.e., have in common many patients and the cardinality of this set is not small. This intuitive procedure can lead, for example, to similar groups of representative symptoms, i.e., at the same level, two groups and the same number of different symptoms in each group. In this case both groups can be considered as representative ones or a consultation with a physician is supposed.

So far so good, but even in this example part of the information used is lost: for instance, the absence of the symptoms are not taken into account. In Table 4,

Table 4: s_1 is more like s_2 than s_3

	p_1	p_2	p_3	p_4
s_1	1	0	1	1
s_2	1	0	1	1
s_3	1	1	1	1

s_1 and s_3 are connected (via p_1, p_3 and p_4) at the same level ($q = 3$) as s_1 and s_2 (via p_1, p_3 and p_4). But obviously, s_1 is more like s_2 than s_3.

2.2 *The negative level of connection*

Due to the tradition of expert systems in medical diagnoses [3, 6, 8, 18, 19, 21] it is important to consider not only an occurrence of symptoms, but a non-occurrence as well.

Let us now get information from the non-occurrence of symptoms in Table 1. For each pair of symptoms the number of patients, that do *not* have in common these symptoms is calculated. Formally, formula (1) can be used here, where matrix R_{SP} is substituted by $\Omega - R_{SP}$, where Ω is a matrix with all elements unity. The results of negative connectivity is represented in Table 5, where q^- are levels of connectivity.

Table 5: Pairwise connections of symptoms that patients have *no* in common from Table 2

q^-	Groups of symptoms
3	$\{s_1\}$
2	$\{s_2\}, \{s_4\}$
1	$\{s_1, s_2\}\{s_2, s_4\}$
0	$\{s_3\}, \{s_3, s_4\}, \{s_2, s_3\}, \{s_1, s_3\}, \{s_1, s_4\}$

The results of this table can be interpreted. For example, the lowest row shows, that there are no patients, that do not have in common $\{s_3\}$, or $\{s_3, s_4\}$, or $\{s_2, s_3\}$, or $\{s_1, s_3\}$, or $\{s_1, s_4\}$, i.e., all patients have at least one of these sets in common. This leads to the assumption, that these groups of symptoms can be considered as candidates to be representative ones. $\{s_1\}$ does not occur by 3 patients, and therefore $\{s_1\}$ can be considered as a representative non-occurrence symptom for the particular disease (if we assume, of course, that one symptom can be representative for non-disease; this case, as a corresponding one from the positive connectivity should be discussed with an expert).

Intuitively it is also clear that if $\{s_1\}$ does not occur by each of three patients, it does not occur by two or one of them. The procedures of inheritance and fusion are the same as in the case of positive connectivities.

The final set of representative symptoms can be differently established. A possible way is to combine representative symptoms for a disease and for non this disease. For example, if a patient has $\{s_2, s_3, s_4\}$ and does not have $\{s_1\}$ we can conclude that the patient has this disease.

2.3 *Some remarks, concerning connectivity analysis*

We did not consider in the previous sections cases when the elements of the matrices R_{SP} take the values from $[0, 1]$ or, moreover, are represented by linguistic terms. In this paper we restrict ourselves to the "crisp" cases to concentrate on the explanation of the idea of the connectivity analysis to demonstrate a way to find similarities between symptoms due to their presence by patients. If, however, the elements of R_{SP} are numbers from $[0, 1]$, they are interpreted as a strength of connection (or compatibility) between a patient and a disease.Q_s is a composition of fuzzy relations. Elements of Q_s can be interpreted as a pairwise strength between symptoms for patients.

It is easy to see that absolute values of the connectivity levels are not decisive for choice of the group of representative

ISBN: 978-989-95079-6-8

symptoms. A corresponding ordinal scale can bear the necessary information for connectivity. It means that for particular calculations we can simply linearly transform obtained connectivity levels into corresponding numbers from $[0, 1]$.

Above we have used a fairly fuzzy procedure to find the representative symptoms from positive and negative connectivities. The main demands are "high level of connection" and rather "wide" set of symptoms, connected at this level. In some sense, more formal algorithm would be preferable.

In the next section we consider if it is possible to formulate the connectivity analysis by similarity relations and corresponding fuzzy cluster analysis to assure more convincing and accurate connectivity analysis.

This idea appears by an intuitive analogy between levels of connections and α-cuts of fuzzy relations. Let us first recall several definitions used to build such analogy.

3 Cluster analysis based on fuzzy equivalence relations

It is known that every fuzzy equivalence relation (sometime called a similarity relation) induces a crisp partition of its α-cuts. And therefore, fuzzy clustering problem can be viewed as the problem of identifying the appropriate fuzzy equivalence relation on given data.

Definition 3.1 *Let R be a fuzzy relation, $R : X \times Y \to [0, 1]$, i.e., $R = \{((x, y), R(x, y))|(x, y) \in X \times Y\}$, the α-cut matrix ${}^{\alpha}R$ is denoted by*

$${}^{\alpha}R = \{(((x, y), {}^{\alpha}R(x, y))|{}^{\alpha}R(x, y) = 1 \text{ if } R(x, y) \geq \alpha;$$
$${}^{\alpha}R(x, y) = 0 \text{ if } R(x, y) < \alpha, (x, y) \in X \times Y, \alpha \in [0, 1]\}$$

Definition 3.2 *Binary fuzzy relation $R : X \times X \to [0, 1]$ is a fuzzy equivalence relation iff it is reflexive, i. e., $R(x,\ x) = 1$; symmetric, i. e., $R(x,\ y) = R(y,\ x)$, and max-min transitive: $R^{(2)} = R \circ R \subset R$ or, more explicitly $R(x,\ z) \geq \max_y \left\{ \min_{x,z}\{R(x,\ y),\ R(y,\ z)\} \right\}$, $\forall x, y, z \in X$*

For practical tasks it is easier first to build a fuzzy compatibility relation and then calculate a transitive closure of this compatibility relation and this way to complete identification of a fuzzy equivalence relation.

Definition 3.3 *A fuzzy relation R on $X \times X$ is called a fuzzy compatibility relation if it satisfies reflexive and symmetric conditions.*

Definition 3.4 *The transitive closure R_T of a fuzzy relation R is defined as the relation that is transitive, contains R ($R_T \supset R$) and has the smallest possible membership grades.*

Theorem 3.1 *[14] Let R be a fuzzy compatibility relation on a finite universal set X with $|X| = n$, then the $\max - \min$ transitive closure of R is the relation $R^{(n-1)}$.*

4 Connectivity analysis in the frame of fuzzy equivalence relations

Ideally, if Q_s (see (1)) would be a similarity relation one can easy see that levels of connections correspond to α-cuts. Thus, the analogy between connectivity analysis and a fuzzy clustering based on fuzzy equivalence relations is established.

NormallyQ_s is symmetric, but its reflexivity and transitivity are questionable. To reach these two properties one can do as follows.

First, as was already told in the section 2.3, we can substitute the elements of Q_s by the corresponding ones from interval $[0, 1]$. Due to the definition of ε-reflexivity, i.e. $R(x,\ x) \geq \varepsilon\ x \in X$, $0 < \varepsilon < 1$ [14], Q_s is ε-reflexive. Thus, Q_s is a fuzzy compatibility relation.

Transitive closure of Q_s can be find due to the following procedure [15] based on the results of the Theorem 3.1: find the transitive closure Q_T of fuzzy compatibility relation Q_s.

> Step 1: Calculate $Q^{(2)}$ if $Q^{(2)} = Q_s$ or $Q^{(2)} \subset Q_s$, then transitive closure $Q_T = Q_s$ and stop. Otherwise, $k = 2$, go to step 2.
>
> Step 2 : If $2^k \geq n - 1$, then $Q_T = Q^{(n-1)}$ and stop. Otherwise, calculate $Q^{(2^k)} = Q^{(2^{k-1})} \circ Q^{(2^{k-1})}$, if $Q^{(2^k)} = Q^{(2^{k-1})}$, then transitive closure $Q_T = Q^{(2^k)}$ and stop. Otherwise, go to step 3.
>
> Step 3: $k = k + 1$, go to step 2.

Due to this procedure a transitive closure of Q_s is as follows (mention a reverse transformation of the previous scale):

$$Q_T = \begin{pmatrix} 2 & 2 & 2 & 2 \\ 2 & 3 & 3 & 3 \\ 2 & 3 & 5 & 3 \\ 2 & 3 & 3 & 3 \end{pmatrix} \tag{3}$$

Following Definition 3.1 the relation (3) induces the partitions of its α-cuts represented in the Table 6:

Table 6:

α-cuts	Groups of symptoms
5	$\{s_3\}$
3	$\{s_2, s_3, s_4\}$
2	$\{s_1, s_2, s_3, s_4\}$
1	$\{s_1, s_2, s_3, s_4\}$

It can be easy see, that these results coincide with results of connectivity analysis represented in the Table 3. Relation (3) represents a transitive closure of Q_s, and the elements of this relation differ from corresponding elements of Q_s. A similar procedure can be done for "negative connectivity" analogy.

Notice, that recently there exist other methods, that transform the initial fuzzy binary relation into reflexive, symmetric and T-transitive fuzzy binary relation, i.e., a similarity relation, that contains the initial relation [9].

Although results from Table 6 satisfy our initial intuition, "the transitive closure method carries a number of major problems, such as the need of storage and computer time required; ... there is no way of controlling the distortion that this method produce on the data sample"[23].

Several possibilities exist to find a more efficient way than this transitive closure method does. For instance, an application of the representation theorem [23] no longer requires reflexive and symmetric fuzzy relation as initial data.

ISBN: 978-989-95079-6-8

Thus, to build a similarity relation from an initial fuzzy relation, can be done differently. Our approach, based on the ε-reflexivity, and transitive closure, return a similarity relation, which α-cuts coincides with group of symptoms from the connectivity analysis. It was checked for data considered in different applications [4, 12, 13, 22] and gives the same results. This is the approach used thus far to establish the validity of the proposed analogy between connectivity analysis and the fuzzy cluster analysis based on fuzzy equivalence relations. Notice that this analogy was found for low dimensional data, large-scale data is a subject of future investigations.

5 Concluding remarks

Let us summarize what we have done. We have shown how connectivity analysis can work in the frame of fuzzy equivalence relation. This means, in particular, that due to the above described analogy, levels of connection in the connectivity analysis can be represented as α-cuts. In turn, the methodology to build clusters based on the α-cuts from fuzzy relations is already established. Therefore, the connectivity analysis can be considered as a fuzzy clustering problem based on the fuzzy compatibility relations.

The question is how to obtain a compatibility relation from initial table (see, for instance, Table 1). In our discussions we used Q_S. As was already told, the connectivity analysis belongs to the family of cluster methods, and particularly, it can be considered as one of the hierarchical methods [1]. Therefore other dissimilarity measures between patterns and features (patients and symptoms) can be used and the problem to find the representative symptoms becomes a problem of the feature selection and feature extraction problem in the cluster analysis. A fuzzy compatibility relation can be defined in terms of an appropriate distance function of, e.g., the Minkowski class [1]. The question is in interpretation of this relation. Therefore the next investigation can be dedicated to the construction of Q_S that allows to inherit useful information from the initial table R_{SP} to build representative symptoms.

References

[1] J. Abonyi and B. Feil. *Cluster Analysis for Data Mining and System Identification*. Birkhäuser Verlag AG, 2007.

[2] K.-P. Adlassnig and G. Kolarz. CADIAG-II: computer-assisted medical diagnosis using fuzzy subsets. In M.M. Gupta and E. Sanchez, editors, *Approximate Reasoning in Decision Analysis*, pages 495–505. North Holland Publishing Company, 1982.

[3] K. Bögl, H. Leitich, G. Kolousek, G. Rothenfluh, and K.-P. Adlassnig. Clinical data interpretation in MedFrame/CADIAG-4 using fuzzy sets. *Biomedical Engineering: Applications, Basis and Communications*, 8(6):488–495, 1996.

[4] F. Criado, T. Gachechiladze, N. Jorjiashvili, Z. Khvedelidze, H. Meladze, J. M. Saacutenchez, G. Sirbiladze, and G. Tsertsvadze. Theory of connectivity and apportionment of representative chains in the problem of decision-making concerning earthquake possibility. *International Journal of General Systems*, 32(2):103–121, 2003.

[5] F. Georgsson and P. Eklund. An identification of handling uncertainties within medical screening: a case study within screening for the breast cancer. In *Fuzzy and Neuro-Fuzzy Systems in Medicine*, pages 161–188. CRC Press, Boca Raton, FL, 1998.

[6] L. Godo, R. López de Mántaras, C. Sierra, and A. Verdaguer. MILORD: The architecture and management of linguistically expressed uncertainty. In D. Dubois, H. Prade, and R.R. Yager, editors, *Fuzzy Information Engineering: A Guided Tour of Applications*, pages 357–387. Wiley Computing Publishing, John Wiley and Sons, Inc, 1990.

[7] P.-J. Horgby, R. Lohse, and N.-A. Sittaro. Fuzzy underwriting: an approach of fuzzy logic to medical underwriting. *Actuarial Practice*, 5(1), 1997.

[8] P. R. Innocent and R. I. John. Computer aided medical diagnoses. *Information Sciences*, 162(2004):81–104.

[9] P. Julián-Iranzo. A procedure for the construction of a similarity relation. In *IPMU2008*, 2008.

[10] T. Kiseliova. A theoretical comparison of Disco and CADIAG-II-like systems for medical diagnoses. *Kybernetika*, 42(6):723–748, 2006.

[11] T. Kiseliova and S. Krajči. Generation of representative symptoms based on fuzzy concept lattices. In B. Reusch, editor, *Computational Intelligence. Theory and Applications: International Conference on Computational Intelligence, 8. Fuzzy Days in Dortmund*, pages 349–356, Dortmund, Germany, September 30 - October 2, 2004. Springer-Verlag Heidelberg.

[12] T. Kiseliova and K. Pagava. New approach to estimate different drugs and/or other medical interventions effectiveness based on fuzzy logic principles. *Georgia Medical News*, 3(156):65–68, March 2008.

[13] T. Kiseliova, K. Pagava, H. Pagava, and A. Pagava. The influence of socio-economical factors on health and health-related behaviour of adolescents in Georgia. In *4th International Conference on Children's Health and the Enviroment. Reducing environmental risks for our children*, volume 10 of *in Book of abstracts in: "MEDYCYNA SRODONOWISKOWA" (Environmental Medicine)*, page 135, Vienna, June, 10-12 2007.

[14] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic. Theory and Applications*. Prentice Hall PTR, 1995.

[15] Gin-Shuh Liang, Tsung-Yu Chou, and Tzeu-Chen Han. Cluster analysis based on fuzzy equivalence relation. *European Journal of Operational Research*, 166(2005):160–171, 2005.

[16] R. Lohse. Fuzzy logic in life underwriting as illustrated by the cardiovascular risk assessment of diabetes melitus type I. In *Hannover Rück*, number 11. http://www.hannover-rueck.de, 2003.

ISBN: 978-989-95079-6-8

[17] M. L. Osborne, W. M. Vollmer, K. L. Pedula, J. Wilkins, A. S. Buist, and M. O'Hollaren. Lack of correlation of symptoms with specialist-assessed long-term asthma severity. *Chest*, 115:85–91, 1999.

[18] N. H. Phuong and D. H. Hung. Tuberdiag: an expert system for pulmonary tuberculosis diagnosis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 7(4):371–381, 1999.

[19] F. Puppe. *Problemlösungsmethoden in Expertensystemen*. Springer-Verlag, Berlin, 1990.

[20] M. A. M. Reis, P. S. P. Silveira, and N. R. S. Orteg. Fuzzy expert system in prediction of neonatal resuscitation. *Brazilian Journal of Medical and Biological Research*, 37(5):755–764, May 2004.

[21] E. H. Shortliffe. *Computer-Based Medical Consultations, MYCIN*. Elsevier, North-Holland, New York, 1976.

[22] A. Sikharulidze. Decision-making aiding fuzzy information analysis in investments. *Proc. of Javakhishvili Tbilisi State University, Applied Mathematics, Computer Sciences,*, 353:77–94, 2003.

[23] L. Varverde and S. Ovchinnikov. Representation of t-similarity relations. *Fuzzy Sets and Systems*, 159(2008):2211–2220, 2008.

ISBN: 978-989-95079-6-8

Identification of Web Information using Concept Hierarchies and On-line Updates of Concept Importance

Zhan Li[1], Ronald R. Yager[2] and Marek Reformat[1]

[1]thinkS[2]: thinking software and system laboratory
Electrical and Computer Engineering
University of Alberta, Edmonton, Canada, T6G 2V4

[2]Machine Intelligence Institute
Iona College, New Rochelle, NY 10801

Abstract—*The Internet users have to perform a lot of search to find web pages with relevant information. The paper proposes an approach for utilization of a hierarchy of concepts to perform identification of web pages in the environment of the Semantic Web. A user provides a hierarchy of concepts that can only partially "cover" their domain of interest. Ontologies related to that domain are used to instantiate the hierarchy with concrete information, as well as to enhance it with new concepts initially unknown to the user. A web page is checked against concepts from the hierarchy and activation levels of those concepts are aggregated using Ordered Weighted Averaging (OWA) operators that are part of the hierarchy of concepts. Mechanisms of aggregations embedded in OWA operators are determined using linguistic quantifiers and importance of concepts. The importance of concepts is constantly changing, and in order to automatically assign importance values and "keep track" of changes a new algorithm for Adaptive Assigning of Term Importance (AATI) is used.*

Keywords—*hierarchy of concepts, ontology, ordered weighted aggregation, text identification*

1 Introduction

The enormous growth of the WWW (World Wide Web) creates a serious challenge in discovering relevant information among all web documents. Those documents vary in quality of information they contain, as well as in their timeliness. It is not easy to find the most relevant documents when the repository is constantly changing.

In general, there are two approaches used for finding relevant web pages, i.e., "Information retrieval" (IR) [4] and "Text Categorization" (TC) [8]. IR retrieves documents based on a query using a "keyword-matching" mechanism without considering different meanings of keywords under different contexts. In TC, there is no query, and the knowledge about a given category is embedded in models developed using machine learning (ML) techniques based on documents representing the category.

In real world scenarios, people relay on Search Engines (SEs) to retrieve required information. SEs accept users' queries and return sets of related web documents applying one of two approaches: building a full-text index-based search (IR-like approach), like Google and Altavista; or constructing taxonomies with web documents populating categories (TC-like approach), such as Yahoo.

People identify documents as belonging to a specific category based on a set of keywords. A person looks at a document and "searches" for keywords and associated with them concepts related to a given category. Moreover, for a person all keywords and concepts are interconnected and constitute a network. Activation of one keyword or concept initiates activation of related concepts. Relations between them can be of different nature: some can "point" to more *specific* concepts, some can relate to concepts that are *parts of* the original concept, and some can relate to concepts that *contain* the original one. More keywords and associated concepts are found in the document more support is collected towards assigning this document to a category described by the keywords.

An introduction of the Semantic Web [1] has led to the increased popularity of ontology as a means of representing knowledge in a way that a human being as well as a machine can understand. An ontology is a formal, explicit specification of a shared conceptualization [2]. It is a set of well-defined classes that describe data models in a specific domain. Together with their individuals (instances of classes), ontologies work as knowledge characters to express individual facts [7]. This new representation of knowledge introduced to the web environment brings new possibilities of utilization of information [5,6].

A different format of knowledge representation has been proposed in [10]. The proposed format, called Hierarchy of Concepts (*HOFC*), represents concepts with atomic attributes or other concepts. As the result a graph-like structure is established where each vertex is a concept, and terminal vertices are attributes. The edges of the hierarchy represent relationships that define concepts with a set of other concepts and/or attributes. An *HOFC* can be used for representing any human-like structure of concepts, for example queries [10].

ISBN: 978-989-95079-6-8

The paper introduces a new idea for constructing a category identifier using an *HOFC* combined with ontology and an algorithm for automatic updating of concept importance. In a nutshell, a user provides a simple hierarchy describing their category of interest. This hierarchy contains concepts that are the most obvious for the user. The concepts are linked together via different types of relationships. The concepts that constitute the hierarchy are generic – they are just definitions of concepts. The specific pieces of information are extracted form ontologies and "attached" to the equivalent concepts of the hierarchy as attributes – in other words ontologies are used to populate the hierarchy. Relations existing between ontology classes are used to find new concepts that are added to the hierarchy as well. This provides a mechanism for dynamic enhancement of hierarchies.

The dynamic character of the proposed approach is evident in a process of determining importance of *HOFC* concepts. Concept importance changes accordingly to changes in the contents of web pages – newly posted documents and texts affect importance of keywords in their ability to "identify" a given category. In order to "follow" these changes we equipped our *HOFC* with *AATI* [3]. The *AATI* is an un-supervised algorithm with ability to determine and update concept importance values automatically.

2 Background

2.1 Hierarchy of Concepts

In a nutshell, an *HOFC* uses a notion of representing concepts with atomic attributes or other concepts. Edges of *HOFC* are of significant importance to the whole idea of *HOFC*. If we assume that a concept C_1 is defined, described by or related to other two concepts C_2 and C_3, then the hierarchy will have two edges connecting C_2 with C_1, and C_3 with C_1. The concept C_1 is called a higher-level concept, and C_2 and C_3 are lower-level concepts. This also means that activation of concepts C_2 and C_3 leads to activation of C_1.

*HOFC*s introduce a very important feature – the process of activation of higher-level concepts by lower-level concepts is fully controlled by a user. There are two controlling components: importance vector $\boldsymbol{M}$ and linguistic quantifier Q. The vector $\boldsymbol{M}$ is an indicator of significance of each lower-level concept in defining a higher-level concept. In other words, $\boldsymbol{M}$ determines a weight of each of the participating (lower-level) concepts in a process of identifying an activation level of higher-level concept. The linguistic quantifier Q guides the aggregation of lower-level concept activations levels. We can have such quantifiers as *some*, *most*, *at least half*, *about 1/3*, or *for all*. Both $\boldsymbol{M}$ and Q determine how activation levels of lower-level concepts should be combined using the Ordered Weighted Averaging (*OWA*) operator [9].

In a formal representation, the *OWA* operator, defined on the unit interval I and having dimension n, is a mapping F_w: $I^n \rightarrow I$ such that:

$$F_w(a_1,\ldots,a_n) = \sum_{j=1}^{n} w_j * b_j \qquad (1)$$

where b_j is the j^{th} largest of all arguments $a_1, a_2,\ldots, a_n$, and w_j is a weight such that w_j is in $[0, 1]$ and the sum of w_j is equal 1. If $\boldsymbol{W}$ is an n-dimensional vector whose j^{th} component is w_j and $\boldsymbol{B}$ is an n-dimensional vector whose j^{th} component is b_j then $F_w(a_1, a_2,\ldots, a_n) = \boldsymbol{W}^T\boldsymbol{B}$. In this formulation $\boldsymbol{W}$ is referred to as the *OWA* weighting vector and $\boldsymbol{B}$ is called the ordered argument vector.

2.2 Ontology

The most important aspect of the ontology used for Semantic Web applications is related to identifying two ontology layers: the ontology definition layer, and the ontology instance layer[1].

The *ontology definition layer* represents a framework used for establishing a structure of ontology and for defining classes (concepts)[2] existing in a given domain. A structure of ontology is built based on a relation *is-a* between classes. This relation represents a *subClassOf* connection between a superclass and a subclass.

The ontology definition contains descriptions of classes. Classes are defined using two types of the properties:

- *datatype property* – this type of property focuses on describing features of a class; this property can be expressed as values of such data types as boolean, float, integer, string, and many more (for example, byte, date, decimal, time);
- *object property* – this property defines other than *is-a* relationships among classes (nodes); these relationships follow the notion of Resource Description Framework (RDF) [11] that is based on a triple *subject-predicate-object*, where: *subject* identifies what object the triple is describing; *predicate* (property) defines the piece of data in the object a value is given to; and *object* is the actual value of the property; for example, the triple "John likes books" has "John" as subject, "likes" as predicate and "books" as object.

Once the ontology definition is constructed, it can be instantiated. The properties of classes are filled out with real data – values are assigned to datatype properties, and links to instances of other classes are assigned to object properties.

2.3 AATI

Knowledge about a specific category is understood as the importance of keywords describing a category. The keywords and their importance are used to identify documents that belong to the category. The primary characteristic of a new algorithm for *Adaptive Assigning of Terms Importance (AATI)* is that it updates weights of keywords – called here terms – at the same time it categorizes web documents. It means that a training phase is not required [3]. The proposed algorithm is able to handle changes of terms and their importance. Knowledge collected by *AATI* is constantly updated, and there is no need to restart the algorithm when terms and their importance change.

AATI works on two sets of terms. One set contains terms that represent a given category, which is called *target* category,

[1] According to the terminology adopted by the Semantic Web community, the term "instance" has been replaced by the term "individual". For the purpose of clarity, the term instance – similar to the term *instantiated* – is used throughout the paper

[2] The term class is used to distinguish ontology concepts from *HOFC* concepts.

ISBN: 978-989-95079-6-8

while the other set contains terms that do not belong to a *target* category. *AATI* updates importance of the terms via changing weights associated with the terms called Term Weights (*TW*). The importance of terms reflects their discriminatory power in identifying a *target* category. Each web document is evaluated by summing *TW* of all terms that the page contains. The sum is called Page Value (*PV*). Please note that *TW* and *PV* are recursively defined. That is, *TW* and *PV* depend on each other.
Let a set $\boldsymbol{P} = \{p_1,p_2,p_3,...,p_m\}$ denotes *PV* of web documents, where each element p_i denotes *PV* of the *i*th document. Let a set $\boldsymbol{T} = \{t_1,t_2,t_3,...,t_n\}$ denotes terms in web documents, where each element t_j represents *TW* of the term *j*. In *AATI*, *TW* defines *PV*:

$$p_i = \sum_{j=0}^{n} \frac{f_{ij}}{N_i} t_j \tag{2}$$

where N_i is the number of words on a page *i*; f_{ij} is the number of occurrences of term *j* on the page *i*. If a term *j* does not occur on page *i*, f_{ij} is zero.
At the same time, *PV* defines *TW*:

$$t_j = \sum_{i=0}^{m} \frac{f_{ij}}{N'_j} p_i \tag{3}$$

where N'_j is the number of occurrences of term *j* in the document set containing *m* documents. If a term does not occur in the page *i*, f_{ij} is zero.
Therefore in the language of linear algebra, the definition between *PV* and *TW* can be written as

$$\begin{matrix} P = A \cdot T \\ T = B \cdot P \end{matrix} \Rightarrow T = (B \cdot A) \cdot T \Rightarrow T = C \cdot T \tag{4}$$

Thus, $\boldsymbol{T}$ is the eigenvector of matrix $\boldsymbol{C}$ with the eigenvalue one. At the same time, because both the sum of each row in the matrix $\boldsymbol{A}$ and that of each row in the matrix $\boldsymbol{B}$ are one, the sum of each row in the matrix $\boldsymbol{C}$ is also one. So, the eigenvector *T* as the solution of (4) exists and is unique.
Therefore, we define the algorithm *AATI* to evaluate *TW* :

- ***Step1***: build a list of terms for a *target* category, and assign zero as *TW* for all terms;
- ***Step2***: take a new web document; if there is no more web documents, ***STOP***;
- ***Step3***: parse the document based on the list of terms;
- ***Step4***: for each term occurring in the document do one of the following:
  (a) If the term value is not zero (it means the term has already appeared in documents), take this value as *TW*;
  (b) If the term value is zero (the term has not been found in any documents), randomly generate a number between 0 and 1, and assume it as *TW* for this term;
- ***Step5***: Calculate the *PV* for this web document by Equation (2)
- ***Step6***: Update the *TW* of those terms found in the web document (3);
- ***Step7***: Normalize TW across all terms;
- ***Step8***: Go back to Step2.

3 HOFC as Keyword Structure for Text Identification

3.1 Description of Domain of Interest

Text categorization uses keywords as representation of a category, and a text identification process checks if a given text contains those keywords. If all or fraction of keywords (depending of the applied technique) are found in a given document then the document is considered as belonging to a category described by the keywords. Usually, all keywords are not equally important and they are associated with weights identifying their importance. The keywords are represented as a flat structure, and there is no indication that they are "related" to each other, and that some of them depend on others.
A human, as it has been already indicated, looks at a document and uses not only keywords but also related keywords to identify a category the document belongs to. The presence of one keyword in the document activates other connected keywords. More keywords are found in the document more support is collected towards assigning this document to a category described by those keywords.
This human-like approach to use a network of keywords describing a specific category can be implemented with an *HOFC* (Section 2.1). An *HOFC* defines a concept with a set of other concepts or attributes[3]. The relations between a defined (higher-level) concept and defining (lower-level) concepts and attributes represent different types of associations that are meaningful for a user. Therefore, a structure of *HOFC* can be treated as a user-defined description of a given category.
This analogy goes even further when we take a closer look at the process of activation of *HOFC* concepts. In a nutshell, an activation level of higher-level concepts depends on activation levels of lower-level concepts. In other words, activation of a higher-level concept emerges via activation of more specific lower-level concepts and attributes. A level of activation of higher-level concepts is determined via an aggregation operator *OWA* defined with a linguistic quantifier (Q) and an importance vector ($\boldsymbol{M}$).

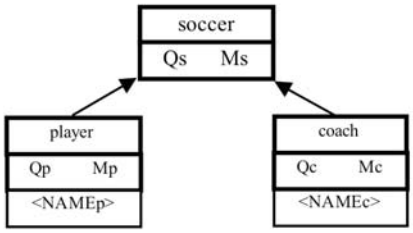


Figure 1: A simple structure of *HOFC* representing the category *soccer*

Fig. 1 shows a simple *HOFC* representing a *soccer* concept defined by two lower-level concepts: *player*, and *coach*. Each

[3] To ensure clarity it should be noted that among all phrases used in the paper the following relations hold: concepts are equivalent to classes; keywords, attributes and terms are equivalent to each other.

ISBN: 978-989-95079-6-8

of them has a single attribute: <NAMEp> and <NAMEc>, respectively. Concepts *soccer*, *player* and *coach* have their own linguistic quantifiers: *Qs*, *Qp*, and *Qc*, and importance vectors ***Ms***, ***Mp***, and ***Mc***. The vector ***Ms*** is of dimensionality 2, dimensionalities of ***Mp*** and ***Mc*** depend on a number of instances of each concept. Instances, that represent concrete pieces of information, are values of attributes <NAMEp> and <NAMEc>.

3.2 HOFC and Ontology

The application of *HOFC* to represent a user's view of the domain of interests brings a number of benefits. The *HOFC* allows us to incorporate different types of relations between concepts and attributes, and build quite complicated and well-interconnected nets of keywords/concepts describing the domain of interest.

However, users who provide *HOFC*s do not have to be experts in the domain. They can provide a very simple hierarchy – called a basic *HOFC* – that contains just a few concepts. This basic *HOFC* is treated as a starting point for building a more comprehensive *HOFC* better suited for identifying a category a given web page belongs to. The process of enriching the user-provided hierarchy is performed with ontology.

An ontology as proposed by the Semantic Web community (Section 2.2) represents a very attractive and powerful approach for knowledge representation. The richness of information that can be expressed with an ontology means that an ontology can be used to provide concrete and additional information for *HOFC*s. In the first case, we talk about instantiation of *HOFC*, and about enrichment of a hierarchy in the second case.

The process of instantiation can be explained with the help of a basic *HOFC*, Fig. 1. This hierarchy is "instantiated" with specific information obtained from the *soccer* category. The ontology is queried about *player*. The result – three instances of the class *player* – is presented in Fig. 2. The process of assigning values to *Q* and ***M*** is described in Section 3.3 and 4.

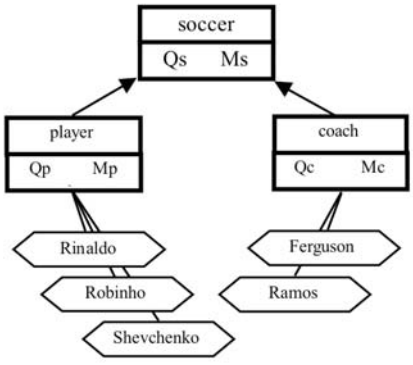


Figure 2: The HOFC of the category *soccer* with instances

An ontology can also be treated as a repository of new concepts that can be added to the concepts already existing in a *HOFC*. This enhancement can be twofold: in the form of addition of lower-level concepts, and in the form of providing concepts related in a "non-trivial" way to concepts already held in the hierarchy.

The concepts provided by a user are used to find corresponding classes in an ontology. The sub-classes of these classes are taken from the ontology and attached to the concepts of *HOFC* as new lower-level concepts. For an *HOFC* presented in Fig. 1, it would mean lower-level concepts: *forward*, *defender*, and *midfielder* are attached to the concept *player*.

Classes defined in an ontology possess object properties (Section 2.2). Object properties are definitions of "non-trivial" relations that exist between pairs of ontology classes. They are used to bring new concepts to an *HOFC*.

For example, the ontology class *player* has an object property *playsFor* that "binds" any instance of *player* with an instance of the ontology class *team*. If we assume that there is an instance of the class *player* called *Rinaldo*. *Rinaldo* has an object property *playsFor* that binds *Rinaldo* with an instance of the class *team – Manchester_United*, Fig. 3.

3.3 Text Identification process

A web page identification process corresponds to an "activation" of the top concept of an *HOFC*. The activation level of the top concept depends on activation levels of lower-level concepts and attributes. The activation of *HOFC* is determined in a bottom-up manner. Attributes at the bottom of *HOFC* are "matched" against words that appear in a web page. Presence of those words means activation of attributes, and activation of attributes means activation of upper-level concepts. This approach requires aggregation of activation levels of all attributes/concepts that define (are attached to) a concept. The aggregation process is performed using *OWA* (Section 2.1). The weights of *OWA* are determined based on *Q* and ***M***.

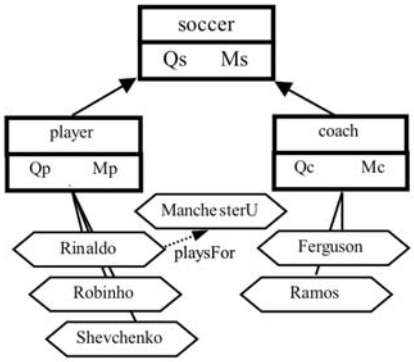


Figure 3: The *HOFC* of the category *soccer* with individuals and "enhancement"

Let $m_i \in [0,1]$ be a value associated with an attribute or concept a_i indicating its importance. If n concepts define a single higher-level concept then ***M*** of this concept is a n-dimensional importance vector $[m_1, m_2, \ldots, m_n]$, and the weighing vector ***W*** has to be calculated based upon both *Q* and ***M***. The first step is to determine the ordered argument vector ***B***, such that b_j is the j^{th} largest of all arguments $a_1, a_2, \ldots, a_n$. The arguments are numbers of occurrences of attributes in a

ISBN: 978-989-95079-6-8

text. Furthermore, we assume μ_j to denote the importance value associated with the attribute or concept that has the j^{th} largest value. Thus if a_5 is the largest value, then $b_1=a_5$ and $\mu_1=m_5$. The next step is to calculate the *OWA* weighing vector $\boldsymbol{W}$ using:

$$w_j = Q(\frac{S_j}{T}) - Q(\frac{S_{j-1}}{T})$$

where

$$S_j = \sum_{k=1}^{j} \mu_k \quad \text{and} \quad T = S_n = \sum_{k=1}^{n} \mu_k \cdot$$

So, S_j is the sum of the importances of the j^{th} most satisfied attributes/concpets, and T is the sum of all importances.
The linguistic quantifier Q is provided by a user at the time they provides their basic *HOFC*. During processes of enhancement, the values of Q for lower-level attributes and concepts are "inherited" from the upper-level concepts. In the case this is not possible a default quantifier *OR* is used.
The values of $\boldsymbol{M}$ representing importance of attributes and concepts of *HOFC* need to reflect a true significance of keywords in their capability to identify a category. The *AATI* algorithm capable of finding those values is used.

4 Importance of HofC Concepts

4.1 HOFC and AATI

Importance levels associated with each concept, keyword/attribute are considered as the knowledge about a domain of interest. Usage of the most accurate and relevant values is important for a good performance of the proposed *HOFC*-based approach. Therefore, the features of *AATI* (Section 2.3) make it a very good candidate for a process of updating and "tracking" changes in importance values of keywords. For the *AATI* algorithm, the importance values $\boldsymbol{M}$ are represented by *TW*.
The *AATI* algorithm modifies *TW* at the same time when it evaluates *PV* for each web document. This process is "controlled" by upcoming pages. If the content of stream of web documents changes, the *TW*s change too.
All concepts/keywords of *HOFC* are input to *AATI*. According to the algorithm their importance is randomly initiated but when time progresses and web pages are being processed – the values of *TW* are being updated and reflect strength of a connection of a term with a domain – strength of a term to "identify" the domain. These *TW* values are fed back to *HOFC* as $\boldsymbol{M}$ values.

4.2 Continuous Changes in Concept Importance

The ability of *AATI* to "follow" the web trends in popularity of a given keyword (popularity associated with a frequency of occurrence) can be illustrated with two examples. The first example represents a change in importance values when a source of web pages changes, while he second focuses on changes in the contents of web page over time.
Fig. 4 and 5 present examples of the adaptability of the *AATI* algorithm. Fig. 4 shows *TW* values of the term "Manchester United" (soccer ontology) as a function of a number of "processed" web pages. The first 1000 pages are from the BBC website (http://www.bbc.co.uk). The next 600 pages are from the CCN website (http://www.cnn.com). We see a decrease in the TW value (solid line after 1000 pages). If the next 600 were from the BBC website the values would follow the dashed line.

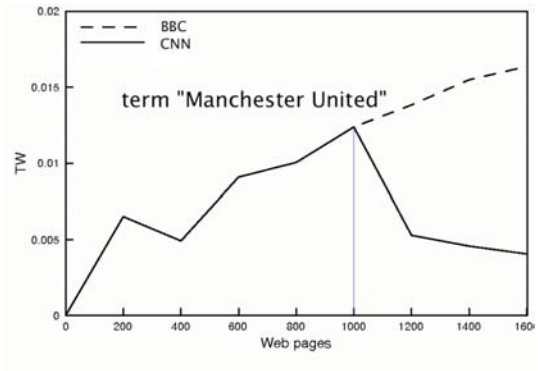


Figure 4: TW changes when the source of upcoming documents changes from BBC to CNN

The term "Manchester United" represents a name of a famous soccer club in England, and its *TW* goes up when the documents are downloaded form BBC. However, when the source of web pages changes and the upcoming documents are pages collected from the CNN site, and the term "Manchester United" is less popular. This decrease in popularity of the term "Manchester United" translates into decrease in its *TW*.
A different kind of change is seen in Fig. 5. The plots represent changes in importance levels when web pages collected at different moments in times are fed into the *AATI* algorithm. The first 2300 pages were collected from the BBC site in February 2007, while the last 2500 pages are from BBC site from January 2009. It can be seen that the importance levels for different terms behave differently: the importance of the player Ronaldo stays more or less at same level – indication that Ronaldo was and still is popular; the importance of the player Robinho increased and then slowly decreased, while the popularity of the player Shevchenko decreased.

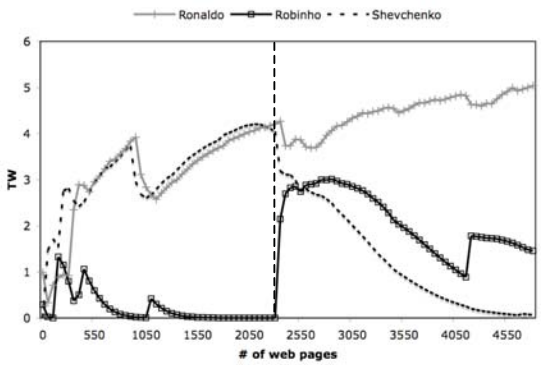


Figure 5: Changes of *TW* when the set of upcoming documents change from Feb07 to Jan09: for Ronaldo, Robinho, and Shevchenko

ISBN: 978-989-95079-6-8

5 Classification System

5.1 Architecture

An implementation of the proposed approach for web content identification requires integration of multiple components executing different tasks. A simplified architecture of a system performing the identification task is shown in Fig. 6. The system contains the following three parts: *HOFC* Preparation and Enhancement Unit, *AATI* Execution Unit, and Identification Unit. The *HOFC* Preperation and Enhancment Unit is responsible for setting up a basic *HOFC* representing user's query, adding new concepts from a domain ontology, and populating an enhanced *HOFC* with concrete pieces of information from the ontology. The *AATI* Unit is invoked on regular basis (based on time interval or a number of incoming web pages) to update values of importance of concepts. The Identification Unit checks the incoming web page against the concepts from the *HOFC*. This process starts with concepts located at the bottom of the *HOFC* – and it progresses towards the top. It is checked if the *HOFC* concepts are "activated". The activation level of concepts is the result of a number of their occurrences in a text, as well as Q and $\boldsymbol{M}$ (Section 3.3).
A prototype of the system has been developed in Java. The Protégé API is used to access ontology classes, their properties and instances. Moreover, API functions allow for creating a new enhanced hierarchy of concepts based on classes and individuals obtained from the domain ontology and the hierarchy of concepts given by a user.

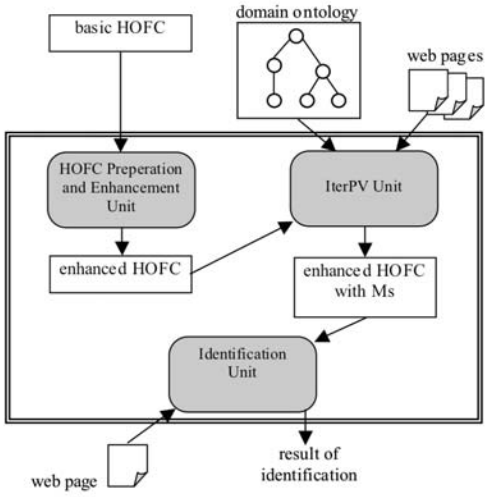


Figure 6: Text Identification System

5.2 Experimental Results

A simple example that illustrates application of the prototype system is presented. The *HOFC* used in the experiment is shown in Fig. 1. It is a very simple hierarchy representing the category *soccer*. A user is a novice to this category, and the only concepts that define soccer are *players* and *coach*. This *HOFC* is later instantiated and enhanced, Fig 3. The enhanced *HOFC* represents the following structure of concepts:

> The concept *soccer* is activated if there are "MOST" of *players* and *coaches*. The concept *players* is activated if there exist "OR" Ronaldo "OR" Robinho "OR" Shevchenko. The concept *coach* is activated when "OR" Ferguson "OR" Ramos are present.

This simple *HOFC* has been checked against a number of different web pages. In all cases the activation levels of the concept *soccer* reflected a human evaluation of pages. The activation levels have been calculated based on aggregations of activations of keywords – Ronaldo, Robinho, Shevchenko, Ferguson, Ramos – using simple linguistic quantifiers MOST and OR, and values of $\boldsymbol{Ms}$ provided by *AATI*.

6 Conclusions

The paper proposes a concept-based structure suitable for text identification. The method uses hierarchies of concepts as category identifiers. The hierarchies use *OWA* operators with linguistic quantifiers and concept importance to control conditions that have to be satisfied by lower-level concepts and attributes in order to activate a higher-level concept.
The concept importance levels are updated using the *AATI* algorithm. This algorithm is able to modify concept and attribute importance values in on-line fashion in an un-supervised manner.
A prototype of the system implementing the proposed approach is briefly described in the paper.

Reference

[1] T. Berners-Lee, J. Hendler, and O. Lassila, The Semantic Web, *Scientific American*, 284: 34-43, May 2001.

[2] T.E. Gruber, A translation approach to portable ontology specifications, *Knowledge Acquisition*, 5: 199-220, 1993.

[3] Z. Li and M. Reformat, An Iterative Algorithm for Text Categorization, submitted to *Information Processing Letters*.

[4] C. Manning, P. Raghavan, and H. Schutze, *Introduction to information retrieval*, Cambridge University Press, New York, NY, 2008

[5] T. Martin, Searching and smushing on the semantic web-challenges for soft computing, in M. Nikravesh, B. Azvine, R. R. Yager, & L. A. Zadeh (Ed.), *Enhancing the Power of the Web*, Heidelberg: Springer, 167-188, 2004.

[6] E. Sanchez, and T. Yamanoi, Fuzzy Ontologies for the Semantic Web, in *Flexible Query Answering Systems*, 691-699, 2006.

[7] F. Scott, W.D. Lewis, and D.T. Langendoen, An ontology for Linguistic Annotation 1-2, in Proceedings of *14th Innovative Applications of AI Conference*, Edmonton, Canada, 11-19, 2002.

[8] F. Sebastiani, Machine learning in automated text categorization, *ACM Computing Survey(CSUR)*, 34(1): 1-47, 2002.

[9] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18: 183-190, 1988.

[10] R.R. Yager, A Hierarchical Document Retrieval Language, *Information Retrieval*, 3: 357-377, 2000.

[11] RDF: http://www.w3.org/RDF/

ISBN: 978-989-95079-6-8

Possibility distribution: a unified representation for parameter estimation

Gilles P. Mauris

LISTIC, Polytech'Savoie, Université de Savoie
Annecy, France
Email: mauris@univ-savoie.fr

Abstract— *the paper presents a possibility formulation of one-parameter estimation that unifies some usual probability formulations. Point and confidence interval estimation are unified in a single theoretical formulation and incorporated into estimators of an omnibus form: a possibility distribution. New relationships between continuous possibility distribution and probability concepts are established. The concept of specificity ordering of a possibility distribution is then used for comparing the efficiency of different estimators. The usefulness of the approach is illustrated on mean and median estimators from data sample of different size and of different probability distribution.*

Keywords— efficiency, parameter estimation, possibility, probability, stochastic ordering, uncertainty.

1 Introduction

Systems designed by engineers are often meant to influence their environment: to manipulate it, to move it, to control it and so on. To enable such actions, these systems need information, e.g. values of physical quantities describing their environment. Generally, two types of information sources are available: prior knowledge and empirical knowledge. The latter is obtained by sensor observations. Prior knowledge is the knowledge that was already there before a given observation became available. The combination of prior knowledge and empirical knowledge leads to posterior knowledge useful for the task at hand. Parameter Estimation deals with this problem to infer a parametric description for an object, a physical process or an event, given measurements that always come with uncertainties due to variability of influence quantities or phenomena. Therefore an assessment of the parameter estimate uncertainty is also required by the application, and prior knowledge is often used to reduce the uncertainty of the final estimate. To tackle such problems a lot of developments have been done in the statistical science [1].

In statistics, a parameter estimator $T_\theta(X)$ is defined as a function of the observable sample data $X_1, X_2, ..., X_n$ that is used to estimate an unknown parameter θ; an estimate $\hat{\theta}$ is the result from the actual application of the function to a particular set of data. Prior knowledge can be that $X_1, X_2, ..., X_n$ is an independent identical distributed (*iid*) sample from a continuous random variable X having a distribution function F. Many different estimators are possible for any given parameter. Two main types of estimators are worldwide used: point estimators and interval estimators. The selection of an estimator of one of the above kinds for purposes of informative inference, including typical applications in scientific research is generally admitted to involve elements of choice which are in some degree arbitrary. Such elements include the choice of a particular confidence level for an interval estimator, and the choice of a location function for a point estimator for a given confidence coefficient. In addition, a point estimate is often desired along with an interval. Such considerations and related ones have led to proposals for simultaneously use of a point confidence estimator and a set of confidence limit or interval estimation having various confidence coefficients [2]. Such estimators may be regarded as a modern formulation of a long standing practice of reporting estimates in the form $\hat{\theta} \pm k\sigma$, where k is some constant and σ the standard deviation of the observations. Note that this expression is recommended by the Guide of Uncertainty in Measurement (GUM [3]) edited by international legal metrology institutions. This form may be interpreted as an ordered set of three point estimators. Tukey proposed in [4] that for general purposes, it would be advantageous to use a set of five points estimators at standard confidence level *2,5%,16,66%, 50%, 83,33% and 97,5%*. Cox proposed to use the full continuous family of confidence levels varying between *0* and *1* by introducing a so called confidence curve estimator [5].

The problem of choosing a good estimator, i.e. an estimator which tends to take values close to the true unknown value, is formulated mathematical by introducing criteria of closeness. The latter has been specified by the introduction of specific loss function (called sometimes risk function). For example, the absolute error criterion was introduced by Laplace, Gauss replaced this by the squared error criterion which proved mathematically much more tractable [2]. Each such definite specification of closeness can be criticized as somewhat arbitrary, except in a context when one postulates the reality of the indicated cost of errors of each possible kind. This point put forth by Galileo leads to comparisons between estimators on the basis of all their probabilities rather than on the basis of certain summaries such as the variance [2]. Birnbaum extended this concept to continuous random variables by introducing the notion of peakedness of a probability distribution, which characterizes the concentration of values around the estimate [6].

ISBN: 978-989-95079-6-8

In this paper, we propose a possibility formulation of one parameter estimation, which unifies the usual previous probability formulation. Our proposition is based on a possibility/probability transformation we previously proposed [7][8], and on the notion of possibility distribution specificity ordering [9]. In section 2, this transformation is recalled and new relationships with the above mentioned probability estimation concepts are exhibited. In particular, the efficiency of estimators is characterized by its stochastic concentration defined by the possibility specificity. The approach is illustrated in section 3 on the problems of mean and median estimation. The advantages of the possibility formulation, which leads to an intuitive graphic representation, are highlighted, particularly when the sample size varies and the probability distribution as well.

2 Possibility versus probability estimation

2.1 Basics of possibility theory

Possibility theory is one of the modern theories available to represent uncertainty when information is scarce and/or imprecise [10]. The basic notion is the possibility distribution [11], denoted π. Here, we consider possibility distributions defined on the real line, i.e. π is an upper semi-continuous mapping from the real line to the unit interval. Thus π is a fuzzy subset but with specific semantics for the membership function. Indeed, a possibility distribution describes the more or less plausible values of some uncertain variable *X*. The possibility theory provides two evaluations of the likelihood of an event, for instance whether the value of a real variable *X* does lie within a certain interval: the possibility Π and the necessity N. The normalized measures of possibility Π and necessity N are defined from the possibility distribution $\pi : R \to [0,1]$ such that $\sup_{x \in R} \pi(x) = 1$ as follows:

$\forall A \subset R, \Pi(A) = \sup_{x \in A} \pi(x)$ and

$\forall A \subset R, N(A) = 1 - \Pi(\bar{A}) = \inf_{x \notin A}(1 - \pi(x))$

The possibility measure Π satisfies:

$\forall A, B \subset R, \Pi(A \cup B) = \max(\Pi(A), \Pi(B))$

The necessity measure N satisfies:

$\forall A, B \subset R, N(A \cap B) = \min(N(A), N(B))$

In fact, possibility measures are set functions similar to probability measures, but they rely on axioms which involve the operations " maximum " and " minimum " instead of the operations "addition" and "product" (if the measures are decomposable).

Moreover we can interpret any pair of dual functions necessity/possibility $[N, \Pi]$ as upper and lower probabilities induced from specific probability families. Let π be a possibility distribution inducing a pair of functions $[N, \Pi]$. We define the probability family $\mathcal{P}(\pi) = \{P, \forall A, N(A) \le P(A)\} = \{P, \forall A, P(A) \le \Pi(A)\}$. In this case, $\sup_{P \in \mathcal{P}(\pi)} P(A) = \Pi(A)$ and $\inf_{P \in \mathcal{P}(\pi)} P(A) = N(A)$ [12]. In other words, the family $\mathcal{P}(\pi)$ is entirely determined by the probability intervals it generates. In the estimation context, this fact can be useful when it is not possible to identify one specific probability distribution for the observations. The concept of specificity can be used to qualify the informativeness of a possibility distribution. Indeed, a possibility distribution π_1 is said more specific than π_2 as soon as $\forall x, \pi_1(x) \le \pi_2(x)$ [9](it is the usual definition of inclusion of fuzzy sets), i.e. π_1 is more informative than π_2.

2.2 Background on probability/possibility transformation

Let us assume that the sample data are issued from a continuous random variable X defined on the set of reals and described by a probability density function f, F being its corresponding cumulative distribution function, with F^{-1} its inverse function if it exists (otherwise the pseudo-inverse function can be considered [13]). For every possible confidence level $\beta \in [0,1]$, the corresponding confidence interval is defined as an interval that contains the parameter to be estimated, i.e. θ, with probability $\ge \beta$. In other words, a confidence interval of confidence level β (denoted I_β) is defined as an interval for which the probability P_{out} to be outside this interval I_β does not exceed $\alpha \overset{def}{=} 1 - \beta$, i.e. $P(\theta \notin I_\beta) = \alpha$.

It is possible to link confidence intervals and possibility distribution in the following way. A unimodal numerical possibility distribution may be viewed as a nested set of confidence intervals, which are the α cuts $[\underline{x}_\alpha, \overline{x}_\alpha] = \{x, \pi(x) \ge \alpha\}$ of π. The degree of certainty that $[\underline{x}_\alpha, \overline{x}_\alpha]$ contains μ is $N([\underline{x}_\alpha, \overline{x}_\alpha])$ (if π continuous). Obviously, the confidence intervals built around the same point x^* are nested. It has been proven in [7] that stacking confidence intervals of a probability distribution on top of one another leads to a possibility distribution (denoted π^* having x^* as modal value). In fact, in this way, the α-cuts of π^*, i.e. $A_\alpha = \{x | \pi^*(x) \ge \alpha\}$ are identified with the confidence interval I^*_β of confidence level $\beta = 1 - \alpha$ around the nominal value x^*. Thus, the possibility distribution π^* encodes the whole set of confidence intervals in its membership function. Moreover, this possibility distribution satisfies:

$\forall A \subset R, \Pi^*(A) \ge P(A)$, with Π^* and P the possibility and probability measures associated respectively to π^* and f.

A closed form expression of the possibility distribution $\pi^M(x)$ induced by confidence intervals around the median $x^* = M$ is obtained for symmetric unimodal continuous probability densities $f(x)$ strictly increasing on the left and decreasing on the right of M [7]:

$$\begin{aligned} &\forall x \in [-\infty, M], \pi^M(x) = 2F(x) \\ &\forall x \in [M, +\infty], \pi^M(x) = 2(1 - F(x)) \end{aligned} \quad (1)$$

Thus, the possibility distribution is closely related to the cumulative distribution function.

ISBN: 978-989-95079-6-8

2.3 Relationships with probability estimation notions

According to the way a possibility distribution is built from the probability distribution F from which the sample data come from, the link with confidence intervals is immediate. Indeed, generally the parameter to be estimated is a location parameter of a distribution function, e.g. the mean. The link with Cox confidence curve $c_\theta(x)$ is also obvious. Indeed the latter is defined by [5]:

$\forall \theta \in [-\infty, x], c_\theta(x) = F(\theta - x)$

$\forall \theta \in [x, +\infty], c_\theta(x) = 1 - F(\theta - x)$

Thus, $c_\theta(x) = \pi_F(x-\theta)/2$.

Other new links concern the quality of estimator that is related to the concentration of values around the estimate, quantified for example by the variance or by the absolute mean deviation. Hereafter a few proposition formalizing for continuous symmetric random variables the links with the absolute mean deviation, the Levy concentration function, and other stochastic orderings.

Proposition 1:

The specificity index $sp(\pi) = \int_{-\infty}^{+\infty} \pi^M(x)dx$ of the possibility $\pi^M(x)$ distribution equivalent to a continuous symmetric random variable X is equal to twice the absolute mean deviation:

$$sp(\pi_X^M) = 2.E|X - M| .$$

Proof:

Assume without loss of generality that *M=0*. Due to symmetry:

$\int_{-\infty}^{+\infty} \pi_X^M(x)dx = \int_{-\infty}^{0} \pi_X^M(x)dx + \int_{0}^{+\infty} \pi_X^M(x)dx =$

$2\int_{-\infty}^{0} \pi_X^M(x)dx = 4\int_{-\infty}^{O} F(x)dx$

In other respect, as $F(-\infty) = 0$:

$2\int_{-\infty}^{0} F(x)dx = 2[xF(x)]_{-\infty}^{0} - 2\int_{-\infty}^{0} xf(x)dx$

$= 0 - \int_{-\infty}^{\circ\infty} xf(x)dx = E(|X|)$

Therefore: $sp(\pi_X^M) = \int_{-\infty}^{+\infty} \pi_X^M(x)dx = 2.E|X - M|$.

Thus the possibility distribution is closely related to the absolute mean deviation, (and not to the conventional standard deviation), of course if the latter is defined. Indeed, when the term $\int_{-\infty}^{0} xf(x)dx$ is infinite, the possibility distribution can be computed but the associated specificity is infinite.

Proposition 2:

The possibility distribution $\pi_X^M(x)$ equivalent to a continuous unimodal symmetric random variable X is related to the Lévy concentration function $Q_X(x)$ by the relation:

$$\forall x \geq 0, Q_X(x) = 1 - \pi_X^M(x + M)$$

Proof:

The concentration function introduced by P. Lévy in 1935 has been first defined by the expression [14, page 44]:

$\forall x \geq 0, Q_X(x) = \sup_{x_0} [F(x_0 + x) - F(x_0 - x)]$

Note that, in a latter definition [14, page 90], which is the one considered usually, the expression has became: $Q_X(x) = \sup_{x_0} [F(x_0 + x) - F(x_0)]$. Here we will consider the first definition, more close to our approach.

If F is a symmetric and unimodal distribution about M then, the supremum in equation is attained for $x_0 = M$ (see [8] for a demonstration). Thus we have:

$\sup_{x_0} [F(x_0 + x) - F(x_0 - x)] = F(M + x) - F(M - x)$.

The symmetry of F gives for $x \geq 0$:

$F(M + x) - F(M - x) = F(M + x) - (1 - F(M + x))$

$= 2F(M + x) - 1 = 1 - 2(1 - F(M + x)) = 1 - \pi(x)$.

Therefore, $Q_X(x) = 1 - \pi_X^M(x + M)$.

In fact, the possibility distribution gathers probability dispersion intervals and it is complementary to the Lévy concentration function for continuous unimodal random variables.

Proposition 3:

The specificity order (defined by the fuzzy subset inclusion) defines a stochastic concentration ordering equivalent for continuous unimodal symmetric random variables both to the peakedness ordering of Birnbaum [6] and to the majorization ordering of Hickey [15]:

$$\pi_X^M(x) \leq \pi_Y^M(x) \Leftrightarrow X \geq^{peaked} Y \Leftrightarrow X \leq^{maj} Y$$

Proof:

According to Birnbaum [6], X is said to be more peaked about θ , than Y about θ if and only if:

$\Pr(|X - \theta| \geq t) \leq \Pr(|Y - \theta| \geq t)$ holds for all $t \geq 0$.

It is clear that we have:

$\pi_Y(\theta - t) = \pi_Y(\theta + t) = 1 - \Pr(\theta - t, \theta + t) = \Pr(|X - \theta| \geq t$

Thus: $\pi_X^M(x) \leq \pi_Y^M(x) \Leftrightarrow X \geq^{peaked} Y$

Note also that peakenedness ordering is equivalent to conventional stochastic ordering of absolute variables:

$X \geq^{peaked} Y \Leftrightarrow |X - \theta| \leq^{sto} |Y - \theta| \overset{def}{\Leftrightarrow} F_{|X-\theta|}(x) \leq F_{|Y-\theta|}(x)$

The majorization order introduced by Hickey is used to compare continuous distribution in terms of randomness. Y is said at least as random as X in the majorization sense if [15]:

$\int_o^t g^*(y)dy \leq \int_o^t f^*(y)dy, \forall t$,

where f^* and g^* are the decreasing rearrangement of f and g respectively; that is:

$f^*(x) = \sup\{c : m(c) > x\}, x > 0$ with,

$m(c) = \mu\{x : f(x) > c\}$, μ denoting Lebesgue measure and f the corresponding density function of F .

ISBN: 978-989-95079-6-8

It is proved in [16] that for unimodal continuous random variables:

$\int_o^t f^*(y)dy = Q_X(t)$.

Thus: $\pi_X^M(x) \le \pi_Y^M(x) \Leftrightarrow Q_X(x) \ge Q_Y(x) \Leftrightarrow X \le^{maj} Y$.

Note also that a similar proposition as the proposition 3, but for discrete probability and possibility distributions has been stated by Dubois and Hüllermeier [17].

Thus given two estimators $T_\theta(X)$ and $U_\theta(X)$ of a parameter θ, T is said to have greater stochastic concentration about than U does if T is more peaked than U . An interesting consequence of greater peakedness proved by Hwang [18] is that if L is any loss function L satisfying: $L(x,\theta) = h(|x-\theta|)$,

where for each θ, $h:[0,\infty) \to [0,\infty)$ is a non decreasing function , we have:

$\pi_{T_\theta(X)}(x) \le \pi_{U_\theta(X)}(x) \Leftrightarrow E_\vartheta(L(T_\theta(X)) \le E_\vartheta(L(U_\theta(X))$

In other words not only is T at least as likely as U to be within a neighborhood $[\theta - x, \theta + x]$ of the true parameter, it is at least as good with regard to a broad class of loss functions which includes absolute mean deviation ($L(x,\theta) = |x-\theta|$), variance ($L(x,\theta) = |x-\theta|^2$), and any other symmetric loss function where the loss is at least as large when x is farther away from θ than when it is closer to θ . It includes also the continuous entropy. Indeed, the latter is defined as: $H(X) = -\int_X f(x)Ln(f(x))$, and the majorization order implies the entropy order [19]. Therefore, for continuous unimodal distribution, the specificity order implies the entropy order.

Thus, the preference ordering of estimators based on greater possibility specificity around the parameter is indeed a strong preference ordering.

2.4 Relationships with other notions

The specificity order (defined by the fuzzy subset inclusion) is also related to second order stochastic dominance (SOSD) in connection with risk-aversion in decision theory. In this context, the notion of dispersion is translated to a notion of risk. Given two symmetric distributions with the same expected value, G is riskier than F if every risk-averse individual prefers F to G. One mathematical definition is:

$F \le_{SOSD} G \Leftrightarrow \int_{-\infty}^t F(x)dx \le \int_{-\infty}^t G(x)dx, \forall t$

We have: $\pi_X^M(x) \le \pi_Y^M(x) \Rightarrow X \le_{SOSD} Y$.

Proof: $\int_{-\infty}^t F(x)dx \le \int_{-\infty}^t G(x)dx$ is equivalent to:

$U(t) = \int_{-\infty}^t (F(x) - G(x))dx \le 0, \forall t$

We have: $U'(t) = F(t) - G(t) \le 0, \forall t \le 0$

due to $\pi_X^M(x) \le \pi_Y^M(x)$. Thus U is decreasing, and since $U(-\infty) = 0$, then $\forall t \le 0, U(t) \le 0$.

By a symmetric reasoning, $\forall t \ge 0, U(t) \le 0$, which completes the proof.

A link exists also with the Value at Risk notion $VaR_\alpha(X)$ that is defined by the 1-α quantile of X, i.e.:

$P(X > VaR_\alpha(X)) = 1 - \alpha$.

It is easy to see that:

$\pi_X^M(x) \le \pi_Y^M(x) \Rightarrow VaR_\alpha(X) \ge VaR_\alpha(Y), \forall \alpha \in [0,1]$

The specificity order is also related to notions of social science, i.e. the Lorenz curve and the Gini index defined by:

$L_X(t) = \frac{1}{E(X)} \int_O^t F(x)dx$ and $G(X) = 1 - 2\int_O^1 L_X(t)dt$

In fact, Lorenz dominance order is equivalent to second order stochastic dominance for distributions with equals means [19]. Thus, we have:

$\pi_X^M(x) \le \pi_Y^M(x) \Rightarrow L_X(x) \le L_Y(x), \forall x$ and,

$\pi_X^M(x) \le \pi_Y^M(x) \Rightarrow G(X) \ge G(Y)$.

The Gini index is a measure of statistical dispersion most prominently used as a measure of inequality of wealth distribution. A low Gini index indicates more equal wealth distribution, while a high Gini index indicates more unequal distribution.

3 Mean and median estimators

Let us consider the case where the parameter to be estimated is a location parameter of a continuous symmetric distribution function. This situation is often encountered in physical entity measurements where the variability of observations is due to variability of influence quantities or phenomena [3].

We considered hereafter the parameter estimation by using the mean and the median estimators.

3.1 Mean estimator

Suppose $X_1, X_2, ..., X_n$ are an *iid* sample from a distribution F . The mean estimator is defined: $\hat{\theta}_n = \left(\sum_{i=1}^n X_i\right)/n$.

Generally, the Gaussian distribution with a deviation σ is used for F , because, in this case the distribution of the mean estimator is also Gaussian but with a standard deviation $\sigma/\sqrt{n}$. Hereafter are plotted the possibility distributions associated to such mean estimators for $n = 1, 3, 30$ and $\sigma = 1$. As expected, the larger the sample size the more specific the possibility distribution.

ISBN: 978-989-95079-6-8

Let us now consider that the sample data are issued from a Cauchy distribution with a scaling parameter of *1*: $f(x) = \frac{1}{\pi(1+x^2)}$.

This distribution has no finite moments, but the median exists and can be used as a measure of central tendency. According to equation (1), the equivalent possibility distribution is defined by:

$$\forall x \le 0, \pi(x) = 1 + \frac{2}{\pi} arctg(x)$$

$$\forall x \ge 0, \pi(x) = 1 - \frac{2}{\pi} arctg(x)$$

The probability distribution of the mean estimator is also a Cauchy distribution with a scale parameter of *1* [1]. Thus the possibility distribution of the mean estimator of any number of data is the same as the one obtained for one datum. Therefore the specificity is not improved by increasing the sample size. Note that this is not in contradiction with the central limit theorem though the latter considers only distribution with finite variance (which is not the case of the Cauchy distribution).

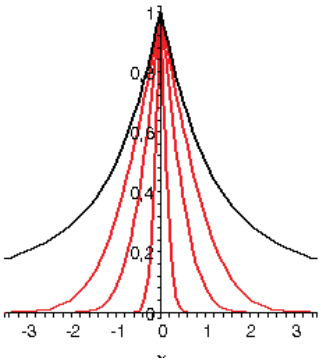


Figure 1: Possibility mean estimators for Normal (red) and Cauchy (black) probability distributions (*n=1, 3, 30*).

3.2 Median estimator

Let M_n be the median of $X_1, X_2, ..., X_n$, θ being the median of the distribution F from which the sample data are issued. For n odd, we have the following relationship [20]:

$$P(M_n - \theta \ge x) = \sum_{i=0}^{(n-1)/2} C_n^i F(x)^i (1 - F(x))^{n-i}.$$

Therefore the expression of the possibility distribution of median estimator is:

$$\forall x \le \theta, \pi_{T_\theta(X)}(x) = 2(1 - \sum_{i=0}^{(n-1)/2} C_n^i F(x)^i (1 - F(x))^{n-i})$$

$$\forall x \ge \theta, \pi_{T_\theta(X)}(x) = 2 \sum_{i=0}^{(n-1)/2} C_n^i F(x)^i (1 - F(x))^{n-i}$$

The possibility median estimators for a Gaussian and a Cauchy distribution are plotted in figure 2 for $n = 1, 3$.

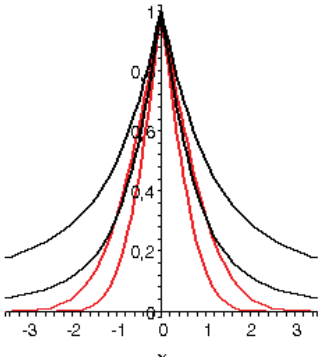


Figure 2: Possibility median estimators for Cauchy (black) and Normal (red) distributions (n=1, 3).

Note that the specificity increases with the sample size in both cases, which is remarkable for the Cauchy distribution.

3.3 Comparison

The preceding results show that the specificity of the possibility distribution associated to a median estimator increases with the sample size (with odd data) even for a long tail probability distribution such as the Cauchy distribution. It is not the case for the mean estimators. This seems indicate that the median estimator is better. Hereafter are plotted the possibility mean and median estimators for a few observations from a normal distribution.

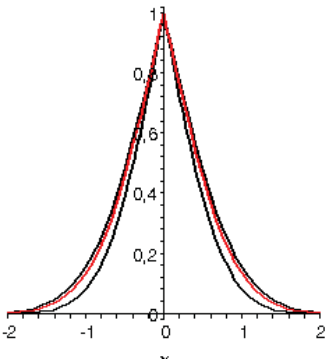


Figure 3: Possibility mean (black, n=2, 3) and median (red, n=3) estimators for a Normal distribution.

The possibility distribution of the median estimators with 3 observations is more specific than the possibility distribution of the mean estimator for 2 observations but less specific than the possibility distribution of the mean estimator for 3 observations. It seems that in general the possibility median estimator for *2n+1* data is more specific than the mean estimator for *2n* data.

ISBN: 978-989-95079-6-8

4 Conclusions

The paper has presented new links between continuous symmetric probability and possibility distributions. The interest of the possibility approach in the context of parameter estimation has been highlighted through the powerful concept of specificity ordering of possibility distribution. The latter allows sounded and intuitive comparisons between different estimators (e.g. mean and median). Further developments should consider dissymmetric distributions and other estimators.

References

[1] J. Rice, *Mathematical Statistics and Data Analysis*, Duxbury Press, 1995, 597 pages.

[2] A. Birnbaum, "A unified theory of estimation", *The Annals of Mathematical Statistics*, Vol. 32, No 1, 1961, p. 112-135.

[3] *Guide for the expression of uncertainty in measurement*, ISO 1993, 1993.

[4] J.W. Tuckey, "Standard confidence points", *Memorandum report 26,* Princeton University, 1949.

[5] D.R. Cox, "Some problems connected with statistical inference", *The Annals of Mathematical Statistics*, Vol. 29, 1958, p. 357-372.

[6] Z.W. Birnbaum, "On random variables with comparable peakedness", *The Annals of Mathematical Statistics*, Vol. 19, 1948, p. 76-81.

[7] G. Mauris, V. Lasserre and L. Foulloy, "A fuzzy approach for the expression of uncertainty in measurement", *Int. Journal of Measurement*, Vol. 29, No 3, March 2001, pp. 165-177.

[8] D. Dubois, L. Foulloy, G. Mauris and H. Prade, "Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities", *Reliable Computing*, Vol. 10, No 4, 2004, pp. 273-297.

[9] D. Dubois, H. Prade and S. Sandri, "On possibility/probability transformations", in*: Fuzzy Logic*, R. Lowen, M. Roubens, Eds, 1993, pp. 103-112.

[10] C. Baudrit and D. Dubois, "Practical representation of incomplete probabilistic information", *Computational Statistics & Data Analysis*, Vol. 51, Issue 1, 2006, pp. 86-108.

[11] L.A. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems*, Vol.1, No1, 1978, pp. 3-28.

[12] D. Dubois and H. Prade," On several representations of an uncertain body of evidence", in: *Fuzzy Information and Decision Processes* (M.M. Gupta, E. Sanchez, eds.), North-Holland, 1982, pp. 167-181.

[13] M. Kendall and A. Stuart, *The advanced theory of statistics*, Ed. Griffin and Co., 1977.

[14] P. Lévy, *Théorie de l'addition des variables aléatoires,* Gauthier-Villars, Paris, 1954, 416 pages.

[15] R.J. Hickey, "Continuous majorization and randomness", *Journal of applied probability*, No. 23, 1984, pp. 924-929.

[16] J.M. Fernandez-Ponce and A. Suarez-Llorens, "An aging concept based on majorization", *Probability in the engineering and informational science*, No 17, 2003, pp. 107-117.

[17] D. Dubois and H. Hüllermeier, "Comparing probability measures using possibility theory: a notion of relative peakedness", *Int. journal of Approximate reasoning,* No 45, 2007, pp. 364-385.

[18] J.T. Hwang, "Universal domination and stochastic domination: estimation simultaneously under a broad class of loss functions", *Annals of Statistics,* No 13, 1985, pp. 295-314.

[19] A. Müller and D. Stoyan, *Comparison methods for stochastic models and risks*, Wiley, 2002, 330 pages.

[20] H.A. David, *Order Statistics*, Wiley, 2003.

ISBN: 978-989-95079-6-8

Cooperative Partner Agent of Seven-Card-Stud Poker

K. Osone[1] T. Onisawa[2]

1,2 Graduate School of Systems and Information Engineering,
University of Tsukuba, 1-1-1, Tennodai, Tsukuba, 305-8573 Japan
Email: 1:kfukuda@fhuman.esys.tsukuba.ac.jp, 2:onisawa@iit.tsukuba.ac.jp

Abstract— *This paper aims at the construction of a cooperative partner agent that plays a seven-card stud poker game with a human partner against an opponent player. If a human player needs some advice on a game, the partner agent gives a human partner player strategy for a game and its grounds. A human player can also ask some questions about the current situation of a game. If a human player tries to take different strategy from the one presented by the partner agent, the partner agent calls human player's attention to the presented strategy with its grounds. Fuzzy theory and case based reasoning are applied to decision-making part in the partner agent. This paper also performs subject experiments to evaluate the effectiveness of the presented partner agent, where each subject plays games with/without the partner agent.*

Keywords— game with imperfect information, poker game, cooperation, fuzzy inferences, case based reasoning

1 Introduction

In our daily life we are often confronted with decision-making problems under uncertain situations. Although we think of many alternatives or various different opinions under such situation [1], it is usually difficult to solve them completely. Therefore, we often get other perspectives or hints to solve problems by exchanging opinions among others or by taking advice from others. An adviser with ripe experience may give better advice on decision-making problems. Exchanging opinions or taking advice may lead to some solution that cannot be found by oneself or to some other understanding on problems. From this point of view, there are many studies on decision support systems [2,3,4,5,6]. However, these systems only support users' decision and do not consider decision-making program from the viewpoint of cooperation with human users.

This paper considers the situation in which human makes decision in cooperation with a computer. In this paper, as an example of decision-making problem, seven-card-stud poker game, a kind of a game with imperfect information [7], is taken because situations estimate is usually difficult under uncertainty of poker game situations. Although there are also some studies about negotiation agent in a game, e.g., a MONOPOLY game [8], or poker playing systems [9,10,11], studies on decision-making from the viewpoint of cooperation with human players in a game are not found.

This paper aims at the construction of not a mere poker playing system but a partner agent, which has a discussion with a human player on game strategy in a seven-card-stud poker game cooperating with a partner player. The presented agent is based on the authors' previous seven-card-stud poker game playing system [12,13,14], and is enlarged from our previous partner agent [15] in a sense that the system gives some advice to a human partner player using case based reasoning. A human player plays poker games with the presented partner agent that cooperates with a human player by not only presenting information on game situations but also performing question and answering of game strategy. Therefore, this paper considers not only gain/loss of point but also the human partner player's evaluation of usefulness/helpfulness.

The organization of the paper is as follows. In Section 2 a seven-card-stud poker game used in this paper is explained. Section 3 presents the partner agent consisting of the decision-making module and the cooperation module. Section 4 mentions subject experiments performed in order to show the usefulness of the presented partner agent and their results. Final section shows conclusions of this paper.

2 Seven-Card-Stud Poker

2.1 Stud Poker

A stud poker game is different from a draw poker one. And although a stud poker game has various kinds of game variations, face-down and face-up cards are dealt to each player in every stud poker game. The card dealt face down is called a hole card that only a card owner can see. On the other hand the card dealt face up is called an up card that all players can see.

2.2 Seven-Card-Stud Poker

This paper considers the situation in which a human player and the partner agent play a poker game together against the previous poker playing system [13,14] as an opponent player. This section describes rules of a seven-card-stud poker game used in this paper.

After both players (a human player and the playing system) pay 5 points as the ante, two hole cards and one up card are dealt to each player, and the first betting round is started. The second player must bet at least the same amount of bet as the first player's one, where one player can raise the betting point only once at each betting round, and the upper limit of betting points is assumed to be total points betted so far. When both players' betting points become even, a next card is dealt to each player and the next betting round is started. The fourth, the fifth and the sixth cards are face-up cards, and the seventh

ISBN: 978-989-95079-6-8

card is a hole card. After all cards are dealt to and betting rounds are ended, each player makes a decision which hole card is revealed. Each player reveals one hole card and the betting round is started. If any players still don't fold a game, the procedure of revealing hole cards and that of betting round are continued. Bluff strategy or slow play is also often taken in the procedure of revealing hole cards. If both players continue a game after the final betting round, they show their final hole cards each other, which is called Showdown, and the player having a stronger poker hand with five cards out of seven is a winner of a game.

2.3 Poker Hands

This paper considers 10 sorts of poker hands including no pair (high card). Scores of these hands are defined as shown in Table 1. In this paper, scores are simply defined as 10 scales from 0 to 9 because the linear scale is simple.

Table 1:Score of poker hands

Poker Hands	Scores	Poker Hands	Scores
Nothing	0	Flush	5
One Pair	1	Full House	6
Two Pairs	2	Four of a Kind	7
Three of a Kind	3	Straight Flush	8
Straight	4	Royal Straight Flush	9

3 Partner Agent

3.1 Relationship between Agent and Players

A human player plays a poker game in cooperation with the partner agent against an opponent player. It is assumed that when the partner agent is called by a human player according to game situations, the agent presents its strategy to a human player. Furthermore, when a human player has some questions on some strategy, the partner agent answers these questions. It is also assumed that the partner agent has same information as the one a human player does. Fig.1 shows the decision-making procedure of the partner agent, which has the decision-making module and the cooperation module.

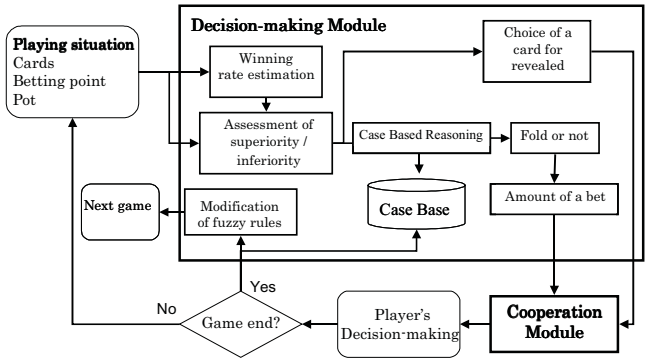


Figure 1: Decision-making procedure of partner agent

3.2 Decision-making Module

The decision-making module of the partner agent consists of six parts, winning rate estimation, assessment of superiority/inferiority, decision-making to fold or not, decision-making on betting points, choice of a card to be revealed and modification of fuzzy rules.

3.2.1 Winning Rate Estimation

In order to estimate the winning rate of a partner player, the agent suppositionally assigns possible combination of cards to human and opponent players' hole cards and to their up cards dealt from now on. The agent makes a judgment on human partner player's winning or losing. The agent repeats this trial and counts the number of the partner player's winning. Let this repeating trial times and the number of human partner player's winning be N_{times} and N_{win}, respectively. The winning rate is defined by

$$P_{win} = \frac{N_{win}}{N_{times}} \cdot \tag{1}$$

In this paper, N_{times} is fixed to be 10000.

3.2.2 Assessment of Superiority/Inferiority

The relative strength of the human partner player's poker hand to the opponent player's one is defined by

$$expect = 2(P_{win} - 0.5) . \tag{2}$$

The *expect* means that the larger the *expect* is, the higher the possibility of the superiority of a human player's hand is. In this paper the obtained *expect* is adjusted by fuzzy inference using fuzzy rules considering the tendency of opponent player's strategy, betting points and the number of turn. Table 2 shows an example of fuzzy rules at the early game stage, where the turn is normalized by 7, which is the maximum number of betting round times in a seven-card-stud poker game, and betting points are normalized by the pot. As Table 2 shows, the fuzzy rules have parameter values $t_{ij} \in [0,100]$ called the tendency values. The tendency value means that if t_{ij} is larger than 50, an opponent player tends to take aggressive strategy, and if t_{ij} is smaller than 50, an opponent player tends to take passive strategy [14]. The initial values of all tendency values are defined to be 50 and the tendency values are changed every game according to the opponent player's strategy.

Table 2: Fuzzy rules for assessment of superiority/inferiority at early game

Adjustment degree (Turn is Early)		Opponent's bet		
		Few	Middle	Much
Hand strength	Weak	Negative t_{11} Positive	Negative t_{12} Positive	Negative t_{13} Positive
	Strong	Negative t_{21} Positive	Negative t_{22} Positive	Negative t_{23} Positive

ISBN: 978-989-95079-6-8

The fuzzy inference using these fuzzy rules is performed as follows. Let the result of fuzzy inference using the Positive consequent part and that using the Negative consequent part be $C_{positive}$ and $C_{negative}$, respectively, where the min-max inference method and the center of gravity are used in the fuzzy inference [18].

$$infer = C_{positive} \times (\frac{t_{ij}}{100}) + C_{negative} \times (1 - \frac{t_{ij}}{100}) . \quad (3)$$

And $expect$ is adjusted in the range of [0,1] using the result of the fuzzy inference (3) by

$$result = \begin{cases} expect + (1 - expect) \times infer \\ \qquad (0 \le infer \le 1) \\ expect + (expect + 1) \times infer \\ \qquad (-1 \le infer \le 0) \end{cases} . \quad (4)$$

The $result$ is the assessment result of the superiority/ inferiority of a human player's poker hand.

3.2.3 Case Based Reasoning

In order to use case examples in past games for decision-making, case based reasoning (CBR) [17] is applied in this paper. The following items are considered as indices of problem features used in CBR: (i) $result$ in Eq. (4), (ii) opponent's bet, (iii) change of $result$ which is the difference between $result$ of the current situation and $result$ of the previous situation, (iv) change of opponent's bet which is the difference between opponent's bet of the current situation and opponent's bet of the previous situation, (v) $potential$ in Eq. (7), (vi) change of potential which is the difference between $potential$ of the current situation and potential of the previous situation, (vii) the number of turn.

The following items are considered as solutions of a problem: (i) a player folds a game or not, (ii) strategy is successful or not. The similarity degree of each item is obtained by

$$Sim_i = 1 - |Past_i - Current_i| , \quad (5)$$

where $Past_i (i = 1,2,...,n)$ is index of the i-th problem feature in a past case, $Current_i (i = 1,2,...,n)$ is the index of the i-th problem feature in a current case, and n is the number of problem features. In this paper, n is fixed at 7. The whole similarity degree is obtained by

$$Sim_{whole} = \frac{\sum_{i=1}^{n} Sim_i}{n} . \quad (6)$$

Case with the largest whole similarity degree larger than 0.9 is defined as the most similar case.

3.2.4 Decision to Fold or Not

From the first through the fifth turns the partner agent has the following decision procedures. When the agent assesses the partner player's hand at superiority and continuing a game is chosen as strategy by CBR, then the partner agent advises to continue a game. However, when folding a game is chosen as strategy by CBR and the whole similarity degree is larger than 0.95, then the partner agent advises to fold a game even if the agent assesses the hand at superiority. When the partner agent assesses the partner player's hand at inferiority, the partner agent makes a decision by the fuzzy inference whether a partner player should fold a game or not. An example of the rules is; if opponent's betting points are large and the number of turn is large, then fold a game with high possibility. If the fuzzy inference result($\in [0,1]$), i.e., the possibility of folding a game, is larger than a uniformly random generated number($\in [0,1]$), the partner agent makes a decision to fold a game. However, when continuing a game is chosen as strategy by CBR and the whole similarity degree is larger than 0.95, then the partner agent advises to continue a game even if the agent assess the partner's hand at inferiority.

In the sixth and the seventh turns, a poker hand is already determined because the seventh card is dealt. However, even if the hand strength is assessed at inferiority, there is sometimes a case where it is possible to take bluff strategy. The presented agent decides whether bluff strategy should be taken or not according to the following procedures.

In order to choose a hole card to be revealed, the potentiality is obtained by the following way. Assuming that one partner player's hole card is revealed, possible combinations of cards are assigned to remaining hole card(s). The expected value of the strength of possible poker hands is obtained by

$$potential = \sum_{i=1}^{m} u_i p(u_i) , \quad (7)$$

where m is the number of partner's possible poker hands, u_i is the score of the possible hand as shown in Table 1, and $p(u_i)$ is the probability of u_i. This expected value is called $potential$ in this paper. In the same way, the agent obtains the expected value of opponent player's hand called $Opp\_potential$. If Eq. (8) holds good, the partner agent decides not to fold and to take a bluff strategy, when Eq. (8) is considered by previous work [15]. However, even if the partner agent decides not to fold, when folding a game is chosen as strategy by CBR and the whole similarity degree is larger than 0.95, then the partner agent advises to fold a game.

$$potential \ge Opp\_potential + 1.3 \quad (8)$$

3.2.5 Decision on Betting Points

Let the upper limit and the lower limit of betting points be $upper$ and $lower$, respectively, where these limits are dependent on the betting round and are provided by the poker game rules. In this paper the upper limit is assumed to be the total points betted so far and the lower limit is assumed to be 5 points. The bet points are obtained by

$$C_{bet} = lower + a(upper - lower) , \quad (9)$$

where a is defined by

ISBN: 978-989-95079-6-8

$$a = \frac{result + 1}{2} + noise\,, \tag{10}$$

where $noise$ is a normally distributed random number with the mean value 0 and the variance 0.1. If bluff strategy is considered by the agent, C_{bet} in Eq. (9) is obtained using a defined by

$$a = \frac{(potential/9) + 1}{2}, \tag{11}$$

where $potential$ is obtained by Eq.(7) and is normalized by 9, i.e., the maximum score of poker hands defined by Table 1. This means that the agent bets large points in order to make the partner player's hand look strong using bluff strategy.

3.2.6 Choice of a Card to be Revealed

In the fifth and the sixth turns, after the betting round is finished, each player chooses one hole card to be revealed and turns it face up. When the agent assesses a partner's hand at superiority and the hand is already shown in face-up cards, a card to be revealed is chosen at random. When the agent assesses a partner's hand at superiority and the hand is hidden in hole cards, the agent advises to reveal the card at the rate of 80%, of which *potential* is the lowest. The agent advises to reveal the card at the rate of 20%, of which *potential* is the second lowest. When the agent assesses a partner's hand at inferiority, the agent advises to reveal the card of which *potential* is the highest.

3.2.7 Modification of Fuzzy Rules

All cards are shown at the game end and both players' hands are revealed. Every one game the modification part of fuzzy rules estimates the change of the tendency value considering the difference between the estimated relative hand strength and actual hand strength by using fuzzy inference. The modification is performed for the fuzzy rule of which satisfaction degree is the highest in the assessment of superiority/inferiority. The fuzzy rules for modifying the fuzzy rules for assessment of superiority/inferiority have the following form of statement. (1) If the estimated relative poker hand strength is lower than the actual poker hand strength, then the tendency value is changed to be small. (2) If the estimated relative poker hand strength is higher than the actual poker hand strength, then the tendency value is changed to be large. (3) If the estimated relative poker hand strength and the actual one are even, then the tendency value is not changed.

The tendency value is actually changed considering the last five estimated changing values of the tendency value and the latest changing value in order to consider opponent player's tendency in a long period.

3.3 Cooperation Module

The cooperation section has two parts. One part gives a human partner player strategy to take in the current game situation and its grounds. The other gives a human player reply to human player's questions about the current game situation.

3.3.1 Strategy Presentation

Agent's strategy presentation to a human partner player is performed at the decision-making process in each betting round or at the process of the human player's choice of a hole card to be revealed after the fifth turn. When the partner agent is called by a human player, the agent gives strategy for the current game situation.

3.3.1.1 Decision to fold or not to fold

At the beginning of each betting round, when a human player calls the partner agent, the partner agent presents its decision and its grounds to a human player according to the assessment of superiority/inferiority obtained by Eq. (4). If the agent decides to fold according to 3.2.4, then the agent presents its grounds to a human player. And when there are case examples similar to the current situation in a game, the partner agent presents a CBR result to a human player. For example, the partner agent says that you have past similar case examples, you folded games in the past similar situations, and that folding a game is good solution for the current situation.

3.3.1.2 Decision of amount of bed

The partner agent presents the point value obtained by 3.2.5. For example, when the agent decides the betting point to be 10, the agent says that 10 points should be bet for the current situation.

3.3.1.3 Decision to call

After an opponent player bets some points, these decision procedures are performed. The agent decides to fold or not as the same procedures in 3.4.1.1

3.3.1.4 Choice of a card to be revealed

The agent presents a card chosen by the procedure in 3.2.6

3.3.1.5 A human player takes another strategy

Although the agent presents strategy to fold a game to a human player, if a human player tries to take different strategy from the presented one, the agent calls the attention to a human player about the game continuation. For example, the agent presents *you should fold this game, because your hand is weak and opponent player's bet is large*.

3.3.2 Answer to Player's Question

When a human partner player asks the agent some question about the current game situation, the agent replies to it.

Table 3: Linguistic expressions corresponding to t_{ij}

t_{ij}	Linguistic Expressions
100-70	Very aggressive
70-60	Aggressive
60-50	Rather aggressive
50-40	Rather passive
40-30	Passive
30-0	Very passive

ISBN: 978-989-95079-6-8

3.3.2.1 Human player's game playing tendency

When the agent gets a question about human player's game playing tendency, the agent presents linguistic expressions based on the tendency value t_{ij} explained in 3.2.2. Although the tendency value t_{ij} is essentially obtained for an opponent player, the partner agent can obtain not only an opponent player's tendency value but also a partner player's one by the same procedures described in 3.2.7. For example, the agent says that a human partner player tends to be passive in the game situation such as the current one. Table 3 shows the linguistic expressions corresponding to t_{ij} .

3.3.2.2 Estimated expected opponent player's hand

The agent performs simulation repeatedly by assigning possible combination of cards to opponent player's hole cards and/or opponent player's cards dealt from now on. Then, the agent estimates the opponent player's poker hand with the highest hand probability and the poker hand with its probability. The agent presents these poker hands and their estimated probabilities to a human partner player. For example, the agent says that the opponent player's hand with the highest probability is One Pair with 46.9% and that the estimated highest score hand is Straight with 10.5%. If a human partner player asks the agent about other poker hands, the agent also replies them.

3.3.2.3 Estimated expected human player's hand

The agent performs simulation repeatedly by assigning possible combination of cards to a human partner player's cards to be dealt from now on, and obtains the partner player's poker hand with the highest hand probability and the highest score poker hand with its probability. Then, the agent presents these poker hands and their estimated probabilities to a partner player. If a human partner player asks the agent about other poker hands, the agent also replies them.

3.3.2.4 Estimated human player's hand strength from opponent player's side

The agent performs simulation repeatedly by assigning possible combination of cards to a human partner player's hole cards and cards to be dealt from now on, and obtains the partner player's poker hand with the highest hand probability and the highest score poker hand with its probability. Then the agent presents these poker hands and their estimated probabilities to a partner player. These probabilities give much information for decision making to a human player whether bluff strategy can be taken or not. If a human partner player asks the agent about other poker hands, the agent also replies them.

4 Experiments

4.1 Experimental Procedures

Subject experiments are carried out in order to evaluate the effectiveness of the presented agent. Each subject plays 30 poker games with the presented partner agent against the poker playing system that has the winning rate estimation part added to our previous playing system, and other 30 poker games without the partner agent. If all 30 games are finished by a showdown, prepared cards are set to have 15 wins and 15 loses, respectively. Subjects are divided into two groups in order to analyze playing strategy in every game with/without the partner agent. Cards in each card set are dealt to all subjects in the same order. That is, all subjects play games on the same condition in the experiments. However, subjects do not know it.

The number of subjects is 12. Although they are novices of a poker game, they have knowledge on poker game rules.

4.2 Experimental Results

4.2.1 Gain/Loss of subjects

Table 4 shows the overall gain/loss points of each subject with/without the partner agent against the playing system. Wilcoxon signed-rank tests [18] of gain/loss points of all subjects with/without the partner agent are carried out. The observed value is 0.98, and significance level 5% is 1.65 on directional test. The null hypotheses are not rejected. This test result shows that there is no difference between gain/loss points with the partner agent and those without the partner agent. However, it is found that nine subjects out of twelve can get more points with the partner agent than without it. Subjects B and H bet large point in betting rounds because both subject and partner agent assess their poker hand at superiority. Therefore, they lose a lot of points. For example, although player hand is tree of a kind, however, opponent player hand is straight, and player and partner agent assess their poker hand at superiority. It is the reason why there is no difference between gain/loss points with the partner agent and those without the partner agent.

Table 4: Gain/Loss of subjects

Subject	without partner agent	with partner agent
A	-17925	2313
B	18217	-79140
C	8193	13817
D	14789	110104
E	12015	-12367
F	-225734	-87563
G	13817	87509
H	-293177	-400371
I	21123	52980
J	19055	21763
K	-52172	-16293
L	-121285	-31182

4.2.2 Evaluations of partner agent

4.2.2.1 Usefulness of partner agent

After the experiments, subjects evaluate the usefulness of the partner agent with a 5-points grade, 5: very useful, 4: useful, 3: neutral, 2: useless, 1: very useless. Fig. 2 shows the evaluation results. Ten subjects out of twelve mark 4 or 5 and the average score among all subjects is 4.2, i.e., useful. And there are no subjects marking 2 or 1. Then subjects' evaluation of the partner agent is high. As a mater of fact, some subjects have comments that the partner agent is useful for the estimation of

ISBN: 978-989-95079-6-8

opponent player's hand and the estimation of superior/inferior of the partner player's hand. And other subjects have comments that decision of betting points becomes easier with the partner agent than without agent because decision of betting points by himself is very difficult for him.

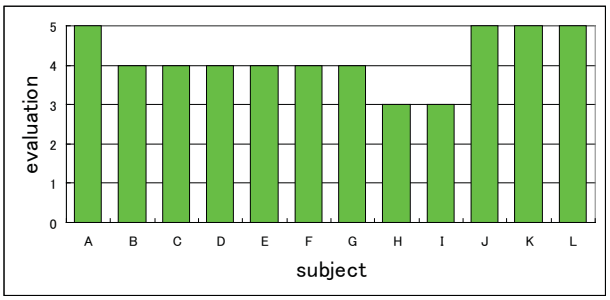


Figure 2: Usefulness of partner agent

4.2.2.2 Learning effect of partner agent

In order to confirm the learning effect of the presented partner agent, after the experiments, the subjects also evaluate whether the partner agent's advices are helpful or not with a 5-points grade, 5: very helpful, 4: helpful, 3: neutral, 2: unhelpful, 1: very unhelpful. Fig. 3 shows the evaluation results. Seven subjects out of twelve mark 4 or 5 and the average score among all subjects is 3.6, i.e., more or less helpful. And there are no subjects marking 1. In this paper, the learning effect is simply evaluated by subjective assessment and subjects' comments analyze the learning effect objectively. Some subjects have comments that they feel that the partner agent's advises are helpful because the partner agent gives them advices referring to case examples them advices are intelligible to them. It is found that the agent gives partner players good advices using CBR. However, some subjects negative comment. Subject H evaluates the partner agent's advices are not helpful because although he wants to play a game more aggressively, the partner agent advises him to fold a game referring to case examples.

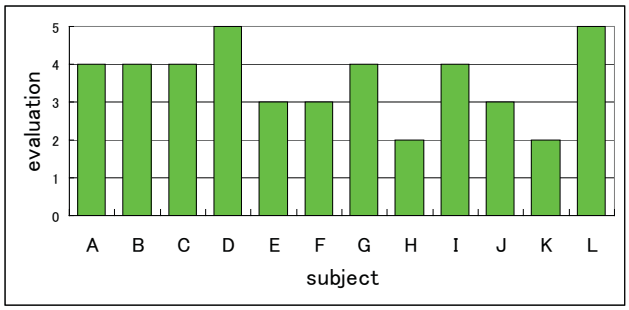


Figure 3: Helpfulness of partner agent's advices

5 Conclusions

This paper constructs a partner agent that supports a human player in playing a seven-card stud poker game. This agent presents its own strategy and/or replies some questions about game strategy in the current game situation when a human player calls it. In order to make use case examples in past games, CBR is also applied. Subject experiments are performed to evaluate the effectiveness of the presented partner agent. Experimental results show that although subjects do not necessarily get more points with the partner agent than without the agent from the viewpoints of gain/loss points, subject's evaluation of usefulness of the partner agent is high. And the evaluation of helpfulness of partner agent's advices is more or less high. It is found that CBR is useful to the presented partner agent.

References

[1] Toshimasa Moriwaki, group, organization, Tokyo Daigaku Shuppan, 2000

[2] Koki Matsura, Negotiation Simulation System Based on Fuzzy Decision-Making, Journal of Japan Society for Fuzzy Theory and System, Vol.8, No.5, pp. 882-890 1996

[3] D.Dubois, M. Grabisch, F. Modave and H. Prade, Relating decision under uncertainty and multicriteria decision making models, Frontiers in Soft Computing and Decision Systems, pp. 6-15, 1997

[4] Gleb Beliakov and Jim Warren, Appropriate Choice of Aggregation Operators in Fuzzy Decision Support System, IEEE TRANSACTION ON FUZZY STSTENS, vol. 9, No. 6, pp. 773-783, 2001

[5] John Zeleznikow, James R. Nolan, Using soft computing to build real intelligent decision support system in uncertain domains, Decision Support 31, pp. 263-285, 2001

[6] Chang-Shing Lee, Chen-Yu Pan, An intelligent fuzzy agent for meeting scheduling decision support system, Fuzzy Set and Systems 142, pp. 467-488, 2004

[7] Norio Sasaki: Introduction of game theory, Nippon Hyoron Sha, 2003

[8] Toshiaki Yasumura, Hidehisa Akiyama, Kunihiko Oguti, Katsumi Nitta: Negotiation Agents in the MONOPOLY Game, Information Processing Society Japan Vol43, No.10, pp. 3048-3055, 200

[9] Darse Billings, Aron Davidson, Jonathan Schaeffer, Duane Szafron, The challenge of poker, Artificial Intelligence 134, pp. 201-240, 2002

[10] Graham Kendall, Mark Willdig, An Investigation of an Adapting Poker Player, Artificial Intelligense, LNAI 2246, pp.189-200, 2001

[11] Arild Sandven and Bjorner Tessem, A Case-Based Learner for Poker, Proceedings of the Ninth Scandinavian Conference on Artificial Intelligence (SCAI 2006) pp. 159-167.

[12] Takehisa ONISAWA, Satoru KAZAMI, Chiharu TAKAHASHI, Construction of Playing System of Game with Imperfect Information -Seven-Card Stud Poke-, Journal of Japan Society for Fuzzy Theory and System, Vol.15, No.1,pp. 127-141, 2003

[13] Takehisa Onisawa and Chiharu Takahashi, Bluff Strategy in Seven-card-Stud Poker Game, Proceedings of the 10th IFSA World Congress, Istanbul, TURKY pp. 491-494, 2003

[14] T. Onisawa and C.Takahashi, Poker Playing system Considering Opponent Player's Strategy, Proceedings of SICE Annual Conference 2004 in Sapporo, pp. 2347-2352, 2004

[15] K. Ohsone, T. Onisawa, Cooperative Partner agent of Imperfect Information Game 2008 IEEE Conference on Soft Computing in Industrial Applications (SMCia/08), June 25-27, Muroran, JAPAN pp. 60-65, 2008

[16] R. Kruse, J. Gebhardt, F.Klawonn Foundations of FUZZY SYSTEMS, JOHN WILEY & SONS, 1994

[17] Yoshio Nakatani, Recent Developments in Case-Based Reasoning, Journal of the Japanese society for Artificial Intelligence, Vol.17 No.1 pp. 28-33, 2002

[18] Ann Hughes, Dennis Grawoig, Statistics: A Foundation for Analysis, ADDISON-WESLEY PUBLISHING COMPANY, 1971

ISBN: 978-989-95079-6-8

Systems of Fuzzy Relation Equations in a Space with Fuzzy Preorder

Irina Perfilieva

Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Ostrava, Czech Republic
Email: Irina.Perfilieva@osu.cz

Abstract— In this paper, we consider the problem of solving systems of fuzzy relation equations in a space with fuzzy preorder. Two types of these systems with different compositions are considered. New solvability criteria are proposed for systems of both types. The new criteria are weaker than all the known ones that are based on the assumption that fuzzy sets on the left-hand side of a system establish a fuzzy partition of a respective universe.

Keywords— system of fuzzy relation equations, fuzzy preorder, criterion of solvability

1 Preliminaries

Let throughout this contribution $\mathscr{L} = \langle L, \vee, \wedge, *, \rightarrow, 0, 1\rangle$ be an integral, residuated, commutative l-monoid (*a residuated lattice*), X a non-empty set and L^X a set of L-valued functions on X. Fuzzy subsets of X are identified with L-valued functions on X (membership functions).

Let X and Y be two universes, not necessary different, $A_i \in L^X$, $B_i \in L^Y$ arbitrarily chosen fuzzy subsets of respective universes, and $R \in L^{X\times Y}$ a fuzzy subset of $X \times Y$. The latter is called a fuzzy relation. Lattice operations $\vee$ and $\wedge$ induce respective union and intersection of fuzzy sets. Two other binary operations $*, \rightarrow$ of $\mathscr{L}$ are used for compositions - binary operations on $L^{X\times Y}$. We will consider two of them: sup-*-composition that is usually denoted by $\circ$, and $\inf$-$\rightarrow$-composition that is denoted by $\triangleleft$. The first one has been introduced by L. Zadeh [16] and the second one by W. Bandler and L. Kohout [1]. We will demonstrate definitions of both compositions on particular examples of set-relation compositions $A \circ R$ and $A \triangleleft R$ where $A \in L^X$ and $R \in L^{X\times Y}$:

$$(A \circ R)(y) = \bigvee_{x\in X} (A(x) * R(x, y)),$$
$$(A \triangleleft R)(y) = \bigvee_{x\in X} (A(x) \rightarrow R(x, y)).$$

Remark 1
Let us remark that both compositions can be considered as set-set compositions where R is assumed to be replaced by a fuzzy set. In this reduced form they are used in instances of systems of fuzzy relation equations below.

By a *system of fuzzy relation equations with sup-*-composition* (SFRE$*$), we mean the following system of equations

$$A_i \circ R = B_i, \quad 1 \le i \le n, \tag{1}$$

that is considered with respect to unknown fuzzy relation $R \in L^{X\times Y}$. Its counterpart is a *system of fuzzy relation equations with* $\inf$-$\rightarrow$*composition* (SFRE$\rightarrow$)

$$A_i \triangleleft R = B_i, \quad 1 \le i \le n, \tag{2}$$

that is considered with respect to unknown $R \in L^{X\times Y}$ also. System (1) and its potential solutions are well investigated in the literature (see e.g. [3, 2, 4, 5, 12, 8, 13, 15]). On the other hand, investigation of solvability of (2) is not so intensive (see [2, 10]).

Both systems of fuzzy relation equations arise when a system of fuzzy IF-THEN rules is modeled by a fuzzy relation (below in (3) it is denoted by R), and continuity of the model [9] is requested. In order to explain this request, we recall that in relation models, a computation of an output value (B) which relates to a given input $A \in L^X$ is performed with the help of sup-*, respectively $\inf$-$\rightarrow$ composition:

$$B = A \circ R \quad \text{or} \quad B = A \triangleleft R. \tag{3}$$

(3) is a computational realization of the *Generalized Modus Ponens* inference scheme in fuzzy logic (in a broader sense). It is often welcome if thus constructed model is *continuous* in the sense that when (input) fuzzy sets $A', A'' \in L^X$ are close to each other (in some space) so do output fuzzy sets $A' \circ R$ and $A'' \circ R$ (respectively, $A' \triangleleft R$ and $A'' \triangleleft R$). We proved in [9] that this is possible if and only if R solves the respective system (1) or (2) of fuzzy relation equations. This fact gives additional importance to the problem of solvability of systems of fuzzy relation equations.

In general, solutions of (1) or (2) may not exist. Therefore, investigation of necessary and sufficient conditions for solvability (or at least, sufficient conditions) is needed. This problem has been widely studied in the literature, and some nice theoretical results have been obtained in the cited above papers.

If the universes of discourse X and Y are infinite then the complexity of verification of necessary and sufficient conditions is comparable with the direct checking of solvability. Therefore, the problem of discovering easy-to-check solvability conditions or criteria is still open (see [11, 14] for some results). Besides other, this paper is a contribution to this topic.

We recall basic facts concerning solvability of system (1) of fuzzy relation equations (similar conditions are known for system (2), so that we will not recall theme (see e.g. [2, 10])).

Theorem 1
(a) [13] If system (1) with respect to unknown fuzzy relation R is solvable then relation

$$\hat{R}(x, y) = \bigwedge_{i=1}^{n} (A_i(x) \rightarrow B_i(y)) \tag{4}$$

is the greatest solution to (1).

ISBN: 978-989-95079-6-8

(b) *[5] Let fuzzy sets $A_i \in L^X$ and $B_i \in L^Y$, $1 \leq i \leq n$, be normal. Then the fuzzy relation*

$$\check{R}(x,y) = \bigvee_{i=1}^{n} (A_i(x) * B_i(y)) \tag{5}$$

is a solution to (1) if and only if

$$\bigvee_{x\in X} (A_i(x) * A_j(x)) \leq \bigwedge_{y\in Y} (B_i(y) \leftrightarrow B_j(y)) \tag{6}$$

holds for all $i, j = 1, \ldots, n$.

(c) *[7] Let fuzzy sets $A_1, \ldots, A_n \in L^X$ be normal and $x_1, \ldots, x_n \in X$ be pairwise different elements such that for all $i = 1, \ldots, n$, $A_i(x_i) = 1$. Let moreover, for all $i, j = 1, \ldots, n$,*

$$\bigvee_{x\in X} (A_j(x) * A_i(x)) \leq \bigwedge_{x\in X} (A_j(x) \leftrightarrow A_i(x)) \tag{7}$$

holds true. Then (1) is solvable if and only if

$$\bigvee_{x\in X} (A_j(x) * A_i(x)) \leq \bigwedge_{y\in Y} (B_i(y) \leftrightarrow B_j(y)).$$

In this contribution, we will prove weaker criteria of solvability for systems (1) and (2) than those ones, consider above in cases (b) and (c) (see Section 5 with the Discussion). Moreover, an algorithm of verifying solvability that is based on each proposed here criterion has a polynomial complexity.

2 Fuzzy Preorders and Their Normal Upper Sets

This section is auxiliary with respect to the problem of solvability discussed above. In this section, we will first recall basic facts that relate to spaces with fuzzy preorder [6]. Then we will prove new results that are used in Sections below where we consider the problem of solvability.

We will fix $\mathscr{L} = \langle L, \vee, \wedge, *, \rightarrow, 0, 1\rangle$ and a non-empty universe X. Recall that a binary fuzzy relation is a *$*$-fuzzy preorder* if it is reflexive and $*$-transitive. Fuzzy preorder $Q : X \times X \longrightarrow L$ on X can be generated by an arbitrary family of fuzzy subsets $(A_i)_{i\in I}$ of X:

$$Q(x,y) = \bigwedge_{i\in I} (A_i(x) \rightarrow A_i(y)).$$

If Q is a fuzzy preorder on X then fuzzy set $A \in L^X$ is called [6] an *upper set* of Q if

$$A(x) * Q(x,y) \leq A(y), \quad x, y \in X.$$

A necessary and sufficient condition that a family of fuzzy subsets of X constitutes a family of upper sets of Q has been proved in [6]. A new result will be proved below in Theorem 2 for a family of normal fuzzy subsets of X. Let us remark that our assumptions are different from those in [6].

Theorem 2
Let $(A_i)_{i\in I} \subseteq L^X$ be a family of normal fuzzy subsets of X and $(x_i)_{i\in I} \subseteq X$ be a family of pairwise different elements such that for all $i \in I$, $A_i(x_i) = 1$. Then the following two statements are equivalent:

(i) there exists fuzzy preorder Q on X such that for each $i \in I$, $A_i(x) = Q(x_i, x)$, $x \in X$.

(ii) For all $i, j \in I$,

$$A_i(x_j) \leq \bigwedge_{x\in X} (A_j(x) \rightarrow A_i(x)). \tag{8}$$

PROOF: It is easy to see that (i) $\Rightarrow$ (ii) so that we will prove the reverse implication. Assume that $(A_i)_{i\in I} \subseteq L^X$ is a family of normal fuzzy subsets of X, and (8) holds true. If $Q(x,y) = \bigwedge_{i\in I}(A_i(x) \rightarrow A_i(y))$ then by the assertion above, Q is a fuzzy preorder on X. We will show that statement (i) is valid for Q.

Let us choose and fix i, $i \in I$. For all $x \in X$, and arbitrary $j \in I$, such that $i \neq j$, we have:

$$Q(x_i, x) = \bigwedge_{j\in I} (A_j(x_i) \rightarrow A_j(x)) \leq$$
$$\leq A_i(x_i) \rightarrow A_i(x) = A_i(x).$$

On the other hand, by (8), $A_j(x_i) \leq A_i(x) \rightarrow A_j(x)$ or, equivalently, $A_i(x) \leq A_j(x_i) \rightarrow A_j(x)$. Therefore,

$$A_i(x) \leq \bigwedge_{j\in I} (A_j(x_i) \rightarrow A_j(x)) = Q(x_i, x).$$

Hence, for all $x \in X$, $A_i(x) = Q(x_i, x)$. □

Corollary 1
Let $(A_i)_{i\in I} \subseteq L^X$ be a family of normal fuzzy subsets of X and $(x_i)_{i\in I} \subseteq X$ be a family of pairwise different elements such that for all $i \in I$, $A_i(x_i) = 1$. Then $Q(x,y) = \bigwedge_{i\in I}(A_i(x) \rightarrow A_i(y))$ is the coarsest fuzzy preorder on X such that (8) holds true.

The following lemma gives another necessary and sufficient condition that a family of normal fuzzy subsets of X constitutes a family of upper sets of Q. We will use that condition in our new criteria of solvability.

Lemma 1
Let A_i, $i \in I$, be a family of normal fuzzy subsets of L^X, such that $A_i(x_i) = \mathbf{1}$ for the respective $x_i \in X$, $i \in I$. Moreover, let for all $i, j \in I$, inequality (8) hold true. Then inequality (8) turns to the equality

$$A_i(x_j) = \bigwedge_{x\in X} (A_j(x) \rightarrow A_i(x)). \tag{9}$$

PROOF: Assume that for all $i, j \in I$, (8) holds true, i.e. $A_i(x_j) \leq \bigwedge_{x\in X}(A_j(x) \rightarrow A_i(x))$. On the other hand,

$$\bigwedge_{x\in X} (A_j(x) \rightarrow A_i(x)) \leq A_j(x_j) \rightarrow A_i(x_j) = A_i(x_j)$$

so that (9) follows.

Assume that for all $i, j \in I$, (9) holds true. Then (8) follows immediately. □

ISBN: 978-989-95079-6-8

3 SFRE$*$ in a Space with a Fuzzy Preorder and Their Solvability

Assume that $\mathscr{L}$, X, Y are as above, and we are given fuzzy sets $A_1, \ldots, A_n \in L^X$ and $B_1, \ldots, B_n \in L^Y$. In this section we will investigate the problem of solvability of system (1) and propose a new criterion that is weaker than all criteria, based on the assumption that fuzzy sets $A_1, \ldots, A_n$, establish a fuzzy partition of X, i.e. that they are classes of a respective similarity on X.

In order to simplify denotation we will choose and fix $y \in Y$ and work with the following instance of the system:

$$A_i \circ r = b_i, \quad 1 \leq i \leq n, \tag{10}$$

where $b_i = B_i(y)$, $1 \leq i \leq n$. In this particular case, system (10) is considered with respect to unknown fuzzy set $r \in L^X$.

Theorem 3
Let fuzzy sets $A_1, \ldots, A_n \in L^X$ be normal and $x_1, \ldots, x_n \in X$ be pairwise different elements such that for all $i = 1, \ldots, n$, $A_i(x_i) = 1$. Let moreover, for all $i, j = 1, \ldots, n$,

$$A_i(x_j) = \bigwedge_{x \in X} (A_j(x) \to A_i(x)) \tag{11}$$

holds true. Then system (10) is solvable if and only if

$$A_i(x_j) \leq (b_j \to b_i). \tag{12}$$

PROOF: By Lemma 1, (11) is equivalent to

$$(\forall i, j)(\forall x \in X) \quad (A_j(x) \leq (A_i(x_j) \to A_i(x))). \tag{13}$$

Assume that system (10) is solvable. Then

$$\hat{r}(x) = \bigwedge_{j=1}^{n} (A_j(x) \to b_j)$$

is a solution so that

$$(\forall j) \quad \bigvee_{x \in X} (A_j(x) * \hat{r}(x)) \geq b_j.$$

Let us fix j, $j = 1, \ldots, n$. By (13),

$$(\forall k) \quad \bigvee_{x \in X} ((A_k(x_j) \to A_k(x)) * \hat{r}(x)) \geq b_j.$$

Then

$$\bigvee_{x \in X} ((A_k(x_j) \to A_k(x)) * (A_k(x) \to b_k)) \geq b_j$$

so that

$$A_k(x_j) \to b_k \geq b_j.$$

The last inequality is equivalent to

$$A_k(x_j) \leq b_j \to b_k,$$

and by arbitrariness of k, j, it is equivalent to (12).

On the other hand, assume that (12) holds true. We will prove that $\hat{r}$ is a solution of system (10). Let us fix i, $i = 1, \ldots, n$ and prove that $\hat{r}$ solves the i-th equation of (10). Indeed,

$$\bigvee_{x \in X} (A_i(x) * \hat{r}(x)) \leq \bigvee_{x \in X} (A_i(x) * (A_i(x) \to b_i)) \leq b_i.$$

Let us prove the opposite inequality.

$$\bigvee_{x \in X} (A_i(x) * \hat{r}(x)) \geq A_i(x_i) * \hat{r}(x_i) = \bigwedge_{j=1}^{n} (A_j(x_i) \to b_j).$$

By (12), $A_j(x_i) \to b_j \geq b_i$, so that the opposite inequality easily follows:

$$\bigvee_{x \in X} (A_i(x) * \hat{r}(x)) \geq b_i.$$

Thus, $\hat{r}$ is a solution of (10), and the system is solvable. □

4 SFRE$\to$ in a Space with a Fuzzy Preorder and Their Solvability

Assume that $\mathscr{L}$, X, Y and fuzzy sets $A_1, \ldots, A_n \in L^X$ and $B_1, \ldots, B_n \in L^Y$ are as above. In this section we will investigate the problem of solvability of system (2) and propose a new criterion. As above we will work with an instance of system (2), i.e. with the following equation:

$$A_i \triangleright r = b_i, \quad 1 \leq i \leq n, \tag{14}$$

where $b_i = B_i(y)$, $1 \leq i \leq n$. System (14) will be considered with respect to an unknown fuzzy set $r \in L^X$.

Theorem 4
Let fuzzy sets $A_1, \ldots, A_n \in L^X$ be normal and $x_1, \ldots, x_n \in X$ be pairwise different elements such that for all $i = 1, \ldots, n$, $A_i(x_i) = 1$. Let moreover, for all $i, j = 1, \ldots, n$, (11) holds true. Then system (14) is solvable if and only if

$$(\forall i, j) \quad A_i(x_j) \leq (b_i \to b_j). \tag{15}$$

PROOF: By Lemma 1, (11) is equivalent to

$$(\forall i, j)(\forall x \in X) \quad A_j(x) \leq (A_i(x_j) \to A_i(x)),$$

and can be equivalently transformed to

$$(\forall i, j)(\forall x \in X) \quad b_i * A_j(x_j) \leq (A_j(x) \to b_i * A_i(x)). \tag{16}$$

Assume that system (14) is solvable. Then

$$\check{r}(x) = \bigvee_{j=1}^{n} (A_j(x) * b_j)$$

is a solution so that

$$(\forall j) \quad \bigwedge_{x \in X} (A_j(x) \to \check{r}(x)) \leq b_j,$$

and hence

$$(\forall j) \quad \bigwedge_{x \in X} (A_j(x) \to A_i(x) * b_i) \leq b_j,$$

ISBN: 978-989-95079-6-8

where $i = 1, \ldots, n$. By (16),

$$(\forall i, j) \quad b_i * A_j(x_j) \leq b_j$$

that is equivalent to (15).

On the other hand, assume that (15) holds true. We will prove that $\check{r}$ is a solution of system (14). Let us fix i, $i = 1, \ldots, n$ and prove that $\check{r}$ solves the i-th equation of (14). First we observe that

$$\bigwedge_{x \in X} (A_i(x) \to \check{r}(x)) \geq \bigwedge_{x \in X} (A_i(x) \to A_i(x) * b_i) \geq b_i.$$

On the other hand,

$$\bigwedge_{x \in X} (A_i(x) \to \check{r}(x)) \leq A_i(x_i) \to \check{r}(x) = \bigvee_{j=1}^{n} (A_j(x_i) * b_j).$$

By (15), for all i, j, $A_j(x_i) * b_j \leq b_i$, so that

$$\bigwedge_{x \in X} (A_i(x) \to \check{r}(x)) \leq b_i.$$

Thus, $\check{r}$ is a solution of (14), and the system is solvable. □

5 Discussion

In this section, we will show how the discovered criteria relate to those existed in literature. Moreover, we will justify our claim in the Abstract that the new criteria are weaker than all known ones that are based on the assumption that fuzzy sets on the left-hand side of a system establish a fuzzy partition of a respective universe. The following criterion is among those to which we have referred to as known ones (it has been recalled in Theorem 1). The formulation below is adapted to the instance (10).

Let fuzzy sets $A_1, \ldots, A_n \in L^X$ be normal and $x_1, \ldots, x_n \in X$ be pairwise different elements such that for all $i = 1, \ldots, n$, $A_i(x_i) = 1$. Let moreover, for all $i, j = 1, \ldots, n$,

$$\bigvee_{x \in X} (A_j(x) * A_i(x)) \leq \bigwedge_{x \in X} (A_j(x) \leftrightarrow A_i(x)) \qquad (17)$$

holds true. Then (10) is solvable if and only if

$$\bigwedge_{x \in X} (A_j(x) \to A_i(x)) \leq (b_i \leftrightarrow b_j).$$

Let us examine condition (7) and compare it with (11). For simplicity, assume that $n = 2$ and sets A_1, A_2 are ordinary (not fuzzy). It is easy to see that in this case, membership functions A_1, A_2 are characteristic functions of the respective sets and $A_1(x) * A_2(x)$ is a characteristic function of the intersection $A_1 \cap A_2$. For this particular case, (7) is fulfilled if and only if either $A_1 = A_2$ or $A_1 \cap A_2 = \emptyset$. Assume that $A_1 \subseteq A_2$ and $A_1 \neq A_2$. Then (7) fails while condition (11) is valid provided that $x_2 \notin A_1$.

Therefore, in Theorems 3, 4 we have obtained new criteria of solvability that have weaker assumptions than all those that explicitly or implicitly use the condition that fuzzy sets $A_1, \ldots, A_n \in L^X$ establish a fuzzy partition of X.

Conclusion

The problem of solvability of systems of fuzzy relation equations with two different compositions has been considered. We established new solvability criteria for systems of both types. The new criteria are based on the assumption that fuzzy sets on the left-hand side of a system are upper sets of a respective $*$-fuzzy preorder. This assumption is weaker than the assumption that fuzzy sets on the left-hand side of a system establish a fuzzy partition of a respective universe.

Acknowledgment

The paper has been supported partially by the grant IAA108270902 of GA AV ČR and partially by the project 1M0572 of the MŠMT ČR.

References

[1] W. Bandler, L. Kohout, Semantics of implication operators and fuzzy relational products, Int. J. Man-Machine Studies 12 (1980) 89–116.

[2] B. De Baets, Analytical solution methods for fuzzy relation equations, in: D. Dubois, H. Prade (eds.), The Handbooks of Fuzzy Sets Series, Vol.1, Kluwer, Dordrecht, 2000, pp. 291–340.

[3] A. Di Nola, S. Sessa, W. Pedrycz, E. Sanchez, Fuzzy Relation Equations and Their Applications to Knowledge Engineering, Kluwer, Boston, 1989.

[4] S. Gottwald, On the existence of solutions of systems of fuzzy equations, Fuzzy Sets and Systems 12 (1984) 301–302.

[5] F. Klawonn, Fuzzy points, fuzzy relations and fuzzy functions, in: V. Novák, I. Perfilieva (eds.), Discovering the World with Fuzzy Logic, Springer, Berlin, 2000, pp. 431–453.

[6] H. Lai, D. Zhang, Fuzzy preorder and fuzzy topology, Fuzzy Sets and Systems 157 (2006) 1865–1885.

[7] I. Perfilieva, Fuzzy function as an approximate solution to a system of fuzzy relation equations, Fuzzy Sets and Systems 147 (2004) 363–383.

[8] I. Perfilieva, S. Gottwald, Solvability and approximate solvability of fuzzy relation equations, Int. J. of General Systems 32 (2003) 361–372.

[9] I. Perfilieva, S. Lemhke, Correct models of fuzzy if–then rules are continuous, Fuzzy Sets and Systems 157 (2006) 3188–3197.

[10] I. Perfilieva, L. Nosková, System of fuzzy relation equations with $\inf \to$ composition: complete set of solutions, Fuzzy Sets and Systems 159 (2008) 2256–2271.

[11] I. Perfilieva, V. Novák, System of fuzzy relation equations as a continuous model of if–then rules, Information Sciences 177 (2007) 3218–3227.

[12] I. Perfilieva, A. Tonis, Compatibility of systems of fuzzy relation equations, Int. J. of General Systems 29 (2000) 511–528.

[13] E. Sanchez, Resolution of composite fuzzy relation equations, Information and Control 30 (1976) 38–48.

[14] M. Štěpnička, B. De Baets, L. Nosková, On additive and multiplicative fuzzy models, in: M. Štěpnička, V. Novák, U. Bodenhofer (eds.), New Dimensions in Fuzzy Logic and Related Technologies. Int. Conference EUSFLAT, Ostrava, 2007, pp. 95–102.

ISBN: 978-989-95079-6-8

[15] X. Wang, Infinite fuzzy relational equations on a complete Brouwerian lattice, Fuzzy Sets and Systems 138 (2003) 657–666.

[16] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning I, II, III, Information Sciences 8-9 (1975) 199–257, 301–357, 43–80.

ISBN: 978-989-95079-6-8

On $\mathbb{M}$-Approximative Operators and $\mathbb{M}$-Approximative Systems

Alexander Šostak
Department of Mathematics, University of Latvia
and Institute of Mathematics and Computer Sciences, University of Latvia
Riga, Latvia
Email: sostaks@latnet.lv

Abstract— *The concept of an $\mathbb{M}$-approximative system is introduced. Basic properties of the category of $\mathbb{M}$-approximative systems and in a natural way defined morphisms between them are studied. It is shown that categories related to fuzzy topology as well as categories related to rough sets can be described as special subcategories of the category of $\mathbb{M}$-approximative systems.*

Keywords— Approximative operator, approximative system, *cl*-monoid, fuzzy topology, L-topology, rough set

1 Introduction and motivation

In 1968, that is only 3 years after L. Zadeh has published his famous work *"Fuzzy Sets"*, thus laying down the principles of what can be called *Mathematics of Fuzzy Sets*, his student C.L. Chang [3] introduced the concept of a fuzzy topological space thus marking the beginning of Fuzzy Topology, the conterpart of General Topology in the context of fuzzy sets. Now Fuzzy Topology is one of the most well developed fields of Mathematics of Fuzzy Sets, and there are published dozens of fundamental works on this subject.

In 1982 Z. Pawlak [14] has introduced the concept of a rough set which can be viewed as a certain alternative for the concept of a fuzzy set for the study of mathematical problems of applied nature. Pawlak's work was followed by many other publications where rough sets and mathematical structures on the basis of rough sets were introduced, studied, and applied.

Although at the first glance it may seem that the concepts of a fuzzy set, of a (fuzzy) topological space and of a rough set are of an essentially different nature and "have nothing in common", this is not the case. Probably, the first one to start studying the intermediate relations between topologies, fuzzy sets and rough sets was J. Kortelainen [11], see also [12], etc. Further a detailed analysis of different relations between fuzzy sets, rough sets and some other related concepts was done in a series of papers by Y. Yao (e.g. [20]), and other researchers.

The aim of this work is to present an alternative view on the relations between fuzzy sets, fuzzy topological spaces and rough sets and to develop a framework allowing to generalize these concepts and corresponding theories. In order to realize this aim we introduce the concept of an $\mathbb{M}$-approximative system (cf [19]) and thus come to the category $\mathbf{AS}^{\mathbb{M}}$ of $\mathbb{M}$-approximative systems. Properties of this category are studied and connections between $\mathbf{AS}^{\mathbb{M}}$ and its subcategories related to fuzzy topology, fuzzy sets and rough sets are described.

2 The context

In our work two lattices will play the fundamental role. The first one is a complete infinitely distributive lattice

$$\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee),$$

whose top and bottom elements are $1_{\mathbb{L}}$ and $0_{\mathbb{L}}$ respectively. Besides sometimes we will assume that the lattice $\mathbb{L}$ is equipped with one of the following operations: a monotone mapping $^c : \mathbb{L} \to \mathbb{L}$ or a binary operation $* : \mathbb{L} \times \mathbb{L} \to \mathbb{L}$.

A lattice $\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee, ^c)$ will be called *adjunctive* if the pair $(^c, ^c)$ is an adjunction

$$(^c, ^c) : \mathbb{L} \vdash \mathbb{L}^{op},$$

that is $a \leq b^c \iff b \leq a^c \quad \forall\, a, b \in \mathbb{L}$, cf e.g. [4]. A lattice $\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee, ^c)$ will be called *involutive* if $^c : \mathbb{L} \to \mathbb{L}$ is an involution, that is if $(a^c)^c = a \quad \forall a \in \mathbb{L}$. One can easily see that in an adjunctive involutive lattice involution $^c : \mathbb{L} \to \mathbb{L}$ is order reversing:

$$a \leq b \implies b^c \leq a^c \; \forall a, b \in \mathbb{L},$$

and conversely, if $^c : \mathbb{L} \to \mathbb{L}$ is order reversing involution, then $(^c, ^c) : \mathbb{L} \vdash \mathbb{L}^{op}$ is an adjunction.

Concerning the second, binary operation $* : \mathbb{L} \times \mathbb{L} \to \mathbb{L}$ (conjunction) it will be assumed that $\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee, *)$ is a commutative cl-monoid (see e.g. [2]), that is
$*$ is commutative: $a * b = b * a$ for all $a, b \in \mathbb{L}$;
$*$ distributes over arbitrary joins:
$a * \left(\bigvee_{i\in\mathcal{I}} b_i\right) = \bigvee_{i\in\mathcal{I}}(a * b_i) \; \forall\, a \in \mathbb{L}, \; \forall \{b_i \mid i \in \mathcal{I}\} \subseteq \mathbb{L}$
and $a * 1_{\mathbb{L}} = a, \quad a * 0_{\mathbb{L}} = 0_{\mathbb{L}}$.

It is well-known (see e.g. [2]) that in a cl-monoid a further binary operation $\mapsto : \mathbb{L} \to \mathbb{L}$ (residuation) is defined related to conjunction $*$ by Galois connection:

$$a * b \leq c \iff a \leq b \mapsto c \; \forall a, b, c \in \mathbb{L}.$$

One can easily see that residuation is nonincreasing by the first argument and nondecreasing by the second argument, and that $b * (b \mapsto a) \leq a \; \forall a, b \in \mathbb{L}$. In particular $b * (b \mapsto 0) \leq 0$, and hence

$$b \leq (b \mapsto 0) \mapsto 0.$$

This allows to conclude, that by setting $a^c = a \mapsto 0$ we obtain an adjunction $(^c, ^c) : \mathbb{L} \vdash \mathbb{L}^{op}$. Indeed, if $a \leq b \mapsto 0$, then

$$b \leq (b \mapsto 0) \mapsto 0 \leq a \mapsto 0.$$

A cl-monoid is called a Girard monoid [10] if $(a \mapsto 0) \mapsto 0 = a \; \forall a \in \mathbb{L}$. Hence in case $\mathbb{L}$ is a Girard monoid, residuation $\mapsto$ induces an order reversing involution $^c : \mathbb{L} \to \mathbb{L}$.

An important situation in our research will be the following. Let $\mathrm{L} = (\mathrm{L}, \leq, \wedge, \vee)$ be a a lattice and X be a set. Then the L-powerset $\mathrm{L}^X =: \mathbb{L}$ becomes a lattice $(\mathbb{L}, \leq, \wedge, \vee)$ by pointwise extending the lattice structure from L to $\mathbb{L}$. Besides $\mathbb{L}$ is

ISBN: 978-989-95079-6-8

infinitely distributive whenever L was infinitely distributive. Moreover, if $\mathrm{L} = (\mathrm{L}, \leq, \wedge, \vee, ^c)$ is an adjunctive (involutive) lattice then by pointwise extending operation ^{c} from L to $\mathbb{L}$, an adjunctive (resp. involutive) lattice $\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee, ^c)$ is obtained. In case $\mathrm{L} = (\mathrm{L}, \leq, \wedge, \vee, *)$ is a cl-monoid, by pointwise extension of $* : \mathrm{L} \times \mathrm{L} \to \mathrm{L}$ to $* : \mathbb{L} \times \mathbb{L} \to \mathbb{L}$ we obtain a cl-monoid $\mathbb{L} = (\mathbb{L}, \leq, \wedge, \vee, *)$

The second lattice belonging to the context of our work is denoted by $\mathbb{M}$. At the moment we assume only its completeness, however sometimes it will be requested that $\mathbb{M}$ is completely distributive. The bottom and the top elements of $\mathbb{M}$ are $0_{\mathbb{M}}$ and $1_{\mathbb{M}}$ resp. As different from $\mathbb{L}$ we do not exclude the case when $\mathbb{M}$ is a one-point lattice and hence in this case $0_{\mathbb{M}} = 1_{\mathbb{M}}$.

3 Basic definitions

Definition 3.1 *An upper $\mathbb{M}$-approximative operator on* L *is a mapping $u : \mathbb{L} \times \mathbb{M} \to \mathbb{L}$ such that*

1. $u(0, \alpha) = 0 \; \forall \alpha \in \mathbb{M}$;
2. $a \leq u(a, \alpha) \; \forall a \in \mathbb{L}, \; \forall \alpha \in \mathbb{M}$;
3. $u(a \vee b, \alpha) = u(a, \alpha) \vee u(b, \alpha)$
4. $u(u(a, \alpha), \alpha) = u(a, \alpha)$;
5. $\alpha \leq \beta, \alpha, \beta \in \mathbb{M} \Longrightarrow u(a, \alpha) \leq u(a, \beta)$.

Operator u is called (upper) semicontinuous (usc) if

- (usc) $u(a, \bigvee_{i \in \mathcal{I}} \alpha_i) = \bigwedge_{i \in \mathcal{I}} u(a, \alpha_i)$;

u is called (upper) weakly semicontinuous (uwsc) if

- (uwsc) *If $u(a, \alpha_i) = \bar{a} \; \forall \alpha_i, i \in \mathcal{I}$ and $\alpha = \bigwedge_{i \in \mathcal{I}} \alpha_i$, then $u(a, \alpha) = \bar{a}$*

Definition 3.2 *A lower $\mathbb{M}$-approximative operator on $\mathbb{L}$ is a mapping $l : \mathbb{L} \times \mathbb{M} \to \mathbb{L}$ such that*

1. $l(1, \alpha) = 1 \; \forall \alpha \in \mathbb{M}$;
2. $a \geq l(a, \alpha) \; \forall a \in \mathbb{L}, \; \forall \alpha \in \mathbb{M}$;
3. $l(a \wedge b, \alpha) = l(a, \alpha) \wedge l(b, \alpha)$
4. $l(l(a, \alpha), \alpha) = l(a, \alpha)$;
5. $\alpha \leq \beta, \alpha, \beta \in \mathbb{M} \Longrightarrow l(a, \alpha) \geq l(a, \beta)$.

Operator l is called (lower) semicontinuous (lsc) if

- (lsc) $l(a, \bigvee_{i \in \mathcal{I}} \alpha_i) = \bigvee_{i \in \mathcal{I}} l(a, \alpha_i)$;

l is called (lower) weakly semicontinuous (lwsc) if

- (lwsc) *If $l(a, \alpha_i) = a^0 \; \forall \alpha_i, i \in \mathcal{I}$ and $\alpha = \bigvee_{i \in \mathcal{I}} \alpha_i$, then $l(a, \alpha) = a^0$*

Definition 3.3 *A triple $(\mathbb{L}, u, l)$, where $u : \mathbb{L} \times \mathbb{M} \to \mathbb{L}$ and $l : \mathbb{L} \times \mathbb{M} \to \mathbb{L}$ are upper and lower $\mathbb{M}$-approximative operators on $\mathbb{L}$, is called an $\mathbb{M}$-approximative system. In case when X is a set* L *is a lattice, $\mathbb{L} = \mathrm{L}^X$ and $(\mathbb{L}, u, l)$ is an approximative system, the quadruple (X, L, u, l) is called an $\mathbb{M}$-approximative space.*

Definition 3.4 *An $\mathbb{M}$-approximative system $(\mathbb{L}, u, l)$ is called semicontinuous (s.c) if u is u.s.c. and l is l.s.c. An $\mathbb{M}$-approximative system $(\mathbb{L}, u, l)$ is called weakly semicontinuous (w.s.c) if u is u.w.s.c. and l is l.w.s.c.*

Definition 3.5 *In case $\mathbb{L}$ is equipped with unary operation $^c : \mathbb{L} \to \mathbb{L}$, an $\mathbb{M}$-approximative system $(\mathbb{L}, u, l)$ is called self dual if*

$$u(a^c, \alpha) = (l(a, \alpha))^c \text{ and}$$
$$l(a^c, \alpha) = (u(a, \alpha))^c \; \forall a \in \mathbb{L}, \; \forall \alpha \in \mathbb{M}$$

Note that in case when $(\mathbb{L}, \leq, \wedge, \vee, ^c)$ is involutive, the system is self-dual iff $(u(a^c, \alpha))^c = l(a, \alpha)$, and $(l(a^c, \alpha))^c = u(a, \alpha), \; \forall a \in \mathbb{L}, \forall \alpha \in \mathbb{M}$;

Remark 3.6 Sometimes we consider $\mathbb{M}$-approximative systems in case of a one-point lattice $\mathbb{M} = \{\cdot\}$. Obviously, in this case the use of the second argument in the notation of approximative systems is redundant and we write just $u(a)$ and $l(a)$ instead of $u(a, \cdot)$ and $l(a, \cdot)$ respectively. Besides, in this case we use the terms *upper and lower approximative operator*, *approximative system*, etc., omitting the prefix $\mathbb{M}$.

4 Lattice of $\mathbb{M}$-approximative systems on a lattice $\mathbb{L}$

Let $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$ stand for the family of $\mathbb{M}$-approximative systems $(\mathbb{L}, u, l)$ where $\mathbb{L}$ and $\mathbb{M}$ are fixed. We introduce an order $\preceq$ on $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$ by setting $(\mathbb{L}, u_1, l_1) \preceq (\mathbb{L}, u_2, l_2)$ iff $u_1 \geq u_2$ and $l_1 \leq l_2$.

Theorem 4.1 *$(\mathcal{AS}^{\mathbb{M}}(\mathbb{L}), \preceq)$ is a complete lattice. Its top and bottom elements are given respectively by*

$$u_\top(a, \alpha) = l_\top(a, \alpha) = a \; \forall a \in \mathbb{L}, \; \forall \alpha \in \mathbb{M};$$

$$u_\bot(a, \alpha) = \begin{cases} 1_{\mathbb{L}} & \text{if } a \neq 0_{\mathbb{L}} \\ 0_{\mathbb{L}}, & \text{if } a = 0_{\mathbb{L}} \end{cases}$$

$$l_\bot(a, \alpha) = \begin{cases} 0_{\mathbb{L}} & \text{if } a \neq 1_{\mathbb{L}} \\ 1_{\mathbb{L}}, & \text{if } a = 1_{\mathbb{L}} \end{cases}$$

The infimum of a family $\{(\mathbb{L}, u_i, l_i) \mid i \in \mathcal{I}\} \subseteq \mathcal{AS}^{\mathbb{M}}(\mathbb{L})$ is given by $(u_0, l_0) = (\bigvee_i u_i, \bigwedge_i l_i)$ and its supremum $\mathbb{L}$ is given by $(u^0, l^0) = (\bigwedge_i u_i, \bigvee_i l_i)$.

Proof Let $\mathcal{S} = \{(\mathbb{L}, u_i, l_i) \mid i \in \mathcal{I}\} \subseteq \mathcal{AS}^{\mathbb{M}}(\mathbb{L})$. Let

$$\bigwedge \mathcal{S} := (\mathbb{L}, u_0, v_0) \text{ where } u_0 = \bigvee_i u_i, \; l_0 = \bigwedge_i l_i.$$

Since $u_0 \geq u_i$ and $l_0 \leq l_i$ for all $i \in \mathcal{I}$, to show the completeness of $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$, it is sufficient to show that u_0 and l_0 are respectively the upper and lower $\mathbb{M}$-approximative operators on $\mathbb{L}$. However this can be established by the direct verification of the conditions (1)-(5) in definitions 3.1, 3.2.

Further, let

$$\bigvee \mathcal{S} = (\mathbb{L}, u^0, l^0) \text{ where } u^0 = \bigwedge_i u_i, l^0 = \bigvee_i l_i.$$

To show that (u^0, l^0) is the supremum of $\mathcal{S}$ it is sufficient to notice that u^0 and l^0 are resp. the upper and lower $\mathbb{M}$-approximative operators. However the validity of properties

ISBN: 978-989-95079-6-8

(1),(2),(4),(5) in definitions 3.2, 3.1 is obvious while the validity of property (3) can be established referring to the distributivity of $\mathbb{L}$.

One can easily establish also the following

Theorem 4.2 *1. The family $(SC\mathcal{AS}^{\mathbb{M}}(\mathbb{L}), \preceq)$ of semicontinuous $\mathbb{M}$-approximative systems is a complete sublattice of $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$.*

2. The family $(WSC\mathcal{AS}^{\mathbb{M}}(\mathbb{L}), \preceq)$ of weakly semicontinuous $\mathbb{M}$-approximative systems is a complete sublattice of $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$.

3. The family $(SD\mathcal{AS}^{\mathbb{M}}(\mathbb{L}), \preceq)$ of self-dual $\mathbb{M}$-approximative systems is a complete sublattice of $\mathcal{AS}^{\mathbb{M}}(\mathbb{L})$.

5 Category $\mathbf{AS}^{\mathbb{M}}$ of $\mathbb{M}$-approximative systems

Let $\mathbb{M}$ be fixed and let $\mathbf{AS}^{\mathbb{M}}$ be the family of all $\mathbb{M}$-approximative systems $(\mathbb{L}, u, v)$. To consider $\mathbf{AS}^{\mathbb{M}}$ as a category whose class of objects are all $\mathbb{M}$-approximative systems we have to specify its morphisms. Given $(\mathbb{L}_1, u_1, l_1), (\mathbb{L}_2, u_2, l_2) \in \mathcal{O}b(\mathbf{AS}^{\mathbb{M}})$ by a morphism

$$f : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_2, u_2, l_2)$$

we call a mapping $f : \mathbb{L}_2 \to \mathbb{L}_1$ such that

1. $f : \mathbb{L}_1 \to \mathbb{L}_2$ is a morphism in the category $\mathbf{LAT^{op}}$ where $\mathbf{LAT}$ is the category of complete infinitely distributive lattices;
2. $u_1(f(b), \alpha) \leq f(u_2(b, \alpha))\ \forall b \in \mathbb{L}_2, \forall \alpha \in \mathbb{M}$;
3. $f(l_2(b, \alpha)) \leq l_1(f(b), \alpha)\ \forall b \in \mathbb{L}_2, \forall \alpha \in \mathbb{M}$

A morphsim $f : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_2, u_2, l_2)$ is aslo referred to as a continuous mapping between the corresponding $\mathbb{M}$-approximative systems

Theorem 5.1 $\mathbf{AS}^{\mathbb{M}}$ *thus obtain is indeed a category.*

Proof Let $f : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_2, u_2, l_2)$ and $g : (\mathbb{L}_2, u_2, l_2) \to (\mathbb{L}_3, u_3, l_3)$ be continuous mappings and let $g \circ f : \mathbb{L}_1 \to \mathbb{L}_3$ be their composition in $\mathbf{LAT^{op}}$. We have to verify that $g \circ f$ satisfies conditions (2) and (3) above. Since it is sufficient to verify these conditions for a fixed $\alpha \in \mathbb{M}$, to simplify the reasonings we omit the second argument in the notation of the approximative operators. Let $c \in \mathbb{L}_3$. Then

$$u_1(f(g(c))) \leq f(u_2(g(c))) \leq f(g(u_3(c))),$$

In a similar way we can show that $f(g(l_3(c))) \leq l_1(g(f(c)))$. Thus the composition $g \circ f : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_3, u_3, l_3)$ is continuous whenever f and g are continuous. We conclude the proof noticing that the identity mapping $f : (\mathbb{L}, u, l) \to (\mathbb{L}, u, l)$ is continuous.
□

In the sequel, when discussing categorical properties of $\mathbf{AS}^{\mathbb{M}}$ and other categories we refer to the monograph [1].

Theorem 5.2 *Every source $f_i : \mathbb{L}_1 \to (\mathbb{L}_i, u_i, l_i), i \in \mathcal{I}$ has a unique initial lift $f_i : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_2, u_2, l_2)$.*

Proof Taking into account Theorem 4.1 it is sufficient to consider the case when the source contains only one morphism $f : \mathbb{L}_1 \to (\mathbb{L}_2, u_2, l_2)$ in $\mathbf{LAT}^{op}$.

Define upper approximative operator $u_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_1$ by

$$u_1(a, \alpha) = \bigwedge \{f(u_2(b, \alpha)) \mid f(b) \geq a\}\ \forall a \in \mathbb{L}_1, \alpha \in \mathbb{M}.$$

Note first that the condition

$$u_1(f(b), \alpha) \leq f(u_2(b, \alpha))\ \forall b \in \mathbb{L}_2\ \forall \alpha \in \mathbb{M}$$

is obviously fulfilled. We verify that u_1 thus defined is indeed an upper approximative operator. As in the previous theorem in our reasoning we fix $\alpha \in \mathbb{M}$ and omit it in notation of approximative operators when verifying the properties (1) - (4).

The first two properties are obvious: $u_1(0_{\mathbb{L}_1}) = 0_{\mathbb{L}_1}$; $u_1(a) \geq a \forall a \in \mathbb{L}_1$.
To verify property (3) let $a_1, a_2 \in \mathbb{L}_1$, then

$$u_1(a_1 \vee a_2) = \bigwedge\{f(u_2(b)) \mid f(b) \geq a_1 \vee a_2\} \leq \bigwedge\{f(u_2(b_1 \vee b_2)) f(b) \geq a_1, f(b_2) \geq a_2\} = \bigwedge\{f(u_2(b_1)) \vee f(u_2(b_2)) \mid f(b_1) \geq a_1, f(b_2) \geq a_2\} = \bigvee_{i=1,2} (\bigwedge\{f(u_2(b_i)) \mid f(b_i) \geq a_i\}) = u_1(a_1) \vee u_1(a_2).$$

The converse inequality is obvious.

To verify the fourth condition notice that $u_1(u_1(a)) = u_1(\bigwedge\{f(u_2(b)) \mid f(b) \geq a\}) \leq \bigwedge\{u_1(f(u_2(b))) \mid f(b) \geq a\} \leq \bigwedge\{f(u_2(u_2(b))) \mid f(b) \geq a\} = u_1(a)$. The converse inequality is obvious

To verify property (5) for u_1 note that

$$\alpha \leq \beta, \alpha, \beta \in \mathbb{M} \Longrightarrow u_1(a, \alpha) \leq u_1(a, \beta)$$

is guaranted by the analogous property of the operator $u_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ and the definition of u_1.

Define lower $\mathbb{M}$-approximative operator $l_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_1$ by

$$l_1(a, \alpha) = \bigvee\{f(l_2(b), \alpha) \mid f(b) \leq a\}\ \forall a \in \mathbb{L}_1 \forall \alpha \in \mathbb{M}.$$

Notice first that

$$f(l_2(b, \alpha)) \leq l_1(f(b), \alpha)\ \forall b \in \mathbb{L}_2, \alpha \in \mathbb{M}.$$

We show that $l_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_1$ thus defined is an lower $\mathbb{M}$-approximative operator. Again, we omit in notation α when it is fixed. The first two conditions from Definition 3.2 are obvious. To verify the third condition let $a_1, a_2 \in \mathbb{L}_1$. Then $l_1(a) \wedge l_1(a_2) = \bigvee\{f(l_1(b_1)) \wedge f(l_1(b_2)) \mid f(l_1(b_i)) \leq a_i, i = 1, 2\} \leq \bigvee\{f(l_1(b_1 \wedge b_2) \mid f(b_1) \wedge f(b_2) \leq a_1 \wedge a_2\} = \bigvee\{f(b) \mid f(b) \leq a_1 \wedge a_2\} = l_1(a_1 \wedge a_2)$,
The converse inequality is obvious, The idempotence of the operator $l_1 : \mathbb{L}_1 \to \mathbb{L}_1$ is establihed as follows:
$l_1(l_1(a)) = l_1(\bigvee\{f(l_2(b)) \mid f(b) \leq a\}) \geq \bigvee\{l_1 f(l_2(b)) \mid f(b) \leq a\} \geq \bigvee\{f(l_2(b)) \mid f(b) \leq a\} = l_1(a)$.
The opposite inequality is obvious.

Finaly, the condition $\alpha \leq \beta, \alpha, \beta \in \mathbb{M} \Longrightarrow l_1(a, \alpha) \geq l_1(a, \beta)$ is guaranted by the analogous property of the operator $l_2 : \mathbb{L}_2 \to \mathbb{L}_2$ and the definition of l_1.

Let $g : (\mathbb{L}_3, u_3, l_3) \to (L_2, u_2, l_2)$ be a morphism in $\mathbf{AS}^{\mathbb{M}}$ and $h : \mathbb{L}_3 \to \mathbb{L}_1$ be a morphism in $\mathbf{LAT}^{op}$ such that $f \circ h = g$.

ISBN: 978-989-95079-6-8

Then from the construction it is clear that $h : (\mathbb{L}_3, u_3, l_3) \to (L_1, u_1, l_1)$ is a morphism in $\mathbf{AS}^{\mathbb{M}}$. Thus $f : (\mathbb{L}_1, u_1, l_1) \to (\mathbb{L}_2, u_2, l_2)$ is indeed the initial lift of $f : \mathbb{L}_1 \to (\mathbb{L}_2, u_2, l_2)$. The uniqueness of the lift is obvious.

Let $A\mathcal{I}\mathbf{LAT}$ denote the category of adjunctive involutive complete infinitely distributive systems.

Theorem 5.3 *Let $\mathbb{L}_1, \mathbb{L}_2$ be adjunctive involutive lattices and $f : \mathbb{L}_1 \to \mathbb{L}_2$ be a morphism in $A\mathcal{I}\mathbf{LAT}^{op}$. If $\mathbb{M}$-approximative operators $l_2 : u_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ are self-dual, then $\mathbb{M}$-approximative operators $l_1, u_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_1$ constructed above are self-dual as well.*

Indeed, let $a \in \mathbb{L}$. Then
$l_1(a^c) = \bigvee\{f(l_2(b)) \mid f(b) \le a^c\} = (\bigwedge\{(f(l_2(b)))^c \mid f(b) \le a^c\})^c = (\bigwedge\{f(u_2(b^c)) \mid f(b) \le a^c\})^c = (\bigwedge\{f(u_2(b^c)) \mid f(b^c) \ge a\})^c = (\bigwedge\{f(u_2(d)) \mid f(d) \ge a\})^c = (u_1(a))^c$.

Theorem 5.4 *If approximative system (u_2, l_2) on $\mathbb{L}_2$ is semi-continuous (weakly semicontinuous), then the approximation system (u_1, l_1) constructed in the previous theorem is semi-continuous (resp. weakly semicontinuous), too.*

Indeed, if (u_2, l_2) is semicontinuous, then $u_1(a, \bigvee_i \alpha_i) = \bigwedge f(u_2(b, \bigvee_i \alpha_i) \mid f(b) \ge a\} = \bigwedge_i \bigwedge\{f(u_2(b, \alpha_i) \mid f(b) \ge a\} = \bigwedge_i u_1(a, \alpha_i)$, and $l_1(a, \bigvee_i \alpha_i) = \bigvee\{f(l_2(b, \bigvee_i \alpha_i) \mid f(b) \le a\} = \bigvee_i \bigvee\{f(l_2(b, \alpha_i) \mid f(b) \le a\} = \bigvee_i l_1(b, \alpha_i)$.
In a similar way one can establish weak semicontinuity of (u_1, l_1) in case (u_2, l_2) was weakly semicontinuous.

Theorem 5.5 *Every sink $f_i : (\mathbb{L}_i, u_i, l_i) \to \mathbb{L}_2,\ i \in \mathcal{I}$ has a unique final lift: $f_i : (\mathbb{L}_i, u_i, l_i) \to (\mathbb{L}_2, u_2, l_2)\ i \in \mathcal{I}$.*

Proof . Taking into account Theorem 4.1 it is sufficient to consider the case of the sink consisting of a single morphism $f : (\mathbb{L}_1, u_1, l_1) \to \mathbb{L}_2$. We define an upper $\mathbb{M}$-approximative operator $u_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ by:

$$u_2(b, \alpha) = \bigwedge\{c \in \mathbb{L}_2 \mid c \ge b, f(c) \ge u_1(f(b), \alpha)\}.$$

It is obvious that $u_1(f(b), \alpha) \le f(u_2(b, \alpha))\ \forall\ b \in \mathbb{L}_2$. We show that $u_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ is an upper $\mathbb{M}$-approximative operator. We omit notation α when it can be fixed.

It is obvious that $u_2(0) = 0$ and $b \le u_2(b)$ for every $b \in \mathbb{L}_2$. Further, let $b_1, b_2 \in \mathbb{L}_2$. Then $u_2(b_1) \vee u_2(b_2) = (\bigwedge\{c_1 \in \mathbb{L}_2 \mid c_1 \ge b_1, f(c_1) \ge u_1(f(b_1))\}) \vee (\bigwedge\{c_2 \in \mathbb{L}_2 \mid c_2 \ge b_2, f(c_2) \ge u_1(f(b_2))\}) = \bigwedge\{c_1 \vee c_2 \mid c_i \ge b_i, f(c_i) \ge u_1(f(b_i)), i = 1, 2\} \ge \bigwedge\{c \mid c \le b_1 \vee b_2, f(c) \ge b_1 \vee b_2, f(c) \ge u_1(f(b_1)) \vee u_1(f(b_2))\} = \bigwedge\{c \mid c \le b_1 \vee b_2, f(c) \ge b_1 \vee b_2, f(c) \ge u_1(f(b_1 \vee b_2))\} = u_2(b_1 \vee b_2)$.
The opposite inequality is obvious.

$u_2(u_2(b)) = \bigwedge\{c \mid c \ge u_2(b), u_1(f(u_2(b))) \le f(c)\}$. Noticing that $u_2(b)$ is among the elements c satisfying the above conditions, we conclude that $u_2(u_2(b)) \le u_2(b)$. The opposite inequality is obvious and hence $u_2(u_2(b)) = u_2(b)$.

Property (5) for u_2 is guaranted by the analogous property of the operator $u_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_1$ and the definition of u_2.

Define lower approximation operator $l_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ by

$$l_2(b, \alpha) = \bigvee\{c \in \mathbb{L}_2 \mid c \le b, l_1(f(c), \alpha) \le f(b)\}.$$

The validity of the first two conditions for $l_2 : \mathbb{L}_2 \times \mathbb{M} \to \mathbb{L}_2$ is obvious. To verify the third property let $b_1, b_2 \in \mathbb{L}_2$. Then $l_2(b_1) \wedge l_2(b_2) = \bigvee\{c_1 \wedge c_2 \mid c_1 \le b_1, c_2 \le b_2, l_1(f(c_1) \le f(b_1), l_1(f(c_2)) \le f(b_2)\} \le \bigvee\{c_1 \wedge c_2 \mid c_1 \wedge c_2 \le b_1 \wedge b_2,\ l_1(f(c_1)) \wedge l_1(f(c_2)) \le f(b_1) \wedge f(b_2)\} \le \bigvee\{c \mid c \le b_1 \wedge b_2, l_1(f(c)) \le f(b_1 \wedge b_2)\} = l_2(b_1 \wedge b_2)$.
The opposite inequality is obvious.

To show the fourth axiom note that

$$l_2(l_2(b)) = \bigvee\{c \mid c \le l_2(b), l_1(f(c)) \le f(l_2(b))\}$$

and since $l_2(b)$ is one of c appearing in the above formula, it holds $l_2(l_2(b)) \ge l_2(b)$. The converse inequality is obvious.

Property (5) for l_2 is guaranted by the analogous property of the operator $l_1 : \mathbb{L}_1 \times \mathbb{M} \to \mathbb{L}_2$ and the definition of l_2.

Theorem 5.6 *If approximation system (u_1, l_1) on $\mathbb{L}_1$ is self-dual, and $f : \mathbb{L}_1 \to \mathbb{L}_2$ is a morphism in the category of adjunkctive involutive lattices, then the approximation system (u_2, l_2) constructed in the previous theorem is self-dual.*

Indeeed, given $b \in \mathbb{L}_2$, we have
$(l_2(b))^c = (\bigvee\{d \mid d \le b, l_1(f(d)) \le f(b)\})^c = \bigwedge\{d^c \mid d \le b,\ l_1(f(d)) \le f(b)\} = \bigwedge\{d^c \mid d^c \ge b^c, (l_1(f(d)))^c \ge f(b^c)\} = \bigwedge\{d^c \mid d^c \ge b^c,\ u_1((f(d))^c) \ge f(b^c)\} = \bigwedge\{d^c \mid d^c \ge b^c,\ u_1((f(d^c))) \ge f(b^c)\} = \bigwedge\{e \mid e \ge b^c,\ u_1(f(e)) \ge f(b^c)\} = u_2(b^c)$.

One can easily establish also the following

Theorem 5.7 *If $\mathbb{M}$-approximation system (u_1, l_1) on $\mathbb{L}_1$ is semicontinuous (weakly semicontinuous), then the approximation system (u_2, l_2) constructed in the previous theorem is semicontinuous (resp. weakly semicontinuous).*

From theorems 5.2, 5.5, we obtain the following important

Corollary 5.8 *Category $\mathbf{AS}^{\mathbb{M}}$ is topological over the category $\mathbf{LAT}^{op}$ lattices with respect to the forgetful functor $\mathfrak{F} : \mathbf{AS}^{\mathbb{M}} \to \mathbf{LAT}^{op}$.*

Besides, taking into account theorems 5.3, 5.6, 5.4, 5.7 we have

Corollary 5.9 *The category $SD\mathbf{AS}^{\mathbb{M}}$ of self-dual $\mathbb{M}$-approximative systems is topological over the category $A\mathcal{I}\mathbf{LAT}^{op}$ with respect to the forgetful functor $\mathfrak{F} : SD\mathbf{AS}^{\mathbb{M}} \to A\mathcal{I}\mathbf{LAT}^{op}$.*

Corollary 5.10 *The categories $SC\mathbf{AS}^{\mathbb{M}}$ and $WSC\mathbf{AS}^{\mathbb{M}}$ of semicontinuous and weakly semicontinuous $\mathbb{M}$-approximative systems are topological over the category $\mathbf{LAT}^{op}$ with respect to the forgetful functor $\mathfrak{F} : (W)SC\mathbf{AS}^{\mathbb{M}} \to \mathbf{LAT}^{op}$.*

6 Categories $\mathbf{AS}^{\mathbb{M}}(\mathrm{L})$ of $\mathbb{M}$-approximative L-spaces

An important subcategory of the category $\mathbf{AS}^{\mathbb{M}}$ is the category whose ojects are L-powersets L^X of arbitrary sets (where L is a fixed lattice) and whose morphisms are induced by mappings of the corresponding sets X. Here are the details:

Let L be a fixed complete infinitely distributive lattice and let the objects of $\mathbf{AS}^{\mathbb{M}}(\mathrm{L})$ be approximation systems of the

ISBN: 978-989-95079-6-8

form (L^X, u, l) where L^X are L-powersets of arbitrary sets X. Sometimes it is more convenient to interpret objects of this category as the corresponding quadruples (X, L, u, l). To define a morphism $F : (\mathrm{L}^{X_1}, u_1, l_1) \to (\mathrm{L}^{X_2}, u_2, l_2)$ consider a mapping $f : X_1 \to X_2$ and let $f^{\leftarrow} : \mathrm{L}^{X_2} \to \mathrm{L}^{X_1}$ be the backward powerset operator induced by f [16]. Now as morphisms in $\mathbf{AS}^{\mathbb{M}}(\mathrm{L})$ we take $F := f^{\leftarrow^{op}} : (\mathrm{L}^{X_1}, u_1, l_1) \to (\mathrm{L}^{X_2}, u_2, l_2)$ in case it is a morphism in the category $\mathbf{AS}^{\mathbb{M}}$.

An important special case is a two point lattice $\mathrm{L} = \mathbf{2}$: in this case we come to the category of $\mathbb{M}$-approximative structures on ordinary sets (of course, for this one has to interpret a subset A of a set X as the characteristic function $\chi_A : X \to \mathbf{2}$). In particular, if $\mathbb{M}$ is a one-point lattice we come to the concept of an approximation system as it was considered by some authors, see e.g. [20].

7 Categories of fuzzy topologies as subcategories of $\mathbf{AS}^{\mathbb{M}}$

7.1 Category of $(\mathrm{L}, \mathbb{M})$*-fuzzy topological spaces*

We start with interpreting the category **FTOP**$(\mathrm{L}, \mathbb{M})$ of $(\mathrm{L}, \mathbb{M})$-fuzzy topological spaces see, e.g. [17], [13], [18], [7] as a subcategory of $\mathbf{AS}^{\mathbb{M}}$. In this section $\mathbb{M}$ is assumed to be completely distributive.

Definition 7.1 *A mapping $\mathcal{T} : \mathrm{L}^X \to \mathbb{M}$ is an $(\mathrm{L}, \mathbb{M})$-fuzzy topology on X if*

1. $\mathcal{T}(0_X) = \mathcal{T}(1_X) = 1$;
2. $\mathcal{T}(U \wedge V) \geq \mathcal{T}(U) \wedge \mathcal{T}(V) \forall U, V \in \mathrm{L}^X$;
3. $\mathcal{T}\left(\bigvee_{i\in\mathcal{I}} U_i\right) \geq \bigwedge_{i\in\mathcal{I}} \mathcal{T}(U_i) \ \ \forall\{U_i \mid i \in \mathcal{I}\} \subseteq \mathrm{L}^X$

A pair $(X, \mathcal{T})$ is called an $(\mathrm{L}, \mathbb{M})$-fuzzy topological space and the value $\mathcal{T}(U), U \in \mathrm{L}^X$ is interpreted as the degree of openess of a fuzzy set U. A mapping $f : (X, \mathcal{T}_X) \to (Y, \mathcal{T}_Y)$ is called continuous if $\mathcal{T}_X\left(f^{-1}(V)\right) \geq \mathcal{T}_Y(V)) \ \ \forall V \in \mathrm{L}^Y$.

Let $(X, \mathcal{T})$ be an $(\mathrm{L}, \mathbb{M})$-fuzzy topological space. By setting

$$\mathrm{int}_{\mathcal{T}}(A, \alpha) = \bigvee\{U \in \mathrm{L}^X \mid U \leq A, \ \mathcal{T}(U) \geq \alpha\},$$

we define the s.c. interior operator $\mathrm{int}_{\mathcal{T}} : \mathrm{L}^X \times \mathbb{M} \to \mathrm{L}^X$.

The relations between (L, M)-fuzzy topologies and lower $\mathbb{M}$-approximative operators are described in the theorem;

Theorem 7.2 *The interior operator* int *is a weakly s.c. lower $\mathbb{M}$-approximative operator on $\mathbb{L} = \mathrm{L}^X$. Conversely, if $l : \mathrm{L}^X \times \mathbb{M} \to \mathrm{L}^X$ is a weakly s.c. lower $\mathbb{M}$-approximative operator, then by setting*

$$T_l(U) = \bigvee\{\alpha \mid l(U, \alpha) \geq U\}$$

we obtain a mapping $\mathcal{T}_l : \mathrm{L}^X \to \mathrm{L}^X$ satisfying conditions (1) and (3) of Definition 7.1. Besides $\mathcal{T}_{\mathrm{int}_{\mathcal{T}}} = \mathcal{T}$ and $l_{\mathcal{T}_l} = l$.

Further, assume that L is an adjunctive involutive lattice and let $^c : \mathrm{L} \to \mathrm{L}$ be the corresponding involution. Then by setting

$$cl_{\mathcal{T}}(A, \alpha) = \bigwedge\{B \mid B \geq A, \mathcal{T}(B^c) \geq \alpha\}$$

a closure operator $cl_{\mathcal{T}} : \mathrm{L}^X \times \mathbb{M} \to \mathrm{L}^X$ is defined. One can easily show that $cl_{\mathcal{T}}$ is a weakly s.c. upper $\mathbb{M}$-approximative operator and prove a theorem establishing relations between weakly s.c. upper $\mathbb{M}$-approximative operators and $(\mathrm{L}, \mathbb{M})$-fuzzy topologies via closure operators, analogous to Theorem 7.2. Besides the $\mathbb{M}$-approximation system $(\mathrm{L}^X, cl_{\mathcal{T}}, int_{\mathcal{T}})$ is self-dual.

Thus in case of an involutive adjunctive lattice L an $(\mathrm{L}, \mathbb{M})$-fuzzy topological space $(X, \mathcal{T})$ can be interpreted as a weakly s.c. $\mathbb{M}$-approximative self-dual system $(\mathbb{L}, \mathrm{cl}, \mathrm{int})$ where $\mathbb{L} = \mathrm{L}^X$.

This allows us to identify the category $\mathbf{FTOP}(\mathrm{L}, \mathbb{M})$ with the subcategory $\mathbf{TopAS}^{\mathbb{M}}(\mathrm{L})$ of the category $\mathbf{AS}^{\mathbb{M}}$ whose objects are self dual weakly s.c. $\mathbb{M}$-approximative systems of the form $(\mathrm{L}^X, \mathrm{int}, \mathrm{cl})$ and the morphisms are $F = f^{\leftarrow^{op}} : (\mathrm{L}^X, \mathrm{int}_X, \mathrm{cl}_X) \to (\mathrm{L}^Y, \mathrm{int}_Y, \mathrm{cl}_Y)$, where $f^{\leftarrow} : \mathrm{L}^Y \to \mathrm{L}^X$ are powerset operators induced by (see e.g. [9]) by coninuous mappings $f : (X, \mathcal{T}_X) \to (Y, \mathcal{T}_Y)$ (cf also the previous section).

7.2 Category of Chang-Goguen L*-topological space*

To obtain characterization of L-topological spaces (see [3], [5]) by means of approximative systems we can restrict the theory developed in 7.1 by taking the two-point lattice $\mathbf{2} = \{0, 1\}$ in the role of $\mathbb{M}$. Then the category of Chang-Goguen L-topological spaces can be identified with the subcategory $\mathbf{TopAS}(\mathrm{L})$ of the category $\mathbf{TopAS^2}(\mathrm{L})$. In particular $\mathbf{TopAS^2}(\mathbf{2})$ can be identified with the classical category **TOP** of ordinary topological spaces and continuous mappings.

7.3 Category of L*-fuzzifying topological space*

To obtain characterization of L-fuzzifying topological spaces, originally defined by U.Hohle [6] and then independently discovered by Mingsheng Ying [21], by means of approximative systems we restrict the theory developed in 7.1 by taking the two-point lattice $\mathbf{2}$ in the role of L (thus $\mathbb{L} = 2^X$) and the lattice L in the role of $\mathbb{M}$. Then the category of L-fuzzifying topological spaces can be identified with the category $\mathbf{TopAS}^{\mathrm{L}}(\mathbf{2})$.

7.4 Category of Hutton fuzzy topological spaces

According to B. Hutton [8], a fuzzy topological space is a pair (L, τ) where L is a completely distributive lattice and $\tau \subseteq \mathrm{L}$ such that $0, 1 \in \tau$; $\ a, b \in \tau \Longrightarrow a \wedge b \in \tau$: $\ a_i \in \tau \forall i \in \mathcal{I} \Longrightarrow \bigvee_{i\in\mathcal{I}} a_i \in \tau$. The morphisms $f : (L_1, \tau_1) \to (L_2, \tau_2)$ in the category **H-TOP** of Hutton fuzzy topological spaces are mappings $f : L_2 \to L_1$ such that $f(\tau_2) \subseteq \tau_1$. One can show that the category **H-TOP** can be identified with the subcategory **HAS** of the category $\mathbf{AS}^{\mathbb{M}}$. whose objects are self dual approximative systems $(\mathbb{L}, l, u)$ where $\mathbb{L}$ is a completely distributive involutive adjunctive lattice and $\mathbb{M} = \{\cdot\}$

7.5 Category of variable basis fuzzy topological spaces

In [15] S.E. Rodabaugh has introduced the notion of a variable-basis fuzzy topological space and defined the corresponding category **R-TOP**. Further the theory of variable basis fuzzy topological spaces and some related categories was developed in a series of papers by S.E. Rodabaugh, P. Eklund and other authors. The category of variable-basis fuzzy topological spaces also can be obtained as a subcategory of the category $\mathbf{AS}^{\mathbb{M}}$. However to describe it in this way and to give an

ISBN: 978-989-95079-6-8

explicite characterization by means of $\mathbb{M}$-approximative systems we need more space than it is allowed here.

8 Categories related to rough sets

8.1 Rough sets

Let $\rho \subseteq X \times X$ be a binary relation on a set X and let $R(x) = \{x' \mid x\rho x'\}$ be the right ρ-class of $x \in X$. Given $A \in 2^X$ let $l(A) = A^{\blacktriangledown} = \{x \mid R(x) \subseteq A\}$, $u(A) = A^{\blacktriangle} = \{x \mid R(x) \cap A \neq \emptyset\}$. In case ρ is reflexive and transitive $u : 2^X \to 2^X$ and $l : 2^X \to 2^X$ are, respectively, upper and lower approximative operators on $2^X = \mathcal{P}(X)$ and $(X, 2, ^{\blacktriangledown}, ^{\blacktriangle})$ is an approximative space. Besides, one can easily see that the system $(\mathbf{2}^X, ^{\blacktriangledown}, ^{\blacktriangle})$ is self dual: $A^{c\blacktriangle} = A^{\blacktriangledown c}$) for any $A \subseteq X$. Such operators and corresponding approximative spaces in case when ρ is an equivalence relation were introduced by Pawlak [14] under the name "rough set". Further approximative operators operators induced by binary relations, either general or satisfying special properties, were studied by different authors, see e.g. [11], [12], [20], etc. Note however, that in case ρ is not reflexive or transitive, this operators may fail to be approximative operators in our sense.

In case when ρ is only reflexive, J. Järvinen and J. Kortelainen [12] along with operators $A^{\blacktriangledown}$ and $A^{\blacktriangle}$ consider also operators $u'(A) = A^{\triangle} = \{x \mid R^{-1}(x) \cap A \neq \emptyset\}$, $l'(A) = A^{\triangledown} = \{x \mid R^{-1}(x) \subseteq A\}$ and show that (u, l') and (u', l) form Galois connection: $u(a) \leq b \iff a \leq l'(b)$; $\quad l(a) \leq b \iff a \leq u'(b)$. Thus in case ρ is also transitive, we obtain "Galoi-connected" approximative systems $(\mathbf{2}^X, ^{\blacktriangledown}, ^{\triangle})$ and $(\mathbf{2}^X, ^{\triangledown}, ^{\blacktriangle})$.

8.2 L*-rough sets*

Generalizing the previous situation let L be a cl-monoid $(\mathrm{L}, \wedge, \vee, *)$, X be a set and $\rho : X \times X \to \mathrm{L}$ be an L-relation on X. Further, assume that ρ is reflexive (that is $\rho(x, x) = 1 \ \forall x \in X$) and transitive (that is $\rho(x, y) * \rho(y, z) \leq \rho(x, z) \ \forall x, y, z \in X$.) For every $x \in X$ we define $\mathcal{R}(x) : X \to \mathrm{L}$ by $\mathcal{R}(x)(x') = \rho(x, x')$ Further, given $A \in \mathrm{L}^X$ let lower and upper approximative operators $l(A) \in \mathrm{L}^X$, and $u(A) \in \mathrm{L}^X$ be defined by $l(A)(x) = \inf_{x' \in X}(\mathcal{R}(x)(x') \mapsto A(x'))$ and $u(A)(x) = \sup_{x' \in X}(\mathcal{R}(x)(x') * A(x'))$ respectively. One can show that (L^X, u, l) is an L-approximative system. We refer to such kind of an approximative system as an L-rough system induced by the L-relation ρ. In case $(\mathrm{L}, \wedge, \vee, *)$ is a Girard monoid, the system (L^X, u, l) is self dual. Further, if $\mathrm{L} = \mathbf{2}$ is a two-point lattice we come to the situation described in the previous subsection. In an natural way we define morphisms for the category $\mathbf{Rgh}(\mathbb{L})$ of L- rough systems and characterize it as a category of approximative systems.

9 Defuzzification approximation operators

Finally we sketch how the concept of an approximative systems can be applied for fuzzy sets themselves.

Let $\mathrm{L} = (\mathrm{L}, \wedge, \vee, \leq)$ be a complete lattice, X be a set and $\mathbb{L} = \mathrm{L}^X$ Define $u : \mathbb{L} \times \mathrm{L} \to \mathbb{L}$ and $l : \mathbb{L} \times \mathrm{L} \to \mathbb{L}$ as follows: Given $A \in \mathrm{L}^X$ let

$$u(A, \alpha) = A \vee 1_{A_\alpha}, \ l(A, \alpha) = \alpha \cdot 1_{A_\alpha} \wedge A,$$

where $A_\alpha = \{x \in X \mid A(x) \geq \alpha\}$ In this way we obtain L-approximation operators on the L-powerset of a set X which can be interpreted as resp. upper and lower level-defuzzification operators .

References

[1] J. Adámek, H. Herrlich and G.E. Strecker, **Abstract and Concrete Categories**, John Wiley & Sons, New York 1990.

[2] G. Birkhoff, **Lattice Theory**, AMS Providence, RI, 1995.

[3] C.L. Chang, *Fuzzy topological spaces*, J. Math. Anal. Appl., **24** (1968), 182-190.

[4] G. Gierz, K.H. Hofmann, K. Keimel, J.D. Lawson, M.W. Mislove, D.S. Scott, **Continuous Lattices and Domains**, Cambridge University Press, Cambridge, 2003.

[5] J.A. Goguen, *The fuzzy Tychonoff theorem*, J. Math. Anal. Appl., **18** (1973), 734-742.

[6] U. Höhle, *Upper semicontinuous fuzzy sets and applications*, J. Math. Anal. Appl., **78** (1980) 659-673.

[7] U. Höhle, A. Šostak, *Axiomatics for fixed-based fuzzy topologies*, Chapter 3 in [8]

[8] B. Hutton, *Normality in fuzzy topologicl spaces*, J. Math. Anal. Appl., **50** (1975), 74-79.

[9] **Mathematics of Fuzzy Sets: Logic, Topology and Measure Theory** , U. Höhle, S.E. Rodabaugh eds. - Handbook Series, vol.3, Kluwer Acad. Publ., - 1999.

[10] S. Jenei, *Structure of Girad Monoids on [0,1]* Chapter 10 In: **Topological and Algebraic Structures in Fuzzy Sets** S.E. Rodabaugh, E.P. Klement eds., Kluwer Acad. Publ. 2003.

[11] J. Korteleinen, On relationship between modified sets, topological spaces and rough sets, Fuzzy Sets and Syst. **61** (1994), 91-95.

[12] J. Järvinen, J. Kortelainen, **A unified study between modal-like operators, Topologies and Fuzzy Sets**, Turku Centre for Computer Sciences, Turku, Finland, Technical Reports **642**, 2004.

[13] T. Kubiak, *On Fuzzy Topologies*, PhD Thesis, Adama Mickiewicz University Poznan, Poland, 1985.

[14] Z. Pawlak, *Rough sets*, International J. of Computer and Information Sciences, **11** (1982), 341-356.

[15] S.E. Rodabaugh, A categorical accomodation of various notions of fuzzy topology, *Fuzzy Sets and Syst.*, **9** (1983), 241-265.

[16] S.E. Rodabaugh, *Powerset Operator Foundations For Poslat Fuzzy Set Theories and Topologies*, Chapter 2 in [8].

[17] A. Šostak, *On a fuzzy topological structure*, Suppl. Rend. Circ. Matem. Palermo, Ser II **11** (1985), 125-186.

[18] A. Šostak, *Basic structures of fuzzy topology*, J. Math. Sci. **78** (1996), 662-701.

[19] A. Šostak, *Approximative fuzzy systems and approximative fuzzy spaces*, In: 9^{th} International Conference on Fuzzy Set Theory and Applications: FSTA2008. Abstracts, p. 118.

[20] Y.Y. Yao, *A comparative study of fuzzy sets and rough sets*, Inf. Sci., **109** (1998), 227 - 242.

[21] Ying Mingsheng, *A new approach to fuzzy topology, Part* I, Fuzzy Sets and Syst., **39** (1991), 303-321.

ISBN: 978-989-95079-6-8

A Fuzzy Weight Representation for Inner Dependence Method AHP

Shin-ichi Ohnishi[12] Takahiro Yamanoi[1] Hideyuki Imai[2]

1 Faculty of Engineering, Hokkai-Gakuen University
Sapporo, 0640926 JAPAN
2 Graduate School of Information Science and Technology, Hokkaido University
Sapporo, 0600814 JAPAN
Email: {ohnishi,yamanoi}@eli.hokkai-s-u.ac.jp, imai@ist.hokudai.ac.jp

Abstract— *AHP (Analytic Hierarchy Process) has been widely used in decision making. Inner dependence method AHP is one technique even in case of criteria have dependency. However using original AHP or Inner dependence method, the results often lose reliability because the comparison matrix does not always have sufficient consistency. In these cases, fuzzy representation for weighting criteria and alternatives using results from a sensitivity analysis is useful. In this paper, we present alternative overall weights by employing some assumptions. Since an idea of less ambiguity is employed, the results show how inner dependence AHP has fuzziness when the comparison matrix is not sufficiently consistent..*

Keywords—Decision making, AHP (Analytic Hierarchy Process), Fuzzy sets, Inner Dependence, Sensitivity analysis.

1 Introduction

AHP (Analytic Hierarchy Process) was proposed by Saaty T.L. in 1977 [1], [2]. The method has been popular and widely used in the domain of decision making, since it can include vagaries such as humans feelings. In addition, it can be developed to ANP (Analytic Network Process) models.

Usually normal AHP must assume independency among criteria, although it is difficult to choose enough independent criteria. Inner dependence method AHP[10] is one technique of solving this kind of problem even in case of criteria have dependency.

However, the comparison matrix often does not have enough consistency when AHP or Inner dependence method is used since, for instance, a problem may contain too many criteria for decision making. In these cases, we consider that answers from decision-makers (i.e. components of the comparison matrix) have ambiguity or fuzziness. For resolving this type of problem, fuzzy reciprocal components have been proposed as components of the data matrix in some research [12]. In this paper, we consider that weights should also have ambiguity or fuzziness. Therefore, it is necessary to represent these weights by use of fuzzy sets.

Sensitivity analysis is applied to Inner dependence AHP to analyze the amount the components of a pairwise comparison matrix influences the weights and consistency of a matrix. This makes it possible to show the magnitude of the fuzziness in the weights.

In previous researches, we proposed a new representation for weights of criteria and alternatives in normal AHP. [7][8][11]. In this paper, a representation of weights of inner dependence method is proposed. It is represented as L-R fuzzy numbers by using the results from the sensitivity analysis. This paper encompasses methodology to represent weights by fuzzy sets. In addition, a representation of fuzziness as a result of inner dependence is presented when a comparison matrix does not have enough consistency.

2 Inner depedence AHP

2.1 Process of Normal AHP

(Process 1) Representation of structure by a hierarchy. The problem under consideration can be represented in a hierarchical structure. The highest level of the hierarchy consists of a unique element that is the overall objective. At the lower levels, there are multiple activities (i.e. elements within a single level) with relationships among elements of the adjacent higher level to be considered. The activities are evaluated using subjective judgments of a decision maker. Elements that lie at the upper level are called parent elements while those that lie at lower level are called child elements. Alternative elements are put at the lowest level of the hierarchy

(Process 2) Paired comparison between elements at each level. A pairwise comparison matrix A is created from a decision maker's answers. Let n be the number of elements at a certain level. The upper

ISBN: 978-989-95079-6-8

triangular components of the comparison matrix a_{ij} ($i< j$ = 1,…,n) are 9, 8, .. , 2, 1, 1/2, …, or 1/9. These denote intensities of importance from activity i to j. The lower triangular components a_{ji} are described with reciprocal numbers as follows

$$a_{ji} = 1/a_{ij}\,, \tag{1}$$

in addition, for diagonal elements, let a_{ii} = 1. The lower triangular components and diagonal elements are occasionally omitted from the written equation as they are evident if upper triangular components are shown. The decision maker should make $n(n\text{-}1)/2$ paired comparisons at a level with n elements.

(Process 3) Calculations of weight at each level. The weights of the elements, which represent grade of importance among each element, are calculated from the pairwise comparison matrix. The eigenvector that corresponds to a positive eigenvalue of the matrix is used in calculations throughout in this paper.

(Process 4) Priority of an alternative by a composition of weights. The composite weight can be calculated from the weights of one level lower. With repetition, the weights of the alternative, which are the priorities of the alternatives with respect to the overall objective, are finally found.

2.2 Consistency

Since components of the comparison matrix are obtained by comparisons between two elements, coherent consistency is not guaranteed. In AHP, the consistency of the comparison matrix A is measured by the following consistency index (C.I.)

$$\text{C.I.} = \frac{\lambda_A - n}{n-1}, \tag{2}$$

where n is the order of matrix A, and λ_A is its maximum eigenvalue.

It should be noted that $\text{C.I.} \geq 0$ holds. And if the value of C.I. becomes smaller, then the degree of consistency becomes higher, and vice versa. The comparison matrix is consistent if the following inequality holds.

$$\text{C.I.} \leq 0.1$$

Also consistency ratio (C.R.) is defined as

$$\text{C.R.} = \frac{\text{C.I.}}{M},$$

Where M is random consistency value. However we only employ C.I., since we mainly use 4 or 5-dimensional data whose random consistency value is not far from 1.

2.3 Inner Dependence Method

Usually normal AHP must assume independency among criteria, although it is difficult to choose enough independent criteria. Inner dependence method AHP[10] is one technique of solving this kind of problem even in case of criteria have dependency.

In the method, using a dependency matrix $F=\{f_{ij}\}$, we can calculate real weights $\boldsymbol{w_n}$ as follows,

$$\boldsymbol{w_n} = F\boldsymbol{w} \tag{3}$$

where $\boldsymbol{w}$ is weights from independent criterion, i.e. normal weights of normal AHP and F is calculated as eigen value of influenced matrix.

3 Sensitivity Analysis

When AHP is used, the comparison matrix is often inconsistent or large differences among the overall weights of the alternatives do not appear. Thus, it is very important to investigate how the components of a pairwise comparison matrix influence the consistency or weights. Sensitivity analysis is used to analyze how results are influenced when certain variables change. Therefore, it is necessary to establish a sensitivity analysis of AHP.

In our research, a previously proposed method [7] is used to evaluate the fluctuation of the consistency index and weights when a comparison matrix is perturbed. This method is useful as it does not change the structure of the data.

Evaluating the consistency index and the weights of a perturbed comparison matrix are performed as follows.

(1) Perturbations $\varepsilon a_{ij}d_{ij}$ are imparted to component a_{ij} of a comparison matrix, and the fluctuation of the consistency index and the weight are expressed by the power series of ε.

(2) Fluctuations of the consistency index and the weights are represented by the linear combination of d_{ij}.

(3) By the coefficient of d_{ij}, it can be shown that how the component of the comparison matrix gives influence on the consistency index and the weight.

Since the pairwise comparison matrix A is a positive square matrix, the following Perron-

ISBN: 978-989-95079-6-8

Frobenius theorem [4] holds.

Theorem 1 (Perron – Frobenius) *For a positive square matrix A, the following holds true.*

1. *Matrix A has a positive eigenvalue. If λ_A is the largest eigenvalue then λ_A is a simple root. The positive eigenvector $\boldsymbol{w}$, corresponding to λ_A, exists. λ_A is called the Frobenius root of A.*
2. *Any positive eigenvectors of A are the constant multiples of $\boldsymbol{w}$.*
3. *The absolute value of the eigenvalues of A, except for λ_A, is smaller than λ_A.*
4. *The Frobenius root of the transposed matrix A' is equivalent to the Frobenius root of A.*

This theorem ensures the existence of a weight vector in a pairwise comparison matrix.

From Theorem 1, the following theorem regarding a perturbed comparison matrix holds true [7].

Theorem2 *Let $A = (a_{ij})$, $i,j = 1,\ldots,n$ be a comparison matrix and let $A(\varepsilon) = A+\varepsilon D_A$, $D_A=(a_{ij}d_{ij})$ be a matrix that has been perturbed. Moreover, let λ_A be the Frobenius root of A with $\boldsymbol{w}_1$ being the corresponding eigenvector. Let $\boldsymbol{w}_2$ be the eigenvector corresponding to the Frobenius root of A', then, the Frobenius root $\lambda(\varepsilon)$ of $A(\varepsilon)$ and the corresponding eigenvector $\boldsymbol{w}_1(\varepsilon)$ can be expressed as follows*

$$\lambda(\varepsilon) = \lambda_A + \varepsilon\lambda^{(1)} + o(\varepsilon), \tag{4}$$

$$\boldsymbol{w}_1(\varepsilon) = \boldsymbol{w}_1 + \varepsilon\boldsymbol{w}^{(1)} + \boldsymbol{o}(\varepsilon), \tag{5}$$

where

$$\lambda^{(1)} = \frac{\boldsymbol{w}_2' D_A \boldsymbol{w}_1}{\boldsymbol{w}_2' \boldsymbol{w}_1}, \tag{6}$$

$\boldsymbol{w}^{(1)}$ is an n-dimension vector that satisfies

$$(A - \lambda_A I)\boldsymbol{w}^{(1)} = -(D_A - \lambda^{(1)} I)\boldsymbol{w}_1, \tag{7}$$

where $\boldsymbol{o}(\varepsilon)$ denotes an n-dimension vector in which all components are $o(\varepsilon)$.

Proof of this theorem can be found in Ohnishi's paper [7].

3.1 Sensitivity analysis of consistency index

Regarding a fluctuation of the consistency index, the following corollary can be obtained from Theorem 2.

Corollary 1 *Using an appropriate p_{ij}, we can represent the consistency index* C.I.(ε) *of the perturbed comparison matrix as follows*

$$\text{C.I.}(\varepsilon) = \text{C.I.} + \varepsilon\sum_i^n\sum_j^n p_{ij}d_{ij} + o(\varepsilon). \tag{8}$$

(Proof)

From the definition of the consistency index (3) and (4),

$$\text{C.I.}(\varepsilon) = \text{C.I.} + \varepsilon\frac{\lambda^{(1)}}{n-1} + o(\varepsilon).$$

Let $\boldsymbol{w}_1=(w_{1i})$ and $\boldsymbol{w}_2=(w_{2i})$ from (6). $\lambda^{(1)}$ is can now be represented as

$$\lambda^{(1)} = \frac{1}{\boldsymbol{w}_2' \boldsymbol{w}_1}\sum_i^n\sum_j^n w_{2i}\, a_{ij}\, w_{1j}\, d_{ij},$$

therefore, the second part of the right side is expressed by a linear combination of d_{ij}. (Q.E.D)

p_{ij} in equation (8) in Corollary 1 shows the influence of comparison matrix components on the consistency.

On the other hand, since the comparison matrix $A(\varepsilon) = (a_{ij}(\varepsilon))$ is reciprocal, then $a_{ji}(\varepsilon) = 1/a_{ij}(\varepsilon)$ and becomes

$$a_{ji} + \varepsilon a_{ji}d_{ji} = \frac{1}{a_{ij}} - \varepsilon\frac{d_{ij}}{a_{ij}} + o(\varepsilon). \tag{9}$$

Here, since $a_{ji} = 1/a_{ij}$,

$$d_{ji} = -d_{ij} \tag{10}$$

is obtained. The impact on the consistency can be easily shown by use of this property.

3.2 Sensitivity analysis of weights

With regards to the fluctuation in weighs, the following corollary can also be obtained from Theorem 2.

Corollary 2 *Using an appropriate $q_{ij}^{(k)}$, we can represent the fluctuation $\boldsymbol{w}^{(1)}=(w_k^{(1)})$ of the weight (i.e. the eigenvector corresponding to the Frobenius root) as follows*

$$w_k^{(1)} = \sum_i^n\sum_j^n h_{ij}^{(k)} d_{ij}. \tag{11}$$

(Proof)

The *k*-th row component of the right side of (7) in

ISBN: 978-989-95079-6-8

Theorem 2 is represented as

$$\sum_i^n \sum_j^n \{\frac{w_{1k}\, w_{2i}\, a_{ij}\, w_{1j}}{\boldsymbol{w}_2'\ \boldsymbol{w}_1} - \delta(i,k) a_{ij}\, w_{1j}\}\, d_{ij},$$

and is expressed by a linear combination of d_{ij}. Here,$\delta(i,k)$ is Kronecker's symbol

$$\delta(i,k) = \begin{cases} 1 & (i = k), \\ 0 & (i \neq k). \end{cases}$$

In contrast, since λ_A is a simple root, Rank(A-$\lambda_A I$) = n-1. Accordingly, the weight vector is normalized as

$$\sum_k^n (w_{1k} + \varepsilon w_k^{(1)}) = \sum_k^n w_{1k} = 1,$$

then the condition is as follows.

$$\sum_k^n w_k^{(1)} = 0. \qquad (12)$$

By using an elementary transformation to formula (7) in the condition above, we also can represent $w_k^{(1)}$ by linear combinations of d_{ij}. (Q.E.D)

As seen in equation (5) in Theorem 2, the component that has a great influence on weight $\boldsymbol{w}_1(\varepsilon)$ is the component which has the greatest influence on $\boldsymbol{w}^{(1)}$. $q_{ij}^{(k)}$ in equation (11) from Corollary 2 shows how the influence by the components of a comparison matrix on the weights can be calculated.

The influence can also be shown easily by use of equation (10).

3.3 Sensitivity for inner dependence method

We can also calculate ith regards to the fluctuation in weighs, the followi

Corollary 3 *Using an appropriate $h_{ij}^{(k)}$,we can represent the fluctuation $\boldsymbol{w}_n^{(1)}$=($w_n^{(1)}$) of the weights of inner dependence AHP as follows*

$$w_n^{(1)} = \sum_i^n \sum_j^n h_{ij}^{(k)} d_{ij}.$$

(Proof)

From the equation (3), weights of inner dependence method $\boldsymbol{w}_n$ is calculated from linear transformation of normal weights $\boldsymbol{w}$ of normal AHP, and from Corollary 2 $\boldsymbol{w}$ is represented as sum of linear combination of d_{ij} .Therefore weights $\boldsymbol{w}_n$ is also represented as sum of linear combination of $d_{ij.}$. (Q.E.D)

4 A Weights Representation

The comparison matrix often has poor consistency (i.e. 0.1<C.I.<0.2) because it encompasses several activities. In these cases, the components of a comparison matrix are considered to have fuzziness since they result from the fuzzy judgment of humans. Therefore, weights should be treated as fuzzy numbers.

To represent fuzziness of weight w_{1k}, an L-R fuzzy number is used.

4.1 L-R fuzzy number

L-R fuzzy number

$$M = (m, \alpha, \beta)_{LR}$$

is defined as fuzzy sets whose membership function is as follows.

$$\mu_M(x) = \begin{cases} R\left(\dfrac{x-m}{\beta}\right) & (x > m), \\ L\left(\dfrac{m-x}{\alpha}\right) & (x \leq m). \end{cases}$$

where $L(x)$ and $R(x)$ are shape function which satisfies

(1) $L(x) = L(-x)$,

(2) $L(0) = 1$,

(3) $L(x)$ is a non increasing function

4.2 Fuzzy weights of criteria

From the fluctuation of the consistency index, the multiple coefficient $g_{ij}h_{ij}^{(k)}$ in Corollary 1 and 3 is considered as the influence on a_{ij}.

Since g_{ij} is always positive, if the coefficient $h_{ij}^{(k)}$ is positive, the real weight of criterion k is considered to be larger than w_{1k}. Conversely, if $h_{ij}^{(k)}$ is negative, the real weight of activity k is considered to be smaller. Therefore, the sign of $h_{ij}^{(k)}$ represents the direction of the fuzzy number spread. The absolute value $g_{ij}|h_{ij}^{(k)}|$ represents the size of the influence.

On the other hand, if C.I. becomes bigger, then the judgment becomes more fuzzy.

Consequently, multiple C.I. $g_{ij}|h_{ij}^{(k)}|$ can be regarded as a spread of a fuzzy weight $\tilde{w}_k$ concerned with

ISBN: 978-989-95079-6-8

a_{ij}.

Definition 1 (fuzzy weight) *Let* w_{nk} *be a crisp weight of criterion k of inner dependence model, and* g_{ij} $|h_{ij}^{(k)}|$ *denote the coefficients found in* Corollary 1 and 3. *If* 0.1<C.I.<0.2*, then a fuzzy weight* $\tilde{w}_k$ *is defined by*

$$\tilde{w}_k = (w_{nk}, \alpha_k, \beta_k)_{LR} \tag{13}$$

where

$$\alpha_k = \text{C.I.} \sum_i^n \sum_j^n s(-, h_{kij}) g_{ij} \mid h_{kij} \mid, \tag{14}$$

$$\beta_k = \text{C.I.} \sum_i^n \sum_j^n s(+, h_{kij}) g_{ij} \mid h_{kij} \mid, \tag{15}$$

$$s(+,h) = \begin{cases} 1, (h \geq 0) \\ 0. (h < 0) \end{cases}, \; s(-,h) = \begin{cases} 1, (h < 0) \\ 0. (h \geq 0) \end{cases}$$

4.2 Fuzzy weights of alternatives

Using the fuzzy weights of criteria defined above and local crisp weights of alternatives with respect to certain criterion, we can calculate overall weights from the viewpoint of the overall objective by extension. However, the results from the operation of fuzzy numbers are frequently too ambiguous to interpret.

Fuzzy weights of activities are normalized thus their sum is 1, therefore we can avoid much ambiguity since this condition has been considered [9].

In general, operating with constraints is difficult but can be accomplished if every fuzzy membership function is linear.

Especially for every normal triangular function with a core u_i, the constraint $\sum_i^n u_i = 1$ holds, and the order of singleton coefficients is assumed. Thus, the upper and lower limit of α -cut sets of linear sum can be easily calculated.

Let $f_t(x_k)$ be a crisp local weight of alternative *t* with respect to activity *k,* and in this paper, assume $0 \leq f_t(x_1) \leq f_t(x_2) \leq \cdots \leq f_t(x_n)$. Then, the overall weight of an alternative *t* is also the L-R fuzzy number and is represented as follows.

$$\tilde{v}_t = (v_t, l_t, r_t)_{LR}$$

where

$$v_t = \sum_k^n w_{1k} f_t(x_k),$$

$$l_t = v_t - \inf \text{supp}(\tilde{v}_t), \quad r_t = \sup \text{supp}(\tilde{v}_t) - v_t$$

In the above equations, inf supp, sup supp are lower and upper limits of support sets and are calculated as follows.

$$\inf \text{supp}(\tilde{v}_t) = \max_j \left[\sum_{i=1}^{j-1} (w_{1i} + \beta_i) f_t(x_i) + \sum_{i=j+1}^{n} (w_{1i} - \alpha_i) f_t(x_i) + \left\{ 1 - \sum_{i=1}^{j-1} (w_{1i} + \beta_i) - \sum_{i=j+1}^{n} (w_{1i} - \alpha_i) \right\} f_t(x_j) \right]$$

$$\sup \text{supp}(\tilde{v}_t) = \min_j \left[\sum_{i=1}^{j-1} (w_{1i} - \alpha_i) f_t(x_i) + \sum_{i=j+1}^{n} (w_{1i} + \beta_i) f_t(x_i) + \left\{ 1 - \sum_{i=1}^{j-1} (w_{1i} - \alpha_i) - \sum_{i=j+1}^{n} (w_{1i} + \beta_i) \right\} f_t(x_j) \right]$$

5. Conclusions

We proposed a representation for the inner dependence overall weights of alternatives by use of fuzzy sets and the result of a sensitivity analysis for cases in which consistency of the comparison matrix is poor. Our approach shows how to represent weights, as well as how the result of AHP has fuzziness, when inconsistency exists. This was due to reduced ambiguity in the representation presented in this work compared to previous normal fuzzy operations.

Acknowledgements

This research is partly supported by a grant from the ministry of education, sports, science, and technology to the national project of 'Advanced improvements of vision, image, speech and language information processing and the application to the technologies for the intelligent instrument and control' in the High-tech Research Center of Hokkai-Gakuen University.

ISBN: 978-989-95079-6-8

References

[1] T. L. Saaty. A scaling method for priorities in hierarchical structures. *J. Math. Psy*., 15(3):234--281, 1977.

[2] T. L. Saaty. The Analytic Hierarchy Process. *McGraw-Hill, New York*, 1980.

[3] T. L. Saaty. Scaling the membership function. European *J. of O.R.*, 25:320--329, 1986.

[4] M. Saito. An Introduction to Linear Algebra. *Tokyo University Press*, 1966.

[5] Y. Tanaka. Recent advance in sensitivity analysis in multivariate statistical methods. *J. Japanese Soc. Comp. Stat*., 7(1):1--25, 1994.

[6] K. Tone. The Game Feeling Decision Making. *Nikka-giren Press, Tokyo*, 1986.

[7] S. Ohnishi, H. Imai, and M. Kawaguchi. Evaluation of a Stability on Weights of Fuzzy Analytic Hierarchy Process using a sensitivity analysis. *J. Japan Soc. for Fuzzy Theory and Sys.,* 9(1):140--147, 1997..

[8] S. Ohnishi, H. Imai, T. Yamanoi. Weights Representation of Analytic Hierarchy Process by use of Sensitivity Analysis, *IPMU 2000 Proceedings*, 2000.

[9] D. Dubois and H. Prade. Possibility Theory An Approach to Computerized Processing of Uncertainty,. *Plenum Press, New York*, 1988.

[10] .T. L. Saaty. Inner and Outer Dependence in AHP. *University of Pittsburgh*, 1991.

[11] S. Ohnishi, T. Yamanoi, H. Imai, A Fuzzy Representation for Weights of Alternatives in AHP, *New Dimensions in Fuzzy Logic and Related Technologies,* Vol. II , pp.311-316 2007

[12] S. Ohnishi, D. Dubois, H.Prade, T.Yamanoi, A Fuzzy Constraint-based Approach to the Analytic Hierarchy Process, *Uncertainty and Intelligent Information Systems,* pp.2171-228 , 2008.

ISBN: 978-989-95079-6-8

On Conceptual Indexing for Data Summarization

Henrik Bulskov[1] Troels Andreasen[1]

1. Department of Computer Science, Roskilde University,
P.O. Box 260, DK-4000 Roskilde, Denmark Email: bulskov@ruc.dk, troels@ruc.dk

Abstract— A summary is a comprehensive description that grasps the essence of a subject. A text, a collection of text documents, a query answer can be summarized by simple means such as an automatically generated list of the most frequent words or "advanced" by a meaningful textual description of the subject. In between these two extremes are summaries by means of selected concepts exploiting background knowledge providing selected key concepts. We address in this paper an approach where conceptual summaries are provided through a conceptualization as given by an ontology. The idea is to restrict a background ontology to the set of concepts that appears in the text to be summarized and thereby provide a structure, a so-called instantiated ontology, that is specific to the domain of the text and can be used to condense to a summary not only quantitatively but also conceptually covers the subject of the text.

Keywords— conceptual clustering, conceptual descriptions, conceptual summaries, ontologies.

1 Introduction

The purpose of a summary is to provide a simplification to highlight the major points from the subject, e.g. a text or a set of texts such as a query answer. The aim is to provide a summary that grasps the essense of the subject.

Most common are summaries as those provided manually by readers or authors as a result of intellectual interpretation. Summaries can however also be provided automatically. One approach, in the Question Answering style, such as this is investigated in for instance the DUC and TREC conferences (see for instance [6],[5],[7]), is to provide a full natural language generation based summary construction while a less ambiguous, in the same tradition, is rather to perform a sentence selection from the text to be summarized.

In the other end the most simple approach is to select a reasonable short list of words among the most frequent and/or the most characteristic words from the set of words found in the text to be summarized. So rather than a coherent text the summary is a simple set of items.

Summaries in the approach presented here are also sets of items, but involves improvements over the simple set of words approach in two respects. First, we go beyond the level of keywords and aim to provide conceptual descriptions from concepts identified and extracted from the text. Second, we involve background knowledge in the form of an ontology. Strictly these two aspects are closely related – to use the conceptualization in the ontology we need means to map from words and phrases in the text to concepts in the ontology.

In this paper we present two different directions to conceptual summaries based on a background ontology. In both cases an ontology plays a key role as reference for the conceptualization. The general idea is from a world knowledge ontology to form a so-called "instantiated ontology" by restricting to a set of instantiated concepts.

First, we consider conceptual clustering over the instantiated concepts based on a semantic similarity measure and second we present an approach based on probabilities.

Below we first introduce to the the ontology notion, then we discuss extraction of conceptual descriptions, and finally we describe the various approaches to conceptual summaries.

2 Representing background knowledge – Ontology

Background knowledge is knowledge that complements the primary target data (the text or text collection / database) that is subject of the summarization with information that is essential to the understanding of this. Background knowledge can take different forms varying from simple lists of words to formal representations. In this context, however, our goal is conceptual summaries provided as sets of words or concepts so background knowledge to support this can range from unstructured lists of words to ontologies.

A simple list of words can be applied as a filter, mapping from a text to the subset of the word list that appears in the text. Such a controlled list of keywords or a vocabulary of topics can by obvious means be improved to capture also morphology by stemming or inflection patterns. Taxonomies, partonomies, semantic networks and ontologies are structures that potentially contribute also to knowledge-based summarization. Our main focus here is on ontologies ordered around taxonomic relationship. Rather than the common description logic based approach we choose here a simpler concept algebraic approach to ontologies.

One important rationale for this is that our goal here is not ontological reasoning in general but rather extraction of sets of mapped concepts and manipulation of such sets (e.g. contraction).

ISBN: 978-989-95079-6-8

2.1 An algebraic approach to ontologies

Given a basis taxonomy that situates a set of atomic term concepts $\mathcal{A}$ in a multiple inheritance hierarchy. Based on this we define a generative ontology by generalization of the hierarchy to a lattice an by introducing a (lattice-algebraic) concept language (description language) that defines an extended set of well-formed concepts, including both atomic and compound term concepts.

The concept language used here, ONTOLOG[9], has as basic elements concepts and binary relations between concepts. The algebra introduces two closed operations *sum* and *product* on concept expressions φ and ψ, where $(\varphi + \psi)$ denotes the concept being either φ or ψ and $(\varphi \times \psi)$ denotes the concept being φ and ψ (also called *join* and *meet* respectively).

Relationships r are introduced algebraically by means of a binary operator (:), known as the Peirce product $(r : \varphi)$, which combines a relation r with an expression φ. The Peirce product is used as a factor in conceptual products, as in $x \times (r\colon y)$, which can be rewritten to form the feature structure $x[r\colon y]$, where $[r\colon y]$ is an *attribution* of the concept x. Thus we can form compound concepts by attribution.

Given atomic concepts $\mathcal{A}$ and semantic relations $\mathcal{R}$, the set of well-formed terms $\mathcal{L}$ is:

$$\mathcal{L} = \{\mathcal{A}\} \cup \{x[r_1 : y_1, ..., r_n : y_n] | x \in \mathcal{A}, r_i \in \mathcal{R}, y_i \in \mathcal{L}\} \quad (1)$$

Compound concepts can thus have multiple as well as nested attributions. For instance with $\mathcal{R}$= {WRT, CHR, CBY, TMP, LOC, ...}[1] and $\mathcal{A}$ ={*entity, physical_entity, abstract_entity, location, town, cathedral, old*} we get:

$$\begin{aligned}&\mathcal{L} = \\ &\{entity, physical\_entity, abstract\_entity,\\ &location, town, cathedral, old,\\ &\ldots, \textit{cathedral}[\text{LOC}\colon \textit{town}, \text{CHR}\colon \textit{old}],\\ &\textit{cathedral}[\text{LOC}\colon \textit{town}[\text{CHR}\colon \textit{old}]], \ldots\}\end{aligned}$$

2.2 Modelling Ontologies

Obviously modelling ontologies from scratch is the best way to ensure that the result will be correct and consistent. However, for many applications the effort it takes is simply not at disposal and manual modeling have to be restricted to narrow and specific subdomains while the major part have to be derived from relevant sources. Sources that may contribute to modeling of ontologies may have various forms. A taxonomy is an obvious choice and it may be supplemented with, for instance, word and term lists as well as dictionaries for definition of vocabularies and for handling of morphology. Among the obviously useful resources are the semantic network WordNet [11] and the Unified Medical Language System (UMLS) [4] and unifies several other resources in the biomedical science area.

To go from a resource to an ontology is not necessarily straightforward, but if the goal is a generative ontology and the given resource is a taxonomy, one option is to proceed as follows. Given a taxonomy $\mathcal{T}$ over the set of atomic concepts $\mathcal{A}$ and a language $\mathcal{L}$, over $\mathcal{A}$ for a given set of relations $\mathcal{R}$, being derived as indicated in (1) above. Let $\hat{\mathcal{T}}$ be the transitive closure of $\mathcal{T}$. $\hat{\mathcal{T}}$ can be generalized to an inclusion relation "$\leq$" over all well-formed terms of the language $\mathcal{L}$ by the following

$$\begin{aligned}\text{“}\leq\text{“} = \quad &\hat{\mathcal{T}}\\ &\cup\{\langle x[\ldots, r\colon z], y[\ldots]\rangle | \langle x[\ldots], y[\ldots]\rangle \in \hat{\mathcal{T}}\}\\ &\cup\{\langle x[\ldots, r\colon z], y[\ldots, r\colon z]\rangle | \langle x[\ldots], y[\ldots]\rangle \in \hat{\mathcal{T}}\}\\ &\cup\{\langle z[\ldots, r\colon x], z[\ldots, r\colon y]\rangle | \langle x, y\rangle \in \hat{\mathcal{T}}\}\end{aligned}$$

where repeated ... denote zero or more attributes of the form $r_i\colon w_i$.

The general ontology $\mathcal{O} = (\mathcal{L}, \leq, \mathcal{R})$ thus encompasses a set of well-formed expressions $\mathcal{L}$ derived in the concept language from a set of atomic concepts $\mathcal{A}$, an inclusion relation generalized from the taxonomy relation in $\mathcal{T}$, and a supplementary set of semantic relations $\mathcal{R}$. For $r \epsilon \mathcal{R}$, we obviously have $x[r\colon y] \leq x$, and that $x[r\colon y]$ is in relation r to y. Observe that $\mathcal{O}$ is generative and that $\mathcal{L}$ therefore is potentially infinite.

An example is given in figure 1 showing a segment of a generative ontology build with WordNet as resource.

3 Referencing the background knowledge – providing descriptions

As already indicated the approach involves surveying text through the ontology provided and delivering summaries on top of the conceptualization of the ontology. For this purpose we need to provide a description of the text to be summarized in terms of the concepts in the ontology. So words and/or phrases must be extracted from the text and mapped into the ontology. This is a knowledge extraction problem, and obviously such knowledge extraction can span from full deep natural language processing (NLP) to simplified shallow processing methods.

Here we will consider the latter due to the counterbalance between the need for a full interpretation and the computational complexity of getting it. A very simple solution would match words in text with labels of concepts in the ontology, hence make a many-to-many relation between words in text and labels in the ontology that just accepts the ambiguity of natural language. Improvements can easily be obtained through pattern based information extraction / text mining and through methods in natural language processing.

First, a heuristic part of speech tagging can be performed on the text, and provided that word classes are assigned to the concepts given in the ontology this enables a word class based disambiguation.

Second, a stemming or, provided lexical information is available, a transformation to a standardized inflectional form can significantly improve the matching.

Third, given part of speech tagged input, simple syntactic natural language grammars can be used to chunk words together forming utterances or phrases [3], that

[1] for *with respect to, characterized by, caused by, temporal, location*, respectively.

ISBN: 978-989-95079-6-8

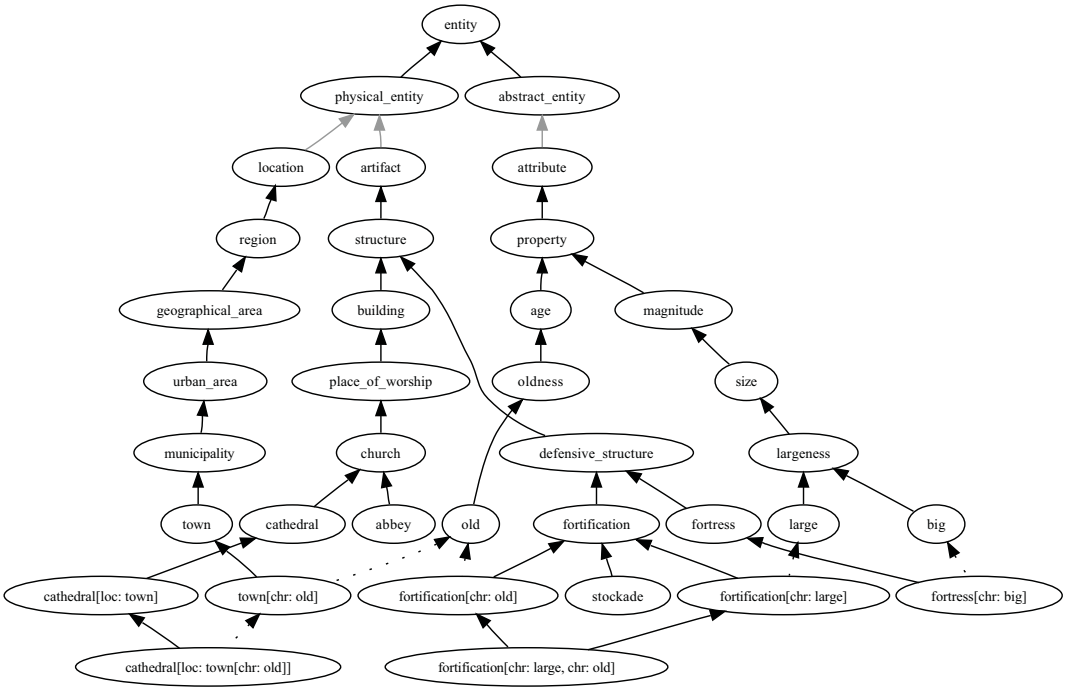


Figure 1: A segment of an ontology based on Wordnet. Does also correspond to an instantiated ontology for the set of instantiated concepts {*cathedral*[LOC : *town*[CHR : *old*]], abbey, *fortification*[CHR : *large*, CHR : *old*], stockade, *fortress*[CHR : *big*]}

.

can be used as the basis for matching against compound concepts in the ontology. Obviously the matching of chunks from the text and concepts in the ontology is in principle the same complex NLP problem over again, but the chunks identified will often correspond to meaningful concepts and therefore lead to a more refined and better result of the matching and, in addition, allow for a simple pattern-based approaches. We refer to [18] and [1] for more refined approaches. Here we will only cover a simple pattern-based approach.

Finally, some kind of word sense disambiguation [20] can be introduced in order to narrow down the possible readings of words, hence ideally mapping words of phrases to exactly one concept in the ontology.

A very simple approach along these lines is the following. Given a part-of-speech tagged and NP-chunked input a grammar for interpretation of the chunks is the following:

$$\begin{aligned} \text{Head} &::= \text{N} \\ \text{NP} &::= \text{A* N* Head} \mid \text{NP P NP} \end{aligned} \tag{2}$$

where A, N and P as placeholders for adjective, noun and preposition respectively. A very course-grained mapping strategy on top of this interpretation can be formed using the following transition rules, where premodifying adjectives relates to the head through *characterized by* (CHR) while premodifying composite nouns and prepositions both relates through *with respect to* (WRT):

$$\begin{aligned} &\text{A}_1, \ldots, \text{A}_n\text{N}_1, \ldots, \text{N}_m\text{Head} \mapsto \\ &\quad \text{Head}[\text{CHR}: \text{A}_1, \ldots, \text{CHR}: \text{A}_n, \text{WRT}: \text{N}_1, \ldots, \text{WRT}: \text{N}_m] \\ &\text{NP (P NP)}_1, \ldots, \text{(P NP)}_n \mapsto \\ &\quad \text{NP}[\text{ WRT}: \text{NP}_1, \ldots, \text{WRT}: \text{NP}_n] \end{aligned} \tag{3}$$

To test this approach we consider the the Metathesaurus in the Unified Medical Language System (UMLS) [13] as resource and build a generative ontology from this. For part of speech tagging and phrase chunking we use the MetaMap application [2].

Consider the following utterance[2] as an example:

> [...] *the plasma patterns of estrogen and progesterone under gonadotropic stimulation simulating early pregnancy* [...]

The first part of the analysis leads to part of speech tagging and phrase recognition as shown in table 1.
By applying the grammar (2) this can be transformed into the following three noun phrases:

plasma/N patterns/N of/P estrogen/N

progesterone/N under/P gonadotropic/A stimulation/N

early/A pregnancy/N

and by using the transition rules (3) we can produce the following compound expressions:

[2]This utterance is from a small 50K abstracts fraction of MEDLINE [14] having both *Homones* and *Reproduction* as major topic keywords.

ISBN: 978-989-95079-6-8

Table 1: Part of speech tagging and phrase recognition

Phrase	Type	Word	POS
Noun Phrase	det	the	det
	mod	plasma	noun
	head	patterns	noun
Preposition	prep	of	prep
	head	estrogen	noun
	conj	and	conj
Noun Phrase	head	progesterone	noun
Preposition	prep	under	prep
	mod	gonadotropic	adj
	head	stimulation	noun
	verb	simulating	verb
Noun Phrase	mod	early	adj
	head	pregnancy	noun

patterns[WRT: plasma, WRT: estrogen]

progesterone[WRT:stimulation[CHR:gonadotropic]]

pregnancy[CHR:early]

3.1 Instantiated Ontology

The description $d_O(T)$ of a text T given the ontology O comprise a set of concepts in O and as indicated the purpose here is to summarize based on relations in the ontology. Now given the set of concepts (the description) $d_O(T)$ an obviously relevant subontology is a subontology that covers all elements of $d_O(T)$. Such a subontology can be consider an instantiation of the text T (or the set of concepts $d_O(T)$).

Given an ontology $\mathcal{O} = (\mathcal{L}, \leq, \mathcal{R})$ and a set of concepts C we define the instantiated ontology $\mathcal{O}_C = (\mathcal{L}_C, \leq_C, \mathcal{R})$ as a restriction of $\mathcal{O}$ to cover only the concepts in C, that is, C and every concept from $\mathcal{L}$ that subsumes concepts in C or attributes for concepts in C. $\mathcal{L}_C$ can be considered an "upper expansion" of C in $\mathcal{O}$. More specifically, with C^+ being C extended with every concept related by attribution from a concept in C:

$$\begin{aligned} \mathcal{L}_C &= C \cup \{x | y \in C^+, x \in \mathcal{L}, y \leq x\} \\ \text{“} \leq_C \text{“} &= \{(x, y) | x, y \in \mathcal{L}_C, x \leq y\} \end{aligned} \quad (4)$$

Thus $\mathcal{O}_C$ is not generative. "$\leq_C$" may be represented by a minimal set "$\leq'_C$"$\subseteq$"$\leq_C$" such that "$\leq_C$" is derivable from "$\leq'_C$" by means of transitivity of "$\leq$" and monotonicity of attribution:

$$\begin{aligned} transitivity : x &\leq y, y \leq z \Rightarrow x \leq z \\ monotonicity : x &\leq y \Rightarrow z[r : x] \leq z[r : y] \end{aligned}$$

Figure 1 shows an example of an instantiated ontology. The general ontology is based on (and includes) WordNet and the ontology shown is "instantiated" wrt. the following set of concepts:

$$\begin{aligned} C = \{&\textit{cathedral}[\text{LOC} : \textit{town}[\text{CHR} : \textit{old}]], \textit{abbey}, \\ &\textit{fortification}[\text{CHR} : \textit{large}, \text{CHR} : \textit{old}], \textit{stockade}, \\ &\textit{fortress}[\text{CHR} : \textit{big}]\} \end{aligned}$$

4 Data summarization through background knowledge

The general idea here is to exploit background knowledge through conceptual summaries, that are to provide a means to survey textual data, for instance a query result. A set of concepts from the background knowledge is first identified in the text and then contracted into a smaller set of, in principle, most representable concepts.

This can be seen as one direction in a more general conceptual querying approach where queries can be posed, and answers be presented, by means of conceptual abstractions. For a general discussion on other means, except from conceptual summaries, of conceptual querying, where also a dedicated language constructs for this purpose is presented we refer to [8]. Here we discuss summaries only.

In the approach to summarization described here we assume an ontology to guide the summarization and, for the text to be summarized, an initial extraction of concepts as described in the previous section. Thus, we can assume an initial set of concepts C and we a facing a challenge to provide a smaller set of representative concepts covering C, that is, an appropriate summary that grasps what´s most characteristic about C. For computation of the summary we restrict to the subontology $\mathcal{O}_C = (\mathcal{L}_C, \leq_C, \mathcal{R})$ corresponding to the instantiated ontology for C.

We introduce two directions for deriving summaries below: one based on clustering of the input concept set C and the other realized as a probability-based restriction of the input ontology $\mathcal{O}_C$. Towards the end we discuss possibilities for combining these approaches.

4.1 Similarity Clustering

Given a similarity measure summaries can be derived from a clustering of concepts applying this measure. Obviously, if the measure is derived from an ontology, and thereby do reflect this, then so will the clustering. We will below assume an ontology-based similarity measure sim. A simple example of such a measure can be derived from the path lenght in the ontology graph (Rada' *Shortest Path Length* [12]). More refined approaches are *Information Content* [16] and *Weighted Shared Nodes* [17].

4.1.1 A hierarchical similarity-based approach

With a given path-length dependent similarity measure derived from the ontology a lub-centered, agglomerative, hierarchical clustering can be performed as follows.

Initially each "cluster" corresponds to an individual element of the set to be summarized. At each particular stage the two clusters which are most similar are joined together. This is the principle of conventional hierarchical clustering. However rather than replacing the two joined clusters with their union as in the conventional approach they are replaced by their lub. Thus given a set of concepts $C = \{c_1, \ldots, c_n\}$ summarizers can be derived as follows.

ALGORITHM – Hierarchical clustering summary

INPUT: Set of concepts $C = \{c_1, \ldots, c_n\}$

ISBN: 978-989-95079-6-8

OUTPUT: Generalizing description $\delta(C)$ for C.

1) Let the instantiated ontology for C be $\mathcal{O}_C = (\mathcal{L}_C, \leq_C \mathcal{R})$
2) Let $T = \{\langle x, y\rangle | sim(x, y) = max_{z,w \in C}(sim(z, w))\}$
3) Let $U = min\left(\{u | u \in \mathcal{L}_C \wedge \exists x, y \in \mathcal{L}_C : x < u \wedge y < u\}\right)$
4) $L = \{x | \langle x, y\rangle \in T \vee \langle y, x\rangle \in T\}$
5) set $\delta(C) = C \cup U/L$

As was also the case with the connectivity clustering, to obtain an appropriate description of C we might have to apply δ several times and at some point m we have that $\delta^m(C) = Top$.

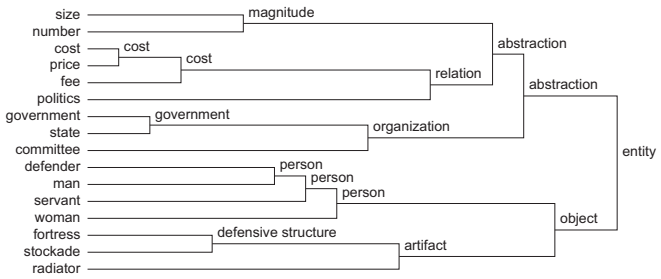


Figure 2: An illustration of the hierarchical clustering summary. The merging of two clusters are shown with their lub.

Figure 2 illustrates the application of δ a total of 15 times to the set of concepts from the previous example, where we might have (depending on the exact similarity values) for instance:

C = *{number, size, committee, government, state, defender, man, servant, woman, bribe, cost, price, fee, fortification, fortress, stockade}*

$\delta^1(C)$ = *{number, size, committee, government, state, defender, man, servant, woman, bribe,* ***cost****, fee, fortification, fortress, stockade}*

$\delta^2(C)$ = *{number, size, committee,* ***government****, defender, man, servant, woman, bribe, cost, price, fee, fortification, fortress, stockade}*

etc.

Thus summaries are generated iteratively and at each step the two closest concepts are clustered and the result is replaced by the corresponding lub.

4.1.2 Simple least upper bound-based approach

The principle of replacing clusters by their least upper bound can by applied on top of in principle any clustering approach. A straightforward similarity based approach is simply to apply a crisp clustering to the set of concepts $C = \{c_1, \ldots, c_n\}$ leading to $\{C_1, \ldots, C_k\}$ and then provide the set of lub's $\{\widehat{c}_1, \ldots, \widehat{c}_k\} = \{lub(C_1), \ldots, lub(C_k)\}$ for the division of C as summary. However to take into account also the importance of clusters in terms of their sizes the summary can be modified by the support of the generalizing concepts, $support(x, C)$, that for a given concept specifies the fraction of elements from the set C covered:

$$support(x, C) = \frac{|\{y | y \in C, y \leq x\}|}{|C|} \quad (5)$$

leading to a fuzzyfied (weighted) summary, based on the division (crisp clustering) of C into $\{C_1, \ldots, C_k\}$:

$$\Sigma_i support\left(lub(C_i), C\right) / lub(C_i) \quad (6)$$

To illustrate this lub-based approach consider table 2. Five groups are given that are derived as clusters of synsets in wordnet [3]

Table 2: A set of clusters and their least upper bounds from WordNet.

cluster	lub
{number, size}	magnitude
{committee, government, state}	organization
{defender, man, servant, woman}	person
{bribe, cost, fee, price}	cost
{fortification, fortress, stockade}	defensive structure

From these clusters the fuzzyfied summary {*.13/magnitude + .19/organization + .25/person + .25/cost + .19/defensive structure*} can be generated.

We may expect a pattern similar to hierarchical clustering in derivation of summaries in an approach based on similarity when the similarity measure reflects simple shortest path in the ontology.

This approach to summarization can be generalized using a fuzzyfied notion in place of the least upper bound as candidate representative. The generalization reduces the sensitivity against noise in the groups resulting from the initial clustering. This approach is described in [8].

4.2 A probability-based approach

We consider above a summary of textual input based on the concepts that appears in the text and how these are related in a background ontology. In the hierarchical approach candidate summarizers are chosen regardless of their coverage of the input, while the lub-based approached is introduced with a support that measures the degree to which all occurring input concepts are summarized. An obvious extension in this direction is to also consider frequencies of terms in the input text and thereby measure the probability of encountering an instance of a concept in the text.

Probabilities provides a means for selecting summarizers without taking other measures into account and thus allows for a straightforward approach as follows.

Given a set of concepts $C = \{c_1, \ldots, c_n\}$ and let $\mathcal{O}_C = (\mathcal{L}_C, \leq_C, \mathcal{R})$ be the instantiated ontology. Let further $child(c)$ denote the set of immediate children and $parent(c)$ the set of immediate parents for any concept $c \in \mathcal{L}_C$. The principle is to accumulate the frequencies to more general concepts but only so that a child c contributes with $\frac{1}{|parent(c)|}$ to each parent. A summary can be derived as follows.

ALGORITHM – Probability-based summary

INPUT: Set of concepts $C = \{c_1, \ldots, c_n\}$, their relative frequencies $C = \{freq(c_1), \ldots, freq(c_n)\}$ and a threshold α

[3] The first four are set of clusters and their least upper bounds in where $C_1, \ldots, C_4$ are from SEMCOR and C_5 is from the example ontology in figure 1.

ISBN: 978-989-95079-6-8

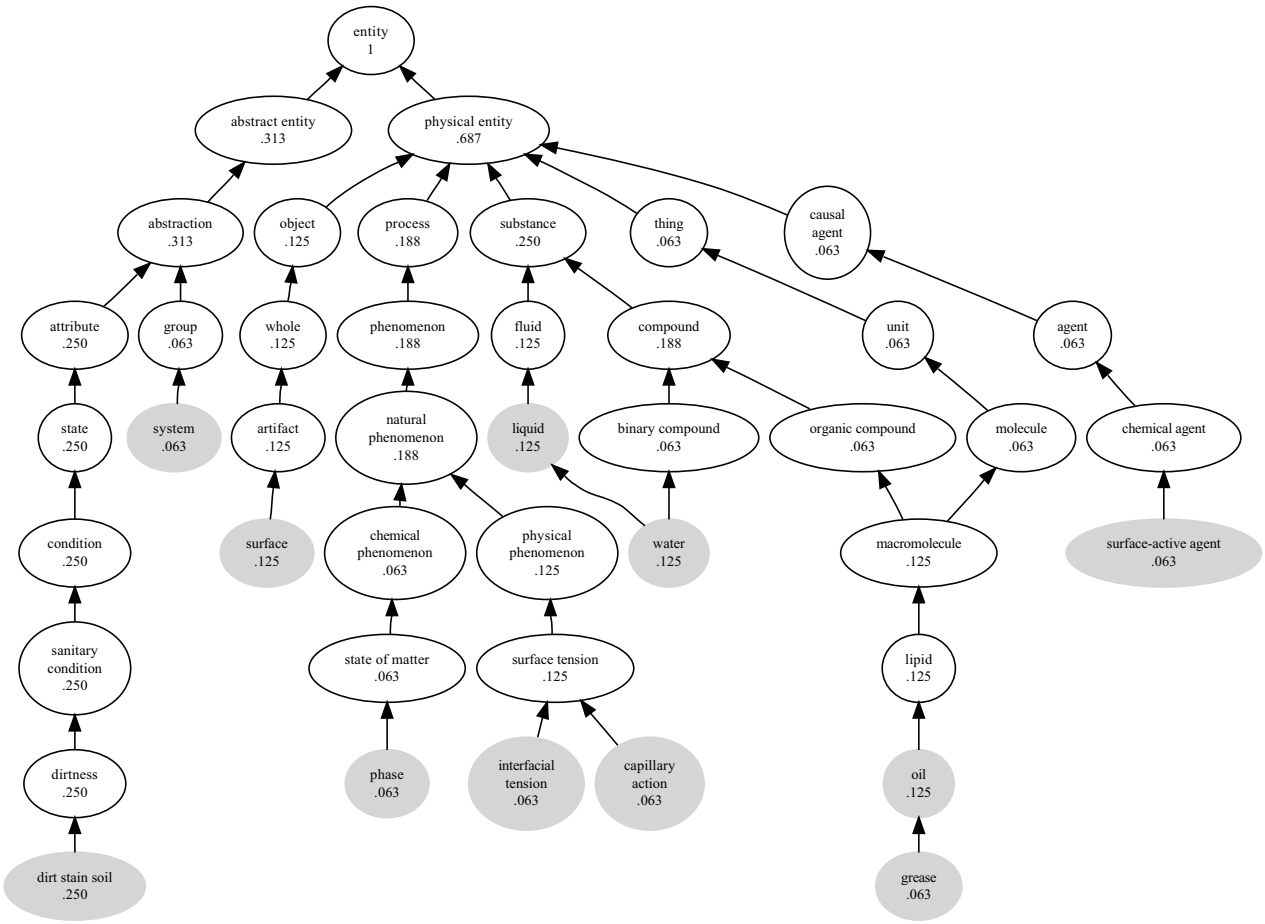


Figure 3: An instantiated ontology based on a paragraph from SEMCOR

OUTPUT: Generalizing description $D(C, \alpha)$ for C.

1) Let the instantiated ontology for C be $\mathcal{O}_C = (\mathcal{L}_C, \leq_C, \mathcal{R})$
2) Accumulate the frequencies in correspondence with the ontology such that $\forall c \in \mathcal{L}_C \backslash C : freq(c) = \sum_{c' \in child(c)} \frac{1}{|parent(c')|} freq(c')$
3) Let $N = |C|$ and $p(c) = freq(c)/N$ be the probability of encountering c.
4) Let $\mathcal{O'}_C = (\mathcal{L'}_C, \leq'_C, \mathcal{R}')$ be the restriction of $\mathcal{O}_C$ to the concepts that appear in $\{c \in \mathcal{L}_C | p(c) \geq \alpha\}$
5) Set the α-level summary of C to the most specific concepts appearing in $\mathcal{O'}_C$, that is $D(C, \alpha) = \{c \in \mathcal{L'}_C | \nexists c' \in \mathcal{L'}_C : c' < c\}$

As an example consider figure 3 that shows an instantiated ontology derived from WordNet for the following paragraph found in SEMCOR[10][4]:

> *Greases*, *stains*, and miscellaneous *soils* are usually sorbed onto the *soiled surface*. In most cases, these *soils* are taken up as *liquids* through *capillary action*. In an essentially static *system*, an *oil* cannot be replaced by *water* on a *surface* unless the *interfacial tensions* of the *water phase* are reduced by a *surface-active agent*.

Words in italics indicate the initial set of concepts[5]. Among the recognized concepts most appear only once, while the frequency of *surface* and *water* is two and the frequency of *soils* is 4 (includes *stains*). We have N =16 and C={*Greases*, *stains*, *soils*, *soiled*, *surface*, *soils*, *liquids*, *capillary action*, *system*, *oil*, *water*, *surface*, *interfacial tensions*, *water*, *phase*, *surface-active agent*} and thus get for instance

> $D(C, 0.1)$={*dirt stain soil*, *surface*, *surface tension*, *water*, *oil*}
>
> $D(C, 0.15)$={*dirt stain*, *soil*, *natural phenomenon*, *compound*}

5 Concluding remarks

In this paper we have considered how to use ontologies to provide data summaries with a special focus on textual data. Such summaries can be used in a querying approach where concepts describing documents, rather than documents directly, are retrieved as query answer. The summaries presented are conceptual due to fact that they exploit concepts from the text to be summarized and ontology-based because these concepts are drawn from a reference ontology.

[4]SEMCOR is a subset of the documents in the Brown corpus which has the advantage of being semantically tagged with senses from WordNet

[5]Notice that due to the use of SEMCOR there is no compound concepts in the initial set here.

ISBN: 978-989-95079-6-8

We have presented three summary principles. Two based on similarity and clustering and the third on probabilities derived from frequencies in the text to be summarized. Obviously a ”meaningful” clustering may lead to good summaries if characterizing subsuming concepts can be found in the ontology, but as indicated also counting occurrences, rather than only recognizing presence, of concepts may contribute to encircling essential concepts. Thus a next step should be to develop a combination that draws on similarity to ensure for instance that close summarizers are joined and on frequencies to ensure that derived summarizers are important.

For a summarization principle to work in practise users clearly need some kind of guidance on how many times to iterate or how to set a threshold. This brings up the important question on evaluation of summaries. Initial considerations on the quality of summaries can be found in [19] but the issue is also an obvious direction for further work in continuation of what has been described here.

References

[1] P. Anker Jensen, J. Fischer Nilsson Ontology-based Semantics for Prepositions, in Syntax and Semantics of Prepositions, P. Paint-Dizier (ed.), Text, Speech and Language Technology, Vol. 29, Springer, 2006.

[2] S. R. Aronson Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program, Proc AMIA Symp, pp 17-21, 2001.

[3] S. Abney Partial parsing via finite-state cascades, in: Proceedings of the ESSLLI´96 Robust Parsing Workshop, 1996.

[4] Olivier Bodenreider The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Research, 32, D267-D270. (2004)

[5] Udo Hahn, Inderjeet Mani The Challenges of Automatic Summarization Computer November 2000

[6] Gabor Melli, YangWang, Yudong Liu, Mehdi M. Kashani, Zhongmin Shi, Baohua Gu, Anoop Sarkar and Fred Popowich: Description of SQUASH, the SFU Question Answering Summary Handler for the DUC-2005 Summarization Task. Proceedings of DUC-2005, Vancouver, Canada pp 103-110 (205).

[7] Zhongmin Shi, Gabor Melli, Yang Wang, Yudong Liu, Baohua Gu, Mehdi M. Kashani, Anoop Sarkar and Fred Popowich Question Answering Summarization of Multiple Biomedical Documents Lecture Notes in Computer Science Springer Berlin / Heidelberg Volume 4509/2007 Advances in Artificial Intelligence pp 284-295 (2007)

[8] Andreasen, T.,Bulskov, H.: Conceptual Querying Through Ontologies. Fuzzy Sets and Systems, Elsevier, to appear (2008).

[9] Nilsson, J.F.: A logico-algebraic framework for ontologies – ONTOLOG. In Jensen, P.A., Skadhauge, P., eds.: First International OntoQuery Workshop, University of Southern Denmark (2001)

[10] Miller, G.A., Chodorow, M., Landes, S., Leacock, C., and Thomas R.G.: Using a semantic concordance for sense identification. In *Proc. of the ARPA Human Language Technology Workshop*, pages 240-243, 1994.

[11] Miller, G.A.: Wordnet: a lexical database for english. Commun. ACM **38**(11) (1995) 39–41

[12] Rada, R., Mili, H., Bicknell, E., Blettner, M.: Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man, and Cybernetics **19**(1) (January 1989) 17–30

[13] Unified Medical Language System U.S. National Library of Medicine http://www.nlm.nih.gov/research/umls/

[14] Medical Literature Analysis and Retrieval System Online U.S. National Library of Medicine http://www.ncbi.nlm.nih.gov/pubmed/

[15] Bulskov, H., Knappe, R., Andreasen, T.: On measuring similarity for conceptual querying. In: FQAS ’02: Proceedings of the 5th International Conference on Flexible Query Answering Systems, Springer-Verlag (2002) 100–111

[16] Resnik, P.: Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language (1999)

[17] Andreasen, T., Knappe, R., Bulskov, H.: Domain-specific similarity and retrieval. Proceedings IFSA 2005, Tsinghua University Press (2005) 496–502

[18] Andreasen, T. and Jensen, P. Anker and Nilsson, J. Fischer and Paggio, P. and Pedersen, B. Sandford and Thomsen, H. Erdman: Content-based text querying with ontological descriptors,Data Knowledge Engineering ,vol 48(2), pp 199–219, Elsevier Science Publishers

[19] Yager, R.R., Petry, F.E.: A Multicriteria Approach to Data Summarization Using Concept Hierarchies IEEE Trans. on Fuzzy Sys. vol 14 no 6 (2006)

[20] Zhou, Xiaohua, Han, Hyoil Survey of word sense disambiguation approaches 18th FLAIRS Conference (2005)

ISBN: 978-989-95079-6-8

Fuzzy analogical model of adaptation for case-based reasoning

Bernadette Bouchon-Meunier [1], Christophe Marsala [1], Maria Rifqi [1]

1 Université Pierre et Marie Curie - Paris 6, CNRS UMR 7606, LIP6,
104 avenue du Président Kennedy, F-75016 Paris, France
Email: {bernadette-bouchon-meunier,christophe.marsala,maria.rifqi}@lip6.fr

Abstract— *Case-based reasoning is a recognized paradigm and has been explored in both applied and methodological directions. In the several phases of CBR, the adaptation phase is certainly the most problematic whereas the most characteristic and interesting phase. We propose to view this task through a fuzzy analogical scheme. The adaptation is realized by focusing on the relation existing between the problem to be solved and the retrieved cases. Two approaches are proposed here: the relation can be captured by a fuzzy linguistic modifier or by a fuzzy interpolation.*

Keywords— Case-based reasoning, analogy, interpolation, adaptation, fuzzy sets.

1 Introduction

Case-based reasoning (CBR) has been explored in both applied and methodological directions. If the retrieval phase of the process has been widely investigated, less modelling has been proposed for the adaptation phase [13]. Derivational analogy [10] has been proposed to solve this problem when general domain knowledge is available, but it remains mainly a domain-dependent approach. More formal analysis of transformational analogy, assuming knowledge about differences in problems and their solutions, has been considered. In particular, a formal model of transformational adaptation has been proposed [2] on the basis of the consideration of a quality function evaluation.

2 Adaptation modeling in CBR

2.1 Presentation of the problem

Let us consider a problem space P and the solution space S and P' the subset of P of already solved problems. For a problem p in P, case-based reasoning can be regarded as the research of a solution s in S which can be associated with p. Let us denote by IAS the application "is a solution of" p and its solution s: $IAS : P \rightarrow S$. A commonly used method is to look for the most resembling elements p' to p in P' already associated with a solution s' in S. Either the best fitting problem is only taken into account and its associated solution is considered, or a group of similar problems is retrieved and various methods can be used to take them into account, by means of an aggregation method or a prototype-based approach for instance.

After the retrieval step, it is generally necessary to incorporate the dissimilarity between p and p' to produce a transformed view s' of the solution s. The transformation is for instance based on a degree of uncertainty attached to the solution s, on the proximity between elements of S, on a modulation of the linguistic expression of s, or on the use of constraints.

Several types of situations occur. In the first type of situation, S is a finite set of crisp solutions and the element s which will be associated with a given p is one of them, associated with the most similar p' in P'.

For *example*, we consider the assignment of a student grant to attend a conference, P is a set of students characterized by several attributes, S the solution space. In this case, S will be reduced to a two-element space "grant, no grant". The student p will be assigned a grant if the most similar element in P' has been assigned a grant.

The only possible adaptation in this case is the assignment of a degree of uncertainty to s'. Obviously, if a decision is to be made, a level of acceptable uncertainty will be chosen and no decision will be possible if this level is not reached.

In the second type of situation, S is infinite. For a given element p in P, the element s which will be associated with p is constructed from the solutions associated with the most similar problems p' in P'. A particular case corresponds to a continuous universe S and a finite number of linguistic values represented by fuzzy sets of S.

For instance, continuing the previous example, we can consider that a grant assignment can be defined as {refused, small, medium, high}. For a given student description p, the grant assignment may be expressed as "rather small" or "very high" if the value of the grant is supposed to depend on his/her merits or characteristics (his/her country is very far away, his/her grades are very good…).

We propose in this paper several adaptation techniques derived from fuzzy set based analogical reasoning and modelling. Our purpose is to propose a general approach based on fuzzy analogy to define adaptation methods in CBR.

2.2 Fuzzy analogy as a model for adaptation in CBR

Problems and solutions are, in many cases, described by means of linguistic terms or approximate values derived from expert knowledge, for instance "if the quality of the paper is very high then the grant assignment is highly recommended". A convenient knowledge representation is thus fuzzy set based. Let us denote by $[0,1]^{\Omega}$ the set of fuzzy sets of any universe Ω.

ISBN: 978-989-95079-6-8

A number of works have already presented various utilizations of fuzzy logic based representations in CBR [1][11][12][14] or analogy [20]. Very little attention has been drawn to the adaptation problem and we focus on this aspect in the present paper.

We consider the universe D of descriptions of problems p present in P and the application $\alpha : P \to [0,1]^D$ such that a problem p is described by a fuzzy set $\alpha(p)$ of D. We consider also the universe E of expressions of solutions in S. Each solution in S is associated by means of a one-to-one mapping γ to a fuzzy set $\gamma(s)$ of E.
Our purpose is to provide a model of transformational adaptation in case-base reasoning, starting from the opportunity to use analogy [18] to describe the assignment of a solution to a problem.

A scheme of analogy has been defined in a fuzzy framework [5] in order to mimic the most classic human means of deriving a solution to a new problem on the basis of past experiences in an automatic approach.
Starting from this scheme, and considering the specific environment of case-based reasoning, we propose the following definition of an analogical scheme in CBR. We consider two similarity relations R_1 on $P \times P$ and R_2 on $S \times S$, such that $R_1(p,p') = 1$ if and only if p and p' are similar, $R_2(s,s') = 1$ if and only if s and s' are similar.

Definition 1. For a given application $IAS : P \to S$ and two similarity relations R_1 on $P \times P$ and R_2 on $S \times S$, an *analogical scheme* on (P,S) is a function (see figure 1):

$$\Re_{IAS,R_1R_2} : P \times S \times P \to S \qquad (1)$$

satisfying $\forall p \in P$ and $\forall s \in S$ such that $s = IAS(p)$
(i) $s = \Re_{IAS,R_1R_2}(p,s,p)$
(ii) $\forall p' \in P$ such that $R_1(p,p') = 1$, $s' = \Re_{IAS,R_1R_2}(p,s,p')$ if and only if $s' = IAS(p')$ and $R_2(s,s') = 1$.

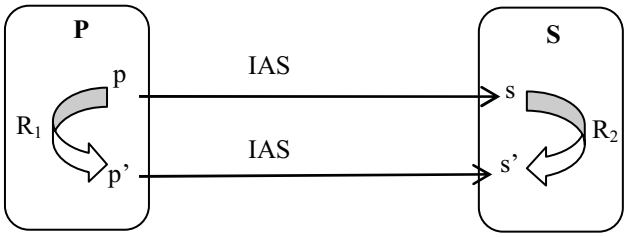


Figure 1. Global analogical scheme for case-based reasoning
x
Definition 1 means that if we want to solve a problem p' and we find a problem p similar to p' with a solution $s = IAS(p)$ in S, then the solution s' of p' will be similar to s. At this point, several problems p can be similar to p' and s' will be similar to all their solutions.
It is now necessary to define an operational way of evaluating the similarity R_1 of problems on the basis of their descriptions and the similarity R_2 of solutions on the basis of their expressions.

2.3 Measures of similitude

In many case-based reasoning applications, gradual evaluations of similarities are used, for instance based on distances between descriptions in D and between solutions in S or based on the compatibility with prototypes [19]. Let S_1 and S_2 be measures of similitude respectively defined on D and E to compare fuzzy sets. Several forms are available, depending of the properties we require from measures of similitude [8].
Following Tversky's seminal work on features of similarity [21], we define a measure of similitude on any universe Ω as follows, for a given *fuzzy set measure* $M : [0,1]^\Omega \to R^+$ such that $M(A) = 0 \Leftrightarrow A = \varnothing$ and M is monotonous with respect to the inclusion $\subseteq$ of fuzzy sets. We suppose also given an operation – of *difference* of fuzzy sets such that $A - A'$ is monotonous with respect to A and $A \subseteq A'$ implies $A - A' = \varnothing$.
With the convention that we use the same symbol for a fuzzy set and its membership function, the most used difference is defined for any y in Ω by:

$$A - A'(y) = \max(0, A'(y) - A(y)) \qquad (2)$$

Definition 2. A *measure of similitude* on Ω is a mapping $S_\Omega : [0,1]^\Omega \times [0,1]^\Omega \to [0,1]$ defined as:

$$S_\Omega(A,A') = F(M(A \cap A'), M(A' - A), M(A - A')) \qquad (3)$$

for a given mapping $F : R^{+3} \to [0,1]$ such that $F(u,v,w)$ is non-decreasing in u, non-increasing in v and w.

A measure of similitude can be regarded as a fuzzy relation on $[0,1]^\Omega \times [0,1]^\Omega$ and a fuzzy version of a similarity relation. Given a function F, we use measures of similitude on D and E to determine the similarities on P and S.
We assume that

$$R_1(p,p') = 1 \Leftrightarrow S_D(\alpha(p), \alpha(p')) \geq \varepsilon \qquad (4)$$

and $$R_2(s,s') = 1 \Leftrightarrow S_E(\gamma(s), \gamma(s')) \geq \varepsilon, \qquad (5)$$

for a chosen threshold ε.

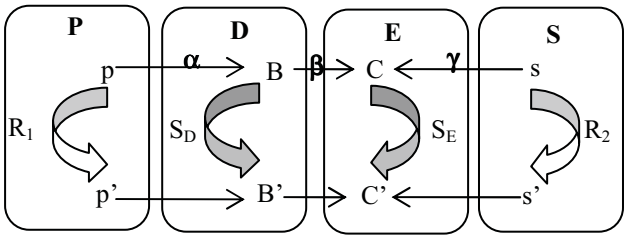


Figure 2. Operational analogical scheme for CBR

We obtain the operational analogical scheme described in Figure 2. The mapping β from D to E is defined in such a way that, for any B in $[0,1]^D$ and C in $[0,1]^E$, $C = \beta(B)$ if

ISBN: 978-989-95079-6-8

and only if there exist p in P and s in S satisfying $B = \alpha(p)$, $C = \gamma(s)$ and $s = IAS(p)$. We have $IAS = \gamma^{-1} \circ \beta \circ \alpha$.

Let us note that, crisp values being particular cases of fuzzy sets, the general scheme remains valid of values of attributes are precise. The measures of similitude must of course be chosen in an appropriate way [15].

2.4 Operational analogical scheme for CBR

It is now necessary to define an operational way of deriving s' from p' on the basis of the two measures of similitude S_D and S_E, and the analogical scheme $\Re_{IAS,R_1,R_2}$. We use a fuzzy version of analogical scheme as follows.

Definition 3. For a given application $\beta : [0,1]^D \to [0,1]^E$ and two measures of similitude S_D on $[0,1]^D \times [0,1]^D$ and S_E on $[0,1]^E \times [0,1]^E$, an $\varepsilon-$*analogical scheme* on (D,E) is a function defined for a given threshold $\varepsilon \in [0,1]$ as:

$$\Re^{\varepsilon}_{\beta,S_D S_E} : [0,1]^D \times [0,1]^E \times [0,1]^D \to [0,1]^E \quad (6)$$

satisfying the following conditions $\forall B \in [0,1]^D$ and $\forall C \in [0,1]^E$ such that $C = \beta(B)$

(iii) $C = \Re^{\varepsilon}_{\beta,S_D S_E}(B, C, B)$

(iv) $\forall B' \in [0,1]^D$ such that $S_D(B,B') \geq \varepsilon$, then

$C' = \Re^{\varepsilon}_{\beta S_D S_E}(B, C, B')$ if and only if $C' = \beta(B')$ and $S_E(C,C') \geq \varepsilon$.

This means that we need to determine a fuzzy $\varepsilon-$analogical scheme $\Re^{\varepsilon}_{\beta,S_D S_E}$ on $\left([0,1]^D, [0,1]^E\right)$ compatible with $\Re_{IAS,R_1R_2}$ (see figure 1). This compatibility is satisfied in the following way.

Property 1: Let us suppose that $IAS = \gamma^{-1} \circ \beta \circ \alpha$. For any p and p' in P, and s in S, the following holds:

$$s' = \Re_{IAS,R_1R_2}(p,s,p') \Leftrightarrow \gamma(s') = \Re^{\varepsilon}_{\beta,S_D S_E}(\alpha(p), \gamma(s), \alpha(p')).$$

Proof. If $p' = p$, then $s' = IAS(p') = IAS(p) = s$ by (i), also $\alpha(p') = \alpha(p)$ and $\gamma(s') = \gamma(s)$. Then (iii) is ensured.

Now, if $s' = \Re_{IAS,R_1R_2}(p,s,p')$, then $R_1(p,p') = 1$, $R_2(s,s') = 1$, and consequently the following inequalities hold:

$S_D(\alpha(p), \alpha(p')) \geq \varepsilon$, $S_E(\gamma(s), \gamma(s')) \geq \varepsilon$,

which entails $\gamma(s') = \Re^{\varepsilon}_{\beta,S_D S_E}(\alpha(p), \gamma(s), \alpha(p'))$ by (iv).

Conversely, if $\alpha(p') = \alpha(p)$, then $\gamma(s) = \gamma(s')$, and $s = s'$. Then (i) is ensured.

Now, if $\gamma(s') = \Re^{\varepsilon}_{\beta,S_D S_E}(\alpha(p), \gamma(s), \alpha(p'))$., then for $S_D(\alpha(p), \alpha(p')) \geq \varepsilon$, we get $\gamma(s') = \beta(\alpha(p'))$ and $S_E(\gamma(s), \gamma(s')) \geq \varepsilon$. For p and p' such that $S_1(p, p') = 1$, $s' = IAS(p')$ since γ is a one-to-one mapping, and $R_2(s, s') = 1$ by (5). Consequently, (ii) holds.

Continuing the previous *example*, we can consider a population P of students p, described by various attributes, such as the quality of their submitted paper, their ages, their countries… Let us consider universes of paper quality Q, of ages A, of countries C, we note D=Q×A×C and we consider fuzzy descriptions of students p, i.e. fuzzy sets of D, for instance (rather good quality, young, far country). Each student p will be assigned an amount of financial support in S which may be expressed by a fuzzy set of the universe of natural numbers E=IN, such as a "small grant".

3 Transformational adaptation in CBR

Since we have established a link between a comparison of problems in P or solutions in S and a comparison of their descriptions in D or expressions in E, we can work on (D,E) instead of (P,S) to determine methods of transformational adaptation in CBR.

3.1 Retrieval of a similar problem

We don't focus in this paper on the retrieval step of CBR. Let us only mention that there exist many ways to use the similarity relations R_1 to retrieve problems in P' which will be used to determine a solution in S to a given problem p.

For instance, the comparison can be restricted to typical problems in P'. The definition of prototypes can easily be solved through methods based on particular measures of similitude and dissimilarity in D [16]. Let P'' be the set of prototypes. They can be particular problems of P' or they can be abstract problems with descriptions in D summarizing the descriptions of a class of problems in P'.

The comparison can also be restricted to clusters of P' determined in agreement with solutions in S.

Let us remark that the threshold ε can be freely chosen by the user. If he/she chooses $\varepsilon = 0$, he/she takes into account all possible problems in P, and the determination of a unique p' associated with a given p lies in this case on the retrieval method. For instance, the retrieval method can be based on an optimization technique and p' will be the most similar to p, corresponding to the greatest value of $S_D(\alpha(p), \alpha(p'))$, under given additional constraints in some cases. Otherwise, the threshold ε can be considered as a level of decidability: if there exists no p' such that $S_D(\alpha(p), \alpha(p')) \geq \varepsilon$, then there is no already solved problem sufficiently similar to p and no solution can be proposed, unless more information is obtained.

If we now suppose that there exists a subset P_0 of such problems p', a method must be picked to determine a unique p'_0 in P_0 which will be used to find a solution to p by adapting $s'_0 = IAS(p'_0)$. Either we use an aggregation

ISBN: 978-989-95079-6-8

method to combine all p'_0 in P_0, or we look for the maximum of $\{S_D(\alpha(p),\alpha(p'))/\, p' \in P_0\}$, under additional constraints if necessary. Another solution is to look for the k-best values among $\{S_D(\alpha(p),\alpha(p'))/\, p' \in P_0\}$, with k chosen by the user, for instance $k = 2$. A refinement of the method considers a restriction of the search to elements of P_0 significantly different, with an additional threshold of decidability ε' such that, if the two best values in $\{S_D(\alpha(p),\alpha(p'))/\, p' \in P_0\}$ correspond to fuzzy subsets b and b' of D such that $|S_D(\alpha(p),b) - S_D(\alpha(p),b')| \le \varepsilon'$ and $\beta(b) \neq \beta(b')$, then no solution can be proposed to solve p, unless additional information is obtained.

We now focus on transformational adaptation techniques to determine the solution s of problem p.

3.2 Transformational adaptation by means of modifiers

We first consider that the retrieval method provides a unique p', such that $R_1(p,p') = 1$, best fitting the considered problem p. The value $S_D(\alpha(p),\alpha(p'))$ of the measure of similitude gives an information on the level of resemblance between p and p'. This value is used to adapt the solution $s = IAS(p)$ in order to define $s' = IAS(p')$. The $\varepsilon-$analogical scheme introduced in Definition 3 indicates that s will be determined in such a way that $S_E(\gamma(s),\gamma(s')) \ge \varepsilon$. It is clear that, $C' = \gamma(s')$ being a fuzzy set of D, there exist many fuzzy sets $C = \gamma(s)$ such that this condition is satisfied.

A convenient way to define C from C' is to use a modifier [4], such that s' will be a linguistically modified form of s. For example, if s is "small", then s' may be "rather small", or "more or less small", depending on the context.
A modifier defined for a universe Ω (Ω being either D or E) is a mapping $m : [0,1]^{\Omega} \to [0,1]^{\Omega}$.
We restrict ourselves to $\varepsilon-$*modifiers* m on Ω such that

$$M(m(X) - X) = 1 - \varepsilon \qquad (7)$$

The modifier is expansive if $X \subseteq m(X)$ for any X in $[0,1]^{\Omega}$. Examples of expansive modifiers are "more or less" or "approximately". Modifiers m are defined by parameterized mathematical transformations such as homotheties [3].
The transformational adaptation process is then the following: given the measure of similitude S_Ω, we define the solution s to problem p in such a way that $S_E(\gamma(s),\gamma(s')) \ge \varepsilon$. The form of m depends on the chosen measure of similitude and the parameters defining m are deduced from ε.

We have proved in [6] that the choice of a particular measure of similitude and the choice of the modifier are linked. Working with expansive modifiers, for instance, leads to use a particular measure of similitude called a measure of *inclusion*, defined as a reflexive measure of similitude (r.m.s.) such that $F(0,v,w) = 0$ for any v and w and $F(u,v,w)$ is independent of v. Classic examples of measures of inclusion are the following:

- $S_\Omega(X,X') = \dfrac{M(X \cap X')}{M(X)}$, (8)

M standing for the fuzzy cardinality of X,

$$M(X) = \int_{y \in \Omega} X(y) \qquad (9)$$

- $S_\Omega(X,X') = 1 - M(X - X')$, (10)

M standing for the height of X:

$$M(X) = \sup_{y \in \Omega} X(y). \qquad (11)$$

Property 2. Let us suppose that the difference of fuzzy sets is defined by (2), the fuzzy set measure by (11) and the similarity by (10). For any given problem p associated with a description B, an $\varepsilon-$modifier m provides the expression of a solution s of p compatible with the $\varepsilon-$analogical scheme i.e. such that $m(C') = \Re^{\varepsilon}_{\beta S_D S_E}(B',C',B)$ for any B,B',C'.

Proof. If m is an $\varepsilon-$modifier, then $M(m(C') - C') = 1 - \varepsilon$, which means that $\sup_{x \in E}(m(C') - C') = 1 - \varepsilon$. In the case where $B = B'$, then $S_D(B,B) = 1$, $\varepsilon = 1$. Then $M(m(C') - C') = 0$ by (7). Then $m(C') = C = C'$ and consequently $C' = \Re^{\varepsilon}_{\beta S_D S_E}(B',C',B')$, satisfying condition (iii). In the case where $\varepsilon \neq 1$, we have $m(C')$ is such that $m(C') = \gamma(s)$ by construction and $S_E(C',m(C')) = \varepsilon$ by (7). Consequently $m(C') = \Re^{\varepsilon}_{\beta S_D S_E}(B',C',B)$.

The properties of a measure of inclusion show that $S_\Omega(X,X')$ is independent of $X' - X$. This means that the choice of a measure of inclusion on E provides $S_E(m(C'),C')$ independent of $C' - m(C')$. This leads to use an expansive modifier m, since it corresponds to $C' - m(C') = \varnothing$ and $C' - m(C')$ has no influence on the value of a measure of inclusion.
An example of such an expansive $\varepsilon-$modifier is the following:

$$C(x) = m(C')(x) = \min\left(1, \frac{C'(x)}{\varepsilon}\right) \text{ for any } x \in E,$$

if we use (10) for M, s expressed as "approximately s'".

Considering again the previous *example*, if a "good" paper's author is assigned a "high grant", a similar paper's author is assigned an "approximately high" grant. The utilization of an expansive modifier reveals a cautious approach.

3.3 Transformational adaptation by means of modifiers in a gradual environment

In a different context, we consider the specific case where there is a gradual link between a variation on D and a

ISBN: 978-989-95079-6-8

variation on E. An example is expressed by "the better the quality of a student's paper, the higher the assigned grant".
Several methods are then possible to obtain a solution s to a problem p. We assume that D and E are included in the universe **R** of real numbers.

We consider again that the retrieval method provides a unique p', such that $R_1(p,p')=1$, best fitting the considered problem p. An appropriate form of modifier corresponds to translations of X to the left or to the right [9], with an amplitude of translation λ (positive or negative), to obtain the modified version $m(X)$. All components of the comparison between X and X' are taken into account, namely $X \cap X'$, $X - X'$ and $X' - X$.
The classic definition of a translatory modifier is the following, if the universe E is $[0,1]$, to simplify:

$m(C')(x) = C'(x+\lambda)$ if $x+\lambda \in [0,1]$

$m(C')(x) = C'(0)$ if $x+\lambda \leq 0$

$m(C')(x) = C'(1)$ if $x+\lambda \geq 1$.

This means that $m(C')$ is generally the translation of C', except at both ends of E. It is easy to check that $M(m(C') - C') = M(C' - m(C')) = |\lambda|$, except at both ends, where $M(m(C') - C') \leq |\lambda|$ or $M(C' - m(C')) \leq |\lambda|$, if we consider the fuzzy set measure (9).
Such a modifier is associated with any form of r.m.s., defined by $F(u,v,w)$, and depending on u, v, w. We can consider the measure of similitude (10) associated with the fuzzy set measure (9).
The amplitude must be defined in our case with respect to ε. Choosing $|\lambda| = 1-\varepsilon$ ensures that $S_E(m(C'),C') \geq \varepsilon$. We then obtain the following result :

Property 3. Let us suppose that the universe is $[0,1]$. If the difference of fuzzy sets is defined by (2), the fuzzy set measure by (9) and the similarity by (10), then a translatory modifier m defined by an amplitude ε such that $|\lambda| = 1-\varepsilon$ satisfies $m(C') = \Re^{\varepsilon}_{\beta S_D S_E}(B',C',B)$ for any B,B',C', according to definition 3.

The solutions we deduce in S from the use of such modifiers in E is more diverse. Depending on the sign and the value of the amplitude λ, depending also on the context, the solution s can take into account a reinforcement or a weakening of $\gamma^{-1}(C')$. In more concrete terms, if a student p' in P' is described by "good quality paper" in D and the solution $s' = IAS(p')$ expressed by "high assigned grant" in E, then a student p similar to p' will correspond to an assignment of "very high grant" (reinforcement) or "rather high grant" (weakening). To determine which of these two possibilities is the right one, more information must be taken into account about the relative position of B and B'.

3.4 Transformational adaptation by means of interpolation

We still consider that there is a gradual link between a variation on D and a variation on E. We now suppose that we take into account two problems in P' similar to p in order to find a solution $s = IAS(p)$ in S in a more comprehensive way.
The problems in P' are associated with descriptions represented by a family $D' = \{\alpha(p_1), \alpha(p_2)...\}$ of fuzzy sets of D and we further suppose that there exists an order $\prec$ on D' such that:

$$\alpha(p_1) \prec \alpha(p_2) \prec ... \prec \alpha(p_i) \prec \alpha(p_{i+1}) \prec ... \prec \alpha(p_n)$$

Let $E' = \{\gamma(s_1), \gamma(s_2)...\}$ be the family of associated expressions of solutions represented by fuzzy sets in E, with $s_i = IAS(p_i)$ for any $i = 1,2...$, supposed to be equipped with the same order $\prec$.
Without any loss of generality, we can then suppose that:

$$\gamma(s_1) \prec \gamma(s_2) \prec ... \prec \gamma(s_i) \prec \gamma(s_{i+1}) \prec ... \prec \gamma(s_n).$$

For a given p in P, we propose to look for the most similar p', namely p_i, and consider the pair (p_i, p_{i+1}) if $\alpha(p_i) \prec \alpha(p)$ or the pair (p_{i-1}, p_i) if $\alpha(p) \prec \alpha(p_i)$. For the sake of simplicity, we consider only the first case, such that $\alpha(p_i) \prec \alpha(p) \prec \alpha(p_{i+1})$.
Because of the assumption of graduality, we will have $\gamma(s_i) \prec \gamma(s) \prec \gamma(s_{i+1})$. The methods to determine s are again various. We propose to use a method based on interpolation [7][18]. The general spirit of the method is based on the basic analogical process "s is to s_i and s_{i+1} as p is to p_i and p_{i+1}".
The method to determine $C = \gamma(s)$ is based on the following steps: first, we compare $\alpha(p)$ to $(\alpha(p_i), \alpha(p_{i+1}))$. Secondly, we construct a family of possible fuzzy sets of E similar to the pair $(\gamma(s_i), \gamma(s_{i+1}))$ in a way analogous to the way $\alpha(p)$ is similar to the pair $(\alpha(p_i), \alpha(p_{i+1}))$. Finally, we deduce one element C from this family.

The detailed process we have proposed takes into account both location in the universe E and shape (understood as their membership function) of fuzzy sets and we proceed as follows.
Step1. We determine the location loc(C) of C, considering that it is to the locations of $\gamma(s_i)$ and $\gamma(s_{i+1})$ as the location of $\alpha(p)$ is to those of $\alpha(p_i)$ and $\alpha(p_{i+1})$.
Step 2. We translate $\alpha(p_i)$ and $\alpha(p_{i+1})$ towards $\alpha(p)$ with respect to locations, to obtain B'_i and B''_i.
Step 3. We compare the shapes of B'_i and B''_i to the shape of $\alpha(p)$.
Step 4. We translate $\gamma(s_i)$ and $\gamma(s_{i+1})$ to loc(C) to obtain respectively C'_i and C''_i.
Step 5. We construct two fuzzy sets C' and C'' of E with location loc(C) such that the shape of C' (resp. C'') can be

ISBN: 978-989-95079-6-8

compared to the shape of C'_i (resp. C''_i) in the same way as the shape of $\alpha(p)$ is compared with the shape of B'_i (resp. B''_i).

Step 6. We aggregate C' and C'' to construct C.

The interest of such a method appears when the family D' is sparse, which means that it is possible to find problems p falling into a "gap" between already solved problems, their similarity being relatively low. This method can also be used in ordinary cases, with a concern of quality of the solution, since taking into account two elements provides a more gradual, and then more robust, treatment of the problem.

4 Conclusion

We have pointed out several methods enabling the user to perform the transformational adaptation of the solution to a similar problem. These methods are diverse and their utilization depends on the context. Their interest is to ensure a gradual passage between cases and the global utilization of the set of already solved problems. Experiments will be available in the future to compare methods.

References

[1] K.S. Aggour, M. Pavese, P.P. Bonissone, W.E. Cheetham. SOFT-CBR: A self-optimizing fuzzy tool for case-based reasoning, *In Proceedings ICCBR 2003*, LNAI 2689, pp. 5-19, 2003.

[2] R. Bergmann, W. Wilke. Towards a New Formal Model of Transformational Adaptation in *Case-Based Reasoning. In Proceedings of the ECAI'98 Conference*, pp.53-57, 1998.

[3] B. Bouchon - Stability of linguistic modifiers compatible with a fuzzy logic ; in *Uncertainty and Intelligent Systems*, Lecture Notes in Computer Sciences n° 313, B. Bouchon, L. Saitta, R.R. Yager (eds.), Springer Verlag, pp. 63-70, 1988

[4] B. Bouchon-Meunier. Fuzzy logic and knowledge representation using linguistic modifier, in *Fuzzy logic for the Management of Uncertainty*, L.A. Zadeh, J. Kacprzyk, (eds.), John Wiley and Sons, 399-414, 1992.

[5] B. Bouchon-Meunier, J. Delechamp, C. Marsala, M. Rifqi. Several Forms of Fuzzy Analogical Reasoning, in *Proceedings 6th IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE'97*, Barcelona, Spain, vol. I, pp. 45-50, 1997.

[6] B. Bouchon-Meunier, C. Marsala. Linguistic modifiers and measures of similarity or resemblance. In *Proceedings of the joint 9th IFSA World Congress and 20th NAFIPS International Conference*, Vancouver, Canada, vol. 4, pp. 2195-2199, 2001.

[7] B. Bouchon-Meunier, C. Marsala, M. Rifqi. Interpolative reasoning based on graduality, in *Proceedings of the FUZZ-IEEE'2000 Conference*, San Antonio, USA, pp.483-487, 2000.

[8] B. Bouchon-Meunier, M. Rifqi, S. Bothorel. Towards general measures of comparison of objects, *Fuzzy Sets and Systems*, 84, 2, 143-153, 1996.

[9] B. Bouchon-Meunier, Yao Jia. Linguistic modifiers and imprecise categories, *International Journal of Intelligent Systems* 7, pp.25-36, 1992

[10] J. G. Carbonell. Derivational Analogy: A theory of reconstructive problem solving and expertise acquisition. In *Machine Learning: An Artificial Intelligence Approach*. Michalski, R., Carbonell, J., & Mitchell, T. (Eds.) Morgan Kaufman Publishers: San Mateo, CA. pp. 371-392, 1986.

[11] D. Dubois, F. Esteva, P. Garcia, L. Godo. R. Lopez de Mantaras, H. Prade. Case-based reasoning: a fuzzy approach, *In Fuzzy Logic in Artificial Intelligence*, LNCS vol. 1566, pp. 79-90,1999.

[12] F. Esteva, P. Garcia, L. Godo, E. Ruspini, L. Valverde. On similarity logic and the generalized modus ponens, in *Proceedings of the FUZZ-IEEE'94 Conference*, Orlando, USA, pp. 1423-1427, 1994.

[13] B. Fuchs, A. Mille. A knowledge-level task model of adaptation in case-based reasoning, in *Proceedings of the ICCBR-99: International Conference on Case-Based Reasoning*, Lecture notes in computer science, vol. 1650, pp. 118-131, 1999.

[14] E. Hüllermeier, D. Dubois, H. Prade. Fuzzy rules in case-based reasoning, *In Conf AFIA99, Raisonnement à partir de cas*, pp. 45-54, 1999.

[15] M.-J. Lesot, M. Rifqi et H. Benhadda. Similarity measures for binary and numerical data: a survey. In *International Journal of Knowledge Engineering and Soft Data Paradigms (KESDP),* 1(1):63-84, 2009

[16] M.-J. Lesot, M. Rifqi, B. Bouchon-Meunier. Fuzzy prototypes: From a cognitive view to a machine learning principle, in H. Bustince, P. Herrera, J. Montero (eds.) *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models. Intelligent Systems from Decision Making to Data Mining, Web Intelligence and Computer Vision*, Springer, pp. 431-453, 2007.

[17] C. Marsala, B. Bouchon-Meunier. Interpolative Reasoning with Multi-Variable Rules. In *Proceedings of the joint 9th IFSA World Congress and 20th NAFIPS International Conference*, Vancouver, Canada, vol. 4, pp. 2476-2481, 2001..

[18] R. Schmidt, O. Vorobieva, L. Gierl. Case-Based Adaptation Problems in Medicine. In *Proceedings of WM2003: Professionelles Wissensmanagement Erfahrungen und Visionen*. Kollen-Verlag. 2003.

[19] R. Schmidt, T. Waligora, O. Vorobieva. Prototypes for Medical Case-Based Applications. In *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*, Springer, pp. 1-15, 2008.

[20] I. B. Turksen, Zhao Zhong. An approximate analogical reasoning approach based on similarity measures, *IEEE Trans. on Systems, Man and Cybernetics*, 18, 6, pp. 1049-1056, 1988.

[21] A. Tversky, Features of similarity. *Psychological. Review,* 84, pp. 327–352. 1977.

ISBN: 978-989-95079-6-8

Embedded Genetic Learning of Highly Interpretable Fuzzy Partitions

Jorge Casillas
CITIC-UGR (Research Center on Communication and Information Technology),
Dept. Computer Science and Artificial Intelligence, University of Granada, E-18071, Granada, Spain
E-mail: casillas@decsai.ugr.es, URL: http://decsai.ugr.es/∼casillas

Abstract— *A new algorithm is proposed to learn fuzzy partitions with a high interpretability degree. The set of input variables, the number of linguistic terms per variable, and the type (triangular or trapezoidal) and parameters of the membership functions are learnt by means of a meta-algorithm that uses a simple learning method to generate a fuzzy rule set from the derived fuzzy partitions. Interpretability constrains and powerful genetic operators are considered. A multi-objective optimization approach is used to generate different interpretability-accuracy tradeoffs. The algorithm is tested in a set of real-world regression problems with successful results compared to other methods.*

Keywords— Fuzzy rule-based systems, fuzzy partitions, granularity, interpretability, multi-objective genetic algorithms.

1 Introduction

In the last few years the interpretability (i.e. the capability of the fuzzy model to express the behavior of the real system in an understandable way) has received more and more attention in the fuzzy community. Indeed, this property of the fuzzy rule-based systems represents one of the main competitive advantages compared to other system modeling techniques.

In the literature we may find many alternatives to improve the interpretability such as reducing the number of fuzzy rules [1], selecting a subset of input variables [2], or using more compact fuzzy rule expressions [3]. Within these options, one of the more important approaches involves to learn the optimal number of linguistic terms per variable [4, 5, 6]. Indeed, to use a reduced number of linguistic terms is crucial to understand the meaning of the variables and directly influences on the fuzzy rule set size.

In this paper we propose a new algorithm to learn the number of input variables, the number of linguistic terms per variable, and the types of membership functions (triangular or trapezoidal) and their parameters with the aim of generating highly interpretable fuzzy partitions. The learning is performed by a wrapper-based embedded process where a meta-algorithm generates different fuzzy partitions and a simple learning method derives fuzzy rule sets from them.

The proposed algorithm, called EGLFP, has some interesting characteristics that make it very competitive: it uses strong fuzzy partition and includes distinguishability constrains for a better interpretability, and it uses variable-length coding schemes, powerful original crossover and mutation operators, and multi-objective optimization for a better search process.

The paper is organized as follows. Section 2 describes the proposed algorithm. Section 3 shows the results obtained in a set of real-world problems compared with other fuzzy rule learning methods. Finally, Sect. 4 concludes and suggests some further works.

2 EGLFP Algorithm

EGLFP is a multi-objective genetic algorithm with a generational evolutionary approach. A crossover operator that recombines membership function parameters, a first mutation operator that tunes these parameters, and a latter mutation operator that changes the fuzzy partition granularity (i.e., the number of linguistic terms) per variable are used in EGLFP. The multi-objective approach is based on the well-known NSGA-II algorithm [7]. The following sections detail the different components of EGLFP.

2.1 Coding Scheme

For the sake of a good interpretability and in order to reduce the search space tackled by the genetic algorithm, we propose the use of strong fuzzy partitions. Each gene (g) is a 2-tuple that contains the information related to a linguistic term of a specific variable. It consists of two real-valued fields (g_{left} and g_{right}) that encode the left and right extremes of the core of the associated fuzzy set (i.e., the semantic rule of the linguistic term) normalized to the interval $[0, 1]$. Figure 1 illustrates an example of gene's coding.

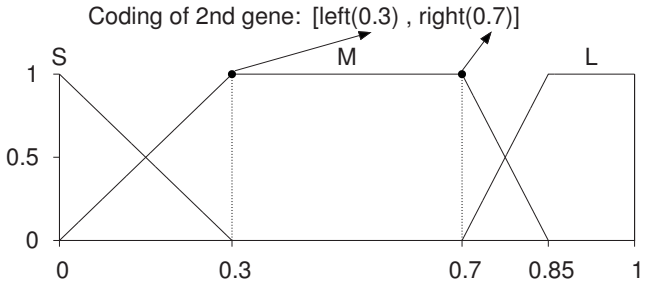


Figure 1: Example of a gene coding. Each gene, which encodes a linguistic term, consists of two fields (*left* and *right*) to encode the extremes of the fuzzy set's core (normalized to $[0, 1]$). Note that, since strong fuzzy partitions are used, the extremes of the core of each fuzzy set coincide with the extremes of the support of the adjacent fuzzy sets

A chromosome is a variable-length string (the length will depend on the number of linguistic terms used in each variable) of these genes that will encode the complete definition of the fuzzy partitions of all the input and output variables:

$$C = \bigcup_{v=1}^{n+m} \bigcup_{i=1}^{l_v} (g^v_{i,left}, g^v_{i,right}) \quad (1)$$

with n being the number of input variables, m the number of output variables, and l_v the number of linguistic terms used in the variable v.

ISBN: 978-989-95079-6-8

It is worth mentioning this coding scheme allows the absence of some variables by using only one linguistic term. Thus, if the variable v is not used, there will be only one gene associated to it with the value $(g^v_{1,left}, g^v_{1,right}) = (0, 1)$.

2.2 Initialization

The initialization process generates the first pool. Among the generated individuals, one of them will encode uniformly distributed triangular-shaped strong fuzzy partitions with the maximum number of linguistic terms allowed for each variable ($maxLV_v$). The remaining pool is filled up at random with individuals of two types.

One the one hand, some individuals (approximately a half of the pool) will preserve in some degree the above mentioned uniformly distributed fuzzy partitions but with a lesser number of linguistic terms. To do so, a fusion operation (which is described below in Sect. 2.5.1) on neighboring fuzzy sets is applied over the original fuzzy partitions. On the other hand, other individuals (approximately the another half of the pool) will be generated completely at random. To do so, firstly the number of linguistic terms is randomly assigned to each variable, then each fuzzy set adopts a triangular or trapezoidal shape at random and, finally, random values are assigned to the extremes of the fuzzy sets' cores.

2.3 Crossover

The crossover generates two individuals that inherit the membership function parameter definition given by two parents, even when these parents hold fuzzy partitions with different number of linguistic terms. This feature, the fact of crossing the parameters of fuzzy partitions with different granularities, provides the proposed algorithm with a powerful search process and is one of the original contributions of the paper.

The proposed crossover is a kind of parent-centric crossover operator [8] where each son is generated from a parent that plays the main role (*dominant parent*) and the another parent playing the secondary role (*recessive parent*). Thus, the son S_1 is generated focused on the parent C_1, but using the parent C_2 to add diversity; analogously with the son S_2.

Since the considered chromosomes have a variable length, a process is followed to select a gene of the recessive parent to be crossed with each gene of the dominant parent. This match is based on the distance (see eq. 5 below) where, given a fuzzy set of the dominant parent to be crossed, the closest fuzzy set in the recessive parent is chosen. Furthermore, to maintain the interpretability and well definition of the fuzzy partitions, an interval variation is fixed every time a gene is going to be crossed. An example is depicted in Fig. 2(a).

Once the dominant and recessive genes are fixed and the interval variation is determined, an original real-coding crossover operator is used. We have named this operator as the Constrained Parent-Centric Crossover (CPCX). Contrary to other previously proposed parent-centric crossover operators [8], CPCX is designed to generate the son's gene constrained to a given interval, which is crucial in our EGLFP algorithm to ensure well-defined and distinguishable fuzzy partitions. The operator is described in Algorithm 1 and an example is given in Figure 2(b).

Algorithm 1: Constrained Parent-Centric Crossover (CPCX) operator

Input: (g, h, min, max, m, M) with g being the dominant value and h the recessive value to be crossed, $[min, max]$ the variation interval of the dominant value, and $[m, M]$ the definition interval of the variable
Output: New value g'
begin
  if $g = h$ **then**
    $g' \leftarrow g$
  else if $h < g$ **then**
    $\alpha \leftarrow (g - h)/(g - m)$
    $l \leftarrow g - \alpha(g - min)$
    $r \leftarrow g$
    $g' \leftarrow U[l, r]$ */* Random number */*
  else
    $\alpha \leftarrow (h - g)/(M - g)$
    $l \leftarrow g$
    $r \leftarrow g + \alpha(max - g)$
    $g' \leftarrow U[l, r]$ */* Random number */*
end

2.4 Parameter Mutation

The parameter mutation changes the real-valued membership function parameter values of the input and output variables encoded in the chromosome. To do that, an original real-coding mutation operator is proposed in this paper. Firstly, a random process is followed to select the field of the gene to be mutated. When the gene to be mutated encodes a triangular fuzzy set, both *left* and *right* fields are mutated with the same value to preserve the original type.

Then, given a gene's field $g^v_{i,e}$ to be mutated, a variation interval around its value is defined with the aim of avoiding lack of distinguishability—i.e., two fuzzy sets very close—and other kinds of deformities that would decrease the interpretability degree. In this way, the following equations are used—Fig. 3(a) shows an example:

$$min_{g^v_{i,e}} = \begin{cases} min_v & \text{if } i = 1 \\ g^v_{i-1,right} + \delta & \text{if } e = \textit{left} \text{ or } g^v_{i,left} = g^v_{i,right} \\ g^v_{i,left} & \text{if } e = \textit{right} \text{ and } g^v_{i,left} \neq g^v_{i,right} \end{cases} \quad (2)$$

$$max_{g^v_{i,e}} = \begin{cases} max_v & \text{if } i = l_v \\ g^v_{i+1,left} - \delta & \text{if } e = \textit{right} \text{ or } g^v_{i,left} = g^v_{i,right} \\ g^v_{i,right} & \text{if } e = \textit{left} \text{ and } g^v_{i,left} \neq g^v_{i,right} \end{cases} \quad (3)$$

with $\delta = 1/(2 \cdot l_v)$ being the allowed minimum distance between the extremes of the cores of the fuzzy sets for the sake of a good distinguishability degree and l_v the number of linguistic terms used in the variable v.

Finally, an original real-coding mutation operator is applied on $g^v_{i,e}$ constrained to the interval $[min_{g^v_{i,e}}, max_{g^v_{i,e}}]$. We have named this operator as the Constrained Asymmetric Mutation (CAM), which is described in Algorithm 2. Figure 3(b) shows an example of the resulting asymmetric probability density function used to mutate a value.

ISBN: 978-989-95079-6-8

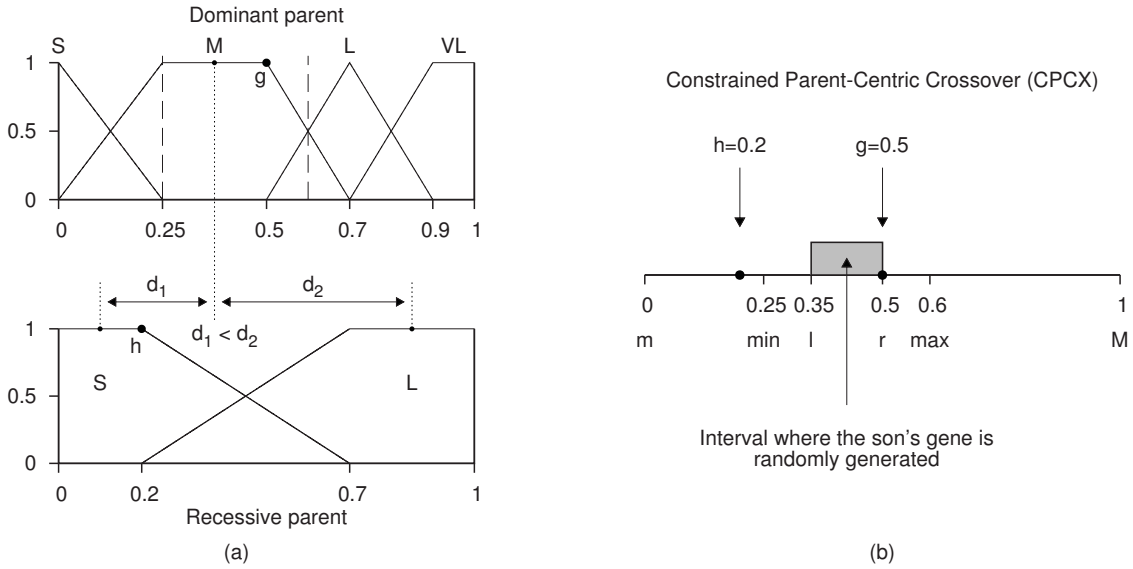


Figure 2: Example of crossover operation: (a) Given the gene g (which encodes the right extreme of the core of the fuzzy set labeled as S_M) of the dominant parent to be crossed, the allowed interval variation $[0.25, 0.6]$ is calculated (dashed lines) and the gene h that encodes the right extreme of the closest fuzzy set in the recessive parent (VS_S_M) is selected; (b) CPCX operator applied on the dominant gene $g = 0.5$, the recessive gene $h = 0.2$, the allowed interval variation $[min, max] = [0.25, 0.6]$, and the definition interval $[m, M] = [0, 1]$; a random value is generated in the interval $[l, r] = [0.35, 0.5]$ ($\alpha = 0.6$)

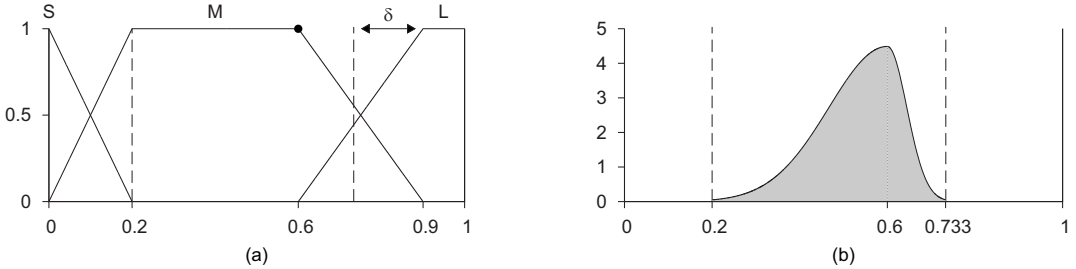


Figure 3: Example of parameter mutation operation: (a) the right extreme of the core of the fuzzy set M is selected to be mutated; then, the interval variation is defined by eq. 2 and 3 with $\delta = 0.167$ since there are three linguistic terms ($l_v = 3$); (b) asymmetric probability density function used in the CAM operator to mutate the right extreme of M's core constrained to the given interval $[0.2, 0.733]$

2.5 Granularity Mutation

The granularity mutation changes the number of linguistic terms in both input and output variables. To do that, it can *fuse* (merge) two neighboring fuzzy sets (thus decreasing in one the number of linguistic terms of the corresponding variable) or *fission* (split) a fuzzy set in two parts (thus increasing in one the number of linguistic terms of the corresponding variable). When fusion is applied on a variable with two linguistic terms, it involves removing the variable from the fuzzy model.

There are some constraints to apply fusion and fission. On the one hand, if fusion is applied on an input variable, this variable must have at least two linguistic terms and there must be more than one input variable with at least two linguistic terms (in order to avoid a fuzzy system without input variables after fusion). When fusion is applied on an output variable, this variable must have at least three linguistic terms since output variable removal is not allowed.

On the other hand, fission on variables with the maximum number of linguistic terms is not allowed. A second constraint is considered to avoid lack of distinguishability after fission. Indeed, if the core width of the fuzzy set associated to the gene to be fissioned is lower than a value inversely proportional to the number of linguistic terms of the corresponding variable, the core of the two fuzzy sets resulting from fission will be too close and, therefore, fission is not allowed. It involves that triangular fuzzy sets can not be fissioned since its core width is zero.

The two following subsections explain in detail how fusion and fission operate.

2.5.1 Fusion (merge) operation

Given a gene g_i^v to be fused, let h be the closest gene to g_i^v:

$$h = \begin{cases} g_{i-1}^v & \text{if } i = l_v \\ g_{i+1}^v & \text{if } i = 1 \\ g_{i-1}^v & \text{if } d(g_i^v, g_{i-1}^v) \leq d(g_i^v, g_{i+1}^v) \\ g_{i+1}^v & \text{otherwise} \end{cases} \quad (4)$$

with

$$d(g, h) = |(g_{left} + g_{right})/2 - (h_{left} + h_{right})/2| \quad (5)$$

To fuse g_i^v and h, the following steps are done:

- Firstly, if $g_{i,left}^v < h_{left}$ then $g_{i,right}^v \leftarrow h_{right}$; otherwise, $g_{i,left}^v \leftarrow h_{left}$.

ISBN: 978-989-95079-6-8

Algorithm 2: Constrained Asymmetric Mutation (CAM) operator

Input: (x, min, max) with x being the value to be mutated and $[min, max]$ its variation interval
Output: Mutated value x'
begin
  $p_{left} \leftarrow (x - min)/(max - min)$
  if $U[0,1] < p_{left}$ **then**
    $\sigma \leftarrow (x - min)/3$
    $x' \leftarrow N[x, \sigma]$ */* Normal random */*
    if $x' > (2x - min)$ **then** $x' \leftarrow 2x - min$
    else if $x' < min$ **then** $x' \leftarrow min$
    if $x' > x$ **then** $x' \leftarrow x - (x' - x)$
  else
    $\sigma \leftarrow (max - x)/3$
    $x' \leftarrow N[x, \sigma]$ */* Normal random */*
    if $x' > max$ **then** $x' \leftarrow max$
    else if $x' < (2x - max)$ **then** $x' \leftarrow 2x - max$
    if $x' < x$ **then** $x' \leftarrow x + (x - x')$
end

- Secondly, gene h is removed from the chromosome and the number of linguistic terms is decreased by one ($l_v \leftarrow l_v - 1$).

2.5.2 Fission (split) operation

The fission operation splits a trapezoidal fuzzy set into two triangular ones. Thus, given a gene g_i^v to be fissioned, the following steps are done:

- A new gene, h, is inserted at the right of g_i^v (h will be the new g_{i+1}^v) with $h_{left} \leftarrow g_{i,right}^v$ and $h_{right} \leftarrow g_{i,right}^v$. Set $g_{i,right}^v \leftarrow g_{i,left}^v$.
- The number of linguistic terms is increased by one ($l_v \leftarrow l_v + 1$).

2.6 Embedded Genetic Learning Approach

Every time a new individual (which encodes the fuzzy partition of each variable) is evaluated, a fuzzy rule set is firstly derived and then the complete fuzzy model (the fuzzy partitions plus the fuzzy rule set) is analyzed. To design the fuzzy rule set an efficient *ad hoc* data-driven method, the well-known Wang-Mendel (WM) algorithm [9] in this paper, is used. Therefore, a wrapper-based embedded genetic learning process is followed [4].

2.7 Inference Mechanism

We consider FITA (first inference, then aggregate) approach, the Max-Min inference scheme (i.e., T-conorm of maximum as aggregation and T-norm of minimum as relational operator), the T-norm of minimum as conjunction, and center-of-gravity as defuzzification. Moreover, the mean of the output domain is returned when no rules are fired for the given input data (this fact may only occur with test data since the algorithm ensures complete fuzzy rule set regarding training data).

2.8 Objective Functions

The multi-objective optimization performed in EGLFP is based on three objective functions to be minimized: O_1 assesses the error of the system, O_2 the complexity of the derived fuzzy rule set, and O_3 the complexity of the learnt fuzzy partitions. Thus, O_1 is focused on improving the accuracy of the fuzzy model while O_2 and O_3 improve its interpretability.

2.8.1 O_1, accuracy objective

The root mean squared error (RMSE) is used to compute the accuracy of the learnt fuzzy model S:

$$O_1(S) = RMSE(S) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(F_S(\mathbf{x}_k) - \mathbf{y}_k)^2} \quad (6)$$

2.8.2 O_2, rule set complexity objective

This second objective simply involves the final number of fuzzy rules obtained after applying the WM algorithm on the decoded fuzzy partitions as explained in Sect. 2.6:

$$O_2(S) = r(S) \quad (7)$$

2.8.3 O_3, fuzzy partition complexity objective

The third objective is the sum of the linguistic terms used in input and output variables:

$$O_3(S) = \sum_{v=1}^{n+m} L_v \quad \text{with} \quad L_v = \begin{cases} l_v & \text{if } l_v > 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

The use of L_v is to not computing the removed input variables, where the number of linguistic terms is $l_v = 1$.

Since the fuzzy partition complexity has a direct influence on the fuzzy rule set complexity, it is expected that a low value of O_3 will imply a low value of O_2. Therefore, O_2 and O_3 are correlated in some degree. To improve the interpretability, O_2 seems to be more influent than O_3. However, we have decided to also use the objective O_3 with the aim of polishing good solutions by reducing the number of linguistic terms as much as possible.

This correlation will involve that, in the practice, the Pareto front shape with three objectives will be similar to the one with the two former ones and therefore, the inclusion of the third objective will not significantly degrade the multi-objective optimization quality.

2.9 Multi-objective Approach

A generational approach with the multi-objective elitist replacement strategy of NSGA-II [7] is used. Crowding distance in the objective function space is considered. Binary tournament selection based on the nondomination rank (or the crowding distance when both solutions belong to the same front) is applied. Instead fixing initial intervals for the objectives (which is not easy with the considered objectives) to compute the crowding distance, it is normalized for each objective according to the extreme values of the solutions contained in the analyzed front.

3 Experimental Analysis

This section includes the obtained results of the proposed EGLFP algorithm in six real-world regression problems where all the input and output variables are real-valued, and compares them with other fuzzy model learning methods.

ISBN: 978-989-95079-6-8

3.1 Problems and Learning Methods

We have considered the following regression problems: *Ele1* (the data set and partitions used in this paper are available at the author's website [1]), *Laser* (available at KEEL website [2]), *Ele2* [1], *DEE* [2], *Concrete* (obtained from the UCI Repository [3]), and *Comp-activ* (obtained from L. Torgo's website[4]).

Table 1 collects the main features of each problem, where *#InputVar* stands for the number of input variables, *#Exam* for the total number of examples, and $maxLV_v$ for the initial number of uniformly distributed triangular-shaped linguistic terms considered for each input variable in the experimental analysis. The output variable is always initially provided with seven uniformly distributed triangular-shaped linguistic terms.

Table 1: Data sets considered in the experimental analysis

Problem	#InputVar	#Exam	maxLV$_v$
Ele1	2	495	7
Laser	4	993	5
Ele2	4	1066	5
DEE	6	365	5
Concrete	8	1030	5
Comp-activ	21	8192	3

The experiments shown in this paper have been performed with a 5-fold cross validation. Thus, the data set is divided into five subsets of (approximately) equal size. The algorithm is then applied five times to each problem, each time leaving out one of the subsets from training, but using only the omitted subset to compute the test error.

We have considered two learning methods for comparison (both of them use the same inference engine described in Sect. 2.7 for our proposal). The *Wang and Mendel (WM)* [9] algorithm is a simple learning method that, in spite of not obtaining accurate results, is a traditional reference in the research community; the algorithm has been implemented by us. The *Cordón, Herrera, and Villar (CHV)* [4] algorithm is a competitive genetic fuzzy system that provides a learning flexibility similar to our EGLFP algorithm since it learns both the number of linguistic terms per variable and the membership function parameters. As in our case, this genetic learning process is also performed by wrapping the WM algorithm. We have used the original algorithm implementation provided by the authors.

Our algorithm has been run with the following parameter values: 1000 iterations, $p = 30$ as population size, $p_c = 0.7$ as crossover probability, and $p_{pm} = p_{gm} = 0.2$ as parameter and granularity mutation probability per chromosome, respectively. We have not performed any previous analysis to fix these values, so better results may probably be obtained by tuning them though we have informally noticed our algorithm is not specially sensitive to any parameter. The same parameter values are also used in the CHV algorithm in addition to $a = 0.35$ (for MMA crossover), $b = 5$ (for non-uniform mutation), and $\alpha = 0.2$ (weight of the number of rules in the fitness function).

3.2 Analysis

Table 2 collects the obtained results according to different quality measures such as approximation error, number of rules, number of linguistic terms, number of input variables, and fitness values. Average values of the five data partitions for each problem are reported.

We have included two versions of our algorithm in the comparative. The first one (EGLFP-1) is guided by a single fitness function in the same way as done in CHV [4] for the sake of a fair comparison. This function is a normalized aggregation of the error and the number of rules as follows:

$$f(S) = 0.8 \cdot MSE(S) + 0.2 \cdot \frac{MSE(S_{WM})}{r(S_{WM})} \cdot r(S) \quad (9)$$

with S_{WM} being the fuzzy model generated by the WM algorithm using uniformly-distributed triangular-shaped strong fuzzy partitions with the maximum number of linguistic terms allowed per variable ($maxLV_v$) and $MSE(S) = RMSE(S)^2$.

The second version (EGLFP-3) is guided by the above three-objectives multicriteria approach. Since in this case several solutions are returned in each run, we show five representative solutions from the final Pareto-optimal set. They are computed by sorting in ascending order the solutions according to the training RMSE and getting the 1 (i.e. best error), 25, 50 (i.e. mean), 75, and 100 (i.e. worst error) percentiles.

From the obtained results we can observe that our algorithm EGLFP-1 clearly overcomes, according to the fitness values (column Fit.), the CHV algorithm in all the considered problems. Indeed, our method generates more accurate fuzzy models in five cases. As regards the interpretability, it is not clear between EGLFP-1 and CHV which method obtains simpler models in terms of number of rules and number of linguistic terms since it depends on the problem. However, our method generates fuzzy models with more easily distinguishable fuzzy partitions thanks to the constraints tackled by the proposed genetic operators. Furthermore, CHV needs, in order to achieve the obtained accuracy values, to use badly formed fuzzy partitions where the extreme fuzzy sets of every variable may not hold normality within the corresponding domain.

If we analyze the results obtained by EGLFP-3 we can see that the process is able to generate a wide range of solutions with different accuracy-interpretability tradeoffs. It worths noticing that the most accurate solution of EGLFP-3 overcomes EGLFP-1 in both accuracy and interpretability in several problems. This fact is due to the niche-based search process caused by the use of the multi-objective approach, which leads the algorithm to a better exploration of the search space. It is also interesting to highlight the significant improvements in interpretability compared with the solutions provided by the WM algorithm. Very simple fuzzy models with a low number of variables, linguistic terms, and fuzzy rules are obtained while preserving a good accuracy.

4 Conclusion and Further Work

We have proposed a competitive genetic algorithm to simultaneously learn many components of a fuzzy model such as the

[1] J. Casillas. FMLib: fuzzy modeling library. http://decsai.ugr.es/~casillas/FMLib/

[2] KEEL: Knowledge extraction based on evolutionary learning. http://www.keel.es

[3] UC Irvine Machine Learning Repository. http://archive.ics.uci.edu/ml/

[4] L. Torgo. Collection of regression datasets. http://www.liacc.up.pt/~ltorgo/Regression/DataSets.html

ISBN: 978-989-95079-6-8

Table 2: Obtained results with $\mathbf{E}_{tra/tst}$ standing for the RMSE on the training/test data set (O_1, eq. 6), **#R** the number of fuzzy rules (O_2, eq. 7), **#L** the sum of the number of linguistic terms of each variable (O_3, eq. 8), **#I** the number of input variables, and **Fit.** the fitness (eq. 9)

Method	$\mathbf{E}_{tra}$	$\mathbf{E}_{tst}$	#R	#L	#I	Fit.
Ele1						
WM	650.743	674.910	22.0	21.0	2.0	
CHV	591.268	627.132	6.8	15.8	2.0	305924
EGLFP-1	558.793	618.857	12.8	15.0	2.0	298660
	560.346	638.527	17.0	16.0	2.0	315981
	577.568	632.652	11.1	12.9	2.0	
EGLFP-3	588.422	618.726	7.7	10.5	2.0	
	643.637	651.702	4.2	6.5	2.0	
	1017.088	988.744	1.8	4.0	1.0	
Laser						
WM	15.994	16.521	58.4	27.0	4.0	
CHV	8.791	10.602	58.8	25.0	4.0	97.694
EGLFP-1	8.509	9.750	33.6	18.6	4.0	87.611
	7.955	9.050	30.2	18.8	4.0	77.168
	9.570	11.099	17.8	14.8	3.8	
EGLFP-3	12.322	13.273	10.7	11.2	3.2	
	22.477	24.230	5.2	8.1	2.4	
	29.411	29.536	2.0	4.0	1.0	
Ele2						
WM	312.446	314.944	65.0	27.0	4.0	
CHV	235.053	238.263	22.6	20.4	4.0	50745
EGLFP-1	185.383	187.804	24.2	18.4	3.2	35371
	183.883	193.290	17.6	17.4	3.0	32462
	211.290	212.999	12.7	14.4	3.0	
EGLFP-3	314.150	314.470	6.9	10.6	3.0	
	553.541	566.163	3.7	8.1	2.0	
	903.109	917.182	2.0	4.0	1.0	
DEE						
WM	0.36302	0.47091	178.4	37.0	6.0	
CHV	0.34580	0.42109	112.0	32.8	0.0	0.11226
EGLFP-1	0.31916	0.41802	142.2	33.0	6.0	0.10254
	0.31795	0.43921	154.8	33.8	6.0	0.10376
	0.34773	0.43448	105.5	28.1	6.0	
EGLFP-3	0.37248	0.43070	50.5	21.4	5.2	
	0.41404	0.43448	13.3	13.5	3.5	
	0.79297	0.79075	2.0	4.0	1.0	
Concrete						
WM	8.2581	9.6227	309.8	47.0	8.0	
CHV	5.5686	7.4688	335.8	43.6	8.0	39.058
EGLFP-1	5.8652	7.1887	235.0	36.0	8.0	37.614
	5.6360	7.3113	324.8	41.2	8.0	39.286
	6.3962	7.1763	210.0	30.8	7.1	
EGLFP-3	7.2349	7.8924	91.4	22.2	5.8	
	8.9515	8.8944	16.8	12.3	3.9	
	14.9830	15.0084	2.0	4.0	1.0	
Comp-activ						
WM	11.9387	11.9762	425.6	70.0	21.0	
CHV	3.6004	3.6022	103.6	51.0	21.0	17.2945
EGLFP-1	3.3237	3.3848	14.0	19.4	6.4	9.7949
	3.3506	3.4277	12.6	17.6	6.6	9.8424
	3.4751	3.5417	8.6	14.1	5.3	
EGLFP-3	4.4530	4.5192	6.3	10.5	3.7	
	6.6188	6.6736	4.0	8.4	2.8	
	18.3997	18.3876	1.8	4.0	1.0	

subset of input variables, the number of linguistic terms per variable, the type of membership functions (trapezoidal or triangular), the membership function parameter values, and the fuzzy rule set. The algorithm is designed to generate highly legible and compact fuzzy partitions thanks to the use of interpretability constraints and powerful genetic operators.

The originality and good performance of the proposal mainly lies in the definition of new real-coding crossover and mutation operators that properly deals with the constraints for distinguishability, orderliness, and non-deformity imposed to the learnt fuzzy partitions; the design of a variable-length coding scheme and a parent-centric crossover capable of handling fuzzy partitions with different number of linguistic terms; and the use of several criteria and a multi-objective optimization scheme that provide with a range of different interpretability-accuracy tradeoffs and endow the algorithm with an effective niche-based search process. Successful results are obtained in six real-world regression problems with up to 21 continuous input variables.

As further work we will investigate the scalability of the algorithm to large-scale problems and the use of more criteria to assess the interpretability quality of the obtained fuzzy models.

Acknowledgment

This work was supported in part by the Spanish Ministry of Science and Innovation (grant no. TIN2008-06681-C06-01) and the Andalusian Government (grant no. P07-TIC-3185).

References

[1] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Selecting fuzzy if-then rules for classification problems using genetic algorithms. *IEEE Transactions on Fuzzy Systems*, 3(3):260–270, 1995.

[2] T.-P. Hong and J.-B. Chen. Finding relevant attributes and membership functions. *Fuzzy Sets and Systems*, 103(3):389–404, 1999.

[3] A. González and R. Pérez. Completeness and consistency conditions for learning fuzzy rules. *Fuzzy Sets and Systems*, 96(1):37–51, 1998.

[4] O. Cordón, F. Herrera, and P. Villar. Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *IEEE Transactions on Fuzzy Systems*, 9(4):667–674, 2001.

[5] J. Espinosa and J. Vandewalle. Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm. *IEEE Transactions on Fuzzy Systems*, 8(5):591–600, 2000.

[6] J.-N. Choi, S.-K. Oh, and W. Pedrycz. Identification of fuzzy models using a successive tuning method with a variant identification ratio. *Fuzzy Sets and Systems*, 159(21):2873–2889, 2008.

[7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarevian. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[8] C. García-Martínez, M. Lozano, F. Herrera, D. Molina, and A. M. Sánchez. Global and local real-coded genetic algorithms based on parent-centric crossover operators. *European Journal of Operational Research*, 185:1088–1113, 2008.

[9] L.-X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427, 1992.

ISBN: 978-989-95079-6-8

Non Additive Measures for Group Multi Attribute Decision Models

Marta Cardin, Silvio Giove

Dept. of Applied Mathematics, University of Venice
Venice, Italy
Email: mcardin@unive.it, sgiove@unive.it

Abstract— *This paper extends the Choquet integral, widely used in multi-attribute decision problems, to the non monotone case in the context of Group Decision Theory. Even if not so often, preference structures which violate the monotonicity axiom can be observed in real applications. Our aim is twofold. First, we propose the Choquet integral with non monotone non additive measure. Then, we apply the Choquet integral in the context of multi person decision problem, a typical framework of many real world applications, for which the Choquet integral was rarely proposed. Thus in our model this aggregation function is applied twice, both in the cases with possible negative interactions. For this reason, our proposal can be defined as two-step signed Choquet integral.*

Keywords— Non additive measures, Choquet integral, Group Decision Theory.

1 Introduction

In Multi Attribute decision models, non additive measures and the Choquet integral have been intensively applied in many real world decision problems. Despite simpler approaches, like simple additive weighting (SAW), non additive measures can model *interactions* between the criteria. This method is a very general aggregation tool, including as special cases many aggregation algorithms, as SAW, min and max operators, k-statistics, OWA, and many others. Nevertheless, monotonicity is usually considered a strict requirement. Only few papers explicitly considered the non monotone case. Even if rarer than for *preferential independence* axiom, some violations of the monotonicity axiom can be observed in the Decision Maker preference structure. Moreover, even if many real world applications exist so far, little results were obtained in the context of Group Decision Theory. This family of problem deals with complex decision about some alternatives that need to be scored by a tool of Experts, see [4] and the references therein. To simplify, suppose that an Experts Committee is demanded to evaluate a finite set of development strategies on the basis of some criteria. Non additive measures can help to solve this type of *multi person - multi attribute* decision problems.
Negative(non monotone) interactions among the criteria can be modeled by means of *signed* measures.
The aim of the paper is twofold; first we consider the possibility of *negative* interactions among the criteria, and then we apply a non additive aggregation algorithm in the context of Group Decision Theory. In this sense, we apply twice the Choquet integral, to aggregate the Expert's opinion for each alternative, and subsequently to aggregate the Expert alternative scores into a global numerical evaluation (*Group alternative score*). The latter is usually performed by a Decision Maker, which, on the basis of his confidence about the Committee members experience and/or capacity, averages their judgements not necessarily in a linear way. Thus we can define our approach as a two-step Choquet integral with *signed* measures, i.e. *two-step signed Choquet integral*. The paper is organized as follows. The next Section briefly resumes some definitions and properties of non additive measures and of the Choquet integral. Section 3 describes our proposed model, which is deeply analyzed in the following Section 4. Section 5 reports the application of the two-step signed Choquet integral to multi person - multi attribute decision problems, and finally in the last section a numerical example is reported.

2 The discrete Choquet integral

Aggregation has for purpose the simultaneous use of different pieces of information provided by several sources, in order to come to a conclusion or a decision. So aggregation functions transform a finite number of inputs, called arguments, into a single output. They are applied in many different domains and, in particular, aggregation functions play an important role in different approaches to decision making, where values to be aggregated are typically preferences or satisfaction degrees. Many functions of different type have been considered in connection with different situations and various properties of these functionals can be imposed by the nature of the considered aggregation problem. A class of aggregation operators can also be introduced axiomatically by means of a set of properties.
We denote by E a non empty real interval. If the integer n represents the number of values to be aggregated an aggregation operator is a function $A : E^n \to E$. To motivate the use of the Choquet integral as an aggregation operator, we present some basic mathematical properties of the aggregation functions.

- **Monotonicity** For all $\mathbf{x}, \mathbf{y} \in E^n$ if $x_i \leq y_i (i = 1, \ldots, n)$ then $A(\mathbf{x}) \leq A(\mathbf{y})$
- **Positive Homogeneity** If $\mathbf{x} \in E^n$ and $a \in \mathbb{R}, a > 0$ then $A(a\mathbf{x}) = aA\mathbf{x}$

Moreover we define x_{-i} the element of $\mathbb{R}^{n-1}$ that is obtained from x by eliminating component i, and we denote (x_{-i}, y_i) as obtained from x by replacing x_i with y_i.
Now we present the concept of comonotonicity.

Definition 2.1 *If x, y are elements of $\mathbb{R}^n$ then x, y are said* comonotonic *if $x_i < x_j$ implies that $y_i \leq y_j$.*

Two vectors x, y are comonotonic if they have the same ranking of their components or there exists a permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that $x_{\sigma(i)}$ and the corresponding $y_{\sigma(i)}$ are arranged in nondecreasing order.

ISBN: 978-989-95079-6-8

Our representation result depends on the following axioms related to the concept of comonotonicity.

- **Comonotonic Monotonicity** If $\mathbf{x}, \mathbf{y} \in E^n$ are comonotonic and $x_i \leq y_i$ $(i = 1, \ldots, n)$ then $A(\mathbf{x}) \leq A(\mathbf{y})$
- **Comonotonic Separability** If $\mathbf{x}, \mathbf{y} \in E^n$ are comonotonic then for every i, $A(x_{-i}, x_i) \geq A(y_{-i}, x_i)$ iff $A(x_{-i}, y_i) \geq A(y_{-i}, y_i)$

The comonotonic separability axiom is obviously a variation of an additive separability axiom and it has been applied successfully in decision making under risk and uncertainty, see [5]and the references therein. It states that preferences between alternatives depend only on the components that differ between the vectors under cons ideration, as long as these alternatives maintain the attributes' ordering.
In order to introduce a non-additive approach to aggregation operators we propose a non-additive integral operator and so we consider the integral as a particular averaging operator. The use of variants of the Choquet integral allows some flexibility in the way criteria are combined. Non additive measures are in the current literature a commonly used method to aggregate numerical information. This is particularly due to the fact that in cooperation with integral aggregation functions, they are a well-founded framework able to capture interactions among the involved variables. Usually the monotonicity property is required in most of practical applications. Monotone measures can model both *synergic* and *redundance* interactions among the criteria. Conversely, they are unable to represent the *neutralization* effect. Such a situation appears, for instance, where increasing one criterion by alone has a positive effect, but the contemporary increase of two criteria has a negative effect. Of course, such a situation cannot be represented by monotone measures.
In particular in this paper we consider a non-monotone Choquet integral as in [5], [8] and [12] and we define a non-monotone Choquet measure and a Choquet integral for a n-dimensional vector.
If N is a finite index set $N = \{1, \ldots, n\}$, a real valued set function $\upsilon : 2^N \to \mathbb{R}$, with $\upsilon(\emptyset) = 0$, is called a *non additive signed measure*. If $A \subseteq B \subseteq N$ implies that $\upsilon(A) \leq \upsilon(B)$, then the function is said to be *monotone* and υ is called a *non-additive measure*. If υ is a non-additive measure then $\upsilon(A) \geq 0$ for all $A \in 2^N$. We say that a measure is *additive* if $\upsilon(A \cup B) = \upsilon(A) + \upsilon(B)$.
We note that if $S \subseteq N$, $\upsilon(S)$ can be viewed as the *importance* of the set of elements S. We introduce now the discrete Choquet integral on N viewed as an aggregation function that generalizes the weighted arithmetic mean.

Definition 2.2 *The* Choquet integral C_υ *of a vector* $\boldsymbol{x} \in \mathbb{R}^n$ *with respect to a non-additive signed measure* υ *is*

$$C_\upsilon(\boldsymbol{x}) := \int_0^{+\infty} \upsilon(\boldsymbol{x} \geq \alpha)\, d\alpha + \int_{-\infty}^{0} (\upsilon(\boldsymbol{x} \geq \alpha) - \upsilon(N))\, d\alpha$$

where $\upsilon(\boldsymbol{x} \geq \alpha) = \upsilon(\{i \in N : x(i) \geq \alpha\})$.

This formula can be interpreted as an expectation operator with respect to a generalized measure. Then, the Choquet integral of a vector $\mathbf{x} \in \mathbb{R}^n$ with respect to a non-additive measure υ can be represented as the following weighted sum:

$$C_\upsilon(\mathbf{x}) = \sum_{1=1}^{n} (x_{(i)} - x_{(i-1)}) \upsilon(A_{(i)}) \tag{1}$$

and $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is such that $x_{\sigma(i)}$ are arranged in nondecreasing order and $A_{(i)} = \{i, \ldots n\}$ and $A_{(n+1)} = \emptyset$. The Choquet integral can be also computed using the Möbius representation of υ, see [8].

Proposition 2.1 *The Choquet integral* $C_\upsilon : \mathbb{R}^n \to \mathbb{R}$ *can be written as*

$$C_\upsilon(\boldsymbol{x}) = \sum_{T \subseteq N} a(T) \bigwedge_{i \in T} x_i, \qquad \boldsymbol{x} \in \mathbb{R}^n$$

where a *is the Möbius representation of* υ.

The Choquet integral has important properties for aggregation.

Proposition 2.2 *Let* A *be an aggregation function defined on* E^n.

i) If A *is a non-monotone Choquet integral then it is positive homogeneous, comonotonic monotone and comonotonic additive.*

ii) A *is a non-monotone Choquet integral if it satisfies positive homogeneity and comonotonic separability.*

iii) A *is a non-monotone Choquet integral if it is continuous and comonotonic additive.*

Proof The proof of part i) is immediate by the definition of Choquet integral.
Now it is important to note that a comonotonic separable aggregation function A is comonotonic additive that is if $\mathbf{x}, \mathbf{y} \in E^n$ are comonotonic then $A(\mathbf{x} + \mathbf{y}) = A(\mathbf{x}) + A(\mathbf{y})$. In fact if $\mathbf{x}, \mathbf{y}, \mathbf{z} \in E^n$ are comonotonic and $A(\mathbf{x}) \geq A(\mathbf{y})$ then by comonotonic separability $A(\mathbf{x} + \mathbf{z}) \geq A(\mathbf{y} + \mathbf{z})$. Then if we consider two comonotonic vectors $\mathbf{x}, \mathbf{y} \in E^n$ there exist $c, d \in E$ such that $A(\mathbf{x}) = \mathbf{c} = (c, \ldots, c)$ and $A(\mathbf{y}) = \mathbf{d} = (d, \ldots, d)$. By the fact that each comonotonic vector is comonotonic with each other vector with equal component we can prove $A(\mathbf{x} + \mathbf{y}) = A(\mathbf{c} + \mathbf{y}) = A(\mathbf{c} + \mathbf{d}) = A(\mathbf{c}) + A(\mathbf{d}) = A(\mathbf{x}) + A(\mathbf{y})$. Now we can conclude since a homogeneous function that satisfies comonotonic additivity is a non-monotone Choquet integral by theorem 1 of [8] and a continuous function that satisfies comonotonic additivity is a non-monotone Choquet integral by corollary 2.2 of [14].

It is elementary verified that if A satisfies the hypothesis of proposition 1.1 and it is also monotone then it is a monotone Choquet integral.

3 The model

We consider a decision problem where $K = \{1, \ldots, k\}$ Experts have to score M alternatives on the basis of some criteria, each of them normalized in the common scale [0,1]. The preference structure of the k-th Expert, $(1 \leq k \leq K)$, is represented by a set of non additive measures m_k, defined on the space of criteria. Despite the majority of Multi Criteria Group decision models, we admit that each Expert is characterized

ISBN: 978-989-95079-6-8

by his own criteria set, even if the sets can partially overlap or coincide as a particular case. Moreover, the preference structure of each Expert will be represented by a *signed* measure, including the possibility of *neutralization* effects among the criteria themselves. Finally, at higher level, an other measure u is used to aggregate the individual opinions into a global score. Assuming that a Decision Maker (DM) has to score the alternatives on the basis of the judgements expressed by the Experts, the measure u can be interpreted as the degree of confidence that the DM feels towards the Experts themselves, or towards each coalition of Experts. Since the Choquet integral is twice applied, we deal with a *two-step* Choquet integral [11][1]. For a better problem formulation, let us consider the following definitions.
i) The criteria set for the k-th Expert E_k is the set $N_i = \{c_{i_1}, \ldots, c_{i_n}\}$, with $n_i = |N_i|$ is the cardinality of N_i. It follows that $\mathbf{x}^i = (x_{i_1}, \ldots, x_{i_n})$ is the *profile* of criteria associated with an alternative by the Expert E_k and an alternative is characterized by a vector $\mathbf{x}$ in $\mathbb{R}^n$ where $n = \sum_{i=1}^{K} n_i$.

ii) The preference structure of the k-th Expert is represented by a non additive signed measure m_k, defined on 2^{N_k}.

iii) There are M alternatives, each of them characterized by the vector $\mathbf{x}_j$ in $\mathbb{R}^n$, $j = 1, .., M$.

Let $v_k(\mathbf{x}) = \int \mathbf{x}^k \, d\, m_k$ for $1 \leq k \leq K$ and $v(\mathbf{x}) = (v_1(\mathbf{x}), \ldots, v_K(\mathbf{x}))$.
Then we define the two-step Choquet integral $C(\mathbf{x})$,

$$C(\mathbf{x}) = \int v(\mathbf{x}) \, d\, u.$$

4 Multi-step non monotone Choquet integral

In this section we give some basic definitions and we present some results on multi-step non monotone Choquet integral. The two-step (monotone) Choquet integral has been investigated mainly in [13] and [11]. Let us now give a formal definition of a multi-step non-monotone Choquet integral.

Definition 4.1 *Let $\Gamma \subseteq \mathbb{R}^n$. For any $i \in N$, the projection $\mathbf{x} \mapsto x_i$ is a* 0-step non-monotone Choquet integral. *Let us consider $F_i : \Gamma \to \mathbb{R}$, $i \in M := \{1, \ldots, m\}$, being k_i-step non-monotone Choquet integrals, and a non-additive signed measure v on M. Then*

$$F(\mathbf{x}) := C_v(F_1(\mathbf{x}), \ldots, F_m(\mathbf{x}))$$

is a k-step non-monotone Choquet integral*, with $k := \max\{k_1, \ldots, k_m\} + 1$. A* multi-step non-monotone Choquet integral *is a k-step non-monotone Choquet integral for some integer $k > 1$.*

Let us first recall the concept of piecewise linear function.

Definition 4.2 *A real-valued function F on a convex closed subset $\Gamma \subseteq \mathbb{R}^n$ is* piecewise linear *if Γ can be written as a union of closed subspaces $\Gamma_1, \ldots, \Gamma_q$ of the same dimension as Γ, such that $F|_{\Gamma_i}$ is linear, $i = 1, \ldots, q$. A linear function G on $\mathbb{R}^n$ which coincides with F on some Γ_i is a* component *of F.*

[1] Anywise, in the quoted reference the two-step Choquet integral was limited to the monotone case only.

The following proposition gives the basic properties of non-monotone multi-step Choquet integrals.

Proposition 4.1 *The non-monotone multi-step Choquet integral is a continuous, positively homogeneous and piecewise linear function.*

Proof The non monotone Choquet integral is a piecewise linear and positively homogeneous function, then also the multi-step non monotone Choquet integral. Clearly, any piecewise linear function is continuous.

Recall that the multi-step Choquet integral is not comonotonic additive, in general, and hence it cannot be described by a 1-step Choquet integral.

Proposition 4.2 *If F is a 2-step non-monotone Choquet integral and the measure of the second level is additive then F coincides with a 1-step non-monotone Choquet integral.*

Proof If we consider two comonotonic vectors $\mathbf{x}, \mathbf{y}$ then for every $i = 1, \ldots, m$ $F_i(\mathbf{x} + \mathbf{y}) = F_i(\mathbf{x}) + F_i(\mathbf{y})$ since F_i is comonotonic additive. Then it is easy to prove that $F(\mathbf{x}+\mathbf{y}) = F(\mathbf{x}) + F(\mathbf{y})$ where $F(\mathbf{x}) = C_v(F_1(\mathbf{x}), \ldots, F_m(\mathbf{x}))$ and C_v is a linear functional. Then F is a positively homogeneous and comonotonic additive functional. Now the positively homogeneous function F that satisfies comonotonic additivity is a non-monotone Choquet integral by theorem 1 of [8].

5 A Multi Criteria Group Decision problem

In the proposed model, the two-step Choquet integral will be used to compute the score of any alternative, proceeding in two sequential phases. Firstly, the individual measures m_k defined on 2^{N_k} are used to compute the individual score. Subsequently all the individual scores are aggregated with the measure u on the space 2^K. Both in the cases signed measure can be admitted, modeling negative interactions.
The measure u represents the preference structure of the DM about the Expert's coalitions. For instance, consider a rule like the following one

If the first and the second Experts score equally an alternative, then score this alternative with their common value, despite any other opinion of the remaining Experts of the group

This rule can be implemented assigning one to the coalition formed by the Experts n. $1, 2$, that is $m\{X\} = 1$ if $A \cap X \neq \emptyset$ with $A = \{1, 2\}$.
The steps of the aggregation algorithms are then the following ones

i) $\forall k$, $1 \leq i \leq k$, aggregate each of the individual measure using the Choquet integral: $v_k(\mathbf{x}) = \int \mathbf{x}^k \, d\, m_k$, obtaining the individual score for the alternative ($\mathbf{x}$)

ii) at the top level (the DM level), aggregate the individual ranking, obtained at the previous step, using the measure m on the space 2^K applying again the Choquet integral: $C(\mathbf{x}) = \int v(\mathbf{x}) \, d\, u$.

ISBN: 978-989-95079-6-8

Let us recall that the *relative importance* for monotone measure of a single criterion is usually measured by a suitable index like the Shapley index, [6]. The Shapley index for the i-th criterion and for the k-th Expert is given by

$$p_k(i) = \sum_{T \subseteq N \setminus \{i\}} \frac{(n_i - t - 1)!t!}{n_i!}[m_k(T \cup i) - m_k(T)] \quad (2)$$

with $t = card(T)$. The Shapley index measures the *average relative importance* of a criterion and varies between zero and one. In the case of non monotone measure, it can happen that the Shapley index is close to zero even if the criterion is *important*. This is due to possible conflicting interactions that compensate positive marginal gains with negative ones. To this purpose, in [1] the *extended* Shapley index was introduced:

$$q_k(i) = \sum_{T \subseteq N \setminus \{i\}} \frac{(n_i - t - 1)!t!}{n_i!}|m_k(T \cup i) - m_k(T)| \quad (3)$$

for the i-th criterion and for the k-th Expert. Both the *extended* Shapley indices measures the *relative* importance of the criterion. If the measure is monotone, the two indices coincide. Otherwise, it can happen that the Shapley index is close to zero, while the *extended* index is high. This means that the criterion is *important*, but, on average, it is *neutral*, it is neither a benefit, nor a cost (for some coalitions it is a cost, while for others it is a benefit). Then, in the non monotone case, both the two indices should be considered.

For a better comprehension of the proposed methodology, consider the following multi person - multi attribute decision problem. Suppose that an Expert Committee is required to evaluate among some different investment projects that are the finite set of the available alternatives. Each project is characterized by a finite set of criteria. Every Expert scores independently any coalition of the criteria. Anywise, the Experts are not forced to consider *all* the criteria. Conversely, each Expert can consider an his *own* subset of criteria. This preference is formalized by a non additive signed measure, which can model *strict conflict* between the criteria. For instance, an Expert can consider three criteria valid for the judgements of the alternatives, and even if all of them, considered by alone, can be seen as *benefit*, the subset of the first and the second one induce a conflict, so that such a coalition receive an inferior score than the minimum between the first and the second criterion scores, which is usually implied in the monotone case.

6 A numerical example

Consider a decision problem where two alternatives have to be scored, on the basis of the four criteria values (alternative *profile*) reported in Table 1.

Suppose that three Experts are involved in the decision process. Then the DM gives a final evaluation of the alternatives on the basis of the Expert's individual scores. As above described, the decision process splits into two subsequent phases: the Expert's scoring, and the aggregation of each Expert's score into an aggregated score. If signed non additive measures are used for the aggregation in both the activities, the problem can be approached with a two-step non monotone Choquet integral.

Table 1: Alternative profile

x_1	x_2	x_3	x_4
0.4	0.2	0.8	0.6
0.1	0.7	0.4	0.5

Table 2: Expert n. 1

$m^1(1)$	$m^1(2)$	$m^1(1,2)$
0.4	0.2	1

In the first phase, suppose that the first Expert takes only the first two criteria into account, the second discharges the first criterion, while the third Expert considers the first and the fourth criteria only. The values assigned by the three Experts directly assigned or implicitly obtained using a suitable questionnaire as proposed by [3] to every coalition are reported in Tables from 2 up to 4. The first Expert considers more important the first criterion with respect to the third one, but the coalition formed by both the two Experts receives a weight greater than the sum of the two criteria weights. Thus the first Expert exhibits a *synergic*, or *disjunctive*, effect. In MCDA literature, this is sometimes knew as *andness*-type effect, while the opposite redundant behavior is named *orness*-type effect, see [7]. The second Expert is characterized by a *linear* behavior, as it can be easily checked. Both the first and the second Expert evaluate monotonically. But this is not true for the last Expert, which considers the first criterion more important than the fourth, but assigns to the coalition formed by the two criteria an inferior weight than the minimum of them. While the first Expert exhibits a *disjunctive* behavior (*orness*-type), and the second is *linear*, we can say that the last Expert is characterized by an *exclusive-orness* behavior. In fact, his preference structure follows a rule like:

The alternative is highly scored if the first or if the fourth criterion is high, but NOT both of them

The scoring of each alternatives can be obtained aggregating using the (single step) Choquet integral for each Expert. The Table 5 reports the results of the aggregation for the three Experts. For instance, the score of the first alternative calculated for the first Expert, 0.28, is obtained as follows, see (1):

$$\mathbf{x}_1^1 = 0.2 \times 1 + (0.4 - 0.2) \times 0.4 = 0.28 \quad (4)$$

In the subsequent phase, the individual Expert's scores need to be aggregated into a final one. To this purpose, suppose that, after a preliminary briefing, the implicit preference struc-

Table 3: Expert n. 2

$m^2(2)$	$m^2(3)$	$m^2(4)$
0.5	0.2	0.3

ISBN: 978-989-95079-6-8

$m^2(2,3)$	$m^2(2,4)$	$m^2(3,4)$	$m^2(2,3,4)$
0.7	0.8	0.5	1

Table 4: Expert n. 3

$m^3(1)$	$m^3(4)$	$m^3(1,4)$
0.7	0.6	0.2

ture of the DM about the Expert's confidence/experience be heuristically expressed by the following rules:

Rule # 1: *If the Experts 1 and 3 agree, I have complete confidence about their choice (in this case, the opinion of the second Expert is inessential).*

Rule # 2: *The coalition formed by Experts 1 and 2 is more reliable than the coalition formed by Experts 2 and 3.*

Rule # 3: *Considering each Experts by alone, the first one is the most important, while the second one is the least.*

These rules can be translated in the coalition weights reported in Table 6.

The individual measures can now be aggregated, obtaining the following aggregated alternative scores for the two considered alternatives:

$$\sigma_1 = 0.648, \sigma_2 = 0.344$$

The score σ_1 is computed as follows:

$$\sigma_1 = 0.2\times 1+(0.28-0.2)\times 0.5+(0.44-0.28)\times 0.3 = 0.648 \quad (5)$$

While for the first and for the second Experts the Shapley and the *extended* Shapley indices coincide (they are monotone), for the third Expert the Shapley index and the *extended* Shapley index are respectively, for the first criterion:

$$p_3 = 0.5 \times \{0.7 + (0.2 - 0.6)\} = 0.15 \quad (6)$$

$$q_3 = 0.5 \times \{0.7 + |0.2 - 0.6|\} = 0.55 \quad (7)$$

Observe that the Shapley index differs from the *extended* Shapley indices. In particular, the extended Shapley index is significatively high (0.55), while the Shapley index is low (0.15). We can conclude that this criterion of the third Expert is *important* even if it is, on average, neither a *benefit*, nor a *cost*.

In a future work, we intend to analyze the properties and the relationships between the Shapley and the *extended* Shapley for the two-step Choquet integral.

Table 5: Alternative scoring of each Expert

Alternatives	Expert n. 1	Expert n. 2	Expert n. 3
1	0.28	0.44	0.2
2	0.22	0.54	0.26

Table 6: Measure of Expert coalitions

u_1	u_2	u_3	$u_{1,2}$	$u_{1,3}$	$u_{2,3}$
0.2	0.3	0.7	0.5	0.7	1

References

[1] Cardin M., Giove S., Aggregation functions with non-monotonic measures, *Fuzzy Economic review*, 13, 2, 3-15, 2008.

[2] Ceberio M., Modave F., An interval valued, 2-additive Choquet Integral for Multi-criteria Decision Making, *10th Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, Perugia, Italy, 2004.

[3] Despic O., Simonovic S.P., Aggregation operators for soft decision making in water resources, *Fuzzy Sets and Systems*, 115, 11-33, 2000.

[4] Bordogna, G., Fedrizzi, M., Pasi, G., A Linguistic Modeling of Consensus for a Fuzzy Majority in Group Decision Making, *IEEE Transactions on Systems, Man and Cybernetics* 27, 1, 126-132, 1997.

[5] De Waegenaere A., Wakker P.P., Nonmonotonic Choquet integrals, *Journal of Mathematical Economics*, 36, 45-60, 2001.

[6] Grabish M., Marichal J.L., Roubens M., Equivalent representations of a set function with applications to game theory and multicriteria decision making. *Mathematics of Operations Research*, vol. 25, 2, 157-178, 2000.

[7] Grabisch M., Roubens M., *Application of the Choquet integral in multicriteria decision making*, in Fuzzy Measures and Integrals - Theory and Applications, M. Grabisch, T. Murofushi, and M. Sugeno (eds), Physica Verlag, 415-434, 2001.

[8] Kajii A., Kojima H., Takashi U., Cominimum additive operators, *Journal of Mathematical Economics*, 43, 218-230, 2007.

[9] Marichal J.L., An axiomatic approachof the discrete Choquet integral as a tool to aggregate interacting criteria *IEEE Transactions on Fuzzy Syssyems*, vol. 8, 6, 800-807, 2000.

[10] Meyer P., Roubens M., On the use of the Choquet integral with fuzzy numbers in multiple criteria decision support, *Fuzzy Sets and Systems*, 157, 927-938, 2006.

[11] Mesiar R., Vivona D., (1999), Two-step integral with respect to fuzzy measure, *Tatra Mountains Mathematical Pubblications*, 16, 359-368, 2006.

ISBN: 978-989-95079-6-8

[12] Murofushi T., Sugeno M., Machida M., Non-monotonic fuzzy measures and Choquet integral, *Fuzzy Sets and Systems*, 64, 73-86, 1994.

[13] Narukawa Y., Murofushi T., The n-step Choquet integral on finite spaces. In *Proc. of the 9th Int. Conf. on Information Proc. and Management of Uncertainty in Knowledge-Based Systems*, Annecy, France, 2002.

[14] Rébillé Y., Sequentially continuous non-monotonic Choquet integrals, *Fuzzy Sets and Systems*, 153, 79-94, 2005.

ISBN: 978-989-95079-6-8

A desktop calculator for parametric fuzzy arithmetic

Laerte Sorini

DEMQ (Department of Economics and Quantitative Methods)
University of Urbino 'Carlo Bo', Italy.
Email: laerte@uniurb.it

Abstract— *In this paper we describe a desktop calculator for the fuzzy arithmetic and the fuzzy extension of basic univariate functions. The fuzzy numbers are modelled indifferently in the parametric LR or LU representations. The operations are performed in the setting of Zadeh's extension principle.*

Keywords— LR-Fuzzy Numbers, LU-Fuzzy Parametrization, Fuzzy Arithmetic.

1 Introduction

About three decades ago, Dubois and Prade developed the arithmetical structure of fuzzy numbers and they introduced the well known LR model and the corresponding formulas for the fuzzy operations (see the recent publication [1] and the references therein).

The arithmetic calculations with fuzzy numbers, according to Zadeh's Extension Principle) can be performed using two general settings:

a. the well known LR fuzzy numbers, for which the operations are performed by calculating the membership function $x \rightarrow \mu(x)$ of the result from the membership functions of the operands (as in [1]).

b. the LU representation of fuzzy numbers, developed in [2], [4] and extensively described in [5], where the operations are performed using the $\alpha-cuts$ representation $\alpha \rightarrow u_\alpha^-, \alpha \rightarrow u_\alpha^+$ of the operands and on the basis of interval analysis for each cut $[u_\alpha^-, u_\alpha^+]$.

It is well known that, for the case of fuzzy numbers (but not for general fuzzy sets), the two representations are equivalent as it is possible to go from LR to LU and from LU to LR by inverting the (x, μ) or the (α, u) axes.

Following the theoretical results on LR and LU fuzzy arithmetic operations, we have developed a desktop calculator which performs the basic operations and the extension of elementary (unidimensional) functions using indifferently one of the two settings.

In this paper we describe the structure of the calculator and how it works.

2 LU and LR fuzzy numbers and arithmetic

Fuzzy numbers are fuzzy sets defined over the real numbers $\mathbb{R}$, having membership functions $(x, \mu_u(x))$ for each $x \in \mathbb{R}$, in the form

$$\mu_u(x) = \begin{cases} L(\frac{b-x}{b-a}) & \text{if } x \in [a,b] \\ 1 & \text{if } x \in [b,c] \\ R(\frac{x-c}{c-d}) & \text{if } x \in [c,d] \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $L, R : [0,1] \longrightarrow [0,1]$ are non-increasing with $R(0) = L(0) = 1$ and $R(1) = L(1) = 0$. The $\alpha - cuts$ are defined by

$$[u]_\alpha = \{x | x \in \mathbb{X},\ \mu_u(x) \geq \alpha\} \quad (2)$$

with

$$[u]_0 = cl\{x | x \in \mathbb{X},\ \mu_u(x) > 0\}.$$

We will denote them by

$$\begin{aligned} [u]_\alpha &= [u_\alpha^-, u_\alpha^+] \quad (3) \\ &\text{where} \\ u_\alpha^- &= b - (b-a)L^{-1}(\alpha) \\ u_\alpha^+ &= c + (d-c)R^{-1}(\alpha) \end{aligned}$$

It is well known that the two functions $\alpha \rightarrow u_\alpha^-$ (the Lower branch) and $\alpha \rightarrow u_\alpha^+$ (the Upper branch) are monotonic (respectively increasing and decreasing) functions for all $\alpha \in [0,1]$.

The LR or LU parametric representations use monotonic interpolation by shape functions $p : [0,1] \longrightarrow [0,1]$ such that $p(0) = 0$ and $p(1) = 1$ with $p(t)$ differentiable and increasing on $[0,1]$; with parameters $\beta_i \geq 0,\ i = 0,1$ we satisfy conditions

$$\begin{aligned} p(0) &= 0,\ p(1) = 1 \\ p'(0) &= \beta_0,\ p'(1) = \beta_1. \end{aligned}$$

An example is the following [4]:

$$p(t; \beta_0, \beta_1) = \frac{t^2 + \beta_0 t(1-t)}{1 + (\beta_0 + \beta_1 - 2)t(1-t)}; \quad (4)$$

Function p in (4) is increasing on [0,1] and is used as model for functions L and R; in fact, if $a \leq b \leq c \leq d$ and $\beta_{0,L}, \beta_{1,L} \geq 0$, $\beta_{0,R}, \beta_{1,R} \geq 0$ are given, an LR-fuzzy number has membership

$$\mu_u(x) = \begin{cases} p(\frac{b-x}{b-a}; \beta_{0,L}, \beta_{1,L}) & \text{if } x \in [a,b] \\ 1 & \text{if } x \in [b,c] \\ 1 - p(\frac{c-x}{c-d}; \beta_{0,R}, \beta_{1,R}) & \text{if } x \in [c,d] \\ 0 & \text{otherwise} \end{cases}.$$

We denote a, b, c, d as $a = u_{0,L}$, $b = u_{1,L}$, $c = u_{1,R}$, $d = u_{0,R}$ so that eight parameters define u:

$$u_{LR} = (u_{0,L}, \beta_{0,L}, u_{0,R}, \beta_{0,R}; u_{1,L}, \beta_{1,L}, u_{1,R}, \beta_{1,R}) \quad (5)$$

provided that $u_{0,L} \leq u_{1,L} \leq u_{1,R} \leq u_{0,R}$ and $\beta_{0,L}, \beta_{1,L} \geq 0$, $\beta_{0,R}, \beta_{1,R} \geq 0$.

ISBN: 978-989-95079-6-8

The LU representation is obtained if the model functions $p(t;\beta_0,\beta_1)$ are used to model the Lower and the Upper branches of the $\alpha-cuts$.

We can also switch the two representations: for example, for a given LR-fuzzy number $u\in\mathbb{F}^{LR}$ given by (5), its approximated LU representation $u\in\mathbb{F}^{LU}$ is

$$\left\{\begin{array}{c} u_{LU}=(u_0^-,\delta u_0^-,u_0^+,\delta u_0^+;\ u_1^-,\delta u_1^-,u_1^+,\delta u_1^+) \\ \text{with} \\ \begin{array}{l} u_0^-=u_{0,L}\,,\delta u_0^-=\frac{1}{\delta u_{0,L}} \\ u_1^-=u_{1,L}\,,\delta u_1^-=\frac{1}{\delta u_{1,L}} \\ u_0^+=u_{0,R}\,,\delta u_0^+=\frac{1}{\delta u_{0,R}} \\ u_1^+=u_{1,R}\,,\delta u_1^+=\frac{1}{\delta u_{1,R}} \end{array} \end{array}\right. \tag{6}$$

(if some $\delta u_{i,L},\delta u_{i,R}$ is zero, the corresponding infinite δu_i^-, δu_i^+ slope can be assigned a BIG number).

If the fuzzy numbers are given in the LR form, then the (LU)-(LR) fuzzy relationship (6) can be used as an intermediate step for LR-fuzzy operations.

Consider two LR-fuzzy numbers u and v ($N=1$ for simplicity)

$$\begin{array}{rcl} u_{LR} & = & (u_{0,L},\delta u_{0,L},u_{0,R},\delta u_{0,R};u_{1,L},\delta u_{1,L},u_{1,R},\delta u_{1,R}) \\ v_{LR} & = & (v_{0,L},\delta v_{0,L},v_{0,R},\delta v_{0,R};v_{1,L},\delta v_{1,L},v_{1,R},\delta v_{1,R}) \end{array} \tag{7}$$

having the LU representations

$$\begin{array}{rcl} u_{LU} & = & (u_0^-,\delta u_0^-,u_0^+,\delta u_0^+;u_1^-,\delta u_1^-,u_1^+,\delta u_1^+) \\ v_{LU} & = & (v_0^-,\delta v_0^-,v_0^+,\delta v_0^+;v_1^-,\delta v_1^-,v_1^+,\delta v_1^+) \end{array} \tag{8}$$

with $u_i^\pm$, $v_i^\pm$, $\delta u_i^\pm$ and $\delta v_i^\pm$ $(i=0,1)$ calculated according to (6).

The model functions above can be adopted not only to define globally the shapes, but also to represent the functions "piecewise", on a decomposition of the interval $[0,1]$ into N subintervals

$$0=\alpha_0<\alpha_1<.....<\alpha_{i-1}<\alpha_i<....<\alpha_N=1.$$

By the transformation $t_\alpha=\frac{\alpha-\alpha_{i-1}}{\alpha_i-\alpha_{i-1}}$, $\alpha\in I_i$, each subinterval I_i is mapped into the standard $[0,1]$ interval to determine each piece independently and obtain general left-continuous LU-fuzzy numbers. Globally continuous or more regular $C^{(1)}$ fuzzy numbers can be obtained directly from the data (for example, $u_{1,i}^-=u_{0,i+1}^-$, $u_{1,i}^+=u_{0,i+1}^+$ for continuity and $d_{1,i}^-=d_{0,i+1}^-$, $d_{1,i}^+=d_{0,i+1}^+$ for differentiability at $\alpha=\alpha_i$).

Let $p_i^\pm(t)$ denote the model function on I_i; we obtain easily

$$p_i^-(t)=p(t;\beta_{0,i}^-,\beta_{1,i}^-)\,,\ p_i^+(t)=1-p(t;\beta_{0,i}^+,\beta_{1,i}^+) \tag{9}$$

with

$$\begin{array}{rcl} \beta_{j,i}^- & = & \dfrac{\alpha_i-\alpha_{i-1}}{u_{1,i}^--u_{0,i}^-}d_{j,i}^- \\ \beta_{j,i}^+ & = & -\dfrac{\alpha_i-\alpha_{i-1}}{u_{1,i}^+-u_{0,i}^+}d_{j,i}^+ \text{ for } j=0,1 \end{array}$$

so that, for $\alpha\in[\alpha_{i-1},\alpha_i]$ and $i=1,2.,,,N$:

$$u_\alpha^- = u_{0,i}^-+(u_{1,i}^--u_{0,i}^-)p_i^-(t_\alpha) \tag{10}$$

$$t_\alpha = \frac{\alpha-\alpha_{i-1}}{\alpha_i-\alpha_{i-1}} \tag{11}$$

$$u_\alpha^+ = u_{0,i}^++(u_{1,i}^+-u_{0,i}^+)p_i^+(t_\alpha) \tag{12}$$

$$t_\alpha = \frac{\alpha-\alpha_{i-1}}{\alpha_i-\alpha_{i-1}}. \tag{13}$$

So, a differentiable shape function requires $4(N+1)$ parameters

$$\begin{array}{rcl} u & = & (\alpha_i;u_i^-,\delta u_i^-,u_i^+,\delta u_i^+)_{i=0,1,...,N} \\ u_0^- & \le & u_1^-\le...\le u_N^-\le u_N^+\le u_{N-1}^+\le...\le u_0^+ \\ \delta u_i^- & \ge & 0,\delta u_i^+\le 0. \end{array} \tag{14}$$

and the branches are computed according to (10) and (12). An example with $N=4$ is in **Table 1.**

Table 1. LU parametrization of a fuzzy number

α_i	u_i^-	δu_i^-	u_i^+	δu_i^+
0.0	-2.0	5.0	2.0	-0.5
0.5	-1.0	1.5	1.2	-2.0
1.0	0.0	2.5	0.0	-0.1

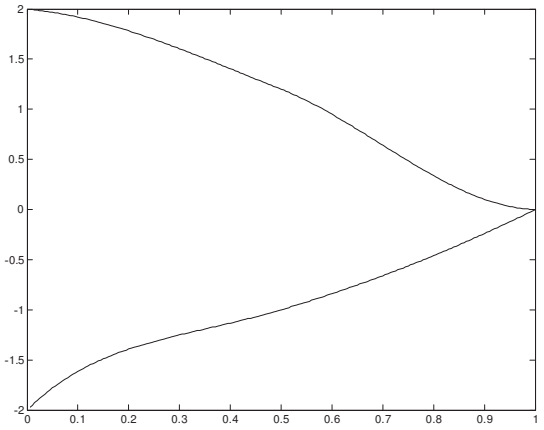


Figure 1. A fuzzy number in LU representation; the parameters are reported in Table 1 and the construction is obtained by the mixed spline with $N=2$.

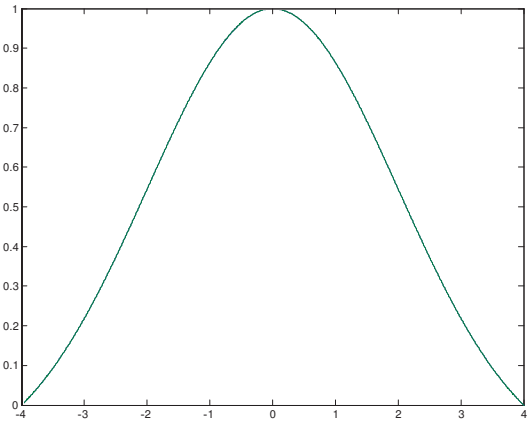


Figure 2. A quasi Gaussian fuzzy number; the parameters are reported in Table 2 and the membership function is obtained by the mixed spline with $N=4$.

For the LR parametrization on a decomposition with N subintervals, we proceed in a similar way; for example, a fuzzy quasi Gaussian number with the following membership

ISBN: 978-989-95079-6-8

function

$$\mu(x)=\begin{cases}\frac{\exp(-\frac{(x-m)^2}{2\sigma^2})-\exp(-\frac{k^2}{2})}{1-\exp(-\frac{k^2}{2})} & \text{if } m-k\sigma\le x\le m+k\sigma\\ 0 & \text{otherwise.}\end{cases} \tag{15}$$

for $m=0,\sigma=2,k=2$ and approximated with $N=4$ (five points) is in **Table 2**:

Table 2. LR parametrization of fuzzy number (15)

α_i	$u_{i,L}$	$\delta u_{i,L}$	$u_{i,R}$	$\delta u_{i,R}$
0.0	-4.0	0.156518	4.0	-0.156518
0.25	-2.8921	0.293924	2.8921	-0.293924
0.5	-2.1283	0.349320	2.1283	-0.349320
0.75	-1.3959	0.316346	1.3959	-0.316346
1.0	0.0	0.0	0.0	0.0

and a hyperbolic tangent fuzzy number with the following membership function:

$$\mu(x)=\begin{cases}\frac{\tanh(-k^2)-\tanh(-\frac{(x-m)^2}{\sigma^2})}{\tanh(-k^2)}\\ \text{if } m-k\sigma\le x\le m+k\sigma\\ 0,\ \text{otherwise}\end{cases}. \tag{16}$$

for $m=0,\sigma=3,k=1$ has the form indicated in **Table 3**:

Table 3.LR parametrization of fuzzy number (16)

α_i	$u_{i,L}$	$\delta u_{i,L}$	$u_{i,R}$	$\delta u_{i,R}$
0.0	-3.0	0.367627	3.0	-0.367627
0.25	-2.4174	0.475221	2.4174	-0.475221
0.5	-1.8997	0.473932	1.8997	-0.473932
0.75	-1.3171	0.370379	1.3171	-0.370379
1.0	0.0	0.0	0.0	0.0

The arithmetic operations associated to the LU representation are performed according to the extension principle and by representing the results by the same model used to represent the operands. The *addition* is defined by:

$$u+v=(u_i^-+v_i^-,\delta u_i^-+\delta v_i^-,u_i^++v_i^+,\delta u_i^++\delta v_i^+)_{i=0,1,\dots,N}.$$

The *scalar multiplication* is defined as follows:
if $k\ge 0$ then

$$ku=(ku_i^-,k\delta u_i^-,ku_i^+,k\delta u_i^+)_{i=0,1,\dots,N};$$

if $k<0$ then

$$ku=(ku_i^+,k\delta u_i^+,ku_i^-,k\delta u_i^-)_{i=0,1,\dots,N}.$$

In particular, if $k=-1$, we have

$$-u=(-u_i^+,-\delta u_i^+,-u_i^-,-\delta u_i^-)_{i=0,1,\dots,N}$$

and the *subtraction* is defined by

$$u-v=u+(-v).$$

We note explicitly that the scalar multiplication is always reproduced exactly in all the models for all $\alpha\in[0,1]$ but, in general, this is not true for the addition as the sum of rational or mixed functions is not always a rational or a mixed function of the same orders.

A particular situation arises for addition (or subtraction) if the mixed model is used. Suppose that the two branches to be added are given by the data $(u_0^{(1)},u_1^{(1)},\delta u_0^{(1)},\delta u_1^{(1)})$ and $(u_0^{(2)},u_1^{(2)},\delta u_0^{(2)},\delta u_1^{(2)})$; the mixed model is characterized by values of w (or a) for each data set

$$w^{(1)}=\frac{\delta u_0^{(1)}+\delta u_1^{(1)}}{u_1^{(1)}-u_0^{(1)}},\ w^{(2)}=\frac{\delta u_0^{(2)}+\delta u_1^{(2)}}{u_1^{(2)}-u_0^{(2)}}.$$

If $w^{(1)+(2)}$ is the w parameter for the addition, then

$$w^{(1)+(2)}=\frac{\delta u_0^{(1)}+\delta u_1^{(1)}+\delta u_0^{(2)}+\delta u_1^{(2)}}{u_1^{(1)}-u_0^{(1)}+u_1^{(2)}-u_0^{(2)}}$$

and it is easy to see that $w^{(1)+(2)}$ is a weighted average of $w^{(1)}$ and $w^{(2)}$ as

$$w^{(1)+(2)}\in\left[\min\left\{w^{(1)},w^{(2)}\right\},\max\left\{w^{(1)},w^{(2)}\right\}\right];$$

if $w^{(1)}=w^{(2)}$ then it follows that $w^{(1)+(2)}=w^{(1)}=w^{(2)}$. So, if the two fuzzy numbers to be added are modelled by a spline of the same *degree,* then the mixed model produces exact addition for all $\alpha\in[0,1]$.

This is true, in particular, for the fuzzy numbers having the slopes d_0 and d_1 not available, then we choose $d_0=n(1-\beta)(u_1-u_0)$ and $d_1=n\beta(u_1-u_0)$ for a fixed integer n and for a parameter $\beta\in[0,1]$. In this case, in fact, if

$$\begin{aligned}\delta u_0^{(k)} &= n(1-\beta^{(k)})(u_1^{(k)}-u_0^{(k)}),\ \delta u_1^{(k)}=\\ &= n\beta^{(k)}(u_1^{(k)}-u_0^{(k)}),\ k=1,2\end{aligned}$$

then

$$\begin{aligned}\delta u_0^{(1)}+\delta u_0^{(2)} &= n(1-\frac{\beta^{(1)}+\beta^{(2)}}{2})(u_1^{(1)}+u_1^{(2)}-u_0^{(1)}-u_0^{(2)})\\ \delta u_1^{(1)}+\delta u_1^{(2)} &= n\frac{\beta^{(1)}+\beta^{(2)}}{2}(u_1^{(1)}+u_1^{(2)}-u_0^{(1)}-u_0^{(2)})\end{aligned}$$

so that $\beta^{(1)+(2)}=\frac{\beta^{(1)}+\beta^{(2)}}{2}$.

For fuzzy *multiplication* we have an easy to implement algorithm, based on the applications of exact fuzzy multiplication at the nodes of the $\alpha-$subdivision; define

$$(uv)_i^-=\min\{u_i^-v_i^-,u_i^-v_i^+,u_i^+v_i^-,u_i^+v_i^+\} \tag{17}$$

$$(uv)_i^+=\max\{u_i^-v_i^-,u_i^-v_i^+,u_i^+v_i^-,u_i^+v_i^+\} \tag{18}$$

and set the following:

$$y=uv=\left(y_i^-,\delta y_i^-,y_i^+,\delta y_i^+\right)_{i=0,1,\dots,N}$$

To implement the multiplication we can proceed as follows: let (p_i^-,q_i^-) be the pair associated to the combination of superscripts $+$ and $-$ giving the minimum $(uv)_i^-$ in (17), and similarly let (p_i^+,q_i^+) the pair associated to the combination of $+$ and $-$ giving the maximum $(uv)_i^+$ in (18), then we obtain:

$$y_i^-=u_i^{p_i^-}v_i^{q_i^-}\ \text{ and }\ y_i^+=u_i^{p_i^+}v_i^{q_i^+}$$

$$\delta y_i^-=\delta u_i^{p_i^-}v_i^{q_i^-}+u_i^{p_i^-}\delta v_i^{q_i^-}\ \text{ and }\ \delta y_i^+=\delta u_i^{p_i^+}v_i^{q_i^+}+u_i^{p_i^+}\delta v_i^{q_i^+}$$

ISBN: 978-989-95079-6-8

where we use the product derivative rule to obtain the new slopes.

Analogous formulas can be deduced for *division:*

$$z = u/v = \left(z_i^-, \delta z_i^-, z_i^+, \delta z_i^+\right)_{i=0,1,...,N}$$

$$\begin{aligned}(u/v)_i^- &= \min\{u_i^-/v_i^-, u_i^-/v_i^+, u_i^+/v_i^-, u_i^+/v_i^+\} \text{ and}\\ (u/v)_i^+ &= \max\{u_i^-/v_i^-, u_i^-/v_i^+, u_i^+/v_i^-, u_i^+/v_i^+\}.\end{aligned}$$

Let (r_i^-, s_i^-) be the pair associated to the combination of $+$ and $-$ giving the minimum in $(u/v)_i^-$ and similarly let (r_i^+, s_i^+) be the pair associated to the combination of $+$ and $-$ giving the maximum in $(u/v)_i^+$, then it follows:

$$z_i^- = u_i^{r_i^-}/v_i^{s_i^-} \qquad z_i^+ = u_i^{r_i^+}/v_i^{s_i^+}$$

$$\begin{aligned}\delta z_i^- &= (\delta u_i^{r_i^-} v_i^{s_i^-} - u_i^{r_i^-}\delta v_i^{s_i^-})/(v_i^{s_i^-})^2\\ \delta z_i^+ &= (\delta u_i^{r_i^+} v_i^{s_i^+} - u_i^{r_i^+}\delta v_i^{s_i^+})/(v_i^{s_i^+})^2.\end{aligned}$$

As pointed out by the results of experimentation reported in [2] and [4] , the operations above are exact at the nodes α_i and have very small global errors on $[0,1]$. Further, it is easy to control the error by using a sufficiently high number of nodes with $\max\{\alpha_i - \alpha_{i-1}\}$ sufficiently small.

The general algorithms for the four arithmetical operations are now detailed.

Let $u = (u_i^-, \delta u_i^-, u_i^+, \delta u_i^+)_{i=0,1,...,N}$ and $v = (v_i^-, \delta v_i^-, v_i^+, \delta v_i^+)_{i=0,1,...,N}$ be given; in order to calculate the LU *addition* $w = u + v$ and the LU *subtraction,* $z = u - v$ with $w = (w_i^-, \delta w_i^-, w_i^+, \delta w_i^+)_{i=0,1,...,N}$ and $z = (z_i^-, \delta z_i^-, z_i^+, \delta z_i^+)_{i=0,1,...,N}$, the following sequence of iterations has to be improved:

for $i = 0, 1, ..., N$
$w_i^- = u_i^- + v_i^-,\ z_i^- = u_i^- - v_i^+$
$\delta w_i^- = \delta u_i^- + \delta v_i^-,\ \delta z_i^- = \delta u_i^- - \delta v_i^+$
$w_i^+ = u_i^+ + v_i^+,\ z_i^+ = u_i^+ - v_i^-$
$\delta w_i^+ = \delta u_i^+ + \delta v_i^+,\ \delta z_i^+ = \delta u_i^+ - \delta v_i^-$
end
test if conditions (14) are satisfied
for $(y_i^-, \delta y_i^-, y_i^+, \delta y_i^+)_{i=0,1,...,N}$

Let $k \in \mathbb{R}$ and $u = (u_i^-, \delta u_i^-, u_i^+, \delta u_i^+)_{i=0,1,...,N}$ be given; the computation of the LU *scalar multiplication* $w = ku$ with $w = (w_i^-, \delta w_i^-, w_i^+, \delta w_i^+)_{i=0,1,...,N}$ is obtained with the following iterations:

for $i = 0, 1, ..., N$
if $k \geq 0$ **then** $w_i^- = ku_i^-,\ \delta w_i^- = k\delta u_i^-$
$w_i^+ = ku_i^+\ ,\ \delta w_i^+ = k\delta u_i^+$
else $w_i^- = ku_i^+,\ \delta w_i^- = k\delta u_i^+$
$w_i^+ = ku_i^-\ ,\ \delta w_i^+ = k\delta u_i^-$
end

Finally, if $u = (u_i^-, \delta u_i^-, u_i^+, \delta u_i^+)_{i=0,1,...,N}$ and $v = (v_i^-, \delta v_i^-, v_i^+, \delta v_i^+)_{i=0,1,...,N}$ are given then the LU multiplication $w = uv$, with $w = (w_i^-, \delta w_i^-, w_i^+, \delta w_i^+)_{i=0,1,...,N}$, is deduced from the following algorithm:

for $i = 0, 1, ..., N$
$m_i = \min\{u_i^- v_i^-, u_i^- v_i^+, u_i^+ v_i^-, u_i^+ v_i^+\}$
$M_i = \max\{u_i^- v_i^-, u_i^- v_i^+, u_i^+ v_i^-, u_i^+ v_i^+\}$
$w_i^- = m_i\ ,\ w_i^+ = M_i$
if $u_i^- v_i^- = m_i$ **then** $\delta w_i^- = \delta u_i^- v_i^- + u_i^- \delta v_i^-$
elseif $u_i^- v_i^+ = m_i$ **then** $\delta w_i^- = \delta u_i^- v_i^+ + u_i^- \delta v_i^+$
elseif $u_i^+ v_i^- = m_i$ **then** $\delta w_i^- = \delta u_i^+ v_i^- + u_i^+ \delta v_i^-$
elseif $u_i^+ v_i^+ = m_i$ **then** $\delta w_i^- = \delta u_i^+ v_i^+ + u_i^+ \delta v_i^+$
endif
if $u_i^- v_i^- = M_i$ **then** $\delta w_i^+ = \delta u_i^- v_i^- + u_i^- \delta v_i^-$
elseif $u_i^- v_i^+ = M_i$ **then** $\delta w_i^+ = \delta u_i^- v_i^+ + u_i^- \delta v_i^+$
elseif $u_i^+ v_i^- = M_i$ **then** $\delta w_i^+ = \delta u_i^+ v_i^- + u_i^+ \delta v_i^-$
elseif $u_i^+ v_i^+ = M_i$ **then** $\delta w_i^+ = \delta u_i^+ v_i^+ + u_i^+ \delta v_i^+$
endif
end

A similar algorithm can be deduced for the division.

3 Computation of unidimensional fuzzy-valued functions with the calculator

We consider first a single variable differentiable function $f : \mathbb{R} \to \mathbb{R}$; its *(EP)*-extension $v = f(u)$ to a fuzzy argument $u = (u^-, u^+)$ has $\alpha - cuts$

$$[v]_\alpha = \left[\min\left\{f(x) \mid x \in [u]_\alpha\right\}, \max\left\{f(x) \mid x \in [u]_\alpha\right\}\right]. \tag{19}$$

If f is monotonic increasing we obtain $[v]_\alpha = [f(u_\alpha^-), f(u_\alpha^+)]$ while, if f is monotonic decreasing, $[v]_\alpha = [f(u_\alpha^+), f(u_\alpha^-)]$; the LU representation of $v = (v_i^-, \delta v_i^-, v_i^+, \delta v_i^+)_{i=0,1,...,N}$ can be obtained.

In the non monotonic (differentiable) case, we have to solve the optimization problems in (19) for each $\alpha = \alpha_i$, $i = 0, 1, ..., N$, i.e.

$$(EP_i)\text{:} \begin{cases} v_i^- = \min\left\{f(x) | x \in [u_i^-, u_i^+]\right\} \\ v_i^+ = \max\left\{f(x) | x \in [u_i^-, u_i^+]\right\}. \end{cases}$$

The $\min$ (or the $\max$) can occur either at a point which is coincident with one of the extremal values of $[u_i^-, u_i^+]$ or at a point which is internal; in the last case, the derivative of f is null and $\delta v_i^- = 0$ (or $\delta v_i^+ = 0$).

To implement the LU-fuzzy calculator, we have written a windows-based frame similar to a standard hand-calculator.

Figure 3 shows a complete view of the calculator; from left to right we can see the grids of the fuzzy numbers X, Y and Z. Z is the result of the operations while X and/or Y are the operands. For each element $u \in \{X, Y, Z\}$the grid contains the values α_i, u_i^-, δu_i^-, u_i^+ and δu_i^+ respectively in the LU View Mode or x, $\mu(x)$, $\delta\mu(x)$ in the LR Mode. To start the calculations, we have implemented a set of predefined types, including triangular, trapezoidal, exponential, gamma, etc. For a given type, it is possible to define the number N of subintervals ($N+1$ points) in the uniform $\alpha-$decomposition.

ISBN: 978-989-95079-6-8

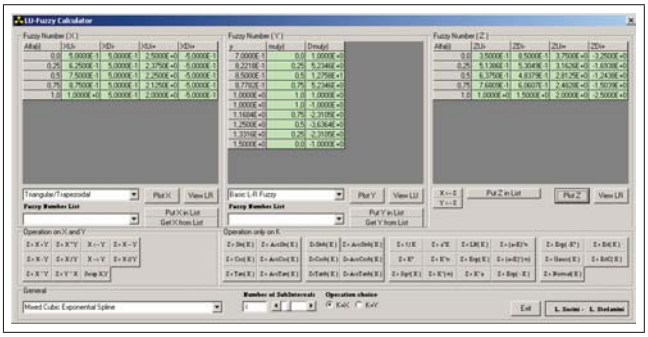


Figure 3. General window of the LU-fuzzy calculator.

The calculations are performed by clicking the button of the corresponding operation. The left group of buttons involves the binary operations (see Figure 4)

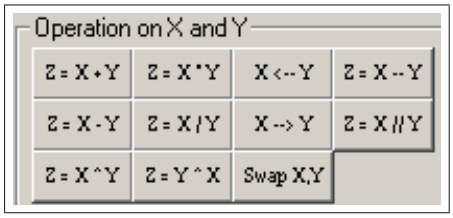


Figure 4. Binary operations and assignments.

The second group of operators (see Figure 5,6,7 and 8.) require the assignment of either X or Y to the temporary K and operate on K itself putting the result into Z.

Figure 5. Extension of univariate functions.

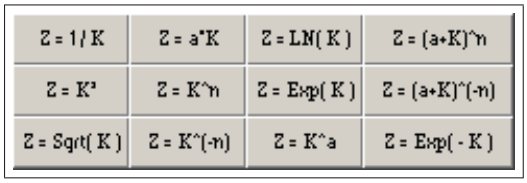


Figure 6. Extension of other univariate functions.

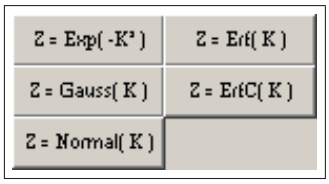


Figure 7. Extension of some other univariate functions.

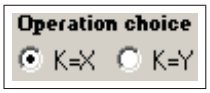


Figure 8. Selection of the argument for fuzzy extension functions.

It is possible to save a given (X, Y or Z) temporary result into a stored list (Put in List button), by assigning a name to it; a saved fuzzy number can be reloaded either in X or Y for further use (Get from List button). The Plot button (see Figure 9.) opens a popup window with the graph of the membership function of the corresponding fuzzy number and is possible select a fuzzy number from a list of predefined types; (View LR,View LU button) allow to switch between LR or LU fuzzy representation.

Figure 9. Fuzzy Number Control Panel

To obtain the graphs or other representations, one of the models can be selected (Figure 10).

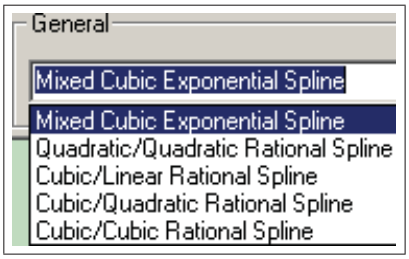


Figure 10. Choosing the monotonic spline model

3.1 *Fuzzy extension of* $(X, Y) \to X^Y$

We illustrate an example to show how the calculator works. We use the definition

$$X^Y = \exp\left(Y \ln\left((X)\right)\right)$$

and we compute

$$Z = X^Y$$

(where the input are the positive fuzzy X and the fuzzy Y), by the sequence of operations:

$$\begin{array}{lll} (i)\text{ (natural logarithm)} & Z & \longleftarrow \quad \ln(X) \\ (ii)\text{ (standard multiplication)} & Z & \longleftarrow \quad YZ \\ (iii)\text{ (exponential)} & Z & \longleftarrow \quad \exp(Z) \end{array}$$

The steps to follow in the calculator are the following.

1. First (see Figure 11) select a trapezoidal fuzzy number here we have 4 sub intervals and 5 $\alpha - cut$.

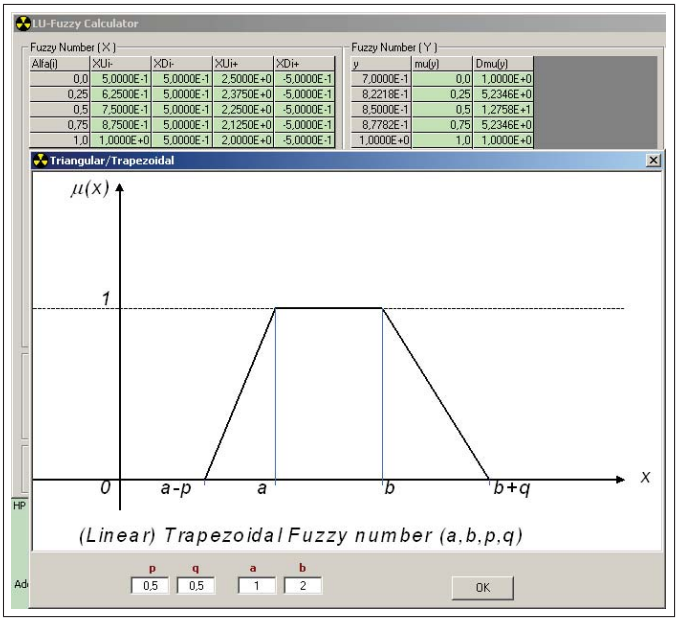


Figure 11. Example for trapezoidal fuzzy number loaded in X

If the selection is loaded into the X-area, the corresponding grid appears as in Figure 12 below and it is possible plot and switch in LR View mode immediately.

ISBN: 978-989-95079-6-8

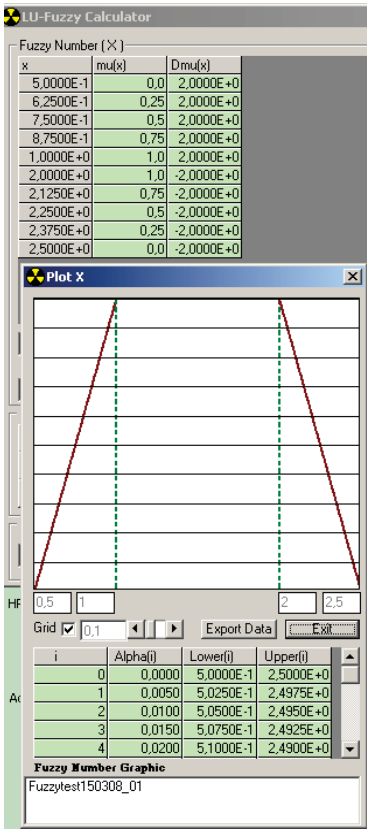


Figure 12. Assigne and plot the trapezoidal fuzzy number to X.

2. A second fuzzy number is loaded into Y and the button corresponding to the operation Z=X^Y is activated. The Z-grid is calculated by the rules of the LU-fuzzy calculus (see Figure 13).

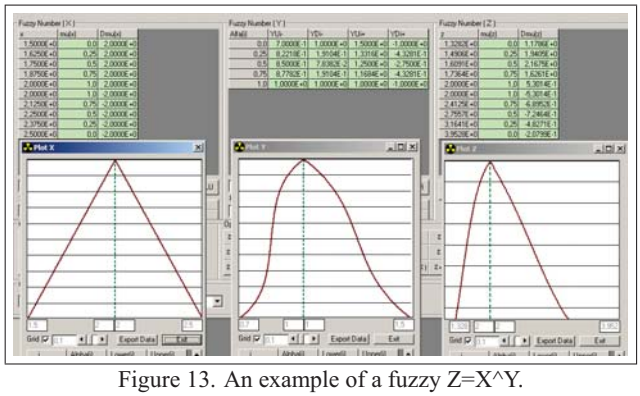


Figure 13. An example of a fuzzy Z=X^Y.

3. To see the graphical representation of X, Y and/or Z, click the corresponding Plot button and the popup windows appear; note that the fuzzy numbers X and Z are represented in LR mode instead Y is in LU mode.

4. Now, we save the result Z of the previous operation and we call it eusflat_Z (figure 14.).

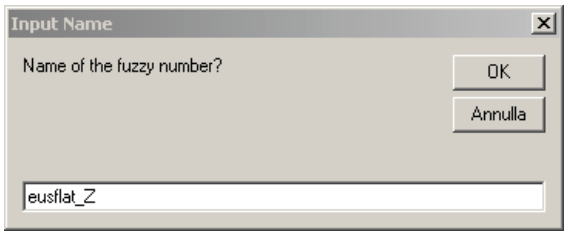


Figure 14. Choose the name for the result.

5. Now we load the saved Fuzzy_Z into the X-area, by getting it from the list of saved elements and with the possibility of reiterating.

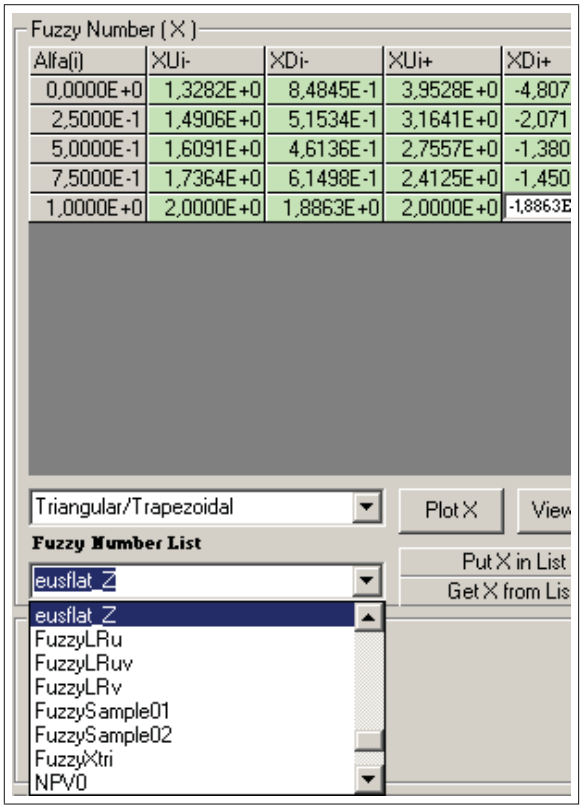


Figure 15. Loading of a saved intermediate fuzzy number into X area.

We believe that the suggested desktop calculator might be of interest in many areas where fuzzy numbers are applied, especially in computer decision support systems. In fact, it enables the interpretation of shapes as inputs and as outputs and the use of LR or LU parametric representations.

References

[1] D. Dubois, H. Prade (ed.), *Fundamentals of Fuzzy Sets*, Kluwer, Boston, The Handbooks of Fuzzy Sets Series, 2000.

[2] M. L. Guerra, L. Stefanini, Approximate Fuzzy Arithmetic Operations Using Monotonic Interpolations, *Fuzzy Sets and Systems*, 150, 2005, 5-33.

[3] L. Stefanini, L. Sorini, An LU-fuzzy calculator for the basic fuzzy calculus, Working Papers 0701, University of Urbino Carlo Bo, Department of Economics, 2007.

[4] L. Stefanini, L. Sorini, M. L. Guerra, Parametric representation of fuzzy numbers and application to fuzzy calculus, *Fuzzy Sets and Systems*, 157, 2006, 2423-2455.

[5] L. Stefanini, L. Sorini, M.L. Guerra, Fuzzy Numbers and Fuzzy Arithmetic, in W. Pedrycz et al. (Eds) "Handbook of Granular Computing" , Chapter 12, J. Wiley & Sons, 2008.

ISBN: 978-989-95079-6-8

Discussions on Interpretability of Fuzzy Systems using Simple Examples

Hisao Ishibuchi, Yusuke Nojima

Department of Computer Science and Intelligent Systems, Osaka Prefecture University
1-1 Gakuen-cho, Naka-ku, Sakai, Osaka, 599-8531 Japan
Email: hisaoi@cs.osakafu-u.ac.jp, nojima@cs.osakafu-u.ac.jp

Abstract—*Two conflicting goals are often involved in the design of fuzzy rule-based systems: Accuracy maximization and interpretability maximization. A number of approaches have been proposed for finding a fuzzy rule-based system with a good accuracy-interpretability tradeoff. Formulation of the accuracy maximization is usually straightforward in each application area of fuzzy rule-based systems such as classification, regression and forecasting. Formulation of the interpretability maximization, however, is not so easy. This is because various aspects of fuzzy rule-based systems are related to their interpretability. Moreover, user's preference should be taken into account when a single fuzzy rule-based system is to be chosen from several alternatives with different accuracy-interpretability tradeoffs. In this paper, we discuss the difficulty in measuring the interpretability of fuzzy rule-based systems using very simple examples. We do not intend to propose any new interpretability measure. Our intention is to help to activate discussions on how to measure the interpretability of fuzzy rule-based systems.*

Keywords—Fuzzy systems, fuzzy rules, accuracy-interpretability tradeoff, multiobjective design of fuzzy systems.

1 Introduction

Handling of the tradeoff between the accuracy maximization and the interpretability maximization has been a hot issue in the design of fuzzy rule-based systems since the mid-1990s [1]-[3]. A number of approaches have already been proposed for improving the accuracy of fuzzy rule-based systems while maintaining their interpretability [1], [2], [4]-[21]. Genetic algorithms have been frequently used in those approaches to search for an accurate and interpretable fuzzy rule-based system. This is because genetic algorithms can perform not only continuous optimization for parameter tuning but also discrete optimization for structure determination. Studies on fuzzy genetics-based machine learning are called genetic fuzzy systems [22]-[24]. In some recent studies [3], [25]-[36], multi-objective genetic algorithms have been used to search for multiple Pareto-optimal fuzzy rule-based systems along the accuracy-interpretability tradeoff surface. Those studies are often referred to as multi-objective genetic fuzzy systems [37]. Recently multi-objective genetic algorithms have also been used for machine learning [38] and data mining [39].

Let us denote a fuzzy rule-based system by S. We can also view S as a set of fuzzy if-then rules. In each application area of fuzzy rule-based systems such as classification, regression and forecasting, the specification of the accuracy of S for the given training data is not difficult (e.g., the number of correctly classified training patterns by S). Let us denote the accuracy measure of S as *Accuracy*(S). A design problem of fuzzy rule-based systems can be formulated as follows:

$$\text{Maximize } \mathit{Accuracy}(S). \quad (1)$$

Due to the accuracy-interpretability tradeoff relation, the accuracy maximization in (1) often leads to the deterioration in the interpretability of fuzzy rule-based systems. This means that we often obtain from (1) an accurate and complicated fuzzy rule-based system with poor interpretability.

In some application areas, not only the accuracy but also the interpretability is very important. Thus we often want to maximize the accuracy of fuzzy rule-based systems without degrading their interpretability. This maximization problem can be formulated as follows:

$$\text{Maximize } \mathit{Accuracy}(S) \text{ subject to } \mathit{Interpretability}(S) \geq \alpha, \quad (2)$$

where *Interpretability*(S) is the interpretability measure of the fuzzy rule-based system S and α is the required minimum level of the interpretability.

Of course, we can formulate the maximization problem of the interpretability under the given minimum accuracy level β as follows:

$$\text{Maximize } \mathit{Interpretability}(S) \text{ subject to } \mathit{Accuracy}(S) \geq \beta. \quad (3)$$

One may want to maximize both the accuracy and the interpretability. In this case, a simple approach is to use a scalarizing function $f(\cdot)$ which combines the accuracy and interpretability measures into a single objective function:

$$\text{Maximize } f(\mathit{Accuracy}(S), \mathit{Interpretability}(S)). \quad (4)$$

A well-known scalarizing function is the weighted sum:

$$\text{Maximize } w_1 \mathit{Accuracy}(S) + w_2 \mathit{Interpretability}(S), \quad (5)$$

where w_1 and w_2 are non-negative weight values. In addition to the weighted sum in (5), we can use various scalarizing functions developed in the field of multiple criteria decision making (MCDM [40]-[42]).

In general, it is not easy for human users to specify an appropriate scalarizing function for multi-objective problems. Users may want to examine some fuzzy rule-based systems with different accuracy-interpretability tradeoffs (instead of a single best solution with respect to a specific scalarizing function). In this case, the design of fuzzy rule-based systems can be formulated as the following multi-objective problem:

$$\text{Maximize } \{\mathit{Accuracy}(S), \mathit{Interpretability}(S)\}. \quad (6)$$

A large number of Pareto-optimal fuzzy rule-based systems can be obtained by multi-objective genetic algorithms such as NSGA-II [43], SPEA [44] and SPEA2 [45].

ISBN: 978-989-95079-6-8

In many cases, the interpretability maximization is handled as the complexity minimization. Thus the above-mentioned formulations in (2)-(6) can be reformulated accordingly. For example, the multi-objective formulation in (6) is rewritten as

Maximize *Accuracy*(*S*) and minimize *Complexity*(*S*), (7)

where *Complexity*(*S*) is a complexity measure.

The main difficulty in the above-mentioned formulations in (2)-(7) for the design of accurate and interpretable fuzzy rule-based systems is the formulation of their interpretability. Whereas the formulation of the accuracy of fuzzy rule-based systems is usually straightforward from their application task such as classification and regression, it is not easy for human users to appropriately formulate the interpretability. This is because various aspects of fuzzy rule-based systems are related to their interpretability [46]-[52]. Moreover, it is not easy for human users to mathematically formulate each of those aspects even when the close relation of each aspect to the interpretability of fuzzy rule-based systems is clear.

In this paper, we explain the difficulty in formulating the interpretability of fuzzy rule-based systems using simple numerical examples. More specifically, we demonstrate the difficulty in comparing different fuzzy rule-based systems with respect to their interpretability even in very simple situations.

2 Interpretability of Fuzzy Partitions

When we use the same type of fuzzy partitions with different granularities (e.g., uniform fuzzy partitions with symmetric triangular membership functions (MFs) in Fig. 1), we can say that the increase in the number of membership functions degrades the interpretability of fuzzy partitions. For example, the fuzzy partition with two membership functions in Fig. 1 (a) is the most interpretable among the four alternatives in Fig. 1. In this case, we can formulate the interpretability of fuzzy partitions by the number of membership functions. That is, the interpretability maximization is realized by minimizing the number of membership functions in fuzzy partitions.

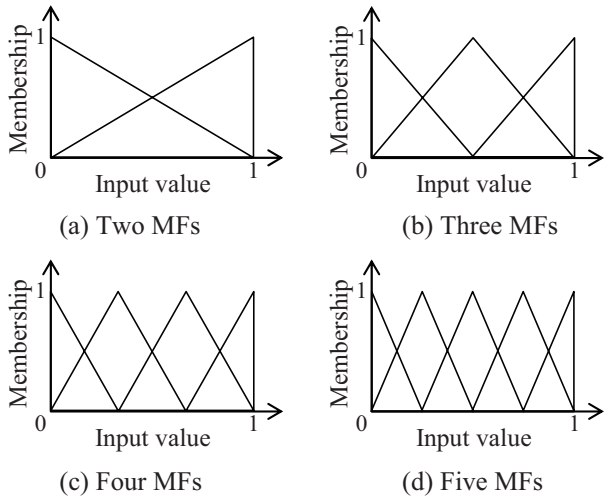


Figure 1: Fuzzy partitions with different granularities.

We can also use the number of membership functions as an interpretability measure (i.e., as a complexity measure to be minimized) when the same fuzzy partition is used for all input variables. For example, we can say that the 3x3 fuzzy grid in Fig. 2 (a) is more interpretable than the 4x4 fuzzy grid in Fig. 2 (b). The comparison, however, becomes difficult when we use different fuzzy partitions for each input variable as shown in the following two examples.

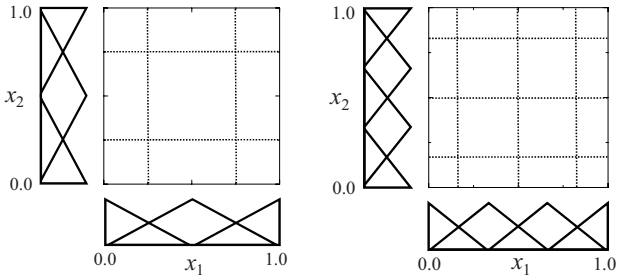


Figure 2: Comparison between the 3x3 and 4x4 fuzzy grids.

Example 1: Let us consider the 4x4 and 3x5 fuzzy grids in Fig. 3. Both fuzzy grids have eight membership functions in total (i.e., 4+4 = 3+5). Thus they are evaluated as having the same interpretability if they are compared using the number of membership functions. The 3x5 fuzzy grid is, however, viewed as being more interpretable than the 4x4 fuzzy grid if we use the number of fuzzy subspaces as an interpretability measure (i.e., 15 < 16). Since a single fuzzy rule is usually generated for each fuzzy subspace, the 3x5 fuzzy grid in Fig. 3 (b) can be viewed as being more interpretable than the 4x4 fuzzy grid in Fig. 3 (a) if we evaluate the interpretability using the number of fuzzy rules. Some human users, however, may intuitively feel that the 4x4 fuzzy grid with the same fuzzy partition for the two input variables is more interpretable than the 3x5 fuzzy grid with the different fuzzy partitions.

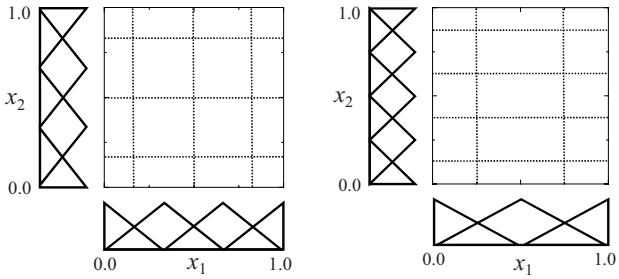
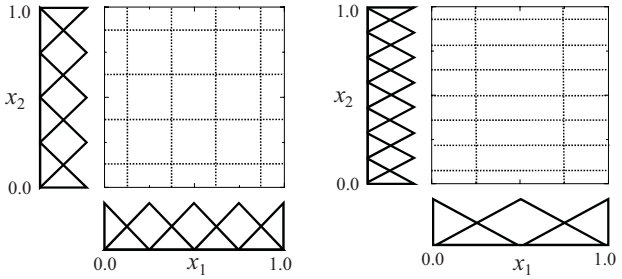


Figure 3: Example 1 with the 4x4 and 3x5 fuzzy grids.

Example 2: Let us consider the 5x5 and 3x8 fuzzy grids in Fig. 4. The 5x5 fuzzy grid in Fig. 4 (a) has less membership functions than the 3x8 fuzzy grid in Fig. 4 (b): 5+5 < 3+8. Thus the 5x5 fuzzy grid is evaluated as more interpretable than the 3x8 fuzzy grid in Fig. 4 (b) if we use the number of membership functions as an interpretability measure. The 5x5 fuzzy grid, however, has more fuzzy subspaces than the 3x8 fuzzy grid: 25 > 24. Thus the 3x8 fuzzy grid is viewed as being more interpretable than the 5x5 fuzzy grid if we use the number of fuzzy subspaces (i.e., the number of fuzzy rules) as an interpretability measure.

ISBN: 978-989-95079-6-8



(a) 5x5 fuzzy grid. (b) 3x8 fuzzy grid.

Figure 4: Example 2 with the 5x5 and 3x8 fuzzy grids.

As we have already explained in this section using the two examples, the comparison of different fuzzy grids with respect to their interpretability is not so easy. This means that the choice of an interpretability measure is difficult. If we use the number of fuzzy subspaces in Fig. 4, the 3x8 fuzzy grid is viewed as being more interpretable than the 5x5 fuzzy grid. When we use not only the number of fuzzy subspaces but also the number of membership functions, these two fuzzy grids are viewed as being non-dominated with each other with respect to the interpretability. In this case, the multi-objective formulation in (7) is handled as the three-objective problem:

$$\begin{aligned} &\text{Maximize } \mathit{Accuracy}(S), \text{ and} \\ &\quad \text{minimize } \{\mathit{Complexity}_1(S), \mathit{Complexity}_2(S)\}, \end{aligned} \quad (8)$$

where $\mathit{Complexity}_1(S)$ and $\mathit{Complexity}_2(S)$ are different complexity measures to be minimized (e.g., the number of fuzzy subspaces and the number of membership functions).

Even if we use these two measures, the 3x5 fuzzy grid is viewed as more interpretable than the 4x4 fuzzy grid. Thus we need another measure if we want to include some bias toward fuzzy grids with the same fuzzy partition for all input variables. For example, the 4x4 fuzzy grid is evaluated as more interpretable than the 3x5 fuzzy grid if the maximum number of membership functions for each input variable is used as an interpretability measure ($\max\{4, 4\} < \max\{3, 5\}$).

3 Interpretability of Fuzzy Rule-Based Systems

In genetic fuzzy systems [22]-[24], almost all aspects of fuzzy rule-based systems can be optimized since genetic algorithms perform continuous, discrete and combinatorial optimization. For example, genetic fuzzy systems can be used for choosing an appropriate type of fuzzy rules (e.g., Takagi-Sugeno, simplified Takagi-Sugeno and Mamdani). In this section, we discuss the interpretability of fuzzy rule-based systems with different fuzzy partitions and different types of fuzzy rules.

Let us consider a simple function approximation problem of a single-input and single-output system $y = f(x)$ in Fig. 5. Our task is to design an accurate and interpretable fuzzy rule-based system from the given input-output data in Fig. 5. For this task, Takagi-Sugeno fuzzy rules are written as

$$\text{Rule } R_i: \text{If } x \text{ is } A_i \text{ then } y = a_i + b_i x, \ i = 1, 2, ..., N, \quad (9)$$

where i is a rule index, A_i is an antecedent fuzzy set, a_i and b_i are real number coefficients of a consequent linear function of each fuzzy rule, and N is the total number of fuzzy rules.

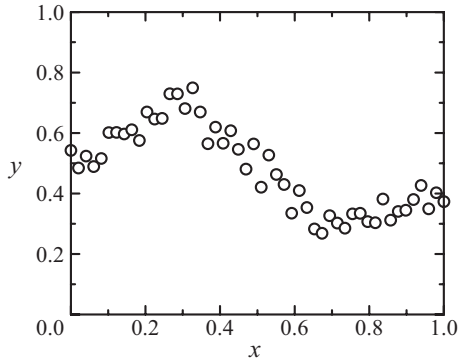


Figure 5: Input-output data in Examples 3 and 4.

When an input value x is presented to the fuzzy rule-based system with the N fuzzy rules in (9), the output value is estimated as follows:

$$y(x) = \frac{\sum_{i=1}^{N} (a_i + b_i x) \cdot \mu_{A_i}(x)}{\sum_{i=1}^{N} \mu_{A_i}(x)}, \quad (10)$$

where $y(x)$ is the estimated output value for the input value x, and $\mu_{A_i}(x)$ is the membership value of the antecedent fuzzy set A_i for the input value x.

We can use the same fuzzy reasoning mechanism for the simplified version of Takagi-Sugeno fuzzy rules:

$$\text{Rule } R_i: \text{If } x \text{ is } A_i \text{ then } y \text{ is } h_i, \ i = 1, 2, ..., N, \quad (11)$$

where h_i is a consequent real number.

Example 3: From the input-output data in Fig. 5, one may think that they can be approximated by a fuzzy rule-based system with three Takagi-Sugeno fuzzy rules. An example of such a fuzzy rule-based system is shown in Fig. 6 where each of the three lines (1), (2) and (3) is the consequent linear function of each of the three fuzzy rules with the trapezoidal antecedent fuzzy sets A_1, A_2 and A_3. The same input-output data can be also approximated by a fuzzy rule-based system with four simplified Takagi-Sugeno fuzzy rules as shown in Fig. 7. Each fuzzy rule in Fig. 7 has a triangular membership function A_i and a consequent real number h_i. The question is which is more interpretable between the two fuzzy rule-based systems in Fig. 6 and Fig. 7.

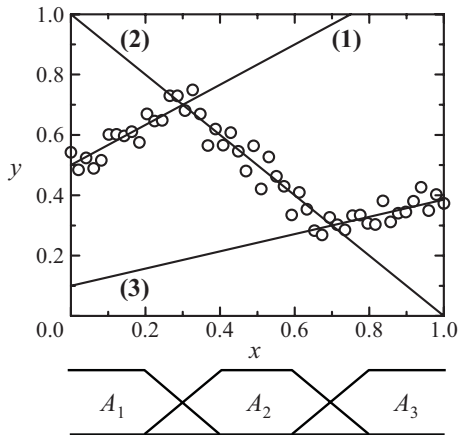


Figure 6: Three Takagi-Sugeno fuzzy rules.

ISBN: 978-989-95079-6-8

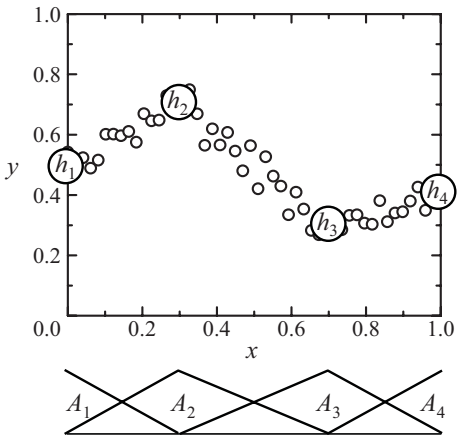


Figure 7: Four simplified Takagi-Sugeno fuzzy rules.

In Fig. 8 and Fig. 9, we show the fuzzy reasoning results by these two fuzzy rule-based systems in Fig. 6 and Fig. 7, respectively. We can see that similar results were obtained from the two fuzzy rule-based systems. Since there is no large difference in the approximation accuracy between Fig. 8 and Fig. 9, the interpretability will play an important role in the selection between the two fuzzy models in Fig. 6 and Fig. 7. If we use the number of fuzzy rules as an interpretability measure, the Takagi-Sugeno model in Fig. 6 is viewed as being more interpretable than the simplified Takagi-Sugeno model in Fig. 7. On the other hand, if we use the total number of parameters (i.e., a_i, b_i and h_i) in the consequent part of the fuzzy rules as an interpretability measure, the simplified Takagi-Sugeno model is evaluated as more interpretable than the Takagi-Sugeno model (i.e., 4 < 6).

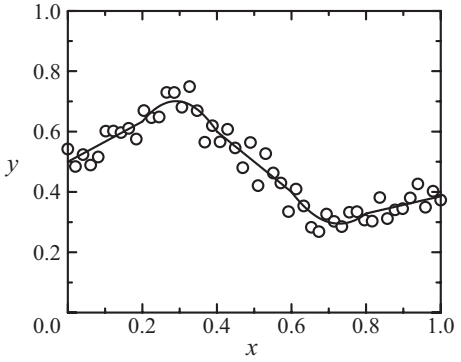


Figure 8: Results by the three Takagi-Sugeno fuzzy rules.

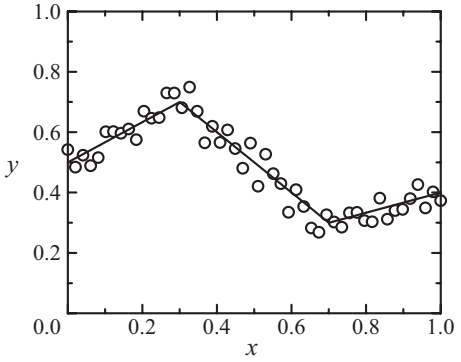


Figure 9: Results by the four simplified fuzzy rules.

Example 4: The given input-output data in Fig. 5 can be also approximated by the two Takagi-Sugeno fuzzy rules in Fig. 10. Whereas data points around $x = 0.5$ are far from the two consequent linear functions in Fig. 10, they can be approximated through the interpolation mechanism of the fuzzy reasoning in Eq. (10). Fig. 11 is the fuzzy reasoning result by the two fuzzy rules in Fig. 10. We can see from Fig. 11 that good approximation was realized by the two fuzzy rules in Fig. 10. Actually, the fuzzy reasoning result in Fig. 11 by the two Takagi-Sugeno fuzzy rules is similar to Fig. 8 and Fig. 9. The question is which is more interpretable between Fig. 6 with the three rules and Fig. 10 with the two rules.

It is clear that Fig. 10 is simpler than Fig. 6 with respect to various aspects of fuzzy rule-based systems (e.g., the number of fuzzy rules, the number of membership functions, and the number of parameters). However, one may think that Fig. 6 is more intuitive than Fig. 10. If we use the local accuracy of each liner function [5] as an interpretability measure, the three Takagi-Sugeno fuzzy rules in Fig. 6 are evaluated as being more interpretable than the two rules in Fig. 10.

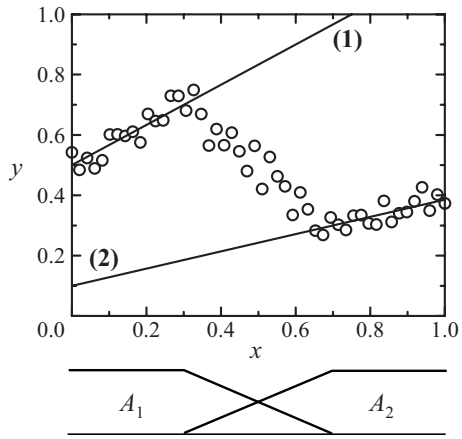


Figure 10: Two Takagi-Sugeno fuzzy rules.

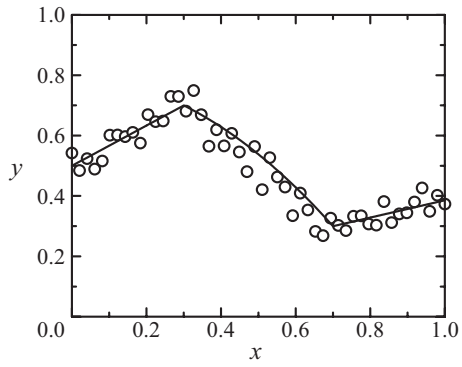


Figure 11: Results by the two Takagi-Sugeno fuzzy rules.

Example 5: We can use different types of fuzzy rules in a single fuzzy rule-based system. We show an example of such a fuzzy rule-based system in Fig. 12 where the second fuzzy rule with the antecedent fuzzy set A_2 has a consequent linear function (line (2)). Each of the other two fuzzy rules with the antecedent fuzzy sets A_1 and A_3 has a consequent real number. As we can expect, good approximation was realized by these three fuzzy rules (due to the page limitation, we can not show

ISBN: 978-989-95079-6-8

the fuzzy reasoning result). The given input-output data in Fig. 12 can be also approximated with a similar accuracy by the four simplified Takagi-Sugeno fuzzy rules in Fig. 13. Since the first two fuzzy rules with the antecedent fuzzy sets A_1 and A_2 in Fig. 13 have the same consequent real number, they can be merged into a single rule. The last two fuzzy rules with A_3 and A_4 in Fig. 13 can be also merged into a single rule. As a result, we have a fuzzy rule-based system with the two simplified Takagi-Sugeno fuzzy rules in Fig. 14.

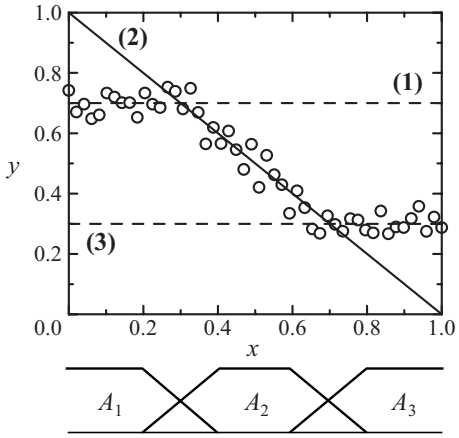


Figure 12: A Takagi-Sugeno rule and two simplified rules.

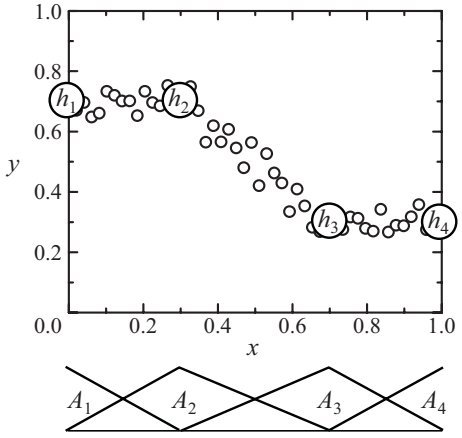


Figure 13: Four simplified Takagi-Sugeno rules.

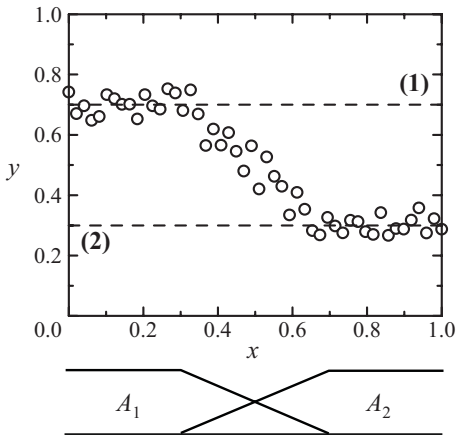


Figure 14: Two simplified Takagi-Sugeno rules.

It is clear that the fuzzy rule-based system with only the two fuzzy rules in Fig. 14 is the simplest one in Figs. 12-14. However, we have no definite answer to the question: Which is the most interpretable in the three models in Figs. 12-14?

4 Conclusions

In this paper, we demonstrated the difficulty in evaluating the interpretability of fuzzy rule-based systems. As shown in this paper, the evaluation of the interpretability is difficult even in very simple situations. Different fuzzy rule-based systems are viewed as being more interpretable according to different interpretability measures. This means that the choice of an appropriate interpretability measure is important in the design of fuzzy rule-based systems. At the same time, such a choice is difficult as shown in this paper. We hope that this paper will activate discussions on the interpretability and help to develop new approaches to fuzzy modelling based on the accuracy-interpretability tradeoff analysis.

References

[1] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms, *Fuzzy Sets and Systems* 65 (2/3) 237-253, 1994.

[2] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, Selecting fuzzy if-then rules for classification problems using genetic algorithms, *IEEE Trans. on Fuzzy Systems* 3 (3) 260-270, 1995.

[3] H. Ishibuchi, T. Murata, and I. B. Turksen, Selecting linguistic classification rules by two-objective genetic algorithms, *Proc. of 1995 IEEE International Conference on Systems, Man and Cybernetics*, 1410-1415, 1995.

[4] M. Setnes, R. Babuska, and B. Verbruggen, Rule-based modeling: Precision and transparency, *IEEE Trans. on Systems, Man, and Cybernetics - Part C* 28 (1) 165-169, 1998.

[5] J. Yen, L. Wang, and G. W. Gillespie, Improving the interpretability of TSK fuzzy models by combining global learning and local learning, *IEEE Trans. on Fuzzy Systems* 6 (4) 530-537, 1998.

[6] Y. Jin, W. von Seelen, and B. Sendhoff, On generating FC^3 fuzzy rule systems from data using evolution strategies, *IEEE Trans. on Systems, Man, and Cybernetics - Part B* 29 (6) 829-845, 1999.

[7] D. Nauck and R. Kruse, Obtaining interpretable fuzzy classification rules from medical data, *Artificial Intelligence in Medicine* 16 (2) 149-169, 1999.

[8] Y. Jin, Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement, *IEEE Trans. on Fuzzy Systems* 8 (2) 212-221, 2000.

[9] M. Setnes and H. Roubos, GA-based modeling and classification: Complexity and performance, *IEEE Trans. on Fuzzy Systems* 8 (5) 509-522, 2000.

[10] L. Castillo, A. Gonzalez, and R. Perez, Including a simplicity criterion in the selection of the best rule in a genetic fuzzy learning algorithm, *Fuzzy Sets and Systems* 120 (2) 309-321, 2001.

[11] H. Roubos and M. Setnes, Compact and transparent fuzzy models and classifiers through iterative complexity reduction, *IEEE Trans. on Fuzzy Systems* 9 (4) 516-524, 2001.

[12] J. Abonyi, J. A. Roubos, and F. Szeifert, Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization, *International Journal of Approximate Reasoning* 32 (1) 1-21, 2003.

[13] R. Alcala, J. R. Cano, O. Cordon, F. Herrera, P. Villar, and I. Zwir, Linguistic modeling with hierarchical systems of weighted linguistic rules, *International Journal of Approximate Reasoning* 32 (2-3) 187-215, 2003.

ISBN: 978-989-95079-6-8

[14] Y. C. Hu, R. S. Chen, and G. H. Tzeng, Finding fuzzy classification rules using data mining techniques, *Pattern Recognition Letters* 24 (1-3) 509-519, 2003.

[15] J. Casillas, O. Cordon, M. J. del Jesus, and F. Herrera, Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction, *IEEE Trans. on Fuzzy Systems* 13 (1) 13-29, 2005.

[16] R. Alcala, J. Alcala-Fdez, J. Casillas, O. Cordon, and F. Herrera, Hybrid learning models to get the interpretability-accuracy trade-off in fuzzy modeling, *Soft Computing* 10 (9) 717-734, 2006.

[17] R. Alcala, J. Alcala-Fdez, F. Herrera, and J. Otero, Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation, *International Journal of Approximate Reasoning* 44 (1) 45-64, 2007.

[18] E. van Broekhoven, V. Adriaenssens, and B. De Baets, Interpretability-preserving genetic optimization of linguistic terms in fuzzy models for fuzzy ordered classification: An ecological case study, *International Journal of Approximate Reasoning* 44 (1) 65-90, 2007.

[19] P. Pulkkinen and H. Koivisto, Identification of interpretable and accurate fuzzy classifiers and function estimators with hybrid methods, *Applied Soft Computing* 7 (2) 520-533, 2007.

[20] S. Destercke, S. Guillaume, and B. Charnomordic, Building an interpretable fuzzy rule base from data using orthogonal least squares - Application to a depollution problem, *Fuzzy Sets and Systems* 158 (18) 2078-2094, 2007.

[21] J. M. Alonso, L. Magdalena, and S. Guillaume, HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism, *International Journal of Intelligent Systems* 23 (7) 761-794, 2008.

[22] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems*, World Scientific, 2001.

[23] O. Cordon, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, Ten years of genetic fuzzy systems: Current framework and new trends, *Fuzzy Sets and Systems* 141 (1) 5-31, 2004.

[24] F. Herrera, Genetic fuzzy systems: Status, critical considerations and future directions, *International Journal of Computational Intelligence Research* 1 (1) 59-67, 2005.

[25] H. Ishibuchi, T. Murata, and I. B. Turksen, Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems, *Fuzzy Sets and Systems* 89 (2) 135-150, 1997.

[26] H. Ishibuchi, T. Nakashima, and T. Murata, Three-objective genetics-based machine learning for linguistic rule extraction, *Information Sciences* 136 (1-4) 109-133, 2001.

[27] H. Ishibuchi and T. Yamamoto, Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining, *Fuzzy Sets and Systems* 141 (1) 59-88, 2004.

[28] H. Wang, S. Kwong, Y. Jin, W. Wei, and K. F. Man, Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction, *Fuzzy Sets and Systems* 149 (1) 149-186, 2005.

[29] H. Ishibuchi and Y. Nojima, Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, *International Journal of Approximate Reasoning* 44 (1) 4-31, 2007.

[30] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni, A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems, *Soft Computing* 11 (11) 1013-1031, 2007.

[31] A. F. Gomez-Skarmeta, F. Jimenez, and G. Sanchez, Improving interpretability in approximative fuzzy models via multiobjective evolutionary algorithms, *International Journal of Intelligent Systems* 22 (9) 943-969, 2007.

[32] R. Alcala, M. J. Gacto, F. Herrera, and J. Alcala-Fdez, A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems, *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 15 (5) 539-557, 2007

[33] J. Gonzalez, I. Rojas, H. Pomares, L. J. Herrera, A. Guillen, J. M. Palomares, and F. Rojas, Improving the accuracy while preserving the interpretability of fuzzy function approximators by means of multi-objective evolutionary algorithms, *International Journal of Approximate Reasoning* 44 (1) 32-44, 2007

[34] M. J. Gacto, R. Alcala, and F. Herrera, Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems, *Soft Computing* 13 (5) 419-436, 2009.

[35] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu, Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index, *Soft Computing* 13 (5) 437-449, 2009.

[36] L. Sanchez, J. Otero, and I. Couso, Obtaining linguistic fuzzy rule-based regression models from imprecise data with multiobjective genetic algorithms, *Soft Computing* 13 (5) 467-479, 2009.

[37] H. Ishibuchi, Multiobjective genetic fuzzy systems: Review and future research directions, *Proc. of 2007 IEEE International Conference on Fuzzy Systems*, 913-918, 2007.

[38] Y. Jin (ed.), *Multi-Objective Machine Learning*, Springer, Berlin, 2006.

[39] A. Ghosh, S. Dehuri, and S. Ghosh (Eds.), *Multi-Objective Evolutionary Algorithms for Knowledge Discovery from Data Bases*, Springer, Berlin, 2008.

[40] K. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer, Boston, 1998.

[41] A. P. Wierzbicki, M. Makowski, and J. Wessels (eds.), *Model-Based Decision Support Methodology with Environmental Applications*, Kluwer, Dordrecht, 2000.

[42] K. Miettinen and M. Makela, On scalarizing functions in multiobjective optimization, *OR Spectrum* 24 (2) 193-213, 2002.

[43] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. on Evolutionary Computation* 6 (2) 182-197, 2002.

[44] E. Zitzler and L. Thiele, Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach, *IEEE Trans. on Evolutionary Computation* 3 (4) 257-271, 1999.

[45] E. Zitzler, M. Laumanns, and L. Thiele, SPEA2: Improving the strength Pareto evolutionary algorithms, *TIK-Report 103*, Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH) Zurich, 2001.

[46] S. Guillaume, Designing fuzzy inference systems from data: An interpretability-oriented review, *IEEE Trans. on Fuzzy Systems* 9 (3) 426-443, 2001.

[47] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Interpretability Issues in Fuzzy Modeling*, Springer, 2003.

[48] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Accuracy Improvements in Linguistic Fuzzy Modeling*, Springer, 2003.

[49] R. Mikut, J. Jakel, and L. Groll, Interpretability issues in data-based learning of fuzzy systems, *Fuzzy Sets and Systems* 150 (2) 179-197, 2005.

[50] C. Mencar, G. Castellano, and A. M. Fanelli, On the role of interpretability in fuzzy data mining, *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 15 (5) 521-537, 2007.

[51] C. Mencar and A. M. Fanelli, Interpretability constraints for fuzzy information granulation, *Information Sciences* 178 (24) 4585-4618, 2008.

[52] S. M. Zhou and J. Q. Gan, Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling, *Fuzzy Sets and Systems* 159 (23) 3091-3131, 2008.

ISBN: 978-989-95079-6-8

Stability Analysis of Continuous PI-like Fuzzy Control Systems based on Vector Norms Approach

Anis Sakly[1] Basma Zahra[1] Mohamed Benrejeb[1]

[1]Unité de recherche L.A.R.A. Automatique[1]
Ecole Nationale d'Ingénieurs de Tunis[1]
BP 37, Belvédère 1002 Tunis, Tunisie
sakly_anis@yahoo.fr, zahra_basma@yahoo.fr, mohamed.benrejeb@enit.rnu.tn

Abstract—*This paper proposes a practical stability analysis of a particular class of continuous PI-like fuzzy control systems based on the convergence of regular vector norms. This approach is based on the comparison, the overvaluing principle and the application of Borne and Gentina criterion. The stability conditions issued from vector norms correspond to a vector Lyapunov function. A comparison system relative to a regular vector norm will be used in order to get the simple arrow form of the state matrix that yields to a suitable use of Borne and Gentina criterion for establishing sufficient conditions for global asymptotic stability.*

Keywords—Continuous PI-like fuzzy control systems, global asymptotic stability, vector norms, Borne and Gentina criterion, Arrow form state matrix.

1 Introduction

Over the last few decades, fuzzy control has received great attention from engineering and science community and has marked rapid growth. However, there isn't a general method for analysis or synthesis of such control strategy. In particular, stability analysis of control systems is an essential step before synthesizing and elaborating the control law.

In this way, many researches have been done on the stability study of fuzzy control systems since their appearance in the middle of the 70's [17].

Indeed, the majority of these studies concern TSK fuzzy systems since consequent fuzzy variables of this type (known as type III) are explicit and many important results have been obtained [13, 18, 23] based on Lyapunov stability theory. Nevertheless, there have been a limited number of stability studies of Mamdani type fuzzy systems (known as type I) due to their complexity, and for this type, fuzzy variables are linguistically understandable in both the premises and consequents. Most of these papers provide a stability analysis of a linear plant controlled by a fuzzy controller and they regard the latter as a nonlinear controller corresponding to a Lur'e system. So, the stability problem of fuzzy control systems comes down to conventional nonlinear stability theory.

Among the methods that have been used, Popov's theorem for time-invariant nonlinearity [16], circle criterion for time-variant nonlinearity [19, 20] and its extended version, the conicity criterion [1].

Other approaches have been used: ones can cite: the hyperstability approach which is equivalent to passivity [9, 10, 24], the input-output stability based on the use of small gain theorem [11, 25] and the Lyapunov function approach [8, 12].

We do remark also that the majority of these papers simplify the study by considering the consequents of the fuzzy system as singletons [22]. So, this type (known as type II) is a special case of both TSK fuzzy systems and those of Mamdani.

In this paper, the stability study of particular class of fuzzy PI controllers of type Mamdani is presented. This study is based on the application of the Borne and Gentina criterion [6] which uses Kotelyanski conditions. In [3], it has been shown that when system state matrix is in arrow form, then Borne and Gentina criterion becomes very simple to apply.

This approach has been used in many previous works [2, 4, 5, 21]. In this way, results proposed in [21] will be used and generalized for stability analysis of continuous Mamdani fuzzy systems in the case of nonlinear processes to be controlled.

The paper is organised as follows. The next section is devoted to the description of the particular class of PI-like fuzzy control system. In section 3, stability conditions of the fuzzy system are established by using Borne and Gentina criterion and vector norms approach. The stability conditions proposed are illustrated with an example presented in section 4. Finally, concluding remarks are drown in section 5.

2 Particular class of PI-like fuzzy controllers

The fuzzy PI control system considered in this study has two inputs the error e and its derivative de and one output the control derivative as shown in Fig. 1, where k_e, k_{de} et k_{du} are scale factors.

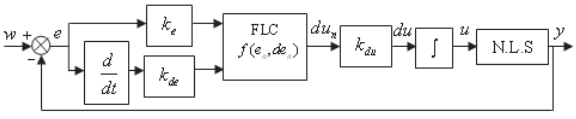


Figure 1: Fuzzy control system.

A particular class of fuzzy PI controllers of type Mamdani is obtained by considering a strong triangular partition of the normalized variables e_n, de_n and du_n presented in Fig. 2.

The rule base considered is an r×r traditional rule table that is of antidiagonal type such that the Mac Vicar-Whelan one (Table 1).

ISBN: 978-989-95079-6-8

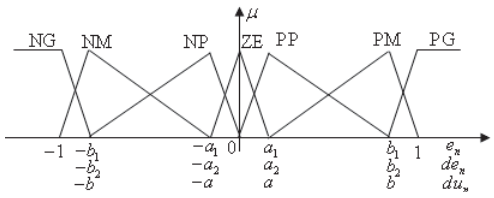


Figure 2: Fuzzy subsets partition.

Table 1 : The Mac Vicar-Whelan rule base.

e_n \ de_n	NG	NM	NP	ZE	PP	PM	PG
NG	NG	NG	NG	NG	NM	NP	ZE
NM	NG	NG	NM	NM	NP	ZE	PP
NP	NG	NM	NP	NP	ZE	PP	PM
ZE	NG	NM	NP	ZE	PP	PM	PG
PP	NM	NP	ZE	PP	PP	PM	PG
PM	NP	ZE	PP	PM	PM	PG	PG
PG	ZE	PP	PM	PG	PG	PG	PG

Let $\sigma(e_n, de_n)$ the surface in the space (e_n, de_n, du_n), verifying the two properties [21]:

$$\begin{cases} \text{i) If } \sigma = 0 \text{ then the input-output characteristic surface } du_n(e_n, de_n) = 0 \\ \text{ii) It exists } k>0 \text{ such as } du_n(k\sigma - du_n) \geq 0 \text{ for all } e_n \text{ and } de_n. \end{cases} \quad (1)$$

The first property means that the intersection of the overvaluing surface σ with the plan (e_n, de_n) is a part of the intersection of the characteristic surface $du_n(e_n, de_n)$ with the same plan. The curve $\sigma = 0$ is a straight line when the fuzzy input partition is identical for the inputs and it represents the second bissectrix as shown in Fig. 3.

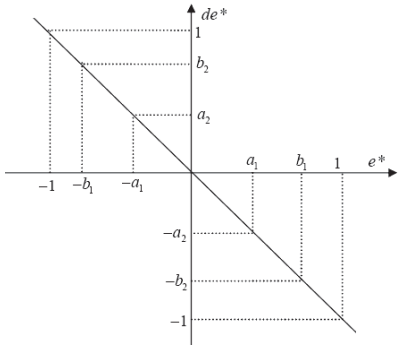


Figure 3: Curve $\sigma = 0$.

According to [21], in the case of an identical partition for the two inputs e_n and de_n (i.e. $a_1 = a_2$ and $b_1 = b_2$) and when using an antidiagonal rule table like the Mac-Vicar Whelan one. This approach is based on the overvaluing and the undervaluing of the fuzzy controller characteristic surface $du_n = (e_n, de_n)$ by two plans crossing the plan (e_n, de_n) in the second bissectrix which is the curve $\sigma = 0$, in this case given by the following equation:

$$e_n + de_n = 0 \quad (3)$$

The slopes of these plans are respectively $k_{\max}$ and $k_{\min}$, those values depend on the geometric parameters of the input variables e_n and de_n (a_i and $b_i, i = 1,2$) and the output variable du_n (a and b), on the other hand they depend on the used inference method.

In this way for each point (e_n, de_n) of $[-1,1]\times[-1,1]$ the controller output du_n is such that:

$$k_{\min}(e_n + de_n) \leq du_n \leq k_{\max}(e_n + de_n).$$

Otherwise $du_n = f(.)(e_n + de_n)$, in such way we can write:

$$k_{\min} \leq f(.) \leq k_{\max} \quad (4)$$

where $f(.)$ is a nonlinear gain.

By programming, we can obtain values of $k_{\max}$ and $k_{\min}$ that allow respectively overvaluing and undervaluing the characteristic surface of the fuzzy controller by PI plans.

In particular, for equidistant fuzzy partition of variables e_n, de_n and du_n ($a_1 = a_2 = a = 0.33$ and $b_1 = b_2 = b = 0.67$), max-min inference method and centroid defuzzification method, we obtain the action surface for the fuzzy controller given by Fig. 4.

This surface is obtained by distorting the discourse universes of e_n and de_n to 100 points, we obtain:

$$k_{\max} = 1,86 \text{ and } k_{\min} = 0,47 \quad (5)$$

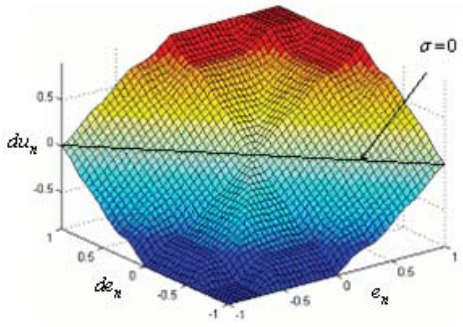


Figure 4: Action surface of the particular Mamdani fuzzy controller.

So the PI fuzzy control system of the Fig. 1 can be set in the following form (Fig. 5):

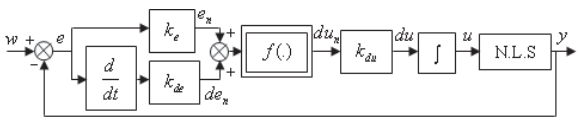


Figure 5: Equivalent fuzzy control system.

3 Proposed stability conditions

3.1 Problem formulatiom

Supposing that the system to be controlled which is nonlinear is represented by the following state matrix given in the Frobenius form such that:

$$\begin{cases} \dot{x} = A(.)x + B(.)u \\ y = C(.)x \end{cases} \quad x \in \mathbb{R}^n \quad (6.a)$$

where:

$$A(.) = \begin{bmatrix} 0 & \dots & 0 & -a_n(.) \\ 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & 0 & \vdots \\ 0 & \dots & 1 & -a_1(.) \end{bmatrix}, B(.) = \begin{bmatrix} b_n(.) \\ \vdots \\ \vdots \\ b_1(.) \end{bmatrix}, C^T(.) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (6.b)$$

From the diagram given in Fig. 5, we have $du = \dot{u}$, supposing that $v = du = \dot{u}$ and $\xi = u$, which leads to : $\dot{\xi} = v$, the nonlinear system equipped with the integration can be represented by the following state matrix:

ISBN: 978-989-95079-6-8

$$\dot{z} = \begin{bmatrix} A(.) & B(.) \\ 0 & 0 \end{bmatrix} z + \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} v \quad z \in \mathbb{R}^{n+1} \tag{7}$$

where: $z = \begin{bmatrix} x \\ \xi \end{bmatrix}$ and $\dot{z} = \begin{bmatrix} \dot{x} \\ \dot{\xi} \end{bmatrix}$

supposing that: $A'(.) = \begin{bmatrix} A(.) & B(.) \\ 0 & 0 \end{bmatrix}$ and $B'(.) = B' = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}$.

So we can write $y = x_n = z_n$ where x_n (respectively z_n) is the n^{th} state variable of state vector x (respectively z), thus: $y = C'(.)z$ where $C'(.) = C' = \begin{bmatrix} 0 & \dots & 1 & 0 \end{bmatrix}$

On the other hand, and by referring to the diagram in Fig. 5 in the autonomous regime, we can write:

$$\begin{aligned} v &= k_{du} f(.)\left[e_n + de_n\right] = -k_{du} f(.)\left[k_e e + k_{de}\dot{e}\right] \\ &= -k_{du} f(.)\left[k_e y + k_{de}\dot{y}\right] \end{aligned} \tag{8}$$

finally: $v = -k_{du} f(.)\left[k_e C'(.)z + k_{de} C'(.)\dot{z}\right]$

then:

$$\begin{aligned} \dot{z} &= A'(.)z + B'(.)v = A'(.)z - B'(.)k_{du} f(.)\left[k_e C'z + k_{de} C'\dot{z}\right] \\ &= A'(.)z - k_{du} f(.)\left[k_e B'C'z + k_{de} B'C'\dot{z}\right] \end{aligned}$$

that leads to:

$$\left[I_{n+1} + k_{du}k_{de} f(.)B'C'\right]\dot{z} = \left[A'(.) - k_{du}k_e f(.)B'C'\right]z \tag{9}$$

supposing now: $\begin{cases} N = I_{n+1} + k_{du}k_{de} f(.)B'C' \\ M = A'(.) - k_{du}k_e f(.)B'C' \end{cases}$

let:

$$N = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & \ddots & 0 \\ 0 & \dots & 0 & k_{du}k_{de} f(.) & 1 \end{bmatrix}, \; M = \begin{bmatrix} 0 & \dots & 0 & -a_n(.) & b_n(.) \\ 1 & \ddots & & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & 1 & -a_1(.) & b_1(.) \\ 0 & \dots & 0 & -k_{du}k_e f(.) & 0 \end{bmatrix}$$

where: $\det(N) = 1$ and $N^{-1} = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & \ddots & 0 \\ 0 & \dots & 0 & -k_{du}k_{de} f(.) & 1 \end{bmatrix}$

Finally we obtain the following description of the closed loop system:

$$\dot{z} = A_C(.)z \tag{10}$$

where:

$$A_C(.) = N^{-1}M = \begin{bmatrix} 0 & \dots & \dots & -a_n(.) & b_n(.) \\ 1 & & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & 1 & -a_1(.) & b_1(.) \\ 0 & \dots & -k_{du}k_{de} f(.) & k_{du} f(.)(k_{de}a_1(.) - k_e) & -k_{du}k_{de} f(.)b_1(.) \end{bmatrix} \tag{11}$$

For establishing stability conditions for the studied system, we have to make a basic change leading to new representation of the system to get best exploitation.

3.2 *New state representation of the system*

We consider the following passage matrix P allowing passing from the matrix $A_C(.)$ to a matrix $A_C^{'}(.)$, such that:

$$P = \begin{bmatrix} 1 & \alpha_1 & \dots & (\alpha_1)^{n-1} & 0 \\ \vdots & \dots & & \vdots & \vdots \\ 1 & \alpha_{n-1} & \dots & (\alpha_{n-1})^{n-1} & \vdots \\ 0 & \dots & & 1 & 0 \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix} \tag{12}$$

We note: $z' = Pz$ and let $z = P^{-1}z'$.

In this way:

$$\dot{z}' = PA_C(.)P^{-1}z' = A_C^{'}(.)z' \tag{13.a}$$

where: $A_C^{'}(.) = PA_C(.)P^{-1}$

the matrix $A_C^{'}(.)$ is given by:

$$A_C^{'}(.) = \begin{bmatrix} \alpha_1 & 0 & \dots & \delta_1(.) & \upsilon_1(.) \\ \vdots & \ddots & & \vdots & \vdots \\ 0 & \dots & \alpha_{n-1} & \delta_{n-1}(.) & \upsilon_{n-1}(.) \\ \beta_1 & \dots & \beta_{n-1} & \sigma(.) & b_1(.) \\ \lambda_1(.) & \dots & \lambda_{n-1}(.) & \varphi(.) & \psi(.) \end{bmatrix} \tag{13.b}$$

where:

$$\beta_i = \frac{1}{\prod_{\substack{j=1 \\ j \neq i}}^{n} (\alpha_i - \alpha_j)} \quad \forall i = 1,...,n-1 \tag{13.c}$$

$$\begin{cases} \delta_i(.) = -D(.,\alpha_i) & \forall i = 1,...,n-1 \\ D(.,\alpha) = \alpha^n + \sum_{j=1}^{n} a_j(.)\alpha^{n-j} \end{cases} \tag{13.d}$$

$$\begin{cases} \upsilon_i(.) = N(.,\alpha_i) & \forall i = 1,...,n-1 \\ N(.,\alpha) = \sum_{j=1}^{n} b_j(.)\alpha^{n-j} \end{cases} \tag{13.e}$$

$$\sigma(.) = -a_1(.) - \sum_{j=1}^{n-1} \alpha_j \tag{13.f}$$

$$\begin{cases} \lambda_i(.) = -k_{du}k_{de} f(.)\beta_i & \forall i = 1,...,n-1 \\ \varphi(.) = k_{du}k_{de} f(.)\left[a_1(.) + \sum_{j=1}^{n-1} \alpha_j\right] - k_{du}k_e f(.) \\ \psi(.) = -k_{du}k_{de} f(.)b_1(.) \end{cases} \tag{13.g}$$

By observing the matrix $A_C^{'}(.)$, we remark that the nonnull elements are located in the diagonal and in both two last rows and two last columns, this matrix is called in the double arrow form. The nonlinear elements are situated in the last row and the two last columns.

We can also deduce that the Borne and Gentina criterion can not be applied in this case, only the usual stability criteria can be used like Holder norms such that max norm, sum norm and module norm.

In order to use suitably the Borne and Gentina criterion, we can think of the isolation of the nonlinear elements in only one row and one column. In this way, we consider a comparison system relative to the following n regular vector norm:

ISBN: 978-989-95079-6-8

$$p(z') = \left[|z'_1|, \ldots, |z'_{n-1}|, \max\left\{ |z'_n|, |z'_{n+1}| \right\} \right]^T \quad (14)$$

The vector z' is of order $(n+1)$ and the vector $p(z')$ is of order n.

Let $Z = p(z')$, so we define the overvaluing system relative to p such that:

$$\dot{Z} = M_C(.)Z \quad (15.a)$$

The matrix $M_C(.)$ is given by:

$$M_C(.) = \begin{bmatrix} \alpha_1 & 0 & \ldots & \gamma_1(.) \\ 0 & \ddots & & \vdots \\ \vdots & \ldots & \alpha_{n-1} & \gamma_{n-1}(.) \\ \mu_1(.) & & \mu_{n-1}(.) & \mu(.) \end{bmatrix} \quad (15.b)$$

where:

$$\gamma_i(.) = |\delta_i(.)| + |\upsilon_i(.)| \qquad \forall i = 1,...,n-1$$
$$= \left| \alpha_i^n + \sum_{j=1}^{n} a_j(.)\alpha_i^{n-j} \right| + \left| \sum_{j=1}^{n} b_j(.)\alpha_i^{n-j} \right| \quad (15.c)$$

$$\mu_i(.) = \max\left\{ |\beta_i|, |\lambda_i(.)| \right\} = |\beta_i| \max\left\{ 1, k_{du}k_{de}f(.) \right\} \quad (15.d)$$

$$\mu(.) = \max\left\{ \sigma(.) + |b_1(.)|, |\varphi(.)| + \psi(.) \right\}$$
$$= \max \left\{ \begin{array}{l} (-a_1(.) - \sum_{j=1}^{n-1} \alpha_j) + |b_1(.)|, \\ \left| k_{du}k_{de}f(.) \left[a_1(.) + \sum_{j=1}^{n-1} \alpha_j \right] - k_{du}k_e f(.) \right| - k_{du}k_{de}f(.)b_1(.) \end{array} \right\} \quad (15.e)$$

The matrix $M_C(.)$ is in the arrow form. The vector norm p allows passing from the matrix in the double arrow form to a matrix in the simple arrow form by decreasing the order of the system from $(n+1)$ to n order.

To apply the Borne and Gentina criterion to the overvaluing matrices, we substitute the nonlinear elements in the last row $(\mu_i(.))$ by constant elements with the following hypothesis:

$$\mu_i(.) = |\beta_i| \text{ for } k_{du}k_{de}f(.) < 1 \quad (16)$$

In these conditions the matrix $M_C(.)$ becomes:

$$M'_C(.) = \begin{bmatrix} \alpha_1 & 0 & \ldots & \gamma_1(.) \\ 0 & \ddots & & \vdots \\ & & \alpha_{n-1} & \gamma_{n-1}(.) \\ |\beta_1| & \ldots & |\beta_{n-1}| & \mu(.) \end{bmatrix} \quad (17)$$

So the nonlinear elements of the matrix $M'_C(.)$ are isolated in the last column.

3.3 Stability conditions

Theorem 1:

If there exist $\alpha_i < 0$ for $i = 1,...,n-1$, $\alpha_i \neq \alpha_j \ \forall i \neq j$ such that $\forall Z \in S$ where S is a neighbourhood domain:

$$i)\ k_{du}k_{de}f(.) < 1$$
$$ii) - \mu(.) + \sum_{i=1}^{n-1} \gamma_i(.)\alpha_i^{-1}|\beta_i| > 0 \quad (18)$$

then the equilibrium point $Z=0$ for the continuous system is asymptotically stable.

If $S = \mathbb{R}^n$, the stability is global. ■

Proof:

The matrix $M'_C(.)$ has its off diagonal elements positive and the ones non constant are isolated in the last line.

Let the following comparison system: $\dot{Z} = M'_C(.)Z$

Thus, by referring to results obtained in [3], the conditions of the previous theorem can be deduced from the Kotelyanski conditions [6]. These conditions require having the principal minors with alternating signs (see Appendix), the (α_i) are chosen all negative.

$\alpha_1 < 0$

$\alpha_1\alpha_2 > 0$

$\vdots$

$(-1)^{n-1} \prod_{i=1}^{n-1} \alpha_i > 0$

and $(-1)^n \det(M_C) > 0$

The $(n-1)$ first conditions are checked because the α_i are negative, however the last condition yields to:

$$(-1)^n \det\left(M'_C(.)\right) = (-1)^n \begin{vmatrix} \alpha_1 & 0 & \ldots & 0 & \gamma_1(.) \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \ldots & 0 & \alpha_{n-1} & \gamma_{n-1}(.) \\ |\beta_1| & \ldots & \ldots & |\beta_{n-1}| & \mu(.) \end{vmatrix}$$
$$= (-1)^n \left[\mu(.) \prod_{i=1}^{n-1} \alpha_i - \sum_{i=1}^{n-1} \left(|\gamma_i(.)\beta_i| \prod_{\substack{j=1 \\ j \neq i}}^{n-1} \alpha_j \right) \right] > 0$$

then the theorem is obtained by dividing this condition by $\left((-1)^{n-1} \prod_{i=1}^{n-1} \alpha_i \right)$ such that: $-\mu(.) + \sum_{i=1}^{n-1} \gamma_i(.)\alpha_i^{-1}|\beta_i| > 0$. ■

In order to simplify the application of the theorem, we have to exploit the expression of $\mu(.)$. We suppose that:

$$\mu(.) = \left| k_{du}k_{de}f(.) \left[a_1(.) + \sum_{j=1}^{n-1} \alpha_j \right] - k_{du}k_e f(.) \right| - k_{du}k_{de}f(.)b_1(.).$$

Thus, we obtain the following corollary.

Corollary 1:

If there exist $\alpha_i < 0$ for $i = 1,...,n$, $\alpha_i \neq \alpha_j \ \forall i \neq j$ such that $\forall Z \in S$ the three following conditions are checked:

$$i)\ k_{du}k_{de}f(.) < 1$$
$$ii)\left(-a_1(.) - \sum_{j=1}^{n-1} \alpha_j \right) + |b_1(.)| < \mu(.) \quad (19)$$
$$iii) - \mu(.) + \sum_{i=1}^{n-1} \gamma_i(.)\alpha_i^{-1}|\beta_i| > 0$$

then the equilibrium point $Z=0$ for the system is asymptotically stable.

If $S = \mathbb{R}^n$, the stability is global. ■

Remark 1 :

For $n = 1$, the stability condition issued from Theorem 1 is such that: $\mu(.) < 0$. ■

ISBN: 978-989-95079-6-8

4 Example

For the validation of the results obtained we consider the stabilization of a fuzzy control system, where the controller is of type PI fuzzy controller and the system to be controlled is a speed of DC motor. This motor is supposed with a shunt excitation represented by a transfer function of a first order system which is preceded by a nonlinear element, corresponding to nonlinearity characteristic of the magnetic flux for example (Fig. 6).

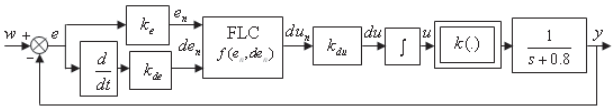


Figure 6: Fuzzy control system for the speed of DC motor with shunt excitation.

The nonlinear gain $k(.)$ is represented by the following allure (Fig. 7):

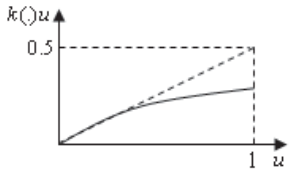


Figure 7: Characteristic of the nonlinear gain.

The system is described by the following state equation:

$$\begin{cases} \dot{x} = A(.)x + B(.)u \\ y = C(.)x \end{cases} \quad (20.a)$$

where:

$$\begin{cases} A(.) = A = -0.8 \\ B(.) = k(.) \\ C(.) = C = 1 \end{cases} \quad (20.b)$$

the matrix $A_C(.)$ is given by :

$$A_C(.) = \begin{bmatrix} -0.8 & k(.) \\ 0.8k_{du}k_{de}f(.) - k_{du}k_{de}f(.) & -k_{du}k_{de}f(.)k(.) \end{bmatrix} \quad (21)$$

By making a basic change to the previous system we obtain:

$$\dot{z}' = A'_C(.)z' \quad (22.a)$$

Such that:

$$\begin{cases} A'_C(.) = PA_C(.)P^{-1} \\ P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2 \end{cases} \quad (22.b)$$

so we can write: $A'_C(.) = A_C(.)$ and:

$$M_C(.) = \max\left\{\left|k_{du}k_{de}f(.)0.8 - k_{du}k_{de}f(.)\right| - k_{du}k_{de}f(.)k(.), -0.8 + k(.)\right\}$$

the factor scales k_e and k_{de} are chosen such that:

$k_e = k_{de} = 1$.

According to Remark 1, by applying Theorem 1 we get the following condition: $\mu(.) < 0$

which leads to:

$$\begin{cases} -0.8 + k(.) < 0 \\ k_{du}f(.)\left[0.2 - k(.)\right] < 0 \end{cases} \text{ and so: } \begin{cases} 0.2 < k(.) < 0.8 \\ k_{du}f(.) > 0 \end{cases}$$

The area representing the stability domain of the fuzzy system is given by Fig. 8:

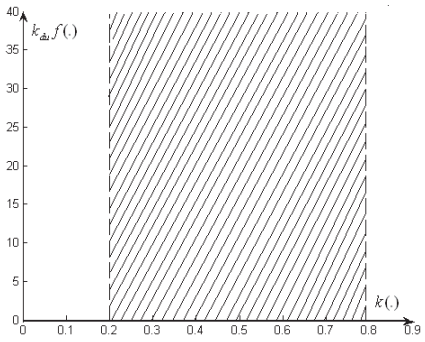


Figure 8: Stability domain of the fuzzy system obtained from Theorem 1.

The application of Corollary 1 allows to deduce the following condition: $-0.8 + k(.) < k_{du}f(.)\left[0.2 - k(.)\right] < 0$

then: $k_{du}f(.) < \dfrac{0.8 - k(.)}{k(.) - 0.2}$ and $0.2 < k(.) < 0.8$.

The area representing the stability domain of the fuzzy system is given by Fig. 9:

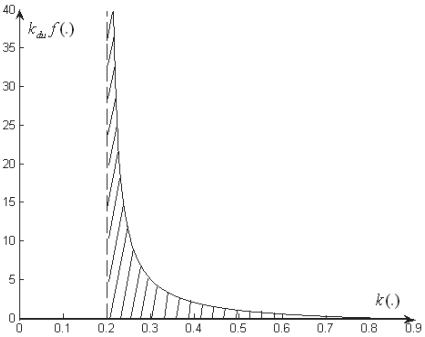


Figure 9: Stability domain of the fuzzy system obtained from Corollary 1.

5 Conclusion

The stability conditions of particular class of PI-like fuzzy control systems were presented in this paper. These conditions were deduced from stability study of overvaluing systems based on vector norms and the application of the Borne and Gentina criterion. After making a basic change on the system, we obtain a state matrix in a double arrow form, to return to the usual arrow form of the matrix and to get matrix with nonlinear elements isolated in only one row or column, we had considered a comparison system relative to a regular vector norm. In this way the Borne and Gentina criterion was used to get sufficient stability conditions. These conditions were applied to nonlinear system given by the control of the speed of DC motor with shunt excitation.

Appendix

Borne-Gentina practical stability criterion [15]

Let consider the nonlinear continuous process described in state space by: $\dot{x} = A(.)x$; $A(.)$ is an $n \times n$ matrix, $A(.) = \{a_{i,j}\}$. If the overvaluing matrix $M(A(.))$ has its non constant elements isolated in only one row, the verification of the Kotelyanski condition enables to conclude to the stability of the initial system.

As an example, if the non constant elements are isolated in only one row of $A(.)$, Kotelyanski lemma applied to the

ISBN: 978-989-95079-6-8

overvaluing matrix obtained by the use of the n regular vector norm $p(x)$ with $x=[x_1,x_2,...,x_n]^T$, such as: $p(x)=[|x_1|,|x_2|,...,|x_n|]^T$, leads to the following stability conditions of initial system:

$$a_{1,1}<0,\begin{vmatrix} a_{1,1} & |a_{1,2}| \\ |a_{2,1}| & a_{2,2} \end{vmatrix}>0,...,(-1)^n\begin{vmatrix} a_{1,1} & |a_{1,2}| & \dots & |a_{1,n}| \\ |a_{2,1}| & a_{2,2} & \dots & |a_{2,n}| \\ \vdots & \vdots & & \vdots \\ |a_{n,1}(.)| & |a_{n,2}(.)| & \dots & a_{n,n}(.) \end{vmatrix}>0$$

The Borne-Gentina practical criterion applied to continuous systems generalizes the Kotelyanski lemma for nonlinear systems and defines large classes of systems for which the linear Aizerman conjecture can be applied, either for the initial system or for its comparison system.

References

[1] J. Aracil, A. Garcia-Cerezo, A. Barreiro and A. Ollero , *Fuzzy Control of dynamical systems and stability analysis based on the conicity criterion*, IFSA World Congress, Brussels, Belgium, 1991.

[2] M. Benrejeb, M. Gasmi, *Stability and Stabilization Conditions of TSK Models using Vector Norms*, International Conference on Communication, Computer and Power (ICCCP'01), Muscat, Février 2001.

[3] M. Benrejeb, P. Borne, F. Laurent, *Sur une application de la représentation en flèche à l'analyse des processus*, RAIRO Automatique, 16 (2), pp 133-146, 1982.

[4] M. Benrejeb, A. Sakly, K. Ben Othman and P. Borne, *Choice of conjunctive operator of TSK fuzzy systems and stability domain study*, Math. And Comp. in Simul., Vol 76, pp 410-421, 2008.

[5] M. Benrejeb, D. Soudani, A. Sakly and P. Borne, *New discrete TSK fuzzy systems characterization and stability domain*, Int. Jour. Of Comp. Comm. and Cont., Vol 1, N 4, pp 9-19, 2006.

[6] P. Borne, *Non linear sytem stability*, vector norm approach, System and control Enceclopedia, t.5, 3402-3406, Perganon Press, 1987.

[7] P. Borne, M. Benrejeb, *On the Representation and the Stability Study of Large Scale Systems*, IJCCC, International Journal of Computers, Communications & Control, Vol 3, Supp. Issue: Proceedings of ICCC, pp 55-66, 2008.

[8] F. Bouslama and A. Ishikawa, *Application of limit fuzzy controllers to stability analysis*, Fuzzy Sets and Systems, Vol 49, pp 103-120, 1992.

[9] G. Calcev, *Some remarks on the stability of Mamdani fuzzy control systems*, IEEE Trans. on Fuzzy Systems, Vol 6, Issue 3, pp 436-442, 1998.

[10] G. Calcev, R. Gorez and M.D. Neyer, *Passivity approach of fuzzy control systems*, Automatica, Vol 34, N 3, pp 339-344, 1998.

[11] G. Chen and H. Ying, *Stability analysis of nonlinear fuzzy PI control systems*, Int. Conf. Fuzzy Logic Applict., Houston, TX, 1993.

[12] Y.Y. Chen, *Stability analysis of fuzzy control-A Lyapunov approach*, IEEE Conf. Syst. Man Cybern., Las Vegas, NV, 1987.

[13] H. Feng-Hsiag, H. Jing-Dong, *Stability analysis of fuzzy large-scale systems*, IEEE Trans. on Systems, Man, and Cybernetics, Part B, Vol 32, Issue 1, pp 122-126, 2002.

[14] C.-C. Fuh, P.-C. Tung, *Robust Stability analysis of fuzzy control System*, Fuzzy Sets and Systems, pp 289-298, 1997.

[15] J.C. Gentina, P. Borne, F. Laurent, "Stabilité des systèmes continus non linéaires de grande dimension, " *RAIRO*, Août, J-3, pp 69-77, 1972.

[16] R. Katoh, T. Yamashita and S. Singh, *Stability Analysis of control system having PD type of fuzzy controller*, Fuzzy Sets and Systems, Vol 74, Issue 2, pp 321-334, 1995.

[17] W.M.J. Kickert and E.H. Mamdani, *Analyses of Fuzzy Logic Controllers*, Fuzzy Sets and Systems, Vol.1, pp 29-44, 1978.

[18] J. Ch. Lo, Y. M. Chen, *Stability issues on Takagi-Sugeno fuzzy model-parametric approach*, IEEE Trans. on Fuzzy Systems, Vol 7 Issue 5, pp 597-607, 1999.

[19] K.S. Ray and D.D. Majumder, *Application of Cercle Criteria for Stability Analysis of SISO and MIMO Systems associated with Fuzzy Logic Controllers*, IEEE Trans. Syst. Man & Cyber., Vol 14, Issue 2, pp 345-349, 1984.

[20] A. Sakly, K. Ben Othman and M. Benrejeb, *Mamdani Type Fuzzy Controllers for the Stabilisation of Linear Systems*, WSEAS Transactions on Systems, Issue 4, Volume 2, pp.820-826, 2003.

[21] A. Sakly, B. Zahra and M. Benrejeb, *Stability Study of Mamdani's Fuzzy Controllers applied to Linear Plants*, Studies in Informatics and Control, SIC Journal , vol 17, Issue 4, December 2008.

[22] M. Sugeno , *On stability of fuzzy systems expressed by fuzzy rules with singleton consequences*, IEEE Trans. on Fuzzy Systems, Vol 7, Issue 2, pp 201-223, 1999.

[23] K. Tanaka, M. Sano, *A Robust Stabilization Problem of Fuzzy Control Systems and its Application to Backing up Control of a Track-Trailer*, IEEE Trans. on Fuzzy Systems, Vol 2, Issue 2, pp 119-134, 1994.

[24] L. Wang and R. Langari, *Fuzzy controller design via hyperstability approach*, IEEE Int. Conf. Fuzzy Syst., Orlando, FL, 1994.

[25] H. Ying, *Conditions on general Mamdani fuzzy controllers as nonlinear, variable gain state feedback controllers with stability analysis*, IFSA World Congress and 20th NAFIPS International Conference, 2001.

ISBN: 978-989-95079-6-8

The Fuzzy Induced Generalized OWA Operator and its Application in Business Decision Making

José M. Merigó, Anna M. Gil-Lafuente

Department of Business Administration, University of Barcelona
Av. Diagonal 690, 08034 Barcelona, Spain
Email: jmerigo@ub.edu, amgil@ub.edu

Abstract—*We present the fuzzy induced generalized OWA (FIGOWA) operator. It is an aggregation operator that uses the main characteristics of the fuzzy OWA (FOWA) operator, the induced OWA (IOWA) operator and the generalized OWA (GOWA) operator. Therefore, it uses uncertain information represented in the form of fuzzy numbers, generalized means and order inducing variables. The main advantage of this operator is that it includes a wide range of mean operators in the same formulation such as the FOWA, the IOWA, the GOWA, the induced GOWA, the fuzzy IOWA, the fuzzy generalized mean, etc. We study some of its main properties. A further generalization by using quasi-arithmetic means is also presented. This operator is called Quasi-FIOWA operator. We also develop an application of the new approach in a strategic decision making problem.*

Keywords— Decision making; OWA operator; Aggregation operators; Fuzzy numbers.

1 Introduction

Different types of aggregation operators are found in the literature for aggregating the information. A very common aggregation method is the ordered weighted averaging (OWA) operator [17]. Since its appearance, the OWA operator has been studied in a wide range of applications [1-3,5,7-8,10-13,15-22]. In [21], Yager and Filev introduced the IOWA operator. It is a generalization of the OWA operator that uses order inducing variables in the reordering of the arguments. In the last years, the IOWA operator has been studied by different authors [5,11-13,16,19,21].

When using the IOWA operator, it is assumed that the available information is exact numbers or crisp values. However, this may not be the real situation found in the decision making problem. Sometimes, the available information is vague or imprecise and it is not possible to analyze it with exact numbers. Then, it is necessary to use another approach to deal with this information such as fuzzy numbers (FN). For these situations, the IOWA is known as fuzzy number induced OWA (FN-IOWA) operator [5].

Recently, [13] have suggested a generalization of the IOWA operator by using generalized means. With this generalization, known as the induced generalized OWA (IGOWA) operator, we are able to include in the same formulation different types of induced aggregation operators such as the IOWA operator, the induced ordered weighted geometric (IOWG) operator and the induced ordered weighted quadratic averaging (IOWQA) operator, among others. Moreover, they also suggested the Quasi-IOWA operator which is a further generalization of the IGOWA operator by using quasi-arithmetic means.

Going a step further, in this paper we present the fuzzy induced generalized OWA operator which generalizes the FN-IOWA by using generalized means. We will call it the fuzzy induced generalized OWA (FIGOWA) operator. Then, we are able to obtain a wide range of fuzzy induced aggregation operators such as the FN-IOWA, the FN-IOWG operator and the FN-IOWQA operator, among others. We study some of the main properties of this generalization and we extend it to a more general formulation by using quasi-arithmetic means. The result is the Quasi-FIOWA operator. We also develop an application of the new approach in a decision making problem about selection of strategies.

This paper is organized as follows. In Section 2, we briefly review some basic concepts such as FN, the FN-IOWA and the IGOWA operator. Section 3 presents the FIGOWA operator and Section 4 studies some of its families. In Section 5 we briefly present the Quasi-FIOWA operator and in Section 6, we develop an application of the new approach in a strategic decision making problem.

2 Preliminaries

2.1 Fuzzy numbers

The FN was first introduced by [4,24]. Since then, it has been studied and applied by a lot of authors such as [6,9-11].

A FN is a fuzzy subset [23] of a universe of discourse that is both convex and normal [9]. Note that the FN may be considered as a generalization of the interval number [14] although it is not strictly the same because the interval numbers may have different meanings.

In the literature, we find a wide range of FNs [6,9]. For example, a trapezoidal FN (TpFN) A of a universe of discourse R can be characterized by a trapezoidal membership function $A = (\underline{a}, \overline{a})$ such that

$$\begin{aligned} \underline{a}(\alpha) &= a_1 + \alpha(a_2 - a_1), \\ \overline{a}(\alpha) &= a_4 - \alpha(a_4 - a_3). \end{aligned} \tag{1}$$

where $\alpha \in [0, 1]$ and parameterized by (a_1, a_2, a_3, a_4) where $a_1 \le a_2 \le a_3 \le a_4$, are real values. Note that if $a_1 = a_2 = a_3 = a_4$, then, the FN is a crisp value and if $a_2 = a_3$, the FN is

ISBN: 978-989-95079-6-8

represented by a triangular FN (TFN). Note that the TFN can be parameterized by (a_1, a_2, a_4).
In the following, we are going to review the FN arithmetic operations as follows. Let A and B be two TFN, where $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$. Then:

1) $A + B = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$
2) $A - B = (a_1 - b_3, a_2 - b_2, a_3 - b_1)$
3) $A \times k = (k \times a_1, k \times a_2, k \times a_3)$; for $k > 0$.

Note that other operations could be studied [6,9] but in this paper we will focus on these ones.

2.2 Fuzzy induced OWA operator

The FIOWA (or FN-IOWA) operator was introduced by [5]. It is an aggregation operator that uses uncertain information represented by FNs. It also uses a reordering process different from the values of the arguments. In this case, the reordering step is based on order inducing variables. It can be defined as follows.

Definition 1. *Let Ψ be the set of FN. A FIOWA operator of dimension n is a mapping FIOWA: $\Psi^n \rightarrow \Psi$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, then*:

$$FIOWA(\langle u_1,\tilde{a}_1\rangle, \ldots, \langle u_n,\tilde{a}_n\rangle) = \sum_{j=1}^{n} w_j b_j \quad (2)$$

where b_j is the $\tilde{a}_i$ value of the FIOWA pair $\langle u_i, \tilde{a}_i\rangle$ having the jth largest u_i, u_i is the order inducing variable and $\tilde{a}_i$ is the argument variable represented in the form FN.
Note that from a generalized perspective of the reordering step it is possible to distinguish between descending (DFIOWA) and ascending (AFIOWA) orders. Note also that this operator provides a parameterized family of aggregation operators that includes the fuzzy maximum, the fuzzy minimum and the fuzzy average (FA), among others.

2.3 Induced generalized OWA operator

The IGOWA operator was introduced in [13] and it represents a generalization of the IOWA operator by using generalized means. Then, it is possible to include in the same formulation, different types of induced operators such as the IOWA operator or the induced OWG (IOWG) operator. It can be defined as follows.

Definition 2. *An IGOWA operator of dimension n is a mapping IGOWA: $R^n \rightarrow R$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, then*:

$$IGOWA(\langle u_1,a_1\rangle,\ldots,\langle u_n,a_n\rangle) = \left(\sum_{j=1}^{n} w_j b_j^{\lambda}\right)^{1/\lambda} \quad (3)$$

where b_j is the a_i value of the IGOWA pair $\langle u_i,a_i\rangle$ having the jth largest u_i, u_i is the order inducing variable, a_i is the argument variable and λ is a parameter such that $\lambda \in (-\infty, \infty)$.
As we can see, if $\lambda = 1$, we get the IOWA operator. If $\lambda = 0$, the IOWG operator and if $\lambda = 2$, the IOWQA operator. Note that it is possible to further generalize the IGOWA operator by using quasi-arithmetic means. The result is the Quasi-IOWA operator.

3 Fuzzy induced generalized OWA operator

The fuzzy induced generalized OWA (FIGOWA) operator is an extension of the GOWA operator that uses uncertain information in the aggregation represented in the form of FNs. The reason for using this operator is that sometimes, the uncertain factors that affect our decisions are not clearly known and in order to assess the problem we need to use FNs. The FN is a very useful technique in decision making because it considers the different uncertain results that could happen in the future. This operator also uses a reordering process based on order inducing variables. It can be defined as follows.

Definition 3. *Let Ψ be the set of FNs. A FIGOWA operator of dimension n is a mapping FIGOWA: $\Psi^n \rightarrow \Psi$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, then*:

$$FIGOWA(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = \left(\sum_{j=1}^{n} w_j b_j^{\lambda}\right)^{1/\lambda} \quad (4)$$

where b_j is the $\tilde{a}_i$ value of the FIGOWA pair $\langle u_i,\tilde{a}_i\rangle$ having the jth largest u_i, u_i is the order inducing variable, $\tilde{a}_i$ is the argument variable represented in the form of FN and λ is a parameter such that $\lambda \in (-\infty, \infty)$.
Note that different types of FNs could be used in the aggregation such as TFNs, TpFNs, L-R FNs, interval-valued FNs, intuitionistic FNs, etc.
As it was explained in [13], when using FN in the OWA operator, we have the additional problem of how to reorder the arguments. In the FIGOWA operator, this is not a problem because the reordering process is developed with order inducing variables and it is independent of the values of the arguments.
From a generalized perspective of the reordering step, it is possible to distinguish between the descending FIGOWA (DFIGOWA) and the ascending FIGOWA (AFIGOWA) operator. The weights of these operators are related by $w_j = w^*_{n-j+1}$, where w_j is the jth weight of the DFIGOWA and w^*_{n-j+1} the jth weight of the AFIGOWA operator.
If B is a vector corresponding to the ordered arguments b_j, we shall call this the ordered argument vector and W^T is the transpose of the weighting vector, then, the FIGOWA operator can be expressed as:

ISBN: 978-989-95079-6-8

$$FIGOWA(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = W^T B \quad (5)$$

Note that if the weighting vector is not normalized, i.e., $W = \sum_{j=1}^{n} w_j \neq 1$, then, the FIGOWA operator can be expressed as:

$$FIGOWA(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = \frac{1}{W}\sum_{j=1}^{n} w_j b_j \quad (6)$$

The FIGOWA operator is monotonic, commutative, bounded and idempotent.

Theorem 1 (Monotonicity). *Assume f is the FIGOWA operator, if $\tilde{a}_i \geq \tilde{e}_i$, for all $\tilde{a}_i$, then*

$$f(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) \geq f(\langle u_1,\tilde{e}_1\rangle,\ldots,\langle u_n,\tilde{e}_n\rangle) \quad (7)$$

Theorem 2 (Commutativity). *Assume f is the FIGOWA operator, then*

$$f(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = f(\langle u_1,\tilde{e}_1\rangle,\ldots,\langle u_n,\tilde{e}_n\rangle) \quad (8)$$

where $(\langle u_1,\tilde{e}_1\rangle,\ldots,\langle u_n,\tilde{e}_n\rangle)$ is any permutation of the arguments $(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle)$.

Theorem 3 (Boundedness). *Assume f is the FIGOWA operator, then*

$$\min\{\tilde{a}_i\} \leq f(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) \leq \max\{\tilde{a}_i\} \quad (9)$$

Theorem 4 (Idempotency). *Assume f is the FIGOWA operator, if $\tilde{a}_i = \tilde{a}$, for all $\tilde{a}_i$, then*

$$f(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = \tilde{a} \quad (10)$$

Note that the proofs of Theorems 1 - 4 are omitted because they are trivial.

Another interesting issue when analysing the FIGOWA operator is the problem of ties in the order inducing variables. In order to solve this problem, we recommend to follow the policy explained in [21]. Basically, the idea is to replace each argument of the tied inducing variables by its fuzzy generalized mean. Then, different types of means may be used to replace the arguments depending on the parameter λ.

As it is explained in [21], different kinds of attributes may be used for the order inducing variables of the FIGOWA operator with the only requirement of having a linear ordering.

4 Families of FIGOWA operators

Basically, we can distinguish between two main groups of FIGOWA operators. The first family represents all the families that may be found in the weighting vector W, while the second family represents all the particular cases coming from the parameter λ.

4.1 Analysing the parameter λ

If we analyze different values of the parameter λ, we obtain another group of particular cases such as the FIOWA operator, the fuzzy IOWG (FIOWG), the fuzzy IOWQA (FIOWQA) and the fuzzy induced ordered weighted harmonic averaging (FIOWHA) operator.

When $\lambda = 1$, the FIGOWA operator becomes the FIOWA operator.

$$FIOWA(\langle u_1,\tilde{a}_1\rangle, \ldots, \langle u_n,\tilde{a}_n\rangle) = \sum_{j=1}^{n} w_j b_j \quad (11)$$

Note that it is possible to study a wide range of families of FIOWA operators by using different weighting vectors in a similar way as it has been explained in Section 4.1. For example, if $w_j = 1/n$, for all $\tilde{a}_i$, we get the FA and if the ordered position of b_j is the same than the position of the values u_i, we get the FOWA operator.

When $\lambda = 0$, we get the FIOWG operator.

$$FIOWG(\langle u_1,\tilde{a}_1\rangle, \ldots, \langle u_n,\tilde{a}_n\rangle) = \prod_{j=1}^{n} b_j^{w_j} \quad (12)$$

Note that in this case we can also study different families of FIOWG operators such as the fuzzy geometric mean or the fuzzy OWG (FOWG) operator. Note also that it is possible to distinguish between descending (DFIOWG) and ascending (AFIOWG) orders.

When $\lambda = -1$, we get the FIOWHA operator.

$$FIOWHA(\langle u_1,\tilde{a}_1\rangle, \ldots, \langle u_n,\tilde{a}_n\rangle) = \frac{1}{\sum_{j=1}^{n} \frac{w_j}{b_j}} \quad (13)$$

From a generalized perspective of the reordering step we find the descending FIOWHA (DFIOWHA) and the ascending FIOWHA (AFIOWHA) operator. Different families of FIOWHA operators are found by using different weighting vectors such as the fuzzy harmonic mean and the fuzzy ordered weighted harmonic averaging (FOWHA) operator.

When $\lambda = 2$, we get the FIOWQA operator.

$$FIOWQA(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = \left(\sum_{j=1}^{n} w_j b_j^2\right)^{1/2} \quad (14)$$

In this case, we can also study a wide range of families of FIOWQA operators such as the fuzzy quadratic mean and the fuzzy OWQA operator, and distinguish between the DFIOWQA and the AFIOWQA operator.

ISBN: 978-989-95079-6-8

4.2 Analysing the weighting vector W

By using a different weighting vector in the FIGOWA operator, we are able to obtain a wide range of aggregation operators. For example, we can obtain the fuzzy maximum, the fuzzy minimum, the FGM, the fuzzy weighted generalized mean (FWGM) and the FGOWA operator.

Remark 1. The fuzzy maximum is obtained if $w_p = 1$ and $w_j = 0$, for all $j \neq p$, and $u_p = \text{Max}\{\tilde{a}_i\}$. The fuzzy minimum is obtained if $w_p = 1$ and $w_j = 0$, for all $j \neq p$, and $u_p = \text{Min}\{\tilde{a}_i\}$. The FGM is found when $w_j = 1/n$, for all $\tilde{a}_i$. The fuzzy weighted generalized mean (FWGM) is obtained if $u_i > u_{i+1}$, for all i, and the FGOWA operator is obtained if the ordered position of u_i is the same than the ordered position of b_j such that b_j is the jth largest of $\tilde{a}_i$.

Remark 2. Other families of FIGOWA operators could be used in the aggregation by using a different manifestation of the weighting vector. For example, we could analyze the step-FIGOWA, the window-FIGOWA, the median-FIGOWA, the olympic-FIGOWA, the centered-FIGOWA, the S-FIGOWA, etc. For more information, see [10-13,17-22].

Remark 3. The step-FIGOWA operator is found when $w_k = 1$ and $w_j = 0$, for all $j \neq k$ and the window-FIGOWA when $w_j = 1/m$ for $k \leq j \leq k + m - 1$ and $w_j = 0$ for $j > k + m$ and $j < k$. Note that k and m must be positive integers such that $k + m - 1 \leq n$.

Remark 4. For the median-FIGOWA, we distinguish between two cases. If n is odd we assign $w_{(n+1)/2} = 1$ and $w_j = 0$ for all others, and this affects the argument $\tilde{a}_i$ with the [$(n + 1)/2$]th largest u_i. If n is even we assign, for example, $w_{n/2} = w_{(n/2)+1} = 0.5$, and this affects the arguments with the $(n/2)$th and [$(n/2) + 1$]th largest u_i.

Remark 5. The olympic-FIGOWA operator is found if $w_1 = w_n = 0$, and for all others $w_j = 1/(n - 2)$. Note that it is possible to develop a general form of the olympic-POWA by considering that $w_j = 0$ for $j = 1, 2, \ldots, k, n, n - 1, \ldots, n - k + 1$, and for all others $w_{j^*} = 1/(n - 2k)$, where $k < n/2$. Note that if $k = 1$, then this general form becomes the usual olympic-POWA.

Remark 6. A further family is the centered-FIGOWA operator. This type of aggregation operator is symmetric, strongly decaying and inclusive. It is symmetric if $w_j = w_{j+n-1}$. It is strongly decaying when $i < j \leq (n + 1)/2$, then $w_i < w_j$ and when $i > j \geq (n + 1)/2$ then $w_i < w_j$. It is inclusive if $w_j > 0$. Note that it is possible to consider a softening of the second condition by using $w_i \leq w_j$ instead of $w_i < w_j$ which is known as softly decaying centered-FIGOWA operator. Note also the possibility of removing the third condition. Then, we shall refer to this type of aggregation as non-inclusive centered-FIGOWA operator.

Remark 7. A further interesting family is the S-FIGOWA operator. In this case, we can distinguish between three types: the "orlike", the "andlike", and the "generalized" S-FIGOWA operator. The generalized S-FIGOWA operator is obtained when $w_p = (1/n)(1 - (\alpha + \beta)) + \alpha$, with $u_p = \text{Max}\{\tilde{a}_i\}$; $w_q = (1/n)(1 - (\alpha + \beta)) + \beta$, with $u_q = \text{Min}\{\tilde{a}_i\}$; and $w_j = (1/n)(1 - (\alpha + \beta))$ for all $j \neq p,q$ where $\alpha, \beta \in [0, 1]$ and $\alpha + \beta \leq 1$. Note that if $\alpha = 0$, we get the andlike S-FIGOWA and if $\beta = 0$, the orlike S-FIGOWA.

Remark 8. Another type is the non-monotonic-FIGOWA operator. It is obtained when at least one of the weights w_j is lower than 0 and $\sum_{j=1}^{n} w_j = 1$. Note that a key aspect of this operator is that it does not always achieve monotonicity.

5 The Quasi-FIOWA operator

The FIGOWA operator may be further generalized by using quasi-arithmetic means. Then, the result is the fuzzy induced ordered weighted quasi-arithmetic averaging operator or Quasi-FIOWA, for short. Note that the Quasi-FIOWA operator is an extension of the Quasi-OWA [1,3,7,11] by using order inducing variables and uncertain information represented with FNs.

Definition 4. *Let Ψ be the set of FNs. A Quasi-FIOWA operator of dimension n is a mapping $f: \Psi^n \rightarrow \Psi$ that has an associated weighting vector W of dimension n such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$, then*:

$$f(\langle u_1,\tilde{a}_1\rangle,\ldots,\langle u_n,\tilde{a}_n\rangle) = g^{-1}\left(\sum_{j=1}^{n} w_j g(b_j)\right) \quad (15)$$

where b_j is the $\tilde{a}_i$ value of the Quasi-FIOWA pair $\langle u_i,\tilde{a}_i\rangle$ having the jth largest u_i, u_i is the order inducing variable, $\tilde{a}_i$ is the argument variable represented in the form of FN and $g(b)$ is a strictly continuous monotone function.

As we can see, when $g(b) = b^{\lambda}$, we get the FIGOWA operator. Note that it is also possible to distinguish between descending (Quasi-DFIOWA) and ascending (Quasi-AFIOWA) orders. Note also that all the properties and particular cases commented in the FIGOWA operator, are also applicable in this case.

6 Application in strategic decision making

In the following, we are going to develop a brief example where we will see the applicability of the new approach. We will focus in a decision making problem about selection of strategies. Note that other business decision making applications could be developed such as financial decision making, human resource selection, etc. Note that the FIGOWA operator may be applied in similar problems than the IOWA and the IGOWA operator.

ISBN: 978-989-95079-6-8

Assume a company that operates in North America and Europe is analyzing the general policy for the next year and they consider 5 possible strategies to follow.

1) A_1 = Expand to the Asian market.
2) A_2 = Expand to the South American market.
3) A_3 = Expand to the African market.
4) A_4 = Expand to the 3 continents.
5) A_5 = Do not develop any expansion.

In order to evaluate these strategies, the group of experts considers that the key factor is the economic situation of the next year. Thus, depending on the situation, the expected benefits will be different. The experts have considered 5 possible situations for the next year: S_1 = Very bad, S_2 = Bad, S_3 = Regular, S_4 = Good, S_5 = Very good. The expected results depending on the situation S_i and the alternative A_i are shown in Table 1. Note that the results are TFN.

Table 1: Available strategies

	S_1	S_2	S_3	S_4	S_5
A_1	(20,30,40)	(60,70,80)	(40,50,60)	(50,60,70)	(50,60,70)
A_2	(30,40,50)	(70,80,90)	(30,40,50)	(30,40,50)	(50,60,70)
A_3	(60,70,80)	(50,60,70)	(40,50,60)	(20,30,40)	(40,50,60)
A_4	(50,60,70)	(30,40,50)	(60,70,80)	(70,80,90)	(10,20,30)
A_5	(40,50,60)	(30,40,50)	(50,60,70)	(60,70,80)	(40,50,60)

In this problem, the experts consider the weighting vector W = (0.1, 0.2, 0.2, 0.2, 0.3). Due to the fact that the attitudinal character is very complex because it involves the opinion of different members of the board of directors, the experts use order inducing variables to express it.

Table 2: Order inducing variables

	S_1	S_2	S_3	S_4	S_5
A_1	7	9	6	5	8
A_2	4	3	6	8	7
A_3	2	8	4	3	6
A_4	5	6	9	2	7
A_5	8	4	3	6	5

With this information, we can aggregate it in order to take a decision. In Table 3 and 4, we show the different results obtained by using different types of FIGOWA operators in the decision process.

Table 3: First aggregation process

	FA	FWA	FOWA
A_1	(44,54,64)	(47,57,67)	(40,50,60)
A_2	(42,52,62)	(44,54,64)	(38,48,58)
A_3	(42,52,62)	(40,50,60)	(38,48,58)
A_4	(44,54,64)	(40,50,60)	(38,48,58)
A_5	(44,54,64)	(44,54,64)	(41,51,61)

Table 4: First aggregation process

	FIOWA	FIOWG	FIOWQA
A_1	(43,53,63)	(40.5,51.1,61.5)	(44.8,54.4,64.2)
A_2	(46,56,66)	(42.8,53.4,63.7)	(49.1,58.6,68.2)
A_3	(43,53,63)	(40.2,50.8,61.2)	(45.2,54.8,64.5)
A_4	(45,55,65)	(36.8,49.1,60.3)	(50.2,59.4,68.7)
A_5	(45,55,65)	(43.7,54.0,64.1)	(46.1,55.9,65.8)

If we establish an ordering of the alternatives, we get the following results shown in Table 5. Note that in this example it is not necessary to establish a criterion for ranking FNs because it is clear which alternative goes first, second and so on, in the ordering process. Note also that "$\succ$" means "preferred to" and "=" means "equal to".

Table 5: Ordering of the investments

	Ordering
FA	$A_1=A_4=A_5 \succ A_2=A_3$
FWA	$A_1 \succ A_2=A_5 \succ A_3=A_4$
FOWA	$A_5 \succ A_1 \succ A_2=A_3=A_4$
FIOWA	$A_2 \succ A_4=A_5 \succ A_1=A_3$
FIOWG	$A_5 \succ A_2 \succ A_1 \succ A_3 \succ A_4$
FIOWQA	$A_4 \succ A_2 \succ A_5 \succ A_3 \succ A_1$

As we can see, depending on the aggregation operator used, the ordering of the investments may be different. Therefore, the decision about which investment select may be also different. For example, the FA gives very similar results between alternatives while the FIOWG and the FIOWQA give more differences between the alternatives.

7 Conclusions

We have presented the FIGOWA operator. It is a generalization of the OWA operator that uses the main characteristics of three well known aggregation operators: the GOWA, the IOWA and the FOWA operator. That is to say, it uses generalized means, order inducing variables and FNs in the aggregation. We have studied some of the main properties of this new aggregation operator. We have further generalized it by using quasi-arithmetic means. Then, we have obtained the Quasi-FIOWA operator. By using this approach, we get a more complete representation of the information because we are using FNs that consider the best and worst result and the possibility that the internal results will occur.

We have also developed an application of the new approach. We have focused in a strategic decision making problem and we have seen that depending on the particular FIGOWA operator used, the results and the decisions may be different.

In future research, we expect to develop further extensions by adding new characteristics [11] in the problem and applying it to other business problems such as financial decision making and human resource management.

Acknowledgment

This paper is partly supported by the Spanish *Ministerio de Asuntos Exteriores y de Cooperación*, *Agencia Española de*

Cooperación Internacional para el Desarrollo (AECID) (project A/016239/08).

References

[1] G. Beliakov, A. Pradera and T. Calvo, *Aggregation Functions: A Guide for Practitioners*. Berlin: Springer-Verlag, 2007.

[2] H. Bustince, F. Herrera and J. Montero, *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*. Berlin: Springer-Verlag, 2008.

[3] T. Calvo, G. Mayor and R. Mesiar, *Aggregation Operators: New Trends and Applications*. New York: Physica-Verlag, 2002.

[4] S.S.L. Chang and L.A. Zadeh, On fuzzy mapping and control. *IEEE Transactions on Systems, Man and Cybernetics*, 2:30-34, 1972.

[5] S.J. Chen and S.M. Chen, A new method for handling multi-criteria fuzzy decision making problems using FN-IOWA operators. *Cybernetics and Systems*, 34:109-137, 2003.

[6] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, 1980.

[7] J. Fodor, J.L. Marichal and M. Roubens, Characterization of the ordered weighted averaging operators. *IEEE Transactions on Fuzzy Systems*, 3:236-240, 1995.

[8] N. Karayiannis, Soft Learning Vector Quantization and Clustering Algorithms Based on Ordered Weighted Aggregation Operators. *IEEE Transactions on Neural Networks*, 11:1093-1105, 2000.

[9] A. Kaufmann, M.M. Gupta, *Introduction to fuzzy arithmetic*. Publications Van Nostrand, Rheinhold, 1985.

[10] J.M. Merigó, Using immediate probabilities in fuzzy decision making. In: *Proceedings of the 22nd ASEPELT Conference*, Barcelona, pp. 1650-1664, 2008.

[11] J.M. Merigó, *New extensions to the OWA operators and its application in decision making* (In Spanish). PhD Thesis, Department of Business Administration, University of Barcelona, 2008.

[12] J.M. Merigó and M. Casanovas, Decision making with distance measures and induced aggregation operators. In: *Proceedings of the 8th FLINS International Conference*, Madrid, pp. 483-488, 2008.

[13] J.M. Merigó and A.M. Gil-Lafuente, The induced generalized OWA operator. *Information Sciences*, 179:729-741, 2009.

[14] R.E. Moore, *Interval Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1966.

[15] V. Torra and Y. Narukawa, *Modeling Decisions: Information Fusion and Aggregation Operators*. Berlin: Springer-Verlag, 2007.

[16] Z.S. Xu and Q.L. Da, An overview of operators for aggregating information. *International Journal of Intelligent Systems*, 18:953-968, 2003

[17] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics B*, 18:183-190, 1988.

[18] R.R. Yager, Families of OWA operators. *Fuzzy Sets and Systems*, 59:125-148, 1993.

[19] R.R. Yager, Induced aggregation operators. *Fuzzy Sets and Systems*, 137:59-69, 2003.

[20] R.R. Yager, Generalized OWA Aggregation Operators. *Fuzzy Optimization and Decision Making*, 3:93-107, 2004.

[21] R.R. Yager and D.P. Filev, Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man and Cybernetics B*, 29:141-150, 1999.

[22] R.R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators: Theory and Applications*. Norwell: Kluwer Academic Publishers, 1997.

[23] L.A. Zadeh, Fuzzy sets. *Information and Control*, 8:338-353, 1965.

[24] L.A. Zadeh, The Concept of a Linguistic Variable and its application to Approximate Reasoning. Part 1. *Information Sciences*, 8:199-249, 1975; Part 2. *Information Sciences*, 8:301-357, 1975; Part 3. *Information Sciences*, 9:43-80, 1975.

ISBN: 978-989-95079-6-8

Solution for Fuzzy Road Transport Traveling Salesman Problem Using Eugenic Bacterial Memetic Algorithm

János Botzheim[1] Péter Földesi[2] László T. Kóczy[3]

1 Department of Automation, Széchenyi István University
H-9026, Győr, Egyetem tér 1. Hungary
2 Department of Logistics and Forwarding, Széchenyi István University
H-9026, Győr, Egyetem tér 1. Hungary
3 Department of Telecommunications, Széchenyi István University
H-9026, Győr, Egyetem tér 1. Hungary
Email: botzheim@sze.hu, foldesi@sze.hu, koczy@sze.hu

Abstract— *The aim of the Traveling Salesman Problem (TSP) is to find the cheapest way of visiting all elements in a given set of cities and returning to the starting point. In solutions presented in the literature costs of travel between nodes (cities) are based on Euclidean distances, the problem is symmetric and the costs are constant. In this paper a novel construction and formulation of the TSP is presented in which the requirements and features of practical application in road transportation and supply chains are taken into consideration. Computational results are presented as well.*

Keywords— Traveling Salesman Problem, time dependent fuzzy costs, eugenic bacterial memetic algorithm

1 Introduction

The aim of the Traveling Salesman Problem (TSP) is to find the cheapest way of visiting all elements in a given set of cities where the cost of travel between each pair of them is given, including the return to the starting point. The TSP is a very good representative of a larger class of problems known as combinatorial optimization problems. If an efficient algorithm (i.e., an algorithm that will guarantee to find the optimal solution in polynomial number of steps) can be found for the traveling salesman problem, then efficient algorithms could be established for all other problems in the NP (nondeterministic polynomial time) class, thus TSP belongs to the so called NP-hard complexity class. TSP is a well studied NP-hard problem [5]. An algorithm can be considered as an effective (good) one if it has a polynomial function of the problem size n, that is, for large values of n, the algorithm runs in time at most Kn^c for some constant number K and c. The question whether or not there is a good algorithm for the TSP has not been settled. For its practical importance and wide range of application in practice [5] many approaches, heuristic searches and algorithms have been suggested [8, 10, 11, 12], while different extensions and variations of the original TSP have been investigated [6, 9, 13].

Solutions presented in the literature most frequently have the following features. Costs of travel between nodes (cities) are based on Euclidean distances, the problem is symmetric, meaning that the cost from $node_i$ to $node_j$ equals to the cost from $node_j$ to $node_i$, and the costs are constant. In this paper a novel construction and formulation of the TSP is presented in which the requirements and features of practical application in road transportation and supply chains are taken into consideration. Since the original formulation of the problem states: the aim is to find the "cheapest" tour, thus the cost matrix that represents the distances between each pair must be determined by calculating the actual costs of transportation processes. The costs of transportation consist of two main elements: costs proportional to transit distances (km) and costs proportional to transit times. Obviously the physical distances can be considered as constant values in a given relation, but transit times are subject to external factors, such as weather conditions, traffic circumstances, etc., so they should be treated as a time-dependent variable. Furthermore the actual costs are rarely constant and predictable, so fuzzy cost coefficient can be applied in order to represent the uncertainty [15, 17]. On the other hand in real road networks the actual distance between two points often alter from the Euclidean distance, furthermore occasionally some extra costs (e.g., ferriage, tunnel fare) can modify the distance-related variable costs. Considering these characteristics the original TSP should be reconstructed, so that realistic solutions can be developed.

For solving the above-mentioned fuzzy road-transport TSP (FRTTSP) in this paper we suggest a eugenic bacterial memetic algorithm (EBMA) since that algorithm is suitable for global optimization of even non-linear, high-dimensional, multi-modal, and discontinuous problems. As numerical example a modified TSP (FRTTSP) instance is considered, in which the elements of cost matrix are dependent on the steps they are selected to carry on with.

2 Formulating and solutions for the classical TSP

In the case of the traveling salesman problem, the mathematical description can be a graph where each city is denoted by a point (or node) and lines are drawn connecting every two nodes (called arcs or edges). A distance (or cost) is associated with every edge. If in a graph edges are drawn connecting any two nodes, then the graph is said to be *complete*. A round-trip of the cities corresponds to a special subset of the lines when each city is visited exactly once, and it is called a tour or a Hamiltonian cycle in graph theory. The length of a tour is the sum of the lengths of the lines in the round-trip.

ISBN: 978-989-95079-6-8

Asymmetric and symmetric TSPs can be distinguished depending on if any edge of the graph is directed or not. To formulate the symmetric case with n nodes $c_{ij} = c_{ji}$, so a graph can be considered where there is only one arc (undirected) between every two nodes. Let $x_{ij} = \{0,1\}$ be the decision variable ($i= 1,2,...,n$ and $j=1,2,...,n$), and $x_{ij} = 1$, means that the arc connecting node$_i$ to node$_j$ is an element of the tour.

Let $x_{ii} = 0 \ (i= 1,2,...,n)$ (1)

meaning that no tour element is allowed from a node to itself. Furthermore

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} = n \quad (2)$$

that is the number of decision variables where $x_{ij} = 1$ is equal to n, and

$$\sum_{i=1}^{n} x_{ij} = 1 \qquad \forall j \in \{1,2,..., n\}, \quad (3)$$

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \forall i \in \{1,2,..., n\}, \quad (4)$$

meaning that each column and row of the decision matrix has a single element with a value 1 (i.e., each city is visited once). For assuring the close circuit, an additional constraint must be set. A permutation of nodes $(p_1,p_2,....p_n)$ has to be constructed so that the total cost $C(p)$ is minimal:

$$\text{minimize} \quad C(p) = \left(\sum_{i=1}^{n-1} c_{p_i, p_{i+1}} \right) + c_{p_n, p_1} . \quad (5)$$

For a symmetrical network there are $1/2\ (n-1)!$ possible tours (because the degree of freedom is $(n-1)$ and tours describing the same sequence but opposite directions are not considered as different tours) and for asymmetric networks where $c_{ij} \neq c_{ji}$ the number of possible tours is $(n-1)!$. Some reduction can be done [14], a guarantee that is proportional to $n^2 2^n$ can be given, however it is clear that an exhaustive search is not possible for large n in practice. Rather than enumerating all possibilities, successful algorithms for solving the TSP problem have been capable of eliminating most of the roundtrips without ever actually considering them. The main groups of engines are [5-12]:

- Mixed-integer programming
- Branch-and-bound method
- Heuristic searches (local search algorithms, simulated annealing, neural networks, genetic algorithms, particle swarm optimization, ant colony optimization, etc.)

In this paper a eugenic bacterial memetic algorithm is proposed as a novel heuristic search for TSP.

3 Variable costs and consequences of their nature

3.1 Costs in real road networks

Considering real road transport networks, especially in city logistics, the actual circumstances and condition of the transit process are subject to not only the topography of the given network but to timing as well. Referring to the phenomenon of cyclic peak-hours and also to the weekly (monthly, yearly) periodicity of traffic on road, it can be stated that the unit cost of traveling is also a variable, and it can be described as a time series rather than a constant value. The main reasons of that are the following:

- The actual cost of 1 km depends on the current fuel consumption, which is partly affected by the speed, but the speed is externally determined by the current traffic.
- A relatively large proportion of transportation cost is the cost of labor (i.e., wages of drivers), which is calculated on driving time. In Europe distance-based payment for commercial drivers are not allowed according to the European Agreement Concerning the Work of Crews of Vehicles Engaged in International Road Transport (AETR).
- The return on equity capital is a vital issue for haulage companies, since the transportation sector of the economy is a capital intensive one, meaning that utilization of vehicle in time is a crucial problem.

In addition, during long-distance shipments the drivers occasionally must stop for a rest (according to AETR) at a minimum 11 hour period (often overnight), and very often week-end traffic restrictions for heavy vehicles are introduced.

Considering the uncertainty of relevant data it can be stated, that one's estimated travel time by automobile between two points is a case of possibility due to the measurement imprecision and perception. It can be seen that the circumstances and conditions are significantly changing in time that is the actual value of cost matrix element c_{ij} should be subject to timing of transit between node$_i$ and node$_j$ and an appropriate representation of imprecision can be the use of fuzzy numbers. In this sense geographical optimization alone is not appropriate, and the road transport operation has to be scheduled in time as well.

3.2 The modified TSP, the Fuzzy Road Transport Traveling Salesman Problem (FRTTSP)

For fulfilling the requirements of realistic road transport processes, we propose the following modification to the classical TSP [23].

Since the overall target is to achieve the cheapest tour (in monetary terms), the constraint that each node (city) is visited exactly once is skipped. Calculating with time-dependent cost coefficients that are not necessarily proportional to distances, a longer route can be a cheaper one, and as a consequence some nodes can be visited more than once. Thus in the FRTTSP we eliminate restrictions (2), (3) and (4).

If a significant improvement in traffic conditions can be expected, that is a future value of a cost c_{ij} will be less then its present value, it is worth waiting (suspending the tour for a while) and continuing in the next step. In this case obviously the cost of staying at a point must be calculated. In this sense we eliminate restriction (1) as well. Very often in the solutions restriction (1) is fulfilled by selecting $c_{ii} = \infty$. In our case c_{ii} is the cost of staying at node$_i$ in a given step.

ISBN: 978-989-95079-6-8

The permutation of nodes (p_1,p_2,….p_n) is being modified as well. As a city may be visited several times, objective function (5) must be rewritten:

$$\text{minimize } C(p) = \left(\sum_{i=1}^{mn-1} c_{p_i, p_{i+1}} \right) + c_{p_{mn}, p_1}, \quad (6)$$

where *m=1,2,...* is the multiplier factor. Objective function (6) is a generalized form of TSP, the multiplier factor equals to *1* in the classical cases.

In order to represent the uncertainty triangular fuzzy numbers are used as cost coefficients. Triangular fuzzy numbers have a membership function consisting of two linear segments joined at a peak, so they can be constructed easily on the basis of little information: the supporting interval $C= [c_1, c_2]$ as the smallest and the largest possible values, and c_M which is the peak value where the membership function equals to *1*. In that case the triangular fuzzy number is denoted by $\boldsymbol{C} = (c_1, c_M, c_2)$.

When the distances between the cities are described by fuzzy numbers, it must be discussed how these fuzzy numbers are summed up in a tour in order to calculate the total distance. The arithmetic of fuzzy numbers is based on the extension principle [18]. We are using triangular shaped fuzzy numbers which can be characterized by three values, the boundaries of the support and the core value. When we calculate the total distance of a tour, then instead of adding fuzzy numbers by the extension principle, we can do an easier calculation based on the defuzzified values of the fuzzy numbers. According to [16], some defuzzification method has invariance properties meaning that the result is invariant under linear transformations, thus there is no need to determine the whole outcome using the extension principle but only to compute the sum of the defuzzified values of each fuzzy number. If we are using triangular shaped fuzzy numbers then the Averaging Level Cuts (ALC) type defuzzification method used in [16] gives the same result as the Center of Gravity (COG) method. So, in the first step the fuzzy numbers are defuzzified by the COG method (which is simply the arithmetic mean of the three characteristic points of the fuzzy number) and then these crisp numbers are summed up providing the total distance of the tour.

4 Eugenic bacterial memetic algorithms

Nature inspired some evolutionary optimization algorithms suitable for global optimization of even non-linear, high-dimensional, multi-modal, and discontinuous problems. The original genetic algorithm was developed by Holland [1] and was based on the process of evolution of biological organisms. It uses three operators: reproduction, crossover and mutation. Later, new kind of evolutionary based techniques were proposed, which are imitating phenomena that can be found in nature.

Bacterial Evolutionary Algorithm (BEA) [2] is one of these techniques. BEA uses two operators; the bacterial mutation and the gene transfer operation. These new operators are based on the microbial evolution phenomenon. Bacteria share chunks of their genes rather than perform a neat crossover in chromosomes. The bacterial mutation operation optimizes the chromosome of one bacterium; the gene transfer operation allows the transfer of information between the bacteria in the population. Each bacterium represents a solution for the original problem. BEA has been applied for wide range of problems, for instance optimizing the fuzzy rule bases [2, 3] or feature selection [4].

Evolutionary algorithms are global searchers, however in most of the cases they give only a quasi-optimal solution for the problem. Local search approaches can give more accurate solution, however they are searching for the solution only in a local area of the search space. Local search approaches might be useful in improving the performance of the basic evolutionary algorithm, which may find the global optimum with sufficient precision in this combined way. Combinations of evolutionary and local-search methods are usually referred to as memetic algorithms [19]. A new kind of memetic algorithm based on the bacterial approach is the bacterial memetic algorithm (BMA) [20].

The algorithm consists of four steps. First, an initial population has to be created. Then, bacterial mutation, a local search and gene transfer are applied, until a stopping criterion is fulfilled. The bacterial mutation is applied to each chromosome one by one. First, N_{clones} copies (clones) of the bacterium are generated, then a certain segment of the chromosome is randomly selected and the parameters of this selected segment are randomly changed in each clone (mutation). Next all the clones and the original bacterium are evaluated and the best individual is selected. This individual transfers the mutated segment into the other individuals. This process continues until all of the segments of the chromosome have been mutated and tested. At the end of this process the clones are eliminated. After the bacterial mutation operator a local search is applied for each individual. This method depends on the given problem. For the TSP it is detailed in the next section.

In the next step the other evolutionary operation, the gene transfer is applied, which allows the recombination of genetic information between two bacteria. First, the population must be divided into two halves. The better bacteria are called the superior half, the other bacteria are called the inferior half. One bacterium is randomly chosen from the superior half, this will be the source bacterium and another is randomly chosen from the inferior half, this will be the destination bacterium. A segment from the source bacterium is chosen randomly and this segment will overwrite a segment of the destination bacterium or it will be added to the destination bacterium. This process is repeated for N_{inf} times. The stopping condition is usually given by a predefined maximum generation number (N_{gen}). When N_{gen} is achieved then the algorithm ends otherwise it continues with the bacterial mutation step. We use eugenic elements also in the algorithm [21]. This reflects the human's decision will, and puts some deterministic element into the algorithm. Details will be given in the next section.

The basic algorithm has four parameters: the number of generations (N_{gen}), the number of bacteria in the population (N_{ind}), the number of clones in the bacterial mutation (N_{clones}), and the number of infections (N_{inf}) in the gene transfer operation.

ISBN: 978-989-95079-6-8

5 EBMA for the modified traveling salesman problem

When applying evolutionary type algorithms first of all the encoding method must be defined. The evaluation of the individuals has to be discussed, too. The operations of the algorithm have to be adapted to the given problem.

5.1 Encoding method and evaluation of the individuals

In the modified traveling salesman problem one city may be visited more than once. Because each city must be visited at least once, one solution of the problem does not need to be a permutation of the cities. The evident encoding of the problem into a bacterium is simply the enumeration of the cities in the order they should be visited. Therefore, a length of the bacterium may be greater than the number of cities (N_{cities}), but an upper bound for the bacterium length has to be defined too, we allow bacteria not longer than $m \cdot N_{cities}$, where m is the multiplier factor, which is a parameter of the algorithm (usually m=2). The initial city is not represented in the bacterium.

The length of the bacteria can be changeable. It can be changing during the evolutionary process and the individuals can have different lengths. In the initial population generation, the length of the bacteria is a random number greater than or equal to N_{cities} and less than or equal to $m \cdot N_{cities}$.

The evaluation of a bacterium is based on the time dependent distance matrix. The distance between the first element of the bacterium and the initial city is taken from the distance matrix at the zeroth time step, the distance between the second element of the bacterium and the first element of the bacterium is taken from the distance matrix at the first time step, and so on, these distances are summed up, and the total distance is obtained in this way.

5.2 Bacterial mutation

In the bacterial mutation there is an additional parameter, the length of the segment to be mutated in the clones. First, the segments of the bacterium are determined, and a random segment order is created. In the clones, the mutation of the given segment is executed. For example in Figure 1, the length of the segment is 3, and there are 4 clones. The random segment order is e.g., {3rd segment, 1st segment, 4th segment, 2nd segment}. This means that in the first sub-cycle of bacterial mutation, the 3rd segment is mutated in the clones. After the mutation of the clones, the best one is being chosen, and this clone (or the un-mutated original bacterium) transfers the mutated segment to the other individuals.

The segments of the bacterium do not need to consist of consecutive elements. The elements of the segments can come from different parts of the bacterium as it can be seen in Figure 2.

Because in the modified TSP the number of visited cities is not predefined, bacteria with different length can occur in the population. Although in the initial population generation bacteria with different length can arise we would like to allow the changes in the length within the bacterial operations, too. Therefore before a clone is mutated a random value is used for determining that after the mutation the length of the clone will increase, decrease or remain the same.

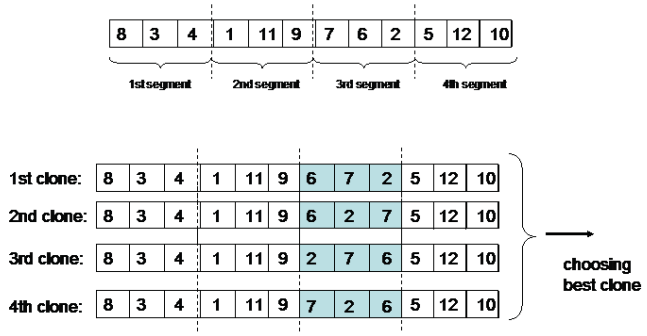


Figure 1: Bacterial mutation

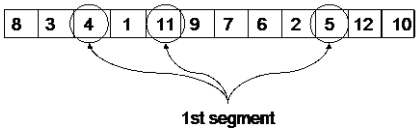


Figure 2: Segment in the bacterial mutation

Increasing is allowed only in the case, when the maximum bacterium length ($m \cdot N_{cities}$) is not exceeded, similarly, decreasing is allowed only in the case, when the minimum bacterium length (N_{cities}) is guaranteed. If the length will increase, then besides changing the positions of the cities in the selected segment of the clone, new cities are added to this clone randomly. If it will decrease, then some cities are deleted from the clone taking care that only those cities are allowed to be deleted, which have at least one other occurrence in the clone. If the length remains the same, then only the positions of the cities in the selected segment are changed.

5.3 Local search method

A tour can be improved by some local heuristics. One of the most successful methods is the Lin-Kernighan algorithm [22] which based on the k-opt algorithm. The k-opt algorithm removes k edges from the tour and reconnects the k paths optimally. We applied the 2-opt and 3-opt technique in our EBMA algorithm. For higher value of k the algorithm would take more time and would provide only small improvements on the 2-opt and 3-opt techniques.

5.4 Gene transfer

In the gene transfer operation there is also an additional parameter, the length of the segment to be transferred from the source bacterium to the destination bacterium. In contrast with the bacterial mutation, in the gene transfer, the segment can contain only consecutive elements within the bacterium. The reason for that is the segment containing consecutive elements representing sub-tours in the bacterium, and transferring good sub-tours is the main goal of the gene transfer operation.

Figure 3 shows the gene transfer in the case of time independent distance matrix. In the case of time dependent

ISBN: 978-989-95079-6-8

distance matrix, the position where the transferred segment goes to in the destination bacterium must be the same as the position of the segment in the source bacterium.

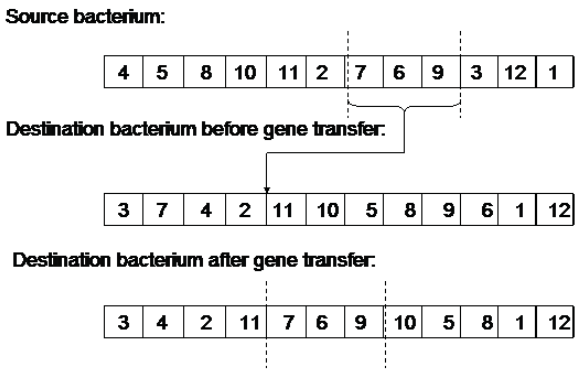


Figure 3: Gene transfer

The different individual lengths are allowed also in the gene transfer operation. After the segment was transferred to the destination bacterium, the elements that occur in the transferred segment and are already in the destination bacterium can be deleted from the destination bacterium. If the same number of elements is deleted from the destination bacterium as the length of the segment, then the length of the destination bacterium remains the same. If less elements are deleted, then its length will be increasing. If more elements are deleted (taking care that each city must have at least one occurrence in the bacterium), then the length will be decreasing. We must also take care that the length of the destination bacterium must be at least N_{cities} and at most $m \cdot N_{cities}$.

5.5 *Eugenic elements*

Eugenic is used in the initial population creation and in the bacterial mutation operator. This means that we put more determinism into the algorithm which contains normally the deterministic local search and the stochastic evolutionary operators. During the initial population creation not only random individuals are generated but also some deterministic ones according to the following rule: there is an individual, which represents the tour in which always the nearest unvisited city is visited. There can be another initial bacterium, which represents the tour in which alternating the nearest and the second nearest city is visited. There can be a third individual, where always the second nearest city is taken.

In the bacterial mutation not only randomly mutated clones are produced, but there will always be a deterministic clone, which performs a reverse ordering permutation on the selected segment. According to our experiences this can be effective in solving TSP like problems.

6 Computational results

First the efficiency of the proposed algorithm is presented. Fig. 4 and 5 compare the computational result against a "classical" reference instance TSP (www.tsp.gatech.edu, XQF131). The total length of the optimal tour is *564*, the solution of our proposed algorithm is *566*, the computational error is *0.35%*.

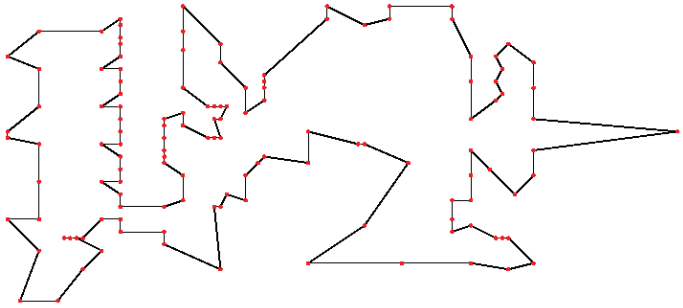


Figure 4: Reference instance (www.tsp.gatech.edu)

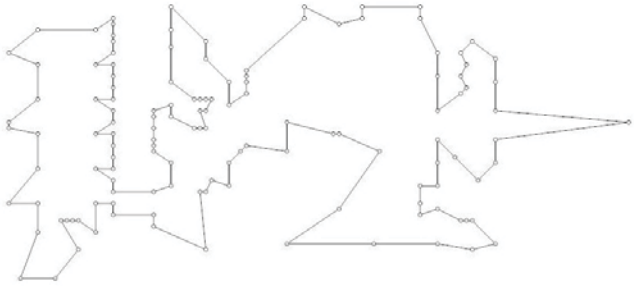


Figure 5: Solution of reference instance by the EBMA

Next in the evaluation process a classic symmetric graph is compared to an asymmetric and time dependent version of the same graph. Solution for the symmetric case is shown in Fig. 6.

The RTTSP version is modified in the following points:

$Node_{16}$ has only one connection to $node_6$.

$Node_6$ has two connections, to $node_{16}$ and to $node_9$.

$$C_{1,23}(t) = C_{1,23} + 0.05t$$
$$C_{20,0}(t) = C_{20,0} + 0.01t$$
$$C_{13,0}(t) = C_{13,0} + 0.02t$$

where t is the step in which the tour visits the node. This is a very simple representation of the time dependency of the graph. The results are shown in Fig. 7.

Since the running time of EBMA is about proportional to the number of generations it is a crucial task to select the most efficient parameters:

- number of bacteria in the population (N_{ind})
- number of clones in the bacterial mutation (N_{clones})
- number of infections (N_{inf}) in the gene transfer operation
- length of the segment to be mutated
- length of the segment to be transferred.

Finally the fuzzy cost coefficients are considered, results are shown in Fig. 8. Let the fuzzy values are:

$$\boldsymbol{C}_{2,0} = (0.015, 0.1986, 0.4)$$
$$\boldsymbol{C}_{0,2} = (0.001, 0.1986, 0.2)$$
$$\boldsymbol{C}_{23,4} = (0.1, 0.4504, 0.5)$$
$$\boldsymbol{C}_{4,23} = (0.4, 0.4504, 0.7)$$
$$\boldsymbol{C}_{26,29}(t) = (0.05, 0.1655, 0.2)\ 0.05\ t$$

ISBN: 978-989-95079-6-8

$C_{29,26}(t)$= (0.11, 0.1655, 0.3) 0.05 t
$C_{10,4}(t)$= (0.03, 0.1449, 0.16) 0.06 t
$C_{4,10}(t)$= (0.12, 0.1449, 0.3) 0.06 t
$C_{17,18}(t)$= (0.15, 0.267, 0.3)+0.05 t
$C_{18,17}(t)$= (0.15, 0.267, 0.3)-0.01 t
$C_{7,23}(t)$= (0.1, 0.18, 0.25)+0.01 t
$C_{23,7}(t)$= (0.1, 0.18, 0.25)-0.01 t
$C_{0,21}(t)$= (0.01, 0.066, 0.10)+0.02 t
$C_{21,0}(t)$= (0.01, 0.066, 0.10)-0.02 t

(Remark: peak values are kept from the crisp matrix.)

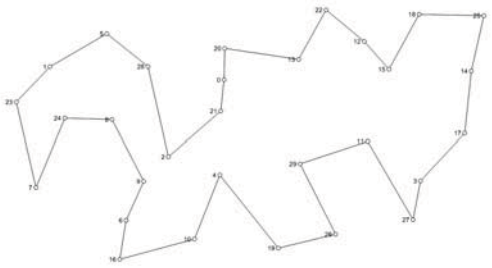


Figure 6: Graphical representation of the best tour of TSP

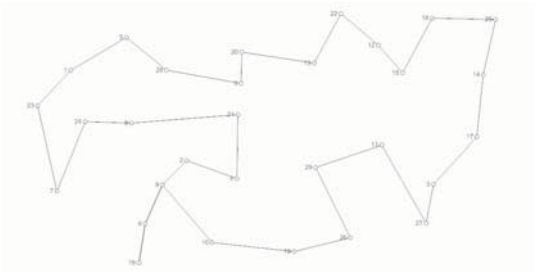


Figure 7: Graphical representation of the best tour in RTTSP

Figure 8: Graphical representation of the best tour in FRTTSP

The scope of future research activity is to set general rules that can give instructions in order to find the most efficient parameters of EBMA according to the size and other (e.g. topographical) features of a given FRTTSP, since running time is significantly affected by those parameters.

Acknowledgment
This paper was supported by the Széchenyi University Main Research Direction Grant 2009, and a National Scientific Research Fund Grant OTKA T048832 and K75711.

References

[1] Holland, J. H.: *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, The MIT Press, Cambridge (1992)

[2] Nawa, N. E., Furuhashi, T.: *Fuzzy System Parameters Discovery by Bacterial Evolutionary Algorithm*, IEEE Tr. Fuzzy Systems 7 (1999) pp. 608-616

[3] Botzheim, J., Hámori, B., Kóczy, L. T., Ruano, A. E.: *Bacterial algorithm applied for fuzzy rule extraction*, in Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, Annecy, France (2002) pp. 1021-1026

[4] Botzheim J., Drobics M., Kóczy L. T.: *Feature selection using bacterial optimization*, in Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, IPMU 2004, Perugia, Italy, (July 2004) pp. 797–804

[5] Applegate, D.L., Bixby, R.E., Chvátal, V., Cook, W.J.: *The Traveling Salesman Problem, A Computational Study*, Princeton University Press, Princeton and Oxford (2006) pp. 10-11

[6] Yu-Hsin Liu : *A hybrid scatter search for the probabilistic traveling salesman problem*, Computers & Operations Research 34 (2007) pp. 2949-2963

[7] Teoh, E.J.,Tan, K.C., Tang, H.J., Xiang, C., Goh, C.K.: *An asynchronous recurrent linear threshold network approach to solving the traveling salesman problem*, Neurocomputing 71 (2008) pp. 1359-1372

[8] Ding, C., Cheng, Y., He, M.: *Two-Level Genetic algorithm for Clustered Traveling Salesman Problem with Application in Large-Scale TSPs*, Tsinghua Science and Technology, Vol.12.No.4 (2007) pp. 459-465

[9] Hsiao-Fan Wang, Yu-Pin Wen: *Time-Constrained Chinese Postman Problems*, Computers and Mathematics with Applications 44 (2002) pp. 375-387

[10] Shi, X.C., Liang, Y.C., Lee, H.P., Lu, C., Wang, Q.X.: *Particle swarm optimization-based algorithms for TSP and generalized TSP*, Information Processing Letters 103 (2007) pp. 169-176

[11] Bontoux, B., Feillet, D.: *Ant colony optimization for the traveling purchaser problem,* Computers & Operations Research 35 (2008) pp. 628-637

[12] Yu-Wan Chen, Yong-Zai Lu, Penf Chen: *Optimization with extremal dynamics for the traveling salesman problem*, Physica A 385 (2007) pp. 115-123

[13] Guan-Chun Luh, Shih-Wei Lee: *A bacterial evolutionary algorithm for the job shop scheduling problem,* Journal of the Chinese Institute of Industrial Engineers, Vol. 23, No. 3, (2006) pp. 185-191

[14] Held, M.R., Karp R.M.: *A dynamic programming approaches to sequencing problems*, Journal of the Society of Industrial and Applied Mathematics 10 (1962) pp. 196-210

[15] Ammar,E.E., Youness, E.A.: *Study on multiobjective transportation problem with fuzzy numbers*, Applied Mathematics and Computation 166 (2005) pp. 241-253

[16] Oussalah, M.: *On the compability between defuzzification and fuzzy arithmetic operations*, Fuzzy Sets and Systems 128 (2002) pp. 247-260

[17] Kikuchi, S., Chakroborty, P.: *Place of possibility theory in transportation analysis*, Transportation Research Part B 40 (2006) pp 595-615

[18] Zadeh, L.A.: *The concept of linguistic variable and its application to approximate reasoning*, Inform. Sci. Part 1, 8 (1975) 199-249, Part 2, 301-357, Part 3, 9 (1975) 43-80.

[19] Moscato P. *On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts: Towards Memetic Algorithms*. Caltech Concurrent Computation Program, Tech. Rep., California, 1989.

[20] Botzheim J., Cabrita C., Kóczy L. T., and Ruano A. E., "*Fuzzy rule extraction by bacterial memetic algorithm*," IFSA 2005, Beijing, China, 2005, pp.1563-1568.

[21] Ye J., Tanaka M., Tanino T.: *Eugenics-based genetic algorithm, IEICE Transactions on Information and Systems*, Vol. E79-D No. 5 pp. 600-607.

[22] Lin S., Kernighan B. W.: *An Effective Heuristic Algorithm for the Traveling-Salesman Problem*, Operations Research, 21, pp. 498-516 (1973).

[23] Földesi P., Botzheim J.: *Solution for modified traveling salesman problem with variable cost matrix using bacterial evolutionary algorithm*. Acta Technica Jaurinensis Series Logistica, 1(2):159–171, 2008.

Feature Selection Using Fuzzy Objective Functions

Susana M. Vieira[1] João M. C. Sousa[1] Uzay Kaymak[2]

1. Instituto Superior Técnico, Technical University of Lisbon
Lisbon, Portugal
2.Econometric Institute, Erasmus School of Economics, Erasmus University of Rotterdam
Rotterdam, The Netherlands
Email: susana@dem.ist.utl.pt, jmsousa@ist.utl.pt, u.kaymak@ieee.org

Abstract *— One of the most important stages in data preprocessing for data mining is feature selection. Real-world data analysis, data mining, classification and modeling problems usually involve a large number of candidate inputs or features. Less relevant or highly correlated features decrease, in general, the classification accuracy, and enlarge the complexity of the classifier. Feature selection is a multi-criteria optimization problem, with contradictory objectives, which are difficult to properly describe by conventional cost functions. The use of fuzzy decision making may improve the performance of this type of systems, since it allows an easier and transparent description of the different criteria used in the feature selection process. In previous work an ant colony optimization algorithm for feature selection was presented, which minimizes two objectives: number of features and classification error. Two pheromone matrices and two different heuristics are used for each objective. In this paper, a fuzzy objective function is proposed to cope with the difficulty of weighting the different criteria involved in the optimization algorithm.*

Keywords— Feature selection, fuzzy decision functions, ant colony optimization.

1 Introduction

Feature selection has been an active research area in data mining, pattern recognition and statistics communities. The main idea of feature selection is to choose a subset of available features, by eliminating features with little or no predictive information and also redundant features that are strongly correlated. Many practical pattern classification tasks (e.g., medical diagnosis) require learning of an appropriate classification function that assigns a given input pattern (typically represented by using a vector of feature values) to one of a set of classes. The choice of features used for classification has an impact on the accuracy of the classifier and on the time required for classification.

The challenge is selecting the minimum subset of features with little or no loss of classification accuracy. The feature subset selection problem consists of identifying and selecting a useful subset of features from a larger set of often mutually redundant, possibly irrelevant, features with different associated importance [1].

Like many design problems, feature selection problem, is characterized by multiple objectives, where a trade-off amongst various objectives must be made, leading to under or over-achievement of different objectives. Moreover, some flexibility may be present for specifying the constraints of the problem. Furthermore, some of the objectives in decision making may be known only approximately. Fuzzy set theory provides ways of representing and dealing with flexible or soft constraints, in which the flexibility in the constraints can be exploited to obtain additional trade-off between improving the objectives and satisfying the constraints.

Various fuzzy optimization methods have been proposed in the literature in order to deal with different aspects of soft constraints. In one formulation of fuzzy optimization due to Zimmermann [2], which is used in the rest of this paper, concepts from Bellman and Zadeh model of fuzzy decision making [3] are used for formulating the fuzzy optimization problem.

In this paper is used a feature selection algorithm which is based in ant colony optimization. The ant feature selection algorithm uses two cooperative ant colonies, which are used to cope with two different objectives. The two objectives we consider are minimizing the number of features and minimizing the classification error. Two pheromone matrices and two different heuristics are used for each objective. These goals are translated into fuzzy sets.

The paper is organized as follows. Section 2 presents a brief description of fuzzy optimization. A brief consideration of the fuzzy models we use is presented in Section 3. The ACO feature selection algorithm is presented in Section 4. In Section 5 the results are presented and discussed. Some conclusions are drawn in Section 6 and the possible future work is discussed.

2 Fuzzy Optimization

Fuzzy optimization is the name given to the collection of techniques that formulate optimization problems with flexible, approximate or uncertain constraints and goals by using fuzzy sets. In general, fuzzy sets are used in two different ways in fuzzy optimization.

1. To represent uncertainty in the constraints and the goals (objective functions).
2. To represent flexibility in the constraints and the goals.

In the first case, fuzzy sets represent generalized formulations of intervals that are manipulated according to rules similar to interval calculus by using the α-cuts of fuzzy sets. In the second case, fuzzy sets represent the degree of satisfaction of the constraints or of the aspiration levels of the goals, given the flexibility in the formulation. Hence, the constraints (and the goals) that are essentially crisp are assumed to have some flexibility that can be exploited for improving the optimization objective. This framework is suitable for the representation of interaction and possible trade-off amongst the constraints and the objectives of the optimization [4].

ISBN: 978-989-95079-6-8

2.1 General formulation

The general formulation for *fuzzy optimization* in the presence of flexible goals and constraints is given by

$$\begin{array}{ll} \underset{\mathbf{x}\in X}{\text{fuzzy maximize}} & [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x})] \\ \text{subject to} & g_i(\mathbf{x}) \tilde{\leq} 0, \quad i = 1, 2, \ldots, m. \end{array} \tag{1}$$

In (1), the tilde sign denotes a fuzzy satisfaction of the constraints. The sign $\tilde{\leq}$ thus denotes that $g_i(\mathbf{x}) \leq 0$ can be satisfied to a degree smaller than 1. The fuzzy maximization corresponds to achieving the highest possible aspiration level for the goals $f_1(\mathbf{x})$ to $f_n(\mathbf{x})$, given the fuzzy constraints to the problem. This optimization problem can be solved by using the approach of Bellman and Zadeh to fuzzy decision making [3].

Consider a decision making problem where the decision alternatives are $\mathbf{x} \in X$. A fuzzy goal F_j, $j = 1, 2, \ldots, n$ is a fuzzy subset of X. Its membership function $F_j(\mathbf{x})$, with $F_j : X \rightarrow [0, 1]$ indicates the degree of satisfaction of the decision goal by the decision alternative $\mathbf{x}$. Similarly, a number of fuzzy constraints G_i, $i = 1, 2, \ldots, m$ can be defined as fuzzy subsets of X. Their membership functions $G_i(\mathbf{x})$, denote the degree of satisfaction of the fuzzy constraint G_i by the decision alternative $\mathbf{x} \in X$. According to Bellman and Zadeh's fuzzy decision making model, the fuzzy decision D is defined as the confluence of the fuzzy goals and constraints, i.e.

$$D(\mathbf{x}) = F_1(\mathbf{x}) \circ \cdots \circ F_n(\mathbf{x}) \circ G_1(\mathbf{x}) \circ \cdots \circ G_m(\mathbf{x}), \tag{2}$$

where $\circ$ denotes an aggregation operator for fuzzy sets. Since the goals and the constraints must be satisfied simultaneously, Bellman and Zadeh proposed to use an intersection operator, i.e. a fuzzy t-norm for the aggregation. The optimal decision alternative $\mathbf{x}^*$ is then the argument that maximizes the fuzzy decision, i.e.

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in X} D(\mathbf{x}). \tag{3}$$

The optimization problem is then defined by

$$\max_{\mathbf{x}\in X} [F_1(\mathbf{x}) \wedge \cdots \wedge F_n(\mathbf{x}) \wedge G_1(\mathbf{x}) \wedge \cdots \wedge G_m(\mathbf{x})]. \tag{4}$$

Note that both the goals and the constraints are aggregated. Hence, the goals and the constraints are treated equivalently, which is why the model is said to be symmetric. The symmetric model is not always appropriate, however, since the aggregation of the goals and the constraints may have different requirements. Often, for example, some trade off amongst the goals is allowed or may even be desirable, which may be modeled by an averaging operation. The constraints, however, should not be violated, i.e. their aggregation must be conjunctive. In that case, the goals and the constraints cannot be combined uniformly by using a single aggregation operator. In the simplest case, the goals must be combined by using one operator and the constraints must be combined by using another operator. The aggregated results must then be combined at a higher level by using a third aggregation operator, which has to be conjunctive (i.e. both the aggregated goals and the aggregated constraints should be satisfied).

Clearly, the above formulation of fuzzy optimization is closely related to the penalty function methods known from classical optimization theory. The aggregated goals correspond to an overall objective function, which is maximized. The constraints are added to this objective function by using fuzzy t-norms, which is similar to the addition of a penalty function to an optimization objective function in classical optimization. After combining the objectives and the constraints, the resulting optimization is unconstrained, but possibly non-convex. Furthermore, gradient descent methods may not be suitable for the maximization due to possible and likely discontinuity in the first derivative of the final aggregated function. Derivative-free search and optimization algorithms such as simulated annealing, evolutionary algorithms or other bio-inspired algorithms, such as ant colony optimization, can be used to solve this type of optimization problems.

2.2 Weighted aggregation in fuzzy optimization

Weighted aggregation has been used quite extensively especially in fuzzy decision making, where the weights are used to represent the relative importance that the decision maker attaches to different decision criteria (goals or constraints). Almost always an averaging operator has been used for the weighted aggregation, such as the generalized means [5].

The averaging operators are suitable for modeling compensatory aggregation. They are not suitable, however, for modeling simultaneous satisfaction of aggregated criteria. Since the goal in fuzzy optimization is the simultaneous satisfaction of the optimization objectives and the constraints, t-norms must be used to model the conjunctive aggregation. In order to use the weighted aggregation in fuzzy optimization, weighted aggregation using t-norms must thus be considered.

The axiomatic definition of t-norms does not allow for weighted aggregation. In order to obtain a weighted extension of t-norms, some of the axiomatic requirements must be dropped. Especially the commutativity and the associativity properties must be dropped, since weighted operators are by definition not commutative.

Weighted t-norms. In [4], the authors used weighted counterparts of several t-norms for fuzzy optimization. The specific operators considered are the weighted extension of the product t-norm given by

$$D(\mathbf{x}, \mathbf{w}) = \prod_{i=1}^{m} [G_i(\mathbf{x})]^{w_i}, \tag{5}$$

the extension of the Yager t-norm given by

$$D(\mathbf{x}, \mathbf{w}) = \max\left(0, 1 - \sqrt[s]{\sum_{i=1}^{m} w_i (1 - G_i(\mathbf{x}))^s}\right), \quad s > 0. \tag{6}$$

The term fuzzy optimization in the remainder of this paper also refers to a formulation in terms of the flexibility of the goals.

3 Fuzzy Models for Classification

Fuzzy modeling for classification, is a technique that allows an approximation of nonlinear systems when there is none or few knowledge of the system to be modeled [6]. The fuzzy modeling approach has several advantages when compared to

ISBN: 978-989-95079-6-8

other nonlinear modeling techniques. In general, fuzzy models can provide a more transparent model and can also give a linguistic interpretation in the form of rules, which is appealing when dealing with classification systems. Fuzzy models use rules and logical connectives to establish relations between the features defined to derive the model. This paper uses Takagi-Sugeno (TS) fuzzy models [7], which consist of fuzzy rules where each rule describes a local input-output relation, typically in an affine form. The affine form of a TS model is given by:

$$R_i : \textbf{If } x_1 \text{ is } A_{i1} \textbf{ and } \ldots \textbf{ and } x_n \text{ is } A_{in} \textbf{ then } y_i = a_{i1}x_1 + \ldots + a_{in}x_n + b_i\,, \quad (7)$$

where $i = 1, \ldots, K$, K denotes the number of rules in the rule base, R_i is the i^{th} rule, $\mathbf{x} = [x_1, \ldots, x_n]^T$ is the antecedent vector, n is the number of features, $A_{i1}, \ldots, A_{in}$ are fuzzy sets defined in the antecedent space, y_i is the output variable for rule i, $\mathbf{a}_i$ is a parameter vector and b_i is a scalar offset. The consequents of the affine TS model are hyperplanes in the product space of the inputs and the output. The model output, y, can then be computed by aggregating the individual rules contribution:

$$y = \frac{\sum_{i=1}^{K} \beta_i y_i}{\sum_{i=1}^{K} \beta_i}, \quad (8)$$

where β_i is the degree of activation of the ith rule:

$$\beta_i = \prod_{j=1}^{n} \mu_{A_{ij}}(x_j), \quad (9)$$

and $\mu_{A_{ij}}(x_j) : \mathbb{R} \rightarrow [0, 1]$ is the membership function of the fuzzy set A_{ij} in the antecedent of R_i.

The performance criterion used to evaluate the fuzzy model is the classification accuracy C_a, given by the percentage of correct classifications:

$$C_a = \frac{(N_n - N_e)}{N_n} \times 100\%, \quad (10)$$

where N_n is the number of used samples and N_e is the number of classification errors in test samples (misclassifications).

4 Ant Feature Selection

Ant algorithms were first proposed by Dorigo [8] as a multi-agent approach to difficult combinatorial optimization problems, such as traveling salesman problem, quadratic assignment problem or supply chain management [9, 10]. The ACO methodology is an optimization method suited to find minimum cost paths in optimization problems described by graphs [11]. This paper presents a new implementation of ACO applied to feature selection, where the best number of features is determined automatically. In this approach, two objectives are considered: minimizing the number of features and minimizing the classification error. Two cooperative ant colonies optimize each objective. The first colony determines the number (cardinality) of features and the second selects the features based on the cardinality given by the first colony. Thus, two pheromone matrices and two different heuristics are used. A novel approach for computing a heuristic value is proposed to determine the cardinality of features. The heuristic value is computed using the Fisher discriminant criterion for feature selection [12], which ranks the features giving them a given relative importance and it is described in more detail in section 4.2.3. The best number of features is called *features cardinality* N_f. The determination of the *features cardinality* is addressed in the first colony sharing the same minimization cost function with the second colony, which in this case aggregates both the maximization of the classification accuracy and the minimization of the features cardinality. Hence, the first colony determines the size of the subsets of the ants in the second colony, and the second colony selects the features that will be part of the subsets.

The algorithm used in this paper deals with the feature selection problem as a multi–criteria problem with a single objective function. Therefore, a pheromone matrix is computed for each criterion, and different heuristics are used.

The objective function of this optimization algorithm aggregate both criteria, the minimization of the classification error rate and the minimization of the features cardinality:

$$J^k = w_1 \frac{N_e^k}{N_n} + w_2 \frac{N_f^k}{n} \quad (11)$$

where $k = 1, \ldots, g$, N_n is the number of used data samples and n is the total number of features. The weights w_1 and w_2 are selected based on experiments.

To evaluate the classification error, a fuzzy classifier is built for each solution following the procedure described in Section 3.

4.1 Fuzzy decision function

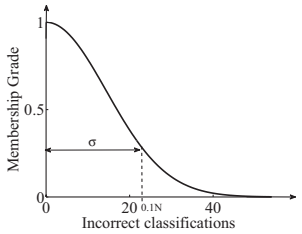
The objective function (11) can be interpreted as follows. The term containing the predicted errors indicates that these should be minimized, while the term containing the size of the features subset used indicates that the number of features should be reduced. Hence, minimizing the output errors and the size of the features subset can be regarded as forcing the model to be less complex and maintain a good performance at the same time. The parameters containing the weights, w_1 and w_2, can be changed so that the objective function is modified in order to lead to a solution where the accuracy of the model is more important or to a much simpler model where the reduction of the number of the features is imperative. This balance always depends on the application.

When fuzzy multicriteria decision making is applied to determine the objective function, additional flexibility is introduced. Each criterion ζ_i is described by a fuzzy set, where $i = 1, \ldots, T$, stands for the different criteria defined. Fuzzy criteria can be described in different ways. The most straightforward and easy way is just to adapt the criteria defined for the classical objective functions. Fig. 1 shows examples of general membership functions that can be used for the error N_e^k and for the features cardinality N_f^k, with $k = 1, \ldots, g$. In this example, the minimization of the classification error is represented by an exponential membership function, given by

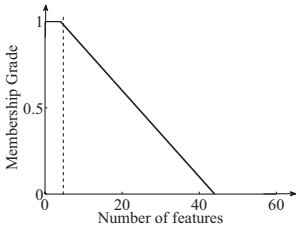
$$\mu_e^k = \exp^{-\left(\frac{N_e^k - a}{2\sigma}\right)^2} \quad (12)$$

This well-known function has the nice property of never reaching the value zero, and the membership value is still quite considerable for an error of 10%. Therefore, this criterion is

ISBN: 978-989-95079-6-8



(a) Membership function for the number of incorrect classifications.



(b) Membership function for the number of selected features

Figure 1: Membership functions for the feature selection goals.

considered to be a fuzzy goal. The features cardinality N_f can be represented, for example, by a trapezoidal membership function around zero, as shown in Fig. 1b. A reduced number of features is considered to be a desired outcome of the optimization algorithm. The membership function is defined so that for a low number of features the membership degree is one and linearly decreases to zero. The membership degree should be zero outside the maximum number of features available. The parameters defining the range of the trapezoidal membership function are application dependent. Sometimes it is convenient to make the upper limit of the membership function much lower than the maximum number of features allowed, specially if a very large number of features is being tested. In general, all the parameters of the different membership functions are application dependent. However, it is possible to derive some tuning guidelines, as will be described here. The membership functions quantify how much the system satisfies the criteria given a particular feature subset solution, bringing various quantities into a unified domain. The use of the membership functions introduces additional flexibility to the goals, and it leads to increased transparency as it becomes possible to specify explicitly what kind of solution is preferred. For instance, it becomes easier to penalize more severely a subset of features that have larger classification errors. Or if we prefer a solution with less features, a higher number of features can also be penalized.

Note that there is no need to scale the several parameters and as in (11), when fuzzy objective functions are used, because the use of membership functions introduce directly the normalization required. For this particular aspect, this feature reduces the effort on defining the objective function, when compared to classical objective functions. After the membership functions have been defined, they are combined by using a decision function, such as a parametric aggregation operator from the fuzzy sets theory (see Section 2).

Table 1: List of symbols.

Variable	Description
General	
n	Number of features
N	Number of samples
N_n	Number of samples used for validation
I	Number of iterations
K	Number of rules/clusters of the fuzzy model
N_c	Number of existing classes in database
g	Number of ants
$\mathbf{x}$	Set with all the features
$\mathbf{w}$	Subset of features selected to build classifiers
J^k	Cost of the solution for each ant k
J^q	Cost of the winner ant q
Ant colony for cardinality of features	
N_f	Features cardinality (number of selected features)
$N_f(k)$	Features cardinality of ant k
I_n	Number of iterations with same feature cardinality
α_n	Pheromone weight of features cardinality
β_n	Heuristic weight of features cardinality
τ_n	Pheromone trails for features cardinality
η_n	Heuristic of features cardinality
ρ_n	Evaporation of features cardinality
Γ_n^k	Feasible neighborhood of ant k (features cardinality availability)
$\mathcal{Q}_i$	Amount of pheromone laid in the features cardinality of the best solution
Ant colony for selecting subset of features	
$L_f^k(t)$	Feature subset for ant k at tour t
α_f	Pheromone weight of features
β_f	Heuristic weight of features
τ_f	Pheromone trails for feature selection
η_f	Heuristic of features
ρ_f	Evaporation of features
Γ_f^k	Feasible neighborhood of ant k (features availability)
$\mathcal{Q}_j$	Amount of pheromone laid in the features of the best solution

4.2 Algorithm description

Let $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$ be the set of given n features, and $\mathbf{w} = [w_1, w_2, \ldots, w_{N_f}]^T$, be a subset of features where $(\mathbf{w} \subseteq \mathbf{x})$. It is desirable that $N_f << n$. Table 1 describes the variables used in the algorithm.

4.2.1 Probabilistic rule

Consider a problem with N_f nodes and two colonies of g ants. First, g ants of the first colony randomly select the number of nodes N_f to be used by the g ants of the second colony. The probability that an ant k chooses the features cardinality $N_f(k)$ is given by

$$p_i^k(t) = \frac{[\tau_{n_i}]^{\alpha_n} \cdot [\eta_{n_i}]^{\beta_n}}{\sum_{l \in \Gamma_n^k} [\tau_{n_l}]^{\alpha_n} \cdot [\eta_{n_l}]^{\beta_n}} \tag{13}$$

where τ_{n_i} is the pheromone concentration matrix and η_{n_i} is the heuristic function matrix, for path (i). The values of the

ISBN: 978-989-95079-6-8

pheromone matrix are limited to $[\tau_{n_{min}}, \tau_{n_{max}}]$, with $\tau_{n_{min}} = 0$ and $\tau_{n_{max}} = 1$. Γ_n^k is the feasible neighborhood of ant k (available number of features to be selected), which acts as the memory of the ants, and contains all the trails that the ants have not passed and can be chosen. The parameters α_n and β_n measure the relative importance of trail pheromone and heuristic knowledge, respectively.

After all the g ants from the first colony have chosen the features cardinality $N_f(k)$, each ant k from the second colony select $N_f(k)$ features (nodes). The probability that an ant k chooses feature j as the next feature to visit is given by

$$p_j^k(t) = \frac{[\tau_{f_j}(t)]^{\alpha_f} \cdot [\eta_{f_j}]^{\beta_f}}{\sum_{l \in \Gamma_f^k} [\tau_{f_l}(t)]^{\alpha_f} \cdot [\eta_{f_l}]^{\beta_f}} \quad (14)$$

where τ_{f_j} is the pheromone concentration matrix and η_{f_j} is the heuristic function matrix for the path (j). Again, the pheromone matrix values are limited to $[\tau_{f_{min}}, \tau_{f_{max}}]$, with $\tau_{f_{min}} = 0$ and $\tau_{f_{max}} = 1$. Γ_f is the feasible neighborhood of ant k (available features), which contains all the features that the ants have not selected and can be chosen. Again, the parameters α_f and β_f measure the relative importance of trail pheromone and heuristic knowledge, respectively.

4.2.2 Updating rule

After a complete tour, when all the g ants have visited all the $N_f(k)$ nodes, both pheromone concentration in the trails are updated by

$$\tau_{n_i}(t+1) = \tau_{n_i}(t) \times (1 - \rho_n) + \Delta\tau_{n_i}(t) \quad (15)$$

$$\tau_{f_j}(t+1) = \tau_{f_j}(t) \times (1 - \rho_f) + \Delta\tau_{f_j}(t) \quad (16)$$

where $\rho_n \in [0, 1]$ is the pheromone evaporation of the features cardinality , $\rho_f \in [0, 1]$ is the pheromone evaporation of the features and $\Delta\tau_{n_i}$ and $\Delta\tau_{f_j}$ are the pheromone deposited on the trails (i) and (j), respectively, by the ant q that found the best solution J^q for this tour:

$$\Delta\tau_{n_i}^q = \begin{cases} \mathcal{Q}_i & \text{if node } (i) \text{ is used by the ant } q \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

$$\Delta\tau_{f_j}^q = \begin{cases} \mathcal{Q}_j & \text{if node } (j) \text{ is used by the ant } q \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

The number of nodes $N_f(k)$ that each ant k has to visit on each tour t is only updated every I_n tours (iterations), in order to allow the search for the best features for each features cardinality N_f. The algorithm runs I times. Both colonies share the same cost function given in (11).

4.2.3 Heuristics

The heuristic value used for each feature (ants visibility) for the second colony, is computed as

$$\eta_{f_j} = 1/N_{e_j} \quad (19)$$

for $j = 1, \ldots, n$. For the features cardinality (first colony), the heuristic value is computed using the

Fisher discriminant criterion for feature selection [12]. Considering a classification problem with two possible classes, class 1 and class 2, the Fisher discriminant criterion is described as

$$F(i) = \frac{|\mu_1(i) - \mu_2(i)|^2}{\sigma_1^2 + \sigma_2^2} \quad (20)$$

Algorithm 1 Ant Feature Selection

```
/*Initialization*/
set the parameters ρ_f, ρ_n, α_f, α_n, β_f, β_n, I, I_n, g.
for t = 1 to I do
  for k = 1 to g do
    Choose the subset size N_f(k) of each ant k using (13)
  end for
  for l = 1 to I_n do
    for k = 1 to g do
      Build feature set L_f^k(t) by choosing N_f(k) features using (14)
      Compute the fuzzy model using the L_f^k(t) path selected by ant k
      Compute the cost function J^k(t)
      Update J^q
    end for
    Update pheromone trails τ_{n_i}(t+1) and τ_{f_j}(t+1), as defined in (15) and (16).
  end for
end for
```

where $\mu_1(i)$ and $\mu_2(i)$ are the mean values of feature i for the samples in class 1 and class 2, and σ_1^2 and σ_2^2 are the variances of feature i for the samples in class 1 and 2. The score aims to maximize the between-class difference and minimize the within-class spread. Other currently proposed rank-based criteria generally come from similar considerations and show similar performance [12]. Since our goal is to work with several classification problems, which can contain two or more possible classes, a one versus-all strategy is used to rank features. Thus, for a C-class prediction problem, a particular class is compared with the other $C-1$ classes that are considered together. The features are weighted according to the total score summed over all C comparisons:

$$\sum_{j=1}^{C} F_j(i), \quad (21)$$

where $F_j(i)$ denotes the Fisher discriminant score for the i^{th} feature at the j^{th} comparison. Algorithm 1 presents the description of the ant feature selection algorithm.

5 Experimental Results

The effectiveness of the proposed method is applied to a data set taken from the UCI repository [13].

The classification error rates of the used classifiers are obtained by performing 10 independent runs. The data set is divided in test and training sets. The experimental results are presented as the best, the worst and the mean value of correct classifications C_a out of ten independent runs. These ten runs were obtained using the same test data set.

Wine data set The wine data set is a widely used classification data available online in the repository of the University of California [13], and contains the chemical analysis of 178 wines grown in the same region in Italy, derived from three different cultivars. Thirteen continuous attributes are available for classification: alcohol, malic acid, ash, alkalinity of ash, magnesium, total phenols, flavanoids, nonflavanoids phenols, proanthocyanism, color intensity, hue, OD280/OD315 of

ISBN: 978-989-95079-6-8

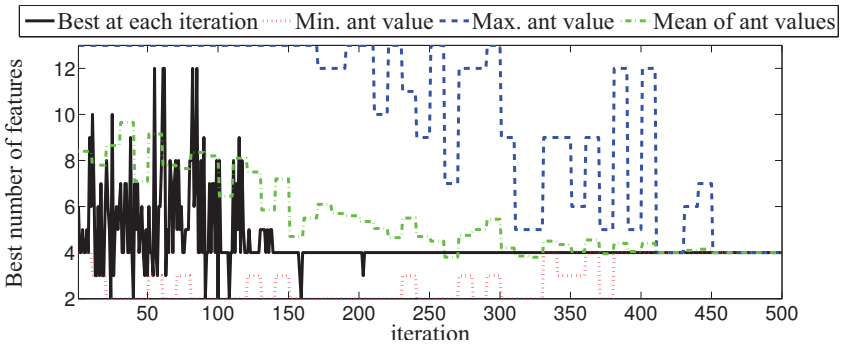


Figure 2: Convergence of the best number of feature during an algorithm run.

Table 2: Classification rates on the Wine data.

Method	**Number of features**	**Classification accuracy** (%)		
		Max.	**Mean**	**Min.**
AFS classic obj. func.	4-8	100	99.8	98.8
AFS fuzzy obj. func.	4	100	99.7	98.6

dilluted wines and proline. This data set has 13 features, three different classes and 178 samples. The AFS algorithm with fuzzy objective function is applied to select the relevant features within the wine classification data set and is compared to the same algorithm with a non fuzzy objective function.

As can be seen in Table 2, the obtained results are better than those in feature selection with classical objective function, once fewer features are selected and because the algorithm always converges to the same number of features.

An example of the process of the ant feature selection with fuzzy objectives searching for optimal solutions for wine data set is given in Fig. 2, where it is possible to observe how all the ants converge to the same solution.

6 Conclusions

A fuzzy objective function for ant feature selection is proposed in this paper. The problem is divided into two objectives: choosing the features cardinality and selecting the most relevant features. The feature selection algorithm uses fuzzy classifiers. The proposed algorithm was applied to a well known classification database that is considered a benchmark. The performance of the proposed algorithm was compared to previous works. The ant based feature selection algorithm yielded similar or better classification rates and the convergence of the solution is better than the previous approach.

In the near future, the proposed feature selection algorithm will be applied to classification problems with a larger number of features.

Acknowledgements

This work is supported by the Portuguese Government and FEDER under the programs: Programa de financiamento Plurianual das Unidades de I&D da FCT (POCTI-SFA-10-46-IDMEC) and by the FCT grant SFRH/25381/2005, Fundação para a Ciência e a Tecnologia, Ministério do Ensino Superior, da Ciência e da Tecnologia, Portugal.

References

[1] H. Motoda H. Liu. *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers, 1998.

[2] H. J. Zimmermann. Description and optimization of fuzzy systems. *International Journal of General Systems*, 2:209–215, 1976.

[3] R. E. Bellman and L. A. Zadeh. Decision-making in a fuzzy environment. *Management Science*, 17(4):141–164, 1970.

[4] U. Kaymak and J.M. Sousa. Weighted constraint aggregation in fuzzy optimization. *Constraints*, 8(1):61–78, January 2003.

[5] U. Kaymak and H. R. van Nauta Lemke. A sensitivity analysis approach to introducing weight factors into decision functions in fuzzy multicriteria decision making. *Fuzzy Sets and Systems*, 97(2):169–182, July 1998.

[6] L. F. Mendonça, S. M. Vieira, and J. M. C. Sousa. Decision tree search methods in fuzzy modeling and classification. *International Journal of Approximate Reasoning*, 44(2):106–123, 2007.

[7] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15(1):116–132, 1985.

[8] M. Dorigo. *Optimization, Learning and Natural Algorithms (in Italian)*. PhD thesis, 1992.

[9] C. A. Silva, J. M. C. Sousa, and T. A. Runkler. Rescheduling and optimization of logistic processes using GA and ACO. *Engineering Applications of Artificial Intelligence*, 21(3):343–352, 2007.

[10] C. A. Silva, J. M. C. Sousa, T. A. Runkler, and J. M. G. Sá da Costa. Distributed optimization of a logistic system and its suppliers using ant colony optimization. *International Journal of Systems Science*, 37(8):503–512, 2006.

[11] Marco Dorigo, M. Birattari, and T. Stützle. Ant colony optimization. *IEEE Computational Intelligence Magazine*, 1(4):28–39, 2006.

[12] R. O. Duda, P. E. Hart, and D. G Stork. *Pattern Classification, Second Edition*. Wiley–Interscience Publication, 2001.

[13] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007.

ISBN: 978-989-95079-6-8

Clustering of Fuzzy Shapes by Integrating Procrustean Metrics and Full Mean Shape Estimation into K-Means Algorithm

Vasile Georgescu

Faculty of Economics and Business Administration, University of Craiova
Craiova, Romania
Email: vgeo@central.ucv.ro, v_geo@yahoo.com

Abstract—*In this paper we propose a generalization of K-means algorithm, which is adapted to integrate Procrustean metrics and full mean shape estimation, with the aim of clustering objects with either multiple or fuzzy contours. First we are concerned with the representation of fuzzy shapes and introduce appropriate shape metrics and descriptors. Next, we discuss Procrustean methods for aligning shapes, finding mutual dissimilarities and estimating shape class centroid. In the case of multiple-contour crisp shapes, we can benefit from the Extended Orthogonal Procrustes method to find mutual distances between shape pairs and from the Generalized Orthogonal Procrustes technique to estimate the Procrustes mean shape of a collection of shapes. On the other hand, dealing with the case of fuzzy shapes needs more advanced Procrustean techniques to consider weighted distances between points placed on* α *– level contours with different membership degrees. This leads to solve a Weighted Orthogonal Procrustes problem, which typically needs to introduce a weighting matrix of residuals (distances). As an application, we suggest using such methods to cluster ultrasound images of lymph nodes, which typically appear as double-contour shapes.*

Keywords— Clustering of fuzzy shapes, Fuzzy shape metrics and descriptors, Procrustes analysis, Mixing K-means algorithm with Procrustean metrics and mean shape estimation.

1 Shape analysis

Shapes and textures are extremely important features in human as well as machine vision and understanding systems. Shape analysis is concerned with two main classes of algorithms: boundary-based (when only the shape boundary points are used for the description) and region-based (when the whole interior of a shape is used). There are many imaging applications where image analysis can be reduced to the analysis of shapes, in contrast to texture analysis. However, many shape/edge detection techniques use texture information during the segmentation process.

There are several methods for extracting data from shapes, each with their own benefits and weaknesses. These include measurement of lengths and angles, landmark analysis and outline analysis. A landmark is a point of correspondence on each object that matches higher dimensionalities between and within populations. Landmark placement consists of locating a finite number of points on the outline.

More advanced techniques have been designed for semiautomatic and automatic feature extractions. Active contour modeling techniques are commonly used for shape analysis and detection. Some of the techniques for texture feature extraction use gray level co-occurrence matrices, fractal dimension, etc.

Morphometric analysis aims to describe the shape of an object in a way that removes extraneous information and thereby facilitates comparison between different objects. In these terms, a shape is referred to as an invariant to similarity transformations (such as scaling, rotation and translation).

The image fuzzification plays a pivotal role in all image processing systems. Several kinds of image fuzzification can be distinguished:

- histogram-based grey-level fuzzification (e.g. brightness in image enhancement);
- local fuzzification (e.g. edge detection);
- feature fuzzification (scene analysis, object recognition).

2 Representation of fuzzy shapes

2.1 Crisp shapes

Crisp shapes represent objects with crisp borders. Furthermore, if a texture is associated with the object, it has to be uniformly represented (e.g. a digitized image, where all pixels are classified as object pixels, or as background pixels).

The coordinates of selected landmarks for a crisp shape can be arranged in a $n \times p$ configuration matrix A, or equivalently on a $np \times 1$ configuration vector $a = vec(A)$.

2.2 Continuous fuzzy shapes

This paper primarily focuses on the representation of fuzzy shapes with fuzzy contour, which are commonly obtained through fuzzy segmentation techniques. In particular, we also consider the case of crisp shapes with multiple contours. In the same way as it is convenient to model binary images as crisp objects, it is possible to model grey-level images directly as fuzzy sets. If the grey-level values of an image are scaled to be between 0 and 1, the grey-level of a pixel can be seen as its membership to the set of high-valued (bright) pixels.

Fuzziness of an image representation can arise from various reasons, such as limited acquisition conditions (scanning resolution), but also as intrinsic property of the image, which may have imprecise borders. In such cases, pixels close to the border of the object have assigned to them a fuzzy membership value according to the extent of their belongingness to the object.

ISBN: 978-989-95079-6-8

Continuous fuzzy shapes can be described as fuzzy geometric objects. A continuous fuzzy geometric object A in $\Re^p$ is defined as a set of pairs $\{(x, \mu_A(x)) \mid x \in \Re^p\}$ where $\mu_A : \Re^p \to [0,1]$ is the membership function of A in $\Re^n$. It is assumed to have a bounded support.

An alternative representation of fuzzy geometric objects is given by a set of $\alpha-$cuts: $C(A) = \{A_\alpha \mid \alpha \in [0,1]\}$, where $A_\alpha = \{x \in \Re^p \mid \mu_A(x) \geq \alpha\}$ is a crisp object, whose $\alpha-$level contour is obtained for $\mu_A(x) = \alpha$. As a characteristic of fuzzy geometric objects, the membership function is non-increasing away from the interior of the object. For example, in figure 1 is shown a fuzzy disk. Its core is a crisp disk defined by $A_1 = \{x \in \Re^2 \mid (x_1^2 + x_2^2 \leq r_1)\}$ and its contour is the circle defined by $A_1^C = \{x \in \Re^2 \mid (x_1^2 + x_2^2 = r_1)\}$, where r_1 is the length of the corresponding radix. In general, for any $\alpha \in [0,1]$, the $\alpha-$cut is defined by $A_\alpha = \{x \in \Re^2 \mid (x_1^2 + x_2^2 \leq r_\alpha)\}$, and the $\alpha-$level contour is defined by $A_\alpha^C = \{x \in \Re^2 \mid (x_1^2 + x_2^2 = r_\alpha)\}$.

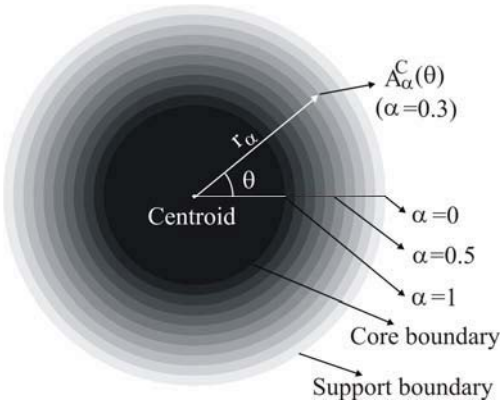


Figure 1: A fuzzy disk: centroid, core, support, α-level contours, radial distance

A shape descriptor based on a one-dimensional functional representation of the two-dimensional shape boundary is called a signature of the shape. The simplest way to generate a signature is to express the radial distance from the centroid to the boundary, as a function of the angle. This is called the centroid distance function. Thus, for crisp objects, the shape signature function corresponds to the Euclidean distance between each boundary point $A(t) = (x(t), y(t))$ and the centroid $A_c = (x_c, y_c)$ of the shape:

$$CD(t) = \sqrt{(x(t) - x_c)^2 + (y(t) - y_c)^2}$$

This shape signature function based on the centroid distance is a convenient choice in the case of star-shaped objects with respect to the centroid (i.e., for each point $y \in A$, the line segment connecting y with the centroid is contained in A).

In the case of a fuzzy object, boundary points are not strictly defined; there is a progressive transition of the membership values from the support outline to the core outline.

The shape signature can be generalized for a continuous fuzzy shape in two possible ways:

- as s radial integral of the membership function:

$$CD_{fuzzy-1}(t) = \int_{A_c}^{A(t)} \mu_A(x(\rho), y(\rho))\, d\rho$$

where $\rho = \rho(t)$ is a parameterization of the straight path between a boundary point and the centroid.

- as an average signature obtained from the $\alpha-$cuts:

$$CD_{fuzzy-2}(t) = \int_0^1 CD_\alpha(t)\, d\alpha$$

where fuzzy star-shaped objects are considered, with all the boundaries of there $\alpha-$cuts jointly indexed by the same parameter t.

A path π in $\Re^p$ from a point $x \in \Re^p$ to another point $y \in \Re^p$ is a continuous function $\pi : [0,1] \to \Re^p$, such that $\pi(0) = x$ and $\pi(1) = y$. The length of a path π in A, denoted by $\Pi_A(\pi)$, is the value of the following integration

$$\Pi_A(\pi) = \int_0^1 \mu_A(\pi(t)) \left| \frac{d\pi(t)}{dt} \right| dt$$

where $\Pi_A(\pi)$ is the integral of membership values (in A) along π.

2.3 Discrete fuzzy shapes

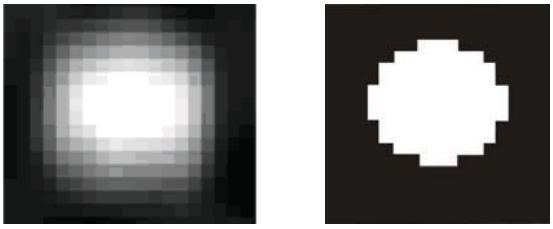
Discrete fuzzy objects can arise from the digitization of scanned images. Generally, the gray-level images will be thresholded to calculate geometrical measures. Since the images or their segments have ill-defined or non-crisp boundaries, it is sometimes appropriate to consider them as fuzzy sets. The concept of fuzzy digital geometry has been introduced by Rosenfeld and plays a key role in many image processing applications: "The standard approach to image analysis and recognition begins by segmenting the image into regions and computing various properties of and relationships among these regions. However, the regions are not always 'crisply' defined; it is sometimes more appropriate to regard them as fuzzy subsets of the image... It is not always obvious how to measure geometrical properties of fuzzy sets, but definitions have been given and basic properties established for a variety of such properties and relationships, including connectedness and surroundedness, convexity, area, perimeter and compactness, extent and diameter" ([12]).

The application areas of fuzzy geometry are image representation, enhancement and segmentation. The process of converting the input image into a fuzzy set by indicating, for each pixel, the degree of membership to the object, is referred to as "fuzzy segmentation". The most straightforward way to perform fuzzy segmentation is to scale grey-levels of an image to be between 0 and 1. Such

ISBN: 978-989-95079-6-8

grey levels reflect the area coverage of a pixel by the object, and can be naturally used as membership values. However, in most cases, more advanced segmentation methods are required, especially since it is rarely sufficient to use only the brightness of pixels to calculate fuzzy membership values. For example, fully segmented image can be generated by combining the optimum automatic thresholding procedure with edge detection to produce continuously connected object border.

The object of interest is represented as a discrete spatial fuzzy subset of a grid. It should be noted, however, that the discrete fuzzy objects obtained from the digitization of scanned images (say, using a grey-level scale) are affected by multiple distortions, due to limited representation resolution. Consequently, their properties are significantly different with respect to those of corresponding continuous fuzzy objects. Figure 2 shows a discrete fuzzy disk (a) and, for comparison, its crisp counterpart (b).



(a) discrete fuzzy disk (b) crisp (binary) disk

Figure 2: Thresholding a fuzzy (gray-level) image: pixels with a membership degree below the threshold are lost

The mapping $\mu_A : \Re^p \to [0,1]$ of a continuous fuzzy shape becomes, by discretization, $\mu_A : Z^p \to \left\{0, \frac{1}{k}, \ldots, 1\right\}$, where k is the maximal number of grey levels available (e.g., k = 255 for 8-bit pixel representation).

A configuration matrix for a discrete p-dimensional fuzzy shape A can be represented by vertical concatenation of its $\alpha-$level contours into a $nk \times p$ block matrix. Each one of the k sub-matrices defined at each level α collects $n \times p$ landmarks: $\mathrm{A}_\alpha = \left(x_1^A(\alpha) \quad x_2^A(\alpha) \quad \ldots \quad x_p^A(\alpha)\right)$, where $\alpha \in \{\alpha_i \mid 0 = \alpha_1 < \cdots < \alpha_i < \cdots < \alpha_k = 1\}$.

A similar $nk \times p$-dimensional configuration matrix can be defined for a crisp shape with multiple (say k) contours.

The shape signature can be also generalized for a discrete fuzzy shape in two possible ways:

- using the distance between the boundary points and the centroid:

$$CD_{discrete\_fuzzy-1}(k) = \mu_A(x_c, y_c) + \\ + \sum_{j=1}^{N_k} \delta_k(j) \cdot \mu_A(x_k(j), y_k(j))$$

- as an average signature obtained from the $\alpha-$cuts:

$$CD_{discrete\_fuzzy-2}(k) = \frac{1}{\alpha_{total}} \sum_{\alpha=1}^{\alpha_{total}} CD_{\alpha-resampled}(k)$$

where $CD_{\alpha-resampled}(k)$ is the k th sample of the resampled signature obtained for one $\alpha-$cut.

2.4 Lymph nodes as an example of crisp double-contour shapes

The ultrasound image of a lymph node (see figure 3) is a typical example of a crisp double-contour shape. It appears as an ovoid-shaped masse with an echogenic center, representing the medullary, and a peripheral, hypoechogenic cortical region, interrupted on the hyllum, which give him a reniform shape. Usually, the normal lymph nodes present a thin cortical peripheral zone, while benign inflammatory changes in lymphadenitis may enlarge the node but with preservation of the ovoidal shape and of the ratio cortical/medullary thickness less than 1.0. Malignant metastatic or infiltrated nodes are more apparent than normal ones as they become larger, rounder, and more uniformly hypoechoic by the regularly/irregularly thickening of the cortical zone with progressive restriction of the hyperechogenic medullar area. The Computed Aided Diagnosis develops ultrasound applications, especially for breast imaging, but the complete characterization must include the analysis of the satellite lymph nodes appearance. Because of the large variability of the shape and cortical-to-medullary ratio, computer vision applications are needed to make possible the automatic diagnosis, especially in breast cancer screening.

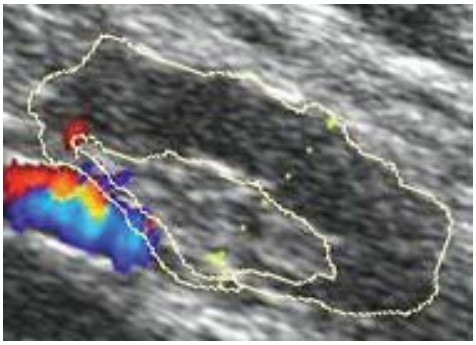


Figure 3: A crisp double-contour shape: the ultrasound image of a lymph node

3 Procrustean shape analysis

A configuration matrix A is not a proper shape descriptor, because it is not pose invariant. For any similarity transformation, i.e. $s \in \Re^+$, $R \in SO(p)$ (the special orthogonal group, i.e. R is $(p \times p)$ matrix, s.t. $R'R = I$) and $t \in \Re^p$, the configuration given by $sAR + 1_p t'$ describes the same shape as A, where 1_p is the $p \times 1$ vector $(1\ 1 \ldots 1)'$. To obtain a true shape representation, location, scale and rotational effects need to be filtered out. This is carried out by shape alignment, i.e. by establishing a *coordinate reference*, commonly known as *pose*. A very popular alignment procedure is *Procrustes shape analysis,* which provides a measure, Procrustes distance, that quantifies the dissimilarity of two configurations, and which is invariant with respect to translation, scaling, and rotation.

ISBN: 978-989-95079-6-8

Procrustes shape analysis also provides a way to define the average shape, the Procrustes mean shape, which can be viewed as a representative class template.

The *Extended Orthogonal Procrustes* (EOP) problem is a least squares method for fitting a given configuration matrix A to another given matrix B. It is based on the functional model $E = sAR + 1_p t' - B$ and consists of minimizing the Procrustes distance between A and B (i.e. $\|E\|_F^2$), under choice of unknown similarity transformation parameters R, t and s. This leads to solving the problem $\min_{R,t,s} E'E$, subject to the orthogonality restriction $R'R = I$.

Generalized Orthogonal Procrustes (GOP) *analysis* is a technique that provides least-squares correspondence of more than two model points. The solution of the problem can be thought as the search of the unknown optimal matrix W (also named *consensus matrix*), defined as follows:

$$M + E_i = \hat{A}_i = s_i A_i R_i + 1_p t_i'; \quad i = 1, \ldots, m$$
$$vec(E_i) \sim N\left(0, \Sigma = \sigma^2 \left(Q_n \otimes Q_p\right)\right)$$

where E_i is the random error matrix in normal distribution, Σ is the covariance matrix, Q_n is the cofactor matrix of the n points, Q_p is the cofactor matrix of the p coordinates of each point, $\otimes$ stands for the Kronecker product, and σ^2 is the variance factor.

Let $C = \sum_{i=1}^{m} \hat{A}_i / m$ be the *geometrical centroid* of the transformed matrices. Therefore, Generalized Orthogonal Procrustes problem can be solved minimizing

$$m \sum_{i=1}^{m} \left\| \hat{A}_i - C \right\|^2 = m \sum_{i=1}^{m} tr\left\{ \left(\hat{A}_i - C\right)' \left(\hat{A}_i - C\right) \right\}$$

Crosilla and Beinat (2002) proved that the shape mean (centroid) C corresponds to the least squares estimation $\hat{M}$ of the true value M: $C = \hat{M} = \sum_{i=1}^{m} \hat{A}_i$.

In the case of *multiple-contour crisp shapes* we can benefit from the *Extended Orthogonal Procrustes* method in order to find mutual distances between shape pairs and from the *Generalized Orthogonal Procrustes* technique in order to estimate the Procrustes mean shape of a collection of shapes. This is illustrated in figures 4 and 5.

On the other hand, dealing with the case of fuzzy shapes needs more advanced Procrustean techniques, which allow us to consider *weighted distances* between points placed on α – level contours with different membership degrees. This leads to solve a *Weighted Orthogonal Procrustes* (WOP) problem.

A weighting matrix W of the residual E (defined above) is now introduced and the minimization problem becomes:

$$\min \|W E\|_F^2$$

subject to orthogonality restriction $R'R = I$, $\det(R) = 1$.

Typically, an iterative method is needed to derive a solution to WOP.

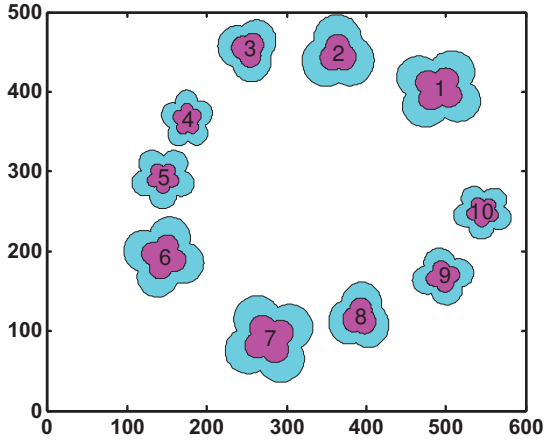


Figure 4: Ten double-contour star-shaped 2D objects with 3, 4 and 5 "lobes"

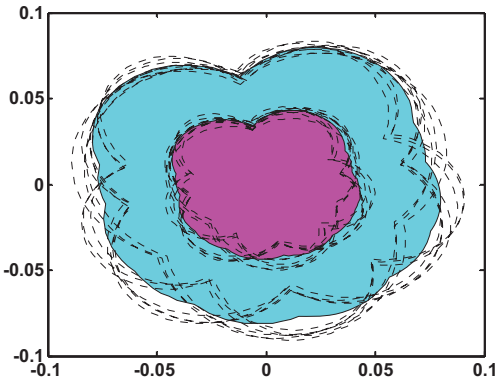


Figure 5: Procrustes mean shape (shape centroid)

4 A generalization of k-means algorithm for clustering fuzzy shapes

K-means is a commonly used data clustering for partitioning data points into disjoint groups such that data points belonging to same cluster are similar, while data points belonging to different clusters are dissimilar. The main idea is to define *k* centroids, one for each cluster, and to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, an early groupage is done. Next, we need to re-calculate *k* new centroids of the clusters resulting from the previous step. After we have these *k* new centroids, a new binding has to be done between the same data set points and the nearest new centroid. We continue this loop until no more changes are done.

Clustering of objects or images of objects, according to the shapes of their boundaries is of a key importance in computer vision and pattern recognition. This paper was intended to pay attention to this reason by proposing a generalization of K-means algorithm in order to integrate Procrustean metrics and full mean shape estimation, in a way

ISBN: 978-989-95079-6-8

making it able of clustering objects with either multiple or fuzzy contours.
We first present the algorithm in pseudo-code, as follows:

- Make initial guesses for the mean shapes v_1, v_2, …, v_k, by choosing the first k shapes from a random permutation.
- While any change still exists in any mean shape
  - Calculate all pair-wise Procrustes distances between shapes using the *Extended Orthogonal Procrustes* algorithm
  - Use the estimated mean shapes to assign the shape samples into clusters
  - For i from 1 to k
    - Replace v_i with the mean shape of all of the samples for cluster i, using the *Generalized Orthogonal Procrustes* algorithm
  - end_for
- end_while

The resulting mean shapes for each one of the 3 clusters are shown below.

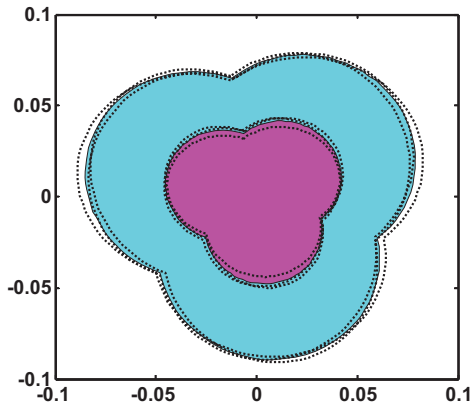


Figure 8: First cluster. Mean shape and three cluster members: {2, 3, 8}

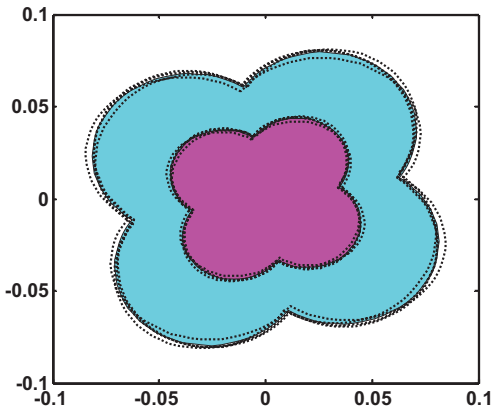


Figure 9: Second cluster. Mean shape and four cluster members: {1, 6, 7, 9}

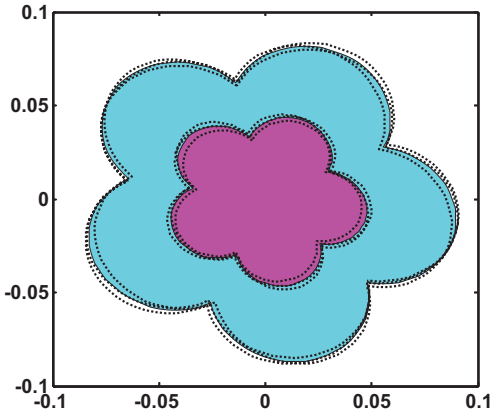


Figure 10: Third cluster. Mean shape and three cluster members: {4, 5, 10}

Our method is thus graphically validated.
As an alternative, one can use a "linkage" method to perform hierarchical shape clustering, i.e. to create a hierarchical tree of clusters starting from the symmetric matrix of Procrustean mutual distances between pairs of shapes. We obtained the same clusters as in the case of using K-means: {6, 7, 9, 1}, {2, 3, 8}, {4, 5, 10}. The dendrogram is shown below.

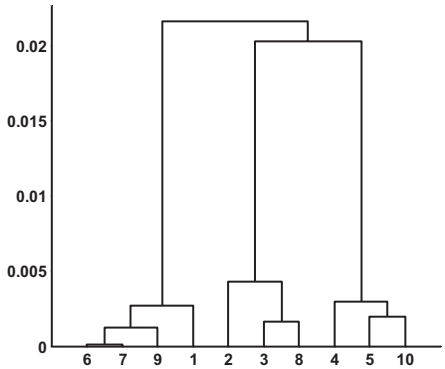


Figure 11: The dendrogram

5 Further remarks

We propose a two stage procedure for landmark placement. In the first stage, landmarks are placed arbitrarily (figure 10).

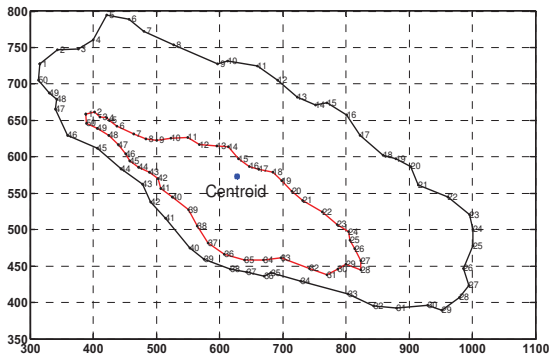


Figure 10: First stage: landmarks are placed arbitrarily

ISBN: 978-989-95079-6-8

However, we need a frame of reference to compare and display the differences in shape. Thus, we use Principal Component Analysis (PCA), which uses similarity transformations to produce a standard shape orientation, based on decomposing the overall variation of data. Each axis on a PCA representation of transformed data is an eigenvector of the covariance matrix of shape variables. In this morphospace, the first axis accounts for maximum variation in the sample, with further axes representing further decreasing variations (figure 11).

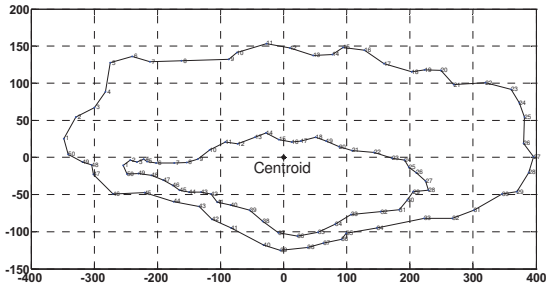


Figure 11: PCA-based alignment along the axes of maximum variation

In the second stage we replace all landmarks based on a standard procedure. The right most object point landmark is taken in the horizontal direction from the centroid. For each pair of α – level contours and for each pair of landmarks on these contours, distances are computed starting from the corresponding points, where corresponding points are those located in the same direction from the centroid. It is essential to use an angular landmark placement procedure on the boundaries, to provide appropriate correspondence between the points on the (parts) of the boundaries having different lengths, when the corresponding boundary subparts of different α – level contours are matched (figure 12).

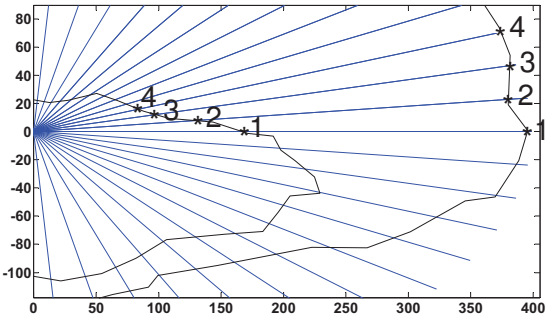


Figure 12: Corresponding points placed on radial directions from the centroid, starting from the right most object point

References

[1] I. Bloch and H. Maître. Fuzzy mathematical morphologies: A comparative study. Pattern Recognition, 28(9):1341–1387, 1995.

[2] J. Buckley and E. Eslami. Fuzzy plane geometry I: Points and lines. Fuzzy Sets and Systems, 86:179–187, 1997.

[3] P. Diamond. A note on fuzzy starshaped fuzzy sets. Fuzzy Sets and Systems, 37:193–199, 1990.

[4] J. Chanussot, I. Nyström, and N. Sladoje. Shape signatures of fuzzy star-shaped set based on distance from the centroid, *Pattern Recognition Letters*, 26(2005) 735-746

[5] B. Chaudhuri. Some shape definitions in fuzzy geometry of space. Pattern Recognition Letters, 12:531–535, 1991.

[6] F. Crosilla, A. Beinat. Use of generalized Procrustes analysis for the photogrammetric block adjustment by independent models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(3), 2002, pp. 195-209.

[7] I.L. Dryden, K.V. Mardia. *Statistical shape analysis*. John Wiley and Sons, Chichester, England, 1998, 83-107.

[8] J. Hartigan, M. Wang. A K-means clustering algorithm. *Applied Statistics*, 28, 1979, 100–108.

[9] M. Machtegael and E. Kerre. Connections between binary, grey-scale and fuzzy mathematical morphologies. Fuzzy Sets and Systems, 124:73–85, 2001.

[10] A. Rosenfeld. Fuzzy digital topology. Information and Control, 40:76–87, 1979.

[11] A. Rosenfeld. The diameter of a fuzzy set. Fuzzy Sets and Systems, 13:241–246, 1984.

[12] A. Rosenfeld. The fuzzy geometry of image subsets. Pattern Recognition Letters, 2:311–317, 1984.

[13] A. Rosenfeld and S. Haber. The perimeter of a fuzzy subset. Pattern Recognition, 18:125–130, 1985.

[14] A. Rosenfeld. Fuzzy geometry: An updated overview. *Information Sciences,* vol. 110, 1998, 127–133.

[15] P. K. Saha, F. W. Wehrli, and B. R. Gomberg. Fuzzy distance transform: Theory, algorithms, and applications. *Computer Vision and Image Understanding*, 86:171–190, 2002.

[16] P.H. Schoenemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1) , 1966, pp. 1-10.

[17] P.H. Schoenemann, R. Carroll. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, 35(2), 1970, pp. 245-255.

ISBN: 978-989-95079-6-8

Development of a Fuzzy Expert System for a Nutritional Guidance Application

Petri Heinonen, Marjo Mannelin, Hannu Iskala[1] Aki Sorsa, Esko Juuso [2]

1 Flow-Team Oy
Oulu, Finland
2 Control Engineering Laboratory, University of Oulu
Oulu, Finland
Email: firstname.lastname@flow-team.fi, firstname.lastname@oulu.fi

Abstract— *The importance of nutritional guidance grows as nutritional problems, such as obesity and type-2 diabetes, are becoming more common. Nutritional guidance is carried out by mapping the nutritional state of an individual using a food diary and by comparing the nutrition intake levels to the recommended reference values. Typically, the expert knowledge of a nutritionist is required to balance the diet.*

This paper presents a fuzzy expert system for a nutritional guidance application. Expert knowledge acquisition is carried out using variable tabulation creating the basis for the rulebase of the fuzzy system. The recommended nutrition intake values are used to generate membership functions for fuzzification. The development of a Mamdani-type fuzzy system is illustrated using a hierarchical structure. The first level system refers to groups of similar foods and the second level model defines the added and reduced foodstuffs.

The validation of the fuzzy model was carried out in three phases: first two types of sensitivity analysis were performed, and then the output was analysed with expert knowledge.. The results from the validation schemes are promising.

Keywords— expert system, fuzzy logic, nutritional guidance

1 Introduction

Unhealthy diets and physical inactivity are the main causes of non-communicable diseases such as Cardiovascular disease, type 2 diabetes, and osteoporosis. Obesity is a large and growing problem in developed countries. [1] There have been regular surveys of Finnish health behaviour since 1978. According to the latest survey, over 50 percent of Finnish men and women are overweight [2]. The best way to avoid nutritional problems is to have a well-balanced diet and to ensure the sufficient intake of essential nutrients [3].

Individual nutrition guidance is becoming more and more important as problems with obesity and type-2 diabetes are increasing rapidly. The nutrition guidance is based on information from food diaries kept by the users where they record all the foods they consume. This information is converted into nutrient intake levels with the help of a food composition database. The intake levels are compared to the recommended intake levels and the expert knowledge of a nutritionist is used to balance the diet to meet the recommendations. There are restrictions, for example allergies, diseases, and special diets which must also be taken in to account when balancing diets.

Some software has been developed for the purpose of nutritional guidance. For example, a French research group developed a nutrition software application for hospital use [4]. They use fuzzy arithmetic to handle imprecise and uncertain information in the computations. Fuzzy logic is first used by the application to determine how well the current meal is balanced, after which a heuristic search is performed to provide nutritional guidance. Fuzzy logic is also used in a nutritional guidance application in [5]. They define the optimality of a nutrient intake level as a fuzzy set (Fig. 1). The values of the points from "a" to "e" in Fig. 1 are based on the recommended intake levels. In [5], the Prerow value (PV) is used to measure how well the diet is balanced. In the calculation of PV, the nutrient with the lowest value has the strongest influence on the result. The gradient optimization method is used to find the shortest route to maximize PV in the discrete multidimensional search space.

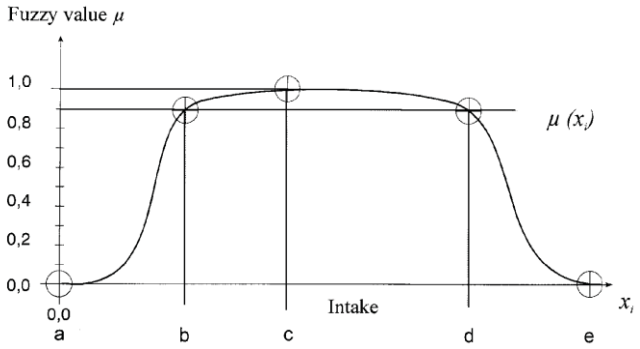


Figure 1: Membership function of optimal nutrient intake level [5].

In [6], the optimization results of two approaches to the nutrition calculation are compared. The first approach considers the use of non-fuzzy numbers and the second approach the use of fuzzy numbers in calculations and optimization. They concluded that optimization with plain numbers always yields results which meet the selected recommendations. The downside of this method is that it recommends drastic changes to eating habits. According to [6], the optimization with fuzzy numbers is not Pareto optimal. The results from fuzzy optimization are sometimes unbalanced. While the most of the nutrient intake levels are on an optimal level, some intake levels are either insufficient or excessive, and some are even less in balance

ISBN: 978-989-95079-6-8

than initially. However, the instructions given are closer to the original eating habits and are therefore easier to follow.

Dietary guidance applications usually present their results in the form of large tables where the suitable amounts of each nutrient are listed. However, this information is not very informative if you want to know what foods you should cut back on and what should you eat more. Therefore, there is a need for software which interprets the nutrient level results into the foodstuff level guidance.

The aim of this study is to develop a fuzzy model for the Nutri-Flow dietary analysis application. The Nutri-Flow application provides nutrition guidance in a user-friendly way: the balanced diet does not alter the original diet too much and the recommended adjustments to the diet are given on a foodstuff level in text format. This paper presents the fuzzy model developed for the Nutri-Flow application.

2 Nutrition recommendations and guidance

The prerequisite for good health is a well balanced diet. In a well balanced diet energy intake is in proportion to energy consumption and the nutrition intake is balanced. This often implies increasing the amount of dietary fibre in the carbohydrate intake and reducing the intake of purified sugars. Dietary guidance is based on the calculation of the user's nutrient intake levels and the nutrition recommendations.

2.1 Nutrition recommendations

The first nutrition recommendations were given to prevent deficiency disorders. The basis of traditional nutrition recommendations is physiological data on nutrient intake requirements. The Finnish Nutrition Recommendations are based on the Nordic recommendations which in turn are based on scientific data. The main objective of the national recommendations is to get Finnish people to balance their diets and to improve their health. The recommendations are defined for healthy and moderately physically active people. Therefore, they should only be used as guidelines for individuals and with great precaution. Nutrition recommendations represent the average daily values of long term intake. [3] Public dietary guidelines are criticized in [7]. They conclude that an adequate scientific background must first be established before public dietary guidelines are declared.

The nutrition recommendations are given as reference values which are defined separately for each nutrient. If there is evidence that the nutrient has an effect on health, a reference value is defined for it. The reference values are determined using estimates and evaluations based on current knowledge of nutritional needs and health. Thus, there will always be a degree of uncertainty and imprecision in the values. The reference values should not be considered as fixed points. Reference values distinguish between the lower intake level (LI), the average requirement (AR), the recommended intake (RI), and the upper intake level (UL). [3] The definitions of the reference values are:

- LI – the minimum requirement of a nutrient without the deficiency symptoms,
- AR – the nutrient intake level that is adequate for maintaining good health for healthy individuals,
- RI – the nutrient intake level that is adequate for maintaining the good nutritional status among 98 % of healthy individuals. RI is calculated based on AR and
- UL – an estimate of the highest level of intake that carries no appreciable risk of adverse health effects [8].

Depending on the nutrient, some reference values may be missing. For example, only RI is defined for magnesium, because the effects of magnesium are not that well known. Whereas, all the reference values have been set for vitamin B_6, because there is reliable scientific data available on how vitamin B_6 affects health with different intake levels. The reference values for vitamin B_6 are presented in Table 1. [3]

Table 1: Reference values for vitamin B_6 [3].

Vitamin B_6 level	Women [mg/d]	Men [mg/d]
LI	0.8	1
AR	1.0	1.3
RI	1.2	1.6
UL	25	25

2.2 Dietary guidance

Nutrient intake is usually assessed using a food composition databases. Food composition information is needed when calculating the composition of menus and recipes. Primary sources of food composition data are government databases, databases provided by academic and other institutions, the food industry, and scientific literature [9]. Finnish food composition information is collected into the Fineli® Nutrition Database which is provided by the National Public Health Institute. The commercial version of the database provides information on 2000 foods and 68 nutrients. The composition data consists of averaged information on Finnish food, and thus the composition of individual foods can vary depending on the given data.

Food composition data and the expert knowledge of a nutritionist are used to balance the user's diet on the basis of a food diary and personal nutrition intake recommendations (reference values). The national recommendations are adjusted individually according to the user's health data to form personal recommendations for evaluation. Typically, a nutritionist is needed to compare the data and to balance a suitable diet. The task is very complex because the information on a nutrient level is not useful in its self and must be converted onto a foodstuff level. In other words, people should know what foods they should eat more and what less. A balanced diet should also be reasonable in comparison to the original diet. Furthermore, some restrictions, for example allergies, diseases, and special diets, must be taken in to account when balancing diets.

3 Fuzzy logic

A rule based fuzzy logic system (FLS) consists of three interconnected main blocks for mapping crisp inputs to crisp outputs. The main blocks are the fuzzifier, the inference machine with the rulebase, and the defuzzifier [10]. The

ISBN: 978-989-95079-6-8

design parameters for a FLS are scaling factors, fuzzification and defuzzification methods, rulebase, and membership function construction, and representation.

Fuzzification is needed to convert the crisp input data into fuzzy sets for the inference machine. The fuzzification is carried out by using membership functions with defined shapes and parameters. In some cases, the input value needs to be normalised before fuzzification.

The mapping from inputs to outputs is done in the fuzzy inference module. The module contains all the necessary information for forming the output of the system. The heart of the inference module is the rulebase. The definition data used to generate and collect the output of the system is stored in the inference machine. The selected implication method and the aggregation method are a part of the definition data. The rulebase is a collection of IF-THEN rules which represent the expert information used to define the behaviour of the system. The antecedent of the rule defines the state of the system and also determines if the rule is triggered. The consequent of the rule is the output of each rule guiding the system output towards the solution. The membership grade of the rule antecedent is utilized through implication to produce the output of each rule. The minimum and the product are typically used as implication methods in Mamdani-type fuzzy systems. In Mamdani-type fuzzy systems, the output of each rule is fuzzy and aggregation is required to obtain the fuzzy output of the system. The fuzzy output needs to be defuzzified to obtain the crisp output of the system. In Takagi-Sugeno and Singleton -type fuzzy systems, the rule outputs are linear functions or constants, respectively. The crisp output of the system in these cases is typically obtained by taking the weighted average of the rule outputs. [11]

The defuzzifier module is needed if the output of the inference machine is fuzzy. There are several defuzzification methods. The most commonly used method is the centre of gravity method (COG). Other typical methods are the first (FOM), the last (LOM), and the middle (MOM) of maxima, and the first (FOS), the last (FOL), and the middle (FOM) of support. The defuzzification method should be chosen carefully because it affects the crisp result considerably. [12]

4 Data acquisition

The Nutri-Flow application uses a commercial version of the Fineli® Nutrition Database. The database is used to convert the information in the food diaries to averaged daily nutrient intake levels. The Finnish recommended nutrient levels (the reference values) are used to define the membership functions for the system inputs. The reference values are updated by nutritionist, and the personal recommendations are revised according to the user's health data.

Expert data acquisition was one of the most challenging parts of this study. The first task was to find an efficient way to convert expert knowledge into the fuzzy rule IF-THEN language. Variable tabulation proved to be a very good tool for this. With tabulation, all the input set conditions with appropriate consequents can be examined separately. Variable tabulation is possible because the number of variables in most of the rules is two or three. If the number of variables connected to a single rule is greater than three, tabulation becomes difficult. Table 2 shows an example of tabulation for carbohydrate and vitamin-C. The "fruits and berries" group is the consequent variable. The notations -, 0, and + represent the fuzzy values "too little", "ideal", and "too much", respectively. In this case, the consequents are limited to "add" and "no action". Naturally, the action "reduce" is also used, however, it is not feasible for the output variable "fruits and berries" group.

Table 2: Variable tabulation for "fruits and berries" group.

	Carbohydrate		
Vitamin-C	-	0	+
-	add	no action	no action
0	no action	no action	no action
+	no action	no action	no action

5 Developed model

The developed fuzzy model is used in the Nutri-Flow application. The schematic diagram of the overall system is provided in Fig. 2. The calculations start by analysing the user's nutrient intake levels based on the food diary and the Fineli® database. The intake levels are then fed into the developed fuzzy system. The fuzzy system is divided into two hierarchical levels. The first level model uses the nutrient intake levels as inputs and produces output on the level of main food groups such as "fruits and berries", "vegetables" and "dairy products". These outputs are then fed into the 2nd level model together with the nutrient intake levels. The 2nd level model produces outputs on a food subgroup level. Fig. 3 illustrates the division of foods into main groups and subgroups.

The fuzzy system requires user information (personal recommendations) to define the membership functions for the input variables. The output variables of the complete fuzzy system are continuous. The behaviour of the fuzzy system is not regulated by user specific information, such as the original diet, allergies, and diseases. The user related specific information is used in the optimization process. As depicted in Fig. 2, the focus of the future development of the Nutri-Flow application is to find an optimization algorithm which takes the outputs of the complete fuzzy system and the user specific information into account. Also, the nutrient density of the recommended foods should be considered when evaluating the final results of the system. Furthermore, the optimization algorithm would be used to search for solutions within a discrete space. The Nutri-Flow application uses the jFuzzyLogic java package for fuzzy inference. This Chapter presents the properties of the fuzzy system in more detail.

5.1 Membership functions

The values of the system inputs are the average daily intake levels of nutrients. The values are crisp numbers and they need to be fuzzified for the fuzzy inference machine. To map the crisp input values into the fuzzy domain, three linguistic variables are used: "too little", "ideal", and "too much" (Fig. 4). The fuzzy output variables are "reduce", "no action", and "add". The selected forms of the MFs are trapezoidal or triangular. The points "A", "B", and "C"

ISBN: 978-989-95079-6-8

correspond to LI, RI, and UL, respectively. If the points "A", "B", and "C" depicted in Fig. 4 are defined, three MFs are used to fuzzify the selected input variables. For example, in case of alcohol, there is no LI value and point “A” is not defined, thus the "too little" MF is not used. It should be noticed that the values "A", "B", and "C" are influenced by the user’s gender, age, mass, height, body mass index, and the level of physical activity. Some other factors, such as diseases and pregnancy, also affect the reference values. The output membership functions are produced similarly but the points from "A" to "C" are kept fixed.

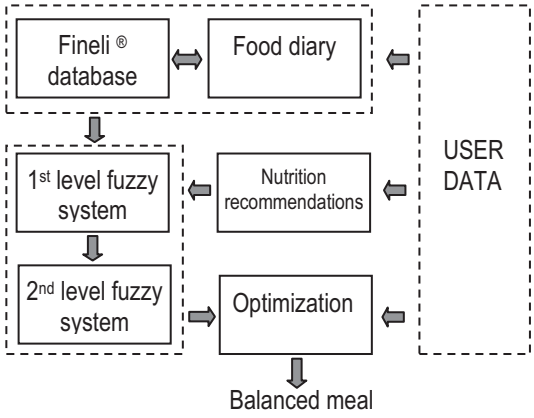


Figure 2: Schematic diagram of overall system.

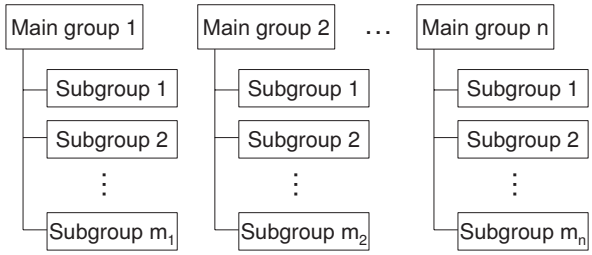


Figure 3: Division of foods to main groups and subgroups.

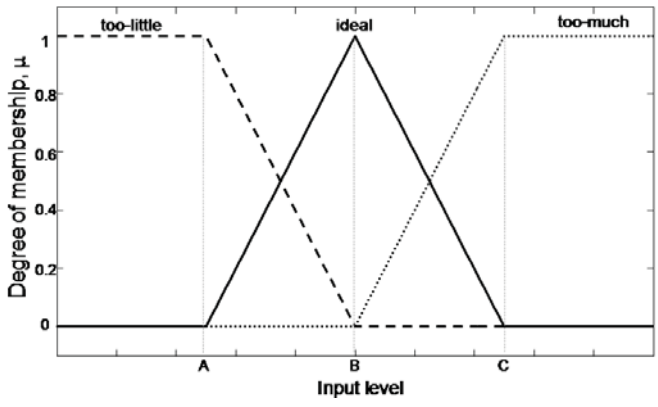


Figure 4: Input membership functions.

5.2 Rulebase and fuzzy inference

The most important and most time consuming phase of this study was building the rulebase. It is the heart of the fuzzy system, and therefore it must work correctly. Every input set condition must trigger at least one rule for each of the output variables. In this study, the rulebase was built on two levels. The first level of the fuzzy system operates on the main food group level and the second one on the subgroup level. The hierarchical structure is used because of the complexity of the system. The purpose of the first level is to roughly evaluate the nutritional status of the user and to produce additional, informative variables for the second level model, which in turn contains more detailed information but uses the additional variables to handle the complexity of the overall system.

As mentioned earlier, expert knowledge is obtained through variable tabulation. As an example, the tabulation of carbohydrate and vitamin-C was presented in Table 2. The consequent variable in this example was the "fruits and berries" -group. The first level rules corresponding to Table 2 are

Rule 1: IF Vitamin-C is "too little" AND Carbohydrate is "too little" THEN fruits and berries group is "add"
Rule 2: IF Vitamin-C is NOT "too little" OR Carbohydrate is NOT "too little" THEN fruits and berries group is "no action"

The output of the first level model directly influences the structure of the second level rules. The second level rules corresponding to the "fruits and berries" group have a general structure which can be written as

IF fruits and berries group is "Add" AND additional propositions THEN ...

If the consequent of the first level model is "no action", then the consequent of the corresponding food subgroups determined on the second level is also "no action".

5.3 Design parameters

In this study, the variables are not scaled and thus no scaling factors are needed. The fuzzy inference uses the minimum as an implication method and the maximum as an aggregation method. The used T-norm is a minimum. The consequents cannot be put in the form of an equation or defined by a single number. Therefore, a Mamdani-type inference is used with COG as the defuzzification method.

6 Results and discussion

All the calculations in this study were carried out in Matlab®. The data for membership function generation was extracted from the Nutri-Flow application. The data contains personal nutrient intake recommendations and the nutrient intake levels of the selected test cases. Also, a set of test cases with varying dietary habits were selected from the software for testing purposes.

6.1 Model validation with expert knowledge

The validation of the developed model was challenging because no "right answers" exist, and thus no meaningful error describing values (such as the sum of squared error) could be calculated. Therefore, only qualitative validation was carried out.

The model outputs are calculated using the developed model with the Nutri-Flow application. The output of the application is in text form and contains information on the

ISBN: 978-989-95079-6-8

diet and instructions on how the diet should be improved. The output is obtained in text form even though the optimization algorithm mentioned earlier has not yet been implemented. The validity of the output of each test case was evaluated by a nutritionist with promising results.

6.2 Uniformly distributed sensitivity test

The input-output behaviour of the model was tested with an uniformly distributed sensitivity test. The system sensitivity of each input variable was examined as follows: the supplies of all nutrients except the one under examination are fixed on the ideal intake level and the test variable gets random values in the range of [0, UL]. The number of test rounds conducted on one variable was 500. The test range is set so that the variable gets values from all the fuzzy sets. The quadratic error of the output is recorded as a sum of the output variables. The test results show how large the independent impact of an input variable is on the output of the whole system.

The value resulting from the test is the sum of the quadratic error which is the Euclidean distance from the system's ideal state. The test result value is zero for a variable with no individual impact on the output variable. The larger the value of a variable the greater its individual impact on the output variable.

A variable affects the output if it is the only variable in the antecedent part or the OR-connective is used. Also, if the tested variable is connected with an AND-connective to other variables and the other variables have a NOT-term in the proposition, the tested variable affects the output.

After the test, the results were given to a nutritionist for analysis. The analysis showed that the input variables only affected the desired output variables with the desired magnitude. An example of the results from the sensitivity analysis is shown in Table 4.

Table 4: Example results from sensitivity analysis.

	output 1	output 2	output 3	output 4
hard fat	0	0	1	0
soft fat	0	0	1	1
fibre	0.51	0	0.51	0

6.3 Normally distributed sensitivity test

Another type of sensitivity analysis was carried out by changing a group of input variables randomly at the same time. In this approach, the values of the input variables vary around the ideal value and they are taken from a normal distribution. The distributions of the output variables are obtained as a result and they are visually inspected. Fig. 5 illustrates a random test set of a variable. The mean and the standard deviation of the random changes are defined individually for all the input variables. The outputs of the random tests are analyzed by visually inspecting the distributions of the output variables. An example of the distribution of an output variable is given in Fig. 6. From the figure, it can be seen that the histogram does not quite follow a normal distribution. That is anticipated because no adding rules are associated with alcoholic beverages. Thus the histogram of the output variable is slightly skewed, and only values equal to or smaller than zero are obtained. Negative values of the output variables indicate that the consumption of a corresponding foodstuff should be reduced.

Some statistical values can also be calculated to evaluate the shape of the histograms. Mean, standard deviation, skewness, and kurtosis are used in this study. The mean and standard deviation values give an overview of the output data behaviour. The skewness is a measure of the asymmetry of the data around the mean value. If the data is spread out more to the left, the skewness value is negative and if the data is spread out more to the right, the skewness value is positive. For normal distribution the skewness value is zero. The distribution can also be evaluated with the kurtosis value which is a measure of how outlier-prone the distribution is. Table 5 presents the statistical values for some output variables. The evaluation of the histograms, both visually and on the basis of statistical values, shows that the developed model behaves correctly.

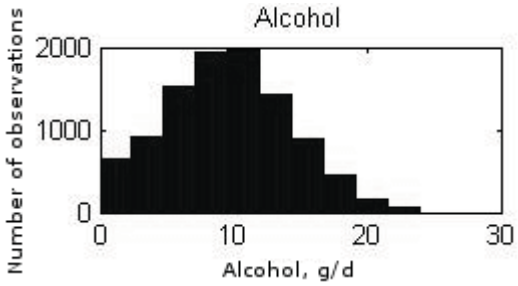


Figure 5: Test set for intake levels of alcohol.

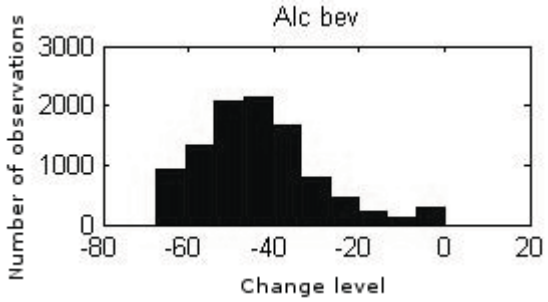


Figure 6: Histogram of output variable.

Table 5: Example statistical values of output variables.

	Mean	Std	Skewness	Kurtosis
Fruits and berries	11.9	19.1	1.1	2.5
Vegetables	-5.0	26.0	-0.2	2.3
Cereals and bakeries	-39.2	17.5	1.4	3.4
Alcoholic beverages	-43.5	14.1	0.9	4.1

6.4 Evaluation of the results

The results obtained from the performed validation are promising but more validation with more efficient methods is required. All of the validation procedures showed that the model performs in correctly, but no solid proof for this was obtained. The addition of the optimization algorithm will

ISBN: 978-989-95079-6-8

make validation easier because the numerical values (distance from the original and the recommended diet) can be calculated.

7 Conclusions

The importance of nutritional guidance is growing as nutritional problems, such as obesity and type-2 diabetes, are becoming more common. Nutrition guidance is carried out by mapping the nutrition intake levels from the user's food diary and comparing them to the recommended levels. Typically, balancing the diet requires expert knowledge on nutrition.

This paper presented the development of a fuzzy expert system for a nutrition guidance application. The expert knowledge acquisition was carried out using variable tabulation, thus creating a basis for the development of the rulebase for the fuzzy system. The recommended nutrition intake values were used to generate the membership functions for fuzzification. A Mamdani-type fuzzy system was developed with a hierarchical structure. The first level system only referred to the groups of similar foods while the second level model referred to food subgroups.

The validation of the fuzzy model was carried out in three phases: first two types of sensitivity analysis were performed, and then the output was analysed with expert knowledge. The results from all the validation schemes are promising.

References

[1] A. Waxman, WHO's global strategy on diet, physical activity and health: response to a worldwide epidemic of non-communicable diseases. Scandinavian Journal of Nutrition, 48(2):58, 2004.

[2] Peltonen et al. Kansallinen FINRISKI 2007-terveystutkimus, Tutkimuksen toteutus ja tulokset. Kansanterveyslaitoksen julkaisuja B34/2008, Helsinki, National Public Health Institute, 2008. (In Finnish)

[3] Nordic Council of Ministers, Nordic Nutrition Recommendations. Nordic Council of Ministers, Copenhagen, 4th, 435, 2004.

[4] J-C. Buisson and A. Garel, Balancing Meals Using Fuzzy Arithmetic and Heuristic Search Algorithms. IEEE Transactions on Fuzzy Systems, 11(1), Feb 2003.

[5] Wirsam et al. Fuzzy sets and fuzzy decision making in nutrition. European Journal of Clinical nutrition, 51(5): 286-296, 1997.

[6] Gedrich et al. How optimal are computer-calculated optimal diets? European Journal of Clinical Nutrition, 53: 309-318, 1999.

[7] Marantz et al. A Call For Higher Standards of Evidence for Dietary Guidelines. American Journal of Preventive Medicine, 34(3): 234-240, 2008.

[8] Institute of Medicine, Dietary Reference Intakes: applications in dietary planning. Washington D.C.: National Academies Press, 2003.

[9] Pennington et al. Food Composition Data: the foundation of dietic practice and research. Journal of the American Dietic Association, 107(12): 2105-2113, 2007

[10] J.M. Mendel, Fuzzy logic systems for engineering: a tutorial. Proceedings of the IEEE, 83(3): 345-377, 1995.

[11] Driankov et al. An Introduction to Fuzzy Control. Springer, Berlin, 2.rev, 316, 1996.

[12] X. Liu, Parameterized defuzzification with maximum entropy weighting function- Another view of the weighting function expectation method. Mathematical and Computer Modelling, 45(1-2): 177-188, 2007.

ISBN: 978-989-95079-6-8

Possible and Necessary h-indices

Marek Gągolewski[1,2] Przemysław Grzegorzewski[1,2]

1. Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
2. Faculty of Mathematics and Information Science, Warsaw University of Technology
pl. Politechniki 1, 00-661 Warsaw, Poland
Email: {gagolews,pgrzeg}@ibspan.waw.pl

Abstract— *The problem of measuring scientific impact is considered. A class of so-called p-sphere indices, which generalize the well known Hirsch index, is used to construct a possibility measure of scientific impact. This measure might be treated as a starting point for prediction of future index values or for dealing with right-censored bibliometric data.*

Keywords— Hirsch's h-index, p-sphere indices, scientific impact, possibility theory, scientometrics.

1 Introduction

Fair and objective assessment methods of individual scientists had become the focus of scientometricians' attention since the very beginning of their discipline. A quantitative expression, i.e. measurement, of some publication-citation process characteristics is assumed to be a predictor of broadly conceived scientific competence.

Among the most popular scientific impact indicators is the h-index, proposed by J. Hirsch in 2005 [1]. It has been defined as follows. An author who had published n papers has the Hirsch index equal to H, if each of his H publications were cited at least H times, and each of the other $n - H$ items were cited no more than H times. This simple indicator quickly received much attention in the academic community [2, 3] and started to be a subject of intensive research. It was noted (see for example [4, 5]) that contrary to earlier approaches, i.e. publication count, citation count etc., the measure both concerns productivity and impact of an individual.

Many modifications of the h-index were later proposed, e.g. Egghe's g-index [6], Kosmulski's $h(2)$-index [7], Jin's R-index [8] or Schreiber's h_m-index [9]. It was also a matter of more formal studies (e.g. [10, 11, 12, 13, 14, 15]). It is worth noting that the h-index can be expressed as the Sugeno integral of some function with respect to a fuzzy counting measure [16]. For more details we refer the reader to the extensive scientometric literature.

2 Difficulties related to the h-index

From now on $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ denotes the set of all natural numbers and zero, while $\mathbb{R}_0^+ = \mathbb{R}^+ \cup \{0\}$ stands for the set of all nonnegative real numbers. For any sets X, Y and Z ($Z \subset X$) and $f : X \to Y$ by $f|_Z$ we mean a mapping satisfying $f|_Z(x) = \mathbf{1}_{x \in Z} f(x)$, where $\mathbf{1}$ is an indicator function.

Let us assume that an individual has published exactly $n \in \mathbb{N}$ papers. The list of papers generates an ordered *citation sequence* $\mathbf{C} = (c_1, c_2, \ldots, c_n)$ such that $c_i \geq c_j$ for $1 \leq i < j \leq n$, where $c_i \in \mathbb{N}_0$ is the number of unique citations received by the i-th article.

The problem with the h-index is twofold. Firstly, it assumes perfect knowledge of the author's citation sequence. In practice we gather bibliometric data from large online academic services, such as Thomson *Web of Science*, Elsevier *Scopus* or Google *Scholar*. The coverage of all digital libraries is limited, so in most cases we are dealing with right-censored data.

On the other hand, citing is a dynamic process. If you were applying for academic tenure and were asked to determine your h-value, would you be sure that the index is not about to increase in a while? That is because the Hirsch coefficient totally ignores the number of citations received by publications represented by $c_1, c_2, \ldots, c_H$ (we only know their citation counts are $\geq H$) and how close to H are $c_{H+1}, c_{H+2}, \ldots, c_n$.

As an illustration of the raised issues, consider the following citation sequences: $\mathbf{C}_1 = (4, 4, 4, 4, 0, 0, 0)$ and $\mathbf{C}_2 = (10, 9, 8, 7, 4, 4, 3, 1)$. Both have the h-index of 4, but in the latter case there is not much needed for $\mathbf{C}_2$'s h to increase even to the value of 6. Such property of the citation sequence could be called *saturation* or even *instability*.

In the next section we recall necessary information on a class of so-called "geometric" scientific impact indices [17]. In Section 4 we suggest how to evaluate the h-index stability by means of possibility theory. In our model a given, but usually uncertain, citation sequence will be treated as a source of information to estimate possible and necessary index values. Section 5 illustrates the construction of the suggested possibility measure for individual's h-index.

3 p-sphere indices

Let us recall some definitions from the paper [17]. The following function, based on the citation sequence, will be useful.

Definition 1. A *citation function* based on a citation sequence $\mathbf{C}$ is a mapping $\pi_{\mathbf{C}} : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ given by

$$\pi_{\mathbf{C}}(x) = \begin{cases} c_i & \text{if} \quad x \in [i-1, i), \quad i = 1, 2, \ldots, n, \\ 0 & \text{if} \quad x \geq n. \end{cases} \tag{1}$$

An exemplary citation function for sequence $\mathbf{C} = (5, 4, 3, 3, 3, 1)$ is depicted in Fig. 1.

Definition 2. Consider a function $f : \mathbb{R}_0^+ \to \mathbb{R}_0^+$. We say that the citation function π *dominates* the function f (denoted $\pi \succeq f$) if $\pi(x) \geq f(x)$ for every $x \in \mathbb{R}_0^+$.

ISBN: 978-989-95079-6-8

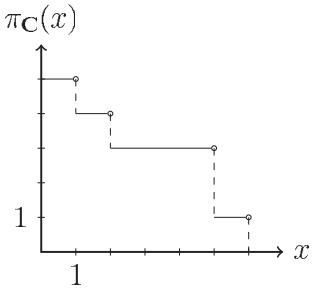


Figure 1: Citation function for $\mathbf{C} = (5, 4, 3, 3, 3, 1)$.

A set of all functions $f : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ dominated by the citation function π will be denoted by L_π.

Among the two new classes of scientific impact indices, the authors defined the p-sphere index.

Definition 3. Given an arbitrary real number $1 \leq p < \infty$ and any real number $r \geq 0$ let $s_{p,r} : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ denote a function

$$s_{p,r}(x) = \begin{cases} (r^p - x^p)^{\frac{1}{p}} & \text{for} \quad x \in [0, r)\,, \\ 0 & \text{for} \quad x \geq r. \end{cases} \tag{2}$$

Moreover, for $p = \infty$ we have

$$s_{\infty,r}(x) = \begin{cases} r & \text{for} \quad x \in [0, r)\,, \\ 0 & \text{for} \quad x \geq r. \end{cases} \tag{3}$$

Intuitively, for $x \in [0, r)$, the graph of $s_{p,r}(x)$ determines a part of an L^p-sphere (i.e. the boundary of an L^p-ball on a plane) of radius r (see Fig. 2) and centered at $(0, 0)$. Therefore $s_{p,r}$ is further on called the *p-sphere function* of *radius* r.

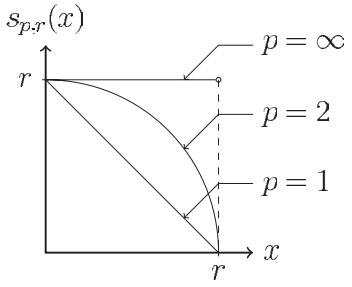


Figure 2: Exemplary p-spheres, for $p = 1, 2, \infty$.

Definition 4. The *p-sphere index* (or, originally, the *maximal p-radius*) of a citation function π is the greatest number $r \geq 0$, for which π still dominates $s_{p,r}$, i.e.

$$r_\pi(p) := \max\{r : \ \pi \succeq s_{p,r}\}\,. \tag{4}$$

The p-sphere index was primarily denoted $r_p(\pi)$. In this article we mainly consider the maximal radius to be a function of p, hence the change. For abbreviation, the p-sphere function of maximal p-radius $s_{p,r_\pi(p)}$ will be denoted by $S_{p,\pi}$ and called the *maximal p-sphere function* for π.

Here are some properties of the p-sphere index.

Lemma 5. *For any given $p \geq 1$ and a citation sequence $\mathbf{C} = (c_1, c_2, \ldots, c_n)$, let $\pi = \pi_{\mathbf{C}}$. Then the following properties hold:*

(i) $r_\pi(p) \leq n$.

(ii) $r_\pi(p) \leq c_1$.

(iii) If $p = \infty$, then $r_\pi(\infty) = H$, where H is the h-index of an individual (see [1]).

(iv) If $p = 1$, then $r_\pi(1) = W$, where W is the individual's w-index, as defined by Woeginger in [13].

(v) For any $q > p$, $r_\pi(q) \leq r_\pi(p) \leq 2r_\pi(q)$.

(ii) $r_\pi(p)$ is nonincreasing with respect to p.

Their proofs were given in [17].

4 Possible and necessary h-index

The theory of fuzzy measures and evidence has been well established. Therefore only definitions and properties that are necessary are recalled (for more details and further references the reader is referred, e.g., to [18]).

Further on we consider measures with respect to $\mathbb{R}_0^+$ and a family of all its subsets $\mathcal{P}(\mathbb{R}_0^+)$.

Definition 6. A function $\mu : \mathcal{P}(\mathbb{R}_0^+) \to [0, 1]$ is a *fuzzy measure* if it satisfies the following requirements:

M1. $\mu(\emptyset) = 0$ and $\mu(\mathbb{R}_0^+) = 1$ (boundary conditions),

M2. for all $A, B \in \mathcal{P}(\mathbb{R}_0^+)$, if $A \subseteq B$, then $\mu(A) \leq \mu(B)$ (monotonicity).

Definition 7. Let Pos denote a fuzzy measure. Then Pos is called a *possibility measure* iff for any family $\{A_k \in \mathcal{P}(\mathbb{R}_0^+) : k \in K\}$ and an arbitrary index set K,

$$\mathrm{Pos}\left(\bigcup_{k\in K} A_k\right) = \sup_{k\in K} \mathrm{Pos}(A_k). \tag{5}$$

For each possibility measure Pos we may associate another fuzzy measure Nec, called *necessity measure*, defined by

$$\mathrm{Nec}(A) := 1 - \mathrm{Pos}(\overline{A}), \tag{6}$$

where $A \in \mathcal{P}(\mathbb{R}_0^+)$.

It may be shown, that for every $A \in \mathcal{P}(\mathbb{R}_0^+)$ and for any possibility measure Pos and the associated necessity measure Nec following relations hold:

(i) $\mathrm{Nec}(A) > 0 \Rightarrow \mathrm{Pos}(A) = 1$,

(ii) $\mathrm{Pos}(A) < 1 \Rightarrow \mathrm{Nec}(A) = 0$.

Proposition 8. *Every possibility measure* Pos *may be uniquely determined by a* possibility distribution function *(abbreviated Pos.D.F.) $R : \mathbb{R}_0^+ \to [0, 1]$ by the formula*

$$\mathrm{Pos}(A) = \sup_{x\in A} R(x), \tag{7}$$

where $A \in \mathcal{P}(\mathbb{R}_0^+)$.

ISBN: 978-989-95079-6-8

Now we are ready to present a class of possibility measures for the h-index values.

Suppose we are given a citation sequence $\mathbf{C} = (c_1, c_2, \ldots, c_n)$ of some individual X. Some of the elements of $\mathbf{C}$ are either right-censored data or are potentially due to increase in a "short time". Let $H = h(\mathbf{C}) = \max\{i \in \mathbb{N}_0 : c_i \geq i\}$ be equal to X's Hirsch index. Thus we have sure evidence that the true value of X's Hirsch-index is *at least* H.

We propose some minimal requirements for a possibility distribution to describe hypothetical values of the h-index.

Definition 9. A *Pos.D.F. for the Hirsch index* is a mapping $R_h : \mathbb{R}_0^+ \to [0,1]$ which satisfies the following axioms:

H1. $R_h(x) = 0$ for $x < H$ or $x \notin \mathbb{N}_0$,

H2. $R_h(H) = 1$ (normalization),

H3. For any $x, x' \in \mathbb{N}_0$, if $H \leq x < x'$ then $R_h(x) > R_h(x')$ or $R_h(x') = 0$.

These axioms seem quite natural. Let us look at some properties of the fuzzy measures defined by such Pos.D.F. The following proposition might be easily proved.

Proposition 10. *Let R_h be an arbitrary Pos.D.F. for the Hirsch index. Furthermore, let* Pos *be the possibility measure determined by R_h and* Nec *be the associated necessity measure. Then for any $[a,b] \in \mathcal{P}(\mathbb{R}_0^+)$ the following properties are satisfied.*

- *If $H \in [a,b]$ then* $\mathrm{Pos}([a,b]) = 1$.
- *If $H \notin [a,b]$ then* $\mathrm{Nec}([a,b]) = 0$.
- *If $H \in [a,b]$ then for any $b < b'$ we get* $\mathrm{Nec}([a,b]) \leq \mathrm{Nec}([a,b'])$.

The postulated measures may give clues for questions such as: What is the (broadly conceived) possibility that the true value of X's h-index is really "equal to $H+1$" or "greater than H". Note that publishing-citing is an extremely complicated process and without drastic simplifications and idealizations it cannot be modeled using stochastic methods. Thus, in general, it is not reasonable to consider the results of individual scientometric measurements by means of the *probability* theory. Therefore, our problem is how to construct appropriate possibilistic measures. In the next section we propose a Pos.D.F. for the Hirsch index defined by means of the $r(p)$-indices.

5 Example

Lemma 11. *Assume we are given a citation function $\mathbf{C} = (c_1, c_2, \ldots, c_n)$, $\pi = \pi_{\mathbf{C}}$ and $r_\pi(\infty) = H \in \mathbb{N}$. Then the properties below are satisfied:*

(i) $c_i \geq H$ *for every* $i \leq H$,

(ii) $c_i \leq H$ *for every* $i > H$,

(iii) $r_\pi(p) \leq 2^{\frac{1}{p}} H$ *for any* $1 \leq p < \infty$.

Proof. Only (iii) is proved here. Let $1 \leq p < \infty$. By Lemma 5, $H \leq r_\pi(p) \leq 2H$. Let $a = \max\{c_1, 2H\}$. Consider a citation sequence $\mathbf{C}' = (c_1', c_2', \ldots, c_n')$, such that $c_i' = a$ for $i \leq H$, and $c_i' = H$ for $H < i \leq n$. Clearly, $r_{\pi_{\mathbf{C}'}}(\infty) = H$, and $c_i \leq c_i'$ for any i. For every $1 \leq p < \infty$, $S_{p,\pi_{\mathbf{C}'}}$ goes through (H, H), so $r_{\pi_{\mathbf{C}'}}(p)$ is the largest possible for fixed H. Solving (2) for r gives $r_{\pi_{\mathbf{C}'}}(p) = 2^{\frac{1}{p}} H$. Hence $r_{\pi_{\mathbf{C}}}(p) \leq 2^{\frac{1}{p}} H$ as stated. □

Let us define the inverse function of r_π. As r_π is not necessarily an injection, we need a special formula. In the sequel we assume that

$$r_\pi^{-1}(\varrho) := \max\{p : r_\pi(p) = \varrho\}, \tag{8}$$

for $\varrho \in [r_\pi(\infty), r_\pi(1)]$, and undefined otherwise. This definition is sensible, as it may be easily shown that for any $1 \leq p' < p''$ and $x \in \mathbb{R}_0^+$, $r_\pi(p') = r_\pi(p'')$ implies $S_{p',\pi}(x) \leq S_{p'',\pi}(x)$, hence the knowledge of maximal p satisfying $r_\pi(p) = \varrho$ is the most informative.

Now, let $R_\pi : \mathbb{R}_0^+ \to [0,1]$ be a function given by:

$$R_\pi(x) = \begin{cases} 0 & \text{for } x < r_\pi(\infty), \\ 2 - 2^{1/r_\pi^{-1}(x)} & \text{for } x \in [r_\pi(\infty), r_\pi(1)], \\ 0 & \text{for } x > r_\pi(1). \end{cases} \tag{9}$$

It is easily seen that for $x \in [r_\pi(\infty), r_\pi(1)]$, $R_\pi(x)$ is a strictly decreasing, but not necessarily continuous function.

For any $1 \leq p < \infty$, if $r_\pi(p) = 2^{\frac{1}{p}} H$, then $r_\pi^{-1}(x) = [\log_2(\frac{1}{H}x)]^{-1}$ and

$$R_\pi(x) = \begin{cases} 2 - \frac{1}{H}x & \text{for } x \in [H, 2H], \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Lemma 12. *Let $\mathbf{C}$ and $\mathbf{C}'$ be citation sequences for which $\pi = \pi_{\mathbf{C}}$, $\pi' = \pi_{\mathbf{C}'}$, $r_\pi(\infty) = r_{\pi'}(\infty) = H$ and for any $1 \leq p < \infty$, $r_\pi(p) = 2^{\frac{1}{p}} H$.*

(i) $R_{\pi'}(x) \leq R_\pi(x)$ *for any x,*

(i) If $R_{\pi'}$ is given by (10)*, then $r_{\pi'}(p) = 2^{\frac{1}{p}} H$ for any $1 \leq p < \infty$.*

The proof is left to the reader.

It may be shown that $R_\pi|_{\mathbb{N}_0}$ is a Pos.D.F. for the Hirsch index for π. Thus the class of p-sphere indices may be used to construct a possibility measure having the properties of our interest. Obviously, one may find uncountably many transforms of r_π^{-1} to $[0,1]$, so that is just an example. Generating meaningful possibility measures by means of the p-sphere indices dependent on a type of the process affecting citation sequences is in scope of our future research.

Let us discuss the behavior of the proposed possible h-index on real-world data and see how it can be used to differentiate between individual's citation information. We consider the output of 3 Polish computer scientists, Prof. A having $n_A = 19$ publications, Prof. B with $n_B = 29$ publications and Prof. C with his $n_C = 18$ publications. Their citation sequences, according to *Scopus*[1], are given in Table 1. See Fig. 3 for the citation functions π_A, π_B, π_C and the maximal p-spheres for $p = 1, 2, \infty$. Each author's h-index equals $H = 7$, but they differ in values of other p-sphere indices, e.g. $r_{\pi_A}(1) = 12$, $r_{\pi_B}(1) = 14$ and $r_{\pi_C}(1) = 10$.

[1] The publication data were gathered on January 20, 2009 and are limited to the field "Computer Science". Authors' names have been intentionally masked.

ISBN: 978-989-95079-6-8

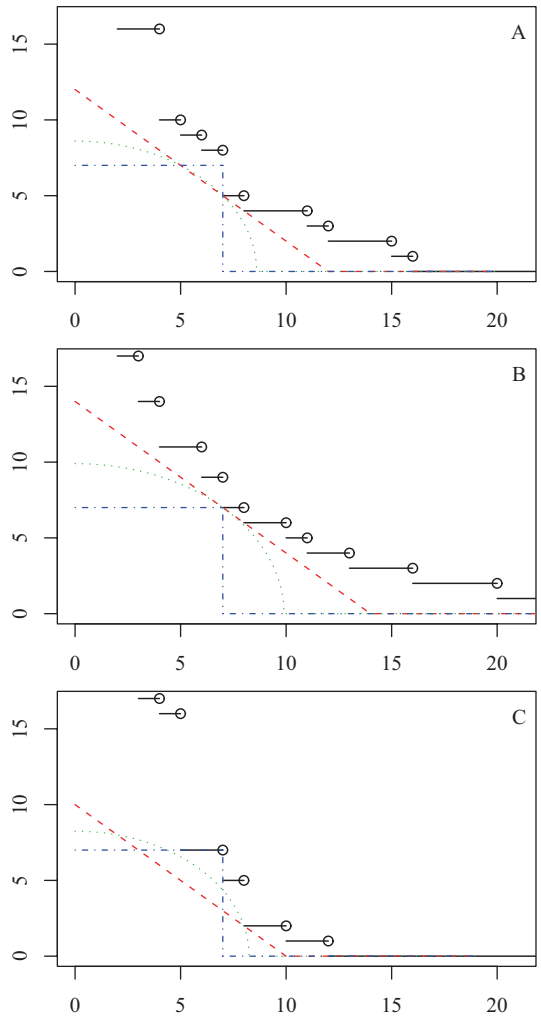


Figure 3: Citation functions of the 3 authors.

The author's R_π functions and resulting possibility distribution functions for the Hirsch-index are depicted in Fig. 4. Please note that R_{π_B} is of the form (10).

We see that Prof. B has the greatest possibility of increasing his h-index. On the contrary, some papers of Prof. C has not got a required number of citations so he can not have high expectations of a greater h-value. The proposed Poss.D.F. clearly differentiates between all the authors and can be used as a complement to the Hirsch index.

6 Conclusions

In the paper we discussed an important problem related to the Hirsch index: it assumes perfect knowledge of author's citation sequence and does not take into account a dynamic essence of the publication/citation process. Therefore we proposed a possibilistic approach to the indicator by setting several axioms for the possibility distribution function for the h-index.

In this model, given citation data are treated as an evidence for the minimal h-value and thus are just a starting point for speculation on its likely value in the case of perfect information.

We used a recently-proposed class of scientometric coefficients, the p-sphere indices, which is a generalization of Hirsch's index, to construct an exemplary possibility measure. A real-world example was presented for illustration.

As it was already mentioned, future work should definitely encompass the construction of different possibility measures, according to the type of a process affecting citation sequences.

Table 1: Citation sequences of the 3 authors.

$\mathbf{C}_A$	(103, 20, 16, 16, 10, 9, 8, 5, 4, 4, 4, 3, 2, 2, 2, 1, 0, 0, 0),
$\mathbf{C}_B$	(56, 30, 17, 14, 11, 11, 9, 7, 6, 6, 5, 4, 4, 3, 3, 3, 2, 2, 2, 2, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0),
$\mathbf{C}_C$	(39, 34, 23, 17, 16, 7, 7, 5, 2, 2, 1, 1, 0, 0, 0, 0, 0, 0).

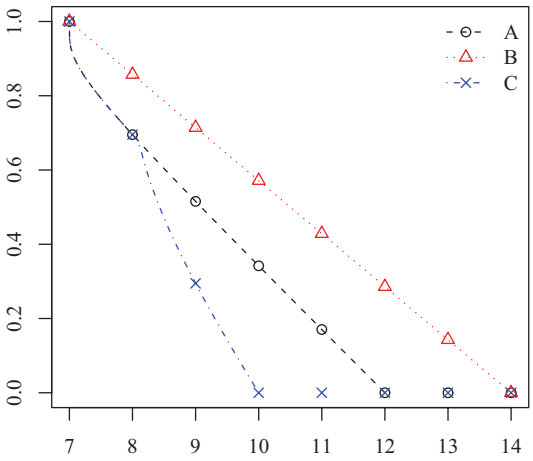


Figure 4: R_π and Pos.D.F. for the h-index of the 3 authors.

References

[1] Jorge E. Hirsch. An index to quantify individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46):16569–16572, 2005.

[2] Philip Ball. Index aims for fail ranking of scientists. *Nature*, 436:900, 2005.

[3] Lutz Bornmann and Hans-Dieter Daniel. The state of h index research. *EMBO Reports*, 10(1):2–5, 2009.

[4] Wolfgang Glänzel. On the h-index — A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67(2):315–321, 2006.

[5] Lutz Bornmann and Hans-Dieter Daniel. What do we know about the h index? *Journal of the American Society for Information Science and Technology*, 58(9):1381–1385, 2007.

[6] Leo Egghe. Theory and practise of the g-index. *Scientometrics*, 69(1):131–152, 2006.

[7] Marek Kosmulski. A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter*, 2(3):4–6, 2006.

[8] BiHui Jin, LiMing Liang, Ronald Rousseau, and Leo Egghe. The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6):855–863, 2007.

[9] Michael Schreiber. A modification of the h-index: The h_m-index accounts for multi-authored manuscripts. *Journal of Informetrics*, 2(3):211–216, 2008.

ISBN: 978-989-95079-6-8

[10] Quentin L. Burrell. On the h-index, the size of the Hirsch core and Jin's A-index. *Journal of Informetrics*, 1:170–177, 2007.

[11] Leo Egghe. Modelling successive h-indices. *Scientometrics*, 77(3):377–387, 2008.

[12] Wolfgang Glänzel. On some new bibliometric applications of statistics related to the h-index. *Scientometrics*, 77(1):187–196, 2008.

[13] Gerhard J. Woeginger. An axiomatic characterization of the Hirsch-index. *Mathematical Social Sciences*, 56(2):224–232, 2008.

[14] Ronald Rousseau. Woeginger's axiomatisation of the h-index and its relation to the g-index, the $h(2)$-index and the r^2-index. *Journal of Informetrics*, 2(4):335–340, 2008.

[15] Antonio Quesada. Monotonicity and the Hirsch index. *Journal of Informetrics*, 3(2):158–160, 2009.

[16] Vicenç Torra and Yasuo Narukawa. The h-index and the number of citations: two fuzzy integrals. *IEEE Transactions on Fuzzy Systems*, 16(3):795–797, 2008.

[17] Marek Gągolewski and Przemysław Grzegorzewski. A geometric approach to the construction of scientific impact indices. *Scientometrics*, 2009. In press. DOI:10.1007/s11192-008-2253-y.

[18] George J. Klir and Bo Yuan. *Fuzzy sets and fuzzy logic. Theory and applications*. Prentice Hall PTR, New Jersey, 1995.

ISBN: 978-989-95079-6-8

Similarity based fuzzy interpolation applied to CAGD*

Joan Jacas Amadeu Monreal Jordi Recasens

Sec. Matemàtiques i Informàtica. E.T.S. d'Arquitectura de Barcelona
Univ. Politècnica de Catalunya. Diagonal, 649 - 08028 Barcelona, Spain
Email: joan.jacas@upc.edu, amadeo.monreal@upc.edu, j.recasens@upc.edu

Abstract— *This paper presents an approach to a system based on fuzzy logic for the* **design** *of interpolative curves and surfaces in the context of Computer Aided Geometric Design (CAGD). Some problems arising from a previous model are studied by using a similarity approach.*

Keywords— Fuzzy interpolation, indistinguishability operator, fuzzy number, Computer Aided Geometric Design,approximate reasoning, t-norm, fuzzy control.

1 Introduction

The aim of computer aided geometric design (CAGD) is to produce a curve or a surface whose shape is controlled by a set of given points. In this sense, it can be understood as an interpolative procedure where the goal object is not a function but its graphical presentation. Therefore, in the design process it is essential to control the degree on smoothness of the solution.

On the other hand, the user of a CAGD system provides the data "approximately" because, after a first trial, he usually adjusts the result to his own taste or aesthetic point of view. So, it is natural to treat the problem under the setting of approximate reasoning considering the input-output data as fuzzy points of $\mathbb{R}$ or $\mathbb{R}^2$ represented by fuzzy numbers.

In order to build such a system, we focus our attention in the fuzzification and defuzzification procedures. To define and control the shapes on the input-output fuzzy numbers, we assume that there exists a similarity relation in $\mathbb{R}$ that granulates and fuzzifies its elements. With respect to the defuzzification process, we can choose the Takagi-Sugeno model that produces a crisp output or select some of the well known defuzzification methods like the center of gravity. The paper is organized in the following way. After this introduction, section 2 recalls the preliminaries related to indistinguishability operators and fuzzy classes. Sections 3 and 4 presents the general model and a model associated to Takagi-Sugeno technique that tries to overcome some of the problems of the model proposed in [17], such as the effective construction of the fuzzy numbers and the solution for the case of sparse data. In section 5 an example is fully developed and we close with conclusions and future works.

2 Preliminaries

The large number of successful applications of fuzzy sets theory (specially in fuzzy control) has produced numerous concepts based on empirical motivations. In order to give a sound justification to the use of these applied concepts, it is necessary to provide in the setting of a clear model, a well-founded interpretation for them. The aim of this section is to introduce an appropriate background that justifies the different definitions and types of fuzzy real numbers as they are interpreted and used in the applications.

As a starting point it is assumed that the concept of fuzzy real number arises from the fact that there exists a certain equality relation in $\mathbb{R}$, in the sense that numbers laying "close" are indistinguishable to the observer [15]. When the defined equality is a T-indistinguishability operator [7,21], triangular and trapezoidal fuzzy numbers among others, appear in a natural way. The model justifies the use of different types of fuzzy numbers depending on the scale defined in $\mathbb{R}$ and also develops the way of generating indistinguishability operators on $\mathbb{R}$ associated to one or several scales.

Fuzzy numbers arise naturally from the underlying indistinguishalility as the singletons associated to the relation [2,12,13,14]. Therefore, only the fuzzy relations whose singletons fulfill a general definition of fuzzy number are considered. Reciprocally, given a "suitable" fuzzy equality in $\mathbb{R}$, only the fuzzy numbers that are the singletons of this relation will be considered as such. As a natural consequence of this setting, we also study the families of fuzzy numbers that are invariant under translations.

Throughout this paper we use the standard definitions and properties of t-norms. In the sequel Π will represent the product t-norm and $\mathcal{L}$ the Łuckasiewicz t-norm. Some definitions on T-indistiguinshability operators and fuzzy numbers are recalled. For a more detailed exposition on these topics, readers are referred to [1,7,15,16,18].

Definition 2.1. Let T be a t-norm. A *T-indistinguishability operator* E on a set X is a reflexive and symmetric fuzzy relation on X such that

$T(E(x,y),E(y,z)) \leq E(x,z)$ (T-transitive property).

If $T = \text{Min}$, then E is called a *similarity*. For $T = \mathcal{L}$, E is termed a *likeness* and when $T = \Pi$, E is a *probabilistic relation*.

T-indistinguishability operators generalize within the fuzzy framework the concept of equivalence relation and are also called equality relations [2,7]. From a semantical point of view, $E(x,y)$ can be interpreted as the degree of similarity between x and y.

Definition 2.2. Given a T-indistinguishability operator E on a set X and an element x_0 of X, the *singleton* induced by x_0 is the fuzzy set f_{x_0} defined by $f_{x_0}(x) = E(x_0, x)$.

The singleton of an element x of X can be thought as its fuzzy equivalence class with respect to E [12] and its fuzzification

*Partially supported by the project TIN2006-14311

ISBN: 978-989-95079-6-8

taking E into account [8].

Definition 2.3. A *fuzzy real number* is a map $f_a : \mathbb{R} \to [0,1]$ such that there exists a number $a \in \mathbb{R}$ with $f_a(a) = 1$ where f_a is non-decreasing on $(-\infty, a)$ and non-increasing on $(a, +\infty)$.

Definition 2.4. A T-indistinguishability operator E on $\mathbb{R}$ is called *admissible* if and only if its singletons are fuzzy numbers.

The fuzzified of a real number via an admissible indistinguishability operator is therefore a fuzzy number.

The link between these operators and some families of fuzzy numbers will be investigated.

2.1 Admissible T-indistinguishability operators

A scale on $\mathbb{R}$ is defined as a monotonic map $f : \mathbb{R} \to \mathbb{R}$. The possibility to distinguish real numbers in a fuzzy environment is based on the use of a scale. Different scales give different degrees of accuracy and determine different fuzzy equalities on $\mathbb{R}$. Theorem 2.1.2 formalizes this idea.

Theorem 2.1.1.[14] Let m be a pseudometric on a set X and t a generator of an archimedean t-norm T. Then $E = t^{[-1]} \circ m$ is a T-indistinguishability operator on X.

N.B. Different generators of the same t-norm generate different T-indistinguishability operators.

Theorem 2.1.2.[14] Let $f : \mathbb{R} \to \mathbb{R}$ be a monotonic map and t a generator of an archimedean t-norm T

$$E_f(x,y) = t^{[-1]}(|f(x) - f(y)|)$$

is an admissible T-indistinguishability operator.

Two monotonic maps f, g generate the same T-indistinguishability operator if and only if there exists a constant k such that $f = \pm g + k$.

In real situations where different experts produce different scales on $\mathbb{R}$ it is necessary to combine their information. The following theorem gives a way to do it.

Theorem 2.1.5. If $\{E_i\}_{i \in I}$ is a family of admissible T-indistinguishability operators on $\mathbb{R}$, then

$$E = \operatorname*{Inf}_{i \in I} E_i$$

is also an admissible T-indistinguishability operator on $\mathbb{R}$.

Therefore, we can combine admissible likeness or probabilistic relations generated by monotonic functions in order to obtain another equality on $\mathbb{R}$ of the same type.

2.2 Fuzzy numbers invariant under translations

Special families of fuzzy numbers are those who are invariant under translations. Roughly speaking, families formed by fuzzy numbers of the same "shape". In the same sense, we can also consider T-indistinguishability operators, invariant under translations i.e.: the degree of similarity $E(x,y)$ between two numbers x, y depends only on their distance.

In this section, we study the link between the families of fuzzy numbers and T-indistinguishability operators invariant under translations.

Definition 2.2.1. A family $\{f_a\}_{a \in \mathbb{R}}$ of fuzzy numbers associated to an admissible T-indistinguishability operator is *invariant under translations* if and only if

$$f_a(x) = f_b(x - a + b).$$

A T-indistinguishability operator E is *invariant under translations* if and only if

$$E(x + a, y + a) = E(x, y).$$

Let us search for the conditions that such a family $\{f_a\}_{a \in \mathbb{R}}$ must fulfill in order to become the set of the singletons associated to a T- indistinguishability operator E, being T Archimedean.

Theorem 2.2.4.[7] Given an archimedean t-norm T with additive generator t, the families $\{f_a\}_{a \in \mathbb{R}}$ of fuzzy numbers invariant under translations corresponding to singletons of a T-indistinguishability operator on $\mathbb{R}$ can be obtained from f_0 (using definition 2.2.1) defined by

$$f_0 = t^{[-1]} \circ F$$

where F is an even and non decreasing in $\mathbb{R}^+$ subadditive function with $F(0) = 0$.

The associated T-indistinguishability operator E is defined by

$$E(x,y) = t^{[-1]} \circ F(y - x)$$

which is trivially invariant under translations.

In this section, we have introduced a general setting in order to generate fuzzy numbers associated to generalized equalities. The model allows the effective construction of families of fuzzy numbers suitable for applications since the underlying equalities can be deduced from scales proposed by experts and specially adapted to a concrete implementation.

3 Interpolation in CAGD

In CAGD, usually the user provides a set of points and the program produces a curve or surface whose shape is controlled by these points. From a classical point of view an interpolative curve means a curve passing through a given set of points. In our proposal we treat this problem in a more relaxed way in the sense that we obtain the values of the curve from a set of data points but we do not impose the condition that the resulting curve should interpolate in the classical sense. Since it is a design process, it is essential to control the degree of smoothness of the solution.

In this paper, we propose the following problems, paying special attention to the case of plane curves:

Problem 1a) *(Non parametric or functional curves).* Given $n+1$ points of control $\overrightarrow{P_i} = (a_i, b_i) \in \mathbb{R}^2$, $i = 0, 1, \ldots, n$, with $a_0 < a_1 < ... < a_n$, we want to construct the graphic of a function $f : [a_0, a_n] \to \mathbb{R}$ such that $\{a_i\}$ are values of the independent variable and $\{b_i\}$ are values or approximations of $\{f(a_i)\}$.

Problem 1b) *(Parametric curves).* Given $n+1$ points of control $\overrightarrow{P_i} = (a_i, b_i) \in \mathbb{R}^2$, or $\overrightarrow{P_i} = (a_i, b_i, c_i) \in \mathbb{R}^3$, $i = 0, 1, \ldots, n$ we want to construct $n+1$ values $t_0 < t_1 < \ldots <$

ISBN: 978-989-95079-6-8

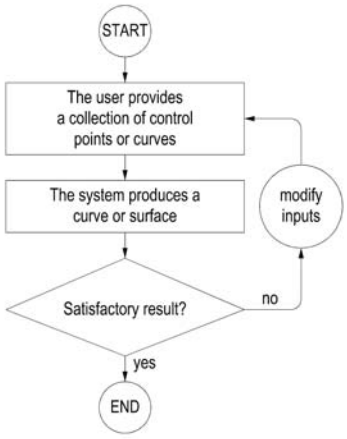


Figure 3.1

t_n of a parameter and a parametric curve $\overrightarrow{f} : [t_0, t_n] \to \mathbb{R}^2$ or $\mathbb{R}^3$ in such a way that the points $\{\overrightarrow{P}_i\}$ are the values or approximations of $\{\overrightarrow{f}(t_i)\}$.

Problem 2a) *(Non parametric or functional surfaces)*. Given $n+1$ arbitrarily distributed points of control $\overrightarrow{P_i} = (a_i, b_i, c_i) \in \mathbb{R}^3$, $i = 0, 1, \ldots, n$, we want to construct a surface passing through or controlled by these points as the graphic of a map $f : A \subset \mathbb{R}^2 \to \mathbb{R}$ (for example, $A = [a_{\min}, a_{\max}] \times [b_{\min}, b_{\max}]$) such that $\{(a_i, b_i)\}$ are the values of the independent variables and $\{c_i\}$ values or approximations of $\{f(a_i, b_i)\}$.

Problem 2b) *(Parametric surfaces)*. Given a net of $(n+1) \times (m+1)$ points $\overrightarrow{P}_{ij} = (a_{ij}, b_{ij}, c_{ij}) \in \mathbb{R}^3 \quad i = 0, \ldots, n$, $j = 0, \ldots, m$, we want to construct $n+1$ values $s_0 < s_1 < \ldots < s_n$ of a parameter s and $m+1$ values $t_0 < t_1 < \ldots < t_m$ of a parameter t and a parametric surface $\overrightarrow{f} : [s_0, s_n] \times [t_0, t_m] \to \mathbb{R}^3$ in such a way that the points $\{\overrightarrow{P}_{ij}\}$ are the values or approximations of $\{\overrightarrow{f}(s_i, t_j)\}$.

There are many standard techniques to treat the preceding problems. For a more complete information, also about methods to construct values of the parameter in parametric cases, the reader is refered to [9,10,11].

The common algorithm of all design methods that interacts between the procedure and the user is shown in the Fig. 3.1.

The algorithm shows the existing vagueness of the data points and its relation with the goal curves. Therefore, it is natural to focus the problem in the setting of approximate reasoning. The model proposed in section 4 uses fuzzy numbers and techniques of Takagi-Sugeno in fuzzy control to design curves and surfaces.

4 The Takagi-Sugeno Model

We propose a system based on fuzzy linguistic rules used in fuzzy control [5,18]. We have adopted the point of view of Takagi and Sugeno [23], in order to avoid the defuzzification process.

The main idea is to use rules of the following type:

If $\overrightarrow{x}$ is **close** to $\overrightarrow{x}_i$ in the domain, then $\overrightarrow{y}$ is **close** to $\overrightarrow{y}_i$ in the image. This idea is developed in the following **model**:

- Use $n+1$ rules

  R_i : If $\overrightarrow{x} \simeq \overrightarrow{x}_i$, then $\overrightarrow{y} \simeq \overrightarrow{y}_i$; $i = 0, 1, \ldots, n$.

- Model the degree of fulfillment of the antecedent of the i^{th} rule with a fuzzy set α_i such that $\alpha_i(\overrightarrow{x}_i) = 1$ and $\alpha_i(\overrightarrow{x})$ is a decreasing function of the distance to $\overrightarrow{x}_i$ [6].

- The consequence function of each rule, $f_i(\overrightarrow{x})$, is the constant $\overrightarrow{y}_i$. (Therefore, the system coincides with the **height defuzzification method**).

- The output of the system is obtained as a mean of the outputs of each rule weighted by the fuzzy sets α_i:

$$\overrightarrow{y}(\overrightarrow{x}) = \frac{\sum_{i=0}^{n} \alpha_i(\overrightarrow{x}) \cdot \overrightarrow{y}_i}{\sum_{i=0}^{n} \alpha_i(\overrightarrow{x})} = \sum_{i=0}^{n} F_i(\overrightarrow{x}) \cdot \overrightarrow{y}_i \quad (2.1.)$$

  where $F_i(\overrightarrow{x}_i) = \frac{\alpha_i(\overrightarrow{x})}{\sum_{j=1}^{n} \alpha_j(\overrightarrow{x})}$.

So, the result is a convex linear combination of the outputs of the rules where the coefficients depend on the data points and the fuzzy sets α_i. This is the general description of our proposal already outlined in [17] that contained the following remarks

4.1 Remarks

4.1.1. If the domain is in $\mathbb{R}$ (the case of a curve), then $\alpha_i = \mu_i$ are **fuzzy numbers**.

4.1.2. If the domain is in $\mathbb{R}^2$ (the case of a surface), then the closeness α_i of a point $\overrightarrow{x} = (x, y)$ to $\overrightarrow{x}_i = (x_i, y_i)$ can be modeled in the following way:
$\overrightarrow{x} = (x, y)$ is **close** to $\overrightarrow{x}_i = (x_i, y_i)$ if and only if x is **close** to x_i **and** y is **close** to y_i. That is, if μ_i and ν_i are fuzzy numbers modeling the closeness of x and y to x_i and y_i respectively, then the fuzzy set α_i can be defined as $\alpha_i(\overrightarrow{x}) = T(\mu_i(x), \nu_i(y))$, where T is a suitable t-norm that plays the semantic role of a conjunction. In order to maintain the degree of smoothness of the resulting surface inherited from the smoothness of the fuzzy numbers, it is convenient to choose a C^∞ t-norm. In our model, we use the **product**.

4.1.3. The fuzzy sets $\{\mu_i\}$ must cover the domain in the sense that for every point $\overrightarrow{x}$ in the domain, there must exists a fuzzy set μ_i with $\mu_i(\overrightarrow{x}) \neq 0$. This is a technical requirement in order to avoid dividing by zero in (2.1) and has the semantical meaning that the image of every point of the domain must be controlled by at least one of the given points $\{\overrightarrow{x}_i\}$.

4.1.4. Curves of type 1b) generated by this model cannot be, simultaneously, interpolative and smooth (i.e. tangent continuous). For interpolative curves $\overrightarrow{y}(t_i) = \overrightarrow{P}_i$, so $\mu_i(t_i) = 1$ and $\mu_j(t_i) = 0 \quad j \neq i$. Due to the monotonicity of the μ_i, $\mu_i(t) \equiv 0 \quad t \notin (t_{i-1}, t_{i+1})$. That means that for every interval $[t_{i-1}, t_i]$ there are only two fuzzy numbers different from zero. Consequently, in this interval the curve is a convex combination of the points $\overrightarrow{P}_{i-1}$ and $\overrightarrow{P}_i$ and we obtain a polygonal curve.

ISBN: 978-989-95079-6-8

4.1.5. It is worth noticing that, whereas in the classical methods used in CAGD only the selection of points is important, here we also have the possibility of choosing the type and size of the fuzzy number associated to each point as a consequence of the underlying indistinguishability, even within the same set of nodes. This aspect gives great versatility to our model.

From these remarks we can detect some problems in order to obtain a general procedure. The first one deals with the effective construction of the fuzzy numbers associated to the crisp data; that is, the fuzzification problem. If we consider that all the given points have the same weight in the construction of the curve or surface, it is natural to accept that the fuzzy points are invariant under translations and we can apply the ideas contained in the subsection 2.2. Otherwise, we can use the scales in order to graduate the importance of the points contained in different intervals of the domain. The fuzzification process of $\mathbb{R}$ is completely meaningful using the underlying equality relations. A second and important problem is mentioned in the remark 4.1.3., since our initial proposal is not applicable when dealing with sparse data. This problem has been extensively treated in the literature[8,10,19,20] from various points of view. Also, we can find many papers where different methods are compared [3,4,6,22]. In order to overcome this problem, we distinguish two different situations:

- a) The data set consists of equidistant x-values. To solve this case, we propose to build a family of parametrized fuzzy equalities in $\mathbb{R}$ whose associated fuzzy numbers cover all the domain for values of the parameter and for every $\vec{x_i}$ of the data set there exists a fuzzy number μ_i such that $\mu_i(\vec{x_j}) = 1$ if $i = j$ and $\mu_i(\vec{x_j}) = 0$ otherwise. Under this hypothesis we can obtain a curve that interpolates the points in the classical sense (see the example in section 5).

- b) The data has sparse non equidistant values. Our proposal consists in constructing a monotonic function f between a subset of the natural numbers and our numerable (finite) data set and using the fuzzy equality $E(x_i, x_j) = t^{-1} \circ F(f^{-1}(x_i) - f^{-1}(x_j))$ in order to define the fuzzy numbers associated to data.

With all this in mind, the solutions proposed to the problems of the introduction are the following:

4.2 Solutions.

Problem 1a) *(Non parametric curves).* In this case, $\overrightarrow{P}_i = (a_i, b_i)$, the domain is some interval in $\mathbb{R}$, $\overrightarrow{x}_i = a_i$, $\overrightarrow{y}_i = b_i$ and $\alpha_i = \mu_i$ are normalized fuzzy numbers. The solution curve is

$$y = f(x) = \frac{\sum_{i=0}^{n} \mu_i(x) \cdot b_i}{\sum_{i=0}^{n} \mu_i(x)} = \sum_{i=0}^{n} F_i(x) \cdot b_i. \qquad (2.2)$$

Problem 1b) *(Parametric curves).* In this case, $\overrightarrow{P}_i = (a_i, b_i) \in \mathbb{R}^2$ or $\overrightarrow{P}_i = (a_i, b_i, c_i) \in \mathbb{R}^3$. The domain is an interval in $\mathbb{R}$, the variable is a parameter t, $\overrightarrow{x}_i = t_i$, $\overrightarrow{y}_i = \overrightarrow{P}_i$ and $\alpha_i = \mu_i$ are normalized fuzzy numbers. The solution curve is

$$\overrightarrow{y} = \overrightarrow{f}(t) = \frac{\sum_{i=0}^{n} \mu_i(t) \cdot \overrightarrow{P}_i}{\sum_{i=0}^{n} \mu_i(t)} = \sum_{i=0}^{n} F_i(t) \cdot \overrightarrow{P}_i. \qquad (2.3)$$

Problem 2a). *(Non parametric surfaces).* In this case, $\overrightarrow{P}_i = (a_i, b_i, c_i) \in \mathbb{R}^3$, the domain is some region on $\mathbb{R}^2$, $\overrightarrow{x}_i = (a_i, b_i)$, $\overrightarrow{y}_i = c_i$ and $\alpha_i(x, y) = \mu_i(x) \cdot \nu_i(y)$ where μ_i, ν_i are normalized fuzzy numbers.

The solution surface is

$$z = f(x, y) = \frac{\sum_{i=0}^{n} \alpha_i(x, y) \cdot c_i}{\sum_{i=0}^{n} \alpha_i(x, y)} = \sum_{i=0}^{n} F_i(x, y) \cdot c_i. \qquad (2.4)$$

Problem 2b). *(Parametric surfaces).*

In this case, $\overrightarrow{P}_{ij} = (a_{ij}, b_{ij}, c_{ij})$, the domain is a rectangle in $\mathbb{R}^2$, the variables are the parameters $s \in [s_0, s_n]$, $t \in [t_0, t_m]$, $\overrightarrow{x}_{ij} = (s_i, t_j)$, $\overrightarrow{y}_{ij} = \overrightarrow{P}_{ij}$ and $\alpha_{ij}(s, t) = \mu_i(s) \cdot \nu_j(t)$ where μ_i, ν_j are normalized fuzzy numbers.

The solution surface is

$$\overrightarrow{P}(s, t) = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} \alpha_{ij}(s, t) \cdot \overrightarrow{P}_{ij}}{\sum_{i=0}^{n} \sum_{j=0}^{m} \alpha_{ij}(s, t)} = \sum_{i=0}^{n} \sum_{j=0}^{m} F_{ij}(s, t) \overrightarrow{P}_{ij} \qquad (2.5)$$

where

$$F_{ij}(s, t) = \frac{\alpha_{ij}(s, t)}{\sum_{i=0}^{n} \sum_{j=0}^{m} \alpha_{ij}(s, t)}$$

Note: Since in this case, the set of points determines a net, we have used a double index for convenience. If we want to use a single index k, we can, for instance, reassign indices in the following way: $\{i, j\} \to \{k\}$ with $k = i \cdot (m + 1) + j$.

5 An example

In this section we develop an example for fuzzy interpolation of curves and surfaces in its functional and parametric forms. In order to generate the underlying fuzzy equality in $\mathbb{R}$ invariant under translations, we need to select a subadditive, even and non decreasing function in $\mathbb{R}^+$. We have choosen $F = \text{Arctan}\,|x|$. This function fulfills all the conditions of theorem 2.2.1. On the other hand we have selected a parametrized family of strictly decreasing functions $f(x, k) : [0, 1] \to \mathbb{R}^+$ that, jointly with its inverses (Fig. 5.1 and Fig. 5.2), allows us to build a parametrized family of t-norms represented in Fig. 5.3 for $k = 5$.

Fig. 5.4 contains a representation of the associated T-indistinguishability that produces the fuzzification of $\mathbb{R}$ depicted in Fig. 5.5 for a discrete sub-family. By using this fuzzification and the procedure explained in section 4, we produce an interpolating curve for the case of equidistant values in functional form (Fig. 5.6).

Fig. 5.7 and Fig. 5.8 present two solutions of a curve constructed by means of its parametrized representation. As it is

ISBN: 978-989-95079-6-8

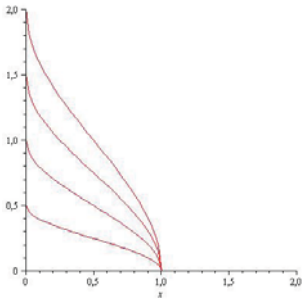


Figure 5.1

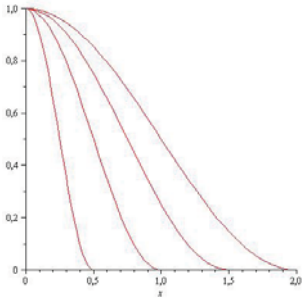


Figure 5.2

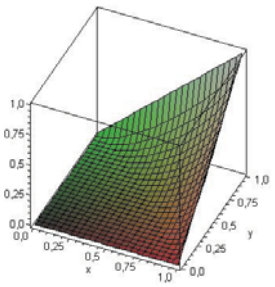


Figure 5.3

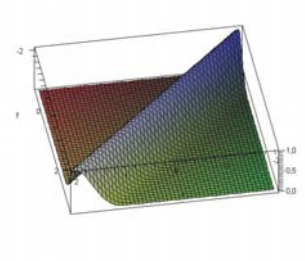


Figure 5.4

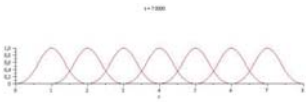


Figure 5.5

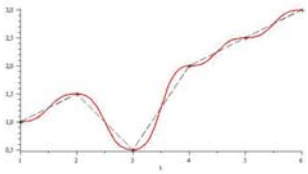


Figure 5.6

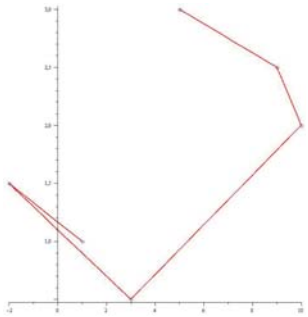


Figure 5.7

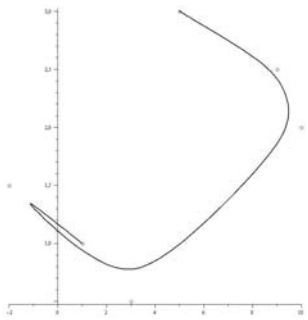


Figure 5.8

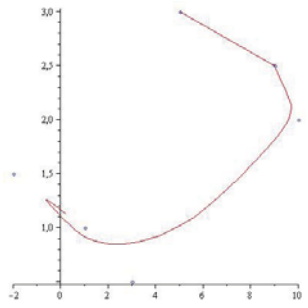


Figure 5.9

ISBN: 978-989-95079-6-8

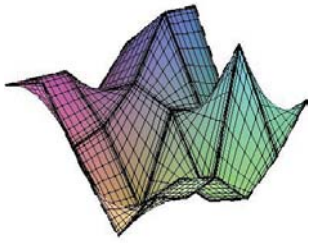


Figure 5.10

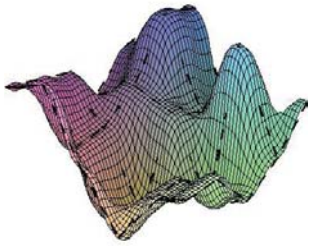


Figure 5.11

pointed out in the remarks, this representation can not be interpolative, in the classical sense, and, at the same time smooth.

Finally, in Fig. 5.9 a parametric curve is represented using a fuzzification of $\mathbb{R}$ defined by a scaled indistinguishability.

In a similar way we can construct surfaces.In this setting,the same incompatibility between interpolation and smoothnes, in the case on parametric representation, applies. Fig. 5.10 shows the case of interpolation for parametric representation. Fig. 5.11 shows the interpolative surface for the functional case.

References

[1] C. Alsina, On a family of connectives for fuzzy sets. *Fuzzy Sets and Systems* 16, (1985) 231-235.

[2] D. Boixader,J. Jacas, J. Recasens, Fuzzy equivalence relations: advanced material. In *Dubois and Prade Eds. Fundamentals of fuzzy sets*.(2000) Kluwer 261-290.

[3] B. Bouchon-Meunier et alt. A comparative view of interpolation methods between sparse fuzzy rules. *Proc. IFSA'01 World Congress*, Vancouver (2001) 2499-2504.

[4] B. Bouchon-Meunier, C. Marsala, M. Rifqi. Interpolative reasoning based on granularity. Proc. FUZZ-IEEE'2000 (2000) 483-487.

[5] J.L.Castro, M. Delgado, Fuzzy Systems with defuzzification are universal approximators, *IEEE Trans. Systems, Man and Cybernetics*, 26 (1996) 149-152.

[6] D. Dubois,H. Prade On Fuzzy interpolation, *Int. J. General Systems* 28 (1999) 103-114.

[7] M. Dermirci, J. Recasens, Fuzzy groups, fuzzy functions and Fuzzy equivalent relations, *Fuzzy Sets and Systems* 144 (2004) 441-450.

[8] F. Esteva et alt. Similarity-based fuzzy interpolation method. *Proc. IPMU 2004*. Perugia (2004) 1443-1449.

[9] G. Farin, *Curves and Surfaces for CAGD.* Academic Press, 1990.

[10] I.D.Faux, M.J.Pratt, *Computational Geometry for Design and Manufacture*, Ellis Horwood Ltd., 1985.

[11] D. Hearn, M.P.Baker, *Computer Graphics*, Prentice Hall, 1994.

[12] U. Höhle, Quotients with respect to similarity relations, *Fuzzy Sets and Systems* 17, (1988) 31-44.

[13] J. Jacas, J. Recasens, On some kind of probabilistic relations. *Proc IPMU Congress*, Palma de Mallorca, (1992) 365-368.

[14] J. Jacas, J. Recasens, Fuzzy Numbers and Equality Relations, *Proc. IEEE International Conference on Fuzzy Systems* San Francisco (1993), 1298-1301..

[15] J. Jacas, J. Recasens, Fixed Points and Generators of Fuzzy Relations, *J. Math. Anal. and Appl.* 186 (1994) 21-29.

[16] J. Jacas, J. Recasens, Fuzzy T-transitive relations: eigenvectors and generators, *Fuzzy Sets and Systems* 72 (1995) 147-154.

[17] J. Jacas, A. Monreal, J. Recasens, A model for CAGD using fuzzy logic, *Int. J. Approximate Reasoning* 72 (1997) 289-308.

[18] F. Klawoon, R. Kruse, Equality Relations as a Basis for Fuzzy Control. *Proc. IEEE International Conference on Fuzzy Systems*, (1992).

[19] L. T. Kòckzy, K. Hirota. Interpolative reasoning with insuficient evidence in sparce rulebases. *Information Sciences* 71(1993) 169-200.

[20] L. T. Kòckzy, K. Hirota. Approximate reasoning by linear rule interpolation and general approximation.*Int. J. Approximate Reasoning* 9(1993) 197-222.

[21] R. Kruse, J. Gebhardt, F. Klawonn, *Foundations of Fuzzy Systems.* John Wiley & Sons, Chichester, 1994.

[22] S. Mizik, P. Baranyi, P. Korondi, H and L. T. Kczy. A comparative of fuzzy interpolation techniques. 4^{th} *Eurofuzzy-Sic'99* (1999) Budapest 544-549

[23] H.Takagi, , M. Sugeno, Fuzzy identification of Systems and its applications to modelling and control, *IEEE Trans. on Systems, Man and Cybernetics*, 15, 116-132, (1985).

ISBN: 978-989-95079-6-8

KM-Fuzzy Approach Space

J. Martínez-Moreno[1] A. Roldán[2] C. Roldán[2]

1. Department of Mathematics. University of Jaén
Jaén, Spain
2. Department of Statistics and Operations Research. University of Jaén
Jaén, Spain
Email: jmmoreno,afroldan,iroldan@ujaen.es

Abstract— *Using the idea of approach space due to Robert Lowen [Approach spaces. The missing link in the topology-uniformity-metric triad], we define the notion of fuzzy approach spaces as a natural generalization of fuzzy metric spaces due to Kramosil and Michalek [Kybernetika , 11, 1975, 336–344] and prove some properties of fuzzy approach spaces.*

Keywords— Approach spaces, Fuzzy approach spaces, Fuzzy metric spaces.

1 Introduction

The notion of fuzzy sets was introduced by Zadeh [1]. Since then, several concepts of fuzzy metric spaces were considered in [2, 3, 4]. In the sequel, we shall adopt the usual terminology, notation and conventions of the theory of fuzzy metric spaces introduced by Kramosil and Michalek [5]. In this section we recall the definitions of a fuzzy metric space, approach space and some elementary properties.

Definition 1.1 *A binary operation* $*\colon [0,1]\times[0,1]\to[0,1]$ *is a continuous t-norm if* $([0,1],*)$ *is a topological monoid with unit 1 such that* $a*b\le c*d$ *whenever* $a\le c$ *and* $b\le d$ *for* $a,b,c,d\in[0,1]$.

Definition 1.2 *A* KM-fuzzy quasi-metric space *(briefly qFM-space) is a triple* $(X,M,*)$ *where* X *is an arbitrary set,* $*$ *is a continuous t-norm and* $M\colon X\times X\times[0,+\infty]\to[0,1]$ *is a (fuzzy) mapping such that for all* $x,y,z\in X$*:*

FM1. $M(x,y,0)=0$*;*

FM2. $x=y$ *if and only if* $M(x,y,t)=1$ *for all* $t>0$*;*

FM3. $M(x,z,t+s)\ge M(x,y,t)*M(y,z,s)$ *for all* $s,t\ge 0$*;*

FM4. $M(x,y,\cdot)\colon[0,+\infty)\to[0,1]$ *is left-continuous.*

If we additionally assume that M satisfies the symmetry condition:

$$M(x,y,t)=M(y,x,t)\ \text{ for all } t>0,$$

then $(X,M,*)$ *is called fuzzy metric space.*

Lemma 1.3 $M(x,y,\cdot)$ *is a nondecreasing function for all* $x,y\in X$.

This paper introduces the concept of a fuzzy approach space. Our definition is inspired in the concept of approach spaces introduced by Lowen [6].

Definition 1.4 *A pair* (X,δ) *where* $\delta: X\times\mathcal{P}(X)\to[0,+\infty]$ *is a function on a set X such that:*

B1. $\delta(x,\{x\})=0$ *for every* $x\in X$.

B2. $\delta(x,\emptyset)=\infty$ *for every* $x\in X$.

B3. $\delta(x,A\cup B)=\min(\delta(x,A),\delta(x,B))$ *for every* $x\in X$ *and* $A,B\in\mathcal{P}(X)$.

B4. $\delta(x,A)\le\delta(x,A^{(r)})+r$ *for every* $p\in X$, $A\in\mathcal{P}(X)$ *and* $r\in[0,\infty]$, *with* $A^{(r)}=\{\,y\in X\ /\ \delta(y,A)\le r\,\}$.

*is called an approach space (**A-space**).*

Lowen ([6], 1997) proved that at least seven different structures can be used to define an approach space, equivalently. Assuming condition B3, B4 can be rewritten as follows (see [7]):

B4′. $(A^{(r)})^{(s)}\subseteq A^{(r+s)}$ for every $A\subseteq X$ and $r,s\in[0,\infty]$.

2 KM-Fuzzy approach spaces

Following, the notions of approach space are generalized to the fuzzy setting and the concept of KM-fuzzy approach spaces is introduced.

Definition 2.1 *A* KM-fuzzy approach space *(briefly FA-space) is a pair* (X,F) *where* X *is an arbitrary set and* $F\colon X\times\mathcal{P}(X)\times[0,+\infty]\to[0,1]$ *is a (fuzzy) mapping satisfying the following conditions:*
For all $x\in X$, $A,B\subset X$ *and* $s,t\in[0,+\infty]$,

FA1. $F(x,A,0)=0$*;*

FA2. $F(x,\emptyset,t)=0$ *for all* $t<+\infty$*;*

FA3. $F(x,\{x\},t)=1$ *for all* $t>0$*;*

FA4. $F(x,A\cup B,t)\ge\max(F(x,A,t),F(x,B,t))$*;*

ISBN: 978-989-95079-6-8

FA5. $F(x,A,t+s) \geq F(x,A^{(r)},t)$, *if* $s > r \geq 0$,
$F(x,A,+\infty) \geq F(x,A^{(+\infty)},t)$,
where $A^{(r)} = \{y \in X : F(y,A,t) = 1, \forall t > r\}$,
and $A^{(+\infty)} = \{y \in X : F(y,A,+\infty) = 1\}$;

FA6. $F(x,A,\cdot)\colon [0,+\infty) \to [0,1]$ *is left-continuous for all* $x \in X$ *and* $A \subset X$.

The map F is symmetric (or the FA-space (X,F) is symmetric) if

$$F(x,\{y\},t) = F(y,\{x\},t) \text{ for all } x,y \in X,\ t > 0.$$

Firstly, we will prove that there are two families of spaces satisfying the above six properties.

Lemma 2.2 (A-space $\hookrightarrow$ FA-space) *Let (X,δ) be an A-space. If we define a fuzzy set F_δ on $X \times \mathcal{P}(X) \times [0,+\infty]$ as follows:*

$$F_\delta(x,A,t) = 0 \qquad \text{if } t \leq \delta(x,A),$$
$$F_\delta(x,A,t) = 1 \qquad \text{if } t > \delta(x,A),$$

then (X,F_δ) is a FA-space.

Proof. FA1, FA2 and FA3 are immediate. In order to prove FA4, if $A,B \subseteq X$ we can suppose $\min(\delta(x,A),\delta(x,B)) = \delta(x,A) \leq \delta(x,B)$ and, then $\max(F(x,A,t),F(x,B,t)) = F(x,A,t)$. Thus $F_\delta(x,A\cup B,t) = 0$ for $t \leq \delta(x,A)$; and $F_\delta(x,A\cup B,t) = 1$ for $t > \delta(x,A)$. This completes the proof of FA4.

For every $r \in [0,+\infty)$, the set $A^{(r)}$ is the same in **A-space** and in **FA-space**, since

$$A^{(r)}_{A-space} = \left\{ x \in X \ / \ \delta(x,A) \leq r \right\}$$

$$= \left\{ x \in X \ / \ F_\delta(x,A,t) = 1, \forall t > r \right\} = A^{(r)}_{FA-space}.$$

Let us suppose that $s > r$. Since $\delta(x,A) \leq \delta(x,A^{(r)}) + r$, thus for $t \leq \delta(x,A) - s$, we have

$$t + r < t + s \leq \delta(x,A) \leq \delta(x,A^{(r)}) + r;$$

i.e., $t < \delta(x,A^{(r)})$ and $F_\delta(x,A,t+s) = F_\delta(x,A^{(r)},t) = 0$ for $t + s \leq \delta(x,A)$.

Otherwise, if $t + s > \delta(x,A)$, then $F_\delta(x,A,t+s) = 1 \geq F_\delta(x,A^{(r)},t)$. This completes the proof.

Theorem 2.3 (FM-space $\hookrightarrow$ FA-space) *Let $(X,M,*)$ be a FM-space. If we define a fuzzy set F on $X \times \mathcal{P}(X) \times [0,+\infty]$ for every $x \in X$, $t \in [0,\infty]$ and $A \subseteq X$ as follows:*

$$F_M(x,A,t) = \sup_{a \in A} M(x,a,t),$$

then (X,F_M) is a FA-space.

Proof. By FM1 and FM2, conditions FA1 and FA3 hold.

The proof of the second condition FA2 follows from the definition of $supA$ for empty sets, $sup\emptyset = 0$. If $x \in X$ and $A,B \subseteq X$ are subsets of X, then

$$\mathbf{Max}\left(\sup_{a \in A} M(x,a,t),\ \sup_{b \in B} M(x,b,t) \right) = \sup_{a \in A \cup B} M(x,a,t).$$

Consequently FA4 is proved.

We will prove now FA5. Let $x \in X$, $A \in \mathcal{P}(X)$ and $r \in [0,\infty]$. If $A = \emptyset$, the condition FA5 is trivial. It is enough to prove that the property FA5 holds true for every $A \neq \emptyset$. Since, for every $x,y,z \in X$, $t,s > 0$

$$M(x,z,t+s) \geq M(x,y,t) * M(y,z,s);$$

in particular,

$$M(x,y,t+s) \geq M(x,a,t) * M(a,y,s)$$

for every $x \in X$, $a \in A^{(r)}$, $y \in A$, $t > 0, s > r \geq 0$.

Taking supremum over $y \in A$, then

$$\begin{aligned} F_M(x,A,t+s) &= \sup_{y \in A} M(x,y,t+s) \\ &\geq M(x,a,t) * \sup_{y \in A} M(a,y,s) \\ &= M(x,a,t) * F_M(a,A,s). \end{aligned}$$

Since, $a \in A^{(r)}$, then $F_M(a,A,s) = 1$ for $s > r$. Thus

$$F_M(x,A,t+s) \geq M(x,a,t).$$

Taking supremum over $a \in A^{(r)}$, we have

$$F_M(x,A,t+s) \geq F_M(x,A^{(r)},t).$$

Note that additionally the following condition is satisfied:

FA7 Let $x,y \in X$ be such that $F_M(x,\{y\},t) = 1$ for every $t > 0$, then $x = y$.

3 Characterization

Now we ask about the maps $F : X \times \mathcal{P}(X) \times [0,+\infty] \to [0,1]$ that could provide to X of a *FA*-space structure. In this cases, it could be interesting to weaken axioms FA4 and FA5.

Proposition 3.1 *Let X be a non-empty set, $F : X \times \mathcal{P}(X) \times [0,+\infty] \to [0,1]$ be a map satisfying axiom FA2 and such that $F(x,A,t) = 1$ for every $t > 0$ and $x \in A \subseteq X$. Then:*
The axiom FA4 is equivalent to

FA4′. $F(x,A\cup B,t) \geq \mathbf{Max}(F(x,A,t),F(x,B,t))$ *for $t > 0$, for every non-empty subsets $A,B \subseteq X$ such that $A \neq B$ and $A \cup B \subset X$, and for each $x \in X \setminus (A \cup B)$.*

The axiom FA5 is equivalent to

FA5′. $F(x,A,t+s) \geq F(x,A^{(r)},t)$ *for $t > 0$, for every $s > r \geq 0$, $\emptyset \neq A \subset X$ and for each $x \in X \setminus A$.*

Proof. We prove that there exists some cases in which FA4 and FA5 are always true.

Property FA4′. If A is empty, then $F(x,A\cup B,t) = F(x,B,t)$ and, by the axiom FA2

$$\mathbf{Max}(F(x,A,t),F(x,B,t)) = \mathbf{Max}(F(x,\emptyset,t),F(x,B,t))$$

$$= \mathbf{Max}(0,F(x,B,t)) = F(x,B,t).$$

A similar conclusion is valid for $B = \emptyset$. We will suppose that A and B are non-empty sets. If $x \in A \cup B$, then, by hypothesis,

$$F(x,A\cup B,t) = 1 \geq \mathbf{Max}(F(x,A,t),F(x,B,t)).$$

ISBN: 978-989-95079-6-8

We can suppose that $x \notin A \cup B$, and thus $A \cup B = X$ is impossible. So, $A \cup B \subset X$. If $A = B$, the inequality is trivial.

Property FA5′. If $r = +\infty$, the inequality $F(x,A,+\infty) \geq F(x,A^{(+\infty)},t)$ is obvious. Therefore we can suppose that $r < +\infty$. If A is empty, then for each $r > 0$,

$$\emptyset^{(r)} = \left\{ x \in X \;/\; F(x,\emptyset,t) = 1, \forall t > r \right\} = \emptyset,$$

and then in FA5 we have the equality. We can suppose that A is non-empty. If $x \in A$, by hypothesis, $F(x,A,t) = 1$ for all $t > 0$. Then

$$F(x,A,t+s) = 1 \geq F(x,A^{(r)},t).$$

Consequently, we can suppose that $x \notin A$.

4 Properties

The *closure* of a subset A of a $FA-space$ (X,F) can be expressed by

$$\overline{A} = A^{(0)} = \{x \in X : F(X,A,t) = 1, \forall t > 0\}.$$

Although the set X has not been endowed with a topology, the following definition can be considered.

Definition 4.1 *A is called **closed** if $\overline{A} \subseteq A$.*

Lemma 4.2 *Let (x,F) be a **FA-space**, $x \in X$ and $A \subseteq X$.*

1. *If $a \in A$, then $F(a,A,t) = 1$ for each $t > 0$.*
2. *If $a \in A$, then $F(x,a,t) \leq F(x,A,t)$ for each $t > 0$.*
3. *If $A \subseteq B \subseteq X$, then $F(x,A,t) \leq F(x,B,t)$ for each $t > 0$.*
4. $\sup_{a \in A} F(x,a,t) \leq F(x,A,t)$ *for each $t > 0$.*
5. *If $A \subseteq B \subseteq X$, then $A^{(r)} \subseteq B^{(r)}$ for each $r \geq 0$.*
6. *Let $0 \leq r \leq s$, then*

$$A \subseteq \overline{A} = A^{(0)} \subseteq A^{(r)} \subseteq A^{(s)} \subseteq \cdots \subseteq X.$$

   In particular, $F(x,A,t) \leq F\left(x,\overline{A},t\right)$.

7. *A is closed if and only if $\overline{A} = A$.*
8. *If $r \geq 0$, then $\emptyset^{(r)} = \emptyset$.*
9. *$x \in A^{(r)}$ if and only if $x \in A^{(r+s)}$ for every $s > 0$.*
10. *The infimum $\inf_{[0,+\infty)}(\{r \geq 0 \,/\, x \in A^{(r)}\})$ is really a minimum, i.e.,*

$$A^{(r)} = \bigcap_{r<s} A^{(s)}.$$

11. *$F(x,A,\cdot)$ is a nondecreasing function.*

Proof.

1. By the axioms FA3 and FA4, if $a \in A$, we have

$$F(a,A,t) = F(a,A \cup \{a\},t)$$

$$\geq \mathbf{Max}(F(a,A,t),F(a,\{a\},t)) = \mathbf{Max}(F(a,A,t),1) = 1.$$

2. By the axiom FA4, if $a \in A$, we have

$$F(x,A,t) = F(x,B \cup \{a\},t)$$

$$\geq \mathbf{Max}(F(x,B,t),F(x,\{a\},t)) \geq F(x,\{a\},t),$$

   where $B = A \backslash \{a\}$.

3. If $A \subseteq B$, then $A \cup B = B$, and thus

$$F(x,B,t) = F(x,A \cup B,t) \geq \mathbf{Max}(F(x,A,t),F(x,B,t))$$

$$\geq F(x,A,t).$$

4. By 2, if $a \in A$, then $F(x,a,t) \leq F(x,A,t)$. Taking supremum

$$\sup_{a \in A} F(x,a,t) \leq F(x,A,t).$$

5. If $A \subseteq B$, by 3

$$F(x,A,t) \leq F(x,B,t).$$

   If $x \in A^{(r)}$, $F(x,A,t) = 1$ for each $t > r$ and then $F(x,B,t) = 1$ for each $t > r$. That is, $x \in B^{(r)}$.

6. If $a \in A$, by 1, $F(a,A,t) = 1$ for each $t > 0$. Thus $a \in \overline{A}$.
   Let $r \leq s$. If $x \in A^{(r)}$, $F(x,A,t) = 1$, for each $t > r$, then, in particular, $F(x,A,t) = 1$ for each $t > s$. Thus $x \in A^{(s)}$.

7. Trivial.

8. By FA2, it is immediate.

9. $x \in A^{(r)}$ if and only if $F(x,A,t) = 1$ for all $t > r$. This condition is equivalent to the following for each $s > r$,

$$F(x,A,t) = 1 \text{ for all } t > s.$$

   Then,

$$x \in A^{(r)} \Leftrightarrow x \in \bigcap_{r<s} A^{(s)} \Leftrightarrow x \in \bigcap_{s>0} A^{(r+s)}.$$

   That is, $x \in A^{(r)}$ if and only if $x \in A^{(r+s)}$ for all $s > 0$.

10. By 3, 6 and FA5

$$F(x,A,t+s) \geq F(x,A^{(r)},t) \geq F(x,A,t),$$

    if $s > 0$.

Next, the previous properties are refined.

Lemma 4.3 *Let (X,F) be a FA-space, $x \in X$, $t \geq 0$ and $A \subseteq X$.*

1. *$F(x,A,t) = F\left(x,\overline{A},t\right)$.*
2. *The closure $\overline{A}$ is a closed set, that is, $\overline{\overline{A}} = \overline{A}$.*
3. *A is closed if and only if for each $x \in X$ satisfying $F(x,A,t) = 1$, with $t > 0$, then $x \in A$.*
4. *$A^{(r)}$ is closed for every $r \geq 0$.*
5. *(X,F) satisfies the axiom FA7 if and only if the points are closed subsets, that is, $\overline{\{x\}} = \{x\}$.*
6. *$(A^{(r)})^{(s)} \subseteq A^{(r+s)}$ for every $r,s \geq 0$.*

ISBN: 978-989-95079-6-8

Proof.

1. Since $A \subseteq \overline{A}$, by 3 of lemma 4.2, $F(x,A,t) \leq F\left(x,\overline{A},t\right)$. Using axiom FA5 with $r=0$, it holds
$$F(x,A,t+s) \geq F(x,A^{(0)},t) = F(x,\overline{A},t)$$
for $s>0$.
Taking infimum on s, we have $F(x,A,t) \geq F(x,\overline{A},t)$.

2. By 6 of lemma 4.2, $\overline{A} \subseteq \overline{\overline{A}}$. Let $x \in \overline{\overline{A}} = A^{(0)^{(0)}}$. This is equivalent to
$$F\left(x,A^{(0)},t\right) = 1$$
for all $t>0$. Using axiom FA5 with $r=0$,
$$F(x,A,t+s) \geq F(x,A^{(0)},t) = 1$$
for $t,s>0$. Then $x \in A^{(0)} = \overline{A}$.

3. Trivial.

4. Let $x \in X$ be such that $F\left(x,A^{(r)},t\right) = 1$ for all $t>0$. By FA5,
$$F(x,A,t+s) \geq F\left(x,A^{(r)},t\right) = 1$$
for all $t>0$, $s>r\geq 0$. In particular $x \in A^{(r)}$. The previous property implies that $A^{(r)}$ is closed.

5. Trivial.

6. Let $r,s \geq 0$ and $x \in (A^{(r)})^{(s)}$, that is,
$$F\left(x,A^{(r)},t\right) = 1$$
for all $t>s$. By FA5,
$$F(x,A,t+s) \geq F\left(x,A^{(r)},t\right) = 1$$
for $t>s>r\geq 0$. Thus $x \in A^{(r+s)}$.

The following proposition is an interesting property related to closures of subsets of FA-spaces.

Proposition 4.4 *Let (X,F) be a FA-space, $r\geq 0$ and $A,B \subseteq X$ subsets of X.*

1. $A^{(r)} \cup B^{(r)} \subseteq (A\cup B)^{(r)}$.
2. *If (X,F) verifies the equality in FA4, then $(A\cup B)^{(r)} = A^{(r)} \cup B^{(r)}$.*

Proof. Since $A \subseteq A\cup B$ and $B \subseteq A\cup B$, by 5 in lemma 4.2,
$$A^{(s)} \subseteq (A\cup B)^{(s)}, \qquad B^{(s)} \subseteq (A\cup B)^{(s)}.$$
Then $A^{(s)} \cup B^{(s)} \subseteq (A\cup B)^{(s)}$.

Suppose that the equality holds in FA4. If $x \in (A\cup B)^{(r)}$, then
$$\mathbf{Max}\left(F(x,A,t),F(x,B,t)\right) = F(x,A\cup B,t) = 1,$$
for all $t>r$.

Then, for each $t>r$,
$$F(x,A,t) = 1 \text{ or } F(x,B,t) = 1.$$
Suppose that $t>r$ is such that $F(x,A,t)=1$. If $s>0$, then
$$F(x,A,t+s) = F(x,A^{(r)},t) \geq F(x,A,t) = 1.$$
Then $F(x,A,t)=1$ for all $t>r$; i.e., $x \in A^{(r)}$.

Consequently, $x \in A^{(r)}$ or $x \in B^{(r)}$and then $x \in A^{(r)} \cup B^{(r)}$.

References

[1] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[2] A. George and P. Veeramani. On some results in fuzzy metric spaces. *Fuzzy Sets Syst*, 64:395–399, 1994.

[3] O. Kaleva and S. Seikkala. On fuzzy metric spaces. *Fuzzy Sets Syst*, 12:215–229, 1984.

[4] D. Mihet. A banach contraction theorem in fuzzy metric spaces. *Fuzzy Sets Syst*, 144:431–439, 2004.

[5] I. Kramosil and J. Michalek. Fuzzy metrics and statistical metric spaces. *Kybernetika*, 11:336–344, 1975.

[6] R. Lowen. *Approach spaces. The missing link in the topology-uniformity-metric triad*. Oxford Mathematical Monographs. Oxford University Press, 1997.

[7] D. Vaughan R. Lowen, M. Sioen. Completing quasi-metric spaces - an alternative approach. *Houston J. Math.*, 29:113–136, 2003.

ISBN: 978-989-95079-6-8

Decision Making in Competitive Location using Fuzzy Sets

Clara M. Campos[1], Lourdes Canós[2], María J. Canós[3],
Vicente Liern[3], José A. Moreno[4], Dolores R. Santos-Peñate[5]

1 Dep. de Economía Financiera y Contabilidad,
University of La Laguna. 38271 La Laguna, Spain
2 Dep. de Organización de Empresas, Economía Financiera y Contabilidad,
Universidad Politécnica de Valencia. 46022 Valencia, Spain
3 Dep. de Economía Financiera y Matemática,
University of Valencia. 46071 Valencia, Spain
4 Dep. de Estadística, I.O. y C. Instituto Universitario de Desarrollo Regional,
University of La Laguna. 38271 La Laguna, Spain
5 Dep. de Métodos cuantitativos en Economía y Gestión,
University of Las Palmas de G.C., 35017 Las Palmas de G.C., Spain

Email: ccampos@ull.es, loucada@omp.upv.es, {maria.j.canos,vliern}@uv.es, jamoreno@ull.es, drsantos@ulpgc.es

Abstract—*This paper deals with the competitive location problems using fuzzy sets. The basic notions on fuzzy optimization and linear programming using fuzzy sets are briefly reviewed. The standard leader-follower location problem and its linear mathematical programming formulation are described. The works appeared in the literature concerning the use of fuzzy sets are analyzed.*

Keywords— Fuzzy Optimization, Competitive Location, Leader-Follower problem.

1 Introduction

Competitive location models represent the competition between two or more firms which provide goods or services to customers. The natural objective of each firm is to maximize its profit, which is often replaced by the maximization of the market share. That is, each firm tries to capture as many customers (or demand) as possible. The problem of locating facilities in a competitive environment has been addressed in several papers in the field of Operations Research, Management Science, Regional Economy and Economic Geography, see e.g. Craig et al. (1984) [9], Drezner (1995) [13], Eiselt and Laporte (1989,1996) [14][15], Friesz et al. (1988) [16], Hakimi (1990) [17], Plastria (2001) [22], Santos-Peñate et al. (2007) [23], Spaulding and Cromley (2007) [25], Till (1992, 2000) [26][27], and the references therein.

In many cases, the real competitive contexts cannot be described with precision. The complex, subjective, and dynamic behavior of real customers produces a high degree of uncertainty related to their decision making process. Consequently the most common scenario corresponds to problems where the information of the parameters and variables is vague or imprecise. Fuzzy sets theory is a rigorous and effective instrument to treat problems where the necessary information is imprecise. In fuzzy optimization two families of problems exist. On the one hand we have problems where the objective function and the constraints are fuzzy and, on the other hand, those where the coefficients are fuzzy. Fuzzy functions are characterized by their membership function. Fuzzy decisions are the intersection of fuzzy sets corresponding to the objective function and to the constraints. Linguistic variables can also be considered (See Zadeh [34]).

In this work we consider the application of fuzzy optimization methodologies to competitive location problems. The paper is organized as follows. In section 2, we describe the fuzzy optimization models. The basic concepts of competitive location are presented in Section 3. Section 4 is devoted to the solution approaches to competitive location problems using fuzzy optimization and fuzzy sets. Finally, Section 5 contains the conclusions.

2 Fuzzy Optimization Problems

Optimization in their most general form involves finding optimal solutions according to stated criteria. This task is usually formulated as optimization problems using objective functions and constraints. The procedures use mathematical properties of the objective and constraints functions. In practice, however, many situations lack the exact information that is needed in the problem, including its objective and constraints, or in other cases, where it is unreasonable to access such specific constraints or clearly defined objective functions. In these situations it is advantageous to model and solve the problem using soft computing and fuzzy techniques.

An *optimization problem* consists of finding the value of the *decision variables* so that an *objective function* is minimized or maximized when the possible values of the variables are subject to a set of *constraints*. The objective function, denoted by f, is defined on a set of solutions denoted by X and the constraints are given by a vector function g in such a way that the problem is formulated as follows:

ISBN: 978-989-95079-6-8

$$\begin{aligned} &\text{Minimize} \\ &\quad f(\vec{x}) \\ &\text{subject to:} \\ &\quad g(\vec{x}) \le \vec{0} \\ &\quad \vec{x} \in X \end{aligned} \tag{1}$$

where $\vec{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ is the vector of decision variables, f and g are defined on $\mathbb{R}^n$ and X is a subset of $\mathbb{R}^n$. The mathematical formulations are presented in this paper using "minimize" operator because the other case, the maximization case, is analogous.

Among the optimization problems that are included in Mathematical Programming, the models that have received the most attention and have offered the most useful applications in different areas are Linear Programming (LP) models, which is the single objective linear case.

The classic problem of LP is to find the maximum or minimum values of a linear function subject to constraints that are represented by linear equations or inequalities. The most general formulation of this problem is:

$$\begin{aligned} &\text{Minimize} \\ &\quad \vec{c}^{\,t}\vec{x} \\ &\text{subject to:} \\ &\quad A\vec{x} \le \vec{b} \\ &\quad \vec{x} \ge \vec{0} \end{aligned} \tag{2}$$

where $\vec{c} = (c_1, c_2, ..., c_n) \in \mathbb{R}^n$ is the objective vector, $\vec{b} = (b_1, b_2, ..., b_m) \in \mathbb{R}^m$ is the right hand side vector, and $A = [a_{ij}]$ is an (n,m)-matrix where a_{ij} is the coefficient of variable x_j in constraint i. In an economic context where the aim is the minimization of the total cost with a limitation of the resources, $\vec{c}$ is the cost vector, $\vec{b}$ is the vector of resources and A is known as the technological matrix. An equivalent formulation of the linear problem is the following:

$$\begin{aligned} &\text{Minimize} \\ &\quad \sum_{j=1}^{n} c_i x_j \\ &\text{subject to:} \\ &\quad \sum_{j=1}^{n} a_{ij} x_j \le b_i \qquad i = 1, 2, ...m \\ &\quad x_j \ge 0, \qquad\quad j = 1, 2, ..., n \end{aligned} \tag{3}$$

In many real situations not all the constraints and objective functions can be valued in a precise way and in these situations we are dealing with the general problem form of Fuzzy Linear Programming (FLP). FLP is characterized as follows: a_{ij}, b_i and c_j can be expressed as *fuzzy numbers*, x_i as variables whose states are fuzzy numbers, addition and multiplication operates with fuzzy numbers, and the inequalities are among fuzzy numbers. Different FLP models can be considered according to the elements that contain imprecise information, this is the criterion used in the classification proposed in Verdegay (1995) [33] and Cadenas and Verdegay (1999) [1].

If we take as a basis the classification proposed in Verdegay [33], we have models with the fuzzy feasible sets, models with fuzzy goals, models with fuzzy coefficients of objective function, models with coefficients of the technological matrix and fuzzy right hand sides, and totally fuzzy models where all the elements of the problem are fuzzy.

1. Models with fuzzy objective function

These models are those whose objective function is not fully known. In LP problems where the costs are known with imprecision they are represented by an n-dimensional fuzzy vector $\vec{\tilde{c}} = (\tilde{c}_1, \tilde{c}_2, ..., \tilde{c}_n)$, leading to the following model:

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \vec{\tilde{c}}^{\,t} x \\ &\text{subject to} \\ &\quad A\vec{x} \le \vec{b} \\ &\quad \vec{x} \ge \vec{0} \end{aligned} \tag{4}$$

Evidently, z is also a fuzzy number, but $\vec{x}$ can be a vector of fuzzy or non-fuzzy numbers, and each fuzzy cost is described by its corresponding membership function $\mu_j(c)$. Each coefficient $\tilde{c}_{kj}$ of the objective function is a plane fuzzy number of the L-R type with modal interval $\left[\underline{c}_j, \overline{c}_j\right]$ and membership functions g_j and h_j (which can be linear, parabolic, etc.). Delgado et al. (1989) [11] prove that the solution can be obtained with the multi-objective auxiliary model:

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \left[\vec{c}_1^{\,t}\vec{x}, \vec{c}_2^{\,t}\vec{x}, ..., \vec{c}_{2^n}^{\,t}\vec{x}\right] \\ &\text{subject to} \\ &\quad A\vec{x} \le \vec{b} \\ &\quad \vec{x} \ge \vec{0}, \\ &\quad \alpha \in [0,1], c_{kj} \in \left\{g_j^{-1}(1-\alpha), h_j^{-1}(1-\alpha)\right\} \\ &\quad k = 1, ..., 2^n, j = 1, ..., n. \end{aligned} \tag{5}$$

2. Models with feasible fuzzy set (fuzzy constraints)

This is the case where constraints can be satisfied, and consequently the feasible region, can be defined as a fuzzy set; it should be defined by means of a membership function $\mu : \mathbb{R}^n \to [0,1]$. In such a situation, for each constraint i, a desirable quantity b_i is considered, but the possibility that it is greater is accepted until a maximum $b_i + t_i$ (t_i is referred

ISBN: 978-989-95079-6-8

to as a violation *tolerance level*). This model is represented by

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \vec{c}^{\,t}\vec{x} \\ &\text{subject to:} \\ &\quad A\vec{x} \lesssim \vec{b} \\ &\quad \vec{x} \geq \vec{0} \end{aligned} \tag{6}$$

where the symbol $\lesssim$ indicates the imprecision of the constraints and where each fuzzy constraint $\vec{a}_i^{\,t} x \lesssim \vec{b}_i$ is specified by a membership function in the form:

$$\mu_i\left(\vec{a}_i^{\,t}\vec{x}\right) = \begin{cases} 1 & \text{if } \vec{a}_i^{\,t}\vec{x} < b_i \\ f_i\left(\vec{a}_i^{\,t}\vec{x}\right) & \text{if } b_i \leq \vec{a}_i^{\,t}\vec{x} \leq b_i + t_i \\ 0 & \text{if } b_i + t_i < \vec{a}_i^{\,t}\vec{x} \end{cases} \tag{7}$$

where $\vec{a}_i = \left(a_{i1}, a_{i2}, \ldots, a_{in}\right) \in \mathbb{R}^n$ is the vector of coefficients for constraint *i*. Expression (7) means that, for each constraint *i*, given the level of tolerance t_i, each point (*n*-dimensional vector) $\vec{x}$ *is* associated with a number $\mu_i(\vec{x}) \in [0,1]$ which is known as the degree of fulfillment (or verification) of the constraint *i*. The functions f_i are assumed continuous and monotonous non-decreasing functions. In particular, Verdegay (1982) [32], using the representation theorem for fuzzy sets, proves that the solutions for the case of linear functions f_i can be obtained from the auxiliary model:

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \vec{c}^{\,t}\vec{x} \\ &\text{subject to} \\ &\quad A\vec{x} \leq \vec{b} + (1-\alpha)\vec{t}^{\,t} \\ &\quad \vec{x} \geq \vec{0}, \alpha \in [0,1] \end{aligned} \tag{8}$$

where $\vec{t} = \left(t_1, t_2, \ldots, t_m\right)$.

3. Models with fuzzy goals

A fuzzy optimization problem with fuzzy goals is one whose goal set is fuzzy, that is, it allows the objective function value to be slightly below the minimum goal for a maximization problem, and similarly for a minimization problem. The corresponding linear model is expressed in the following way:

$$\begin{aligned} &\widetilde{\text{Minimize}} \\ &\quad z = \vec{c}^{\,t}\vec{x} \\ &\text{subject to} \\ &\quad A\vec{x} \leq \vec{b} \\ &\quad \vec{x} \geq 0 \end{aligned} \tag{9}$$

If t_0 is the maximum quantity that the function objective should be inferior to the minimum goal c_0, then each vector $\vec{x}$ is associated with a number $\mu_0\left(\vec{x}\right)$, which represents the degree that the decision maker considers to be an achieved goal. It is defined according to the following function:

$$\mu_0\left(\vec{x}\right) = \begin{cases} 1 & \text{if } \vec{c}^{\,t}\vec{x} > c_0 \\ f_0\left(\vec{c}^{\,t}\vec{x}\right) & \text{if } c_0 - t_0 \leq \vec{c}^{\,t}\vec{x} \leq c_0 \\ 0 & \text{if } \vec{c}^{\,t}\vec{x} < c_0 - t_0 \end{cases} \tag{10}$$

where f_0 is a continuous monotonous non-decreasing function. The corresponding satisfactory solutions can be obtained by solving the equivalent problem when a level for the α-cuts is chosen.

4. Models with fuzzy coefficients in the technological matrix

Consider a problem of this type:

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \vec{c}^{\,t}\vec{x} \\ &\text{subject to} \\ &\quad \underset{\sim}{A}\vec{x} \lesssim \underset{\sim}{\vec{b}} \\ &\quad \vec{x} \geq \vec{0} \end{aligned} \tag{11}$$

where the values of the technological matrix and the right hand sides are fuzzy numbers. Fuzzy constraints can also be included. Delgado et al. (1987) [10] also include imprecision in the constraints. They consider the fuzzy relations of the constraints with the application of a ranking function *g* to compare the fuzzy terms. This new formulation is expressed by the auxiliary problem:

$$\begin{aligned} &\text{Minimize} \\ &\quad z = \vec{c}^{\,t}\vec{x} \\ &\text{subject to} \\ &\quad \underset{\sim}{\vec{a}_i}\vec{x} \leq_g \underset{\sim}{b_i} + \underset{\sim}{t_i}(1-\alpha),\ i = 1,\ldots m \\ &\quad \vec{x} \geq \vec{0}, \alpha \in [0,1] \end{aligned} \tag{12}$$

3 Competitive Location

In competitive location problems two or more firms compete by mean of their locations for providing services or products to customers. Several scenarios may be considered:

- No firm operates in the market and firm F_1 wants to enter the market with p_1 facilities.
- There are *s* firms F_i operating in the market, each firm F_i has p_i facilities and a new firm F_{s+1} wants to enter the market with p_{s+1} facilities.
- There are *s* firms F_i operating in the market, each firm F_i has p_i facilities and a firm F_k wants to open $\overline{p}_k$ facilities or to close $\underline{p}_k$ facilities.

In particular, the Stackelberg location model is a standard two-stage problem where a firm, the leader, chooses its location points and then a competitor, the follower, enters

ISBN: 978-989-95079-6-8

the market and decides its locations taking into account the leader position. The problem of the follower is to find the optimal locations given the position of the leader, the problem of the leader is to determine the best locations taking into account the reaction of the follower to any possible strategy of a competitor.

Different objectives can be considered:

- The maximization of the market share.
- The minimization of the competitor market share.
- The maximization of the difference between its market share and the competitor market share.

For modelling the customer's behaviour in the problem we can use different rules:

- *Binary choice rule*. A choice rule where the customers patronize their closest facility.
- *Partially binary choice rule*. A choice rule where each customer patronizes the closest facility of each firm.
- *Proportional choice rule.* With this choice rule, customers patronize all the facilities according to a non-increasing function of the travel distance.
- *Threshold choice rule.* A customer choice rule where a threshold-sensitive behaviour is assumed.

Moreover, whether goods are essential or not, demand are said to be inelastic (constant demand) or elastic (demand varies with distance). The customer behaviour is modelled by their optimization problems that result from the application of the choice rule for them.

For binary inelastic demand, the leader, follower and customer decision problems can be formulated as linear optimization models (see Campos-Rodríguez et al. [6]). The problems are stated as follows.

Let $m = |L|$ be the number of possible facility locations and $n = |C|$ be the number of customer locations. Let $d_{ki} = d(c_k,f_i)$ be the distance between the k-th customer location c_k and the i-th facility point f_i. In addition, let h_k be the total demand of the customers located at c_k. Finally, let $K = \{1, 2, ..., n\} = [1..n]$ denote the index set for the customer locations and $I = \{1, 2, ..., m\} = [1..m]$ denote the index set for the facility locations.

A set Z of location points is identified by a binary vector z with size m. This vector is $z = (z_1, z_2,..., z_m) = (z_i, i \in [1..m])$ where $z_i = 1$ if $f_i \in Z$ and $z_i = 0$ otherwise. Thus the corresponding set is given by $Z = \{ f_i \in L: z_i = 1 \}$.

The decision variables in the leader and follower location problems are the m-vectors x and y of 0-1 or binary decision variables corresponding to the sets X and Y.

For the customer problem, let z_{ki} be the 0-1 decision variables indicating whether the customers located at the k customer location c_k prefer the location f_i for the facility. However, in each customer problem, the sets X and Y are data that are represented by corresponding m-vectors of 0-1 values $\bar{x}$ and $\bar{y}$ for the variables x and y.

For the follower location problem we have an m-vector of values $\bar{x}$ and an m-vector of variables y. The binary choice rule oriented to the leader implies that the scalar product $\bar{x} \cdot \bar{y}$ is zero.

Using the coefficients b_{ij}^k and c_{ij}^k given, for any $i,j \in [1..m]$, $k \in [1..n]$, respectively, by:

$$b_{ij}^k = \begin{cases} 1 & d_{ki} \le d_{kj} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

and

$$c_{ij}^k = \begin{cases} 1 & d_{ki} < d_{kj} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

The customer decision problem of the customers at location c_k, for every $k \in [1..n]$, is the linear feasible problem in the variables z_{ki}, for any $i \in [1..m]$, given by:

$$\begin{aligned} &\sum_{i=1}^{m} z_{ki} = 1 \\ &z_{ki} \le \bar{x}_i + \bar{y}_i && i \in [1..m] \\ &z_{ki} \le 1 - c_{ji}^k \bar{y}_j - b_{ji}^k \bar{x}_j - b_{ij}^k \bar{x}_i && i, j \in [1..m] \\ &z_{ki} \in \{0,1\} && i \in [1..m] \end{aligned} \tag{15}$$

For the follower problem consider the coefficients c_{ki} given, for any $k \in [1..n]$ and $i \in [1..m]$, by

$$c_{ki} = \begin{cases} 1 & d_{ki} < \min\{d_{kj} : \bar{x}_j = 1\} \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

Then, the follower problem is the linear optimization problem in the variables z_{ki}, $k \in [1..n]$, $i \in [1..m]$, given by:

maximize

$$\sum_{i=1}^{m} \sum_{k=1}^{n} h_k z_{ki}$$

such that:

$$\begin{aligned} &\sum_{i=1}^{m} y_i = r \\ &\sum_{i=1}^{m} z_{ki} \le 1 && k \in [1..n] \\ &z_{ki} - c_{ki} y_i \le 0 && k \in [1..n]; i \in [1..m] \\ &z_{ki}, y_i \in \{0,1\} && k \in [1..n]; i \in [1..m] \end{aligned} \tag{17}$$

Finally, consider the set J of all the possible selection for the follower. The leader problem is the linear optimization problem in the variables w and for any $k \in [1..n]$, $i \in [1..m]$ and $j \in J$, by

ISBN: 978-989-95079-6-8

$$z_{ki}^{j} = \begin{cases} 1 & d_{ki} \le \min\{d_{kj} : j \in J\} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

given by:

minimize

w

such that:

$$\begin{aligned} & \sum_{i=1}^{m} x_i = p \\ & \sum_{k=1}^{n} h_k - \sum_{i=1}^{m}\sum_{k=1}^{n} h_k z_{ki}^{j} \le w \quad j \in J \\ & \sum_{i=1}^{m} z_{ki}^{j} \le 1; j \in J \qquad k \in [1..n] \\ & z_{ki}^{j} - c_{ki}^{j} x_i \le 0 \qquad k \in [1..n]; i \in [1..m]; j \in J \\ & z_{ki}^{j}, x_i \in \{0,1\} \qquad k \in [1..n]; i \in [1..m]; j \in J \\ & w \ge 0 \end{aligned} \quad (19)$$

See Campos-Rodríguez et al. (2009) [6] for details.

4 Solution Approaches

Several kind of uncertainty have been considered in Location problems (Benati (2000) [2], Benati and Hansen (2000) [1], Devletoglou (1965) [12], and Sikdera et al. (2007) [24]). Fuzzy sets have been used in several location problems (see Canós et al. (1998) [7]), In general, the user preferences have uncertainty and the corresponding utilities are uncertain. Therefore their choice must be described using a membership function and the competitive models become fuzzy. On the other hand, travel times to reach the facilities are also uncertain and they must be expressed as triangular fuzzy number. The distances d_{ki} should be replaced in the corresponding model by a triangular fuzzy distance $\tilde{d}_{ki} = (d_{ki}^{1}, d_{ki}^{2}, d_{ki}^{3})$. The using of the α-cuts will deal to a crisp model and the solution will be a function of this parameter α. There are several works in Literature using fuzzy methods for competitive location problems.

Liang *et al.* [18] analyze the optimum output quantity decision analysis of a duopoly market under a fuzzy decision environment. To efficiently handle the fuzziness of the decision variables, the linguistic values, subjectively represented by the trapezoidal fuzzy numbers, are used to act as the evaluation tool of decision variables such as fixed cost and unit variable cost.

Osumi *et al.* [21] investigate a competitive facility location problem where there are two facilities on a linear market. Customers at a demand point utilize the facility which seems to be the nearest one from them. They do not distinguish the small difference between two distances. This preference is formulated by introducing relative fuzzy difference based on the actual distance between two points. This paper considers the problems to find the optimal location for the follower and for the leader.

Uno and Katagiri [31] study a new optimal location problem, called defensive location problem (DLP). In the DLPs, a decision maker locates defensive facilities in order to prevent her/his enemies from reaching an important site, called a core; the DLPs are formulated as bilevel 0-1 programming problems to find Stackelberg solutions. an interactive fuzzy satisfying method is proposed,

Uno *et al.* [28] y [29] analyse competitive facility location problems and consider the case where the set of customers is divided into several subsets or types by investigating their preferences and criteria. Since the preferences and criteria often include the vagueness of human's judgement they express such parameter by fuzzy numbers.

Uno and Masatoshi [31] propose a multi-objective approach for competitive facility location models with fuzzy numbers. In cases where the objective of each firm that locates its own facilities is only to maximize its reward, the location of all facilities is usually crowded on some regions which have many populations or are a hub for all regions. Such a location forces the firms to compete extremely and the market about facilities is insecure. Therefore, they formulate a multi-objective problem in which the decision maker is an arbitrator and whose objectives are rewards for all firms. By using a solution to the problem, they find a good restriction such that more firms can survive.

Considering the imprecision or vagueness in the customer choice rule, the appropriate model for dealing problem (19) is that with fuzzy coefficients in the technological matrix since the coefficients c_{ki}^{j}. Several relaxed choice rules have been proposed to allow a soft behavior of the customers. These rules will result in corresponding membership functions for the coefficients c_{ki}^{j} that provide a particular linear programming problem with fuzzy technological coefficients. The problem is then solved by the corresponding auxiliary problem.

5 Conclusions and further research

Fuzzy sets constitute an appropriate approach to manage the uncertainty that appears in real competitive location problems. Fuzzy competitive location problems need to be clearly formalized and classified. There are several research works in some competitive location problems with fuzzy elements. However, practical solution procedures have to be tested for real problems Chance-constrained or Possibility theory and the classical use of α-cuts.

Acknowledgements

The authors have supported by projects TIN2008-06873-C04-01, TIN2008-06873-C04-02, ECO2008-05589 of Spanish govern and FEDER, and PI042005/044 of Canary govern.

ISBN: 978-989-95079-6-8

References

[1] S. Benati, P. Hansen, 2000. Competitive Location with Random Utilities: Problem Formulation and Algorithms. *Cahiers du GERAD*

[2] S. Benati 2000. NP-hardness of some competitive location models with probabilistic choice rules. *Studies Locational Analysis* 14, 211-231.

[3] J.M. Cadenas, J.L. Verdegay, 1999. Modelos de optimización con datos imprecisos. *Ed. Universidad de Murcia*.

[4] J.M. Cadenas, J.L. Verdegay, 2006. A Primer on Fuzzy Optimization Models and Methods. *Iranian Journal of Fuzzy Systems* 3(1):1-21.

[5] J.M. Cadenas, J.L. Verdegay, 1997. Using fuzzy numbers in linear programming. *IEEE Transactions on Systems, Man, and Cybernetics*, Part B 27(6): 1016-1022.

[6] C. Campos-Rodríguez, J.A. Moreno-Pérez, D.R. Santos Peñate, 2009. An exact procedure and LP formulation for the leader-follower location problem. To appear in TOP

[7] M.J. Canos, C. Ivorra, V. Liern, 1998. Fuzzy methods in location problems. *X Meeting of the EURO Working Group on Locational Analysis*, February 10-13, Murcia (Spain)

[8] A. Charnes, W.W. Cooper, 1959. Chance-Constrained Programming. *Management Science,* 6:73-79.

[9] S. Craig, Ghosh A., McLafferty S. 1984. Models of the retail location process: A review. *Journal of Retailing* 60(1), 5-36.

[10] M. Delgado, J.L. Verdegay, M.A. Vila, 1987. Imprecise costs in mathematical programming problems. *Control and Cybernetics* 16(2):113–121.

[11] M. Delgado, J.L. Verdegay, M.A. Vila, 1989. A general model for fuzzy linear programming. *Fuzzy Sets and Systems* 29:21-29.

[12] N.E. Devletoglou, 1965. A dissenting view of duopoly and spatial competition. *Economica* 32(126), 140-160.

[13] T. Drezner, 1995. Competitive facility location in the plane. In *Facility Location. A survey of applications and methods*, Z. Drezner (Ed.), Springer, Berlin, 285-300.

[14] H.A. Eiselt, G. Laporte, 1989. Competitive spatial models. *European Journal of Operational Research* 39, 231-242.

[15] H.A. Eiselt, G. Laporte, 1996. Sequential location problems. *European Journal of Operational Research* 96, 217-231.

[16] T.L. Friesz, T. Miller, R.L. Tobin, 1988. Competitive network facility location models: A survey. *Papers of the Regional Science Association* 65, 47-57.

[17] S.L. Hakimi,. 1990. Location with spatial interactions: Competitive locations and games. In *Discrete Location Theory*, Mirchandani, P.B. and R.L. Francis (eds.), John Wiley \& Sons, New York, 439-478.

[18] G.-S. Liang, L.-Y. Lin, C.-F. Liu, 2008. The optimum output quantity of a duopoly market under a fuzzy decision environment *Computers & Mathematics with Applications* 56(5):1176-1187

[19] B. Liu, 2001. Fuzzy random Chance-Constrained Programming. *IEEE Transactions on Fuzzy systems* 9(5):713-720

[20] B. Liu, 2009. *Theory and practice of uncertain programming* (2nd ed.). Springer-Verlag.

[21] S. Osumi, S. Shiode, 2003. Competitive facility location problem with fuzzy relative distance *Journal of Nonlinear Convex Analysis*. 4(1):141-149.

[22] F. Plastria, 2001. Static competitive facility location: An overview of optimisation approaches. *European Journal of Operational Research* 129:461-470.

[23] D.R. Santos-Peñate, R. Suárez-Vega, P. Dorta-González 2007. The leader-follower location model. *Networks and Spatial Economics* 7:45-61.

[24] I.U. Sikdera, A. Gangopadhyay, 2007. Managing uncertainty in location services using rough set and evidence theory. *Expert Systems with Applications* 32(2):386-396

[25] B.D. Spaulding, R.G. Cromley, 2007. Integrating the maximum capture problem into a GIS framework. *Journal of Geographical Systems. Geographical Information, Analysis, Theory and Decision* 9(3): 267-288.

[26] J-C. Till, 1992. Competitive strategies for multi-establishment firms. *Economic Geography* 68(3): 290-309.

[27] J-C. Till, 2000. Network competition and branch differentiation with consumer heterogeneity. *The Annals of Regional Science* 34:451-468.

[28] T. Uno, H. Ishii, S. Saito, S. Osumi, 2001. An introduction to competitive facility location problems with vagueness. *Knowledge-Based Intelligent Information Engineering System and Allied Technologies* 69:1375-1379.

[29] T. Uno, H. Ishii, S. Saito, S. Osumi, 2004. Competitive facility location problem: An algorithm for the problem concerning the existence of multi-type customers, *Central European Journal of Operations Research*, 12(1):79-85.

[30] T. Uno, H. Katagiri, 2008. Single- and multi-objective defensive location problems on a network. *European Journal of Operational Research*, 188(1):76-84

[31] T. Uno, S. Masatoshui, 2003. Multi-Objective Approach for Competitive Facility Location Model with Fuzzy Numbers *Faji Shisutemu Shinpojiumu Koen Ronbunshu* 19:711-714

[32] J.L. Verdegay, 1982. Fuzzy mathematical programming *In: Fuzzy Information and Decision Processes,* Gupta MM, Sanchez E (eds). North-Holland.

[33] J.L. Verdegay, 1995. Fuzzy optimization: models, methods and perspectives. *6th IFSA-95 World Congress*. p. 39-71.

[34] L.A. Zadeh, 1965. Fuzzy sets. *Information and Control*, 8:338-353.

ISBN: 978-989-95079-6-8

Using Syntactic Possibilistic Fusion for Modeling Optimal Pessimistic Qualitative Decision

Salem Benferhat[1] Hadja Faiza Khellaf-Haned[2] Aicha Mokhtari[2] Ismahane Zeddigha[2]

1.CRIL, Université d'Artois
Rue Jean Souvraz, S.P.18 62307 Lens, Cedex France
2.LRIA, Université des Sciences et de la Technologie Houari Boumediene
BP 32 EL ALIA 16111 Bab Ezzouar Alger Algérie
Email: benferhat@cril.univ-artois.fr, hanedfaiza@yahoo.com, {aissani_mokhtari,izeddigha}@yahoo.fr

Abstract— *This paper describes the use of syntactical data fusion to computing possibilistic qualitative decisions. More precisely qualitative possibilistic decisions can be viewed as a data fusion problem of two particular possibility distributions (or possibilistic knowledge bases): the first one representing the beliefs of an agent and the second one representing the qualitative utility. The proposed algorithm computes a pessimistic optimal decisions based on data fusion techniques. We show that the computation of optimal decisions is equivalent to computing an inconsistency degree of possibilistic bases representing the fusion of agent's beliefs and agent's preferences.*

Keywords— Data Fusion, Decision theory, Pessimistic Criteria, Possibilistic Logic.

1 Introduction

The problem of decision on uncertainty is crucial for many applications in artificial intelligence. A decision theory must provide some criteria for representing behaviors of an agent in order to take the optimal decision amongst a set of decisions. Possibility theory represents one of the theory which deals with uncertain information. This paper presents a computation of pessimistic decisions based on syntactic possibilistic fusion operations[1]. A qualitative possibilistic decision model [2] allows a gradual expression of both agent's preferences and knowledge. The preferences and the available knowledge about the world are expressed in ordinal way. In [2], the authors have proposed two qualitative criteria for ordinal decision approaches under uncertainty: the pessimistic and the optimistic decisions criteria. The first one being more cautious than the second one for computing optimal decisions. A method for computing optimal decisions, based on ATMS (Assumption-based Truth Maintenance System), has been proposed in [3]. Using the pessimistic criteria, the procedure is translated into an ATMS problem [4, 5].

Recently, in a companion paper [6] a method has been proposed for computing optimal optimistic decisions using possibilistic fusion modes. This paper is also developed in the same spirit. Indeed, qualitative possibilistic decisions can be viewed as a data fusion problem of two particular possibility distributions: the first one representing the beliefs of an agent and the second one representing the qualitative utility (or agent's preferences). The handling of possibilistic decision raises new issues and it requires additional steps such that the computation of negated possibilistic bases.

The rest of this paper is organized as follow. Section 2 gives a brief backgrounds on possibilistic logic, on qualitative decision problems under uncertainty based on possibilistic logic where agent's beliefs and the preferences are expressed by means of possibilistic bases. In Section 3, we provide a preliminary step for our work. Section 4 presents an efficient and unified way of computing pessimistic qualitative decisions based on syntactic counterpart of data fusion problem. The paper ends with some conclusions in Section 5.

2 Backgrounds

2.1 Possibilistic Logic

This section gives a brief refresher on possibilistic logic. See [7] for more details on possibilistic logic. Let $\mathcal{L}$ be a finite propositional language and Ω be the set of all propositional interpretations. $\phi, \psi, \ldots$ denote propositional formulas. For an interpretation ω and a propositional formula $\phi, \omega \models \phi$ means that ω is a model (in the sense of propositional logic) of ϕ. A possibility distribution [7, 8] π is a mapping from a set of interpretations Ω into the unit interval [0,1]. $\pi(\omega)$ represents the degree of compatibility (or consistency) of the interpretation ω with available pieces of information. Given a possibility distribution π, two dual measures are defined on the set of propositional formulas:

- The possibility (or consistency) measure of a formula ϕ, defined by:

$$\Pi(\phi) = max\{\pi(\omega) : \omega \models \phi \text{ and } \omega \in \Omega\} \quad (1)$$

  which evaluates the extent to which ϕ is consistent with available beliefs expressed by π.

- The necessity (or certainty) measure of a formula ϕ, defined by:

$$N(\phi) = 1 - \Pi(\neg\phi) \quad (2)$$

  which evaluates the extent to which ϕ is entailed by available beliefs.

A possibilistic knowledge base Σ is a set of weighted formulas:

$$\Sigma = \{(\phi_i, \alpha_i) : i = 1, ..., n\},$$

where ϕ_i is a propositional formula and $\alpha_i \in]0, 1]$ represents the certainty level of ϕ_i. Each piece of information (ϕ_i, α_i) of a possibilistic knowledge base can be viewed as a constraint that restricts possibility degrees associated with interpretations [7].

If an interpretation ω satisfies ϕ_i then its possibility degree is equal to 1 (ω is completely compatible or consistent with

ISBN: 978-989-95079-6-8

the belief ϕ_i), otherwise it is equal to $1 - \alpha_i$ (the more ϕ_i is certain, the less ω is possible). In particular, if $\alpha_i = 1$, then any interpretation falsifying ϕ_i, is such that its possibility degree is equal to 0, namely is impossible.

More formally, the possibility distribution associated with a weighted formula (ϕ_i, α_i) is [7] $\forall \omega \in \Omega$:

$$\pi_{(\phi_i,\alpha_i)}(\omega) = \begin{cases} 1 - \alpha_i & if\ \omega \not\models \phi_i \\ 1 & \text{otherwise} \end{cases} \quad (3)$$

More generally, the possibility distribution associated with a qualitative possibilistic knowledge base Σ is: $\forall \omega \in \Omega$:

$$\pi_\Sigma(\omega) = min\{\pi_{(\phi_i,\alpha_i)}(\omega) : (\phi_i, \alpha_i) \in \Sigma\}. \quad (4)$$

2.2 Possibilistic Qualitative Decision

Several works have been proposed for dealing with qualitative decision theory under uncertainty. Some approaches consider only all-or-nothing notions of utility and plausibility [9], others use in addition a preference ordering on consequences [10]. However, in many applications the knowledge bases may be pervaded with uncertainty, and the goals may not have equal priority. Dubois, Le Berre, Prade and Sabbadin [3] have enriched logical view of the decision problem by assigning levels of certainty to formulas in the knowledge bases, and levels of priority to the goals. They have proposed two syntactic approaches based on possibilistic logic, applied on two stratified logical bases that model gradual knowledge and preferences. The first one being more cautious than the second, for computing optimal decisions [3].

It has been shown in [3] that the semantics underlying the two syntactic approaches are in agreement with the two qualitative utility functions proposed in [2].

Let K be a stratified base which represents level of certainty of the knowledge, $K = \{(\phi_i, \alpha_i) : i = 1, ..., n\}$ where $\alpha_i \in S$ such that $(\alpha_i > 0)$ denotes the degree of certainty, and the ϕ_i 's are formulas in $\mathcal{L}$ where decision literals may appear.

Let P be a stratified base expressing preferences, $P = \{(\psi_i, \beta_i) : i = 1, ..., n\}$ where $\beta_i \in S$ such that $(\beta_i > 0)$ is a degree of priority, and the ψ_i's are formulas in $\mathcal{L}$ where decision literals may also appear.

Let $K_{\geq\alpha}$ denotes the set of formulas with certainty at least equal to α . Let $P_{\geq\beta}$ denotes the set of formulas with priority at least equal to β. We also recall that $K_{>\alpha}$ (with $\alpha < 1$) denotes the set of formulas with certainty strictly greater than α and $P_{>\beta}$ (with $\beta < 1$) denotes the set of formulas with priority strictly greater than β.

The certainty degrees and the priority degrees are assumed to be commensurate and assessed to the same linearly order scale S [2]. The top element of S will be denoted by 1, and the bottom element will be denoted by 0.

For any set A of formulas, $A^\wedge$ denotes the logical conjunction of the formulas in A. Let $D = \{l_i\}$ be a set of decision variables, where l_i are distinguished variables of the language $\mathcal{L}$. Let $d \subseteq D$, then the decision $d^\wedge$ is the logical conjunction of literals in the chosen subset. The variables that are not in D are state variables.

Each set of decision d induces a possibility distribution π_{K_d} in the following way [7]:

$$\pi_{K_d}(\omega) = \begin{cases} 1 \text{ if } \forall(\phi_i, \alpha_i) \in K, \omega \models \phi_i \text{ and } \omega \models d^\wedge \\ min_{(\phi_i,\alpha_i)\in K/\omega\models\neg\phi_i}(1 - \alpha_i) \text{ if } \omega \models d^\wedge \\ 0 \text{ if } \omega \not\models d^\wedge \end{cases}$$

Where α_i represents the degrees of necessity of the formulas in the corresponding layers of K_d.

In other hand, the utility function μ_P associated to the preferences base P is built over Ω in a similar way:

$$\mu_P(\omega) = \begin{cases} 1 \text{ if } \forall(\psi_j, \beta_j) \in P, \omega \models \psi_j \\ min_{(\psi_j,\beta_j)\in P/\omega\models\neg\psi_j}(1 - \beta_j) \end{cases}$$

where β_j represents a degree of priority of a formulas in P. Making a decision amounts to choosing a subset d of the decision set D. The objective is to rank-order decisions on the basis of K and P. The pessimistic utility function is expressed in terms of the possibility distribution π_{K_d} and the utility function μ [3]:

$$u_*(d) = min_{\omega\in\Omega} max(1 - \pi_{K_d}(\omega), \mu(\omega)) \quad (5)$$

In the pessimistic case, the decision d must satisfy [1]:

$$K^\wedge_\alpha \wedge d^\wedge \vdash P^\wedge_{>(1-\alpha)} \quad (6)$$

The decision d associated with the most certain part of K entails the satisfaction of the goals, even those with low priority.

3 Syntactic Counterparts of Negated Preference Bases

This section presents a first result of this paper needed for the development of our algorithm. It consists in a characterization of a syntactic counterpart of negated possibilistic base defined by: $\forall \omega \in \Omega, \pi_{n_\Sigma}(\omega) = 1 - \pi_\Sigma(\omega)$.

Let $\Sigma = \{(\phi_i, \alpha_i) : i = 1, ..n\}$ be a possibilistic base. We assume that: $\alpha_0 = 0 < \alpha_1 < ... < \alpha_n$. The following definition gives the possibilistic knowledge base associated with the negation of Σ.

Definition 1. The negated base of $\Sigma = \{(\phi_i, \alpha_i) : i = 1, ..., n\}$ is a possibilistic base, denoted by n_Σ, and defined by:

$$n_\Sigma = \{(d_i, 1 - \alpha_i) : i = 1, ..., n\} \cup \{(\bot, 1 - \alpha_n)\}$$

where $d_i = \neg\phi_i \vee \neg\phi_{i+1} \vee ... \vee \neg\phi_n$.

Example 1. Let $\Sigma = \{(\neg a \vee b, 0.3), (b \vee \neg c, 0.6)\}$. Then, $n_\Sigma = \{((a \wedge \neg b) \vee (\neg b \wedge c), 0.7), (\neg b \wedge c, 0.4)\}$

The following proposition shows that n_Σ is indeed encodes the negation of Σ:

Proposition 1. Let $\Sigma = \{(\phi_i, \alpha_i) : i = 1, ..., n\}$ be a preference base, and n_Σ its negated base obtained using Definition 1. Let π_Σ and π_{n_Σ} be the utility distributions associated with Σ and n_Σ respectively. Then:

$$\forall \omega \in \Omega, \pi_\Sigma(\omega) = 1 - \pi_{n_\Sigma}(\omega).$$

Proof. Recall first that when $\pi_\Sigma(\omega) = \alpha$ (resp $\pi_{n_\Sigma} = \beta$) then ω falsifies formulas of Σ (resp of n_Σ) having a weight equal to α (resp β), but satisfies all formulas of Σ (resp n_Σ) having a weight strictly larger than α (resp β). We distinguish 3 cases:

ISBN: 978-989-95079-6-8

1. $\pi_\Sigma(\omega) = 1$, which means that $\forall(\phi_i, \alpha_i) \in \Sigma, \omega \models \phi_i$. This means that ω falsifies all formulas of n_Σ, and in particular the highest formula α_1, namely $\omega \not\models \alpha_1$. Hence, $\pi_{n_\Sigma}(\omega) = 0 = 1 - \pi_\Sigma(\omega)$.

2. $\pi_\Sigma(\omega) = \alpha_n$. This means that ω satisfies all formulas in Σ except α_n. This also means that ω satisfies all formulas d_i (since d_i is a disjunction of $\neg\phi_n$ and others terms). The only formula falsified by $\omega \in n_\Sigma$ is $(\bot, 1 - \alpha_n)$. Hence, $\pi_{n_\Sigma}(\omega) = 1 - (1 - \alpha_n) = \alpha_n = \pi_\Sigma(\omega)$.

3. $\pi_\Sigma(\omega) = \alpha_j$ with $j < n$. This means that $\omega \not\models \phi_j$. Hence, $\forall i \leq j, \omega \models \phi_j$ (since d_j contains $\neg\phi_j$). $\mu_P(\omega) = \alpha_j$ also means that $\forall k > j, \omega \models \phi_k$, namely $\omega \models \phi_n \wedge ... \wedge \phi_{j+1}$. Hence, $\omega \not\models \neg\phi_n \vee ... \vee \phi_{j+1}$. Namely, $\omega \not\models d_{j+1}$. Therefore: $\pi_{n_\Sigma}(\omega) = 1 - (1 - \alpha_{(j+1)-1}) = \alpha_j = \pi_\Sigma(\omega)$.

The obtained base n_Σ must be put in a clausal form to get C_{n_Σ}.

Example 2. Let n_Σ be the base obtained in example 1. The clausal form of n_Σ is then: $C_{n_\Sigma} = \{(c, 0.4), (a \vee \neg b, 0.7), (a \vee c, 0.7), (\neg b, 0, 7), (\neg b \vee c, 0.7)\}$

4 Computation of Pessimistic Decision

In this section, we propose an algorithm for computing qualitative optimal decision in the pessimistic case. The knowledge base K can be the result of merging several knowledge bases $K_1, K_2, ..., K_n$. The possibilistic distribution associated to K is then obtained by merging possibility distributions associated with $K_1, , K_2, ..., K_n$ using some merging operator /oplus, namely: $\pi_K = \oplus(\pi_{K1}, ..., \pi_{Kn})$. When $\oplus = min$, it has been shown that $K = K_1 \cup ...K_n$. If $\oplus = max$, then $K = \{(\phi_1 \vee ...\phi_n, min(\alpha_1, ..., \alpha_n)) : (\phi_1, \alpha_1) \in K_1, ..., (\phi_n, \alpha_n) \in K_n\}$. For syntactic counterparts of more general operator, see [1]. Similarly, P can be also the result of merging several preference bases.
Now, once K and P are fixed, we propose a syntactic computation of optimal decisions. We recall that an optimal pessimistic decision d maximizing $u_*(d)$ is such that:

$$u_*(d) = min_{\omega\in\Omega} max(1 - \pi_{K_d}(\omega), \mu(\omega)) \quad (7)$$

From equation (6), a first way to syntactically compute optimal decisions is first to compute the counterpart of $1 - \pi_{K_d}$, then compute the counterpart of $max(1 - \pi_{K_d}(\omega), \mu(\omega))$ and lastly compute $u_*(d)$. This approach is not interesting since first in general Σ is more important than P and computation of $max(1 - \pi_{Kd}(\omega), \mu(\omega))$ is more expensive than the computation of the minimum. We propose to explore another possibility by noting that equation (6) is equivalent to:

$$u_*(d) = 1 - max_{\omega\in\Omega} min(\pi_{K_d}(\omega), 1 - \mu(\omega)) \quad (8)$$

Besides, the syntactic counterpart of $min(\pi_1(\omega), \pi_2(\omega))$ is the possibilistic base $\Sigma_{min} = \Sigma_1 \cup \Sigma_2$ [1].
Thus, combining these results, the corresponding base Σ_{min} associated to $min(\pi_{K_d}(\omega), 1 - \mu(\omega))$ is the possibilistic base $K \cup n_P \cup \{(d, 1)\}$, such that n_P is the possibilistic base corresponding to the utility function $1 - \mu(\omega)$. We recall that:

$$u_*(d) = 1 - max_{\omega\in\Omega}\{min(\pi_{K_{d(\omega)}}, 1 - \mu(\omega))\} \quad (9)$$

Lemma 1. Let n_P be the negated based, obtained using Definition 1, of the preferences base P . Then, the syntactic counterparts of $min(\pi_{K_d(\omega)}, 1 - \mu(\omega))$ is $K \cup n_P$.

On the other hand, the inconsistency degree $Inc(K \cup C_{n_P})$ of a possibilistic base $K \cup C_{n_P}$, where C_{n_P} represents the conjunctive form of the base n_P, is defined as follow [7]:

$$Inc(K \cup C_{n_P}) = 1 - max\{\pi_{K_d}(\omega)\}$$

Lemma 2. The pessimistic utility function associated to decision d is :

$$u_*(d) = Inc(\Sigma_{min})$$

Where $Inc(\Sigma_{min})$ represents the inconsistency degree of the base $K \cup C_{n_P} \cup \{(d, 1)\}$.
Then, the computation of optimal pessimistic decisions is obtained using the Algorithm 1.
The computation of inconsistency degree is performed by a call to the function $incons(B \cup \{(\neg\phi, 1)\}, Inc, bool)$. This function, given by the Algorithm 2, has three parameters: a stratified knowledge base, an integer representing current inconsistency degree and a boolean variable.
The base $B^*_{\geq\alpha_r}$ is defined as the classical projection of the $\alpha\_cut\ B_{\geq\alpha_r}$. $B^*_{\geq\alpha_r} = \{\phi : (\phi, \beta) \in B, \beta \geq \alpha_r\}$.
The function incons is on the extension of dichotomous algorithm for computing inconsistency proposed in [11] on which some adaptations have been made in order to compute optimal decision. Indeed, one the inconsistency degree of $K \cup C_{n_P} \cup \{(d_i, 1)\}$ can be greater than the inconsistency degree of $K \cup C_{n_P} \cup \{(d_{j\in[1,i-1]}, 1)\}$, the algorithm stops.
In terms of complexity, the proposed algorithm is based on

Algorithm 1 Computation of optimal pessimistic decision

```
Input: K knowledge base
       C_{n_P} conjunctive form of n_p
       N number of decision variables
       D = {d_{i∈[1,n]}} set of decisions
Output: decision
Begin
i ← 1
max ← 0
inc ← 1
for i = 1 to N do
  incons(K ∪ C_{n_P} ∪ {(d_i, 1)}, inc, bool)
  if (bool=true) then
    if (inc > max) then
      max ← inc
      decision ← {d_i}
    else if inc = max then
      decision ← decision ∪ {d_i}
    end if
  end if
end for
RETURN (decision)
End
```

ISBN: 978-989-95079-6-8

an inconsistency degree computation which is known to be NP-hard and requires in the worst case $[log_2 m]$ satisfiability checks, where m is the number of different valuations involved in $K \cup C_{n_P} \cup \{(d_i, 1)\}$ using any prouver for the propositional satisfiability problem SAT.

Algorithm 2 Function $incons(B \cup \{(\neg\phi, 1)\}, inc, bool)$

Input B:stratified base
ϕ:weighted formula
n: number of strates in base B
Output inc: inconsistency degree
bool:boolean
Begin
$l \leftarrow 0$ {initially pointed on the last strate of the base}
$u \leftarrow m$ {initially pointed on first strate of the base}
$bool \leftarrow true$
while $((l < u)$ and $(bool = true))$ **do**
  $r \leftarrow [(l + u)/2]$ {pointer uses for dichotomy}
  if $(B^*_{\geq \alpha_r} \wedge \neg\phi$ consistent) **then**
    $u \leftarrow r - 1$ {check the inconsistency in the most big base}
  else
    if $(inc \geq \alpha_r)$ **then**
      $l \leftarrow r$
    else
      $bool \leftarrow false$
    end if
  end if
end while
if ($\alpha_r \geq inc$) **then**
  $bool \leftarrow false$
end if
if (bool = true) **then**
  $inc \leftarrow \alpha_r$
end if
RETURN(inc,bool)
End

Example 3. Let us exemplify the algorithm on an example initially proposed in [12] and reused in [2]. The example is about taking an umbrella or not, knowing that the sky is cloudy.
The literals of the language are:

- It rains: r
- Being wet: w
- Taking an umbrella: um
- The sky is cloudy: c

The stratified knowledge base is $K = \{(\neg um \vee \neg w, 1), (\neg r \vee um \vee w, 1), (r \vee \neg w, 1), (c, 1), (\neg c \vee r, 0.6)\}$.
The stratified preference base is $P = \{(\neg um, 0.2), (\neg w, 1)\}$. We do not like to take an umbrella, but it is more important to be dry.
The set of decisions is $D = \{um, \neg um\}$, i.e taking an umbrella or not.

The preliminary step consists to compute the conjunctive normal form C_{n_P} of the negated form n_P of the preference base P. Then, the algorithm applies as follow:

- initially, $i \leftarrow 1$, $inc \leftarrow 1$, max $\leftarrow 0$ and $d_1 \leftarrow um$. We have to call to the function $incons(K \cup C_{n_P} \cup \{(um, 1)\}, inc, bool)$. This function return inc=0.1 and bool=true. In this case, $inc > max$, so $decision \leftarrow \{um\}$ and $max \leftarrow inc$.
- In the next step, $i \leftarrow 2$ and $d_2 \leftarrow \neg um$. The call of the function $incons(K \cup C_{nP} \cup \{(\neg um, 1)\}, inc, bool)$ returns $inc = 0.1$ and bool is true. As inc=max, the set of optimal decisions do not change.

Thus, the best pessimistic solution of the umbrella problem is to take an umbrella.

5 Conclusion

The main contribution of this paper is a proposition of a new approach to compute a qualitative possibilistic pessimistic decision problem. This problem is viewed as the one of computing inconsistency degrees of particular bases in the framework of possibilistic logic. The application exploits the syntactic counterparts of data fusion techniques.
Our approach is an alternative to the ATMS-based solution, proposed in [3] and it avoids the use of the ATMS to compute the pessimistic optimal qualitative decision which is known to be not efficient when it deals with an important number of variables. Then, in the pessimistic case, the knowledge base and the preferences base are fused and the decision problem is translated into the computation inconsistency degree of a specific base. This process leads to use a logical technique more adapted than the ATMS one.

References

[1] S. Benferhat, D. Dubois, and H. Prade. From semantic to syntactic approaches to information combination in possibilistic logic. *Aggregation and Fusion of Imperfect Information, Physica Verlag*, pages 141–151, 1997.

[2] D. Dubois and H. Prade. Possibility theory as a basis for qualitative decision theory. In *14th International Joint Conference on Artificial Intelligence (IJCAI'95), Montreal*, pages 1924–1930, 1995.

[3] D. Dubois, D. Le Berre, H. Prade, and R. Sabbadin. Logical representation and computation of optimal decisions in a qualitative setting. In *AAAI-98*, pages 588–593, 1998.

[4] J. De Kleer. An assumption-based truth maintenance systems. *Artificial Intelligence Jounral*, 28:127–162, 1986.

[5] J. De Kleer. Extending the assumptions-based truth maintenance systems. *Artificial Intelligence Journal*, 28:163–196, 1986.

[6] S.Benferhat, F.Khellaf-Haned, A.Mokhtari, and I.Zeddigha. On the use of syntactic possibilistic fusion for computing optimistic qualitative decisions. In *Proceedings of JCIS/FTT Salt Lake City, USA*, 2007.

[7] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. *In Handbook of Logic in Artificial Intelligence and Logic Programming*, D. Gabbay et al., eds, 3, Oxford University Press:pages 439–513, 1994.

ISBN: 978-989-95079-6-8

[8] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

[9] B. Bonet and H. Geffner. Arguing for decisions: a qualitative model of decision making. In *Proceedings of the 12th conference on uncertainty in Artificial Intelligence (UAI'96)*, pages 98–105. Portland, Oregon: Morgan Kaufman, 1996.

[10] R. Brafman and M. Tennenholtz. On the axiomatization of qualitative decision criteria. In *Proceedings of the 14th National Conference on Artificial Intelligence (AAAI'97)*, pages 76–81, 1997.

[11] J. Lang. Possibilistic logic: Complexity and algorithms. *In Handbook of Defeasible Reasoning and Uncertainty Management Systems*, 5:179–220, 2000.

[12] C. Boutilier. Toward a Logic for Qualitative Decision Theory. In *Proceedings of the 4th International Conference on Principles of Knowledge Representation, (KR'94)*, pages 75–86, 1994.

ISBN: 978-989-95079-6-8

A Fuzzy-Semiotic Framework for Modeling Imprecision in the Assessment of Depression

Mila Kwiatkowska[1] Krzysztof Kielan, MD[2] Krzysztof Michalik[3]

1 Department of Computing Science, Thompson Rivers University
Kamloops, British Columbia, Canada
2 Department of Psychiatry, Diana, Princess of Wales Hospital, Grimsby, UK
3 Department of Knowledge Engineering, University of Economics, Katowice, Poland
Email: mkwiatkowska@tru.ca, krystof.kielan@nelctp.nhs.uk, k.michalik@aitech.pl

Abstract—*This paper presents a new framework to model an assessment process for a complex and multidimensional syndrome such as depression. Since the measurements of depression are inherently imprecise, we explicitly model the context of the assessment process, and we analyze various aspects of imprecision (syntactic, semantic, and pragmatic). The framework is based on fuzzy logic and semiotics. The fuzzy-logic approach allows for the representation of quantitative imprecision of the measurements and the semiotic approach allows for the representation of the qualitative imprecision of the concepts. We have applied this fuzzy-semiotic framework to two types of clinical measurements: the rating by the experts and the filling out of self-administered questionnaires. The proposed framework provides a conceptual foundation for the construction of a medical decision support tool.*

Keywords— Assessment of depression, fuzzy logic, imprecision, modelling, medical decision, semiotics.

1 Introduction

Depression is a common mental disorder characterized by sad moods, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, low energy, and poor concentration. Depression represents a major public health problem, and it has a high prevalence worldwide and high societal costs [1]. The World Health Organization (WHO) reports that "depression is affecting about 121 million people worldwide;" furthermore, WHO ranks depression as one of the most burdensome diseases in terms of disability costs [1, 2]. The assessment and treatment of depression are lengthy and complex, and often the patients wait a long time to gain access to psychiatrists. We believe that specialists as well as general practitioners may be assisted by a decision-support tool, which may reduce the waiting time and facilitate earlier medical assistance. The medical decision support tool is not intended to make therapeutic and diagnostic decisions, and it should never replace proper psychiatric examination or harm the crucial relationship and trust between a doctor and a patient. The effectiveness of such a support system depends profoundly on the underlying models of the measurement process. Thus, we present first a framework based on fuzzy-logic and semiotic approaches to model the complex and multi-dimensional concept of depression and its measurement methods. Second, we then use this framework to model two general types of measurements: (1) rating by an expert using widely accepted rating scales such as the Hamilton Rating Scale for Depression (HRSD) and (2) self-administered questionnaires filled out by patients such as the Beck Depression Inventory (BDI).

This paper is structured as follows. Section 2 surveys various classifications of depression and methods used to measure the severity of depression. Section 3 presents the definition of imprecision in medical data and defines three characteristics. Section 4 presents a fuzzy-semiotic framework for explicit representation of imprecision in medical data. Section 5 and 6 describe the implementation of the framework for the assessment of depression in the context of treatment evaluation and in the context of screening. Finally, the last section presents conclusions and suggests directions for future research.

2 Depression and its measurement process

Depression is a term used to cover a wide range of states, from feeling sad or helpless, through minor depression to major depression (MD). There are many approaches to the definitions, classifications, diagnostic criteria, and measurements of depression. These diverse approaches reflect the fact that depression has a complex etiology and presents itself with a variety of symptoms, which differ in different patients. In this paper, we focus on modeling the measurement process; thus we concentrate on a classification of symptoms, their frequency, and duration. However, the measurement process is determined by the definition of depression (its conceptualization). Thus, we cannot model the measurement process without specifying first the various conceptualization approaches.

2.1 Conceptualization of Depression

Several conceptual approaches to depression exist, and many authors view depression as a syndrome rather than a single diagnostic entity. In this paper, we focus on major characteristics: (1) depression can be defined by a set of presenting symptoms, which display specific severity, frequency, and duration, (2) depression can be viewed as a dimensional concept in which symptoms may be grouped or clustered into specific dimensions, and (3) depression may be conceptualized as a state or trait.

The symptomatic approach to depression identifies more than 10 symptoms, which have varied definitions and which

ISBN: 978-989-95079-6-8

are used in different ways by the diagnostic criteria. The symptoms are generally grouped into three classes: *affective* (crying, sadness, apathy), *cognitive* (thoughts of hopelessness, helplessness, suicide, worthlessness, guilt), and *somatic* (sleep disturbance, changes in energy level, changes in appetite, and elimination). Not all symptoms are present at the same time, and the severity of symptoms differs. Moreover, the symptoms may vary in their "directions." For example, two subtypes of depression are distinguished: depression with vegetative symptoms (e.g., appetite loss, weight loss, insomnia) and depression with reverse vegetative symptoms (e.g., appetite increase, weight gain, hypersomnia). The second subtype, according to the epidemiological studies, is characteristic of one-fourth to one-third or of all people with major depression, and it is more common among women.

2.2 Operationalization of Depression

In our discussion, we refer to two general diagnostic criteria: WHO's International Classification of Disease (ICD-10 in Great Britain) and the Diagnostic and Statistical Manual of Mental Diseases Fourth Edition (DSM-IV in North America). Both criteria are based on a symptomatic approach and have many similarities and differences [3]. They differ in the set of symptoms; however, both have eight symptoms in common: depressed mood, loss of interest, decrease in energy or increased fatigue, sleep disturbance, appetite disturbance, recurrent thoughts of death, inability to concentrate or indecisiveness, psychomotor agitation or retardation. The ICD-10 and DSM-IV criteria have a significant overlap in diagnosis, yet in some cases the patient may meet the diagnostic criteria in one system but not in the other. For example, the ICD-10 criteria do not take into account bereavement, while DSM-IV excludes a diagnosis of major depression if the symptoms may be linked with the bereavement process. Both diagnostic criteria have been modified several times and remain subjects of ongoing discussions. These modifications clearly indicate the difficulties in defining such a heterogeneous syndrome.

2.3 Assessment of Depression

Assessment of depression must be placed in a broader context of clinical decision making. Thus, for example, Nezu et al. [4] define the following questions to guide the selection and interpretation of the appropriate measurement process: (1) What are the goals of assessment? (2) Who is to be assessed? (3) What is the value of a given measure and who is the source of the information? The goals, for example, are screening, diagnosis and classification, clinical hypothesis testing, treatment planning, prediction of behavior, or outcome evaluation. The assessed groups are heterogeneous in terms of age, comorbidity, cultural background, and gender. The measures can be divided along the lines of *idiographic* and *nomothetic* philosophies of measurement. The idiographic approach assumes that each person is unique and thereby requires an individualized method of measurement. Thus, the measuring process requires an assessment by a clinician who follows the general structure of a rating scale, but who can vary the specific questions according to individual needs of the patient. The nomothetic approach has the goal of generalization, and the measurements strictly follow standardized questionnaires. Moreover, various measures are characterized by varied reliability, validity, and practical utility.

3 Imprecision

Imprecision is an intrinsic part of many medical concepts and their measurements. Concepts such as *quality of life*, *state of health,* and *depression* are difficult to define, measure, and quantify. Moreover, a certain level of imprecision is characteristic of all medical data. Imprecision is an inherent part of conceptualization, operationalization, and the measurements themselves. Moreover, imprecision is distinct from incompleteness (absence of value), inaccuracy (value is not close to the "true" value), inconsistency (dissimilar values from several sources), and uncertainty (probability or belief that the value is the "right" value. In our discussion, we focus on three characteristics:

1) Imprecision has two aspects: *qualitative* and *quantitative*. The qualitative imprecision is a result of a vagueness of the concept (e.g., quality of life, status of health) and the inability to precisely measure the concept (e.g., inherent imprecision of self-administered depression questionnaires). The quantitative imprecision is a result of a lack of precision in a measurement. We view these aspects of imprecision as pragmatic (vagueness of the concept), semantic (lack of precise measures), and syntactic (lack of precision in a measurement).

2) Imprecision is highly *contextual* and *interpretative*, i.e., a statement "I feel sad some of the time" may be sufficiently precise in a specific situation or a more precise statement such as "I feel sad 2-3 times a week for more than 2 hours without any particular reason" may be needed. Thus, imprecision is a quality of a specific value used in reference to a concept in a specific context. Often, imprecise values are sufficient, since the precision may be impossible to attain, impractical, too expensive, or unneeded.

3) Imprecision should be viewed in *terms of degrees.* Each concept, its representation, and its interpretation have a certain degree of imprecision or precision. Imprecision in a sense has continuous values. The values can be ordered in increasing or decreasing order of precision. For example, the severity of depressive mood can be described in terms of its frequency and duration with an increasing precision: "occasional sadness" "sadness 2-3 times a day" and "sadness 2-3 times a day more than for 2 hours."

4 A fuzzy-semiotic framework

In this section, we present a conceptual framework for modeling imprecision and its qualitative and quantitative aspects. Our framework is based on a complementary combination of the semiotic approach and the fuzzy-logic approach. The semiotic approach provides a model for *qualitative* imprecision, whereas the fuzzy-logic approach

ISBN: 978-989-95079-6-8

provides explicit representation for *quantitative* imprecision. Our framework addresses also the contextual and interpretative characteristics of imprecision.

4.1 Semiotic Approach

We based the framework on Peircean semiotics, which defines "sign" as any entity carrying some information and used in a communication process. Peirce divided semiotics into three categories [5]: syntax (the study of relations between signs), semantics (the study of relations between signs and the referred objects), and pragmatics (the study of relations between the signs and the agents who use the signs to refer to objects in the world). This triadic distinction is represented by Peirce's semiotic triangle: the *object*, *representamen* (the form which the sign takes), and *interpretant*. Fig. 1 shows the semiotic representation for the assessment of depression. In this model, the "interpretant" refers to the purpose of the assessment (e.g., screening, diagnosis, treatment evaluation), the agents (e.g., patients, clinicians), the perspectives (e.g., health care costs, accessibility, ethics), the biases (e.g., specific subgroups of agents), and the views (e.g., variations in the diagnostic criteria).

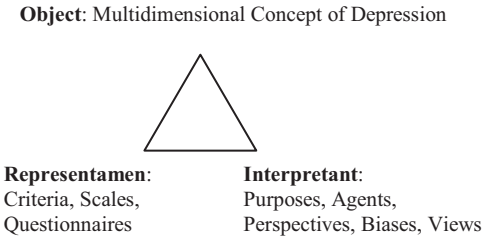


Figure 1: Semiotic triangle.

The assessment involves three aspects: conceptualization (what to measure), operationalization (how to measure), and utilization (how to use the measure). These three aspects constitute a triplet: $<O,M,U>$, where O represents a set of objects, M a set of measures, and U a set of utilization parameters (interpretants). For example, the representation of depression has a set of objects such as depressive episode, cognitive symptoms, affective symptoms, somatic symptoms, a set of diagnostic criteria and assessment instruments, and a set of clinical guidelines for the assessment process in a particular context.

4.2 Fuzzy- Logic Approach

We based the framework on a fuzzy-logic approach (1) to assess the severity of depressive episode as "absent," "mild," "moderate," and "severe" and (2) to assess the severity of particular dimensions of depression. Since the assessment process depends on the clinical context, the accepted definition of depression, and the specific diagnostic criteria, we apply the fuzzy-logic framework to the treatment evaluation and screening.

5 Modeling the assessment of depression in the context of treatment evaluation

In this section, we describe the treatment evaluation process, which typically is used in psychiatric clinics to assess clinical situations and to evaluate the effectiveness of pharmaceutical treatment. We have applied the fuzzy-semiotic framework to model the assessment of depression in the context of treatment evaluation. The evaluation protocol is based on our earlier work on a fuzzy-logic based system to support a depressive episode therapy [6]. In the following subsections, we discuss diagnostic criteria, an assessment instrument, a measure for treatment effectiveness, and an implementation of the fuzzy-semiotic framework.

5.1 Diagnostic Criteria

The treatment evaluation requires at least two consultations: pre- and post-treatment. During the first consultation, the clinician evaluates the patients according to diagnostic criteria accepted by the clinic, such as DSM-IV or ICD-10. In this specific application, the clinician uses the Research Diagnostic Criteria (RDC) for ICD-10 to assess the severity of a depressive episode as *mild*, *moderate*, or *severe*, and to classify it as an episode with or without psychotic symptoms and with or without somatic symptoms. During the second consultation, the clinician repeats the assessment, compares the results with the results from the first consultation, and evaluates the clinical situation as a *recovery*, *partial improvement*, *lack of improvement*, and *deterioration*. The RDC for ICD-10 have 16 items, which are rated by the clinician using the Hamilton Rating Scale for Depression (HRSD) and eight additional questions for the items not covered by HRSD.

5.2 Assessment Instrument

The Hamilton Rating Scale for Depression (HRDS) is an assessment instrument, which has been the most frequently used clinical rating scale since its inception in 1960 [7]. The HRDS is completed by a clinician, and it used to indicate the severity of depression in patients already diagnosed with a depressive disorder. The HRDS has 21 items; 17 items are usually used for scoring. The items are measured on a three–point scale (0, 1, 2) or a five-point scale (0, 1, 2, 3, 4). The items are based on symptoms. The items on the three-point scale are quantified as 0 = "symptom absent," 1 = "slight or doubtful," and 2 = "clearly present." The items on the five-point scale are quantified in terms of increasing intensity: 0 = "symptom absent," 1 = "doubtful or trivial," 2 = "mild," 3 = "moderate," and 4 = "severe." For example, the symptom *depressed mood* is quantified as 0 = "absent," 1 = "feeling of sadness, gloomy attitude, pessimism about future," 2 = "occasional weeping," 3 = "frequent weeping," 4 = "extreme symptoms." Typically, the scores from 17 HRDS's items are added together, and the final score ranges from 0 to 52 points. Originally, Hamilton did not specify cutoff points; however, generally the scores lower than 7 indicate an absence of depression, scores 7 to 17 indicate mild depression, scores 18 to 24 indicate moderate depression, and scores of 25 and above indicate severe depression. Generally, the HRDS has high reliability and validity. On the other hand, the literature on the HRSD includes many papers criticizing the scale on a number of grounds. One of the important issues for our explicit model of imprecision is a critical assumption about the type of a measurement. The problem is related to the quantification of the concept: is the

ISBN: 978-989-95079-6-8

HRDS an *ordinal measurement* scale or is it an *interval measurement* scale? Thus, although from a theoretical perspective, the HRDS is an ordinal scale, from the practical perspective, the HRDS is perceived as an interval measurement, which means that one unit on the scale represents the same magnitude of change across the entire scale. For example, the symptom *depressed mood* rated from 0 to 4 should have equal distances between the four points. The distance of two units between "absence of depressed moods" (rated as 0) and "occasional weeping" (rated as 2) should be the same as the distance between "occasional weeping" (rated as 2) and "extreme symptoms, when the patient reports only sad feelings in verbal and non-verbal communication" (rated as 4).

We have observed that (1) the HRDS represents an ordinal scale, and the assumption of the equity of distances between units introduces a large measurement imprecision, and (2) the range of values on three-point scale and five-point scale is not sufficient for scaling small changes in symptoms during the treatment. To address these two problems, we have introduced an 11-point rating scale – a range of values from a discrete set: {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}. The value 0 means "normal state" (not precisely defined), and the value 1 means extreme pathology (also not specified), the medium values are based on the HRDS range of scores 0–4 (1 = 0.2, 2 = 0.4, 3 = 0.6, and 4 = 0.8). Each item has the same rating range. The 16 items on the RDC for ICD-10 and an example of values for patient p_1 are presented in Fig. 2.

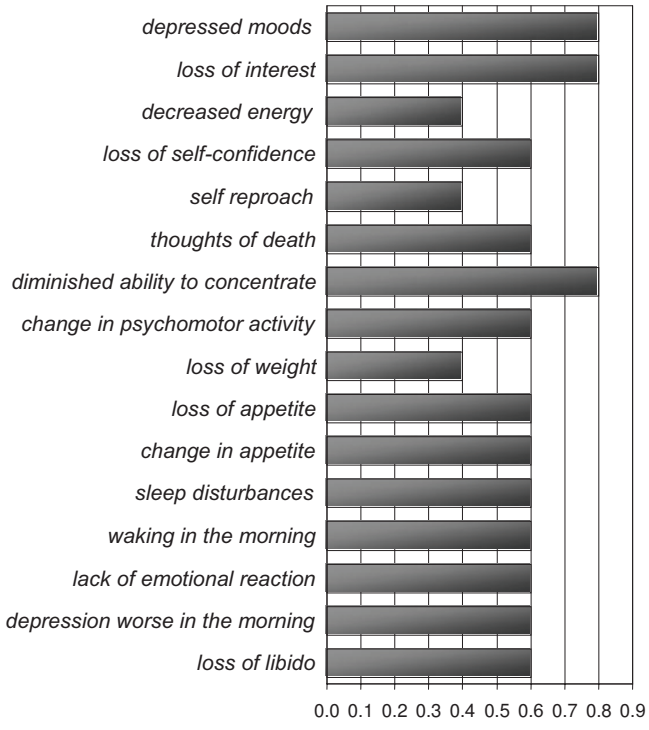


Figure 2: Assessment of depressive episode for patient p_1 using the RDC for ICD-10.

5.3 Measure of the Treatment Effectiveness

The treatment evaluation is based on the comparison between the pre-treatment and post-treatment assessment outcomes. In the assessment process, a clinician (a rater) evaluates the intensity of depression for each of the 16 items, and then calculates the relative intensity of depression in percent. The *Relative Depression Intensity*, *RDI,* is calculated for a specific patient, p_1, rated by a specific rater, r_1, at given point in time, t_1, using (1). The RDC criteria for ICD-10 are represented as C_1,..., C_{16} (n = 16). The maximum value for each C_i is 1.

$$\mathrm{RDI}\ (\mathrm{p}_1, \mathrm{r}_1, \mathrm{t}_1) = \frac{\sum_{i=1}^{n} C_i(p_1, r_1, t_1)}{\sum_{i=1}^{n} \mathrm{C}_\mathrm{i}^{\max}} 100\% \qquad (1)$$

The RDI values range from 0% to 100% and are interpreted using linguistic descriptors as follows: "absent" corresponds to interval <0, 10), "mild" corresponds to <10, 40), "moderate" corresponds to <40, 70), and "severe" corresponds to <70, 100>.

The efficiency of therapy is measured using the *Therapy Effectiveness Index*, *TEI*. The TEI is based on the RDI assessed during the initial consultation at time t_1, and the RDI assessed during a later consultation at time t_2, where $t_1 < t_2$, using (2).

$$\mathrm{TEI} = \frac{RDI(p_1, r_1, t_1) - RDI(p_1, r_1, t_2)}{RDI(p_1, r_1, t_1)} 100\% \qquad (2)$$

The values of the TEI relate to four clinical situations: "no improvement," TEI = 0; "recovery," TEI = 1, "partial improvement," 0 < TEI < 1; and "deterioration," TEI < 0.

5.4 Fuzzy-Semiotic Model for the Treatment Evaluation

From the semiotic perspective, the assessment process involves a representation and an interpretation of the assessment outcomes. The outcomes of the assessment are interpreted with reference to specific patients, raters (clinicians), and specific timing. Thus, we conceptually model the assessment process as a tuple shown in (3).

$$\mathrm{Assessment} = < P, R, \{(O,V)\}_{i=1}^{n}, T > \qquad (3)$$

Where P represents a finite set of references to the patients, R represents a finite set of references to the raters, O represents a finite set of outcomes, V represents a finite set of values for the outcomes, and T represents a finite set of time points. The outcomes, O, consist of the criteria, C, and the measurements, M. The measurements can be *quantitative* and *qualitative*.

The quantitative measurements are absolute scores (calculated as the sum of all items from a diagnostic scale) and relative scores or indexes (calculated as a ratio between the patient's score and the possible maximum score value). For example, the relative intensity of depression, the RDI (shown earlier in Fig. 2) is 60 %.

The qualitative measurements are linguistic descriptions such as "symptoms do not meet criteria for any depressive episode," "mild," "moderate," and "severe." Additionally, the linguistic descriptions may include a phrase concerning somatic and psychotic symptoms. Thus, a qualitative

ISBN: 978-989-95079-6-8

outcome may be specified, for example as "severe with exclusion of psychotic symptoms."

Using our framework, an assessment of patient, p_1, by a clinician, r_1, at time, t_1, with outcomes o_1 and o_2 is represented as (p_1, r_1, (o_1, 60) (o_2, "moderate"), t_1).

We present the fuzzy-logic aspect of the framework based on the standardized rating guidelines for the administration of the HRDS, GRID-HAMD-17 [8]. The GRID-HAMD-17 overcomes several shortfalls in the original HRDS, particularly, the high level of imprecision in measures of frequency and intensity. The GRID-HAMD-17 provides specific instructions for the evaluation of the *frequency* and *intensity* for the 12 items and *intensity* only for 5 items (frequency is not applicable for these items). The frequency is represented by four linguistic terms: "absent or clinically insignificant," "occasional," "much of the time," and "almost all of the time." The GRID-HAMD-17 guidelines specify the mapping between the linguistic terms and the frequency of symptoms measured in days/week. The mapping is defined as follows: "absent" = not occurring, "occasional" = less than 3 day/week, "much of the time" = 3–5 days/week, "almost all the time" = 6–7 days/week. The same definition of frequency is used for all applicable items. The intensity is represented by five linguistic terms: "absent," "mild," "moderate," "severe," and "very severe." The terms have specific qualitative mappings for each item.

We present the fuzzy-logic approach, using an example of the first item on the GRID-HAMD-17, "depressed moods." We have constructed two linguistic variables: *Frequency* and *Intensity of Depressed Moods*. The variable *Frequency* has four terms: *absent*, *occasional*, *much of the time*, *almost all the time*. The membership functions, MFs, for the symptom frequency are shown in Fig. 3. They have been created based on the frequency measured by days per week.

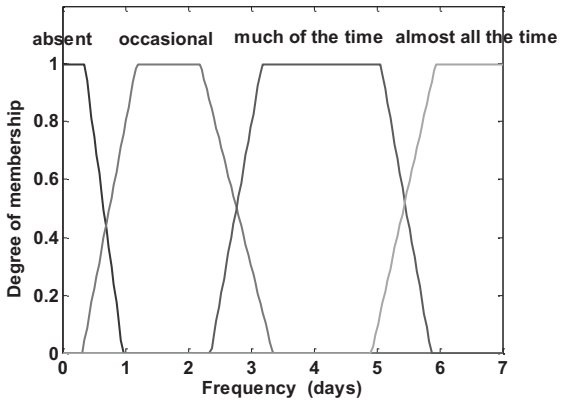


Figure 3: MFs for the frequency of a symptom.

The MFs for the intensity of a depressed mood are shown in Fig. 4. The MFs are based on a continuous scale from 0 to 1, which corresponds to the clinician's rates on a discrete scale from 0 to 1 with an increment of 0.1. The intensity of the symptom is rated by the clinician based on the GRID-HAMD-17 specification. For example, the depressed mood with a "severe" intensity is described as "intense sadness; hopelessness about most aspects of life, feeling of complete helplessness or worthlessness."

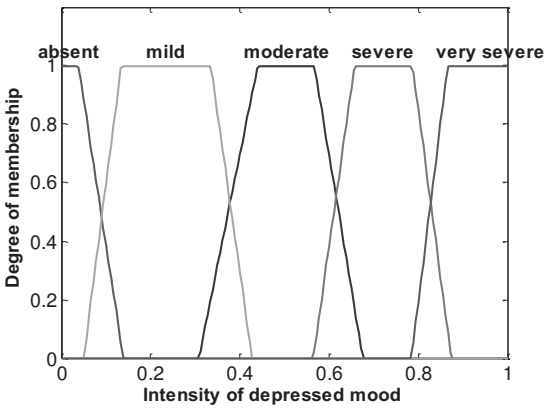


Figure 4: MFs for the depressed mood intensity.

6 Modeling the assessment of depression in the context of screening

In this section, we describe the screening process using the Beck Depression Inventory (BDI). The BDI was originally developed by Beck in 1961 [9], and since then, it has been used to measure the intensity of depression in psychiatric patients, to screen for depression in family practice, and to screen for depression in research studies. BDI has 21 items evaluating 21 symptoms of depression. Each item is rated on a four-point intensity scale (0 to 3). The currently used BDI version refers to the last 7 days and is typically self-administered [10]. The scores are added together and given a total score between 0 and 63. The common guidelines for the interpretation of scores specify that scores less than 10 indicate "absent" depression, 10 to 18 indicate "mild to moderate", 19 to 29 indicate "moderate to severe," and scores of 30 and above indicate "severe" depression.

Using our framework, an assessment of patient, p_1, by self-reporting, $r_1 = p_1$ at time, t_1, with outcomes o_1 and o_2 is represented as (p_1, r_1, (o_1, 35) (o_2, "severe"), t_1). Using the fuzzy-logic approach, we have constructed a linguistic variable *Severity of Depression* with four terms: absent, mild, moderate, and severe. The MFs for the severity of depression are shown in Fig. 5.

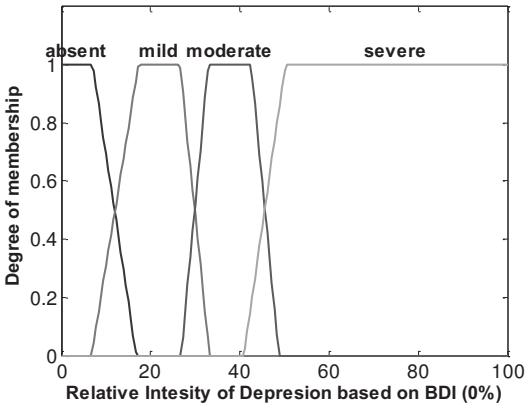


Figure 5: MFs for the severity of depression based on BDI.

We have observed that the assessment of depression is highly contextual and depends on (1) the conceptual basis of the measures, (2) the use of specific guidelines or approaches to the measuring instruments, and (3) the interpretation of the

ISBN: 978-989-95079-6-8

outcomes. To demonstrate the inherent imprecision of interpretation, we have compared the guidelines for the RDC for ICD-10 (described in Section 5), HAMD-17, and BDI. We have converted the HAMD-17 scale and BDI to indexes, by dividing the HAMD-17 score by its maximum score of 52, and dividing the BDI score by its maximum score of 63. Table 1 shows the mappings between qualitative descriptors and corresponding ranges of scores for the RDC-based evaluation, HAMD-17, and BDI. We have assumed that the scales are interval measurements, and they measure intensity of similar conceptual constructs of depression. Interestingly, we have noticed that the patient p_1, scoring 60% = "moderate" on RDI, will be classified as "severe" on HAMD-17 and BDI.

Table 1: Depression severity for three assessment measures.

Severity	RDC ICD-10(%)	Indexed HAMD-17 %	Indexed BDI (%)
Absent	< 10	< 14	< 16
Mild	10 -39.9	14 – 33.9	16 – 29.9
Moderate	40 -69.9	34 – 46.9	30 – 46.9
Severe	> 69.99	> 46.9	> 46.9

7 Conclusions

In this paper, we have examined three essential characteristics of imprecision: (1) qualitative and quantitative aspects, (2) contextual and interpretative nature, and (3) graduality of imprecision. We have demonstrated that imprecision is an intrinsic part of the assessment of such a complex syndrome as depression. Furthermore, we have presented a fuzzy-semiotic framework for the explicit representation of qualitative, quantitative, contextual, and interpretative aspects of imprecision. We have used a semiotic approach to represent the concept of depression, its symptomatic representations, and the clinical utilization of the measures. Furthermore, we have applied a fuzzy-logic approach to explicitly represent the imprecision of various measures.

We are planning to expand and further formalize the proposed framework and to build a comprehensive data model for the medical concept of "depression" and its assessment in treatment evaluation and screening. We will apply this model in a clinical decision support system for the diagnosis and treatment of depression, as well as in a support system for the treatment of sleep disorders.

Acknowledgment

The work presented in this paper is partially supported by NSERC Grant # 327545-06.

References

[1] R.C. Kessler, Epidemiology of depression [In] I.H. Gotlib and C.L. Hammen (Eds) *Handbook of depression*, New York: Guilford Press, 2002.

[2] World Health Organization, Mental Health: Depression http://www.who.int/mental_health/management/depression/definition/en, accessed on January 19, 2009.

[3] A.M. Gruenberg, R.D. Goldstein, and H.A. Pincus, Classification of Depression: Research and Diagnostic Criteria: DSM-IV and ICD-10 [In] *Biology of Depression. From Novel Insights to Therapeutic Strategies*. J. Licincio, Ma-Li Wong (Eds), WIILEY-VCH, Weinheim, 2005.

[4] A.M. Nezu, C.M. Nezu, K.S. McClure et al., Assessment of Depression [In] I.H. Gotlib and C.L. Hammen (Eds) *Handbook of depression*, New York: Guilford Press, 2002.

[5] T.A. Sebeok, *Signs: An introduction to semiotics,* University of Toronto Press, 1999.

[6] K. Kielan, The Salomon advisory system supports a depressive episode therapy. *Polish Journal of Pathology*, 54, 3, 215-218, 2003.

[7] M. Hamilton, A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23 (1), 56-61, 1960.

[8] A. Kalali, J.B.W. Williams, K.A. Kobak, et al., The new GRID HAM-D: pilot testing and international field trials. *International Journal of Neuropsychopharmacology*, 5, S147–S148, 2002.

[9] A.T. Beck, C.H Ward, M. Mendelson, et al., An Inventory for measuring depression, *Archives of General Psychiatry*, 4, 561-571, 1961.

[10] I. McDowell and C. Newell, *Measuring Health: A Guide to Rating Scales and Questionnaires*, New York: Oxford University Press, 1996.

An Axiomatical Approach to *T*–Multi-indistinguishabilities

Javier Martín Gaspar Mayor

Department of Mathematics and Computer Sciences
University of the Balearic Islands
Palma (Mallorca), Spain
Email: javier.martin@uib.es, gmayor@uib.es

Abstract— *Axiomatic definitions for T–multi-indistinguishabilities and multidistances are introduced, with the objective to measure the degree of similarity not only between two elements but also between all of the elements of a finite list. The consistency of these definitions with respect to classic T–indistinguishabilities and distances allows us to translate the close relationship existing between these two concepts.*

Keywords— additive generator, distance, multidistance, pseudo-distance, T–indistinguishability, T–multi-indistinguishability, t-norm.

1 Introduction

Fuzzy equivalence relations, i.e. reflexive, symmetric and Min–transitive relations, were introduced in 1971 by L.A. Zadeh [1] with the name of similarities. Later, the Minimum used in their definition was replaced by a t–norm T [2] and they were called T–indistinguishabilities. Since then, fuzzy equivalences have received many other denominations, depending on the author, the context or sometimes on the particular t-norm used to define them: likeness relations, fuzzy equalities, T–equivalence relations (the one preferred in [3]), proximity relations, etc. Here we will use the term T–indistinguishability.

T–indistinguishabilities are then the fuzzy concept corresponding to classic equivalence relation, and they have become from the first beginning as a useful tool in situations with a certain degree of uncertainty where the transitivity of the equivalence relations is too strong. They have been widely used in many applications: Fuzzy Control, Cluster Analysis, Pattern Recognition, Approximate Reasoning, etc. They are also an important branch of Fuzzy Sets Theory and Fuzzy Logic.

For a given set, T–indistinguishabilities measure the degree of similarity between two of its elements. In this paper we propose an extension of this concept in order to be used to measure the similarity between all of the elements of a finite list. This extension to multiple arguments will be called T–multi-indistinguishabilitiy, and we will deal also with its counterpart in the field of metric spaces: multidistance.

The paper is organized as follows. First, in Section 2, well known definitions and properties of t–norms are recalled. In Sections 3 and 4 axiomatical definitions for T–multi-indistinguishabilities and related pseudo-multidistances and multidistances are proposed, respectively. After, the close relationship existing between T–indistinguishabilities and pseudodistances, by means both of an additive generator of the t–conorm and strong negations, is recalled and then translated to our multidimensional definition in Section 5. Finally, an example of how to construct T–multi-indistinguishabilities from multidistances by using this relationship is detailed in Section 6.

2 Definitions and basic properties of triangular norms

Triangular norms are involved in the definition of T–indistinguishabilities, and so a quick overview on definitions, examples and properties is convenient (see [3] for a more complete study).

Definition 1 *A* triangular norm*, or a* t–norm *in short, is a binary operation $T\colon [0,1]^2 \to [0,1]$ satisfying the following properties, for all $x, y, z \in [0,1]$:*

(i) commutativity: $T(x,y) = T(y,x)$,

(ii) associativity: $T(x, T(y,z)) = T(T(x,y), z)$,

(iii) monotonicity: $y \leqslant z \Rightarrow T(x,y) \leqslant T(x,z)$,

(iv) neutral element: $T(x,1) = x$.

A t-conorm *is a binary operation satisfying (i), (ii),(iii), with* 0 *as a neutral element.*

Definition 2 *A triangular norm is called* Archimedean *if for each $(x,y) \in (0,1)^2$ there exists an $n \in \mathbb{N}$ such that $T(\overbrace{x, \ldots, x}^{n}) < y$. Analogous definition for Archimedean t–conorm.*

Basic t-norms are:

- minimum: $T_M(x,y) = \min\{x,y\}$,
- product: $T_P(x,y) = xy$,
- Łukasiewicz t-norm:

$$T_L(x,y) = \max\{x+y-1, 0\},$$

- drastic t-norm:

$$T_D(x,y) = \begin{cases} x & \text{if } y = 1, \\ y & \text{if } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

In this work we will deal with continuous Arquimedean t–norms, whose characterization is given in the following result.

ISBN: 978-989-95079-6-8

Proposition 1 *A t–norm T is a continuous Archimedean t–norm if and only if it has a continuous additive generator, i.e., there exists a continuous, strictly decreasing function $t\colon [0,1] \to [0,\infty]$ with $t(1) = 0$, which is uniquely determined up to a positive multiplicative constant, such that*

$$T(x,y) = t^{(-1)}(t(x) + t(y)),$$

for all $(x,y) \in [0,1]^2$.

Remember that the *pseudo-inverse* $t^{(-1)}$ of an additive generator t of T is defined as:

$$t^{(-1)}(x) = \sup\{z \in [0,1] : t(z) > x\}.$$

3 *T*–indistinguishabilities and *T*–multi-indistinguishabilities

Let $\mathbb{X}$ be a non empty set. T–indistinguishabilities are defined as follows (see for example [3]).

Definition 3 *Let T be a t–norm. We say that a function $E\colon \mathbb{X}^2 \to [0,1]$ is a* T–indistinguishability *on $\mathbb{X}$ if and only if for all $x_1, x_2, y \in \mathbb{X}$:*

(Ti1) $E(x_1,x_2) = 1$ *when* $x_1 = x_2$,

(Ti2) $E(x_1,x_2) = E(x_2,x_1)$,

(Ti3) $E(x_1,x_2) \geqslant T(E(x_1,y),E(x_2,y))$.

And finally, a T–indistinguishability E separates points *if, in addition,*

(Ti4) $E(x_1,x_2) = 1 \Rightarrow x_1 = x_2$.

Special cases are the *similarity relations* of Zadeh , the *equality relations* of Menger and the *likeness relations* of Ruspini, corresponding to the minimum, product and Łukasiewicz t-norms, respectively.

The elements (lists) of the cartesian product $\mathbb{X}^n$ will be denoted by $(x_1,\ldots,x_n)$, or $\mathbf{x}$.

A *partition* of $(x_1,\ldots,x_n)$ is a set of lists $\{\mathbf{x_i} = (x_{i_1},\ldots,x_{i_{n_i}}) \in \mathbb{X}^{n_i}; i = 1,\ldots,k\}$ such that the sets of subindexes $\{\{i_1,\ldots,i_{n_i}\}, i = 1,\ldots,k\}$ constitute a partition of $\{1,\ldots,n\}$.

Example 1 *We consider the list $\mathbf{x} = (1,0,0,3,1,1)$ and the partition $\{\{1,4,5\},\{2,3,6\}\}$ of $\{1,2,3,4,5,6\}$. The corresponding partition of the list $\mathbf{x}$ is $\{\mathbf{x_1} = (1,3,1), \mathbf{x_2} = (0,0,1)\}$.*

We introduce now our proposal for an axiomatic definition of T–multi-indistinguishability.

Definition 4 *Let T be a t–norm. We say that a function $\mathcal{E}\colon \bigcup_{n\geqslant 1}\mathbb{X}^n \to [0,1]$ is a* T–multi-indistinguishability *on $\mathbb{X}$ if and only if the following properties hold for all n, for all $x_1,\ldots,x_n,y \in \mathbb{X}$.*

(Tmi1) $\mathcal{E}(x_1,\ldots,x_n) = 1$ *when* $x_1 = \ldots = x_n$,

(Tmi2) $\mathcal{E}(x_1,\ldots,x_n) = \mathcal{E}(x_{\pi(1)},\ldots,x_{\pi(n)})$ *for any permutation π of* $1,\ldots,n$,

(Tmi3) $\mathcal{E}(x_1,\ldots,x_n) \geqslant T(\mathcal{E}(x_1,y),\ldots,\mathcal{E}(x_n,y))$.

We say that $\mathcal{E}$ is a strong *T–multi-indistinguishability if it fulfills (Tmi1), (Tmi2) and:*

(Tmi3') $\mathcal{E}(x_1,\ldots,x_n) \geqslant T(\mathcal{E}(\mathbf{x_1},\mathbf{y}),\ldots,\mathcal{E}(\mathbf{x_k},\mathbf{y}))$ *for any partition $\{\mathbf{x_1},\ldots,\mathbf{x_k}\}$ of $\boldsymbol{x} = (x_1,\ldots,x_n)$, for all $\mathbf{y} \in \bigcup_{n\geqslant 1}\mathbb{X}^n$.*

A T–multi-indistinguishability $\mathcal{E}$ separates points *if, in addition,*

(Tmi4) $\mathcal{E}(x_1,\ldots,x_n) = 1 \Rightarrow x_i = x_j$ *for all* $i,j = 1,\ldots n$.

To conclude this definition, we will say that $\mathcal{E}\colon \bigcup_{n=1}^{k}\mathbb{X}^n \to [0,1]$ is a k–T–multi-indistinguishability *if it satisfies the corresponding properties in $\bigcup_{n=1}^{k}\mathbb{X}^n$. In this sense, classic T–indistinguishabilities are 2–T–multi-indistinguishabilities.*

Remark 1 *We denote by $\mathcal{E}(\mathbf{x},\mathbf{y})$ the T–multi-indistinguishability between the elements of the joint list $\mathbf{x},\mathbf{y}$:*

$$\mathcal{E}(\mathbf{x},\mathbf{y}) = \mathcal{E}(x_1,\ldots,x_n,y_1,\ldots,y_m).$$

However, this does not means that $\mathcal{E}$ is a T–indistinguishability between lists of $\bigcup_{n\geqslant 1}\mathbb{X}^n$ because (Ti1) is not fulfilled. But T–indistinguishabilities on this set of lists can be constructed from $\mathcal{E}$ and a T–indistinguishability E on $[0,1]$ as follows:

$$E_{\mathcal{E}}(\mathbf{x},\mathbf{y}) = E(\mathcal{E}(\mathbf{x}),\mathcal{E}(\mathbf{y})).$$

In this case, $E_{\mathcal{E}}$ does not separate points even if E and $\mathcal{E}$ do, because two different lists $\mathbf{x},\mathbf{y}$ with the same degree of indistinguishability verify $E_{\mathcal{E}}(\mathbf{x},\mathbf{y}) = 1$.

4 Distances and multidistances

The definition of classic distances and metric spaces [4] goes back to 1906. We recall it here.

Definition 5 *A function $d\colon \mathbb{X}^2 \to [0,\infty]$ is a* pseudo–distance *on a set $\mathbb{X}$ if for all $x_1,x_2,y \in \mathbb{X}$:*

(d1) $d(x_1,x_2) = 0$ *when* $x_1 = x_2$,

(d2) $d(x_1,x_2) = d(x_2,x_1)$,

(d3) $d(x_1,x_2) \leqslant d(x_1,y) + d(x_2,y)$.

If the following property also holds:

(d4) $d(x_1,x_2) = 0 \Rightarrow x_1 = x_2$,

then we say that d is a distance.

The related properties we propose [5] to define pseudo–multidistances and multidistances are the following.

Definition 6 *A function $D\colon \bigcup_{n\geqslant 1}\mathbb{X}^n \to [0,\infty]$ is a* pseudo-multidistance *on a set $\mathbb{X}$ if it fulfills the following properties for all n, for all $x_1,\ldots,x_n,y \in \mathbb{X}$:*

(md1) $D(x_1,\ldots,x_n) = 0$ *when* $x_1 = \ldots = x_n$,

(md2) $D(x_1,\ldots,x_n) = D(x_{\pi(1)},\ldots,x_{\pi(n)})$ *for any permutation π of* $1,\ldots,n$,

ISBN: 978-989-95079-6-8

(md3) $D(x_1,\dots,x_n) \leqslant D(x_1,y)+\dots+D(x_n,y)$.

We say that D is a strong *pseudo–multidistance if it fulfills (md1),(md2) and:*

(md3') $D(x_1,\dots,x_n) \leqslant D(\mathbf{x_1},\mathbf{y})+\dots+D(\mathbf{x_k},\mathbf{y})$ *for any partition* $\{\mathbf{x_1},\dots,\mathbf{x_k}\}$ *of* $\mathbf{x}=(x_1,\dots,x_n)$*, for all* $\mathbf{y}\in\bigcup_{n\geqslant 1}\mathbb{X}^n$.

If in addition the following property holds:

(md4) $D(x_1,\dots,x_n)=0 \Rightarrow x_i=x_j$ *for all* $i,j=1,\dots n$,

we will say that D is a multidistance.

Finally, $D\colon \bigcup_{n=1}^{k}\mathbb{X}^n \to [0,\infty]$ *is a* k–pseudo-multidistance*, or* k–multidistance*, if it satisfies the corresponding properties in* $\bigcup_{n=1}^{k}\mathbb{X}^n$*. In this sense, classic distances are 2–distances.*

Remark 2 *Again (see remark 1 after Definition 4), expressions like $D(\mathbf{x},\mathbf{y})$ in condition (Tmi3') have to be interpreted as:*

$$D(\mathbf{x},\mathbf{y})=D(x_1,\dots,x_n,y_1,\dots,y_m),$$

but D is not an ordinary pseudodistance on $\bigcup_{n\geqslant 1}\mathbb{X}^n$ because (md1) does not hold. However, pseudodistances d_D in this set can be constructed by means of D and a pseudodistance d on $\mathbb{R}$ in this way:

$$d_D(\mathbf{x},\mathbf{y})=d(D(\mathbf{x}),D(\mathbf{y})).$$

This distance function d_D does not fulfill (md4): different lists $\mathbf{x},\mathbf{y}$ such that $D(\mathbf{x})=D(\mathbf{y})$ verify $d_D(\mathbf{x},\mathbf{y})=0$, and then it is not a multidistance.

Example 2 *The drastic distance* $d\colon \mathbb{X}^2\to\{0,1\}$*, given by:*

$$d(x,y)=\begin{cases}1 & \text{if } x\neq y,\\ 0 & \text{if } x=y,\end{cases}$$

can be extended in several ways. For example, with the multidistances $D_1,D_2,D_3\colon \bigcup_{n\geqslant 1}\mathbb{X}^n\to[0,\infty]$ *defined as follows:*

- $D_1(x_1,\dots,x_n)=\begin{cases}0 & \text{if } x_i=x_j\ \forall i,j,\\ 1 & \text{otherwise},\end{cases}$
- $D_2(x_1,\dots,x_n)=|\{x_1,\dots,x_n\}|-1$,
- $D_3(x_1,\dots,x_n)=n-m$*, where m is the number of times that appears the most repeated element of the list.*

It can be proved that D_1,D_2 are strong multidistances. But D_3 is not strong: for instance, if we take the partition $\{(0,0),(1,1)\}$ of $(0,0,1,1)$,

$$D_3(0,0,1,1)=2\not\leqslant D_3(0,0,0)+D_3(1,1,0)=1.$$

Example 3 *For any metric space $(\mathbb{X},d)$, the function D_M defined by:*

$$D_M(x_1,\dots,x_n)=\max\{d(x_i,x_j);\ i<j\},$$

is a strong multidistance on $\mathbb{X}$.

Another interesting examples and some properties of multidistances can be found in [5].

There exist another definitions for multidistances, or multimetrics. For example the one in [6], in the field of Probability and Statistics, which is more general because it allows to measure the distance between functions. Reduced to lists, multidistances there correspond with strong multidistances here. On the other hand, symmetry is not an axiom in [6], and hence it can not be avoided, preventing the definition of *asymmetric* multidistances.

5 *T*–multi-indistinguishabilities and pseudo multidistances

One way to construct T–indistinguishabilities is to start from pseudodistances, because there is a close relationship between these two concepts. Here we recall two important conexions that relate them [7] and, in both cases, we do the translation to the multidimensional case.

5.1 Relationship via an additive generator of the t–norm

The first relationship between T–indistinguishabilities and pseudo-distance is established by means of an additive generator of the t–norm T.

Proposition 2 *Let T be a continuous Archimedean t–norm and t an additive generator of T.*

(i) If $d\colon \mathbb{X}^2\to[0,\infty]$ *is a pseudo-distance on a set $\mathbb{X}$, then the function* $E_d\colon \mathbb{X}^2\to[0,1]$ *defined by* $E_d=t^{(-1)}\circ d$ *is a T–indistinguishability on $\mathbb{X}$. Furthermore, E_d separates points if and only if d is a distance.*

(ii) If $E\colon \mathbb{X}^2\to[0,1]$ *is a T–indistinguishability on $\mathbb{X}$, then the function* $d_E\colon \mathbb{X}^2\to[0,\infty]$ *given by* $d_E=t\circ E$ *is a pseudo-distance; d_E is a distance if and only if E separates points.*

The literal translation of this result to our multidimensional proposal is as follows.

Proposition 3 *Let T be a continuous Archimedean t–norm with continuous additive generator t.*

(i) If $D\colon \bigcup_{n\geqslant 1}\mathbb{X}^n\to[0,\infty]$ *is a pseudo-multidistance on a set $\mathbb{X}$, then the function* $\mathcal{E}_D\colon \bigcup_{n\geqslant 1}\mathbb{X}^n\to[0,1]$ *defined by* $\mathcal{E}_D=t^{(-1)}\circ D$ *is a T–multi-indistinguishability on $\mathbb{X}$. Furthermore, $\mathcal{E}_D$ separates points if and only if D is a multidistance.*

(ii) If $\mathcal{E}\colon \bigcup_{n\geqslant 1}\mathbb{X}^n\to[0,1]$ *is a T–multi-indistinguishability on $\mathbb{X}$, then the function* $D_{\mathcal{E}}\colon \bigcup_{n\geqslant 1}\mathbb{X}^n\to[0,\infty]$ *given by* $D_{\mathcal{E}}=t\circ\mathcal{E}$ *is a pseudo-multidistance; $D_{\mathcal{E}}$ is a multidistance if and only if $\mathcal{E}$ separates points.*

Proof

For (i), condition (Tmi1) is fulfilled:

$$\mathcal{E}_D(x,\dots,x)=t^{(-1)}(D(x,\dots,x))=t^{(-1)}(0)=1,$$

ISBN: 978-989-95079-6-8

(Tmi2) is immediate, due to the symmetry of D, and the inequality (Tmi3) also holds:

$$\begin{aligned}\mathcal{E}_D(x_1,\ldots,x_n) &= t^{(-1)}(D(x_1,\ldots,x_n))\\ &\geqslant t^{(-1)}(\textstyle\sum_{i=1}^n D(x_i,y))\\ &= t^{(-1)}(\textstyle\sum_{i=1}^n t(\mathcal{E}_D(x_i,y)))\\ &= T(\mathcal{E}_D(x_1,y),\ldots,\mathcal{E}_D(x_n,y)).\end{aligned}$$

Finally,

$$\begin{aligned}\mathcal{E}_D(x_1,\ldots,x_n)=1 &\Leftrightarrow t^{(-1)}(D(x_1,\ldots,x_n))=1\\ &\Leftrightarrow D(x_1,\ldots,x_n)=0,\end{aligned}$$

which suffices to verify that conditions (Tmi4) and (md4) are equivalent.

For (ii), the reasoning is the same. From:

$$M_{\mathcal{E}}(x,\ldots,x)=t(\mathcal{E}(x,\ldots,x))=t(1)=0,$$

we have (md1). The symmetry of $\mathcal{E}$ ensures (md2). Condition (md3) is also satisfied:

$$\begin{aligned}M_{\mathcal{E}}(x_1,\ldots,x_n) &= t(\mathcal{E}(x_1,\ldots,x_n))\\ &\leqslant t(T(\mathcal{E}(x_1,y),\ldots,\mathcal{E}(x_n,y))\\ &= \textstyle\sum_{i=1}^n t(\mathcal{E}(x_i,y))\\ &= \textstyle\sum_{i=1}^n M_{\mathcal{E}}(x_i,y),\end{aligned}$$

and (Tmi4) and (md4) are equivalent on account of this:

$$\begin{aligned}D_{\mathcal{E}}(x_1,\ldots,x_n)=0 &\Leftrightarrow t(\mathcal{E}(x_1,\ldots,x_n))=0\\ &\Leftrightarrow \mathcal{E}(x_1,\ldots,x_n)=1.\end{aligned}$$

■

Note that we have used this property of the additive generator: for all n,

$$T(x_1,\ldots,x_n)=t^{(-1)}(t(x_1)+\ldots+t(x_n)).$$

It always holds in the case of continuous t–conorms. But there exist classes of non–continuous t–conorms not fulfilling it, such as non Archimedean t–norms. More details can be found in [8].

5.2 Relationship via strong negations

Strong negations [9] are decreasing functions $n\colon [0,1]\to[0,1]$, such that $n\circ n=\mathrm{Id}_{[0,1]}$.

If T is a t-norm, then the binary operation $T^*\colon [0,1]^2\to[0,1]$ defined by:

$$T^*(x,y)=n\circ T(n(x),n(y)),$$

is a t–conorm, called the *dual* of T with respect to n.

It is said that distance functions (distances, multidistances, etc.) are *normalized* if $\mathrm{Ran}D\subseteq[0,1]$. In this case, the sum in inequalities (d3), (md3) and (md3') in Definitions 5 and 6 can be replaced by a t-conorm S. For example, for a distance d or a multidistance D we have, respectively,

(Sd3) $d(x_1,x_2)\leqslant S(d(x_1,y),d(x_2,y))$,

(Smd3) $D(x_1,\ldots,x_n)\leqslant S(D(x_1,y),\ldots,D(x_n,y))$.

So, d is an *S–distance* [7] and D is an *S–multidistance*.

In the case that S is the Maximum t–conorm, we have *ultradistances* or *multi-ultradistances*.

The next result states the close relationship between T–indistinguishabilities and S–distances. It admits more general formulations but at this point the following suffices.

Proposition 4 *Let T be a t–norm and T^* its dual t–conorm with respect to a strong negation n.*

(i) If E is a T–indistinguishability on a set $\mathbb{X}$, then $d_E=n\circ E$ is a normalized T^*–pseudodistance; d_E is a distance if and only if E separates points.

(ii) If d is a normalized T^*–pseudodistance on $\mathbb{X}$, then $E_d=n\circ d$ is a T–indistinguishability; E_d separates points if and only if d is a distance.

There exists a complete analogy with the multidimensional case.

Proposition 5 *Let T be a continuous t–norm and T^* its dual t–conorm with respect to a strong negation n.*

(i) If $\mathcal{E}$ is a T–multi-indistinguishability on a set $\mathbb{X}$, then $D_{\mathcal{E}}=n\circ\mathcal{E}$ is a normalized T^*–pseudo- multidistance on the same set; $D_{\mathcal{E}}$ is a multidistance if and only if $\mathcal{E}$ separates points.

(ii) If D is a normalized T^*–pseudo-multidistance on $\mathbb{X}$, then $\mathcal{E}_D=n\circ D$ is a T–multi-indistinguishability on $\mathbb{X}$; $\mathcal{E}_D$ separates points if and only if D is a multidistance.

Proof

A straightforward verification is only needed. If D and $\mathcal{E}$ are related by means of a strong negation,

$$\begin{aligned}\mathcal{E}(x,\ldots,x)=1 &\Leftrightarrow n\circ\mathcal{E}(x,\ldots,x)=n(1)\\ &\Leftrightarrow D(x,\ldots,x)=0,\end{aligned}$$

and thus (Tmi1) and (md1) are equivalent. If one of them is symmetric, the other one also is. Also (Tmi3) and (md3) are equivalent: for all $y\in\mathbb{X}$,

$$\begin{aligned}&\mathcal{E}(x_1,\ldots,x_n)\geqslant T(\mathcal{E}(x_1,y),\ldots,\mathcal{E}(x_n,y))\\ &\Leftrightarrow n\circ\mathcal{E}(x_1,\ldots,x_n)\leqslant n\circ T(\mathcal{E}(x_1,y),\ldots,\mathcal{E}(x_n,y))\\ &\Leftrightarrow D(x_1,\ldots,x_n)\leqslant n\circ T(n\circ D(x_1,y),\ldots,n\circ D(x_n,y))\\ &\Leftrightarrow D(x_1,\ldots,x_n)\leqslant T^*(D(x_1,y),\ldots,D(x_n,y)).\end{aligned}$$

Finally, for (Tmi4) and (md4):

$$\begin{aligned}\mathcal{E}(x_1,\ldots,x_n)=1 &\Leftrightarrow n\circ\mathcal{E}(x_1,\ldots,x_n)=1\\ &\Leftrightarrow D(x_1,\ldots,x_n)=0\end{aligned}$$

■

Obviously, if $\mathcal{E}$ is a T–multi-indistinguishability then it is a T'–multi-indistinguishability for all t–norm T' such that $T'\leqslant T$. If D is a normalized S–multidistance then it is an S'–multidistance for all t–conorm S' such that $S'\geqslant S$.

In particular, if E is a Min–multi-indistinguisability (E is a multisimilarity) then it is a T–multi-indistinguishability for all t–norm T. If D is a normalized Max–multidistance (D is an ultramultidistance) then it is an S–multidistance for all t–conorm S.

ISBN: 978-989-95079-6-8

It is interesting to observe that the ways to normalize an ordinary distance d with the formulas $d' = \frac{d}{1+d}$ and $d'' = \min\{1, d\}$ also work for multidistances, as we can see in the following result.

Proposition 6 *If D is a multidistance on $\mathbb{X}$ with values in $[0,\infty]$, then the functions D', D'' defined by*

$$D' = \frac{D}{1+D} \text{ and } D'' = \min\{1, D\}$$

are multidistances on $\mathbb{X}$ with RanD',RanD'' $\subseteq [0,1]$.

Proof
Conditions (md1), (md2) and (md4) are immediate. Let us see that condition (md3) also holds.

For $D' = \frac{D}{1+D} = 1 - \frac{1}{1+D}$ we must prove $D'(\mathbf{x}) \leqslant \sum_{i=1}^n D'(x_i, y)$ for all y, that is,

$$\begin{aligned} 1 - \tfrac{1}{1+D(\mathbf{x})} &\leqslant \textstyle\sum_{i=1}^n (1 - \frac{1}{1+D(x_i,y)} \\ &= n - \textstyle\sum_{i=1}^n \frac{1}{1+D(x_i,y)}. \end{aligned}$$

From $D(\mathbf{x}) \leqslant \sum_{i=1}^n D(x_i, y)$ we have

$$1 - \tfrac{1}{1+D(\mathbf{x})} \leqslant 1 - \tfrac{1}{1+\sum_{i=1}^n D(x_i,y)}.$$

But

$$1 - \tfrac{1}{1+\sum_{i=1}^n D(x_i,y)} \leqslant n - \textstyle\sum_{i=1}^n \frac{1}{1+D(x_i,y)},$$

because (suppose that $D(x_1, y) \leqslant D(x_i, y)$ for all i)

$$\begin{aligned} &\textstyle\sum_{i=1}^n \frac{1}{1+D(x_i,y)} - \frac{1}{1+\sum_{i=1}^n D(x_i,y)} \\ &\quad\leqslant \textstyle\sum_{i=1}^n \frac{1}{1+D(x_i,y)} - \frac{1}{1+D(x_1,y)} \\ &\quad= \textstyle\sum_{i=2}^n \frac{1}{1+D(x_i,y)} \\ &\quad\leqslant n-1. \end{aligned}$$

And for the function $D'' = \min\{D, 1\}$, the inequality $D(\mathbf{x}) \leqslant \sum_{i=1}^n D(x_i, y)$ implies that

$$\min(1, D(\mathbf{x})) \leqslant \min(1, \textstyle\sum_{i=1}^n D(x_1, y)).$$

But

$$\min(1, \textstyle\sum_{i=1}^n D(x_1, y)) \leqslant \sum_{i=1}^n \min(1, D(x_1, y)),$$

that is, $D''(\mathbf{x}) \leqslant \sum_{i=1}^n D''(x_i, y)$ for all $y \in \mathbb{X}$. ∎

6 An example: *averaged* multidistances and related *T*–multi-indistinguishabilities

Let $(\mathbb{X}, d)$ be a metric space. Let us consider the function $G\colon \bigcup_{n\geqslant 1} \mathbb{X}^n \to [0,\infty]$ defined by

$$\begin{cases} G(x_1) = 0, \\ G(x_1, \ldots, x_n) = \sum_{i<j} d(x_i, x_j), \end{cases}$$

that is, the sum of all of the pairwise ordinary distances. It is not a multidistance: it fulfills (md1), (md2) and (md4) but not (md3). For example, in the set $\mathbb{R}^2$ equipped with the Euclidean distance we take the vertexes A, B, C of a equilateral triangle of side of length 1. With these points and taking $y = A$, condition (md3) is not fulfilled:

$$G(A, B, C) = 3 \not\leqslant d(A, A) + d(B, A) + d(C, A) = 2.$$

In order to obtain multidistances based on G, we have to multiply it by a factor λ which depends on the length of the list: $\lambda = \lambda(n) \in \mathbb{R}^+$, $n \geqslant 1$.

Proposition 7 *The function $D_\lambda\colon \bigcup_{n\geqslant 1} \mathbb{X}^n \to [0,\infty]$ defined by*

$$\begin{cases} D_\lambda(x_1) = 0, \\ D_\lambda(x_1, \ldots, x_n) = \lambda(n) \sum_{i<j} d(x_i, x_j), \end{cases}$$

is a multidistance if and only if:

(i) $\lambda(2) = 1$,

(ii) $0 < \lambda(n) \leqslant \frac{1}{n-1}$ *for all* $n > 2$.

Proof
The conditions are necessary: the first is needed to fulfill $D_\lambda|_{\mathbb{X}^2} = d$ and the first inequality of the second condition, $0 < \lambda(n)$, is needed to fulfill $D_\lambda \geqslant$ (md3).

Let us see that the second inequality of (ii), $\lambda(n) \leqslant \frac{1}{n-1}$, is also necessary with a counterexample. Let us suppose that there exists n_0 such that $\lambda(n_0) > \frac{1}{n_0-1}$. We can consider the drastic distance, two different points $x, y \in \mathbb{X}$ and the list $(\overbrace{x, \ldots, x}^{n_0-1}, y) \in \mathbb{X}^{n_0}$. We have:

$$D_\lambda(\overbrace{x, \ldots, x}^{n_0-1}, y) = \lambda(n_0) \cdot (n_0 - 1) > 1,$$

but the sum of distances from the components of $(\overbrace{x, \ldots, x}^{n_0-1}, y)$ to x is 1, and therefore (md3) does not hold.

Conditions (i) and (ii) are also sufficient: condition (i) ensures that $D_\lambda \geqslant 0$. The first inequality of condition (ii) ensures (md1) and condition (md2) is fulfilled from the definition of D_λ.

Finally, let us see that condition (md3) follows from $\lambda(n) \leqslant \frac{1}{n-1}$: for any $y \in \mathbb{X}$,

$$\begin{aligned} D_\lambda(x_1, \ldots, x_n) &= \lambda(n) \textstyle\sum_{i<j} d(x_i, x_j) \\ &\leqslant \lambda(n) \textstyle\sum_{i<j} (d(x_i, y) + d(x_j, y)) \\ &= \lambda(n)(n-1) \textstyle\sum_i d(x_i, y) \\ &\leqslant \textstyle\sum_i d(x_i, y). \end{aligned}$$

∎

These multidistances are in some sense like averaged sums of the pairwise distance values for all pairs of elements of the list. In fact, if $\lambda(n) = \frac{1}{\binom{n}{2}} \leqslant \frac{1}{n-1}$, D_λ is the arithmetic mean of these distances.

Let us find the T_L and T_P–indistinguishabilities corresponding to the multidistance $D = D_\lambda$.

The generator of the Łukasiewicz t-norm is $t(x) = 1 - x$, and its pseudo-inverse is:

$$t^{(-1)} = \begin{cases} 1 - x & \text{if } x \in [0, 1], \\ 0 & \text{if } x > 1, \end{cases}$$

Thus, the T_L–multi-indistinguishability is given by $\mathcal{E}_D = t^{(-1)} \circ D$, that is:

$$\mathcal{E}_D(x_1, \ldots, x_n) = \max\{1 - \lambda(n) \sum_{i<j} d(x_i, x_j), 0\}.$$

For the product t–norm T_P, with additive generator $t = -\ln x$ and pseudo-inverse $t^{(-1)}(x) = e^{-x}$, the T_P–multi-indistinguishability is:

$$\mathcal{E}_D(x_1, \ldots, x_n) = e^{-\lambda(n) \sum_{i<j} d(x_i, x_j)}.$$

ISBN: 978-989-95079-6-8

To finalize, we are going to translate the multidistance D into T–multi-indistinguishabilities through the strong negation $n(x) = 1 - x$. First, the normalizations of D given in Proposition 6:

$$D' = \frac{D}{1+D} = 1 - \frac{1}{1+D},$$

$$D'' = \min\{D, 1\},$$

and then the negations:

$$n \circ D' = \frac{1}{1+D}.$$

$$n \circ D'' = 1 - \min\{D, 1\} = \max\{1 - D, 0\}.$$

Thus, the T–multi-indistinguishabilities are

$$\mathcal{E}'_D(x_1, \ldots, x_n) = \frac{1}{1 + \lambda(n) \sum_{i<j} d(x_i, x_j)},$$

$$\mathcal{E}''_D(x_1, \ldots, x_n) = \max\{1 - \lambda(n) \sum_{i<j} d(x_i, x_j), 0\}.$$

7 Conclusions

The concept of T–indistinguishability relation, which measures the degree of similarity between two elements of a set, has been extended with the aim that it also measures the similarity between all of the elements of a finite list. Thus, we have stated an axiomatical definition for T–multi-indistinguishabilities, and a parallel definition for multidistances and pseudo-multidistances. The close relationship between them has been studied and an example of how to obtain T–multi-indistinguishabilities from multidistances has been explained. But this is only the basis; a deeper development from both theoretical and practical points of view must be done.

Acknowledgment

The authors acknowledge the support of the Govern Balear grant PCTIB2005GC1-07 and the Spanish DGI grant MTM2006-08322.

References

[1] L. A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

[2] E. Trillas and L. Valverde. An inquiry into indistinguishability operators. In H.J. Skala, S. Termini, and E. Trillas, editors, *Aspects of Vagueness*, pages 231–256. Reidel, Dordrecht, 1983.

[3] E.P. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Publishers, The Netherlands, 2000.

[4] M. Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo*, 22:1–74, 1906.

[5] J. Martín and G. Mayor. How separated Palma, Inca and Manacor are? In *Proceedings of AGOP 2009, Fifth International Summer School on Aggregation Operators*.

[6] D.H. Wolpert. Metrics for more than two points at once. In Y. Bar-Yam, editor, *Proceedings of the International Conference on Complex Systems ICCS 2004*. Perseus books, New York, in press.

[7] L. Valverde. On the structure of F–indistinguishability operators. *Fuzzy Sets and Systems*, 17(3):313–328, 1985.

[8] A. Mesiarová-Zemánková. Ranks of additive generators. *Fuzzy Sets and Systems*, in press.

[9] E. Trillas. Sobre funciones de negación en la teoría de conjuntos difusos. *Stochastica*, III(1):47–59, 1979.

ISBN: 978-989-95079-6-8

Mining Linguistic Information for Configuring a Visual Surface Inspection System

D. Martín, D.M. Guinea, M.C. García-Alegre, D. Guinea

Industrial Automation Institute (IAI), Spanish Council for Scientific Research (CSIC)
28500 Arganda del Rey, Madrid, Spain
Email: {dmartin, dmguinea, maria, domingo}@iai.csic.es

Abstract— *The configuration of a surface inspection vision system as a complex task, requires mining associations among attributes due to the variability of the surface and the environment in real-time production process. The surface inspection task has to change to deal with different elements such as, wood, stainless steel or paper inspection and, in the case of stainless steel, with reflectance and thickness variations.*

This work is an approach to mine linguistic information based on the fuzzy transform method, capable of finding associations among features. We are interested in the linguistic form of the observed associations, derived from numerical data, as the configuration task needs such linguistic knowledge. This linguistic knowledge is handled by the dynamic cognitive architecture (ARDIS), here propose, to deal with the system knowledge, represented by means of IF-THEN rules, required to configure a specific inspection system.

Keywords— Cognitive Architecture, Fuzzy Transform, Knowledge-based Vision Systems, Configuration Task, Surface Defects, Visual Inspection.

1 Introduction

The configuration of a surface inspection system requires mining associations among attributes, to avoid the data analyses performed by a human expert, due to the variability of the surface in the real-time production process. The surface inspection task changes, for different materials such as wood, stainless steel or paper and, in the case of stainless steel, the reflectance, thickness or stainless steel type varies.

This work proposes the use of the fuzzy transform method [1,2], capable of finding associations from a set of numerical features. The linguistic expressions of the observed associations, derived from numerical data is an input for the configuration of the dynamic knowledge-based architecture (ARDIS) [3], which uses IF-THEN rules to configure a specific inspection system.

Visual systems for inspecting surface defects have always been present in the laminated materials industry [4]. Nevertheless, these systems still show several drawbacks, such as the vision system reusing, as they are designed for a particular surface inspection application and do not offer the possibility of changing either the objectives or the inspection goals. They are designed for a specific application and are not ready to hold unexpected variations. This fact entails a high cost as each new visual inspection task has to be redesigned, from the superficial defect analysis to the overall inspection system by a human expert. The solution here proposed points to extract linguistic knowledge on surface visual inspection, which is easily differentiated in types, by means of associations. This would allow changing or reconfiguring the components related to the inspection process that vary, in the production line, due to changes on either camera, surface or defect type.

To this aim a visual inspection architecture, namely ARDIS, is proposed to surface dynamic inspection in laminated materials based on the configuration of a specific visual system to obtain a good quality control of the manufacturing surface. The configuration task for surface inspection is analysed at the knowledge level and the task is decomposed into simple subtasks to reach the inference level, the most basic tasks. This task is solved as a Configuration-Design task following the CommonKADS methodology [5] leading to the knowledge-based dynamic architecture (ARDIS), which can account with all the knowledge involved in the process. The surface inspection generic knowledge is differentiated among environment, image quality, real-time and computer vision techniques to be integrated in ARDIS.

The injection of surface inspection knowledge from the human expert is an essential issue in the configuration process of any visual inspection system. For instance, the change of the surface thickness or lighting in a production line implies a variation that is not considered in current inspection systems. In this case, the distance between the camera and the surface varies and the camera gets out of focus, generating blurred images. Thus, expert knowledge is necessary to solve the problem as human experts do. But often, we lack knowledge from the expert, so mining linguistic information from numerical data will aid the architecture to re-configure, just in time, the inspection system.

In ARDIS, the surface inspection task, at the domain knowledge level, is analysed and decomposed into simple subtasks to reach the inference level that allows differentiating the Environment (E), Image Quality (IQ), Real-Time (RT) and Computer Vision Techniques (CVT) requirements.

This work is a first approach, to obtain the associations among E, IQ, RT and CVT, to configure the parameters of the inspection system to carry out a specific surface inspection task. In stainless steel inspection, real-time parameters could affect the more the CVT parameters.

ISBN: 978-989-95079-6-8

To reach the proposed goal, numerical data from E, IQ, RT and CVT are obtained and evaluated over a set of stainless steel and wood images. Results are shown where CVT is selected as a dependent attribute and E, IQ, RT are independent attributes.

2 ARDIS a cognitive architecture

The cognitive architecture, ARDIS, is proposed to deal with the complex task of visual inspection of laminated surfaces in industrial environments displaying a high variability in lighting, reflectance and real-time defect detection conditions. The generic expert knowledge on surface visual inspection and the deep knowledge on the inspection restrictions and defects characteristics that has the line inspector, is used in the design of a visual inspection system.
The architecture has to offer an adaptive behaviour for detecting different defects or inspecting different type of surfaces. To this aim the cognitive architecture is designed so as to be able to configure in real-time a visual inspection system, adapted for each specific type of laminated surface and defect.
The proposed architecture allows using generic knowledge on surface inspection to configure a specific system of visual surface inspection for different laminated materials such as, metal or wood. This knowledge, in the static roles of the architecture, is codified by means of crisp and fuzzy rules.
The configured visual inspection system is composed of a set of components, design elements, and their parameters. Each configured component depends on the type of surface and defect to be inspected. An environment component can be the lighting device or the camera, which could be composed of several sub-components, each one described by a set of parameters.
The method Propose-&-Revise used to solve the visual surface inspection task is displayed in Figure 1, showing the diagram of subtasks and inferences. This main method shows how the ARDIS architecture operates in three steps: (1) the initialization process, (2) the extension of the initial inspection skeleton and, (3) the revision process of the overall inspection system.

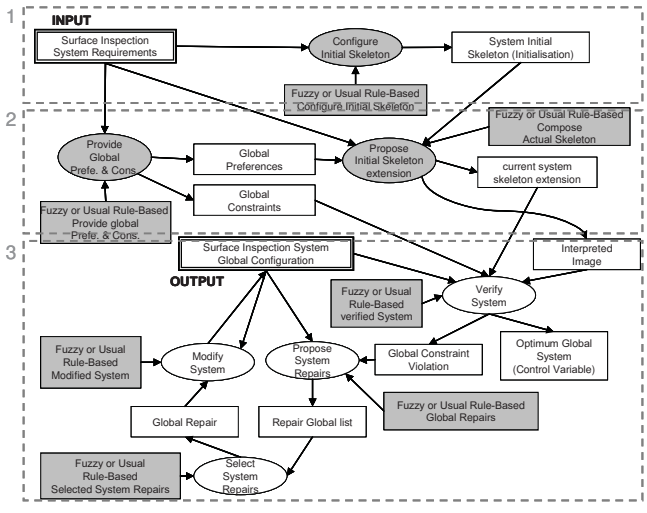


Figure 1: Subtask and inference diagram of the method Propose-&-Revise to solve the surface inspection task.

2.1 The domain generic knowledge on surface inspection

The domain generic knowledge on surface inspection is composed of different types of generic knowledge on real-time restrictions, image quality control restrictions, environment conditions and computer vision techniques. This differentiation of the type of knowledge makes possible to distinguish the knowledge used in each inference, making easier the specification of components and their interactions. Thus, the ARDIS architecture has been provided with all the necessary functionalities required to configure a specific complete system of visual inspection where knowledge is partitioned to be ease reused.
The relationships between the knowledge types are described in Figure 2. Here, the environment knowledge is related to real-time knowledge, such as image acquisition rate or lighting components that can influence real-time components in one or another way. So, during the configuration step of an inspection system, if an environment component is configured this has to be taken into account in the configuration of the real-time components.

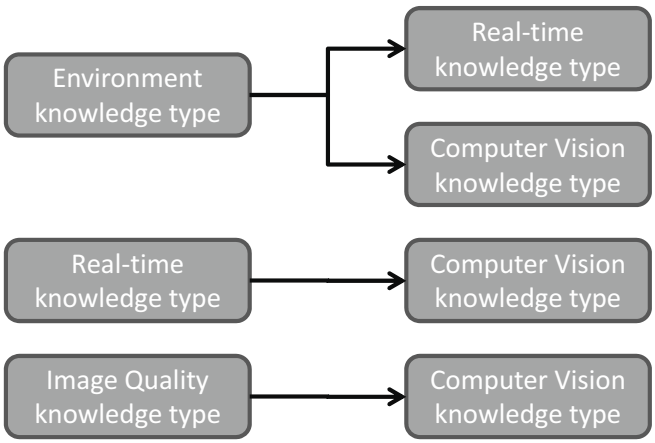


Figure 2: Influence diagram in the surface inspection process

2.2 Dynamic configuration using ARDIS architecture

The dynamic configuration process relies on (i) the cognitive architecture ARDIS, (ii) the domain generic knowledge on surface inspection and (iii) the line inspector knowledge who selects the requirements of the inspection system. This dynamic configuration process allows the inspection of each specific surface in a production line. Consequently, it is possible to inspect stainless steel or wood laminates in the same production line. Following the control structure of the ARDIS architecture, requirements and generic knowledge are integrated into crisp and fuzzy rules. This process gives rise to the configuration of the components of the surface inspection system. The Table 1 shows the set of rules used to configure the environment components of the inspection system.

ISBN: 978-989-95079-6-8

Table 1: Knowledge Base for configuring the surface inspection system.

KNOWLEDGE BASE USED IN THE INFERENCES OF THE SUBTASK Propose Environment Skeleton Extension
IF ProductionLineSpeed is high THEN ExpositionTime is short
IF IlluminationSystem is ExteriorLaserIllumination THEN CameraSensorGain is automatic
IF IlluminationSystem is ExteriorLaserIllumination THEN Camera-IlluminationRelativePosition is 45degrees
IF IlluminationType is GreenLaserSource THEN ImageChannel is Green

3 Stainless steel and wood images

The work departs from a set of images from inspections of residual oxide scale on cold stainless steel strip and wood. The first visual system is configured to detect 50 microns size defects and the second one to inspect larger defects in a wood surface. The images are obtained with a robust laser technique for diffuse illumination which allows stainless steel or wood industrial inspection with the same system. This visual technique for surface inspection consists on a smart vision system based on a green laser diode diffuse illumination, thus images display a green predominant colour.

The Figure 3 shows a set of images of stainless steel to inspect micro oxide defects, and Figure 4 displays a new laminated material, wood, where knot defects and wood irregularities have to be inspected in the same production line. The wood and stainless steel images have been acquired with an experimental system based on green laser illumination which allows inspecting defects ranging micro to five millimetres size. The acquisition visual system utilizes magnification 1.

4 Defining parameters from images

The analysis of the images has been accomplished with a HP xw4600 Workstation and "Intel(R) Core(TM)2 Quad CPU Q6600 @ 2.40GHz" processor. The numeric parameters have been calculated using the image processing toolbox of MATLAB [6].

Parameters of environment (E), image quality (IQ), real-time (RT) and computer vision techniques (CVT) are differentiated and obtained from the original image acquired by the visual inspection system. The E, IQ, RT and CVT parameters are all derived from an image, either stainless steel or a wood, and the evaluation is performed under the same conditions.

Several features are obtained from images but no associations among them. At this point, some additional knowledge is required to establish the priority on the selected parameter (E, IQ, RT) for the configuration of the inspection system.

The numerical data and F-transform will aid to mine the influence of the E, IQ and RT parameters on the CVT parameters. This knowledge will be, later on, used to set the priority in the configuration process. Then, fixing a CVT segmentation technique, we will try to find the influence between "E, IQ and RT" and the CVT technique. Thus, a segmentation technique could be more influenced by real-time inspection than a non-uniform lighting in the environment.

The linguistic information obtained for configuring a visual surface inspection system has the following structure: IF SegmentationTechnique is CVT and SurfaceInspection is StainlessSteel THEN ParameterPriority is Real-Time.

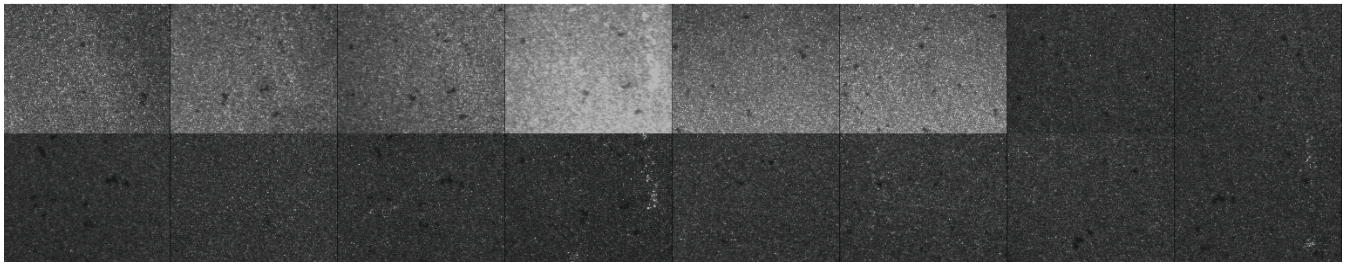


Figure 3: Images of stainless steel with micro oxide defects

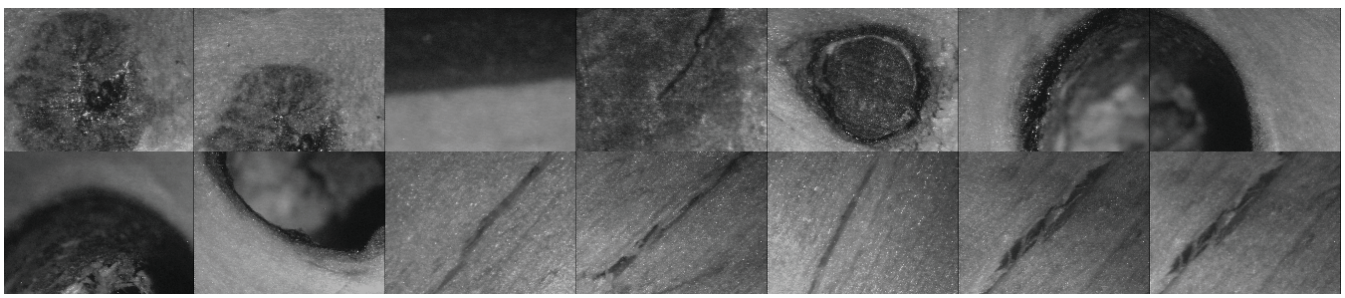


Figure 4: Images of wood with knot defect and other wood irregularities

ISBN: 978-989-95079-6-8

The parameters chosen to evaluate the E, IQ, RT and CVT properties over the acquired image, selecting the Mean-Shift algorithm [7] as the segmentation technique, are:

1) Evaluation of the Environment (E): the lighting non-uniformity of the image is used to measure the influence of the environment in the image. This environment evaluation will be an independent attribute or parameter and the corresponding contexts are $E_{StainlessSteel}$= [0.0736 8.1698], E_{wood}= [8.0138 50.2337]. Figure 5 shows data obtained from stainless steel and wood images.

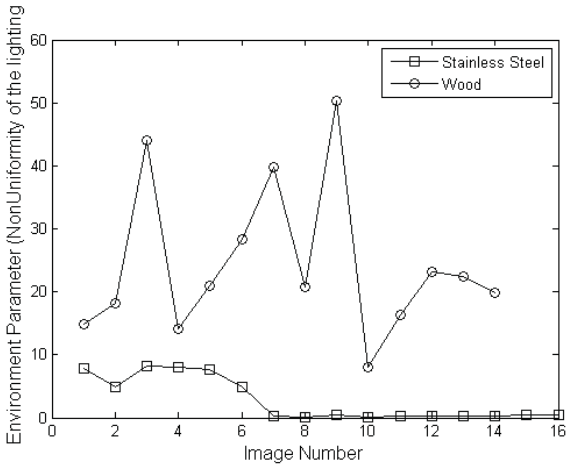


Figure 5: Evolution of the lighting in Stainless Steel and Wood images.

2) Evaluation of the Real-Time (RT): the computation time of the Mean-Shift algorithm [7] is used to estimate a Real-Time parameter in the image. The Real-Time evaluation will be an independent attribute or parameter and the corresponding contexts are $RT_{StainlessSteel}$ = [8.3594 12.3281], RT_{Wood}= [10.1719 13.7813]. Figure 6 shows data obtained on a set of stainless steel and wood images.

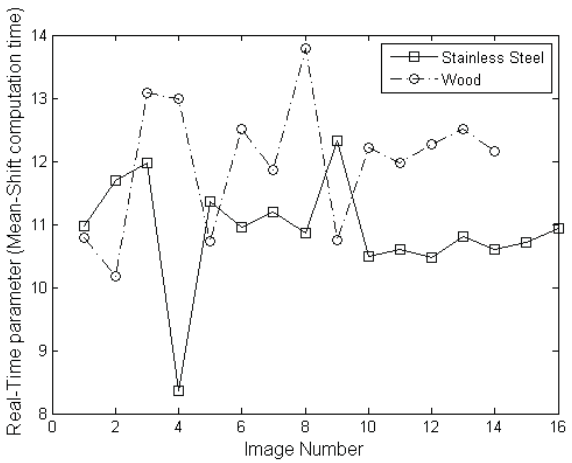


Figure 6: Evolution of the Mean-Shift computation time in Stainless Steel and Wood images.

3) Evaluation of Image Quality (IQ): gray-level co-occurrence matrix from images, was used to calculate the contrast property from the gray-level co-occurence matrix which returns a measure of the intensity contrast between a pixel and its neighbour over the whole image. The Image Quality evaluation will be an independent parameter and the corresponding contexts are $IQ_{StainlessSteel}$ = [5.0702 35.2404], IQ_{Wood} = [0.2268 10.5327]. Figure 7 shows data obtained from stainless steel and wood images.

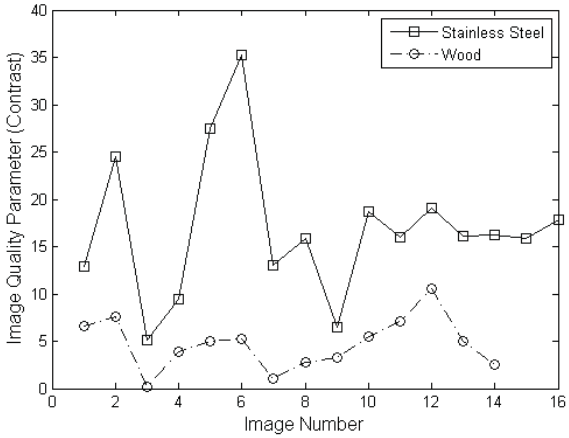


Figure 7: Evolution of the contrast in Stainless Steel and Wood images.

4) Evaluation of the segmentation technique (CVT): the Mean-Shift technique is used as the segmentation technique and the evaluation is based on the number of Mean-Shift regions located by the algorithm. This CVT evaluation will be the dependent parameter and the corresponding contexts are $CVT_{StainlessSteel}$ = [919 2968], CVT_{Wood} = [8 243]. Figure 8 shows data obtained from stainless steel and wood images.

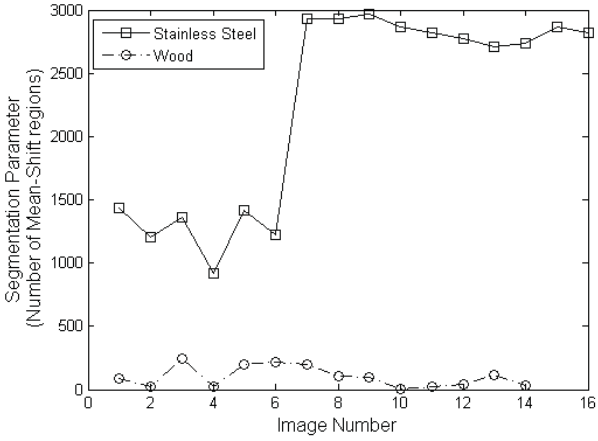


Figure 8: Evolution of the number of Mean-Shift regions in Stainless Steel and Wood images.

ISBN: 978-989-95079-6-8

5 Membership functions for data analysis

The membership functions map points in the input space to a membership value (or degree of membership) between 0 and 1.

If X is the input space and its elements are denoted by x, then a fuzzy set A in X is defined as a set of ordered pairs.

$$A = \{x, \mu_A(x) \mid x \in X\}$$

Where $\mu_A(x)$ is the membership degree of x in A.

Gaussian functions are selected to represent membership functions, due to their smoothness and concise notation. These curves have the advantage of being smooth and nonzero at all points.

The membership functions selected for the environment (E), real-time (RT), image quality (IQ) and computer vision techniques (CVT) parameters are shown in the Figures 9, 10 and the vectors of nodes $\langle l_1,...,l_k \rangle$ which define the membership functions of the induced fuzzy sets on the E, IQ, RT and CVT parameters for Stainless Steel and Wood, are displayed in Table 2.

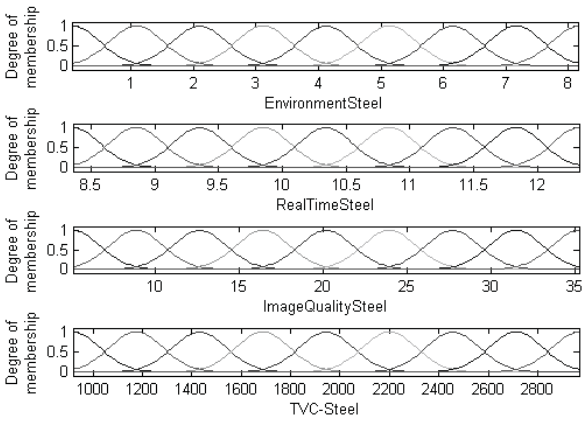


Figure 9: Membership functions with 9 nodes for E, IQ, RT and CVT parameters on Stainless Steel images.

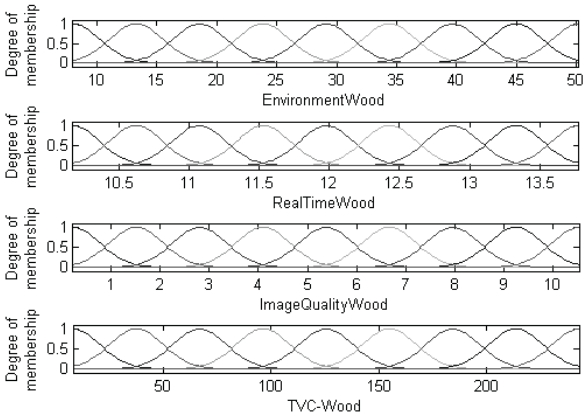


Figure 10: Membership functions with 9 nodes for E, IQ, RT and CVT parameters on Wood images.

Table 2: Vectors of nodes of the membership functions.

	Stainless Steel	*Wood*
Environment **(E)**	[0.0736, 1.086, 2.098, 3.11, 4.122, 5.134, 6.146, 7.158, 8.17]	[8.014, 13.29, 18.57, 23.85, 29.12, 34.4, 39.68, 44.96, 50.23]
Real-Time **(RT)**	[8.359, 8.855, 9.352, 9.848, 10.34, 10.84, 11.34, 11.83, 12.33]	[10.17, 10.62, 11.07, 11.53, 11.98, 12.43, 12.88, 13.33, 13.78]
Image Quality **(IQ)**	[5.07, 8.841, 12.61, 16.38, 20.16, 23.93, 27.7, 31.47, 35.24]	[0.2268, 1.515, 2.803, 4.092, 5.38, 6.668, 7.956, 9.244, 10.53]
Computer Vision Techniques **(CVT)**	[919, 1175, 1431, 1687, 1944, 2200, 2456, 2712, 2968]	[8, 37.38, 66.75, 96.13, 125.5, 154.9, 184.3, 213.6, 243]

6 Mining associations between environment, real-time, image quality and computer vision techniques parameters: A case study

The proposed method is based on F-transform to search dependence between the CVT segmentation technique and the E, IQ, RT parameters. These associations allow to determine the priority of the parameters E, IQ, RT for configuring a visual surface inspection system.

Following the applications of F-transform to data analysis [2], we look for dependences among some attributes having a general expression of the type: $X_z = H(X_1,...,X_k)$, where X_z is the dependent attribute, such as the CVT segmentation technique and $X_1,\ldots,X_k$ are the independent attributes as E, IQ, RT parameters. The evaluation of these attributes is carried out on a set of 16 images (objects $\{o_1,\ldots,o_j,\ldots,o_{16}\}$) for Stainless Steel inspection and 14 images (objects $\{o_1,\ldots,o_j,\ldots,o_{14}\}$) for Wood inspection.

This first approach selects a computer vision technique for image segmentation, such as the Mean-Shift algorithm and afterwards, the F-transform is used, as well as the "*degree of support*" (r) and "*confidence*" (γ) parameters [2], for mining associations in the data, finding hypothesis of potential dependences among the selected segmentation technique and the E, RT, IQ parameters. The *degree of support* (r) is defined by (1):

$$r = \frac{|\{o_j \mid Fn[l_1...l_k](o_j) > 0\}|}{m} \quad (1)$$

Where m is the number of objects, Fn $[l_1\ldots l_k](o_j)$ is the membership function of an induced fuzzy set on the set of objects $\{o_1,...,o_m\}$ and $|\cdot|$ denotes a number of elements of the given set.

The *degree of confidence* (γ) using the F-transform and inverse F-transform is formulated in (2):

$$\gamma = \sqrt{\frac{\sum_{j=1}^{m}(f_F(o_j) - F_{l_1...l_k})^2 \cdot Fn[l_1...l_k](o_j)}{\sum_{j=1}^{m}(f_{jz} - F_{l_1...l_k})^2 \cdot Fn[l_1...l_k](o_j)}} \quad (2)$$

Where, $f_F(o_j)$ is an auxiliary function on the set of objects $\{o_1,...,o_m\}$ induced by the inverse F-transform with components $F_{l_1 \ldots l_k}$ of F-transform, and f_{jz} is the value of the dependent attribute X_z measured on the object o_j.

Finally, each association is supported by the mean of the "*degrees of support*" (r_{mean}) and "*confidence*" (γ_{mean}) parameters, as each r and γ corresponding to the *k*-tuple of fuzzy sets $\langle Fn[l_1],\cdots,Fn[l_k]\rangle$ is analysed, therefore the mean value represents a meaningful value of r and γ.

The results are presented under the format (3):

$$(X_1 \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{mean\; r,\gamma}{\overset{F}{\sim}} X_z(segmentation) \quad (3)$$

Then applying the F-transform, three associations for Stainless Steel and Wood are obtained, using the segmentation algorithm (Mean-Shift):

$$(E_{\substack{Stainless\\Steel}} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.3194\\\gamma=0.00021982}}{\overset{F}{\sim}} CVT_{\substack{Stainless\\Steel}}(Mshift)$$

$$(RT_{\substack{Stainless\\Steel}} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.3889\\\gamma=0.00040048}}{\overset{F}{\sim}} CVT_{\substack{Stainless\\Steel}}(Mshift)$$

$$(IQ_{\substack{Stainless\\Steel}} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.3750\\\gamma=0.0015}}{\overset{F}{\sim}} CVT_{\substack{Stainless\\Steel}}(Mshift)$$

$$(E_{Wood} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.3968\\\gamma=0.0547}}{\overset{F}{\sim}} CVT_{Wood}(Mshift)$$

$$(RT_{Wood} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.4048\\\gamma=0.0084}}{\overset{F}{\sim}} CVT_{Wood}(Mshift)$$

$$(IQ_{Wood} \; is \; \langle Fn[l_1],\cdots,Fn[l_k]\rangle) \underset{\substack{mean\\r=0.3810\\\gamma=0.0142}}{\overset{F}{\sim}} CVT_{Wood}(Mshift)$$

The linguistic approach of these associations for configuring a visual surface inspection system can be formulated as:

Stainless Steel:
"IF $CVT_{StainlessSteel}$ is Mean-Shift THEN ConfigurationPriority is IQ"
with *degree of support* r=0.3750 and *confidence* γ=0.0015.

The three Stainless Steel associations, highlight the fact that Mean-Shift technique has a stronger influence on image quality than do the others parameters, so the priority of the Stainless Steel configuration process is the image quality parameters. This association is quite obvious as a consequence of the influence of small defects and huge number of regions, Figure 8, detected by the Mean-Shift algorithm and the high frequency of the images.

Wood:
"IF CVT_{Wood} is Mean-Shift THEN ConfigurationPriority is E"
with the *degree of support* r= 0.3968 and *confidence* γ= 0.0547

The three Wood associations points that Mean-Shift has a larger influence on environment than the others parameters, so the priority of the Wood configuration process is the environment parameters. This association is also obvious due to the influence of huge defects and small number of regions, Figure 8, detected by the Mean-Shift algorithm and the low frequency of the images.

7 Conclusions

This work proposes a first approach to derive associations among environment, real-time, image quality and computer vision techniques, based on the F-transform and inverse F-transform.
This is the first step to tune the parameters of a visual inspection system to accomplish a specific surface inspection task, either on Stainless steel or Wood.
The mean of r and γ, is used to summarize the results instead of using individual r and γ values. The derived associations are quite obvious and suggest hypothesis that correspond to expert knowledge.
Future work will be performed on applications that use the linguistic form of the observed associations from numerical data as the configuration task for surface inspection that requires such linguistic knowledge.

Acknowledgment

Authors express their gratitude to ACERINOX S.A. (Spanish stainless steel manufacturing company) for fruitful cooperation and financial support and CICYT- DPI - 2006-14497 for insightful knowledge on cognitive architectures. Authors also thank Eugenio Villanueva for his valuable aid and comments in technical aspects of this work.

References

[1] I. Perfilieva. Fuzzy transforms: Theory and applications. Fuzzy Sets and Systems, 157:993-1023, 2006.

[2] I. Perfilieva, V. Novák and A. Dvořák. Fuzzy transform in the analysis of data. International Journal of Approximate Reasoning, 48:36-46, 2008.

[3] D. Martín. Arquitectura dinámica para inspección visual de superficies en tiempo real. PhD thesis, Madrid, Spain: UNED, 2008.

[4] T. S. Newman and A. K. Jain, 'A survey of automated visual inspection', Computer Vision and Image Understanding, 61:231–262, 1995.

[5] G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. V. de Velde, and B.Wielinga, Knowledge Engineering and Manage-ment: The CommonKADS Methodology, The MIT Press, 2000.

[6] Matlab, The MathWorks Inc. Natick, MA, United States, http://www.mathworks.com/

[7] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5): 603-619, May 2002.

ISBN: 978-989-95079-6-8

A construction of an L-fuzzy valued measure of L-fuzzy sets

Vechislav Ruzha, Svetlana Asmuss

Department of Mathematics, University of Latvia
Zellu street 8, Riga, LV-1002, Latvia
Institute of Mathematics and Computer Science of University of Latvia
Rainis blvd. 29, Riga, LV-1459, Latvia
Email: vecislavs.ruza@ge.com, svetlana.asmuss@lu.lv

Abstract— *We develope a construction of an L-fuzzy valued measure by extending a measure defined on a σ-algebra of crisp sets to an L-fuzzy valued T_M-countably additive measure defined on a T_M-clan of L-fuzzy sets in the case when operations for L-fuzzy sets and L-fuzzy real numbers are based on the minimum t-norm T_M.*

Keywords— L-fuzzy set, L-fuzzy real number, L-fuzzy valued measure.

1 Introduction

There are many works devoted to measures of fuzzy sets. The most important concepts and results concerning this topic are considered by D. Butnariu and E.P. Klement [2]. Our interest is in developing a theory of measure and integral where not only sets are fuzzy, but also measure and integral take fuzzy real values. To do this we need a concept of a fuzzy real number. There are several (quite different) ways to define a fuzzy real number. For our purposes we use fuzzy real numbers as they were first defined by B. Hutton [3], and then studied thoroughly in a series of papers (see e.g.[4],[5]). The preference of using this approach for defining fuzzy real numbers is motivated by our intention to develop the investigations from [6],[7].

In our previous work [1] we suggested the construction of a T-measure of L-fuzzy sets by extension of a measure defined on a σ-algebra of crisp sets to a T-measure on a T-tribe in the case of the minimum t-norm T_M and $L=[0,1]$. The main purpose of the present paper is to show that the construction still can be generalized in the case when L is complete and completely distributive lattice $L(\wedge,\vee,0_L,1_L)$. It was achieved in the case of the minimum t-norm T_M: $T_M(x,y)=x\wedge y$.

2 Preliminaries

Let X be a nonempty set. The class of all L-fuzzy subsets of X (in the sequel L-fuzzy sets) is denoted L^X. The operations for L-fuzzy sets $A,B\in L^X$ such as intersection, union and difference are defined by using a minimum triangular norm T_M, its corresponding triangular conorm S_M and an involution N:
$(A\cap B)(x)=A(x)\wedge B(x)$, $(A\cup B)(x)=A(x)\vee B(x)$,
$(A\setminus B)(x)=A(x)\wedge N(B(x))$.

2.1 Classes of L-fuzzy sets

In order to define an L-fuzzy valued function of L-fuzzy sets we consider such classes of L-fuzzy sets as T_M-semirings (defined by analogy with classical case, see e.g. [8]) and T_M-clans (see e.g. [2]).

Definition 2.1. *A subclass $\wp\subset L^X$ is called a T_M-semiring on X if the following properties are satisfied:*

(i) $\emptyset\in\wp$;

(ii) for all $A,B\in\wp$ *we have* $A\wedge B\in\wp$;

(iii) for all $A,B\in\wp$ *there exist such T_M-disjoint L-fuzzy sets* $A_1,A_2,...,A_n\in\wp$ *that* $A\setminus B=\bigvee_{i=1}^{n}A_i$.

A finite family of L-fuzzy sets $A_1,A_2,\ldots,A_n$ is said to be *T_M-disjoint* (see e.g.[2]) iff for each $k\in\{1,\ldots,n\}$

$$(\bigvee_{j=1,j\neq k}^{n}A_j)\wedge A_k=\emptyset.$$

A countable family of L-fuzzy sets is said to be T_M-disjoint iff every finite subfamily of this family is T_M-disjoint.

Definition 2.2. *A subclass $\mathcal{A}\subset L^X$ is called a T_M-clan on X iff the following properties are satisfied:*

(i) $\emptyset\in\mathcal{A}$;

(ii) for all $A\in\mathcal{A}$ *we have* $N(A)\in\mathcal{A}$;

(iii) for all $A,B\in\mathcal{A}$ *we have* $A\wedge B\in\mathcal{A}$.

2.2 L-fuzzy real numbers

For our purposes we use the L-fuzzy real numbers as they were first defined by B. Hutton [3].

Definition 2.3. *An L-fuzzy real number is a function $z:\mathbb{R}\to L$ such that*

(i) z is non-increasing;

(ii) $\bigwedge_x z(x)=0_L$, $\bigvee_x z(x)=1_L$;

(iii) z is left semi-continuous, i.e. $\bigwedge_{t<x} z(t)=z(x)$.

The set of all L-fuzzy real numbers is called *the L-fuzzy real line* and it is denoted by $\mathbb{R}(L)$. An L-fuzzy number z is called *non-negative* if $z(0)=1_L$. The set of all non-negative L-fuzzy real numbers we denote by $\mathbb{R}_+(L)$.

The ordinary real line $\mathbb{R}$ can be identified with the subspace

$$\{z_a\mid a\in\mathbb{R}\}\subset\mathbb{R}(L)$$

ISBN: 978-989-95079-6-8

by assigning to a real number $a \in \mathbb{R}$ the fuzzy real number z_a defined by

$$z_a(x) = \begin{cases} 1_L, \text{ if } x \leq a, \\ 0_L, \text{ if } x > a. \end{cases}$$

The operation of L-fuzzy addition $\oplus$ defined by

$$(z_1 \oplus z_2)(x) = \bigvee_t \left(\{ z_1(t) \wedge z_2(x-t) \} \right)$$

whenever $z_1, z_2 \in \mathbb{R}(L)$, is a jointly continuous extension of addition $+$ from the real line $\mathbb{R}$ to the L-fuzzy real line $\mathbb{R}(L)$.

Given a sequence of non-negative L-fuzzy real numbers $(z_n)_{n\in\mathbb{N}} \subset \mathbb{R}_+(L)$ we consider the countable sum

$$\bigoplus_{n\in\mathbb{N}} z_n = \bigvee_{n\in\mathbb{N}} (z_1 \oplus z_2 \oplus ... \oplus z_n).$$

2.3 L-fuzzy valued functions

Let $\aleph$ be a class of L-fuzzy sets. Within this section some basic properties of an L-fuzzy valued function $\eta : \aleph \to \mathbb{R}_+(L)$ are considered.

Definition 2.4. *An L-fuzzy valued function η is called T_M-additive iff*

for all $A, B \in \aleph$ such that $A \wedge B = \emptyset, A \vee B \in \aleph$ it holds $\eta(A \vee B) = \eta(A) \oplus \eta(B)$.

Definition 2.5. *An L-fuzzy valued function η is called T_M-countably additive iff*

for all $(A_n)_{n\in\mathbb{N}} \subset \aleph$ such that
$(\forall i, j \in \mathbb{N} : i \neq j \Rightarrow A_i \wedge A_j = \emptyset)$ and $\bigvee_{n=1}^{\infty} A_n \in \aleph$
it holds $\eta(\bigvee_{n=1}^{\infty} A_n) = \bigoplus_{n\in\mathbb{N}} \eta(A_n)$.

Definition 2.6. *Let $\wp$ be a T_M-semiring. A function $m : \wp \to \mathbb{R}_+(L)$ is called an L-fuzzy valued elementary T_M-measure if it satisfies the following conditions:*

(i) $m(\emptyset) = z_0$;

(ii) m is T_M-additive.

Definition 2.7. *Let $\mathcal{A}$ be a T_M-clan. A function $\mu : \mathcal{A} \to \mathbb{R}_+(L)$ is called an L-fuzzy valued T_M-countably additive measure if it satisfies the following conditions:*

(i) $\mu(\emptyset) = z_0$;

(ii) μ is T_M-countably additive.

3 A construction of an L-fuzzy valued elementary T_M-measure

Let Φ be a σ-algebra of "crisp" subsets of X and ν is a finite measure $\nu : \Phi \to [0, +\infty[$. Our aim is to construct an L-fuzzy valued T_M-countably additive measure $\tilde{\mu}$ on a T_M-clan by extension a crisp measure ν. To achieve this we generalize the well known construction of "classic" measure theory (see e.g. [8]) to the L-fuzzy case.

3.1 T_M-semiring $\tilde{\wp}$ of L-fuzzy sets

To realise the construction we use the special type of L-fuzzy sets with respect to σ-algebra Φ of crisp sets on X. For $M \in \Phi$, $\alpha \in L$ we define L-fuzzy set $A(M, \alpha)$:

$$(A(M,\alpha))(x) = \begin{cases} \alpha, x \in M, \\ 0_L, x \notin M. \end{cases}$$

For convenience we denote this special type of L-fuzzy sets by $A(M, \alpha)$ (the choice of a letter in the notation is not crucial: $B(N, \beta)$ is still the same type of an L-fuzzy set).

Proposition 3.1. *The class of L-fuzzy sets*

$$\tilde{\wp} = \{A(M,\alpha) | M \in \Phi \text{ and } \alpha \in L\}$$

is a T_M-semiring.

Proof. For all $A(M,\alpha), B(K,\beta) \in \tilde{\wp}$ we have
$A(M,\alpha) \wedge B(K,\beta) = C(M \cap K, \alpha \wedge \beta) \in \tilde{\wp}$ and
$A(M,\alpha) \setminus B(K,\beta) = A_1(M \setminus K, \alpha) \vee A_2(M \cap K, \alpha \wedge N(\beta))$,
where $A_1(M \setminus K, \alpha), A_2(M \cap K, \alpha \wedge N(\beta)) \in \tilde{\wp}$.
Taking into account that also $\emptyset = A(\emptyset, 0_L) \in \tilde{\wp}$, it is proved that $\tilde{\wp}$ is a T_M-semiring. □

3.2 L-fuzzy valued elementary T_M-measure $\tilde{m}$

We define an L-fuzzy valued function $\tilde{m}$ on the T-semiring $\tilde{\wp}$ by the formula $\tilde{m}(A(M,\alpha)) = z_{\nu(M),\alpha}$, where

$$z_{\nu(M),\alpha}(t) = \begin{cases} 1_L, t \leq 0, \\ \alpha, 0 < t \leq \nu(M), \\ 0_L, t > \nu(M), \end{cases}$$

is an L-fuzzy real number.

Proposition 3.2. *$\tilde{m}$ is an L-fuzzy valued elementary T_M-measure.*

Proof. It is easy to see that

$$\tilde{m}^*(\emptyset) = z_{\nu(\emptyset),0_L} = z_0$$

and the equality

$$\tilde{m}(A(M,\alpha)) \oplus \tilde{m}(B(K,\beta)) = \tilde{m}(A(M,\alpha) \vee B(K,\beta))$$

is true if $A(M,\alpha) = \emptyset$ or $B(K,\beta) = \emptyset$.

Let us consider $A(M,\alpha), B(K,\beta) \in \tilde{\wp}$ such that

$$A(M,\alpha) \wedge B(K,\beta) = \emptyset,\ A(M,\alpha) \cup B(K,\beta) \in \tilde{\wp} \text{ and } \alpha \neq 0_L, \beta \neq 0_L.$$

It follows that $M \cap K = \emptyset$ or $\alpha \wedge \beta = 0_L$.

If $M \cap K = \emptyset$ then $\alpha = \beta$ and in this case

$$\tilde{m}(A(M,\alpha)) \oplus \tilde{m}(B(K,\alpha)) = z_{\nu(M),\alpha} \oplus z_{\nu(K),\alpha} =$$
$$= z_{\nu(M)+\nu(K),\alpha} = z_{\nu(M\cup K),\alpha} = \tilde{m}(A(M,\alpha) \vee B(K,\alpha)).$$

If $M \cap K \neq \emptyset$ then it is sufficient to consider the case when $\alpha \wedge \beta = 0_L$. In this case α and β are incomparable (due to the assumptions that $\alpha \neq 0_L, \beta \neq 0_L$). Because of

$$A(M,\alpha) \vee B(K,\beta) =$$
$$= C_1(M \cap K, \alpha \vee \beta) \vee C_2(M \setminus K, \alpha) \vee C_3(K \setminus M, \beta) \in \tilde{\wp}$$

we obtain that $K \setminus M = \emptyset$, $M \setminus K = \emptyset$ and hence $M = K$. Then

$$\tilde{m}(A(M,\alpha)) \oplus \tilde{m}(B(M,\beta)) = z_{\nu(M),\alpha} \oplus z_{\nu(M),\beta} =$$
$$= z_{\nu(M),\alpha\vee\beta} = C(M, \alpha \vee \beta) = \tilde{m}(A(M,\alpha) \vee B(M,\beta)).$$

By this we prove that $\tilde{m}$ is T_M-additive. □

ISBN: 978-989-95079-6-8

4 Measurable L-fuzzy sets

4.1 An extension of an L-fuzzy valued elementary T_M-measure

We define an L-fuzzy valued function $\tilde{m}^* : L^X \to \mathbb{R}_+(L)$ by the following formula

$\tilde{m}^*(E) = \bigwedge\{\bigoplus_{n=1}^{\infty} \tilde{m}(E_n) \mid (E_n)_{n\in\mathbb{N}} \subset \tilde{\wp}:$

$E \le \bigvee_{n=1}^{\infty} E_n$ and $(E_n)_{n\in\mathbb{N}}$ is T_M-disjoint$\}, E \in L^X$.

Remarks.

(i) For every $E \in L^X$ there always exists such a sequence $(E_n)_{n\in\mathbb{N}} \subset \tilde{\wp}$ that $E \le \bigvee_{n=1}^{\infty} E_n$. It is enough to take $E_1(X, 1_L)$. Thus $\tilde{m}^*$ is bounded from above in the following sense: $\tilde{m}^*(E) \le z_{\nu(X),1_L}$ for all $E \in L^X$.

(ii) For all $E \in \tilde{\wp}$ we have $\tilde{m}^*(E) = \tilde{m}(E)$.

Proposition 4.1. *For all $A, B \in L^X$ we have*

$$\tilde{m}^*(A) \oplus \tilde{m}^*(B) \ge \tilde{m}^*(A \wedge B) \oplus \tilde{m}^*(A \vee B).$$

Proof. For a given L-fuzzy sets A and B we consider two T_M-disjoint sequences:

$$(C_i(M_i, \alpha_i))_{i\in\mathbb{N}} \subset \tilde{\wp} : A \le \bigvee_{i=1}^{\infty} C_i,$$

$$(D_j(K_j, \beta_j))_{j\in\mathbb{N}} \subset \tilde{\wp} : B \le \bigvee_{j=1}^{\infty} D_j$$

and define a new T_M-disjoint sequence

$$(H_{ij}(L_{ij}, \gamma_{ij}))_{i,j\in\mathbb{N}}, \text{ where } L_{ij} = M_i \cap K_j \text{ and } \gamma_{ij} = \alpha_i \wedge \beta_i.$$

Obviously, $A \wedge B \le \bigvee_{i\in\mathbb{N}} \bigvee_{i\in\mathbb{N}} H_{ij}$.

As to L-fuzzy set $A \vee B$ we can cover it with elements of three sequences:

$$\begin{aligned} (\tilde{H}_{ij}(L_{ij}, \lambda_{ij}))_{i,j\in\mathbb{N}}, &\quad \text{with} \quad L_{ij} = M_i \cap K_j, \lambda_{ij} = \alpha_i \vee \beta_i, \\ (\tilde{C}_i(\tilde{M}_i, \alpha_i))_{i\in\mathbb{N}}, &\quad \text{with} \quad \tilde{M}_i = M_i \setminus \bigcup_{j\in\mathbb{N}} K_j, \\ (\tilde{D}_j(\tilde{K}_j, \beta_j))_{j\in\mathbb{N}}, &\quad \text{with} \quad \tilde{K}_j = K_j \setminus \bigcup_{i\in\mathbb{N}} M_i. \end{aligned}$$

Now let us transform the sum $\bigoplus_{i=1}^{\infty} \tilde{m}^*(C_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(D_j)$. We use the following notation:

$$F_{ij}^{\alpha} = (L_{ij}, \alpha_i), F_{ij}^{\beta} = (L_{ij}, \beta_j), ij \in \mathbb{N}.$$

Because of σ-additivity of measure ν we get

$$\nu(M_i) = \nu(\tilde{M}_i) + \sum_{j=1}^{\infty} \nu(L_{ij}), i \in \mathbb{N},$$

$$\nu(K_j) = \nu(\tilde{K}_j) + \sum_{i=1}^{\infty} \nu(L_{ij}), j \in \mathbb{N}.$$

It follows

$$\tilde{m}^*(C_i) = z_{\nu(M_i),\alpha_i} = z_{\nu(\tilde{M}_i) + \sum_{j=1}^{\infty} \nu(L_{ij}),\alpha_i} =$$

$$= z_{\nu(\tilde{M}_i),\alpha_i} \oplus \bigoplus_{j=1}^{\infty} z_{\nu(L_{ij}),\alpha_i} = \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(F_{ij}^{\alpha}),$$

$$\tilde{m}^*(D_j) = z_{\nu(K_j),\beta_j} = z_{\nu(\tilde{K}_j) + \sum_{i=1}^{\infty} \nu(L_{ij}),\beta_j} =$$

$$= z_{\nu(\tilde{K}_j),\beta_j} \oplus \bigoplus_{i=1}^{\infty} z_{\nu(L_{ij}),\beta_j} = \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \tilde{m}^*(F_{ij}^{\beta}).$$

Now we obtain

$$\bigoplus_{i=1}^{\infty} \tilde{m}^*(C_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(D_j) =$$

$$= \bigoplus_{i=1}^{\infty} (\tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(F_{ij}^{\alpha})) \oplus \bigoplus_{j=1}^{\infty} (\tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \tilde{m}^*(F_{ij}^{\beta})) =$$

$$= \bigoplus_{i=1}^{\infty} \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} \tilde{m}^*(F_{ij}^{\alpha}) \oplus \bigoplus_{j=1}^{\infty} \bigoplus_{i=1}^{\infty} \tilde{m}^*(F_{ij}^{\beta}) =$$

$$= \bigoplus_{i=1}^{\infty} \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} (\tilde{m}^*(F_{ij}^{\alpha}) \oplus \tilde{m}^*(F_{ij}^{\beta})).$$

Since

$$\tilde{m}^*(F_{ij}^{\alpha}) \oplus \tilde{m}^*(F_{ij}^{\beta}) = \tilde{m}^*(H_{ij}) \oplus \tilde{m}^*(\tilde{H}_{ij}),$$

we continue

$$\bigoplus_{i=1}^{\infty} \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} (\tilde{m}^*(F_{ij}^{\alpha}) \oplus \tilde{m}^*(F_{ij}^{\beta})) =$$

$$= \bigoplus_{i=1}^{\infty} \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{H}_{ij}) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} \tilde{m}^*(H_{ij}).$$

Now taking into account the fact that

$$A \wedge B \le \bigvee_{i=1}^{\infty} \bigvee_{j=1}^{\infty} H_{ij} \text{ and } A \vee B \le (\bigvee_{i=1}^{\infty} C_i) \vee (\bigvee_{j=1}^{\infty} D_j) \vee (\bigvee_{i=1}^{\infty} \bigvee_{j=1}^{\infty} \tilde{H}_{ij}),$$

we get

$$\bigoplus_{i=1}^{\infty} \tilde{m}^*(\tilde{C}_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{D}_j) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} \tilde{m}^*(\tilde{H}_{ij}) \oplus \bigoplus_{i=1}^{\infty} \bigoplus_{j=1}^{\infty} \tilde{m}^*(H_{ij}) \ge$$

$$\ge \tilde{m}^*(A \wedge B) \oplus \tilde{m}^*(A \vee B).$$

So independent of the choice of sequences $(C_i)_{i\in\mathbb{N}}$, $(D_j)_{j\in\mathbb{N}}$ it holds

$$\bigoplus_{i=1}^{\infty} \tilde{m}^*(C_i) \oplus \bigoplus_{j=1}^{\infty} \tilde{m}^*(D_j) \ge \tilde{m}^*(A \wedge B) \oplus \tilde{m}^*(A \vee B).$$

Finally, by taking infimums of sums we get

$$\tilde{m}^*(A) \oplus \tilde{m}^*(B) \ge \tilde{m}^*(A \wedge B) \oplus \tilde{m}^*(A \vee B).$$

□

ISBN: 978-989-95079-6-8

4.2 $\tilde{m}^*$-measurable L-fuzzy sets

We generalize the concept of measurability in the sense of Caratheodory (see e.g. [8]).

Definition 4.2. *A set $A \in L^X$ is called an $\tilde{m}^*$-measurable L-fuzzy set, if it satisfies the following conditions for all sets $E \in L^X$:*

(i) $\tilde{m}^*(A) \oplus \tilde{m}^*(E) = \tilde{m}^*(A \wedge E) \oplus \tilde{m}^*(A \vee E)$,

(ii) $\tilde{m}^*(N(A)) \oplus \tilde{m}^*(E) = \tilde{m}^*(N(A) \wedge E) \oplus \tilde{m}^*(N(A) \vee E)$.

We denote $\mathcal{A}_{\tilde{m}^*}$ the class of all $\tilde{m}^*$-measurable L-fuzzy sets. Some obvious properties of $\mathcal{A}_{\tilde{m}^*}$:

$\emptyset, X \in \mathcal{A}_{\tilde{m}^*}$,

$A \in \mathcal{A}_{\tilde{m}^*} \Longrightarrow N(A) \in \mathcal{A}_{\tilde{m}^*}$.

Theorem 4.3. *Class $\mathcal{A}_{\tilde{m}^*}$ is a T_M-clan.*

Proof. For given L-fuzzy sets $A_1, A_2 \in \mathcal{A}_{\tilde{m}^*}$ and $E \in L^X$ we will prove the equality

$$\begin{aligned} &\tilde{m}^*(E) \oplus \tilde{m}^*(A_1 \wedge A_2) = \\ &\tilde{m}^*(E \wedge (A_1 \wedge A_2)) \oplus \tilde{m}^*(E \vee (A_1 \wedge A_2)) \end{aligned} \tag{1}$$

Since A_1, A_2 are $\tilde{m}^*$-measurable we have

$$\tilde{m}^*(A_1) \oplus \tilde{m}^*(A_2) = \tilde{m}^*(A_1 \wedge A_2) \oplus \tilde{m}^*(A_1 \vee A_2),$$

$$\tilde{m}^*(A_1 \wedge E) \oplus \tilde{m}^*(A_1 \vee E) = \tilde{m}^*(A_1) \oplus \tilde{m}^*(E).$$

Now by summing up these two equalities and adding one more additional summand $\tilde{m}^*(A_1 \wedge (E \vee A_2))$ to both sides of the equality we obtain

$$\begin{aligned} &\tilde{m}^*(A_1) \oplus \tilde{m}^*(A_2) \oplus \tilde{m}^*(A_1 \wedge E) \oplus \\ &\oplus \tilde{m}^*(A_1 \vee E) \oplus \tilde{m}^*(A_1 \wedge (E \vee A_2)) = \\ &= \tilde{m}^*(A_1 \wedge A_2) \oplus \tilde{m}^*(A_1 \vee A_2) \oplus \tilde{m}^*(A_1) \oplus \\ &\oplus \tilde{m}^*(E) \oplus \tilde{m}^*(A_1 \wedge (E \vee A_2)). \end{aligned} \tag{2}$$

Let us transform now the left part of (2). To do this we use (3) and (4):

$$\begin{aligned} &\tilde{m}^*(A_1) \oplus [\tilde{m}^*(A_2) \oplus \tilde{m}^*(E \wedge A_1)] = \\ &= \tilde{m}^*(A_1) \oplus [\tilde{m}^*(E \wedge A_1 \wedge A_2) \oplus \tilde{m}^*((E \wedge A_1) \vee A_2)] = \\ &= \tilde{m}^*(E \wedge A_1 \wedge A_2) \oplus \tilde{m}^*((E \wedge A_1) \vee A_2 \vee A_1) \oplus \\ &\oplus \tilde{m}^*(((E \wedge A_1) \vee A_2) \wedge A_1) = \\ &= \tilde{m}^*(E \wedge A_1 \wedge A_2) \oplus \tilde{m}^*(A_1 \vee A_2) \oplus \tilde{m}^*((E \vee A_2) \wedge A_1) \end{aligned} \tag{3}$$

and

$$\begin{aligned} &\tilde{m}^*(E \vee A_1) \oplus \tilde{m}(A_1 \wedge (E \vee A_2)) = \\ &= \tilde{m}^*(E \vee (A_1 \wedge A_2) \vee A_1) \oplus \tilde{m}^*(E \vee A_1 \wedge A_2 \wedge A_1) = \\ &= \tilde{m}^*(E \vee (A_1 \wedge A_2)) \oplus \tilde{m}^*(A_1). \end{aligned} \tag{4}$$

Next we substitute (3) and (4) in (2).

$$\tilde{m}^*(E \wedge A_1 \wedge A_2) \oplus \tilde{m}^*(A_1 \vee A_2) \oplus \tilde{m}^*(A_1 \wedge (E \vee A_2)) \oplus$$

$$\oplus \tilde{m}^*(E \vee (A_1 \wedge A_2)) \oplus \tilde{m}^*(A_1) =$$

$$= \tilde{m}^*(A_1 \wedge A_2) \oplus \tilde{m}^*(A_1 \vee A_2) \oplus \tilde{m}^*(A_1) \oplus \tilde{m}^*(E) \oplus$$

$$\oplus \tilde{m}^*(A_1 \wedge (E \vee A_2)).$$

Finally, canceling like terms we obtain (1).
By analogy it can be proved that

$$\tilde{m}^*(E) \oplus \tilde{m}^*(A_1 \vee A_2) = \tilde{m}^*(E \wedge (A_1 \vee A_2)) \oplus \tilde{m}^*(E \vee (A_1 \vee A_2)).$$

Taking into account that $N(A_1 \vee A_2) = N(A_1) \wedge N(A_2)$ and $N(A_1 \wedge A_2) = N(A_1) \vee N(A_2)$, we get that $A_1 \vee A_2$ and $A_1 \wedge A_2$ are $\tilde{m}^*$-measurable. □

Proposition 4.4. $\tilde{\wp} \subset \mathcal{A}_{\tilde{m}^*}$

Proof. First we will show that for a given $A(H, \alpha) \in \tilde{\wp}$ and all $E \in L^X$ it holds

$$\tilde{m}^*(A) \oplus \tilde{m}^*(E) = \tilde{m}^*(A \vee E) \oplus \tilde{m}^*(A \wedge E).$$

Because of Proposition 4.1 it is sufficient to prove the inequality

$$\tilde{m}^*(A) \oplus \tilde{m}^*(E) \leq \tilde{m}^*(A \vee E) \oplus \tilde{m}^*(A \wedge E).$$

We consider two T_M-disjoint sequences $(C_k(M_k, \beta_k))_{k \in \mathbb{N}} \subset \tilde{\wp}$ and $(D_k(K_k, \gamma_k))_{k \in \mathbb{N}} \subset \tilde{\wp}$ such that

$$A \vee E \leq \bigvee_{k=1}^{\infty} C_k \text{ and } A \wedge E \leq \bigvee_{k=1}^{\infty} D_k$$

and define two the new sequences

$$(F_k(M_k \setminus H, \beta_k))_{k \in \mathbb{N}} \text{ and } (G_k(H \cap M_k, \beta_k))_{k \in \mathbb{N}}.$$

By T_M-additivity of $\tilde{m}$ it holds

$$\tilde{m}(G_k) \oplus \tilde{m}(F_k) = \tilde{m}(C_k), k \in \mathbb{N}.$$

Taking into account

$$A \leq \bigvee_{k=1}^{\infty} G_k \text{ and } E \leq (\bigvee_{k=1}^{\infty} F_k) \vee (\bigvee_{k=1}^{\infty} G_k),$$

we obtain

$$\tilde{m}^*(A) = \tilde{m}(A) \leq \bigoplus_{k=1}^{\infty} \tilde{m}(G_k),$$

$$\tilde{m}^*(E) \leq \bigoplus_{k=1}^{\infty} \tilde{m}(F_k) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(D_k).$$

Summing up

$$\tilde{m}^*(A) \oplus \tilde{m}^*(E) \leq \bigoplus_{k=1}^{\infty} \tilde{m}(G_k) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(F_k) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(D_k) =$$

$$= \bigoplus_{k=1}^{\infty} (\tilde{m}(G_k) \oplus \tilde{m}(F_k)) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(D_k) = \bigoplus_{k=1}^{\infty} \tilde{m}(C_k) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(D_k).$$

So independent of a choice of sequences $(C_k)_{k \in \mathbb{N}}$ and $(D_k)_{k \in \mathbb{N}}$ it holds

$$\bigoplus_{k=1}^{\infty} \tilde{m}(C_k) \oplus \bigoplus_{k=1}^{\infty} \tilde{m}(D_k) \geq \tilde{m}^*(A) \oplus \tilde{m}^*(E).$$

Finally, by taking infimums of sums we get

$$\tilde{m}^*(A) \oplus \tilde{m}^*(E) \leq \tilde{m}^*(A \vee E) \oplus \tilde{m}^*(A \wedge E).$$

Since

$$N(A) = B_1(H, N_\alpha) \vee B_2(X \setminus H, 1_L),$$

where $B_1(H, N_\alpha), B_2(X \setminus H, 1_L) \in \tilde{\wp}$, the equality

$$\tilde{m}^*(N(A)) \oplus \tilde{m}^*(E) = \tilde{m}^*(N(A) \wedge E) \oplus \tilde{m}^*(N(A) \vee E)$$

can be proved by analogy with the proof of Theorem 4.3. □

ISBN: 978-989-95079-6-8

5 A construction of an L-fuzzy valued T_M-countably additive measure

Let us denote by $\tilde{\mu}$ the restriction of $\tilde{m}^*$ to $\mathcal{A}_{\tilde{m}^*}$:

$$\tilde{\mu} : \mathcal{A}_{\tilde{m}^*} \to \mathbb{R}_+(L).$$

Theorem 5.1. *$\tilde{\mu}$ is T_M-countably additive L-fuzzy valued function.*

Proof. First we will show that for a T_M - disjoint sequence $A_1, A_2, ..., A_n \in \mathcal{A}_{\tilde{m}^*}$ it holds $\tilde{\mu}(\bigvee_{i=1}^{n} A_i) = \bigoplus_{i=1}^{n} \tilde{\mu}(A_i)$.

For $n = 2$ by usinq $\tilde{m}^*$-measurability and T_M - disjointness of A_1, A_2 we obtain

$$\tilde{\mu}(A_1) \oplus \tilde{\mu}(A_2) = \tilde{\mu}(A_1 \vee A_2) \oplus \tilde{\mu}(A_1 \wedge A_2) =$$
$$= \tilde{\mu}(A_1 \vee A_2) \oplus \tilde{\mu}(\emptyset) = \tilde{\mu}(A_1 \vee A_2).$$

Now we assume that the equality holds for a given $n \in \mathbb{N}$ and we prove it for $n+1$ T_M-disjoint sets $A_1, A_2, ..., A_{n+1} \in \mathcal{A}_{\tilde{m}^*}$

$$\tilde{\mu}(\bigvee_{k=1}^{n+1} A_k) = \tilde{\mu}((\bigvee_{k=1}^{n} A_k) \vee A_{n+1}) = \tilde{\mu}(\bigvee_{k=1}^{n} A_k) \oplus \tilde{\mu}(A_{n+1}) =$$
$$= \bigoplus_{k=1}^{n} \tilde{\mu}(A_k) + \tilde{\mu}(A_{n+1}) = \bigoplus_{k=1}^{n+1} \tilde{\mu}(A_k).$$

Now let us consider a T_M-disjoint sequence $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}_{\tilde{m}^*}$ such that $\bigvee_{n=1}^{\infty} A_n \in \mathcal{A}_{\tilde{m}^*}$. To prove the inequality

$$\tilde{\mu}(\bigvee_{n=1}^{\infty} A_n) \leq \bigoplus_{n=1}^{\infty} \tilde{\mu}(A_n)$$

we take T_M-disjoint sequences $(B_k^n)_{k\in\mathbb{N}} \in \tilde{\wp}$ such that $A_n \leq \bigvee_{k=1}^{\infty} B_k^n, n \in \mathbb{N}$. Then $\bigvee_{n=1}^{\infty} A_n \leq \bigvee_{n=1}^{\infty} \bigvee_{k=1}^{\infty} B_k^n$ and hence

$$\tilde{m}^*(\bigvee_{n=1}^{\infty} A_n)) \leq \bigoplus_{n=1}^{\infty} \bigoplus_{k=1}^{\infty} \tilde{m}(B_k^n).$$

Taking into account that this inequality holds independent of the choice of sequences $(B_k^n)_{k\in\mathbb{N}}$ we obtain

$$\tilde{m}^*(\bigvee_{n=1}^{\infty} A_n) \leq \bigoplus_{n=1}^{\infty} \tilde{m}^*(A_n).$$

Now we will prove inverse inequality. For a given $n \in \mathbb{N}$ we have

$$\tilde{\mu}(\bigvee_{k=1}^{n} A_k) \leq \tilde{\mu}(\bigvee_{k=1}^{\infty} Ak).$$

and

$$\tilde{\mu}(\bigvee_{k=1}^{n} A_k)) = \bigoplus_{k=1}^{n} \tilde{\mu}(A_k),$$

hence

$$\bigoplus_{k=1}^{n} \tilde{\mu}(A_k) \leq \tilde{\mu}(\bigvee_{k=1}^{\infty} A_k).$$

Finally by taking supremum we obtain

$$\bigoplus_{k=1}^{\infty} \tilde{\mu}(A_k) \leq \tilde{\mu}(\bigvee_{k=1}^{\infty} A_k).$$

□

As the final result, by extension of a crisp measure ν we obtain L-fuzzy valued T_M-countably additive measure $\tilde{\mu} : \mathcal{A}_{\tilde{m}^*} \to \mathbb{R}_+(L)$ such that

(i) $\tilde{\mu}/\tilde{\wp} = \tilde{m}$;

(ii) $\tilde{\mu}/\Phi = \nu$.

The last equality means that for every $M \in \Phi$ it holds $\tilde{\mu}(A(M, 1_L)) = z_{\nu(M)}$.

References

[1] Ruzha V., Asmuss S., *A Construction of a Fuzzy Valued Measure Based on Minimum t-norm.* In: New Dimensions in Fuzzy Logic and Related Technologies. Proceedings of the 5th EUSFLAT Conference, Ostrava, 2007, 175-178.

[2] Klement E. P., Butnariu D., *Triangular norm based measures.* In: Endre Pap (Hrsg.), *"Handbook of Measure Theory"*, Kapitel 23, Seite(n) 947-1010, Elsevier, Amsterdam, 2002.

[3] Hutton B., *Normality in fuzzy topological spaces*, J.Math.Anal.Appl., 50 (1975), 74-79.

[4] Gantner T.E., Steinlage R.C., Warren R.H. *Compactness in fuzzy topological spaces*, J.Math.Anal.Appl., 62 (1978), 547-562.

[5] Rodabaugh S.E., *Fuzzy addition and the L-fuzzy real line*, Fuzzy Sets and Systems, 8 (1982), 39-52.

[6] Asmuss S., Šostak A., *Extremal problems of approximation theory in fuzzy context*, Fuzzy Sets and Systems, 105 (1999), 249-257.

[7] Asmuss S., Šostak A., *On central algorithms of approximation under fuzzy information*, Fuzzy Sets and Systems, 155 (2005), 150-163.

[8] Halmosh P., *Measure theory*, Moscow, IL (1953) (in russian).

ISBN: 978-989-95079-6-8

Distributed Genetic Tuning of Fuzzy Rule-Based Systems

Ignacio Robles, Rafael Alcalá, José Manuel Benítez and Francisco Herrera

Department of Computer Science and Artificial Intelligence
University of Granada, Spain
Email: ignaciorobles@gmail.com, {alcala, J.M.Benitez, herrera}@decsai.ugr.es

***Abstract*— *The tuning of Fuzzy Rule Base-Systems is necessary to improve their performance after the extraction of rules. This optimization problem can become a hard one when the size of the considered system in terms of the number of variables, rules and data samples is big. To alleviate this growth in complexity, we propose a distributed genetic algorithm which explotes the nowadays available parallel hardware (multicore microprocessors and clusters). The empirical performance in solution quality and computing time is assessed by comparing its results with those from a highly effective sequential tuning algorithm. Both methods are applied for the modeling of four well-known regression problems.**

***Keywords*—** 2-tuples, Distributed Genetic Algorithms, Fuzzy Rule-based Systems Tuning, Lateral Tuning

1 Introduction

Fuzzy rule based-systems (FRBS) have become a wide choice when addressing modeling and system identification problems. An essential component in these systems is the fuzzy rule base, together with the data base compose the knowledge base. The construction of this knowledge base is a key step for obtaining a correct system.

It is very difficult for human beings to obtain appropiate rules when dealing with real-world complex problems with many variables and where the necessary number of rules is high. When an expert determines the rule set for a determined problem, generally it will not be the optimal set in terms of performance. Performance is an important goal design for whatever the system is intended to be used: either with approximation purposes or when searching for an interpretable system. To cope with this problem a refining process that adjusts the system is required. This process is widely known as *tuning*.

Classically, tuning processes involve changing the shape of the Membership Functions (MFs) associated to the labels in the database so that the best cooperation among rules is reached. However as the number of variables and rules increases, tuning methods show poor performance due to the growing complexity of the search space. Moreover, the comptuing time consumed by these approaches grows with the complexity of the search space which would result in a procedures that are not useful in practice.

Nowadays parallel hardware and software has become very affordable. They are broadly available which makes them perfect to deal with complex search spaces in order to improve the poor performance achieved with classical tuning approaches. Clear examples in this line are multicore processors and linux clusters.

In this paper, we address the tuning problem and present a distributed method for lateral FRBS tuning. The paper is structured as follows: in the second section of the paper the lateral tuning of FRBSs problem is stated and an efficient sequential specialized algorithm is reviewed. The third section describe our proposal for the distributed tuning for FRBS. An empirical evaluation of the distributed algorithm is presented in the fourth section. Finally some conclusions and future work is commented in the last section.

2 Lateral Tuning of FRBSs

The tuning of FRBS is a problem long studied by researchers in the community [1, 2]. A quite efficient procedure fuzzy systems tuning was presented recently by Alcalá *et al.* [3]. We choose that as a reference to compare the perfomance of our proposal with. We describe this method briefly in this section.

2.1 Lateral Tuning: The Linguistic 2-Tuples Representation

In [3], a new procedure for FRBSs tuning was proposed. It is based on the linguistic 2-tuples representation scheme introduced in [4], which allows the lateral displacement of the support of a label and maintains the interpretability at a good level. This proposal introduces a new model for rule representation based on the concept of symbolic translation [4]. The symbolic translation of a label is a number in [-0.5, 0.5), expressing this interval the domain of a label when it is moving between its two adjacent lateral labels (see Figure 1.a). Let us consider a generic linguistic fuzzy partition $S = s_0, \ldots, s_{L-1}\}$ (with L representing the number of labels). Formally, we represent the symbolic translation of a label s_i in S by means of the 2-tuple notation,

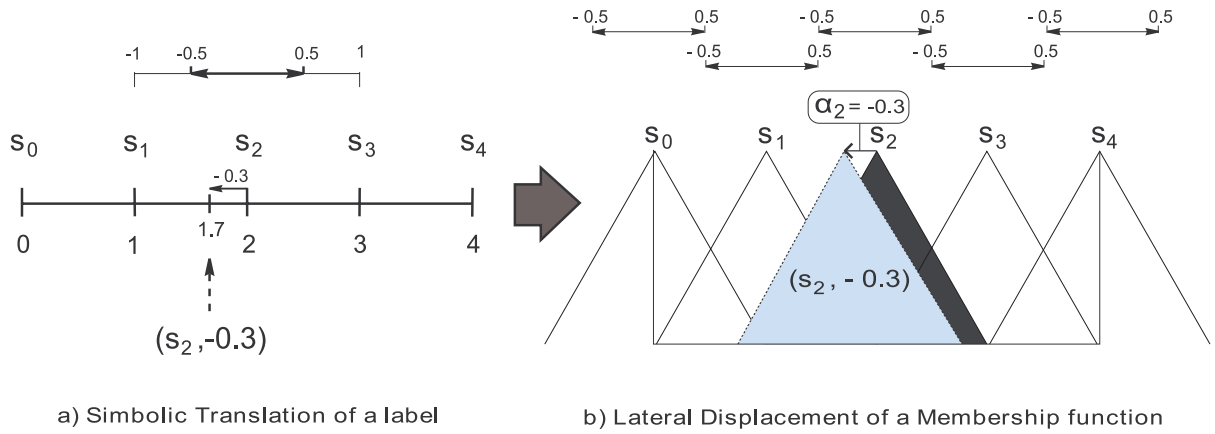
$$(s_i, \alpha_i), \;\; s_i \in S, \; \alpha_i \in [-0.5, 0.5).$$

The symbolic translation of a label involves the lateral variation of its associated MF. Figure 1 shows the symbolic translation of a label represented by the 2-tuple $(s_2, -0.3)$ together with the associated lateral variation.

In the context of FRBSs, the linguistic 2-tuples could be used to represent the MFs comprising the linguistic rules. This way to work, introduces a new model for rule representation that allows the tuning of the MFs by learning their respective lateral displacements. With respect to the classic tuning, usually considering three parameters in the case of triangular MFs, this way to work involves a reduction of the search space that eases a fast derivation of optimal models, improving the convergence speed and avoiding the necessity of a large number of evaluations.

In [3], two different rule representation approaches have been proposed, a global approach and a local approach. The global approach tries to obtain more interpretable models,

ISBN: 978-989-95079-6-8



a) Simbolic Translation of a label

b) Lateral Displacement of a Membership function

Figure 1: Symbolic Translation of a Label and Lateral Displacement of the associated MF

while the local approach tries to obtain more accurate ones. In our case, tuning is applied at the level of linguistic partitions (global approach). By considering this approach, the label s_i^v of a variable v is translated with the same α_i^v value in all the rules where it is considered, i.e., a global collection of 2-tuples is used in all the fuzzy rules. Notice that from the parameters α_i^v applied to each label we could obtain the equivalent triangular MFs. Thus, an FRBS based on linguistic 2-tuples can be represented as a classic Mamdani FRBS [5]. Refer to [3] for further details on this approach.

2.2 Sequential Algorithm for the Lateral Tuning of FRBSs

In [3], a sequential genetic algorithm was proposed to perform a lateral tuning of previously obtained FRBSs. A short description of this algorithm is given in the next (see [3] for a detailed description).

The used model was the genetic model of CHC [6]. CHC makes use of a "Population-based Selection" approach. N parents and their corresponding offsprings are combined to select the best N individuals to take part of the next population. The CHC approach makes use of an incest prevention mechanism and a restarting process to provoke diversity in the population, instead of the well known mutation operator.

This incest prevention mechanism is considered in order to apply the crossover operator, i.e., two parents are crossed if their hamming distance divided by 2 is over a predetermined threshold, T. Since a real coding scheme is considered, each gene is transformed by considering a Gray Code with a fixed number of bits per gene ($BITSGENE$) determined by the system expert. In our case, the threshold value is initialized as:

$$T = (\#GenesC_T * BITSGENE)/4.0.$$

Following the original CHC scheme, T is decreased by one when the population does not change in one generation. In order to avoid very slow convergence, T is also decreased by one when no improvement is achieved with respect to the best chromosome of the previous generation. The algorithm restarts when T is below zero.

In the following, the components used to design the evolutionary tuning process are explained. They are: DB codification, chromosome evaluation and genetic operators.

2.3 DB Codification

A real coding scheme is considered, i.e., the real parameters are the GA representation units (genes). Let us consider n system variables and a fixed number of labels per variable L. Then, a chromosome has the following form (where each gene is associated to the tuning value of the corresponding label),

$$(\alpha_1^1, \ldots, \alpha_1^L, \alpha_2^1, \ldots, \alpha_2^L, \ldots, \alpha_n^1, \ldots, \alpha_n^L)$$

To make use of the available information, the initial FRBS obtained from an automatic fuzzy rule learning method is included in the population as an initial solution. To do so, the initial pool is obtained with the first individual having all genes with value '0.0', and the remaining individuals generated at random in [-0.5, 0.5).

2.4 Chromosome Evaluation

To evaluate a determined chromosome the well-known Mean Square Error (MSE) is used:

$$\text{MSE} = \frac{1}{2 \cdot N} \sum_{l=1}^{E} (F(x^l) - y^l)^2,$$

with E being the data set size, $F(x^l)$ being the output obtained from the FRBS decoded from the said chromosome when the l-th example is considered and y^l being the known desired output.

2.5 Genetic Operators

The genetic operators considered in CHC are crossover and restarting approach (no mutation is used). A short description of these operators comes next:

- Crossover. The crossover operator is based on the the concept of environments. These kinds of operators show a good behavior in real coding. Particularly, the Parent Centric BLX (PCBLX) operator [7] (an operator based on BLX-α) is considered.
- Restarting. To get away from local optima, this algorithm uses a restart approach [6]. In this case, the best chromosome is maintained and the remaining are generated at random within the corresponding variation intervals [-0.5, 0.5). It follows the principles of CHC [6], performing the restart procedure when threshold L is below zero.

ISBN: 978-989-95079-6-8

3 A distributed genetic algorithm for FRBS tuning

The availability of extremely fast and low cost parallel hardware in the last few years benefits the investigation on new approaches to existing optimization algorithms. The key to these new approaches is achieving not only gains in time, which is somehow inherent to distributed algorithms, but gains in quality of the solutions found.

Distributed Genetic Algorithms (DGA) are excellent optimization algorithms and have proven to be one of the best options when trying to cope with large scale problems and when the classic approaches take too much time to give a proper solution.

One procedure for the parallelization of GA comes from the consideration of spatial separation of populates. Schematically:

```
1. Generate a random population, P.
2. Divide P into n subpopulations: SPi,
   i = 1, ..., n.
3. Define a topology for SP1, ..., SPn.
4. For i = 1 to n do:
4.1. Apply in parallel during FM
     generations the genetic operators.
4.2. Send in parallel NM chromosomes to
     neighbour subpopulations.
4.3. Receive in parallel chromosomes
     from neigbour subpopulations.
5. If stopping criteria is not meet then
   go back to step 4.
```

In this section the DGA [8] used for Fuzzy Rule-based System tuning is described.

3.1 Gradual Distributed Real-Coded Genetic Algorithm used

Gradual Distributed Real-Coded Genetic Algorithms (GDRCGAs) are a kind of heterogeneous DGAs based on real coding where subpopulations apply genetic operators in different levels of exploitation/exploration. This heterogeneous application of genetic operators produce a *parallel multiresolution* which allows a wide exploration of the search space and effective local precision. Due to appropiate connections between subpopulations in order to gradually exploit multiresolution,these algorithms achieve refinement or expansion of the best emerging zones of the search space.

The GDRCGA [8] used for FRBS tuning employs 8 subpopulations in a hypercube topology as seen in Figure 2.

In this topology two important groups of subpopulations can be clearly identified:

1. **Front side**: this side of the hypercube is oriented to explore the search space. In this side, four subpopulations, $E_1, ..., E_4$, apply exploration tuned genetic operators in a clockwise increasing degree.

2. **Back side**: subpopulations in the back side of the hypercube, $e_1, ..., e_4$, apply exploitation oriented genetic operators in a clockwise increasing degree.

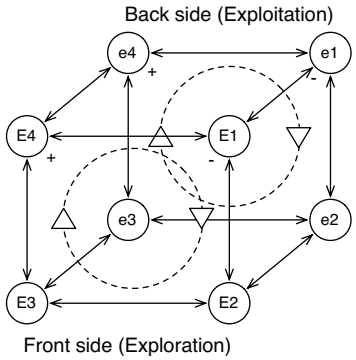


Figure 2: Hypercube topology for GDRCGA

One of the key elements of DGAs is the migration policy of individuals between subpopulations. In this particular model, a *immigration* process [9] is achieved when the best chromosome in every subpopulation abandons it and moves to an immediate neighbour. Due to this immigration policy, three different immigration movements can be identified depending on the subpopulations involved:

1. **Refinement migrations**: individuals in the back side move clockwise to the immediate neighbor, i.e. from e_2 to e_3. Chromosomes in the front side move counterclock from a more exploratory subpopulation to a less exploratory oriented one.

2. **Expansion migrations**: individuals in the back side move counterclock to the immediate neighbor and chromosomes in the front side move clockwise from a less exploratory subpopulation to a more exploratory oriented one, i.e. from E_4 to E_1.

3. **Mixed migrations**: subpopulations from one side of the hypercube exchange their best individual with the counterpart subpopulation in the other side: interchange between E_i and e_i, $i = 1 \dots 4$.

Figure 3 shows the three different migration movements described above.

The genetic operators used in the distributed model are:

- **Selection mechanism**: linear ranking selection (LRS) [10] with stochastic universal sampling [11]. Values of LRS parameter is shown if Table 1.

- **Crossover operator**: BLX-α [12] operator using values for α shown in Table 2.

- **Mutation operator**: non-uniform mutation operator applied with probability $P_{mut} = 0,125$.

Table 1: LRS parameter values for each subpopulation

Exploitation				Exploration			
$+$	$\leftarrow$		$-$	$-$	$\rightarrow$		$+$
e_4	e_3	e_2	e_1	E_1	E_2	E_3	E_4
0,9	0,7	0,5	0,1	0,9	0,7	0,5	0,1

ISBN: 978-989-95079-6-8

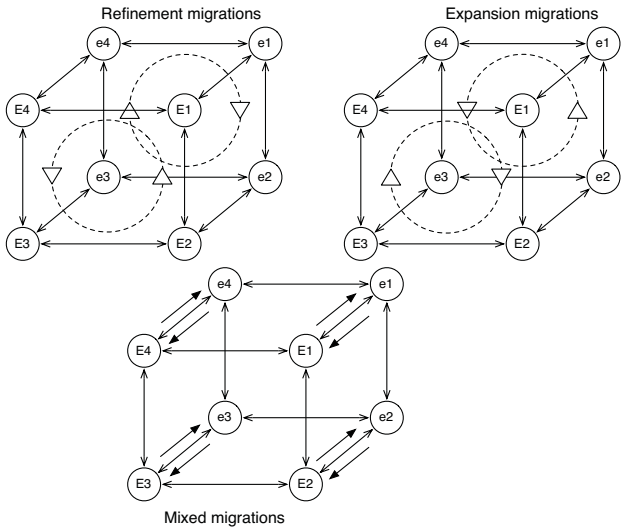


Figure 3: Three different migration movements

Table 2: Values of α for each subpopulation

Exploitation				Exploration			
$+$	$\leftarrow$		$-$	$-$	$\rightarrow$		$+$
e_4	e_3	e_2	e_1	E_1	E_2	E_3	E_4
0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8

As stated in [8], the frequency in which migration movements occur is crucial to avoid the classic withdraws of DGAs: the conquest and noneffect problems. In order to reduce the negative effect of these problems, immigrants stay in the receiving subpopulations for a brief number of generations. Besides, a restart operator is used to avoid stagnation of the search process. This restart operator randomly reinitializes all subpopulations if non-significant improvement of the best element is achieved for a number of generations. Also an elitism strategy is used in order to keep the best adapted individual of every subpopulation.

4 Empirical evaluation

In this section we describe the empirical evaluation we performed in order to assess the merits of our proposal.

We consider the tuning of FRBSs constructed for the modeling of some well-known regression problems. We have compared the quality of the tuning performed by the sequential algorithm and our proposal. We have selected four problems, some of their salient features are shown in Table 3 displayed in increasing complexity order.

Table 3: Data sets used to evaluate the algorithm

Data set	Variables	Instances
Electrical Maintenance	5	1056
Trasury	16	1049
Weather-Izmir	10	1461
Abalone	8	4177

A five-fold cross-validation approach has been used. So five runs with different independent test sets have been carried out for each problem. The performance of a fuzzy rule is measured as mean squared error (MSE) over the test set.

Our proposed distributed algorithm is compared with a specialized sequential genetic algorithm (CHC) [3] in terms of quality of the solutions achieved (MSE) as well as in time. As a starting point we have used the results achieved by applying Wang & Mendel fuzzy rule learning method [13]. These initial results are shown in Table 4.

Table 4: Initial results using Wang & Mendel fuzzy rule learning method

Dataset	Training	σ_{tra}	Test	σ_{test}
Electrical M.	57605.83	2840.78	57934.25	4732.66
Treasury	1.636	0.121	1.632	0.182
Weather-Izmir	6.944	0.720	7.368	0.909
Abalone	3.341	0.130	3.474	0.247

Generally when comparing a distributed or parallel approach with some other sequential algorithm an interesting measure is the execution time gain ratio. This ratio could be defined as follows:

$$R = \frac{T_{seq}}{T_{dist}} \quad (1)$$

where T_{seq} is the time spent by the sequential algorithm and T_{dist} is the execution time of the distributed approach. The higher the value of R, the better. Time gain ratio values obtained in the empirical experimentation are shown in Table 5.

An interesting point to which to pay attention is the *evolution* of the mean squared error as the number of evaluations increases. So, three different numbers of evaluations have been choosen: 10000, 25000 and 50000 evaluations per run. The results in terms of quality of the solutions attained are shown in Table 6. An important fact to notice is that the mean squared error achieved with the distributed method is lower than the error obtained with the specialized genetic algorithm in all data sets at 50000 evaluations. The distribuited also obtains good results with fewer iterations in some cases (e.g. Electrical Maintenance, 25000 it.), but clearly its real effectiveness will be reached when the computation load is higher. Actually, we are running further experiments with datasets of higher complexity.

Table 5: Time gain ratio with 50000 evaluations

Data set	T_{seq}	T_{dist}	R
Electrical Maintenance	187,3	391,6	0,479
Trasury	525,3	739,7	0,710
Weather-Izmir	849,8	867,1	0,980
Abalone	1980,9	942,5	2,101

As shown in Table 5, the time gain ratio, R, increases with the problem complexity. In the less complex data sets the ratio obtained is substantially low because the sequential specialized genetic algorithm is very fast and the time spent in communications of the distributed approach slows it down in comparision. As the complexity of the data set increases the time gain ratio also increases, showing that the distributed approach in the most complex data set is more than two times faster than the sequential specialized genetic algorithm.

ISBN: 978-989-95079-6-8

Table 6: Mean squared errors in training and test sets. The winner for each pair of training is *in italics*. The winner for each pair of test is **boldfaced**

		CHC		GDRCGA	
Data set	Evaluations	Training	Test	Training	Test
Electrical Maintenance	10000	*2.59363671E+04*	2.92591821E+04	2.65539710E+04	**2.89024830E+04**
	25000	2.48690100E+04	2.80510895E+04	*2.39248797E+04*	**2.67720415E+04**
	50000	2.46214328E+04	2.78282761E+04	*2.26682075E+04*	**2.54097540E+04**
Abalone	10000	*2.61355003E+00*	2.79981355E+00	2.65916770E+00	**2.79026550E+00**
	25000	2.60333453E+00	2.79298130E+00	*2.59992700E+00*	**2.76143590E+00**
	50000	2.60303744E+00	2.79117626E+00	*2.57035010E+00*	**2.75904570E+00**
Weather-Izmir	10000	*1.68875432E+00*	**1.89318352E+00**	1.89195950E+00	1.95458830E+00
	25000	*1.64117336E+00*	**1.86996710E+00**	1.66238700E+00	1.87669540E+00
	50000	1.64010963E+00	1.86891124E+00	*1.57019250E+00*	**1.86195430E+00**
Treasury	10000	*1.71238672E-01*	**1.86722425E-01**	2.12486800E-01	2.16882700E-01
	25000	*1.33618274E-01*	**1.50895419E-01**	1.42194200E-01	1.67407400E-01
	50000	1.20604483E-01	1.37784224E-01	*1.15845500E-01*	**1.31803000E-01**

5 Conclusions and final remarks

We have developed and presented a very promising distributed algorithm for FRBSs lateral tuning that achieves better results than a specialized genetic algorithm. Due to its distributed nature and consequently the spatial separation implied, it needs more evaluations to converge than a classic sequential algorithm. It always presents the same behaviour in comparision to the specialized sequential algorithm: with a small number of evaluations it offers a higher error than the sequential approach but when the evaluations are high it gives better quality solutions.

Empirical results show that when the complexity of the problem grows, our distributed method takes adventage of the large computing times and converges to a better solution in less time. This property makes the distributed tuning algorithm very convenient when dealing with large scale data sets.

The distributed algorithm takes longer than the sequential algorithm when dealing with small size data sets mainly due to two reasons: interprocess communication in the distributed approach implies additional execution time which can not be paralelized and the specialized algorithm is optimized for small size data sets where the search space is not too complex.

Since execution time and quality of the results are two properties always in conflict somehow, our approach is very convenient since it can be graduated in order to achieve faster execution times with a small cost in quality and viceversa.

We expect to achieve great execution time gains when applying the distributed FRBSs tuning algorithm with large scale data sets with a non significant loss of quality.

Acknowledgment

Supported by the Spanish Ministry of Education and Science under grant no. TIN2008-06681-C06-01.

References

[1] R. Alcalá, J. Alcalá-Fdez, M. J. Gacto, and F. Herrera. Rule base reduction and genetic tuning of fuzzy systems based on the linguistic 3-tuples representation. *Soft Computing*, 11(5):401–419, 2007.

[2] F. Herrera, M. Lozano, and J. L. Verdegay. Tuning fuzzy logic controllers by genetic algorithms. *International Journal of Approximate Reasoning*, 12:299–315, 1995.

[3] R. Alcalá, J. Alcalá-Fdez, and F. Herrera. A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. *IEEE Trans. Fuzzy Syst.*, 15(4):616–635, 2007.

[4] F. Herrera and L. Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Trans. Fuzzy Syst.*, 8(6):746–752, 2000.

[5] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7:1–13, 1975.

[6] L. J. Eshelman. The CHC adaptive search algorithm: How to have safe serach when engaging in nontraditional genetic recombination. In G.J.E. Rawlin, editor, *Foundations of genetic Algorithms*, volume 1, pages 265–283. Morgan Kaufman, 1991.

[7] F. Herrera, M. Lozano, and A. M. Sánchez. A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study. *International Journal of Intelligent Systems*, 18:309–338, 2003.

[8] F. Herrera and M. Lozano. Gradual distributed real-coded genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 4(1):43–63, April 2000.

[9] B. Kröger, P. Schwenderling, and O. Vornberger. Parallel genetic packing on transputers. *Parallel Genetic Algorithms: Theory and Applications: Theory Applications*, pages 151–186, 1993.

[10] J. E. Baker. Adaptative selection methods for genetic algorithms. In MA Ed. L. Erlbraum Associates: Hillsdale, editor, *Proceedings of the First International Conference on Genetic Algorithms and their Applications*, pages 101–111, 1985.

[11] J. E. Baker. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the 2nd International Conference on Genetic Algorithms (ICGA'87)*, 1987.

[12] L.J. Eshelman and J.D. Schaffer. Real-coded genetic algorithms and interval-schemata. *Foundations of Genetic Algorithms*, 2:187–202, 1993.

[13] L. X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Trans. Syst. Man and Cybernetics*, 22(6), 1992.

ISBN: 978-989-95079-6-8

Forecasting Exchange Rates: A Neuro-Fuzzy Approach

Meysam Alizadeh[1] Roy Rada[2] Akram Khaleghei Ghoshe Balagh[3] Mir Mehdi Seyyed Esfahani[1]

1 Department of Industrial Engineering, Amirkabir University of Technology (Tehran Polytechnic)
Tehran, Iran
2 Department of Information Systems, University of Maryland Baltimore County
Baltimore, USA
3 Researches & Planning Center of Bank Mellat
Tehran, Iran
Email: alizadeh.m@ieee.org, rada@umbc.edu, a.khaleghei@bankmellat.ir, msesfahani@aut.ac.ir,

Abstract—*This paper presents an adaptive neuro-fuzzy inference system (ANFIS) for USD/JPY exchange rates forecasting. Previous work often used time series techniques and neural networks (NN). ANFIS can be used to better explain solutions to users than completely black-box models, such as NN. The proposed neuro-fuzzy rule based system applies some technical and fundamental indexes as input variables. In order to generate membership functions (MFs), we make use of fuzzy clustering of the output space. The neuro-fuzzy model is tested with 28 candidate input variables for both currencies. For the purpose of comparison, Sugeno-Yasukawa model, feedforward multi-layer neural network, and multiple regression are benchmarked. The comparison demonstrates that the presented algorithm shows its superiority in terms of prediction error minimization, robustness and flexibility.*

Keywords— ANFIS Network, Exchange Rates, Forecasting, Neuro-Fuzzy Systems, Nonlinear Identification.

1 Introduction

The Efficient Market Hypothesis ([1], [2], [3]) states that one cannot predict the behavior of efficient markets, as the markets follow a random walk. However, other studies ([4], [5]) found evidences against the Efficient Market Hypothesis. Exchange rate prediction is one of the most challenging applications of modern time series forecasting. To become an efficient dealer in foreign exchange market, one must be conversant with the factors that are responsible for a currency to appreciate or depreciate.

Methods that have been proposed for currency exchange rates forecasting fall into three categories: (i) econometric methods, (ii) time series methods, and (iii) soft computing methods. The difficulty in predicting currency exchange rates, due to their high volatility and complexity, has long been a concern in international financial markets, as many econometric methods are unable to produce significantly better forecasts than the random walk model [6]. Time series methods have their limitations for multidimensional time series with mutual non-linear dependencies. Most recently, researchers have applied soft computing tools such as multilayer Artificial Neural Networks (ANN), Fuzzy Logic (FL), and their hybrids to exchange rate forecasting.

ANN support time-series analysis and forecasting and multivariate analysis. Medeiros et al [7] presented and compared alternatives to model and forecast monthly exchange rates time series. Medeiros et al [7] simulated ANN, neuro-coefficient smooth transition autoregressive, linear autoregression, and random walk models. Kamruzzaman and Sarker [8] have investigated ANN modeling of foreign exchange rates using three learning algorithms, namely, Standard Back Propagation, Scaled Conjugate Gradient, and Back Propagation with Bayesian Regularization. Majhi et al [9] proposed two low complexity ANN exchange rate prediction models for US Dollar to British Pound, Indian Rupees, and Japanese Yen exchange rates.

The literature has shown fuzzy sets as a superior modeling tool for many problems. The statistical fuzzy interval NN is proposed by [10] to perform knowledge discovery and predict currency exchange rate. To overcome the difficulty in finding matching rules for forecasting in the fuzzy time series model, [11] proposed to use the Euclidean distance between two fuzzy logic relationships as a metric for selecting matching rules and applied their model to exchange rate time series prediction.

At the computational level, a fuzzy system is a layered structure, similar to an ANN of the radial basis function type [12]. In order to optimize parameters in a fuzzy system, ANN learning algorithms can be employed. This neuro-fuzzy modeling approach can be better used to explain solutions to users than completely black-box models, such as ANN [13].

For the sake of clarity and completeness, section 2 gives a brief overview of the adaptive neuro-fuzzy inference system (ANFIS) approach. Section 3 systematically discusses neuro-fuzzy systems modeling. Section 4 carefully applies neuro-fuzzy modeling to US Dollar to Japanese Yen exchange rate forecasting. Finally, conclusions are given in Section 5.

2 Adaptive Neuro-Fuzzy Inference System

The Takagi-Sugeno-Kang (TSK) model uses fuzzy logic with crisp functions in consequent that is convenient for complex applications [14]. TSK systems are widely used in the form of a neuro-fuzzy system called Adaptive Neuro-Fuzzy Inference System (ANFIS). The structure of ANFIS and its learning algorithms are described in the following:

ISBN: 978-989-95079-6-8

2.1 ANFIS Structure

An ANFIS is a fuzzy inference system that can be trained to model some collection of input/output data. The training module allows the system to tune its parameters to learn the input/output relationships hidden in the data set. ANFIS uses two approaches: ANN and fuzzy modeling. By composing these two approaches, a suitable reasoning in quality and quantity might be achieved [15]. In ANFIS fuzzy logic is used to determine decision surfaces rather than to determine uncertainty associated with particular linguistic terms [16]. The rule-based representation of neuro-fuzzy systems offers transparency.

For pedagogical purposes, one can imagine a fuzzy inference system with two inputs x and y and one output z. The equivalent ANFIS architecture (Type-3 ANFIS) is shown in Fig. 1. The node functions in the same layer are from the same function family. The first layer implements a fuzzification, the second layer executes the T-norm of the antecedent part of the fuzzy rules, the third layer normalizes the membership functions, the fourth layer calculates the consequent parameters, and finally the last layer computes the overall output as the summation of all incoming signals. The feed forward equations of this ANFIS are as follows:

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \tag{1}$$

$$\overline{w}_i = \frac{w_1}{w_1 + w_2}, i = 1,2. \tag{2}$$

$$\begin{cases} f_i = p_1 x + q_1 y + r_1 z \\ f_2 = p_2 x + q_2 y + r_2 z \end{cases} \Rightarrow \tag{3}$$

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2} = \overline{w}_1 f_1 + \overline{w}_2 f_2 \tag{4}$$

where,

- x is the input to node i,
- $\mu_{A_i}(x)$ is the node i node function,
- A_i is the linguistic label associated with node functions,
- w_i is the firing strength of the ith rule,
- $\overline{w}_i$ is the ratio of the ith rule's firing strength to the sum of all rules' firing strength,
- $\{p_i, q_i, r_i\}$ is the parameter set, and
- f_i is the consequent value.

Note that the network's output y is nonlinear in the weights w**.** The training of this ANN is thus a *nonlinear optimization problem* to which various methods can be applied [13].

2.2 Learning Algorithms

The neuro-fuzzy inference system is optimized by adapting the antecedent parameters and consequent parameters so that a specified objective function (usually a difference between the model output and the actual output) is minimized. A number of methods have been proposed for learning rules. For example, Mascioli, Varazi, and Martinelli [17] have proposed merging of Min-Max and ANFIS models to determine the optimal set of fuzzy rules. Jang and Mizutani [18] have presented an application of the Lavenberg-Marquardt method, which is essentially a nonlinear least-squares technique, for learning in an ANFIS network. Tang, Quek, and Ng [19] proposed a hybrid system combining a Fuzzy Inference System and Genetic Algorithms to tune the parameters in the TSK fuzzy ANN.

Jang [20] proposed methods to update the ANFIS parameters involving gradient descent and Least Square Error (LSE). High complexity is one of these methods' features. Several popular training algorithms for tuning parameters of ANFIS membership functions are compared in [21]. In this paper we use the hybrid learning algorithm proposed in [20] which is a combination of least square estimation and backpropagation algorithms.

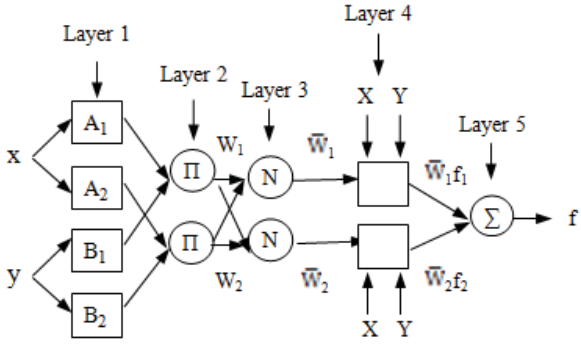


Figure 1: The equivalent ANFIS structure

3 Designing a Systematic Neuro-Fuzzy System

The two basic steps in neuro-fuzzy systems modeling are *system identification* and *fuzzy reasoning*. In the system identification stage, the significant input variables are determined, the fuzzy if–then rules are generated, and the parameters of the model, such as the number of clusters, the level of fuzziness, and the operators to be used in the reasoning, are selected. The second, *fuzzy reasoning*, is used to infer new knowledge from the identified rule base [22].

In fuzzy system modeling, the relationships between variables are represented with *if-then* rules with imprecise predicates. The basic steps are:

1) Fuzzy clustering of the output,
2) Input selection,
3) Rule base construction,
4) Tuning the parameters of membership function of input and output variables, and
5) Inference.

3.1 Fuzzy Clustering of the Output

To determine the number of rules of the initial ANFIS, one should choose the optimum number of clusters. For this purpose, a validity index which is proposed by [23] and modified by [24] is used. In other words, one minimizes:

$$V_{FNT}(U,V;X) = \frac{2}{c(c-1)} \sum_{p \neq q}^{c} S_{rel}(A_P, A_q) \tag{5}$$

The optimal number of clusters is obtained by minimizing $V_{\mathrm{FNT}}(U, V; X)$ over the range of c values: 2, …, $c_{\max}$; where

ISBN: 978-989-95079-6-8

$S_{rel}(A_P, A_q)$ is the relative similarity between two fuzzy sets A_p and A_q and is defined as:

$$S_{rel}(A_{P,}, A_q) = \sum_{i=1}^{n} S_{rel}(x_j : A_p, A_q) h(x_j) \quad (6)$$

where,

$$h(x_j) = -\sum_{p=1}^{c} u_{A_p}(x_j) \log(u_{A_p}(x_j)) \quad (7)$$

And $S_{rel}(x_j : A_p, A_q)$ is the relative similarity between two fuzzy sets A_p and A_q at x_j which is defined as:

$$S_{rel}(x_j : A_p, A_q) = \frac{f(x_j : A_p \cap A_q)}{f(x_j : A_p \cap A_q) + f(x_j : A_p - A_q) + f(x_j : A_q - A_p)} \quad (8)$$

Here, $h(x_j)$ is the entropy of datum x_j and $u_{A_p}(x_j)$ is the membership value with which x_j belongs to the cluster A_p.

After determining the optimum number of clusters, Gustafson-Kessel (GK) clustering algorithm is used for generating membership functions.

3.2 Input Selection

The performance of non-linear identification techniques is often determined by the appropriateness of the selected input variables and the corresponding time lags. High correlation coefficients between candidate input variables in addition to a non-linear relation with the output signal induces the need for an appropriate input selection methodology [25].

For variables selection, the Sugeno and Yasukawa [26] method is used. They proposed a combinatorial approach in which all possible combinations of input candidates are considered. For each combination, they built two fuzzy models based on two separated sets of data and calculated a performance index called "Regularity Criterion" (RC). After that a combination of input variables is chosen which has the minimum value of the performance index.

3.3 Rule Base Construction

There are three alternative strategies for incorporating the fuzzy clustering during rule base construction; cluster the:

1. output space and obtain the fuzzy membership functions based on the projections of the output clusters onto the input space [26].
2. input space, relating the output variables to each input cluster based on the degree of possibility [27].
3. joined input and output spaces and then project these multidimensional clusters to the separate input and output spaces [28].

Kilic, Uncu, and Turksen [22] compared these three approaches. This paper uses the first approach.

3.4 Membership Parameters Tuning

ANFIS has two kinds of parameters that needed to be trained: the antecedent parameters and the premises parameters. In this study, Gaussian membership functions are located in the antecedent part:

$$\mu_{A_i}(x) = \exp\left\{-\left[\left(\frac{x - c_i}{a_i}\right)^2\right]^{b_i}\right\} \quad (9)$$

It has three types of parameters: $\{a_i, b_i, c_i\}$, where a_i is the variance, b_i is the crossover slop and c_i is the center of MFs.

4 Exchange Rates Forecast Modeling

The authors used ANFIS in forecasting US Dollar/Japanese Yen exchange rates. While other researchers primarily used a few technical inputs ([8], [9], [10], [29], [30], [31]), (see Table 1) this work uses significant fundamental and technical inputs.

Table 1: List of some previous inputs data used in exchange rate forecasting

Authors	Model	Inputs
Tenti (1996)	Recurrent Neural Network	Compound returns of the last n periods (n=1, 2, 3, 5, 8) The running standard deviation of the k last periods (k=13, 21, 34) Average directional movement index (ADX) Trend movement index (TMI) Rate of change (ROC) Ehlers leading indicator (ELI)
Lisi and Schiavo (1999)	Chaotic models and NN	Monthly exchange rate of Franc, Deutschmark, Lira, and Pound- all against US Dollar
Kamruzzaman and Sarker (2004)	ANN	Moving average of one week, two weeks, one month, one quarter, half year closing rate Last week's closing rate
Preminger and Franck (2007)	Regression	Monthly exchange rate of Yen and Pound against the US dollar
Zhang and Wan (2007)	Fuzzy interval NN	Three, two, and one weeks ago rate of exchange
Majhi, Panda, and Sahoo (2009)	ANN	Normalized rate on the first day of a month Mean of the monthly rate Variance of the monthly rate

The candidate inputs are shown in Table 2. Similar to [32] the inputs are divided into categories. However, in each category, this study used some other indices than [32]. Additionally, some of the most popular technical input variables are used. As a result, 28 fundamental and technical candidate input variables, among which there exist 20 fundamental indices for both countries and the remaining ones have been allocated to technical indices, are fed to ANFIS. The full sample compromises daily observation for the Jan-2001 to Aug-2008 period from the Reuters 3000 Xtra Hosted Terminal Platform, where 60% of data points are used for training and the rest for testing the model.

After collecting candidate input data, the Sugeno-Yasukawa [26] input selection method is used. The method selected the following input variables:

- US Federal Reserve Bank interest rate,
- US M2 Money Supply: it technically defined as sum of M1, savings deposits, small denomination time

ISBN: 978-989-95079-6-8

deposits (where small is less than $100,000), and retirement account, where M1 is the sum of the tender that is held outside banks, travelers checks, checking accounts (but not demand deposits), minus the amount of money in the Federal Reserve float,

- West Texas Intermediate oil price,
- Rate of Change (10 days) of the USD/JPY time series,
- Momentum (5 days) of the USD/JPY time series, and
- Stochastic K% (5 days) of the USD/JPY time series.

Table 2: Candidate Input Data

No.	Factor	Indexes
1	Economic Activities	Gross domestic product, Industrial Production
2	Price	Consumer price index, Producer price index
3	Interest Rates	Short term interest rate (overnight, 1 week), Federal rate
4	Money Supply	M1, M2, M3
5	Trade Balance	Trade balance, BOP
6	Employment	Unemployment rate, Nonfarm payroll
7	Personal Consumption	Retail sale, Personal income
8	Oil	Brent, WTI oil
9	Stock	Dow Jones, Nikkei
10	Technical Indexes	Stochastic K%(5, 10 days), ROC(5, 10 days), Momentum(5, 10 days), William R%(5, 10 days)

The corresponding RC values are shown in Fig. 2. Next, based on an introduced Cluster Validity Index, we determine that the optimum number of clusters is 7 (see Fig. 3). GK fuzzy clustering algorithm is applied for initialization of antecedent membership functions. Then ANFIS with hybrid learning algorithm (combination of least square estimation and backpropagation algorithms) is used for parameters tuning. We use Mamdani-style inference, min–max operators, and centroid defuzzification method. The rules that ANFIS generated make sense relative to the economics of the foreign exchange market, as shown in Fig. 4.

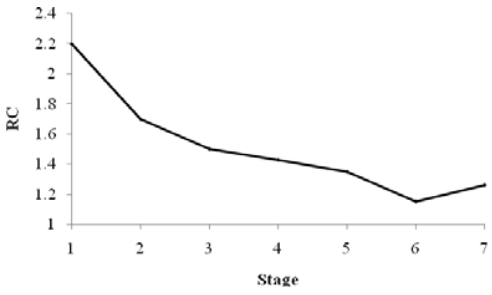


Figure 2: Behavior of RC in proposed model

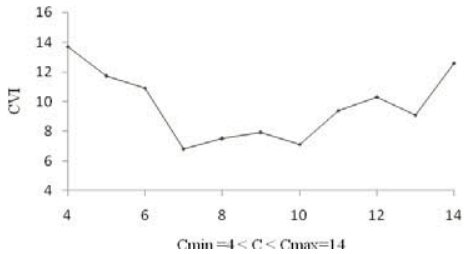


Figure 3: Identification of the optimum number of clusters

To further validate this approach for currency exchange rates forecasting, the results were compared to the results with a) the Sugeno-Yasukawa [26] approach, b) a feedforward multi-layer neural network, and c) multiple regression. The comparison uses the root mean square of the prediction error (RMSE) and mean error of the prediction (BIAS). The comparison results for training and testing dataset are shown in Table 3, 4, 5, and 6. The characteristics of the mentioned test models are described next.

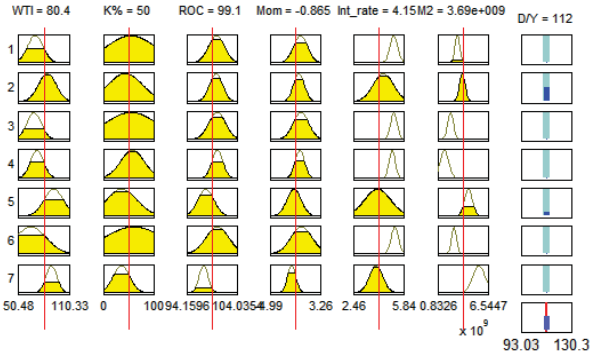


Figure 4: Fuzzy rule base of USD/JPY exchange rates forecasting

4.1 Sugeno-Yasukawa Approach

This Sugeno-Yasukawa [26] approach produces a fuzzy model with 6 rules, 6 inputs, and 1 output. The inputs are the same as with the ANFIS model. Other aspects of the approach include a Mamdani-style inference, min–max operators, and centroid defuzzification.

4.2 Feedforward Multi-layer Neural Network

A 10×5×1 feedforward multi-layer network with gradient descent learning algorithm is used to forecasting US Dollar against Japanese Yen currency exchange rates. A tangent sigmoid activation function is used in each node. The tests were performed for a maximum iteration number of 500.

4.3 Multiple Regression

The multiple regression analysis with Matlab ® is used for prediction. The regression equation is:

$$y = 1.9164\, x_1 - 0.00001\, x_2 - 0.0602\, x_3 + 1.7423\, x_4 - 0.2103\, x_5 - 0.0208\, x_6 \quad (10)$$

The output is computed by the above equation and the modeling performance measure is obtained.

Table 3: RMSE of different models for training set

Model	# Rules	# Runs	RMSE		
			Min.	Max.	Ave.
Multiple regression	-	10	6.1002	9.1770	7.2391
Neural network	-	10	1.3055	2.7595	2.1306
Sugeno-Yasukawa	6	10	3.1397	5.6691	3.8723
ANFIS	7	10	1.3805	2.4980	1.7447

ISBN: 978-989-95079-6-8

Table 4: RMSE of different models for testing set

Model	# Rules	# Runs	RMSE		
			Min.	Max.	Ave.
Multiple regression	-	10	6.3060	9.4483	7.9221
Neural network	-	10	1.3820	4.9647	2.8082
Sugeno-Yasukawa	6	10	3.9835	8.2009	4.8290
ANFIS	7	10	1.9284	5.4007	2.6301

Table 5: BIAS of different models for training set

Model	# Rules	# Runs	BIAS		
			Min.	Max.	Ave.
Multiple regression	-	10	1.7193	2.0918	1.7201
Neural network	-	10	0.5311	0.8375	0.7611
Sugeno-Yasukawa	6	10	0.8129	1.1602	0.9069
ANFIS	7	10	0.5591	0.7190	0.6207

Table 6: BIAS of different models for testing set

Model	# Rules	# Runs	BIAS		
			Min.	Max.	Ave.
Multiple regression	-	10	1.7660	2.2183	1.9219
Neural network	-	10	0.5005	1.1020	0.9011
Sugeno-Yasukawa	6	10	0.6291	1.5918	1.4820
ANFIS	7	10	0.5197	0.9518	0.8526

5 Conclusions

The average RMSE and BIAS results are better for the ANFIS approach than for the Sugeno-Yasukawa, ANN, or multiple regression approaches. However, the difference between the ANFIS and ANN average RMSE and BIAS does not appear significant. One might say that the performance of ANFIS and ANN was comparable as regards accuracy. However, with ANN one does not gain the benefit of the semantically meaningful presentation of rules that is possible with ANFIS. Given the choice between two computer systems that produce comparable test accuracies, humans are likely to prefer the system whose logic is transparent.

Future work would explore refined hypotheses in various directions. One hypothesis is that by taking advantage of further information about currency exchange rates, we can improve the initialization of antecedent membership functions and the clustering technique. Another hypothesis is that the application of other learning or optimization algorithms, such as genetic algorithms and particle swarm optimization, would demonstrate interesting variations on the representation and reasoning for the currency exchange problem. By implementing new hybrid learning programs, researchers might gain further insights about how to solve the currency exchange problem. Vast amounts of available data, clear definitions of success in testing, and practical importance are hallmarks of this problem. The problem of forecasting currency exchange rates provides a valuable test bed for soft computing methods, and researchers can thus elaborate the merit and the shortcomings of each method.

References

[1] F. Eugene, The Behavior of Stock Market Prices, *Journal of Business*, 38: 34–105, 1965.

[2] P. Samuelson, Proof That Properly Anticipated Prices Fluctuate Randomly, *Industrial Management Review*, 6: 41–49, 1965.

[3] F. Eugene, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance*, 25: 383–417, 1970.

[4] A.D. Hsieh, Chaos and non-linear dynamics: application to financial markets, *Journal of Finance,* 46:1833–1877, 1991.

[5] G. Tsibouris and M. Zeidenberg, Testing the efficient market hypothesis with gradient descent algorithm, in: A. Refenes (Ed.), *Neural networks in the capital markets*, John Wiley and Sons Ltd., England, 127–136, 1996.

[6] A.S. Chen and M.T. Leung, Regression neural network for error correction in foreign exchange forecasting and trading, *Computers and Operations Research*, 31:1049-1068, 2004.

[7] M.C. Medeiros, Á. Veiga, and C. E. Pedreira, Modeling Exchange Rates: Smooth Transitions, Neural Networks, and Linear Models, *IEEE Transactions on Neural Networks*, 12(4), July 2001.

[8] J. Kamruzzaman and R.A. Sarker, ANN-Based Forecasting of Foreign Currency Exchange Rates, *Neural Information Processing-Letters and Reviews*, 3(2), May 2004.

[9] R. Majhi, G. Panda, and G. Sahoo, Efficient prediction of exchange rates with low complexity artificial neural network models, *Expert Systems with Applications*, 36:181–189, 2009.

[10] Y.Q. Zhang and X. Wan, Statistical fuzzy interval neural networks for currency exchange rate time series prediction, *Applied Soft Computing*, 7:1149–1156, 2007.

[11] Y. Leu, C.P. Lee, and Y.Z. Jou, A distance-based fuzzy time series model for exchange rates forecasting, *Expert Systems with Applications*, doi:10.1016/j.eswa.034, 2008.

[12] J.S.R. Jang and C.T. Sun, Functional equivalence between radial basis function networks and fuzzy inference systems, *IEEE Transactions on Neural Networks*, 4(1): 156–159, 1993.

[13] R. Babuška and H. Verbruggen, Neuro-fuzzy methods for nonlinear system identification. *Annual reviews in control*, 27:73-85, 2003.

[14] R. Alcala, J. Casillas, O. Cordon, and F. Herrera, Learning TSK rule-based system from approximate ones by mean of MOGUL methodology. Granada University of Spain, 2000.

[15] M. Teshnehlab, M.A. Shoorehdeli, and A.K. Sedigh, Novel hybrid learning for tuning ANFIS parameters as an identifiers using fuzzy PSO, *IEEE International Conference on Networking, Sensing and Control, Sanya, China*, 111-116, 2008.

[16] G.S. Atsalakis and K.P. Valavanis, Surveying stock market forecasting techniques, PartII: Soft computing methods. *Expert Systems with Applications*, doi:10.1016/j.eswa.2008.07.006, 2008.

[17] F.M. Mascioli, G.M. Varazi, and G. Martinelli, Constructive Algorithm for Neuro-Fuzzy Networks, Proceedings of the *Sixth IEEE International Conference Fuzzy Systems, Spain*, 1: 459- 464, 1997.

[18] J.S.R. Jang and E. Mizutani, Levenberg-Marquardt Method for ANFIS Learning, *Biennial Conference of the North American Fuzzy Information Processing Society*, 87-91, 1996.

[19] A.M. Tang, C. Quek and G.S. Ng, GA-TSKfnn: Parameters tuning of fuzzy neural network using genetic algorithms. *Expert Systems with Applications,* 29: 769-781, 2005.

[20] J.S.R. Jang, ANFIS: Adaptive-network-based fuzzy inference systems, *IEEE Transactions on Systems, Man & Cybernetics,* 23(3): 665–685, 1993.

[21] M.S. Chen, A Comparative Study of Learning Methods in Tuning Parameters of Fuzzy Membership Functions, *IEEE SMC '99 Conference Proceedings, Tokyo, Japan*, 40-44,1999.

[22] K. Kilic, O. Uncu, and I.B. Turksen, Comparison of different strategies of utilizing fuzzy clustering in structure identification, *Information Sciences,* 177:5153-5162, 2007.

ISBN: 978-989-95079-6-8

[23] Y.I. Kim, D.W. Kim, D. Lee, and K.H. Lee, A cluster validation index for GK cluster analysis based on relative degree of sharing, *Information Sciences,* 168: 225-242, 2004.

[24] M.H. Fazel Zarandi, B. Rezaee, I.B. Turksen, and E. Neshat, A type-2 fuzzy rule-based expert system model for stock price analysis, *Expert Systems with Applications,* 36: 139–154, 2009.

[25] K. Maertens, J.D. Baerdemaeker, and R. Babuška, Genetic polynomial regression as input selection algorithm for non-linear identification, *Soft Computing,* 10:785-795, 2006.

[26] M. Sugeno and T. Yasukawa, A fuzzy logic based approach to qualitative modeling, *IEEE Transactions on Fuzzy Systems,* 1:7–31, 1993.

[27] G. Castellano, A. Fanelli, and C. Mencar, A fuzzy clustering approach for mining diagnostic rules, in: Proceedings of IEEE Conference on Systems, Man and Cybernetics, Washington, USA, 2003.

[28] O. Uncu, I.B. Turksen, and K. Kilic, Localm-fsm: A new fuzzy system modeling approach using a two-step fuzzy inference mechanism based on local fuzziness level, in: Proceedings of *International Fuzzy Systems Association World Congress*, 191–194, 2003.

[29] P. Tenti, Forecasting foreign exchange rates using recurrent neural network. *Applied Artificial Intelligence* 10: 567-581, 1996.

[30] F. Lisi and R.A. Schiavo, A comparison between neural networks and chaotic models for exchange rate prediction. *Computational Statistics & Data Analysis* 30: 87-102, 1999.

[31] A. Preminger and R. Franck. Forecasting exchange rates: A robust regression approach. *International Journal of Forecasting* 23: 71-84, 2007.

[32] K. Izumi and K. Ueda, Phase Transition in a Foreign Exchange Market—Analysis Based on an Artificial Market Approach, *IEEE Transactions On Evolutionary Computation*, 5(5), October 2001.

ISBN: 978-989-95079-6-8

Genetic Learning of Serial Hierarchical Fuzzy Systems for Large-Scale Problems

Alicia D. Benítez and Jorge Casillas[1]

[1] Department of Computer Science and Artificial Intelligence
CITIC-UGR (Research Center on Information and Communications Technology)
University of Granada, E-18071 Granada, Spain
Email: aliciades@gmail.com, casillas@decsai.ugr.es

Abstract— *When we face a problem with a high number of variables by a standard fuzzy system, the number of rules increases exponentially and then the obtained fuzzy system is scarcely interpretable. This problem can be handled by arranging the inputs in hierarchical ways. The paper presents a multi-objective Genetic Algorithm that learns Serial Hierarchical Fuzzy Systems with the aim of coping the curse of dimensionality. By means of an experimental study, we have observed that our algorithm obtains good results of interpretability and precision with problems in which the number of variables is relatively high.*

Keywords— Curse of dimensionality, hierarchical fuzzy systems, multi-objective genetic algorithms, variable selection.

1 Introduction

If a conventional fuzzy system is applied to large-scale problems (i.e, those with a high number of input variables), the number of rules grows exponentially with respect to the number of inputs received [1, 2]. Indeed, if we have n variables and k linguistic terms per variable, it requires up to n^k rules to build a complete Mamdani-based fuzzy system, and consequently, the accuracy-interpretability balance would be broken. This problem is know as the "curse of dimensionality."

In order to solve it, several approaches have been suggested such as variable selection [3, 4] and rule set reduction [5, 6]. Nevertheless, when the number of variables increases considerably, this kind of reduction is not enough to solve this problem. There exists a different approach which deals with this problem: Hierarchical Fuzzy Systems (HFS). An HFS is made up of a set of fuzzy subsystems or modules. These modules are linked in such a way that the output of a module is the input of other ones. Thanks to the decomposition of the fuzzy system made in an HFS, the complexity of each module is significantly reduced. There are several kinds of modules:

- SISO (Single Input Single Output): It has one input and one output.
- MISO (Multiple Inputs Single Output): It has several inputs and a single output [7]. We can find FLU (Fuzzy Logic Unit) in this kind of modules. One FLU special case is the one with two inputs and one output, which is equally found in the literature [2, 8–11].
- MIMO (Multiple Inputs Multiple Outputs): They have several inputs and outputs [12].

Apart from distinguishing several kinds of modules, it is also possible to find different types of hierarchical structures. There are different classifications [9, 13], though the most general one is the following [11]:

- Serial HFS (SHFS): The input of one module is the output of the previous ones, along with external variables [10].
- Parallel HFS: This system is organized in layers. The first one is made up of a set of modules receiving the input variables. Each variable is used as input only in a single module. The output of the modules in the first layer is the input of the modules which constitute the next layer, and so on. An aggregate operation might also exist in order to combine the outputs of one layer [2, 12].
- Hybrid HFS: This type of HFS is a mixture of the two previous ones [7, 14].

Other approaches study the best way to deal with the output variables in the modules (through fuzzy numbers [9] and matrixes modeling [11]) and the rules in different hierarchical levels [15]. Besides, some metaheuristics have been applied in this field with the purpose of obtaining an HFS capable of getting the best balance between precision and interpretability. For instance, Differential Evolution [10] has been employed to find the best membership functions, Genetic Algorithms [8], Ant Systems [14], Descendent Gradient Method [7] have been used to learn the hierarchical structure.

In this paper, a multi-objective genetic fuzzy system is suggested to obtain the best distribution of input variables and modules belonging to SHFS removing variables by crossover and mutation operators, so it can face the "curse of dimensionality." The paper is organized in the following sections: in Section 2, the suggested algorithm is described; in Section 3, the empirical study is shown; in Section 4, conclusions and future works are referred.

2 GSHFS algorithm

We propose a SHFS structure learning algorithm. These systems are a set of linked modules where the output of a module is one of the inputs of the next module. A module has its own fuzzy rule set which allows to infer the output variable from the input variables.

The algorithm uses a multi-objective genetic algorithm called Genetic SHFS (GSHFS) where there are two objectives to minimize: the root mean squared error (RMSE) and the number of rules. Our algorithm has three specific genetic operators: one crossover operator and two mutation operators.

ISBN: 978-989-95079-6-8

The following subsections detail the different components of the algorithm.

2.1 Coding scheme

In our algorithm, each individual represents a SHFS. The coding consists of a gene concatenation. Each gene has two fields: a variable index and a flag. The flag takes binary values (0 if it is an exogenous variable and 1 if it is an endogenous variable). Figures 1, 2, and 3 show some examples of the coding scheme.

A variable with 1 in the flag field means that is an endogenous variable (variable linking two modules) and a module exists. This variable is the input of the next module. All treated variables are variables of the problem. The algorithm decides if a variable of the problem will be endogenous or exogenous, but does not create new endogenous variables. In this way, the correlation between the variables of the problem is considered. The length of chromosome is fixed to the number of input variables of the problem. The highest hierarchy level happens when all modules are SISO. In this case, all variables will be endogenous except the first and the number of levels is equal to the number of input variables.

2.2 Initialization

The algorithm randomly generates an initial population. The variable index and the endogenous/exogenous flag is randomly chosen for each gene. The only restriction is that the variable can not be repeated and the flag of the first gene can not be 1 (i.e., the first variable have to be exogenous).

2.3 Crossover operator

The crossover operator is applied according to a probability between paired parents and provides exploitation. When crossover operator is applied to two parents, P_1 and P_2, there are a set of types of combinations which we must consider: 1) both parents, P_1 and P_2, have several modules; 2) parent P_1 has several modules and parent P_2 has one module; 3) parent P_1 has one module and parent P_2 has several modules (this case is similar to previous case); 4) both parents, P_1 and P_2, have one module. A parent-centric crossover has been used in some cases where the offspring mainly inherits the information of one of the parent and takes the secondary parent in order to add diversity.

The crossover is applied to individuals depending on the types of variables that two parents have in common. Each case has a priority. In this way, the crossover operator is applied from high to low priority as follows:

2.3.1 Both parents, P_1 and P_2, have several modules

- *Priority 1.- Endogenous-Endogenous case.* If P_1 parent has common endogenous variables with P_2 parent, a common endogenous variable is selected at random. This variable is the crossover point. The offspring O_1 is generated centered on P_1. Thus, the offspring O_1 inherits from P_2 the variables and modules from the beginning to the crossover point, but excluding the crossover point. The remaining is taken from P_1 and repeated variables are removed from the part inherited from P_2. The offspring O_2 is generated in the same way but centered on P_2.

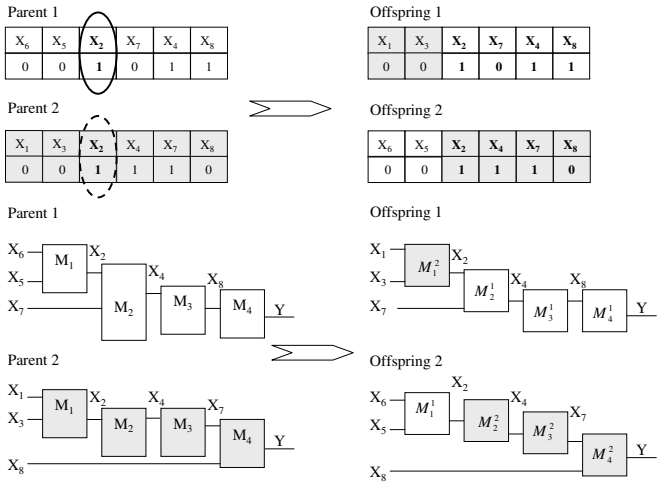


Figure 1: Example of crossover of two parents with several modules, Endogenous-Endogenous case (priority 1)

- *Priority 2.- Endogenous-Exogenous case.* If P_1 has endogenous variables which are exogenous in P_2, a common variable is randomly selected as crossover point. The offspring O_1 is generated centered on P_1. The common endogenous variable inherited from P_1 is converted into exogenous and the rest of previous modules to this variable are removed. The idea of this type of crossover is that if an endogenous variable of P_1 is equal to an exogenous variable of P_2, it indicates us that if we convert this endogenous variable into exogenous variable, RMSE will decrease, because the SHFS does not carry any error, taking the variable directly as an input.

- *Priority 3.- Exogenous-Endogenous case.* If P_1 has exogenous variables which are endogenous in P_2, one of those variables is randomly selected as crossover point. The offspring O_1 is created centered on P_1, but excluding the crossover point. The first part of O_1 is inherited from the first part of P_2 (from the first gene to the crossover point). The repeated variables are removed in O_1 from the part inherited from P_2. Figure 2 shows an example of this case. The generated SHFS offspring has four modules.

- *Priority 4.- Exogenous-Exogenous case.* If P_1 has a set of exogenous variables in common with exogenous variables in P_2, the offspring O_1 is generated as copy of P_1, but with a change: a common exogenous variable between P_1 and P_2 is randomly chosen. This variable in P_2, along with others variables, generates an endogenous variable. This endogenous variable in P_2 is selected to replace the random exogenous selected variable in O_1, which is common in P_1 and P_2. Later, the repeated variables in O_1 in part inherited from P_1 are removed. If this case comes true in P_2, O_2 offspring will be generated with the same procedure, but centered on P_2. Notice that the common exogenous variables of P_1 are looked for in P_2 excluding the exogenous variables of the module with output Y in P_2, because if this module is included, when an exogenous variable is selected there is not endogenous variable as output because it coincides with the output of

ISBN: 978-989-95079-6-8

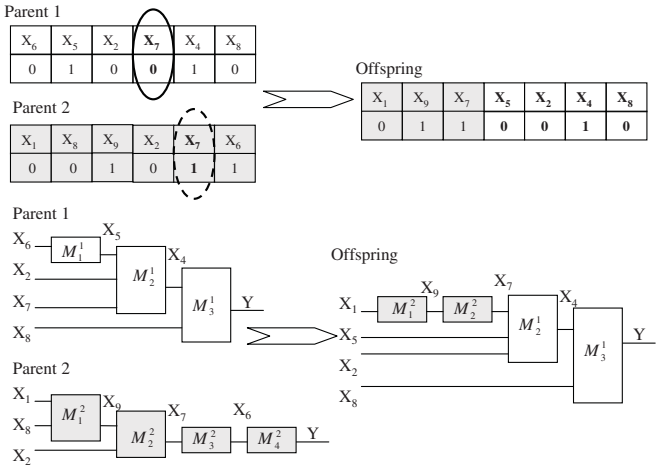


Figure 2: Example of crossover of two parents with several modules, Exogenous-Endogenous case (priority 3)

the SHFS.

- *Priority 5.- Different variables case.* If variables in P_1 and P_2 are different, an endogenous variable from P_1 and P_2 is chosen at random as crossover point. The offsprings are generated like priority 1.

2.3.2 *Parent P_1 has several modules and parent P_2 has one module*

- *Priority 1.- Endogenous-Exogenous case.* If P_1 has common endogenous variables with the exogenous variables of P_2, the offspring O_1 is generated like Endogenous-Exogenous case when P_1 and P_2 have several modules.

- *Priority 2.- Exogenous-Exogenous case.* If P_1 and P_2 have common exogenous variables, the offspring O_1 centered on P_1 is created as follows. Firstly, O_1 is generated as a copy of P_1. Then, an exogenous common variable of P_1 and P_2 is randomly chosen and moved to module with output Y. This type of crossover carries lower error and the output of SHFS is better.

- *Priority 3.- Different variables case.* In this case, O_1 is a copy of P_1 and later, an exogenous variable of P_2 is selected at random and inserted in the module of O_1 with the output Y.

2.3.3 *Parent P_1 has one module and parent P_2 has several modules*

- *Priority 1.- Exogenous-Endogenous case.* If P_1 has exogenous variables which are endogenous in P_2, the offspring O_1 is generated like Exogenous-Endogenous case when P_1 and P_2 have several modules.

- *Priority 2.- Exogenous-Exogenous case.* If P_1 has a set of exogenous variables in common with exogenous variables in P_2, the offspring O_1 is generated like Exogenous-Exogenous case when P_1 and P_2 have several modules.

- *Priority 3.- Different variables case.* If variables in P_1 and P_2 are different, an exogenous variable from P_1 and an endogenous variable from P_2 are chosen at random as crossover point. The offsprings are generated like priority 1.

2.3.4 *Both parents, P_1 and P_2, have one module*

This case, the offsprings O_1 and O_2 are generated as follows. The common exogenous variables of P_1 and P_2 are inserted in both offsprings. The rest of variables (no common variables between parents) are inserted in a set V with $|V|$ size. A number r is randomly generated: if there are not common variables between P_1 and P_2 then $r \in \{1, \ldots, n-1\}$; otherwise, $r \in \{0, \ldots, n\}$. r variables are randomly took out and inserted in O_1. Finally, the rest of variables are inserted in O_2.

2.4 *Mutation operator*

The mutation operator makes local changes in the chromosome. Two kinds of mutations have been designed:

2.4.1 *Exchange Mutation*

This operator makes an exchange of variables in a module. It chooses an exogenous variable at random in a module and exchanges between this exogenous variable and the endogenous variable of the module.

2.4.2 *Insertion Mutation*

The mutation operator distinguishes between: used and unused variables in the individual. According to it and by means of a probabilistic decision (Algorithm 1 shows a scheme), the mutation operator will choose between inserting an unused variable in a module, removing an used variable, or moving an used variable to other module.

Algorithm 1 Insertion Mutation

```
r = U[0,1];
if (r < 0.5 and there are unused variables) then
    v = Choose_at_random_a_unused_variable();
    Insert_unused_variable(v);
else
    t = U[0,1];
    if (t < 0.5 and the SHFS has only a SISO module) then
        Remove_one_used_variable();
    else
        Move_used_variable();
    end if
end if
```

- The insertion of an unused variable is as follows. Given an unused variable v, a module m is chosen randomly. Next, the operator decides if v is inserted into m with a probability of 0.5. If so, v is inserted as exogenous variable. Otherwise, if m has at least an exogenous variable as input, one of them, e, is selected at random. The mutation operator creates a new SISO module previous to m where the input is e and the output is v. In this case, v will be endogenous variable and input of module m. If m has not exogenous variables, the m's output is converted into a new exogenous variable of m and v will be the new output of m.

- To remove an used variable, first a module m of the SHFS is randomly chosen. If m has exogenous variables then one of them is removed at random. Otherwise, the module m is removed, thus linking the output of the module before m with the input of the module after m.

ISBN: 978-989-95079-6-8

- To move an used variable from a module to other, two modules of the chromosome are chosen at random: a source module m_1 where a variable is removed and a destination module m_2 where the variable is inserted. An exogenous variable of the m_1 is randomly removed and inserted as exogenous variable of the m_2. If there are not exogenous variables into the m_1 and consequently m_1 has one input and one output (a SISO module), its endogenous variable is chosen. In this situation, m_1 disappears and the chosen variable is inserted as an exogenous variable into m_2. If $m_1 = m_2$, an exogenous variable v is randomly selected, then a new module is created after m_1/m_2. The input of this new module (and therefore the output of the module m_1/m_2) is v while the output of the new module will be old output of m_1/m_2.

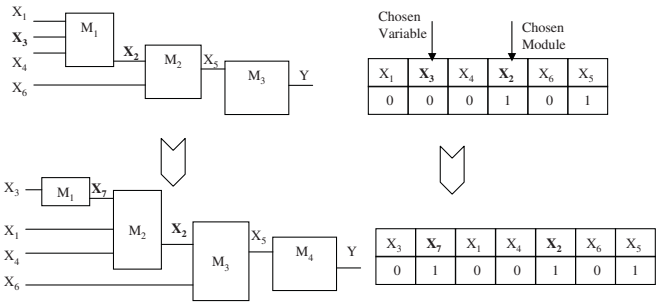


Figure 3: Example of insertion mutation of a unused variable (X_7) as endogenous variable, thus creating a new module

2.5 Fuzzy rule set learning

The learning fuzzy rule set of SHFS is as follows. Each module learns its fuzzy rule set with Mamdani rules depending on the inputs variables by WM [16]. The endogenous variables in SHFS are inferred by its respective module according to the input variables of the module. The obtained system is evaluated by training data examples, choosing the correspondent values of exogenous variables in that module. A module, according to learned fuzzy rule set, generates a value of output (an endogenous variable). The value is the next input of the module along with others exogenous variables. The membership functions are triangular shape distributed strong fuzzy partition.

2.6 Multi-objective approach

A generational approach with the multi-objective elitist replacement strategy of NSGA-II [17] is used. Crowding distance [17] in the objective function space is considered. Binary tournament selection based on the non-domination rank (or the crowding distance when both solutions belong to the same front) is applied. The crowding distance is normalized for each objective according to the extreme values of the solutions contained in the analyzed front.

2.7 Objective functions

We consider two objective functions to be minimize and so get a better precision and interpretability of the system.

2.7.1 Accuracy

It is computed as the RMSE:

$$F_1(S) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S(x^i) - y^i)^2} \quad (1)$$

with S being the fuzzy system to be evaluated, N the data set size and (x^i, y^i) the ith input-output pair of the data set. The objective is to minimize this function to get good precision.

2.7.2 Interpretability

It is computed as follows:

$$F_2(S) = r(S) \quad (2)$$

with $r()$ being the number of rules of the fuzzy system.

Although the number of modules or the number of variables by module are notable criteria to evaluate the learned quality of the SHFS, the number of rules is more intuitive to validate the interpretability because it combines both criteria in one objective. If the number of rules is considered, the number of variables and modules is controlled automatically: as the number of rules decreases, the number of modules and variables increase.

3 Experimental Results

This section presents the experimental results. The objective of experimentation is to prove the reduction of number of rules and so, to assess the interpretability that can be compared to a referential learning method (WM in our case).

3.1 Problems

We have considered five regression problems with a moderate and high number of input real-valued variables: *Dee*[1], *Concrete*[2], *Elevators*[3], *Computer Activity (Comp-activ)*[3], and *Ailerons*[3].

Table 1 collects the main features of each problem, where *#InputVar* stands for number of input variables, *#Exam* for total number of examples, and *#LingTerms* for the number of triangular-shaped uniformly distributed linguistic terms considered for each variable in this experimental analysis.

Table 1: Data sets considered in the experimental analysis

Data set	*#InputVar*	*#Exam*	*#LingTerms*
Dee	6	365	5
Concrete	8	1030	5
Elevators	18	16559	3
Comp-activ	21	8192	3
Ailerons	40	13750	3

The experiments shown in this paper have been performed by a five fold cross validation with a total of 30 runs per problem (six runs for each data set partition). Thus, the data set is divided into five subsets of (approximately) equal size. The algorithm is then applied five times to each problem, each time

[1] KEEL: Knowledge extraction based on evolutionary learning. http://www.keel.es

[2] UCI Machine Learning Repository. Collection of regression datasets. http://archive.ics.uci.edu/ml/datasets.html

[3] L. Torgo, Collection of regression datasets. http://www.liacc.up.pt/ ltorgo/Regression/DataSets.html

ISBN: 978-989-95079-6-8

leaving out one of the subsets from training, but using only the omitted subset to compute the test error.

Our algorithm has been run with the following parameter values: 1,000 iterations, 60 as population size, 0.7 as crossover probability, and 0.2 as mutation probability per chromosome. We have not performed any previous analysis to fix these values, so better results may probably be obtained by tuning them though we have informally noticed our algorithm is not specially sensitive to any parameter.

3.2 Obtained results

For each problem we present the five validation test results (Table 2), where $RMSE_{tra}$ and $RMSE_{tst}$ are the approximation error (eq. 1) values over the training and test data set respectively, *#M*, *#R*, *#V* stand for the number of modules, the number of fuzzy rules, and the number of variables respectively.

Since our algorithm performs multi-objective optimization, several solutions are returned in each run. Therefore, we show five representative solutions from the final Pareto-optimal set: the first row of each problem is the best solution (highest precision), the second row is the first quartile, the third row is the median (it is not the average), the fourth row is the third quartile, and the fifth row is the worst solution (lowest precision).

Table 2: Obtained Results

	Method	$RMSE_{tra}$	$RMSE_{tst}$	*#M*	*#R*	*#V*
DEE	WM	0.375594	0.477303	1.0	178.4	6.0
	GSHFS	0.372482	0.474365	1.2	157.3	5.6
		0.396565	0.460321	1.8	95.6	5.2
		0.430456	0.484292	1.8	66.1	4.6
		0.495178	0.507092	2.1	28.6	3.7
		0.681968	0.695144	1.2	5.5	1.2
CONCRETE	WM	8.548268	9.822000	1.0	310.4	8.0
	GSHFS	8.479054	9.356946	1.4	257.1	7.2
		9.436752	9.812343	2.2	114.7	6.0
		10.700607	10.839647	2.3	53.3	5.2
		13.111270	13.450416	2.4	25.4	4.2
		16.669822	16.477185	1.3	5.6	1.3
ELEVATORS	WM	0.567771	0.542072	1.0	511.0	18.0
	GSHFS	0.487507	0.488641	2.3	39.7	7.7
		0.498203	0.498650	2.2	27.5	6.5
		0.517594	0.517673	2.4	18.9	5.7
		0.563252	0.563026	2.0	11.7	4.3
		1.421620	1.422627	1.6	3.8	1.8
COMP-ACTIV	WM	9.050474	9.084798	1.0	425.6	21.0
	GSHFS	5.398862	5.400107	2.5	37.5	8.1
		5.880516	5.874316	2.3	25.1	6.6
		6.708214	6.719212	2.2	16.8	5.5
		8.888890	8.935257	2.1	9.7	3.9
		31.900506	31.756379	1.7	4.1	2.0
AILERONS	WM	0.253971	0.255557	1.0	1080.6	40.0
	GSHFS	0.238855	0.239763	2.8	57.1	10.0
		0.254118	0.254958	2.7	37.7	8.2
		0.276632	0.277216	2.6	24.0	6.8
		0.318274	0.318159	2.3	12.9	5.0
		0.724993	0.731567	2.2	5.9	2.9

3.3 Analysis

The main objective is to obtain a fuzzy system with good tradeoffs between precision and interpretability. When the number of rules of a fuzzy system decreases, the system gets better interpretability, but the precision is lower and vice versa. This event can be observed clearly in the results obtained as problems with either low or high number of variables. Notice that the fuzzy rule set is learned individually for each module.

First set of Table 2 shows five solutions of the Pareto obtained by our algorithm in *Dee* problem. We can observe that the precision is high when the number of rules is high and, consequently, the fuzzy system has a lower interpretability. If we compare our algorithm with WM method, the best solution (highest precision) is better in precision, having a number of rules and a number of variables lower. In first quartile, the precision has a little increase according to solution obtained by WM in the $RMSE_{tra}$, but the $RMSE_{tst}$ is lower and the interpretability is decreased in more than a half.

However, if we observe the problems with a considerable number of variables (more than eight), our algorithm has better performance. Lets look at *Elevators* problem. It has 18 variables. The best solution (highest precision) obtained by our algorithm is better than WM solution. The $RMSE_{tra}$ and $RMSE_{tst}$ have been decreased. The interpretability is better because the SHFS has a lower number of variables and a higher number of modules. The number of rules has been decreased by 92%. The third quartile of the $RMSE_{tra}$ is better in precision than WM. The third quartile of the $RMSE_{tst}$ is a little worse but the number of rules obtained by our algorithm has decreased by 95%, 98%, and 96%, respectively.

In *Ailerons* problem (40 variables), we can see that our solution with highest precision is better than the solution obtained by WM. We want to emphasize that the number of rules is decreased by 95% and the number of variables by 75%. The median of the $RMSE_{tra}$ is a solution with a precision a little higher than WM solution, but the median of the $RMSE_{tst}$ is lower and the interpretability is better than the solution of WM, due to decrease of the number of variables and its distribution in modules.

To sum up, we have seen that in problems with a low moderate number of variables, the precision is worse when the number of rules decrease. It does not happen in problems with a higher number of variables because with a lower number of variables, the correlation between them is low, because they are important. Nevertheless, in problems with a higher number of variables exists more dependence between them: a hierarchical structure and a lower number of variables help to decrease the complexity of the learned fuzzy model.

Figure 4 shows the average Pareto front obtained in Ailerons problem by our algorithm. In it we can see that if there are many rules (lower interpretability), then the precision is higher. The SHFS gets a better interpretability with a lower precision.

Finally, we show an example of GSHFS hierarchical structure obtained by our algorithm in Figure 5. It is an example of non-dominated solution obtained by GSHFS in a data set partition of *Ailerons*. The obtained values are: $RMSE_{tra}$ = 0.214068, $RMSE_{tst}$ = 0.219682, #M = 3, #R = 63, #V = 9. The results obtained by WM in the same data set partition are: $RMSE_{tra}$ = 0.254144, $RMSE_{tst}$ = 0.261472, #M = 1, #R = 1072, #V = 40. We can observe that $RMSE_{tra}$ and $RMSE_{tst}$ obtained by our algorithm is lower than the values obtained by WM, decreasing the number of rules by 94%.

ISBN: 978-989-95079-6-8

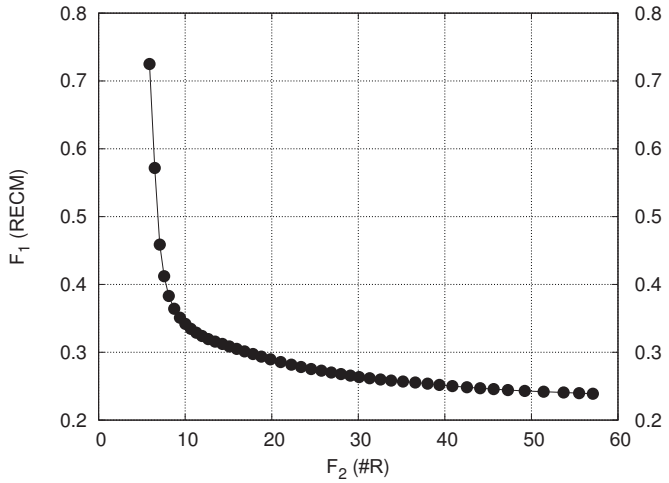


Figure 4: Average Pareto front obtained in Ailerons problem

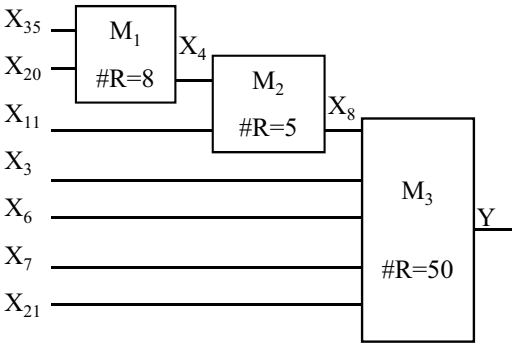


Figure 5: Example of a solution obtained by GSHFS in the *Ailerons* problem with #M = 3, #R = 63, #V = 9

4 Conclusion and Further Work

We have proposed a multi-objective algorithm applied to learning SHFS for palliate exponential increase of the number of rules when number of variables increases. The set of variables is divided into modules by the algorithm. We have proved that this division by SHFS can obtain good results in problems with a higher number of variables.

As further work, we suggest to add mechanisms for a better precision, for example, a global learning of fuzzy rule set, a detailed study of the interpretability of each rule of the fuzzy system, and to extend this algorithm for parallel and hybrid structure learning.

Acknowledgment

This work was supported in part by the Spanish Ministry of Science and Innovation (grant no. TIN2008-06681-C06-01), and by Andalusian Government (grant no. P07-TIC-3185).

References

[1] G.V.S. Raju, J. Zhou, and R.A. Kisner. Hierarchical fuzzy control. *International Journal Control*, 54(5):1201–1216, 1991.

[2] M.G. Joo and J.S. Lee. Hierarchical fuzzy control scheme using structured takagi-sugeno type fuzzy inference. In *Proceedings of IEEE International Fuzzy Systems Conference*, pages 78–83, Seoul, Korea, 1999.

[3] T.P. Hong and J.B. Chen. Finding relevant attributes and membership functions. *Fuzzy Sets and Systems*, 103(3):389–404, 1999.

[4] A. González and R. Pérez. Selection of relevant features in a fuzzy genetic learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 31(3):417–425, 2001.

[5] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Selecting fuzzy if-then rules for classification problems using genetic algorithms. *IEEE Transactions in Fuzzy Systems*, 3(3):260–270, 1995.

[6] T. Taniguchi, K. Tanaka, H. Ohtake, and H.O. Wang. Model construction, rule reduction, and robust compensation for generalized form of takagi-sugeno fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 9(4):525–538, 2001.

[7] D. Wang, X.J. Zeng, and J.A. Keane. Learning for hierarchical fuzzy systems based on gradient-descent method. In *Proceedings of IEEE International Conference on Fuzzy Systems*, pages 92–99, 2006.

[8] K. Shimojima, T. Fukuda, and Y. Hasegawa. Self-tuning fuzzy modeling with adaptive membership function, rules, and hierarchical structure based on genetic algorithm. *Fuzzy Sets and Systems*, 71:295–309, 1995.

[9] A.E. Gaweda and R. Scherer. Fuzzy number-based hierarchical fuzzy system. *Lecture Notes in Computer Science*, 3070:302–307, 2004.

[10] F. Cheong. A hierarchical fuzzy system with high input dimensions for forecasting foreign exchange rates. *IEEE Congress on Evolutionary Computation, CEC*, pages 1642–1647, 2007.

[11] S. Aja-Fernández and C. Alberola-López. Matriz modeling of hierarchical fuzzy systems. *IEEE Transactions in Fuzzy Systems*, 16(3):585–599, 2008.

[12] P. Salgado. Rule generation for hierarchical collaborative fuzzy system. *Applied Mathematical Modelling, Science Direct*, 32:1159–1178, 2008.

[13] V. Torra. A review of the construction of hierarchical fuzzy systems. *International Journal of Intelligent Systems*, 17:531–543, 2002.

[14] Y. Chen, J. Dong, and B. Yang. Automatic design of hierarchical ts-fs model using ant programming and pso algorithm. In C. Bussler and D. Fensel, editors, *Proceedings 12th International Conference on Artificial Intelligence, Methodology, Systems and Applications, Lecture Notes on Artificial Inteligence, LNAI 3192*, pages 285–294, 2004.

[15] T.M. Jelleli and A.M. Alimi. Improved hierarchical fuzzy control scheme. *Adaptive and Natural Computing*, 1:128–131, 2005.

[16] L.X. Wang and J. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427, 1992.

[17] K. Deb, A. Pratap, S. Agarwal, and T. Meyarevian. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

ISBN: 978-989-95079-6-8

An Introduction to Parameterized IFAM Models with Applications in Prediction

Peter Sussner[1] Rodolfo Miyasaki[2] Marcos Eduardo Valle[3]

1.Department of Applied Mathematics, University of Campinas
Campinas, State of São Paulo, Brazil
2.Credit Planning and Management Department, Banco Itaú
São Paulo, Brazil
3.Department of Mathematics, University of Londrina
Londrina, Brazil
Email: sussner@ime.unicamp.br, rodolfo.miyasaki@itau.com.br, valle@uel.br

Abstract— *Fuzzy associative memories (FAMs) and, in particular, the class of implicative fuzzy associative memories (IFAMs) can be used to implement fuzzy rule-based systems. In this way, a variety of applications can be dealt with. Since there are infinitely many IFAM models, we are confronted with the problem of selecting the best IFAM model for a given application. In this paper, we restrict ourselves to a subclass of the entire class of IFAMs, namely the subclass of IFAMs that are associated with the Yager family of parameterized t-norms. For simplicity, we speak of the class of Yager IFAMs. In this setting, we formulate the problem of choosing the best Yager IFAM for a given application as an optimization problem. Considering two problems in time series prediction from the literature, we solve this optimization problem and compare the performance of the resulting Yager IFAM with the performances of other fuzzy, neural, neuro-fuzzy, and statistical techniques.*

Keywords— Fuzzy associative memory, implicative fuzzy associative memory, Yager family of parameterized t-norms, time-series prediction, hydroelectric plant, monthly streamflow prediction.

1 Introduction

Implicative fuzzy associative memories (IFAMs) belong to the class of *fuzzy morphological associative memories* (FMAMs) [1, 2]. The theoretical background for FMAMs can be found in *fuzzy mathematical morphology* (FMM) [3, 4]. More precisely, FMAMs can be viewed as fuzzy neural networks whose neurons perform elementary operations of fuzzy mathematical morphology [2].

Recently, Sussner and Valle [5] showed that many well-known fuzzy associative memory (FAM) models such as the FAMs of Kosko, Junbo et al., Liu, and Bělohlavék [6, 7, 8, 9] represent particular instances of FMAMs. In this paper, we focus on the FMAM subclass of IFAMs. Upon presentation of an input pattern $\mathbf{x}$, an IFAM model performs max-T products, where T is a t-norm, at every node. A certain continuous t-norm determines a particular IFAM model. For example, the Lukasiewicz IFAM is associated with the Lukasiewicz t-norm. Learning in IFAM models occurs by forming the R-implication of the underlying t-norm and by computing a matrix product that corresponds to the Bandler-Kohout superproduct [10, 11]. In the context of fuzzy associative memories, we speak of *implicative fuzzy learning*. As we have pointed out in previous articles, implicative fuzzy learning is based on the greatest solution of a fuzzy relational inequality [1, 12, 13]. In the general setting of FMAMs, implicative fuzzy learning generalizes to *learning by adjunction*, a recording strategy that was derived from a concept in mathematical morphology known as the duality relationship of adjunction [2, 14].

We have successfully applied IFAMs to some problems in time series forecasting, in particular a problem of forecasting the demand of man power in steel manufacturing industry in the state of West Bengal, India, and a problem of stream flow prediction in a hydroelectric plant in southeastern Brazil [5, 15]. In these prediction problems, the Lukasiewicz IFAM exhibited the best performance compared to a variety of other FAM models and to a number of statistical, neural, fuzzy, and neuro-fuzzy approaches from the literature. The goal of this paper is to optimize our results by considering not only a finite number of IFAMs but an infinite number of IFAM models.

To this end, we introduce the class of Yager IFAMs, a class of parameterized IFAM models depending on a single parameter. We simply define a Yager IFAMs as an IFAM that is associated with a Yager t-norm [16]. Recall that a Yager t-norm is of the form $T^d(x, y) = 1 - \{1 \wedge [(1-x)^d + (1-y)^d]^{1/d}\}$, where the symbol $\wedge$ denotes the minimum operation and where $d > 0$.

Thus, we can formulate an optimization problem depending on the parameter d that consists in minimizing the distance between the target pattern and the pattern produced by the IFAM corresponding to the parameter d. As a result of this optimization process, we obtain the Yager IFAM (or the set of Yager IFAMs) that provides the best fit with respect to the training data.

The paper is organized as follows. After presenting a brief introduction to IFAMs, we introduce the Yager class of parameterized IFAMs in Section 3. Subsequently, we discuss the aforementioned prediction problems and determine the Yager IFAMs that minimize the mean squared distance between the outputs produced by the memory cues $\mathbf{x}^\xi$ and the desired outputs. We finish the paper with some concluding remarks and suggestions for further research.

2 A Brief Introduction to Implicative Fuzzy Associative Memories

Implicative Fuzzy Associative Memories (IFAMs) can be defined in terms of certain matrix products, namely the max-T product and the min-I product, where T is a t-norm and I is a fuzzy implication [2]. We will make use of the following

ISBN: 978-989-95079-6-8

t-norms:

$$T_M(x,y) = x \wedge y\,, \qquad \text{(Minimum)} \quad (1)$$

$$T_P(x,y) = x \cdot y\,, \qquad \text{(Product)} \quad (2)$$

$$T_L(x,y) = 0 \vee (x+y-1)\,, \qquad \text{(Lukasiewicz)} \quad (3)$$

$$T_W(x,y) = \begin{cases} x \wedge y, & x \vee y = 1\,, \\ 0, & x \vee y < 1\,. \end{cases} \qquad \text{(Nilpotent Min.)} \quad (4)$$

Reverse fuzzy implications can be easily derived from fuzzy implications. Recall the following popular implications:

$$I_M(x,y) = \begin{cases} 1, & x \le y \\ y, & x > y \end{cases}\,, \qquad \text{(Gödel)} \quad (5)$$

$$I_P(x,y) = \begin{cases} 1, & x \le y \\ y/x, & x > y \end{cases}\,, \qquad \text{(Product)} \quad (6)$$

$$I_L(x,y) = 1 \wedge (y - x + 1)\,. \qquad \text{(Lukasiewicz)} \quad (7)$$

Let $A \in [0,1]^{m\times p}$ and $B \in [0,1]^{p\times n}$. The max-T product of A and B is given by the matrix $C = A \circ B$ and the min-I product of A and B is given by the matrix $D = A \circledast B$, where C and D are defined as follows (note that the min-I product of A and B can also be viewed as a Bandler-Kohout superproduct [10, 11]).

$$c_{ij} = \bigvee_{k=1}^{p} C(a_{ik}, b_{kj}) \;\text{ and }\; d_{ij} = \bigwedge_{k=1}^{p} I(b_{kj}, a_{ik})\,. \quad (8)$$

A FAM is a fuzzy neural network that is designed to store a fundamental memory set, i.e., a set of associations $\{(\mathbf{x}^\xi, \mathbf{y}^\xi) : \xi = 1,\ldots,p\}$, where $\mathbf{x}^\xi \in [0,1]^n$ and $\mathbf{y}^\xi \in [0,1]^m$. This task can be achieved by means of a synaptic weight matrix $W \in [0,1]^{m\times n}$. Let $X = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p] \in [0,1]^{n\times p}$ denote the matrix whose columns are the input patterns and let $Y = [\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^p] \in [0,1]^{m\times p}$ denote the matrix whose columns are the output patterns. We suggested the following rule for synthesizing the weight matrix W of an IFAM [1, 12, 13]:

$$W = Y \circledast X^T\,. \quad (9)$$

Note that (9) depends on the choice of a fuzzy implication. In an IFAM model, we require that the underlying fuzzy implication I is the R-implication of a continuous t-norm T. In other words, the corresponding fuzzy implication I must be adjoint to a continuous t-norm T [2, 4]. For example, the pairs (I_M, T_M), (I_P, T_P), and (I_L, T_L) represent pairs of adjoint operators.

There are infinitely many IFAM models. In the special cases where the fuzzy implications occurring in (9) are I_M, I_P, and I_L, we speak of the Gödel IFAM, the Goguen IFAM, and the Lukasiewicz IFAM, respectively.

Once the recording phase has been completed, the IFAM weight matrix can be applied to an arbitrary input pattern $\mathbf{x} \in [0,1]^n$. If θ denotes a threshold or bias vector that is given by the entry-wise minimum over all $\mathbf{y}^\xi$, where $\xi = 1,\ldots,p$, then we obtain the following output pattern $\mathbf{y} \in [0,1]^m$:

$$\mathbf{y} = (W \circ \mathbf{x}) \vee \boldsymbol{\theta}\,, \qquad \text{where} \qquad \boldsymbol{\theta} = \bigwedge_{\xi=1}^{p} \mathbf{y}^\xi\,. \quad (10)$$

Here, the t-norm T occurring in the max-T product is such that its R-implication I is associated with the product "$\circledast$" of (9). In other words, the t-norm T of (10) and the fuzzy implication I of (9) are adjoint [2, 4].

3 The Yager Class of Parameterized IFAMs

Note that a particular IFAM model is determined by a continuous t-norm. An entire subclass of IFAM models is given by the class of Yager t-norms or intersections which are obviously continuous. Recall that the class of Yager t-norms represents a family of parameterized t-norms given by the following equation [16]:

$$T^d(x,y) = 1 - \left\{1 \wedge [(1-x)^d + (1-y)^d]^{1/d}\right\},\; d > 0\,. \quad (11)$$

For $d = 1$, the Yager t-norm coincides with the Lukasiewicz t-norm T_L. For $t \to 0$ the Yager t-norm approaches the nilpotent minimum T_W, which is the pointwise smallest t-norm. For $t \to \infty$ the Yager t-norm approaches the minimum t-norm T_M, the pointwise largest t-norm [17, 18].

For $d > 0$, the symbol I^d denotes the R-implication of T^d. The operator I^d can be computed as follows:

$$I^d(x,y) = 1 - \left\{0 \vee [(1-y)^d - (1-x)^d]\right\}^{1/d},\; d > 0\,. \quad (12)$$

The resulting IFAMs are called Yager IFAMs. These parameterized IFAM models will be considered in the applications of the next section. In fact, we will tackle the problem of choosing the best Yager IFAM for a given application in terms of an optimization problem.

To this end, let us consider the fundamental memory set $\{(\mathbf{x}^\xi, \mathbf{y}^\xi) : \xi = 1,\ldots,p\}$ which can be viewed as the set of training patterns. Let W^d denote the synaptic weight matrix of the Yager IFAM with parameter $d > 0$. A possible approach towards finding the best Yager IFAM consists in minimizing the distances between $(W^d \circ^d \mathbf{x}^\xi) \vee \theta$ and $\mathbf{y}^\xi$ for $\xi = 1,\ldots,p$, where the symbol "$\circ^d$" denotes the max-T^d product. This optimization problem can be formulated as follows:

$$\begin{cases} \text{minimize } \sum_{\xi=1}^{p} \left\|\mathbf{y}^\xi - [(W^d \circ^d \mathbf{x}^\xi) \vee \theta]\right\|_2 \\ \text{subject to } \quad d > 0 \end{cases}\,. \quad (13)$$

If $\mathbf{y}^{\xi,d} = (W \circ^d \mathbf{x}^\xi) \vee \theta$ and Y^d is the matrix $Y^d = [\mathbf{y}^{1,d}, \ldots \mathbf{y}^{p,d}]$ then the optimization problem of (13) can be expressed in terms of the Frobenius norm as follows:

$$\begin{cases} \text{minimize } \left\|Y - Y^d\right\|_F \\ \text{subject to } \quad d > 0 \end{cases}\,. \quad (14)$$

In the applications as fuzzy rule-based systems that are described below, an input fuzzy set $\mathbf{x} \in [0,1]^n$ is derived from a real-valued input vector $\mathbf{v}$ and the respective output fuzzy set $\mathbf{y} = (W^d \circ \mathbf{x}) \vee \theta$ is defuzzified yielding a real-valued output $s = defuzz(\mathbf{y})$. The training patterns are given in the form $(\mathbf{v}^1, s^1), \ldots, (\mathbf{v}^p, s^p)$. Let $s^{\xi,d}$ denote the defuzzification of $\mathbf{y}^{\xi,d}$, i.e. $s^{\xi,d} = defuzz(\mathbf{y}^{\xi,d})$ for all $\xi = 1,\ldots,p$. If we have $\mathbf{s} = (s^1,\ldots,s^p)^T$ and $\mathbf{s}^d = (s^{1,d},\ldots,s^{p,d})^T$ then we choose to minimize the following expression instead of (14):

ISBN: 978-989-95079-6-8

$$\begin{cases} \text{minimize } \left\|\mathbf{s}-\mathbf{s}^d\right\|_2 \\ \text{subject to } \quad d>0 \end{cases} . \tag{15}$$

In other words, we minimize the Euclidean distance between the desired results $\mathbf{s}$ and the results produced by the Yager IFAMs for $d > 0$. To this end, we applied the routine FMINBND of MATLAB's Optimization Toolbox in order to determine a local minimum of the real-valued objective function $f(d) = \left\|\mathbf{s}-\mathbf{s}^d\right\|_2$ in an interval. Note that plotting the objective function $f : \mathbb{R} \to \mathbb{R}$ allows us to select an interval $[x_1, x_2]$ such that an application of FMINBND yields a candidate for a global minimum. The resulting parameter d yields the Yager IFAM that exhibits the best performance on the training data. The next sections describes applications of this strategy to two forecasting problems that can be found in the literature.

4 Applications of Yager IFAMs in Prediction

Fuzzy associative memories such as IFAMs can be used to implement mappings of fuzzy rules. In this case, a set of rules in the form of human-like IF-THEN conditional statements are stored. In this subsection, we consider two problems in time-series prediction. We use (15) to select the Yager IFAM that produces the least MSE on the training data.

4.1 Prediction of Manpower Requirement in Steel Manufacturing

Let us consider the problem of predicting the manpower requirement in steel manufacturing industry in the state of West Bengal, India [19]. Initially, we have five linguistic values A_i, $i = 1, \ldots, 5$ and a set of fuzzy conditional statements such as "If the manpower of year n is A_i, then that of year $n + 1$ is A_j". The linguistic values A_i correspond to fuzzy sets. Table 1 shows the set of input-output pairs that we stored in a number of different FAM models including the Gödel, the Goguen, the Lukasiewicz, and the Yager IFAMs with $d > 0$.

If W is the synaptic weight matrix of an IFAM model and $\boldsymbol{\theta}$ is the threshold obtained after the learning process, then the predicted manpower of year $n + 1$ is given by the following equation:

$$A_{n+1} = (W \circ A_n) \vee \boldsymbol{\theta} , \tag{16}$$

where A_n is the manpower of year n and $\circ$ is the max-t composition. Note that here we have $\boldsymbol{\theta} = (0, 0, 0, 0, 0)^T$ and therefore

$$A_{n+1} = W \circ A_n . \tag{17}$$

The fuzzy set A_n is computed by fuzzifying the numerical input value v corresponding to the manpower requirement of year n according to the method described by Choudhury et. al. [19]. We calculated the predicted value s of the required manpower for year $n + 1$ by applying (17) to A_n and by defuzzifying the result A_{n+1} using the "mean of maxima" (MOM) scheme.

Choudhury et. al. used this problem to compare the average or mean percentage error (MPE) produced by Kosko's FAM [6, 20] and by the statistical methods ARIMA1 and ARIMA2 [21, 22]. In a recent paper [1], we included some IFAM models, in particular the Lukasiewicz IFAM, as well as the Lukasiewicz Generalized FAM (GFAM) of Chung and Lee [23] and the max-min FAM with threshold of Liu [8] in this comparison. The Lukasiewcz IFAM, which is closely related to the original morphological associative memory (MAM) [1, 24, 25], exhibited the best performance among all the models mentioned above. Recall that the Lukasiewicz IFAM coincides with the Yager IFAM for $d = 1$.

Table 1: Set of input and output pairs used in the forecasting application

ξ	$\mathbf{x}^\xi$	$\mathbf{y}^\xi$
1	$[1.0, 0.5, 0, 0, 0]^T$	$[0.5, 1.0, 0.5, 0, 0]^T$
2	$[0.5, 1.0, 0.5, 0, 0]^T$	$[0.5, 1.0, 0.5, 0, 0]^T$
3	$[0.5, 1.0, 0.5, 0, 0]^T$	$[0, 0.5, 1.0, 0.5, 0]^T$
4	$[0, 0.5, 1.0, 0.5, 0]^T$	$[0.5, 1.0, 0.5, 0, 0]^T$
5	$[0, 0.5, 1.0, 0.5, 0]^T$	$[0, 0.5, 1.0, 0.5, 0]^T$
6	$[0, 0.5, 1.0, 0.5, 0]^T$	$[0, 0, 0.5, 1.0, 0.5]^T$
7	$[0, 0, 0.5, 1.0, 0.5]^T$	$[0, 0, 0.5, 1.0, 0.5]^T$
8	$[0, 0, 0.5, 1.0, 0.5]^T$	$[0, 0, 0, 0.5, 1.0]^T$
9	$[0, 0, 0, 0.5, 1.0]^T$	$[0, 0, 0, 0.5, 1.0]^T$

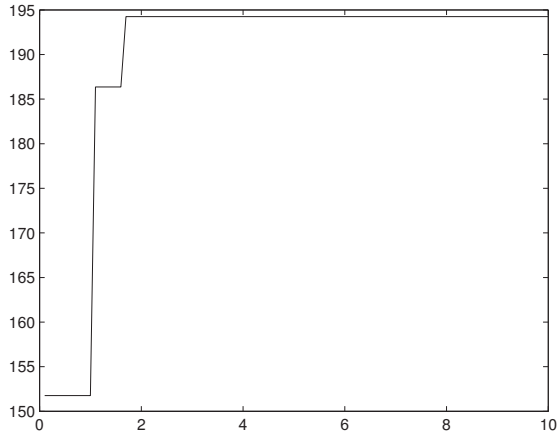


Figure 1: The distance $\left\|\mathbf{s}-\mathbf{s}^d\right\|_2$ between the actual demand of manpower between 1984 and 1995 and the demand of manpower predicted by the Yager IFAM as a function of d.

In this paper, we determine the best Yager IFAM for this prediction problem by solving the optimization problem of (15). Since the Lukasiewicz IFAM is taken into account in the optimization process, the IFAMs representing solutions of (15) are guaranteed to perform at least as well as the Lukasiewicz IFAM with respect to the MSE. In fact, the minimum mean squared error (MSE) is adopted for any value of d in the interval $(0, 1]$ as shown in Figure 1.

Table 2 presents a comparison of the FAM models and the statistical autoregressive integrated moving average (ARIMA) methods in terms of the MSE, the MPE, and the mean absolute error (MAE). Figure 2 provides for a graphical interpretation of the predictions.

The results of this experiment indicate the utility of IFAMs for prediction problems and the validity of our approach for determining the best Yager IFAM. However - as the reader may have noticed - this prediction problem does not include

ISBN: 978-989-95079-6-8

Table 2: MSE, MAE and MPE produced by the prediction models

Method	MSE ($\times 10^5$)	MAE (m^3/s)	MPE (%)
Yager IFAM ($d \in (0,1]$):	1.92	32.75	2.29
Lukasiewicz IFAM:	1.92	32.75	2.29
Kosko's FAM:	2.57	38.75	2.67
Lukasiewicz GFAM:	2.57	38.75	2.67
Gödel IFAM:	2.89	38.58	2.73
Max-min FAM/threshold:	2.89	38.58	2.73
Goguen IFAM:	3.14	42.75	2.99
ARIMA2	9.36	83.55	5.48
ARIMA1	23.26	145.25	9.79

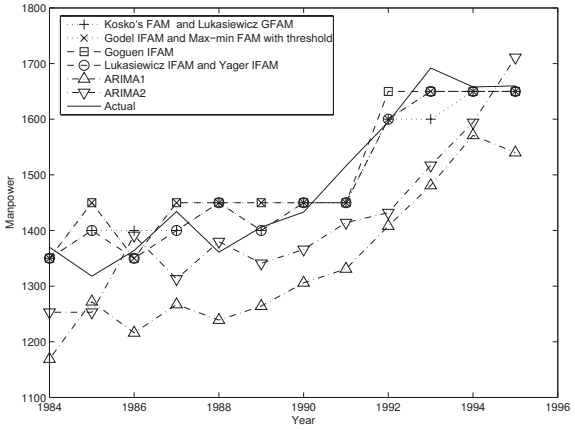


Figure 2: Comparison in forecasting manpower. The continuous line represents the actual manpower. The dashed line marked by '$\circ$' corresponds to the Lukasiewicz IFAM model, i.e., the Yager IFAM for $d = 1$ and the dashed line marked by '$\square$' corresponds to the Goguen IFAM. The dotted line marked by '$\times$' refers to Kosko's FAM model as well as the Lukasiewicz Generalized FAM, and the dotted line marked by '$+$' refers to the max-min FAM with threshold and the Gódel IFAM. The lines marked by '$\triangle$' and '$\triangledown$' represent ARIMA1 and ARIMA2.

any test data that would allow us to assess the performance of the models for years beyond the year of 1995. More realistic problems such as the one described below consist of training data and test data. In this case, the optimization should be performed on the training data and the test data should serve to validate this approach (possibly, separate validation data can be extracted from the training data and the validation data can be used to choose the best model).

4.2 Prediction of Average Monthly Streamflow of a Hydroelectric Plant

Furnas is a large hydroelectric plant that is located in southeastern Brazil. Magalhaẽs et al. as well as Sussner and Valle have previously discussed the problem of forecasting the average monthly streamflow [26, 15]. The seasonality of the monthly streamflow suggests the use of 12 different models, one for each month of the year. The training data correspond to the years $1931 - 1990$ and the test data correspond to the years $1991 - 1998$.

Let s^ξ, where $\xi = 1, \ldots, p$, be samples of a seasonal streamflow time series. The goal is to estimate the value of s^γ from a subsequence of $(s^1, s^2, \ldots, s^{\gamma-1})$. Here, we employ subsequences that correspond to a vector of the form

$$\mathbf{v}^\gamma = (s_{\gamma-h}, \ldots, s_{\gamma-1})^T, \tag{18}$$

where $h \in \{1, 2, \ldots, \gamma - 1\}$. In this experiment, we employed a fixed number of three antecedents in our IFAM models. For example, only the values of January, February, and March were taken into account to predict the streamflow of April.

The uncertainty that is inherent in hydrological data suggests the use of fuzzy sets to model the streamflow samples. For $\xi < \gamma$, a fuzzification of $\mathbf{v}^\xi$ and s^ξ using Gaussian membership functions yields fuzzy sets $\mathbf{x}^\xi : \mathcal{U} \to [0, 1]$ and $\mathbf{y}^\xi : \mathcal{V} \to [0, 1]$, respectively, where $\mathcal{U}$ and $\mathcal{V}$ represent finite universes of discourse. A subset S of the resulting input-output pairs $(\mathbf{x}^\xi, \mathbf{y}^\xi)$, where $\xi = 1, \ldots, p$, is implicitly stored in an IFAM model (we only construct the parts of the weight matrix that are actually used in the recall phase). We employed the *subtractive clustering method* to determine the set S [27]. The centers and the widths of the Gaussian-type membership functions of the input patterns $\mathbf{x}^\xi$ and the output patterns $\mathbf{y}^\xi$ in S were estimated by means of the MATLAB function `subclust`. Here, we used a constant width of $r = 0.5$.

Upon presentation of the input pattern $\mathbf{x}^\gamma$, the IFAM with parameter d yields the corresponding output pattern $\mathbf{y}^{\gamma,d}$. For computational reasons, $\mathbf{x}^\gamma$ is modeled as a discrete Dirac-δ (impulse) function. A defuzzification of $\mathbf{y}^{\gamma,d}$ using the centroid method yields $s^{\gamma,d}$.

As before, we generated the vectors $\mathbf{s} = (s^1, \ldots, s^p)^T$ and $\mathbf{s}^d = (s^{1,d}, \ldots, s^{p,d})^T$. We employed the MATLAB function FMINBND to solve the optimization problem given by (15). This optimization process resulted in the parameter $d = 4.3568$. Figure 3 depicts $\left\|\mathbf{s} - \mathbf{s}^d\right\|_2$ for $d \in (0, 10]$. For the training data concerning the years $1931 - 1990$, the MSE of the Yager IFAM with $d = 4.3568$ is 12087.3 whereas the MSE for the Lukasiewicz IFAM, i.e., the Yager IFAM with $d = 1$, is 12387.4.

Table 3 provides the MSEs, the MPEs, and the MAEs produced by some IFAMs and other models during the testing phase. The values of the periodic auto-regressive moving average model (PARMA) [21], the fuzzy neural network NFN [28], and the predictive fuzzy clustering method FPM-PRP were drawn from the literature [26].

Figure 4 shows the forecasted streamflows estimated by the prediction model based on the FMAM for the Furnas reservoir from 1991 to 1998. Table 3 compares the errors that were generated by the IFAMs and several other models [26]. In contrast to the IFAM models, the MLP, NFN, and FPM-PRP models were initialized by optimizing the number of the parameters for each monthly prediction. For example, the MLP considers 4 antecedents to predict the streamflow of January and 3 antecedents to predict the streamflow for February. Moreover, the FPM-PRP model also takes advantage of slope information which requires some additional "fine tuning". Nevertheless, the Yager IFAM resulting from the min-

ISBN: 978-989-95079-6-8

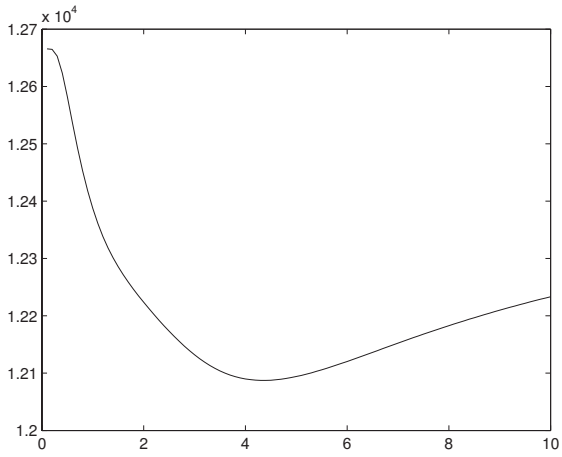


Figure 3: The distance $\left\|\mathbf{s}-\mathbf{s}^{d}\right\|_{2}$ between the actual streamflows given by the training data and the streamflows predicted by the Yager IFAM as a function of d.

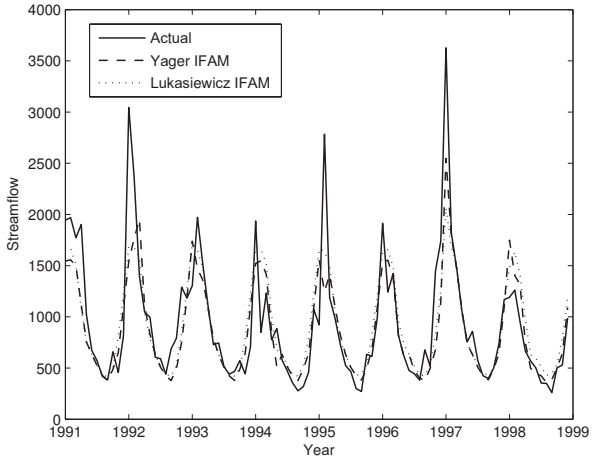


Figure 4: The streamflow prediction for the Furnas reservoir from 1991 to 1998. The continuous line corresponds to the actual values and the dashed line corresponds to the prediction of the Yager IFAM with $d = 4.3568$. The dotted line refers to the values predicted by the Lukasiewcz IFAM.

Table 3: Mean square, mean absolute, and mean relative percentage errors produced by the prediction models.

Method	MSE ($\times 10^5$)	MAE (m^3/s)	MPE (%)
FPM-PRP	1.20	200	18
Yager IFAM ($d = 4.3568$)	1.28	216	21
Lukasiewicz IFAM	1.27	229	24
PARMA	1.85	280	28
MLP	1.82	271	30
NFN	1.73	234	20

imization performed in (15) and the Lukasiewicz IFAM produced very satisfactory predictions that are visualized in Figure 4. Note that in the testing phase the MSE produced by the Yager IFAM with $d = 4.3568$ is slightly higher than the one produced by the Lukasiewicz IFAM although the Yager IFAM with $d = 4.3568$ outperforms the Lukasiewicz IFAM in terms of MAE and MPE. This fact may be due to overfitting to the training data. With some fine tuning of the parameters, it is possible to experimentally determine a Lukasiewicz IFAM (and thus a Yager IFAM) that outperforms the FPM-PRP with respect to the three error measures MSE, MAE, and MPE [15]. In our opinion, however, the fine tuning of the parameters should preferably not be performed experimentally but should be part of the optimization process.

5 Concluding Remarks

This paper represents the first attempt of tackling the problem of selecting the best IFAM for a given application. To this end, we introduced parameterized IFAM models, specifically the Yager class of IFAMs, and we formulated the problem of determining the Yager IFAM that yields the best fit to the training data as an optimization problem.

Recall that an IFAM is uniquely determined by a continuous t-norm, such as a Yager t-norm T^d with parameter d for $0 < d < \infty$, which gives rise to an R-implication. Although other continuous parameterized t-norms could have been chosen [29], the Yager t-norms T^d have the advantage that T^d approaches T_W, the pointwise smallest t-norm, for $d \to 0$ and T^d approaches T_M, the pointwise largest t-norm, for $d \to \infty$. Moreover, T^1 equals the Lukasiewicz t-norm which implies that the Lukasiewicz IFAM, i.e., the IFAM that exhibited the best performance in our previous experiments, is taken into account in the optimization process.

We applied our approach to two prediction problems from the literature. In the first problem concerning the prediction of the manpower requirement in steel manufacturing industry in West Bengal, India, only training data are available. The minimization of the MSE for the Yager IFAMs resulted in an infinite number of Yager IFAMs corresponding to $d \in (0, 1]$ that exhibit the same errors on the training data as the Lukasiewicz IFAM ($d = 1$). For the streamflow prediction problem in a major Brazilian hydroelectric plant, we dispose of training data and test data. The optimization process generated a Yager IFAM that outperforms the Lukasiewicz IFAM with respect to the MSE on the training data but exhibits a slightly higher MSE on the test data. This result may be due to overfitting and in fact we know that better Yager IFAMs exist for this problem. Nevertheless, the particular Yager IFAM derived from the optimization process exhibits very satisfactory results (which are better than those of the Lukasiewicz IFAM with respect to the MAE and MPE error measures), especially in view of the fact that we used a fixed number of antecedents and other parameters in conjunction with the Yager IFAM.

In the near future, we intend to incorporate the selection of additional parameters into the optimization process and we intend to use validation and/or regularization techniques to avoid overfitting. Furthermore, we plan to investigate other classes of parameterized IFAMs that are based on other types of parameterized t-norms.

ISBN: 978-989-95079-6-8

References

[1] P. Sussner and M. Valle. Implicative fuzzy associative memories. *IEEE Transactions on Fuzzy Systems*, 14(6):793–807, 2006.

[2] M. Valle and P. Sussner. A general framework for fuzzy morphological associative memories. *Fuzzy Sets and Systems*, 159(7):747–768, 2008.

[3] M. Nachtegael and E. Kerre. Connections between binary, gray-scale and fuzzy mathematical morphologies. *Fuzzy Sets and Systems*, 124(1):73–85, Nov. 2001.

[4] P. Sussner and M. Valle. Classification of fuzzy mathematical morphologies based on concepts of inclusion measure and duality. *Journal of Mathematical Imaging and Vision*, 32(2):139–159, 2008.

[5] P. Sussner and M. Valle. Fuzzy associative memories and their relationship to mathematical morphology. In W. Pedrycz, A. Skowron, and V. Kreinovich, editors, *Handbook of Granular Computing*, chapter 33. John Wiley and Sons, Inc., New York, 2008.

[6] B. Kosko. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, Englewood Cliffs, N.J., 1992.

[7] F. Junbo, J. Fan, and S. Yan. A learning rule for fuzzy associative memories. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, volume 7, pages 4273 – 4277, June 1994.

[8] P. Liu. The fuzzy associative memory of max-min fuzzy neural networks with threshold. *Fuzzy Sets and Systems*, 107:147–157, 1999.

[9] R. Belohlávek. Fuzzy logical bidirectional associative memory. *Information Sciences*, 128(1-2):91–103, 2000.

[10] W. Bandler and L. Kohout. Fuzzy relational products and fuzzy implication operators. In *International Workshop on Fuzzy Reasoning Theory and Applications*, Queen Mary College, University of London, London, 1978.

[11] L. Běhounek and M. Daňková. Relational compositions in fuzzy class theory. *Fuzzy Sets and Systems*, 160(8):1005–1036, April 2009.

[12] J. Drewniak. Fuzzy relation equations and inequalities. *Fuzzy Sets and Systems*, 14(3):237–247, December 1984.

[13] M. Valle, P. Sussner, and F. Gomide. Introduction to implicative fuzzy associative memories. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, pages 925 – 931, Budapest, Hungary, July 2004.

[14] H. Heijmans. *Morphological Image Operators*. Academic Press, New York, NY, 1994.

[15] P. Sussner and M. Valle. Morphological and certain fuzzy morphological associative memories for classification and prediction. In V. Kaburlassos and G. Ritter, editors, *Computational Intelligence Based on Lattice Theory*, volume 67, pages 149 – 173. Springer Verlag, Heidelberg, Germany, 2007.

[16] R. Yager. On a general class of fuzzy connectives. *Fuzzy Sets and Systems*, 4(3):235–242, 1980.

[17] G. Klir and T. Folger. *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, Englewood Cliffs, NJ, 1988.

[18] C. Lin and C. Lee. *Neural Fuzzy Systems*. Prentice Hall, Upper Saddle River, NJ, 1996.

[19] J. Choudhury, B. Sarkar, and S. Mukherjee. Forecasting of engineering manpower through fuzzy associative memory neural network with arima: a comparative study. *Neurocomputing*, 47:241–257, 2002.

[20] B. Kosko. *Fuzzy Engineering*. Prentice Hall, Upper Saddle River, N.J., 1992.

[21] J. Box, G. Jenkins, and C. Reinsel. *Time Series Analysis, Forecasting, and Control*. Holden-Day, Oakland, CA, third edition, 1994.

[22] T. Mills. *Time Series Techniques for Economists*. Cambridge University Press, Cambridge, United Kingdom, 1990.

[23] F. Chung and T. Lee. On fuzzy associative memory with multiple-rule storage capacity. *IEEE Transactions on Fuzzy Systems*, 4(3):375–384, August 1996.

[24] G. Ritter, P. Sussner, and J. Diaz de Leon. Morphological associative memories. *IEEE Transactions on Neural Networks*, 9(2):281–293, 1998.

[25] P. Sussner and M. Valle. Grayscale morphological associative memories. *IEEE Transactions on Neural Networks*, 17(3):559–570, May 2006.

[26] M. Magalhães, R. Ballini, R. Gonçalves, and F. Gomide. Predictive fuzzy clustering model for natural streamflow forecasting. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 390–394, Budapest, Hungary, 2004.

[27] S. Chiu. Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*, 1994.

[28] M. Figueiredo and F. Gomide. Adaptive neuro-fuzzy modeling. *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, 3:1567–1572, 1997.

[29] E. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, 2001.

ISBN: 978-989-95079-6-8

Defining A Fuzzy Partition for Coarseness Modelling in Texture Images

J. Chamorro-Martínez, P. Martínez-Jiménez[1] J.M. Soto-Hidalgo [2]

1.Department of Computer Science and Artificial Intelligence, University of Granada
Granada, Spain
2.Department of Computer Architecture, Electronics and Electronic Technology, University of Córdoba
Córdoba, Spain
Email: {jesus,pedromartinez}@decsai.ugr.es, jmsoto@uco.es

Abstract— *In this paper, the texture feature "coarseness" is modelled by means of a fuzzy partition on the domain of coarseness measures. The number of linguistic labels to be used, and the parameters of the membership functions associated to each fuzzy set are calculated relating representative coarseness measures (our reference set) with the human perception of this texture property. A wide variety of measures is studied, analyzing its capability to discriminate different coarseness categories. Data about the human perception of fineness is collected by means of a pools, performing an aggregation of their assessments by means of OWA operators. This information is used to obtain a fuzzy partition adapted to the human perception of coarseness-fineness*

Keywords— Coarseness, fineness, fuzzy partition, fuzzy texture, image features, texture features.

1 Introduction

For analyzing an image several kind of features can be used. From all of them, texture is one of the most popular and, in addition, one of the most difficult to characterize due to its imprecision. For describing texture, humans use vague textural properties like *coarseness/fineness*, *orientation* or *regularity* [1, 2]. From all of them, the *coarseness/fineness* is the most common one, being usual to associate the presence of fineness with the presence of texture. In this framework, a *fine* texture corresponds to small texture primitives (e.g. the image in figure 1(A)), whereas a *coarse* texture corresponds to bigger primitives (e.g. the image in figure 1(I)).

There are many measures in the literature that, given an image, capture the fineness (or coarseness) presence in the sense that the greater the value given by the measure, the greater the perception of texture [3]. However, given a certain measure value, there is not an immediate way to decide whether there is a fine texture, a coarse texture or something intermediate; in other words, there is not a textural interpretation.

To face this problem, fuzzy logic has been recently employed for representing the imprecision related to texture. In many of these approaches, fuzzy logic is usually applied just during the process, being the output a crisp result [4, 5]. Other approaches try to model the texture and its semantic by means of fuzzy sets defined on the domain of a given texture measure. In this last framework, some proposals model the texture property by means of an unique fuzzy set [6], and other approaches define fuzzy partitions providing a set of linguistic terms [7, 8].

Focusing our study in the last type of approaches, two questions need to be faced for defining properly a fuzzy partition: (i) the number of linguistic labels to be used, and (i) the parameters of the membership functions associated to each fuzzy set (and, consequently, the kernel localization). However, these question are not treat properly in the literature. Firstly, the number of fuzzy sets are often chosen arbitrarily, without take into account the capability of each measure to discriminate between different categories. Secondly, in many of the approaches, just an uniform distribution of the fuzzy sets is performed on the domain of the measures, although is wellknown that measure values corresponding to representative labels are not distributed uniformly. In addition, from our knowledge, none of the fuzzy approaches in the literature considers the relationship between the computational feature and the human perception of texture, so the labels and the membership degrees do not necessarily will match with the human assessments.

In this paper, we propose a fuzzy partition taking into account the previous questions. Firstly, in order to select the number of linguistic labels, we analyze the ability of each measure to discriminate different coarseness categories. For this purpose, data about the human perception of fineness is collected by means of a pools. This information is also used to localize the position and size of the kernel of each fuzzy set, obtaining a fuzzy partition adapted to the human perception of coarseness-fineness

The rest of the paper is organized as follows. In section 2 we present our methodology to obtain the fuzzy partition. Results are shown in section 3, and the main conclusions and future work are sumarized in section 4.

2 Fuzzy Partitions for Coarseness

As it was pointed, there is not a clear perceptual interpretation of the value given by a fineness measure. To face this problem, we propose to define a fuzzy partition on the domain of a given fineness measure. For this purpose, several questions will be faced: (i) what reference set should be used for the fuzzy partition, (ii) how many fuzzy sets will compound the partition, and (ii) how to obtain the membership functions for each fuzzy set.

Concern to the reference set, we will define the partition on the domain of a given coarseness-fineness measure. From now on, we will note $\mathcal{P} = \{P_1, \ldots, P_K\}$ the set of K measures analyzed in this paper, Π_k the partition defined on the domain of $\mathcal{P}_k$, N_k the number of fuzzy sets which compounds the partition Π_k, and $\mathcal{T}_k^i$ the i-th fuzzy set in Π_k. In this paper, the set $\mathcal{P} = \{P_1, \ldots, P_K\}$ is formed by the $K = 17$ mea-

ISBN: 978-989-95079-6-8

Figure 1: Some examples of images with different degrees of fineness

sures shown in the first column of table 1. It includes classical statistical measures, frequency domain approaches, fractal dimension analysis, etc. All of them are automatically computed from the texture image.

With regard to the number of fuzzy sets which compounds the partition, we will analyze the ability of each measure to distinguish between different degrees of fineness. This analysis will based on how the human perceives the fineness-coarseness. To get information about human perception of fineness, a set of images covering different degrees of fineness will be gathered. This images will be used to collect, by means of a pool, human assessments about the perceived fineness. From now on, let $\mathcal{I} = \{I_1, \ldots, I_N\}$ be the set of N images representing fineness-coarseness examples, and let $\Gamma = \{v^1, \ldots, v^N\}$ be the set of perceived fineness values associated to $\mathcal{I}$, with v^i being the value representing the degree of fineness perceived by humans in the image $I_i \in \mathcal{I}$. The way to obtain Γ will be describe in section 2.1

Using the data about human perception, and the measures values obtained for each image $I_i \in \mathcal{I}$, we will apply a set of multiple comparison tests in order to obtain the number of fineness degrees that each measure can discriminate (section 2.2). In addition, with the information given by the tests, we will define the fuzzy sets which will compund the partition (2.3).

2.1 *Assessment collection*

In this section, the way to obtain the set $\Gamma = \{v1, \ldots, v^N\}$ of perceived fineness values associated to $\mathcal{I}$ will be described. For this purpose, firstly the image set $\mathcal{I}$ will be selected. After that, a poll for getting assessments about the perception of fineness will be designed. Finally, for a given image, the assessments of the different subjects will be aggregated.

2.1.1 *The texture image set*

A set $\mathcal{I} = \{I_1, \ldots, I_N\}$ of $N = 80$ images representative of the concept of *fineness* has been selected. Figure 1 shows some images extracted from the set $\mathcal{I}$. Such set has been selected satisfying the following properties:

1. It covers the different presence degrees of fineness.
2. The number of images for each presence degree is representative enough.
3. Each image shows, as far as possible, just one presence degree of fineness.

Due to the third property, each image can viewed as "homogeneous" respect to the fineness degree represented, i.e., if we select two random windows (with a dimension which does not "break" the original texture primitives and structure), the perceived fineness will be the same for each window (and also respect to the original image). In other words, we can see each image $I_i \in \mathcal{I}$ as a set of lower dimension images (windows) with the same fineness degree of the original one.

As we explained, given an image $I_i \in \mathcal{I}$, a set of measures $\mathcal{P}$ will be applied on it. In fact, and thanks to the third property, we really can apply these measures for each subimage, assuming that the human assessment associated to that subimage will be the human assessment associated to the whole image. From now on, we will note as $\mathbf{M}_w^i = [m_1^{i,w}, \ldots, m_{K'}^{i,w}]$ the vector of measures for the w-th window of the image I_i, with $m_k^{i,w}$ being the result of applying the measure $P_k \in \mathcal{P}$ to the w-th window of the image I_i.

2.1.2 *The poll*

Given the image set $\mathcal{I}$, the next step is to obtain assessments about the perception of fineness from a set of subjects. From now on we shall note as $\Theta^i = [o_1^i, \ldots, o_L^i]$ the vector of assessments obtained from L subjects for the image I_i. To get Θ^i, subjects will be asked to assign images to classes, so that each class has associated a perception degree of fineness. In particular, 20 subjects have participated in the poll and 9 classes have been considered. The first nine images in figure 1 show the nine representative images for each class used in this poll. It should be noticed that the images are decreasingly ordered according to the degree of fineness.

As result, a vector of 20 assessments $\Theta^i = [o_1^i, \ldots, o_{20}^i]$ is obtained for each image $I_i \in \mathcal{I}$. The degree o_j^i associated to the assessment given by the subject S_j to the image I_i is computed as $o_j^i = (9 - k) * 0.125$, where $k \in \{1, \ldots, 9\}$ is the index of the class C_k to which the image is assigned by the subject.

ISBN: 978-989-95079-6-8

Algorithm 1 Distinguishable clusters selection

Input:

$Part^0 = C_1, C_2, \ldots, C_n$: Initial Partition

δ: distance function between clusters

ϕ: Set of multiple comparison tests

NT: Number of positive tests to accept distinguishability

1.-Initialization

$k = 0$

$distinguishable = false$

2.- While (distinguishable = false) and (k < n)

Apply the multiple comparison tests ϕ to $Part^k$

If for each pair $C_i, C_j \in Part^k$ more than NT of the multiple comparison tests ϕ show distinguishability

$distinguishable = true$

Else

Search for the pair of clusters C_r, C_{r+1}, verifying $\delta(C_r, C_{r+1}) = \min\{\delta(C_i, C_{i+1}), C_i, C_{i+1} \in Part^k\}$

Join C_r and C_{r+1} on a cluster $C_u = C_r \cup C_{r+1}$

$Part^{k+1} = Part^k - C_r - C_{r+1} + C_u$

$k = k + 1$

3.- Output: $\widetilde{Part}_k = C_1, C_2, \ldots, C_{n-k}$

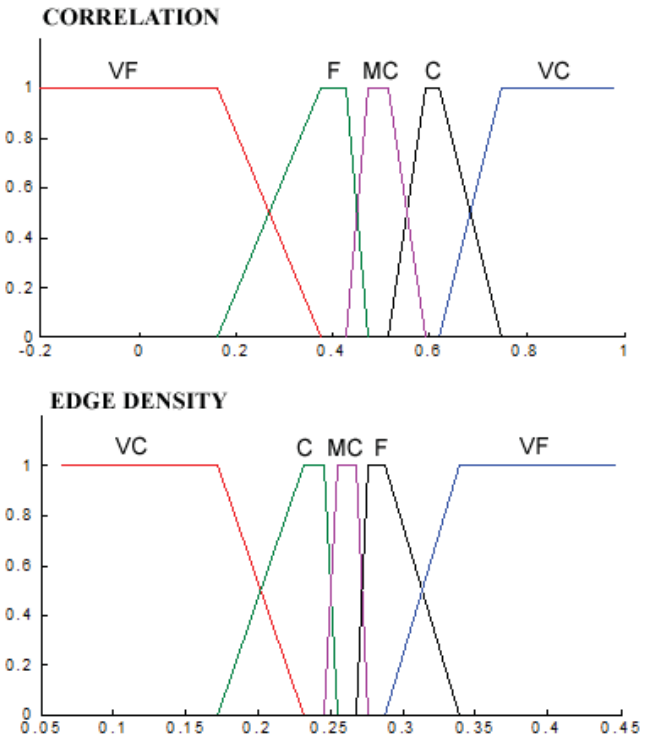


Figure 2: Fuzzy partitions for the measures Correlation and Edge Density. The linguistic labels are VC = very coarse, C = coarse, MC = medium coarse, F = fine, VF = very fine

2.1.3 *Assessment aggregation*

Our aim at this point is to obtain, for each image in the set $\mathcal{I}$, one assessment v^i that summarizes the assessments Θ^i given by the different subjects about the presence degree of fineness.

To aggregate opinions we have used an OWA operator guided by a quantifier [9]. Concretely, the quantifier "the most" has been employed, which allows to represent the opinion of the majority of the polled subjects. This quantifier is defined as

$$Q(r) = \begin{cases} 0 & \text{if } r < a, \\ \frac{r-a}{b-a} & \text{if } a \leq r \leq b, \\ 1 & \text{if } r > b \end{cases} \quad (1)$$

with $r \in [0, 1]$, $a = 0.3$ and $b = 0.8$. Once the quantifier Q has been chosen, the weighting vector of the OWA operator can be obtained following Yager [9] as $w_j = Q(j/L) - Q((j-1)/L), j = 1, 2, ..., L$. According to this, for each image $I_i \in \mathcal{I}$, the vector Θ^i obtained from L subjects will be aggregated into one assessment v^i as follows:

$$v^i = w_1\hat{o}^i_1 + w_2\hat{o}^i_2 + ... + w_L\hat{o}^i_L \quad (2)$$

where $[\hat{o}^i_1, \ldots, \hat{o}^i_L]$ is a vector obtained by ranking in nonincreasing order the values of the vector Θ^i.

2.2 *Distinguishability Analysis of the Fineness Measures*

As it was expected, some measures have better ability to represent fineness-coarseness than the others. To study the ability of each measure to discriminate different degrees of fineness-coarseness (i.e. how many classes can P_k actually discriminate), we propose to analyze each $P_k \in \mathcal{P}$ by applying a set of multiple comparison tests following the algorithm 1. This algorithm starts with an initial partition[1] and iteratively joins clusters until a partition in which all classes are distinguishable is achieved. In our proposal, the initial partition will be formed by the 9 classes used in our poll (where each class will contain the images assigned to it by the majority of the subjects), as δ the Euclidean distance between the centroids of the involved classes will be used, as ϕ a set of 5 multiple comparison tests will be considered (concretely, the tests of Scheffé, Bonferroni, Duncan, Tukey's least significant difference, and Tukey's honestly significant difference [10]), and finally the number of positive tests to accept distinguishability will be fixed to $NT = 3$.

From now on, we shall note as $\Upsilon_k = C^k_1, C^k_2, \ldots, C^k_{N_k}$ the N_k classes that can be discriminated by P_k. For each C^k_r, we will note as $\bar{c}^k_r$ the class representative value. In this paper, we propose to compute $\bar{c}^k_r$ as the mean of the measure values in the class C^k_r.

Table 1 shows the parameters obtained by applying the proposed algorithm 1 with the different measures considered in this paper. The second column of this table shows the N_k classes that can discriminated each measure and the third column shows how the initial classes have been grouped. The columns from fourth to eighth show the representative values $\bar{c}^k_r$ associated to each cluster.

2.3 *The Fuzzy Partitions*

In this section we will deal with the problem of defining the membership function for each fuzzy sets compounding the partition. As it was explained, the number of fuzzy sets will be given by the number of categories that each measure can discriminate.

In this paper, trapezoidal functions are used for defining the

[1] Let us remark that this partition is not the "fuzzy partition". In this case, the elements are measure values and the initial clusters the ones given by the pool

ISBN: 978-989-95079-6-8

Table 1: Result obtained by applying the algorithm 1

Measure	N_k	Classes	$\bar{c}_5 \pm KW_5/2$	$\bar{c}_4 \pm KW_4/2$	$\bar{c}_3 \pm KW_3/2$	$\bar{c}_2 \pm KW_2/2$	$\bar{c}_1 \pm KW_1/2$
Correlation [3]	5	{1,2-4,5-6,7-8,9}	0.122±0.038	0.403±0.0272	0.495±0.0225	0.607±0.0133	0.769±0.0210
ED [11]	5	{1,2,3-5,6-8,9}	0.348±0.0086	0.282±0.0064	0.261±0.0063	0.238±0.0066	0.165±0.0061
Abbadeni [12]	4	{1,2-6,7-8,9}	-	5.672±0.2738	9.208±0.4247	11.12±0.2916	25.23±1.961
Amadasun [1]	4	{1,2-6,7-8,9}	-	4.864±0.271	7.645±0.413	9.815±0.230	19.62±1.446
Contrast [3]	4	{1,2-5,6-8,9}	-	3312±265.5	2529±295.5	1863±94.84	790.8±129.4
FD [13]	4	{1,2,3-8,9}	-	3.383±0.0355	3.174±0.0282	2.991±0.0529	2.559±0.0408
Tamura [2]	4	{1,2-6,7-8,9}	-	1.540±0.0634	1.864±0.0722	2.125±0.0420	3.045±0766
Weszka [14]	4	{1,2-6,7-8,9}	-	0.153±0.0064	0.113±0.0093	0.099±0.0036	0.051±0.0041
DGD [15]	3	{1,2-8,9}	-	-	0.020±0.0010	0.038±0.0017	0.091±0.0070
FMPS [16]	3	{1,2-8,9}	-	-	0.256±0.0477	0.138±0.0122	0.0734±0.0217
LH [3]	3	{1,2-8,9}	-	-	0.023±0.0010	0.052±0.0025	0.127±0.0096
Newsam [17]	3	{1,2-6,7-9}	-	-	0.1517±0.0425	0.2654±0.0466	0.4173±0.0497
SNE [18]	3	{1,2-8,9}	-	-	0.879±0.0182	0.775±0.0087	0.570±0.0232
SRE [19]	3	{1,2-8,9}	-	-	0.995±0.00026	0.987±0.00066	0.966±0.0030
Entropy [3]	2	{1,2-9}	-	-	-	9.360±0.124	8.656±0.301
Uniformity[3]	2	{1,2-9}	-	-	-	$1.3_{E^{-4}}$ ±$2.6_{E^{-5}}$	$3.9_{E^{-4}}$ ±$1.9_{E^{-4}}$
Variance[3]	1	-	-	-	-	-	-

membership functions. In addition, a fuzzy partition in the sense of Ruspini is proposed. Figure 2 shows some examples of the type of fuzzy partition used. To establish the localization of each kernel, the representative value $\bar{c}_r^k$ will be used (in our case, the mean). Concretely, this value will be localized at the center position of the kernel.

To establish the size of the kernel, we propose a solution based on the multiple comparison tests used in section 2.2. As it is known, in these tests confidence intervals around the representative value of each class are calculated (being accomplish that these intervals do not overlap for distinguishable classes). All values in the interval are considered plausible values for the estimated mean. Based on this idea, we propose to set the kernel size as the size of the confidence interval.

The confidence interval CI_r^k for the class C_r^k is defined as

$$CI_r^k = \bar{c}_r^k \pm 1.96 \frac{\bar{\sigma}_r^k}{\sqrt{||C_r^k||}} \quad (3)$$

where $\bar{c}_r^k$ is the class representative value, and $\bar{\sigma}_r^k$ is the estimated standard deviation for the class. Thus, the kernel size KW_r^k is

$$KW_r^k = 3.92 \frac{\bar{\sigma}_r^k}{\sqrt{||C_r^k||}} \quad (4)$$

and the endpoints of the kernel will be given by $\bar{c}_r^k \pm KW_r^k/2$ Table 1 shows these values for each measure and each class.

Figure 2 shows the fuzzy partitions for the measures of correlation and ED (the ones with higher capacity to discriminate fineness clases).

3 Results

In this section, the fuzzy partition defined for the measure "Correlation" (showed in Figure 2) will be applied in order to analyze the performance of the proposed model.

Let's consider Figure 3(A) corresponding to a mosaic made by several images, each one with a different increasing degree of fineness. Figure 3(B-F) shows the membership degree to the fuzzy sets "very coarse", "coarse", "medium coarse", "fine" and "very fine", respectively, using the proposed model. For each pixel in the original image, a centered window of size 32×32 has been analyzed and its membership degree to each fuzzy set has been calculated. Thus, Figure 3(B) represents the degree in which the texture is perceived as "very coarse", with a white level meaning maximum degree, and a dark one meaning zero degree. It can be noticed that our model captures the evolution of the perception degrees of fineness.

Figure 4 presents an example where the proposed fuzzy partition has been employed for pattern recognition. In this case, Figure shows a microscopy image (Figure 4(A)) corresponding to the microstructure of a metal sample. The lamellae indicates islands of eutectic, which are to be separated from the uniform light regions. The brightness values in regions of the original image are not distinct, so texture information is needed for extracting the uniform areas. This fact is showed in Figure 4(B1,B2), where a thersholding on the original image is displayed (homogeneous regions cannot be separated from the textured ones as they "share" brightness values). Figure 4(C1) shows a mapping from the original image to its membership degree to the fuzzy set associated "very coarse". Thus, Figure 4(C1) represents the degree in which the texture is perceived as "very coarse" and it can be noticed that uniform regions correspond to areas with the maximum degree (bright grey levels), so if only the pixels with degree upper than 0.9 are selected, the uniform light regions emerge with ease (Figure 4(C2,C3)).

4 Conclusions and future works

In this paper, a fuzzy partition for representing the fineness/coarseness concept have been proposed. The number of fuzzy sets and the parameters of the membership functions have been defined relating fineness measures with the human perception of this texture property. Pools have been used for collecting data about the human perception of fineness, and the capability of each measure to discriminate different coarseness degrees has been analyzed. The results given by our approach show a high level of connection with the human perception of fineness/coarseness. As future work, the perfor-

ISBN: 978-989-95079-6-8

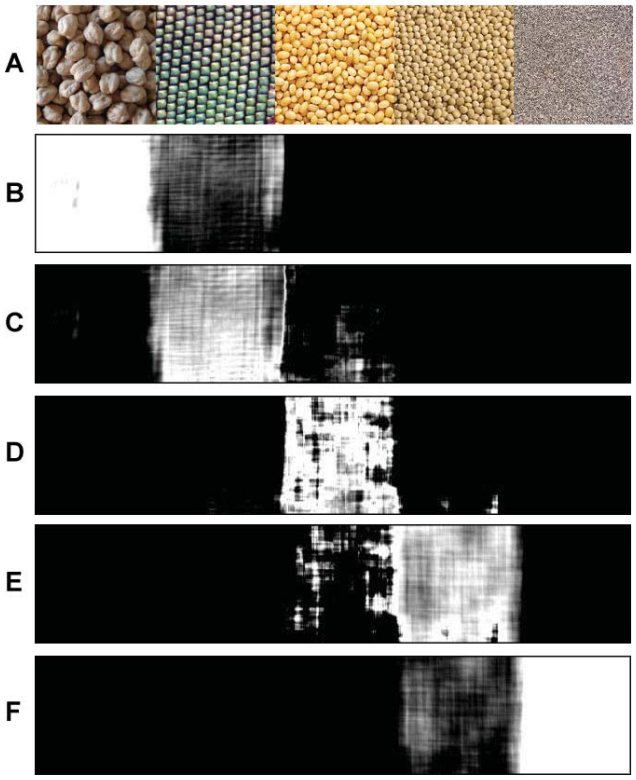


Figure 3: Results for a mosaic image. (A) Original image (B)(C)(D)(E)(F) Membership degree of each pixel to the sets ”very coarse”, ”coarse”, ”medium coarse”, ”fine” and ”very fine”, respectively (the darker the pixel, the lower the membership degree)

mance of the fuzzy partition will be analyzed in applications like textural classification or segmentation.

References

[1] M. Amadasun and R. King. Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man and Cybernetics*, 19(5):1264–1274, 1989.

[2] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8:460–473, 1978.

[3] R.M. Haralick. Statistical and structural approaches to texture. *Proceedings IEEE*, 67(5):786–804, 1979.

[4] M. Hanmandlu, V. K. Madasu, and S. Vasikarla. A fuzzy approach to texture segmentation. In *Proc. International Conference on Information Technology: Coding and Computing*, volume 1, pages 636–642, 2004.

[5] A. Barcelo, E. Montseny, and P. Sobrevilla. Fuzzy texture unit and fuzzy texture spectrum for texture characterization. *Fuzzy Sets and Systems*, 158:239–252, 2007.

[6] J. Chamorro-Martnez, E. Galn-Perales, J.M. Soto-Hidalgo, and B. Prados-Suarez. Using fuzzy sets for coarseness representation in texture images. In *Proceedings IFSA'07*, pages 783–792, 2007.

[7] S. Kulkarni and B. Verma. Fuzzy logic based texture queries for cbir. In *Proc. 5th Int. Conference on Computational Intelligence and Multimedia Applications*, pages 223–228, 2003.

[8] H.C. Lin, C.Y. Chiu, and S.N. Yang. Finding textures by textual descriptions, visual examples, and relevance feedbacks. *Pattern Recognition Letters*, 24(14):2255–2267, 2003.

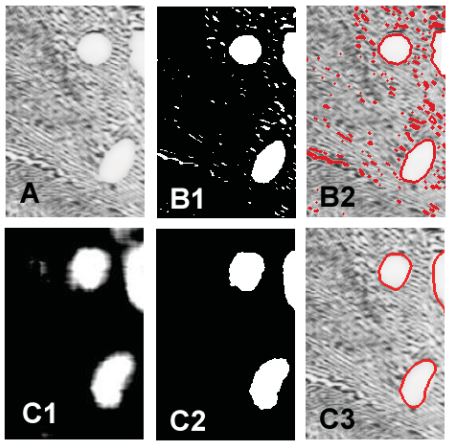


Figure 4: Example of pattern recognition (A) Original image (B1) Binary image obtained by thresholding the original one (B2) Region outlines of B1 superimposed on original image (C1) Membership degrees to the set ”very coarse” obtained with our model from the original image (C2) Binary image obtained by thresholding C1 (C3) Region outlines of C2 superimposed on original image

[9] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man and Cybernetics*, 18(1):183–190, 1988.

[10] Y. Hochberg and A.C. Tamhane. *Multiple Comparison Procedures*. Wiley, 1987.

[11] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[12] N. Abbadeni, N. Ziou, and D.S. Wang. Autocovariance-based perceptual textural features corresponding to human visual perception. In *Proc. of 15th International Conference on Pattern Recognition*, volume 3, pages 901–904, 2000.

[13] S. Peleg, J. Naor, R. Hartley, and D. Avnir. Multiple resolution texture analysis and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (4):518–523, 1984.

[14] J.S. Weszka, C.R. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man and Cybernetics*, 6:269–285, 1976.

[15] S.I. Kim, K.C. Choi, and D.S. Lee. Texture classification using run difference matrix. In *Proc. of IEEE 1991 Ultrasonics Symposium*, volume 2, pages 1097–1100, December 1991.

[16] H. Yoshida, D.D. Casalino, B. Keserci, A. Coskun, O. Ozturk, and A. Savranlar. Wavelet-packet-based texture analysis for differentiation between benign and malignant liver tumours in ultrasound images. *Physics in Medicine and Biology*, 48:3735–3753, 2003.

[17] S.D. Newsam and C. Kammath. Retrieval using texture features in high resolution multi-spectral satellite imagery. In *Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI, SPIE Defense and Security*, April 2004.

[18] C. Sun and W.G. Wee. Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics and Image Processing*, 23:341–352, 1983.

[19] M.M. Galloway. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, 4:172–179, 1975.

ISBN: 978-989-95079-6-8

A Fuzzy Particle Swarm Optimization Algorithm for a Cell Formation Problem

Esmaeil Mehdizadeh[1], Reza Tavakkoli-Moghaddam [2]

1Faculty of Industrial and Mechanical Engineering, Islamic Azad University, Qazvin Branch, Iran
2 Department of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran
Email: mehdizadeh@qazviniau.ac.ir

Abstract— *Group technology (GT) is a useful way to increase productivity with high quality in cellular manufacturing systems (CMSs), in which cell formation (CF) is a key step in the GT philosophy. When boundaries between groups are fuzzy, fuzzy clustering has been successfully adapted to solve the CF problem; however, it may result uneven distribution of parts/machines where the problem becomes larger. In this case, particle swarm optimization (PSO) can be used to tackle such a hard problem. This paper proposes a hybrid algorithm based on the fuzzy clustering and particle swarm optimization (FPSO) to solve the given CF problem. We experiment a number of examples to show the efficiency of the proposed algorithm and find that our proposed FPSO algorithm is able to obtain good results at reasonable time.*

Keywords— Cellular manufacturing systems, cell formation, fuzzy clustering, particle swarm optimization (PSO)

1 Introduction

Cell formation (CF) is a key step in Group technology (GT) that is used to design a good cellular manufacturing system by using the similarities of parts related to machines so that it can identify part families and machine groups. The parts in the same machine group have similar requirement so that GT can reduce travel and setup time. In CF the part/machine matrix, which has $m \times p$ dimensions with binary components, is usually described and given. The m rows indicate machines and p columns represent p parts that need to be operated upon. In matrix m, ‘‘1’’ (‘‘0’’) represents that this part should be (not) worked on the machine. The matrix exhibits parts requirement relative to machines. Our objective is to group parts and machines just like a cell.

Cell formation problem (CFP) is shown to be NP-hard [1] in the strong sense, and obtaining an optimal solution for the large-sized problems in reasonable computational time is extremely difficult.

Clustering involves the task of dividing data points into homogeneous classes or clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible. In real applications there is very often no sharp boundary between clusters so that fuzzy clustering is often better studied for the data. Membership degrees between zero and one are used in fuzzy clustering instead of crisp assignments of the data to clusters. In fuzzy clustering, the data points can belong to more than one cluster, and associated with each of the points are membership grades which indicate the degree to which the data points belong to the different clusters. In deterministic CF methods assumed well- defined boundaries between part–machine cells. These crisp boundary assumptions may fail to fully describe the case where the part–machine cell boundaries are fuzzy. This is why fuzzy clustering algorithms were applied for CF.

There are many CF methods in the literature [2]. The CF models can also be categorized into those of crisp or fuzzy. Crisp models assume that there are well-defined boundaries between groups and therefore assign each part or machine to only one family. In reality, some parts may belong to one part family, but there may have parts whose linkages are much less evident.

Various Clustering methods have been proposed to solve the CF problem. The fuzzy *c*-means (FCM) algorithm was first used in part-family formation by Xu and Wang [3]. FCM algorithm performs well with small and well-structured data sets. However, when the data set becomes larger, the algorithm may result in clustering errors, infeasible solutions, and uneven distribution of parts/machines. Then, several researchers have proposed alternatives to remedy these weaknesses with mixed success. For example, Chu and Hayya [4] improved the study carried out by Xu and Wang [3].

Al- Ahmari [5], Yang et al. [6] and Feng et al. [7] applied concepts of fuzzy clustering on the cell formation problem. Li et al. [8] improved fuzzy clustering algorithm to overcome the deficiencies of FCM. Since large instances are so difficult to optimally solve, approximate methods are needed. Perhaps, meta-heuristics are the most successful approximate methods that have been used so far. Thus, for example, Boctor [9] and Chen and Srivastava [10] used simulated annealing. Genetic algorithms have been used by Kazerooni et al. [11], Brown and Sumichrast [12], Ravichandran et al. [13]. Aljaber et al. [14] and Lozano et al. [15] applied tabu search. Attila [16] proposed an ant system algorithm. Zhao et al. [17] used swarm intelligence, and finally Andres and Lozano [18] presented particle swarm optimization (PSO) algorithm to solve the cell formation problem addressed in GT.

In the previous work, fuzzy clustering and PSO have been applied in the CF in separate. Our aim is to design a fuzzy particle swarm optimization (FPSO) clustering algorithm to solve the part–machine grouping problem,

ISBN: 978-989-95079-6-8

which is known as a hard combinatorial problem.

The rest of this paper is given bellow. The fuzzy cell formation model is presented in Section 2. In Section 3, the PSO algorithm is presented. The proposed FPSO algorithm is presented in Section 4. Computational results with a number of test problems taken from the literature are shown in Section 5. Finally, Section 6 draws conclusions, suggests directions for future research and discusses the limitations of the research.

2 Fuzzy clustering problem

The fuzzy cell formation (FCF) problem described by Li et al. [8] as follows: Given the routing information of n parts and m machines, the goal of CF is to cluster the parts into c part families and the corresponding machines into machine cells. The classification result can be expressed as a matrix $U=[\mu_{ik}]_{c\times n}$, (k= 1, 2,…, n and i= 1, 2,…, c), and μ_{ik} is the membership degree of part k to group i, which satisfies:

$$0 \le \mu_{ik} \le 1 \tag{1}$$

$$\sum_{i=1}^{c} \mu_{ik} = 1 \tag{2}$$

$$0 < \sum_{k=1}^{n} \mu_{ik} \le n \tag{3}$$

The FCM clustering algorithm is based on the minimization of the following equation:

$$J(p) = \sum_{k=1}^{n} \sum_{i=1}^{c} [\mu_{ik}]^m \|x_k - V_i\|^2 \tag{4}$$

where, m>1 is a real number governing the influence of membership grades, V_i is the cluster center of the part family i, and x_k is the vector of part k. The necessary conditions for minimizing $J(p)$ are the following update equations:

$$V_i = \frac{\sum_{k=1}^{n} [\mu_{ij}]^m x_k}{\sum_{k=1}^{n} [\mu_{ij}]^m} \tag{5}$$

which, c= 1, 2, … , n.

$$\mu_{ik}^{(t+1)} = \left[\sum_{j=1}^{c} \left(\frac{\|x_k - V_i^{(t)}\|^2}{\|x_k - V_j^{(t)}\|^2} \right)^{\frac{1}{m-1}} \right]^{-1} \tag{6}$$

where, $\|x_k - V_i\|^2$ represents the Euclidean distance between x_k and V_i , and $\mu_{ik}^{(t+1)}$ is the membership degree of part k in group i.

Many variations of FCM algorithms can be found in Bezdek [19]. The algorithm is based on the assumption that the desired number of clusters c, real number m, stopping criterion ε, and a particular distance are given.

Step 1: Let t=0 and select an initial fuzzy pseudo-partition $p^{(0)}$.

Step 2: Calculate c cluster centers, $V_1^{(t)}$, …, $V_c^{(t)}$, by (5) for $p^{(t)}$ and the chosen value of m.

Step 3: Define $\mu_i^{(t+1)}$ by (6) and update $p^{(t+1)}$

Step 4: Compare $p^{(t)}$ and $p^{(t=1)}$. If $\left| p^{(t+1)} - p^{(t)} \right| \le \varepsilon$, then stop the algorithm; otherwise, increase t by one and then return to Step 2.

Since the fuzzy clustering problem is a combinatorial optimization problem that is hard to solve [20]. Large instances are so difficult to optimally solve, approximate methods are then needed.

3 Particle swarm optimization

In particle swarm optimization (PSO) a number of simple entities—the particles—are placed in the search space of some problem or function, and each evaluates the objective function at its current location. Each particle then determines its movement through the search space by combining some aspect of the history of its own current and best (best-fitness) locations with those of one or more members of the swarm, with some random perturbations. The next iteration takes place after all particles have been moved. Eventually the swarm as a whole likes a flock of birds collectively foraging for food is likely to move close to an optimum of the fitness function.

There are many variants of the PSO proposed in the literature so far, when Eberhart and Kennedy [21] first introduced this technique. A version of this algorithm is used for part-machine grouping by Andres and Lozano [18].

For description of The PSO algorithm, first, let me define the notation adopted in this paper: the i-th particle of the swarm is represented by the D-dimensional vector $x_i=(x_{i1}, x_{i2}, \ldots, x_{iD})$ and the best particle of the swarm (i.e., the particle with the smallest function value) is denoted by index p_g. The best previous position (i.e., the position giving the best function value) of the i-th particle is recorded and represented $p_i=(\rho_{i1}, \rho_{i2}, \ldots, \rho_{iD})$, and the position change (velocity) of the i-th particle is $Vel_i=(Vel_{i1}, Vel_{i2}, \ldots, Vel_{iD})$. The particles are then manipulated according to the following equations:

$$Vel_{id}(t+1) = \chi\{wVel_{id}(t) + c_1\phi_1[\rho_{id}(t) - x_{id}(t)] + c_2\phi_2[(\rho_{gd}(t) - x_{id}(t)]\} \tag{7}$$

$$x_{id}(t+1) = x_{id}(t) + Vel_{id}(t+1) \tag{8}$$

where, d=1, 2, …, D ; i=1, 2, …, n; and n is the size of the swarm; w is the inertia weight; c_1 and c_2 are two positive acceleration constants; ϕ_1 and ϕ_2 are two random values into the range [0, 1]; χ is a constriction factor that is used in constrained optimization problems in order to control the magnitude of the velocity (in unconstrained optimization problems it is usually set to 1.0).

4 Proposed FPSO algorithm

In the fuzzy clustering, a single particle represents a cluster center vector and a swarm represents a number of candidates clustering for the current data vector. Here; each point or data vector belongs to every various cluster by different membership function, thus; assign a fuzzy membership to each point or data vector. Each cluster has a cluster center, and per iteration presents a solution that gives

ISBN: 978-989-95079-6-8

a vector of cluster centers. We determine the position of each vector for every particle and update it, then change the position of cluster centers based of particles. For the purpose of our algorithm, we define the following notations:

n	Number of part
c	Number of cluster center
$V_l^{(t)}$	Position of l-th particle' cluster center in stage t
$Vel_l^{(t)}$	Velocity of l-th particle in stage t
x_k	Vector of parts ($k = 1,2,.....,n$)
$\rho_l^{(t)}$	Best position funded by l-th particle in stage t
$\rho_g^{(t)}$	Best position funded by all particles in stage t
$P^{(t)}$	Fuzzy pseudo partition in stage t
$\mu_{ik}^{(t)}$	Membership function of k-th part in stage t into i-th cluster

The fitness of particles is easily measured by (4). The c-means algorithm tends to converge faster than the proposed FPSO algorithm with a less accurate clustering. In this section, the performance of the PSO clustering algorithm is improved by seeding the initial swarm with the result of the c-means algorithm. The FPSO algorithm first executes the c-means algorithm once. In this case, the c-means clustering algorithm is terminated by one of two stopping criteria: I) the maximum number of iterations; or II) $\left|p^{(t+1)} - p^{(t)}\right| \le \varepsilon$. The result of c-means algorithm is then used as one of the particles, while the rest of the swarms are initialized randomly. The following algorithm can use to finding cluster for each data vector or part:

Step 1: Let t=0, select initial parameters such as number of cluster center c, initial position of particle by the FCM, initial velocity of particles, c_1, c_2, w, χ, and a real number $m \in (1, \infty)$, and a small positive number ε for stopping criterion.
Step 2: Calculate $\mu_{ik}^{(t)}$ for all particles and all i=1, 2, …, c and k=1, 2, …, n by (6) and update $p^{(t+1)}$.
Step 3: For each particle, calculate the fitness by using (4).
Step 4: Update the global best and local best position.
Step 5: Update $Vel_l^{(t)}$ and $V_l^{(t)}$ for all l=1, 2, …, *n_particle* by using (7) and (8).
Step 6: Update $p^{(t+1)}$ by the Step 2.
Step 7: Compare $p^{(t)}$ and $p^{(t+1)}$. If $\left|p^{(t+1)} - p^{(t)}\right| \le \varepsilon$, then stop; otherwise, increase t by one and continue form Step 3.

5 Numerical example

In this section, examples from the literature are considered to illustrate the application of the proposed fuzzy algorithm in the cell formation problem. We compare the results of the FCM and FPSO algorithms on various problems taken from the literature. Their performances are measured by the objective function value given in (4) and CPU time. A general rule of thumb is that a clustering result with lower *J(p)* and lower CPU time is preferable. For a comparable assessment, we code these methods by using the fuzzy tools available in MATLAB 7 and the FPSO, respectively, with 10 particles, w=0.72, and $c_1 = c_2 = 1.49$. For our experimental tests, we use a PC Pentium III, CPU 1133 MHz and 256 MB of RAM for the same parameters for all algorithms implementations: m=2, the maximum number of iterations is 100 and $\varepsilon = 0.00001$.

Table 1: Data from Chu and Hayya [4]

	Parts								
Machines	1	2	3	4	5	6	7	8	9
1	1	1	0	0	1	0	0	0	0
2	1	1	0	0	0	1	0	0	1
3	0	0	1	0	0	0	1	1	0
4	0	1	1	1	0	0	0	1	0
5	1	0	0	1	1	0	0	1	0
6	0	1	0	0	0	1	0	0	1
7	0	0	1	0	0	0	1	1	0
8	0	0	1	1	1	0	1	1	0
9	0	1	0	0	0	1	0	0	1

Table 2: The membership matrix for cells for first example

Machines	Memberships for cells		
	1	2	3
1	.2239	.2755	.5006
2	.0912	.7838	.1250
3	.9335	.0249	.0417
4	.4413	.1871	.3716
5	.0712	.0392	.8895
6	.0160	.9666	.0174
7	.9335	.0249	.0417
8	.5359	.0947	.3694
9	.0160	.9666	.0174

Table 3: The membership matrix for part families for first example

Parts	Memberships for part families		
	1	2	3
1	.7110	.1656	.1234
2	.2502	.6234	.1264
3	.0698	.0441	.8861
4	.5026	.1370	.3603
5	. 7659	.0922	.1419
6	.0437	.9346	.0216
7	.2061	.1233	.6706
8	.0751	.0330	.8919
9	.0437	.9346	.0216

Table 4: Comparison of Chu and Hayya's approach and FPSO approach

	Chu and Hayya's approach		The FPSO approach	
	Machine cell	Part families	Machine cell	Part families
Cell 1	M1,M5	P1,P4,P5	M1,M5	P1,P4,P5
Cell 2	M2,M6,M9	P2,P6,P9	M2,M6,M9	P2,P6,P9
Cell 3	M3,M4 M7,M8	P3,P7,P8	M3,M4 M7,M8	P3,P7,P8
$J(p)$	3.729600	3.859139	3.729600	3.859139

ISBN: 978-989-95079-6-8

The first example taken from Chu and Hayya [4] consists of nine machines and nine parts is illustrated in Table 1. Tables 2 and 3 show the membership matrix values for cells and part families, respectively. The final membership matrix values indicates the degree of membership of each machine associated with the machine cell (MC) and can be configured as: MC (1) = {M1, M5}, MC (2) = {M2, M6, M9}, and MC (3) = {M3, M4, M7, M8}. Similarity, the part family (PF) can be configured as: PF (1) = {P1, P4, P5}, PF (2) = {P2, P6, P9}, and PF (3) = {P3, P7, P8}. Table 4 compares Chu and Hayya's approach with the FPSO approach results. As illustrated in the table, no changes are observed between machine cells and part families.

Table 5: Data from Susanto et al. [22]

Machine	Parts									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	0	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0	0
3	1	1	1	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0
5	0	0	0	0	1	1	1	0	0	0
6	0	0	0	0	1	1	1	0	0	0
7	0	0	0	1	1	1	0	0	0	0
8	0	0	0	0	0	0	0	1	1	1
9	0	0	0	0	0	0	0	1	1	1
10	0	0	0	0	0	0	0	1	1	0
11	1	1	0	0	1	1	0	0	0	0
12	1	1	0	0	1	1	0	0	0	0

The second example is taken from Susanto et al. [22] as shown in Table 5. The obtained results are illustrated in Tables 6 and 7 and can be configured as: MC (1) = {M5, M6, M7, M11, M12}, MC (2) = {M8, M9, M10}, and MC (3) = {M1, M2, M3, M4}, $J_{(p)} = 4.844414$. Part families are as follows: PF (1) = {P5, P6, P7}, PF (2) = {P8, P9, P10}, and PF (3) = {P1, P2, P3, P4}, J (p) = 5.020256.

Table 6: The membership matrix for cells for second example

Machines	Degree of membership		
	1	2	3
1	.0788	.0701	.8511
2	.0788	.0701	.8511
3	.0842	.0730	.8428
4	.1233	.1082	.7685
5	.9017	.0460	.0522
6	.9017	.0460	.0522
7	.6359	.1578	.2062
8	.0541	.8940	.0519
9	.0541	.8940	.0519
10	.01159	.7661	.1180
11	.4311	.1561	.4128
12	.4311	.1561	.4128

Table 7: The membership matrix for part families for second example

Parts	Degree of membership		
	1	2	3
1	0.9265	0.0315	0.0420
2	0.9265	0.0315	0.0420
3	0.6430	0.1975	0.1595
4	0.4342	0.2743	0.2916
5	0.0238	0.0236	0.9525
6	0.0238	0.0236	0.9525
7	0.1765	0.2902	0.5333
8	0.0218	0.9534	0.0248
9	0.0218	0.9534	0.0248
10	0.0702	0.8483	0.0815

Table 8: Solution to Al-Ahmari

Cell	Machine Cells	$J(p)$ FCM	CPU Time FCM	$J(p)$ FPSO	CPU Time FPSO
2	{1,3,4,9,10,13,14,16,17,20,21,22},{2,5,6,7,8,11,12,15,18,19,23,24}	54.583351	0.300	54.583351	0.190
3	{1,3,13,16,20,21,22},{6,7,9,10,14,17,23,24},{2,4,5,8,11,12,15,18,19}	36.388919	0.641	36.388919	0.170
4	{2,3,4,6,9,10,11,12,15,16,17,18},{1,8,14,20,23,24},{7,13,21,22},{5,19}	22.515392	0.490	22.515392	0.431
5	{1,13,21,22},{2,5,11,19},{3,4,7,14,16,20,23,24},{9,10,17},{6,8,12,15,18}	14.691166	0.852	14.691166	0.280
6	{3,20},{1,13,21,22},{2,5,11,19},{7,14,23,24},{4,6,8,12,15,16,18},{9,10,17}	9.046538	0.571	9.046538	0.350
7	{4,16},{9,10,17},{6,8,12,15,18},{3,20},{2,5,11,19},{1,13,21,22},{7,14,23,24}	6.172838	0.611	6.172838	0.581
8	{1,13,21,22},{7,23,24},{4,16},{14},{9,10,17},{6,8,12,15,18},{3,20},{2,5,11,19}	4.139911	0.651	4.139911	0.591
9	{9,10},{14},{17},{6,8,12,15,18},{3,20},{2,5,11,19},{1,13,21,22},{7,23,24},{4,16}	2.772339	0.671	2.772339	0.411
10	{4,16},{6,7,23,24},{17},{8,12,15,18},{2,5,11,19},{1,14},{13,21,22},{6},{3,20},{9,10}	1.694722	0.551	1.694722	0.541

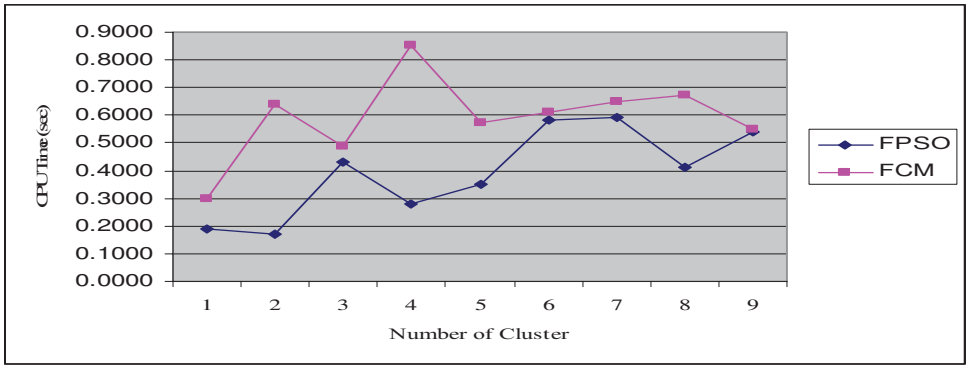


Figure 1: CPU Time comparison between FPSO and FCM algorithms

ISBN: 978-989-95079-6-8

The proposed approach is also tested on a large-sized problem taken from Al-Ahmari [5]. In the example, the problem with 24 machines and 40 parts is considered. Different numbers of manufacturing cells are used (i.e., 2, 3, 4, 5, 6, 7, 8, 9, and 10). The results summarized in table 8 and Figure 1. This example demonstrates the possibility of using the approach for large-scale CF problems, and evaluates the obtained results using $J(P)$.

6 Conclusion

Group technology (GT) is a useful way to increase productivity with high quality in flexible manufacturing systems; and CF is a key step in GT. It is used to design a good cellular manufacturing that uses the similarities of parts related to machines so that it can identify part families and machine groups. The cell formation problem is NP-complete and different -heuristic methods have been used to solve it. Particle swarm optimization is one of them. On the other hand, the crisp models fail to fully reflect the complex nature of part features or routing data, where boundaries between groups are fuzzy. Fuzzy clustering has been successfully adapted to solve the cell formation problem, but when the problem becomes larger, the fuzzy clustering algorithms may result uneven distribution of parts/machines. In the previous works, fuzzy clustering and particle swarm optimization have been applied in CF individually. We have designed a fuzzy particle swarm optimization clustering algorithm (FPSO) to solve the part–machine grouping problem, which is a hard combinatorial problem. The presented numerical examples confirm the effectiveness of the proposed approach. It is found that this algorithm provides a good solution to CF problems at reasonable time and allows the user in formulating the required size of machine cells and part families.

References

[1] A. Ballakur and H.J. Steudel, "A Within Cell Utilization Based Heuristic for Designing Cellular Manufacturing Systems", *Int. J. of Prod. Res.*, 25: 639-55, 1987.

[2] Y. Yin and K. Yasuda, "Similarity Coefficient Methods Applied to the Cell Formation Problem: A Taxonomy and Review", *Int. J. Prod. Economics,* 101: 329-352, 2006.

[3] H. Xu and H.P. Wang, "Part Family Formation for GT Applications Based on Fuzzy Mathematics", *Int. J. of Prod. Research,* 27: 1637-1651, 1989.

[4] C.H. Chu and J.C. Hayya, "A Fuzzy Clustering Approach to Manufacturing Cell Formation", *Int. J. of Prod. Research*, 29: 1475-1487, 1991.

[5] A.M.A. Al-Ahmari, "A Fuzzy Analysis Approach for Part-Machine Grouping in Cellular Manufacturing Systems", *Integrated Manufacturing System,* 13:7: 489-497, 2002.

[6] M.S. Yang, W.L. Hung and F.C. Cheng, "Mixed-Variable Fuzzy Clustering Approach to Part Family and Machine Cell Formation for GT Applications", *Int. J. Prod. Economics,* 103: 185-198, 2006.

[7] C.Y. Chen and F. Ye, "Adaptive Hyper-Fuzzy Partition Particle Swarm Optimization Clustering Algorithm", *Cybernetics and Systems,* 37:463-479, 2006.

[8] J. Li, C.H Chu, Y. Wang and W. Yan, "An Improved Fuzzy Clustering Method for Cellular Manufacturing*", Int. J. of Prod.. Research,* 45: 1049-1062, 2007.

[9] F.F. Boctor, "A Linear Formulation of the Machine-Part Cell Formation Problem", *Int. J. Prod. Res.,* 29: 343-56, 1991.

[10] W. Chen and B. Srivastava, "Simulated Annealing Procedures for Forming Machine Cells in Group Technology", *Eur. J. Oper. Res.*, 75: 100-111, 1994.

[11] M. Kazerooni, L. Luong and K. Abhary," Cell Formation using Genetic Algorithms. *Int. J. of Flexible Autom. Integr Manuf*, 3: 283-299, 1995.

[12] E.C. Brown and R.T. Sumichrast, "A Grouping Genetic Algorithm for the Cell Formation Problem", *Int. J. Prod. Res.,* 39: 3651-3669, 2001.

[13] K.S. Ravichandran, K. Chandra, S. Rao and R. Saravanan, "The Role of Fuzzy and Genetic Algorithms in Part Family Formation and Sequence Optimization for Flexible Manufacturing Systems", *Int. J. of Adv. Manuf. Technology,* 19: 879-888, 2002.

[14] N. Aljaber, W. Baek and C. L. Chen, "A Tabu Search Approach to the Cell Formation Problem", *Computers and Industrial Eng.*, 32: 169-185, 1997.

[15] S. Lozano, B. Adenso-Dı, I. Eguia, and L. Onieva, "A One-Step Tabu Search Algorithm for Manufacturing Cell Design", *J. of Oper. Res. Soc.,* 50: 509-516, 1999.

[16] A.A. Islier, "Group Technology by an Ant System Algorithm", *Int. J. of Prod. Research,* 43: 913-932, 2005.

[17] G. Zhao, P. Jiang, M. Zheng, G. Zhou, and B. Chen, "Using Swarm Intelligence Algorithm to Solve a Part Clustering Model Based on Weighted Directed Graph", *Int. J. of Prod. Res.,* 2007. A. Andres and S. Lozano, "A Particle Swarm Optimization Algorithm for Part–Machine Grouping", *Robotics and Computer-Integrated Manufacturing,* 22: 468-474, 2006.

[18] J.C. Bezdek, "*Pattern Recognition with Fuzzy Objective Function Algorithms*", Plenum Press: New York, 1981.32

[19] J.G. Klir and B. Yuan, "*Fuzzy Sets and Fuzzy Logic, Theory and Applications"*, Prentice Hall of India, 257-262, 2003.

[20] J. Kennedy and R.C. Eberhart, "Particle Swarm Optimization", In: *Proceedings of IEEE Int. Conf. on Neural Networks,* Perth, Australia, 1942-1948, 1995.

[21] S. Susanto, R.D. Kennedy and J.W.H. Price, "A New Fuzzy-*C*-Means and Assignment Technique-Based Cell Formation Algorithm to Perform Part-Type Clusters and Machine-Type Clusters Separately", *Prod. Planning & Control,* 10: 375-388, 1999.

ISBN: 978-989-95079-6-8

Multiple Negations in Fuzzy Interval Logic

Eunjin Kim[1*], Ladislav J. Kohout[2]

1.Dept. of Computer Science, University of North Dakota
Grand Forks, ND 58202-9015, USA
2.Dept. of Computer Science, Florida Sate University
Tallahassee, FL 32306-4530, USA
Email: ejkim@cs.und.edu, kohout@cs.fsu.edu

Abstract— *This paper continues our study in fuzzy interval logic based on the Checklist Paradigm(CP) semantics of Bandler and Kohout. We investigate the fuzzy interval system of negation which was defined by the Sheffer(*NAND*), the Nicod(*NOR*) and the implication connectives of m_1 interval system in depth. The top-bottom(*TOP-BOT*) pair of fuzzy negation interval shows non-involutive property; however, it shows various interesting properties as the negation is iterated to each* TOP *and* BOT *fuzzy negation connective. As the iteration reaches to infinity, the* TOP-BOT *pair of iterated negation shows nearly involutive property.*

Keywords— Checklist Paradigm, Fuzzy Interval, Fuzzy Negation, Non-Involution, Gap-Theorem.

1 Introduction

The convention that fuzzy logics default into 2-valued logic for the values 0 and 1 is generally accepted. The question how connectives of the same type differ outside the boundary points, within the open interval $(0,1)$ has been well researched. For example the implications such as Lukasiewicz($\overset{L}{\rightarrow}$), Kleene-Dienes($\overset{KD}{\rightarrow}$), Reichenbach($\overset{KDL}{\rightarrow}$), Goguen($\overset{G43}{\rightarrow}$), G$\ddot{o}$del($\overset{S*}{\rightarrow}$), Early-Zadeh($\overset{EZ}{\rightarrow}$) and Wilmott($\overset{W}{\rightarrow}$) are interrelated by the following inequalities [1]:

$$a \overset{S*}{\rightarrow} b \leq a \overset{G43}{\rightarrow} b \leq a \overset{L}{\rightarrow} b, \text{ and}$$

$$a \overset{W}{\rightarrow} b \leq a \overset{EZ}{\rightarrow} b \leq a \overset{KD}{\rightarrow} b \leq a \overset{KDL}{\rightarrow} b \leq a \overset{L}{\rightarrow} b, \text{ where}$$

$a \overset{L}{\rightarrow} b = min(1, 1-a+b)$; $a \overset{KDL}{\rightarrow} b = min(1, 1-a+ab)$; $a \overset{KD}{\rightarrow} b = max(1-a, b)$, etc[1]. However, much less is known about inter-relations of interval systems of connectives.

In 1979 Bandler and Kohout derived five interesting systems of fuzzy logic, $m_1, m_2, \ldots, m_5$, based on the Checklist paradigm [2]. Since then, the systems of connectives that can be generated from implicational intervals by group transformations have been investigated systematically; in particular, the m_1 interval logic system of 16 connectives has been investigated in depth. Among them, ten 2-argument connectives such as $\wedge, \vee, \rightarrow, \ldots \equiv$, *oplus*(XOR) yield the interval pairs of connectives ($conbot \leq m_1 \leq contop$) where its implication($\rightarrow$) yields Lukasiewicz and Kleene-Dienes implication for its TOP-BOT pair of interval, in particular. However, a unary connective, a negation, did not yield interval but just singleton: i.e. $\neg a = 1-a$. From this question, we proposed the alternative negations in [3] by the Sheffer(NAND), the Nicod(NOR) and a pair of <implication, 0>, which fuzzifies $\neg a$, generating the intervals like the fuzzified 2-ary connectives. The TOP-BOT pair of generated negation interval lacks of involutive property on the surface, but a question still remains if they show a nearly involutive property converging to a certain value as an iteration of negation proceeds.
Thus, we further investigate this non-involutive property of TOP-BOT pair of interval negations as well as the relationship between three definitions of negation which are equivalent in two-valued logic but bifurcate into different negations in interval many-valued logic in this paper.

[1]For the rest of definitions, refer to [1],[2]

2 Interval logic system generated by the checklist paradigm

The structure of five fuzzy interval systems m_1 - m_5, based on the Checklist paradigm by Bandler and Kohout in [2] is generated by a distinct measure that performs the *summarization* of the information contained in certain well-defined binary structures called *fine structures*. The interval produced by a measure m_i pair of connectives of one type can be generically characterized by the following inequality:

$$CON_{Bot} \leq m_i \leq CON_{Top}, \quad i \in \{1, 2, \ldots, 5\}$$

For the details of the checklist paradigm and its various uses, see the papers [2],[4],[5],[6],[7],[8].

2.1 Five Implication Operator Based Interval Systems of Bandler and Kohout

The following five inequalities linking the interval bounds for implication operators $[\rightarrow_{bot}, \rightarrow_{top}]$ with corresponding measures, m_i, $i = \{1,2,3,4,5\}$ have been listed in [2].

1. The Kleene-Dienes implication(KD) and Łukasiewicz implication (Ł) respectively, are attainable lower and upper bounds of m_1:
   $max(1-a, b) \leq m_1(\rightarrow) \leq min(1, 1-a+b)$.

2. A certain new function of (a,b) and the Goguen-Gaines (G43) implication (the left-hand side) are respectively attainable lower and upper bounds of m_2:
   $max(0, (a+b-1)/a) \leq m_2(\rightarrow) \leq min(1, b/a)$.

3. Another function of (a,b) and the Early Zadeh implication (EZ) are respectively attainable lower and upper bounds of m_3:
   $max(a+b-1, 1-a) \leq m_3(\rightarrow) \leq max[min(a,b), 1-a]$.

ISBN: 978-989-95079-6-8

4. Still another function of (a,b) and the Wilmott implication (W) respectively, are attainable lower and upper bounds of m_4:
$min[max(a+b-1,1-a),max(b,1-a-b)] \leq m_4(\rightarrow)$
$\leq min[max(1-a,b),(max(a,(1-b),min(b,1-a)))]$.

5. Yet another function of (a,b) and one of G43 respectively, are attainable lower and upper bounds of m_5:
$max[(a+b-1)/a,1-a] \leq m_5(\rightarrow) \leq max[min(1,b/a),1-a]$.

In m_1 system, a Kleene-Dinese logic system forms a a lower bound of the interval($m1(\rightarrow_{bot})$) while a Łukasiewicz logic system forms a upper bound of the interval ($m1(\rightarrow_{top})$).

From this implicational interval pair $[m1(\rightarrow_{bot}), m1(\rightarrow_{top})]$, one can generate all 16 pairs of connectives of corresponding interval systems by Kohout-Bandler group of logic transformations. 10 of these pairs of connectives are genuine interval pairs and remaining 6 connectives collapse into a single point, not an interval. It has been investigated systematically over the years in the series of papers [9], [10]. These 16 connectives of measure $m_1, \ldots, m_5$ can be generated by group transformation. 8 of these interval pair of connectives of m_1 system are shown in section 3.

3 Characterization of logics by group transformation

Definition 1 *Group Transformations in Logic:*
Let f be any one of the 10 two-argument propositional connectives of a logic system and $\neg$ be an involutive negation. Then, we define the following transformations over f:

1. $I(f) = f(x,y)$: *Identity transformation*
2. $D(f) = \neg f(\neg x, \neg y)$: *Dual transformation*
3. $C(f) = f(\neg x, \neg y)$: *Contradual transformation*
4. $N(f) = \neg f(x,y)$: *Negation transformation*

It is well known that for the crisp (2-valued) logic these $\mathcal{T} = \{I,D,C,N\}$ transformations determine the Piaget group which is a realization of the abstract Klein 4-element group.

The new 4 non-symmetrical transformations below which are added to the above 4 basic transformations by Bandler and Kohout in 1979 [11],[12] enriches the algebraic structure of logical transformations.

Definition 2 *8-element Group Transformations: Bandler-Kohout*

5. $LC(f) = f(\neg x, y)$: *Left Contradual*
6. $RC(f) = f(x, \neg y)$: *Right Contradual*
7. $LD(f) = \neg f(\neg x, y)$: *Left Dual*
8. $RD(f) = \neg f(x, \neg y)$: *Right Dual*

This enlarged set of transformations $\mathcal{T} = \{I, D, C, N, LC, RC, LD, RD\}$ forms 8-element group $\{T, \circ\}$ called $\mathcal{S}_{2\times2\times2}$ group. The equation, $N^2 = C^2 = D^2 = LC^2 = LD^2 = RC^2 = RD^2 = I$, provides sufficient information to identify this group. Its group operations are shown in [13] in detail.

When $\mathcal{T} = \{I, D, C, N, LC, RC, LD, RD\}$ are applied to Łukasiewicz implication($\rightarrow_{top}$) and to Kleene-Dienese implication($\rightarrow_{bot}$) of m_1 system of logic, respectively, they yield the closed set of connectives as below [13],[10].

The definitions of group transformation use an involutive negation $\neg a = 1-a$ which satisfies all of four axioms below. An involutive negation, $\neg$, can be generated by a pair of Łukasiewicz implication($\rightarrow_{Top}$) and a constant 0 or by that of Kleene-Dienese implication($\rightarrow_{Bot}$) and constant(0):

$$\neg a = \begin{cases} a \rightarrow_{Top} 0 = a \xrightarrow{Ł} 0 = min(1, 1-a+0) = 1-a \\ a \rightarrow_{Bot} 0 = a \xrightarrow{KD} 0 = max(1-a, 0) = 1-a \end{cases}$$

However, both of them fall into a single point; there is no genuine interval of negation.

Table 1: Group transformation of m_1 system: $<\rightarrow_{top}, \rightarrow_{bot}>$.

Transformation of Connective	Type of Interval Bound
$g_{1t} = I(\rightarrow_{Top}) = min(1, 1-a+b)$	$\rightarrow_{Top}$
$g_{2t} = C(\rightarrow_{Top}) = min(1, 1+a-b)$	$\leftarrow_{Top}$
$g_{3t} = D(\rightarrow_{Top}) = max(0, b-a)$	$\not\leftarrow_{Bot}$
$g_{4t} = N(\rightarrow_{Top}) = max(0, a-b)$	$\not\rightarrow_{Bot}$
$g_{5t} = LC(\rightarrow_{Top}) = min(1, a+b)$	$\vee_{Top}$
$g_{6t} = LD(\rightarrow_{Top}) = max(0, 1-a-b)$	$\downarrow_{Bot}$
$g_{7t} = RC(\rightarrow_{Top}) = min(1, 2-a-b)$	$\mid_{Top}$
$g_{8t} = RD(\rightarrow_{Top}) = max(0, a+b-1)$	$\wedge_{Bot}$
$g_{1b} = I(\rightarrow_{Bot}) = max(1-a, b)$	$\rightarrow_{Bot}$
$g_{2b} = C(\rightarrow_{Bot}) = max(a, 1-b)$	$\leftarrow_{Bot}$
$g_{3b} = D(\rightarrow_{Bot}) = min(1-a, b)$	$\not\leftarrow_{Top}$
$g_{4b} = N(\rightarrow_{Bot}) = min(a, 1-b)$	$\not\rightarrow_{Top}$
$g_{5b} = LC(\rightarrow_{Bot}) = max(a, b)$	$\vee_{Bot}$
$g_{6b} = LD(\rightarrow_{Bot}) = min(1-a, 1-b)$	$\downarrow_{Top}$
$g_{7b} = RC(\rightarrow_{Bot}) = max(1-a, 1-b)$	$\mid_{Bot}$
$g_{8b} = RD(\rightarrow_{Bot}) = min(a, b)$.	$\wedge_{Top}$

Axiom 3 *Axioms of Fuzzy Negation [1]:*

Let a negation $\neg$ a be defined by a function

$$\neg : [0,1] \rightarrow [0,1]$$

A function $\neg$ should satisfy at least two of axiomatic requirements below:

A1. $\neg 0 = 1$ *and* $\neg 1 = 0$. : *boundary condition.*
A2. *For all* $a, b \in [0,1]$, *if* $a \leq b$, *then* $\neg a \geq \neg b$. : *monotonicity.*
A3. $\neg$ *is a continuous function.* : *continuity.*
A4. $\neg$ *is involutive, i.e.* $\neg(\neg a) = a$, $\forall a \in [0,1]$. : *involution.*

In the next section, we define various negations which yield intervals and investigate their properties.

4 Interval Negations in Many-Valued Logics

In [3], we have shown that the negations can also be defined by the Sheffer or the Nicod connectives as well as by the implication:

Definition 4 *Negation [3]:*

1. $\neg_S a = a \mid a$: by Sheffer
2. $\neg_N a = a \downarrow a$: by Nicod
3. $\neg_{PLY} a = a \rightarrow 0$: by Implication

In classical logic, negations by above definitions are all equal: $\neg a = \neg_S a = \neg_N a = \neg_{PLY} a = 1-a$. In many_valued logic, however, they do not yield the equal results – $\neg_S a \neq \neg_N a \neq \neg_{PLY} a$ – but these negations in checklist paradigm m_1 system generate a genuine interval negation [3].

ISBN: 978-989-95079-6-8

4.1 Negation on m_1 Defined by the Sheffer Connective

Since the Sheffer connective, $a \mid b$, of the interval system m_1 appears a TOP-BOT pair by means of RD transformation of $<\rightarrow_{Top}, \rightarrow_{Bot}>$ as it is shown at the table in section 3,

$$a \mid b = \begin{cases} a \mid_{Top} b = min(1, 2-a-b) \\ a \mid_{Bot} b = max(1-a, 1-b), \end{cases}$$

the connective type $\neg_S a$ also also appears in two forms:

$$\neg_S a = \begin{cases} a \mid_{Top} a = min(1, 2(1-a)) \\ a \mid_{Bot} a = 1-a \end{cases}$$

Thus, $\neg_S a$ defined by *Sheffer* generates the interval

$$[\, \mid_{Bot},\ \mid_{Top}\,] = [\, 1-a,\ \ min(1, 2(1-a))\,].$$

4.2 Negation in m_1 Defined by the Nicod Connective

Since the Nicod connective, $a \downarrow b$, appears as a TOP-BOT pair of connectives in the interval system m_1 , similarly

$$a \downarrow b = \begin{cases} a \downarrow_{Top} b = min(1-a, 1-b) \\ a \downarrow_{Bot} b = max(0, 1-a-b), \end{cases}$$

the connective type $\neg_N a$ also also appears in two forms:

$$\neg_N a = \begin{cases} a \downarrow_{Top} a = 1-a & = D(a \mid_{Bot} a) \\ a \downarrow_{Bot} a = max(0, 1-2a) & = D(a \mid_{Top} a). \end{cases}$$

Thus, the interval of $\neg_N a$ by *Nicod* is:

$$[\downarrow_{Bot},\ \downarrow_{Top}] = [max(0, 1-2a),\ 1-a].$$

4.3 Negation in m_1 Defined by the Implication Connective

Since $a \overset{KD}{\rightarrow} b \leq m_1(\rightarrow) \leq a \overset{Ł}{\rightarrow} b$, in $a \rightarrow b$ of m_1 system, i.e. the BOT connective is the Kleene-Dienes implication operator while the TOP id Łukasiewicz, we have:

$$\neg_{PLY} a = \begin{cases} a \rightarrow_T 0 = a \overset{Ł}{\rightarrow} 0 = min(1, 1-a+0) = 1-a \\ a \rightarrow_B 0 = a \overset{KD}{\rightarrow} 0 = max(1-a, 0) = 1-a \end{cases}$$

4.4 Interval of Negation in m_1 logic system

Two intervals of negation defined by the *Sheffer* connective and by the *Nicod* connective may be combined to yield a genuine interval of negation since $1-a$ is the lower bound of *Sheffer* negation and the upper bound of *Nicod* negation, respectively. These two intervals are concatenated as below:

$$\begin{aligned} [\neg_{BOT},\ \neg_{TOP}] &= [\downarrow_{Bot},\ \downarrow_{Top}] \cup [\, \mid_{Bot},\ \mid_{Top}\,] \\ &= [\ \max(0,\ 1\text{-}2a),\ \ \min(1,\ 2(1\text{-}a))\ ] \\ &= \begin{cases} [\, 0,\ 2(1-a)\,], & 1 \geq a \geq .5 \\ [\, 1-2a,\ 1\,] & .5 > a \geq 0 \end{cases} \\ \neg_{MID} &= \neg_{PLY} a = a \downarrow_{Top} a = a \mid_{Bot} a = \ 1-a. \end{aligned}$$

Thus, *Sheffer*($\mid$) and *Nicod*($\downarrow$) form the TOP system of negation and the BOT system of negation, respectively. The $\neg_{PLY}$ $a = 1-a$, is a median value of the interval negation, $\neg_{Mid}$. Since both $\neg_{Top}$ and $\neg_{Bot}$ satisfy A1 - A3 of *Axioms of fuzzy negation* except A4 of sec.3, they are non-involutive TOP-BOT pair of fuzzy interval negations while $\neg_{Mid}$ is an involutive fuzzy negation.

Theorem 5 *Gap Theorem 1. (Bandler and Kohout [6])*

$$\begin{aligned} a \text{ AND}_{Top}\ b - a \text{ AND}_{Bot}\ b &= a \text{ OR}_{Top}\ b - a \text{ OR}_{Bot}\ b \\ &= a \text{ PLY}_{Top}\ b - a \text{ PLY}_{Bot}\ b \\ &= min(\varphi a,\ \varphi b). \\ a \text{ IFF}_{Top}\ b - a \text{ IFF}_{Bot}\ b &= a \text{ EOR}_{Top}\ b - a \text{ EOR}_{Bot}\ b \\ &= 2min(\varphi a,\ \varphi b). \end{aligned}$$

Similarly, we can describe Gap Theorem for a TOP-BOT pair of negation.

Theorem 6 *Gap Theorem 2.*

$$\neg_{TOP}\ a - \neg_{BOT}\ a = min(2a, 2(1-a)) = 2\varphi a$$

Hence, the margins of imprecision can be directly measured by the degree of fuzziness φ where $\varphi a = min(a, 1-a)$.

To further investigate a non-involutive properties of $[\neg_{Bot}, \neg_{Top}]$ pair, we apply multiple nagation. It leads to the iterative negations, interrupting mathematical properties and mathematical limit, but yields an interesting sequence of values of interval. In the following sections, we use the symbols $\neg_T$ and $\neg_B$ for $\neg_{Top}$ and $\neg_{Bot}$, respectively, for simplicity.

5 Multiple negations

As both $\neg_{Bot}(= \neg_B)$ and $\neg_{Top}(= \neg_T)$ are non-involutive, a question arise what would happen if a negation is iterated to $\neg_B$ and $\neg_T$, respectively: namely, a multiple negation.

Definition 7 *Multiple Negations.*

1. $\neg^1 a = \neg\, a.$
2. $\neg^n a = (\neg\, a)^n = \neg\, (\neg^{n-1}\, a)$
$= \neg(\neg(\neg^{n-2}\, a))$
$= \underbrace{\neg(\neg(\neg \ldots, (\neg}_{n}\, a))), \quad 1 \leq n$

Thus, multiple negations will be applied to $\neg_T$ and $\neg_B$, respectively in the following subsections in order to explore their properties from the computational results. The following two subsections summarize the computational results of multiple negations of $\neg_{Top}a$ and $\neg_{Bot}a$, respectively .

5.1 Multiple negation of $\neg_{TOP}$

$$\begin{aligned} \neg_T a = a \mid_T a &= min(1, 2(1-a)) \\ &= \begin{cases} 2(1-a), & a \in [\frac{1}{2},\ 1] \\ 1 & a \in [0, \frac{1}{2}) \end{cases} \end{aligned} \tag{1}$$

$$\begin{aligned} \neg_T^2 a = \neg_T(\neg_T a) &= min(1,\ 2(1-\neg_T a)) \\ &= \begin{cases} 1, & a \in (\frac{3}{4}, 1] \\ 2(2a-1), & a \in [\frac{1}{2},\ \frac{3}{4}] \\ 0 & a \in [0,\ \frac{1}{2}) \end{cases} \end{aligned} \tag{2}$$

$$\begin{aligned} \neg_T^3 a = \neg_T(\neg_T^2 a) &= min(1,\ 2(1-\neg_T^2 a))) \\ &= \begin{cases} 0, & a \in (\frac{3}{4},\ 1] \\ 2(3-4a), & a \in [\frac{5}{8},\ \frac{3}{4}] \\ 1 & a \in [0,\ \frac{5}{8}) \end{cases} \end{aligned} \tag{3}$$

$$\begin{aligned} \neg_T^4 a = \neg_T(\neg_T^3 a) &= min(1,\ 2(1-\neg_T^3 a))) \\ &= \begin{cases} 1, & a \in (\frac{11}{16},\ 1] \\ 2(8a-5), & a \in [\frac{5}{8},\ \frac{11}{16}] \\ 0 & a \in [0,\ \frac{5}{8}) \end{cases} \end{aligned} \tag{4}$$

ISBN: 978-989-95079-6-8

$$\neg_T^5 a = \neg_T(\neg_T^4 a) \quad = min(1,\ 2(1-\neg_T^4 a))) = \begin{cases} 0, & a \in (\frac{11}{16},\ 1] \\ 2(11-16a), & a \in [\frac{21}{32},\ \frac{11}{16}] \\ 1 & a \in [0,\ \frac{21}{32}] \end{cases} \tag{5}$$

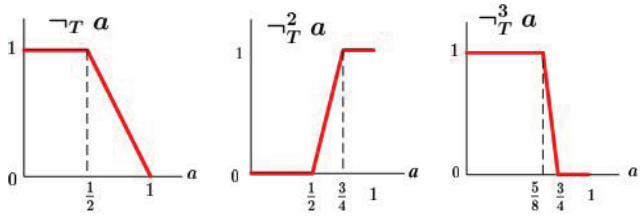


Figure 1: $\neg_T\ a,\ \neg_T^2\ a,\ \neg_T^3\ a$

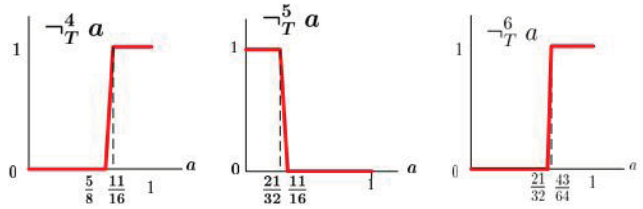


Figure 2: $\neg_T^4\ a,\ \neg_T^5\ a,\ \neg_T^6\ a$

If we further generalize it to $\neg_T^n\ a$ both for $n = 2k$ and for $n = 2k+1, \forall k \geq 1$, respectively, we get the following formulas (6)-(7).

$$\neg_T^{2k}\ a = \neg_T(\neg_T^{2k-1} a) \quad = \begin{cases} 1, & a \in (u_1,\ 1] \\ v_0, & a \in [u_2,\ u_1] \\ 0 & a \in [0,\ u_2) \end{cases} \tag{6}$$

where $v_0 = 2^{2k}a - \frac{2}{3}(2^{2k}-1)$,

$u_1 = \frac{2\cdot 4^k+1}{3\cdot 4^k}$,

$u_2 = \frac{2(4^k-1)}{3\cdot 4^k}$,

and

$$\neg_T^{2k+1}\ a = \neg_T(\neg_T^{2k} a) \quad = \begin{cases} 0, & a \in (u_3,\ 1] \\ v_1 & a \in [u_4,\ u_3] \\ 1 & a \in [0,\ u_4) \end{cases} \tag{7}$$

where $v_1 = \frac{2}{3}(2^{2k+1}+1) - 2^{2k+1}a$,

$u_3 = u_1 = \frac{2\cdot 4^k+1}{3\cdot 4^k}$,

$u_4 = \frac{4^{k+1}-1}{6\cdot 4^k}$.

Thus, any multiple negations in the odd sequence of $\neg_T^{2k+1}$, $\forall k \geq 0$, i.e. $< \neg_T a, \neg_T^3 a, \ldots, \neg_T^{2k+1} a, \ldots >$, are monotonically decreasing(**Axiom A2**) while those in the even sequence of $\neg_T^{2k}$, $\forall k \geq 1$, $< \neg_T^2 a, \neg_T^4\ a, \ldots, \neg_T^{2k}\ a, \ldots >$, are monotonically increasing(**A2**). All of multiple negations satisfy the boundary condition(**A1**) and continuity(**A3**) .

5.2 *Multiple negation of* $\neg_{BOT}$

$$\neg_{bot} a = a \downarrow_B a \quad = max(0, 1-2a) = \begin{cases} 0, & a \in (\frac{1}{2},\ 1] \\ 1-2a & a \in [0,\ \frac{1}{2}] \end{cases} \tag{8}$$

$$\neg_B^2 a = \neg_B(\neg_B a) \quad = max(0, 1-2\cdot\neg_B a) = \begin{cases} 1, & a \in (\frac{1}{2},\ 1] \\ 4a-1, & \in [\frac{1}{4},\ \frac{1}{2}] \\ 0 & a \in [0,\ \frac{1}{4}) \end{cases} \tag{9}$$

$$\neg_B^3 a = \neg_B(\neg_B^2\ a) \quad = max(0, 1-2\cdot\neg_B^2\ a) = \begin{cases} 0, & a \in (\frac{3}{8},\ 1] \\ 3-8a, & a \in [\frac{1}{4},\ \frac{3}{8}] \\ 1 & a \in [0,\ \frac{1}{4}) \end{cases} \tag{10}$$

$$\neg_B^4 a = \neg_B(\neg_B^3\ a) \quad = max(0,\ 1-2\cdot\neg_B^3 a) = \begin{cases} 1, & a \in (\frac{3}{8},\ 1] \\ 16a-5, & a \in [\frac{5}{16},\ \frac{3}{8}] \\ 0 & a \in [0,\ \frac{5}{16}) \end{cases} \tag{11}$$

$$\neg_B^5 a = \neg_B(\neg_B^4\ a) \quad = max(0,\ 1-2\cdot\neg_B^4 a) = \begin{cases} 0, & a \in (\frac{11}{32}\ 1] \\ 11-32a, & a \in [\frac{5}{16},\ \frac{11}{32}] \\ 1 & a \in [0,\ \frac{5}{16}) \end{cases} \tag{12}$$

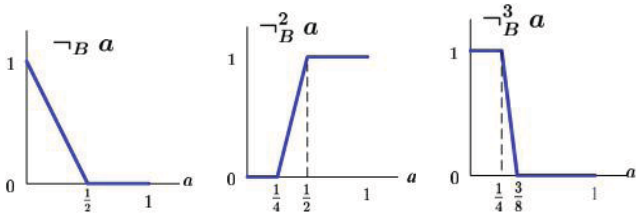


Figure 3: $\neg_B\ a,\ \neg_B^2\ a,\ \neg_B^3\ a$

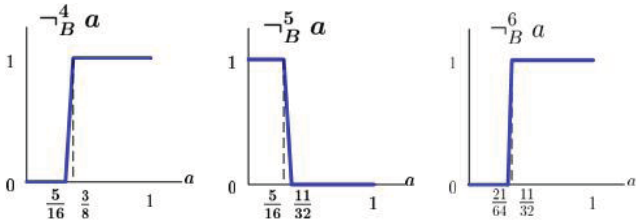


Figure 4: $\neg_B^4\ a,\ \neg_B^5\ a,\ \neg_B^6\ a$

If we further generalize it to $\neg_T^n\ a$ both for $n = 2k$ and for $n = 2k+1, \forall k \geq 1$, respectively, we get the following formulas (13)-(14).

$$\neg_B^{2k}\ a = \neg_B(\neg_B^{2k-1} a) \quad = \begin{cases} 1, & a \in (w_1,\ 1] \\ y_0, & a \in [w_2,\ w_1] \\ 0 & a \in [0,\ w_2) \end{cases} \tag{13}$$

where $y_0 = 2^{2k}a - \frac{1}{3}(2^{2k}-1)$,

$w_1 = \frac{4^k+2}{3\cdot 4^k}$,

$w_2 = \frac{4^k-1}{3\cdot 4^k}$,

ISBN: 978-989-95079-6-8

and

$$\neg_B^{2k+1} a = \neg_B(\neg_B^{2k} a) = \begin{cases} 0, & a \in (w_3,\ 1] \\ y_1 & a \in [w_4,\ w_3] \\ 1 & a \in [0,\ w_4) \end{cases} \tag{14}$$

where $y_1 = \frac{1}{3}(2^{2k+1}+1) - 2^{2k+1}a$,

$w_3 = \frac{2\cdot 4^k+1}{6\cdot 4^k}$,

$w_4 = w_2 = \frac{4^k-1}{3\cdot 4^k}$.

Similar to TOP system of negation in sec. 5.1., any multiple negations in the odd sequence are monotonically decreasing while those in the even sequence are monotonically increasing. All of them satisfy the boundary condition and continuity. Both TOP-BOT systems of negation also show the following property in the interval of a.

Property 8 *The width of interval of a:*
The width of interval of a in eq.(6)-(7) and eq.(13)-(14) is:

1. *For* $n = 2k$: $\neg_T^{2k} a$ *and* $\neg_B^{2k} a$,
$$| u_1 - u_2 | = | w_1 - w_2 | = \tfrac{1}{2^{2k}} = \tfrac{1}{2^n},$$

2. *For* $n = 2k+1$: $\neg_T^{2k+1} a$ *and* $\neg_B^{2k+1} a$,
$$| u_3 - u_4 | = | w_3 - w_4 | = \tfrac{1}{2^{2k+1}} = \tfrac{1}{2^n}.$$

Theorem 9 *A relationship between* $\neg^n$ *and* $\neg^{n+1}$*:*

(1) $1 - \neg_T^n a = \frac{1}{2}\neg_T^{n+1} a,\ \forall n \in N,\ a \in [max(u_2,u_4),\ u_1]$ *in eq.(6)-(7).*

(2) $1 - \neg_B^{n+1} a = 2 \cdot \neg_B^n a,\ \forall n \in N,\ a \in [w_2,\ min(w_1,w_3)]$ *in eq.(13)-(14).*

Proof:

(1) (*a*) Let $n = 2k,\ \forall k \geq 1$.
$$\begin{aligned} 1 - \neg_T^{2k} a &= 1 - (2^{2k}a - \tfrac{2}{3}(2^{2k}-1)) \\ &= \tfrac{1}{3}(2^{2k+1}+1) - 2^{2k}a \\ &= \tfrac{1}{2}(\tfrac{2}{3}(2^{2k+1}+1) - 2^{2k+1}a) \\ &= \tfrac{1}{2}\cdot\neg_T^{2k+1} a \end{aligned}$$
(*b*) Let $n = 2k+1, \forall k \geq 0$.
$$\begin{aligned} 1 - \neg_T^{2k+1} a &= 1 - (\tfrac{2}{3}(2^{2k+1}+1) - 2^{2k+1}a) \\ &= 2^{2k+1}a - \tfrac{1}{3}(2^{2k+2}-1) \\ &= \tfrac{1}{2}(2^{2k+2}a - \tfrac{2}{3}(2^{2k+2}-1)) \\ &= \tfrac{1}{2}\neg_T^{2k+2} a \end{aligned}$$
Therefore, $1 - \neg_T^n a = \frac{1}{2}\neg_T^{n+1} a,\ ,\ \forall n \in N$. □

(2) (*a*) $n = 2k-1, \forall k \geq 1$.
$$\begin{aligned} 1 - \neg_B^{2k} a &= 1 - (2^{2k}a - \tfrac{1}{3}(2^{2k}-1)) \\ &= \tfrac{2}{3}(2^{2k-1}+1) - 2^{2k}a \\ &= 2(\tfrac{1}{3}(2^{2k-1}+1) - 2^{2k-1}a) \\ &= 2\cdot\neg_B^{2k-1} a \end{aligned}$$
(*b*) $n = 2k, \forall k \geq 1$.
$$\begin{aligned} 1 - \neg_B^{2k+1} a &= 1 - (\tfrac{1}{3}(2^{2k+1}+1) - 2^{2k+1}a) \\ &= 2^{2k+1}a - \tfrac{2}{3}(2^{2k}-1) \\ &= 2(2^{2k}a - \tfrac{1}{3}(2^{2k}-1)) \\ &= 2\cdot\neg_B^{2k} a \end{aligned}$$
Therefore, $1 - \neg_B^{n+1} a = 2 \cdot \neg_B^n a,\ \forall n \in N$. □

5.3 *Convergence of multiple negations*

In order to investigate a convergence of multiple negations, let us apply lim to $\neg_T^{2k} a$, $\neg_T^{2k+1} a$, $\neg_B^{2k} a$ and to $\neg_B^{2k+1} a$ as $k \to \infty$, respectively. It derives the following formula of convergence in each case:

$$\lim_{k\to\infty} \neg_T^{2k} a = \begin{cases} 1, & a \in (\frac{2}{3}^+,\ 1] \\ z_0, & a \in [\frac{2}{3}^-,\ \frac{2}{3}^+] \\ 0 & a \in [0,\ \frac{2}{3}^-) \end{cases} \tag{15}$$

where $z_0 = \lim_{k\to\infty} v_0 = \lim_{k\to\infty} 2^{2k}a - \frac{2}{3}(2^{2k}-1)$, in eq.(6)
$z_0 = \frac{2}{3}$ if $a = \frac{2}{3}$, in particular,
and $\frac{2}{3}^+ = \lim_{k\to\infty} u_1 = \frac{2}{3} + \varepsilon_1$,
$\frac{2}{3}^- = \lim_{k\to\infty} u_2 = \frac{2}{3} - \varepsilon_1$, for a very small $\varepsilon_1 > 0$.

$$\lim_{k\to\infty} \neg_T^{2k+1} a = \begin{cases} 0, & a \in (\frac{2}{3}^+,\ 1] \\ z_1 & a \in [\frac{2}{3}^-,\ \frac{2}{3}^+] \\ 1 & a \in [0,\ \frac{2}{3}^-) \end{cases} \tag{16}$$

where $z_1 = \lim_{k\to\infty} v_1 = \lim_{k\to\infty} \frac{2}{3}(2^{2k+1}+1) - 2^{2k+1}a$, in eq.(7)
$z_1 = \frac{2}{3}$ if $a = \frac{2}{3}$, in particular,
and $\frac{2}{3}^+ = \lim_{k\to\infty} u_3 = \frac{2}{3} + \varepsilon_2$,
$\frac{2}{3}^- = \lim_{k\to\infty} u_4 = \frac{2}{3} - \varepsilon_2$, for a very small $\varepsilon_2 > 0$
where $\varepsilon_2 < \varepsilon_1$.

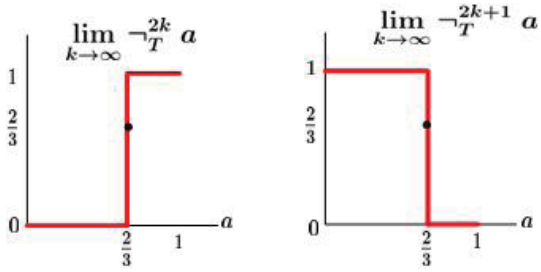


Figure 5: $\lim_{k\to\infty} \neg_T^{2k} a$ and $\lim_{k\to\infty} \neg_T^{2k+1} a$

As $k \to \infty$. both lim $\neg_T^{2k} a$ and lim $\neg_T^{2k+1} a$ converge either to 0 or to 1, depending on the values of a in the most of intervals, except $a \in [\frac{2}{3} - \varepsilon_i,\ \frac{2}{3} + \varepsilon_i]$ for $i = 1, 2$ where ε_i is in eq.(15)-(16). For an a in the interval $[\frac{2}{3} - \varepsilon_i,\ \frac{2}{3} + \varepsilon_i]$, the values of $\lim_{k\to\infty} \neg_T^{2k} a$ or those of $\lim_{k\to\infty} \neg_T^{2k+1} a$ increases or decreases drastically in monotonicity, respectively.

We can investigate the limit negations of $\neg_B^{2k} a$ and $\neg_B^{2k+1} a$ of BOT system, similarly.

$$\lim_{k\to\infty} \neg_B^{2k} a = \begin{cases} 1, & a \in (\frac{1}{3}^+,\ 1] \\ z_2, & a \in [\frac{1}{3}^-,\ \frac{1}{3}^+] \\ 0 & a \in [0,\ \frac{1}{3}^-) \end{cases} \tag{17}$$

where $z_2 = \lim_{k\to\infty} y_0 = \lim_{k\to\infty} 2^{2k}a - \frac{1}{3}(2^{2k}-1)$, in eq.(13)
$z_2 = \frac{1}{3}$ if $a = \frac{1}{3}$, in particular,
and $\frac{1}{3}^+ = \lim_{k\to\infty} w_1 = \frac{1}{3} + \delta_1$,
$\frac{1}{3}^- = \lim_{k\to\infty} w_2 = \frac{1}{3} - \delta_1$, for a very small $\delta_1 > 0$.

ISBN: 978-989-95079-6-8

$$\lim_{k\to\infty} \neg_B^{2k+1} a = \begin{cases} 0, & a \in (\frac{1}{3}^+,\ 1] \\ z_3, & a \in [\frac{1}{3}^-,\ \frac{1}{3}^+] \\ 1 & a \in [0,\ \frac{1}{3}^-) \end{cases} \quad (18)$$

where $z_3 = \lim_{k\to\infty} y_1 = \lim_{k\to\infty} \frac{1}{3}(2^{2k+1}+1) - 2^{2k+1}a$, in eq.(14)
$z_3 = \frac{1}{3}$ if $a = \frac{1}{3}$, in particular,
and $\frac{1}{3}^+ = \lim_{k\to\infty} w_3 = \frac{1}{3} + \delta_2$,
$\frac{1}{3}^- = \lim_{k\to\infty} w_4 = \frac{1}{3} - \delta_2$, for a very small $\delta_2 > 0$
where $\delta_2 < \delta_1$.

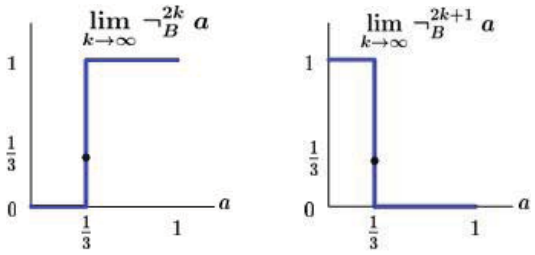


Figure 6: $\lim_{k\to\infty} \neg_B^{2k} a$ and $\lim_{k\to\infty} \neg_B^{2k+1} a$

As $k \to \infty$. both $\lim \neg_B^{2k} a$ and $\lim \neg_B^{2k+1} a$ converge either to 0 or to 1 for the most of $a \in [0,\ 1]$, except $a \in [\frac{1}{3} - \delta_i,\ \frac{1}{3} + \delta_i]$, for $i = 1, 2$ where δ_i is in eq.(17)-(18). For an a in $[\frac{1}{3} - \delta_i,\ \frac{1}{3} + \delta_i]$, the values of $\lim_{k\to\infty} \neg_B^{2k} a$ or those of $\lim_{k\to\infty} \neg_B^{2k+1} a$ are in the drastic monotonic increase or decrease, respectively. Thus, we can further describe the following properties.

Property 10 *For $n \in N$,*

P1. $\lim_{n\to\infty} \neg_T^{n+1} a \approx 1 - \lim_{n\to\infty} \neg_T^n a, \quad a \in [0,\ 1]$.

P2. $\lim_{n\to\infty} \neg_B^{n+1} a \approx 1 - \lim_{n\to\infty} \neg_B^n a, \quad a \in [0,\ 1]$.

P3. $\lim_{k\to\infty} \neg_T^n a \approx \lim_{n\to\infty} \neg_B^n (a + \frac{1}{3}), \quad a \in [0,\ 1]$.

P4. $\lim_{n\to\infty} \neg_B^n a \approx \lim_{k\to\infty} \neg_T^n (a - \frac{1}{3}), \quad a \in [0,\ 1]$.

From the above **P1** and **P2**, we can notice that both $\neg_T$ and $\neg_B$ are *nearly limit involutive* because $\lim_{n\to\infty} \neg_T^{n+1} a \approx 1 - \lim_{n\to\infty} \neg_T^n a \approx 1 - (1 - \lim_{n\to\infty} \neg_T^{n-1} a) \approx \lim_{n\to\infty} \neg_T^{n-1} a, \quad a \in [0,\ 1]$, i.e. $\neg_T(\neg_T(\lim_{n\to\infty} \neg_T^{n-1}) \approx \lim_{n\to\infty} \neg_T^{n-1}$. It holds for $\neg_B$, similarly.

6 Conclusions

The involutive fuzzy negation, $\neg a = 1 - a$, has been collapsed into a single point while ten of binary connectives yield TOP-BOT interval pairs in m_1 logic system of 16 connectives by group of logic transformations since Bandler-Kohout proposed five checklist paradigm based logic systems $m_1 - m_5$. A BOT-TOP pair of connectives of fuzzy negation interval $[\neg_{BOT},\ \neg_{TOP}]$, however, could be successfully generated from the interval of *Nicod* system $[\downarrow_{top},\ \downarrow_{bot}]$ and that of *Sheffer* system $[|_{bot},\ |_{top}]$. These pairs of negation connective are non-involutive negation pairs by themselves. When a negation is applied iteratively, those multiple negations, $\neg_T^n a, \neg_B^n a$ change their values from 0 to 1 or 1 to 0 drastically within a very small interval whose width is $\frac{1}{2^n}$ and reveals a property of nearly limit involution.

References

[1] G.J. Klir and B. Yuan Fuzzy sets and fuzzy logic: Theory and applications. Prentice Hall, 1995.

[2] W. Bandler and L.J. Kohout. Semantics of implication operators and fuzzy relational products. *Internat. Journal of Man-Machine Studies*, 12:89–116, 1980. Reprinted in Mamdani, E.H. and Gaines, B.R. eds. *Fuzzy Reasoning and its Applications*. Academic Press, London, 1981, pp. 219-246.

[3] E. Kim and L.J. Kohout. Interval Negation in Fuzzy Logics. In *Proc. of NAFIPS (North American Fuzzy Information Processing Society)*, pp. 537-542, San Diego, CA, June 24-27 2007. IEEE.

[4] W. Bandler and L.J. Kohout. Unified theory of multiple-valued logical operators in the light of the checklist paradigm. In *Proc. of the 1984 IEEE Conference on Systems, Man and Cybernetics*, pp. 356–364, New York, 1984. IEEE.

[5] W. Bandler and L.J. Kohout. The interrelations of the principal fuzzy logical operators. In M.M. Gupta, A. Kandel, W. Bandler, and J.B. Kiszka, eds., *Approximate Reasoning in Expert Systems*, pp. 767–780. North-Holland, Amsterdam, 1985.

[6] W. Bandler and L.J. Kohout. The use of checklist paradigm in inference systems. In C.V. Negoita and H. Prade, eds., *Fuzzy Logic in Knowledge Engineering*, chapter 7, pp. 95–111. Verlag TÜV Rheinland, Köln, 1986.

[7] P. Hájek and L.J. Kohout. Fuzzy implications and generalized quantifiers. *Internat. Journal of Uncertainty, Fuzziness and Knowledge Based Systems*, 4(3):225-233, 1996.

[8] L.J. Kohout and E. Kim. On the emergence of fuzzy logics from elementary percepts: the checklist paradigm and the theory of perceptions. In *Forging New Frontiers: Fuzzy Pioneers I. Studies in Fuzziness and Soft Computing*, 217:437-460, Springer, 2007.

[9] L.J. Kohout and E. Kim. Group transformations of systems of logic connectives. In *Proc. of IEEE-FUZZ'97*, 1:157–162, IEEE, New York, July 1997.

[10] L.J. Kohout and E. Kim. Characterization of interval fuzzy logic systems of connectives by group transformations. *Reliable Computing*, 10:299–334, 2004.

[11] L.J. Kohout and W. Bandler. Checklist paradigm and group transformations. *Technical Note EES-MMS-ckl91.2*, 1979. Dept. of Electrical Engineering, University of Essex. U.K.

[12] L.J. Kohout and W. Bandler. How the checklist paradigm elucidates the semantics of fuzzy inference. In *Proc. of the IEEE Internat. Conference on Fuzzy Systems 1992*, pp. 571–578. IEEE, New York, 1992.

[13] L.J. Kohout and E. Kim. Global characterization of fuzzy logic systems with para-consistent and gray set features. In P. Wang, ed., *Proc. 3rd Joint Conf. on Information Sciences JCIS'97 (5th Int. Conf. on Fuzzy Theory and Technology)*, pp. 238–241, Research Triangle Park, NC, March 1997. Duke University. In volume 1: Fuzzy Logic, Intelligent Control and Genetic Algorithms.

ISBN: 978-989-95079-6-8

Development of Multiple Linguistic Equation Models with Takagi-Sugeno Type Fuzzy Models

Esko K. Juuso
Control Engineering Laboratory
Department of Process and Environmental Engineering
University of Oulu, Finland
Email: esko.juuso@oulu.fi

Abstract— *Multimodel approaches are widely used with linear submodels, but border areas around submodels are problematic. Special cases of fuzzy linguistic equation models, which can be understood as linguistic Takagi-Sugeno (LTS) type fuzzy models, can be used to solve these problems in many cases. These models use a special nonlinear scaling approach for both inputs and outputs. The LTS models are robust solutions for applications where the same variables can be used for defining operating areas and also in the submodels. No special smoothing algorithms are needed.*

Keywords— linguistic equations, nonlinear systems, multimodels, Takagi-Sugeno fuzzy models.

1 Introduction

Fuzzy set systems enable the use of expert system techniques in uncertain and vague systems [1]. The traditions of physical modelling on the basis of understanding system behaviour are maintained with fuzzy rules and membership functions, which can represent not only gradually changing nonlinear mappings but also abrupt changes [2]. Various approaches using either expertise or data are used in constructing these mappings, but as the complexity of the application increases more and more combined approaches are required. These approaches are presented in the upper right-hand corner of Fig. 1. Heuristic knowledge and know-how can be introduced to fuzzy set systems with a trial and error based approach. Data-based approaches rely usually on automatic generation of rules from predefined simple sets of membership functions. Both knowledge and data need to be used together to develop practical applications.

Linguistic fuzzy models [3] are mainly used in the knowledge-based approach, whereas Takagi-Sugeno (TS) type fuzzy models [4] and fuzzy relational models [5] are mainly used for data-driven methods. TS models are constructed by combining supervised and unsupervised learning. The antecedent membership functions can be initialised by grid partioning, iterative search, or fuzzy clustering [6]. The neuro-fuzzy ANFIS method [7] is widely used in tuning. The structure and the parameters can also be updated recursively when new data become available [8]. The trade-off between global model accuracy and interpretability of local models as the linearisations of a nonlinear system is important in the development of a TS model. To restrict the freedom of the parameters a multiobjective identification for dynamic TS models is presented in [9].

This paper presents a set of fuzzy linguistic equation models, which combine nonlinear scaling and Takagi-Sugeno type fuzzy models.

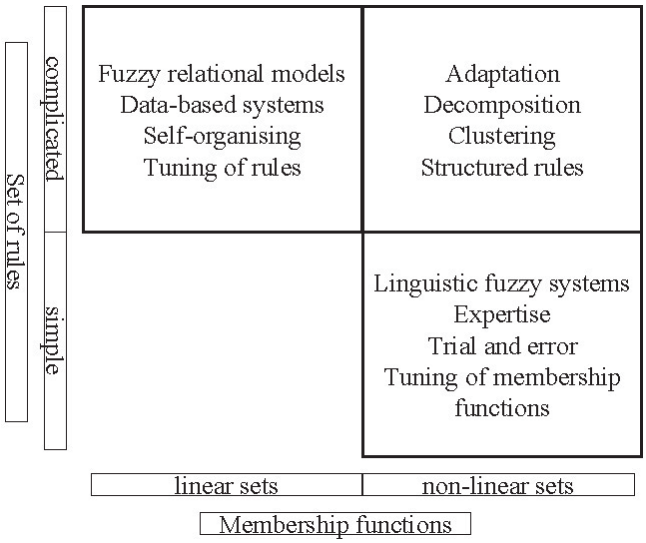


Figure 1: Classification of fuzzy set systems [2].

2 Methodologies

2.1 Linguistic equations

The linguistic equation (LE) approach originates from fuzzy set systems: rule sets are replaced with equations, and the effects of membership functions are handled with scaling [10]. For nonlinear models, the scaling technique must be nonlinear as the model equations are linear. The scaling functions are called membership definitions as they are closely connected to the membership functions used in fuzzy set systems. [2]

Nonlinear scaling. Nonlinear scaling is carried out using functions which denote membership definitions. The mapping function is performed by converting variable values from the variable range into a range [-2, 2], known as the *linguistic range*. The new range describes the distribution of variable values over the original range fairly accurately. Membership definitions are presented by a variable specific function

$$x_j = f(X_j) \,\forall\, \min(x_j) \le x_j \le \max(x_j),\ X_j \in [-2, 2], \tag{1}$$

where x_j is the value of variable j, and X_j is the corresponding value within the range $[-2,\ 2]$, which includes the normal operation range $[-1,\ 1]$ and areas for warnings and alarms. The values X_j are called *linguistic values* because the scaling function is based on the membership functions of fuzzy set systems: values -2, -1, 0, 1 and 2 can be associated to the

ISBN: 978-989-95079-6-8

linguistic labels, e.g.

$$\{very\,low,\ low,\ normal,\ high,\ very\,high\} \qquad (2)$$

are defined with membership functions (Fig. 2). The number of membership functions is not limited to five: the values between these integers correspond to finer partitions of the fuzzy set system. Early applications of linguistic equations only used integer values [10].

In the case of polynomial membership definitions,

$$\begin{array}{ll} f_j^- = a_j^- X_j^2 + b_j^- X_j + c_j, & X_j \in [-2, 0), \\ f_j^+ = a_j^+ X_j^2 + b_j^+ X_j + c_j, & X_j \in [0, 2], \end{array} \quad . \qquad (3)$$

the linguistic level of the input variable j is calculated according to the equation

$$X_j = \begin{cases} 2 & \text{with } x_j \geq \max x_j \\ \frac{-b_j^+ + \sqrt{b_j^{+2} - 4a_j^+ (c_j - x_j)}}{2a_j^+} & \text{with } c_j \leq x_j \leq \max(x_j) \\ \frac{-b_j^- + \sqrt{b_j^{-2} - 4a_j^- (c_j - x_j)}}{2a_j^-} & \text{with } \min(x_j) \leq x_j \leq c_j \\ -2 & \text{with } x_j \leq \min(x_j). \end{cases} \qquad (4)$$

where a_j^-, b_j^-, a_j^+, and b_j^+ are coefficients of the polynomials (3), c_j is the real value corresponding to the linguistic value 0, and x_j is the real value. $\min(x_j)$ and $\max(x_j)$ are the minimum and maximum values of the real data corresponding to the linguistic values -2 and 2.

After the linguistic level of the model output, X_{out}, is calculated using the linguistic equation model, it is converted into a real value of output, x_{out}, using the following equation:

$$x_{out} = \begin{cases} a_{out}^- {X_{out}}^2 + b_{out}^- X_{out} + c_{out} \\ \qquad \text{with } X_{out} < 0 \\ a_{out}^+ {X_{out}}^2 + b_{out}^+ X_{out} + c_{out} \\ \qquad \text{with } X_{out} \geq 0 \end{cases} \qquad (5)$$

where a_{out}^-, b_{out}^-, a_{out}^+ and b_{out}^+ are coefficients of the polynomials (3), and c_{out} is the real value corresponding to the linguistic value 0.

The coefficients of the polynomials can be represented by

$$\begin{array}{rcl} a_j^- & = & \frac{1}{2}(1 - \alpha_j^-)\,\Delta c_j^-, \\ b_j^- & = & \frac{1}{2}(3 - \alpha_j^-)\,\Delta c_j^-, \\ a_j^+ & = & \frac{1}{2}(\alpha_j^+ - 1)\,\Delta c_j^+, \\ b_j^+ & = & \frac{1}{2}(3 - \alpha_j^+)\,\Delta c_j^+, \end{array} \qquad (6)$$

where $\Delta c_j^- = c_j - (c_l)_j$ and $\Delta c_j^+ = (c_h)_j - c_j$. Membership definitions may contain linear parts if some coefficients α_j^- or α_j^+ equal one.

Membership definitions are determined by the centre point c_j, and the core and support areas, which guarantee that the resulting membership definitions are monotonously increasing functions. An easier way is to define the centre point, the core $[(c_l)_j,\ (c_h)_j]$ and the ratios

$$\begin{array}{l} \alpha_j^- = \frac{(c_l)_j - \min(x_j)}{c_j - (c_l)_j}, \\ \alpha_j^+ = \frac{\max(x_j) - (c_h)_j}{(c_h)_j - c_j}, \end{array} \qquad (7)$$

from the range $\frac{1}{3} \ldots 3$, and calculate the support $[\min(x_j),\ \max(x_j)]$. The membership definitions of each variable are configured with five parameters, which can presented in three consistent sets. The working point (centre point) c_j belongs to all these sets, and the others are:

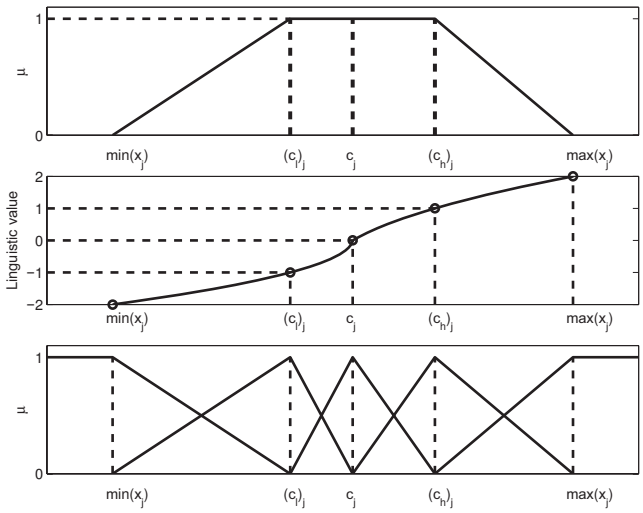


Figure 2: Feasible range, membership definitions and membership functions [2].

- the corner points $\{\min(x_j),\ (c_l)_j,\ (c_h)_j,\ \max(x_j)\}$ are good for visualisation;
- the parameters $\{\alpha_j^-,\ \Delta c_j^-,\ \alpha_j^+,\ \Delta c_j^+\}$ are suitable for tuning;
- the coefficients $\{a_j^-,\ b_j^-,\ a_j^+,\ b_j^+\}$ are used in calculations.

The upper and the lower parts of the scaling functions can be convex or concave, independent of each other. Simplified functions can also be used: a linear membership definition only requires two parameters: c_j and $b_j = b_j^+ = b_j^-$ or $\Delta c_j = \Delta c_j^+ = \Delta c_j^-$, since $\alpha_j^+ = \alpha_j^- = 1$ and $a_j^+ = a_j^- = 0$; an asymmetrical linear definition has $\Delta c_j^+ \neq \Delta c_j^-$ and $b_j^+ \neq b_j^-$.

Interactions. The basic element of a linguistic equation (LE) model is a compact equation

$$\sum_{j=1}^{m} A_{ij} X_j + B_i = 0, \qquad (8)$$

where X_j is a linguistic level of the variable j, $j = 1 \ldots m$. The direction of the interaction is represented by the interaction coefficients A_{ij}. The bias term B_i was introduced for fault diagnosis systems. A LE model with several equations is represented as a matrix equation

$$AX + B = 0, \qquad (9)$$

where the interaction matrix A contains all the coefficients A_{ij} and the bias vector B all the bias terms B_i.

The model is represented by

$$x_{out} = f_{out}\left(-\frac{1}{A_{i\,out}} \Big(\sum_{j=1, j \neq out}^{m} A_{ij}\, f_j^{-1}(x_j) + B_i \Big) \right), \qquad (10)$$

ISBN: 978-989-95079-6-8

where the functions f_j and f_{out} are membership definitions. In the general case, the weight factors

$$w_{ij} = -\frac{A_{ij}}{A_{i\,out}}, \tag{11}$$

and the bias term

$$B_i = -\frac{B_i}{A_{i\,out}}, \tag{12}$$

Altogether, there are $n_i + 1$ parameters. As the scaling functions of each variable require three additional parameters for every two parameters needed for normalisation, the total number of additional parameters is $4 + 4n_i$ for n_i input variables.

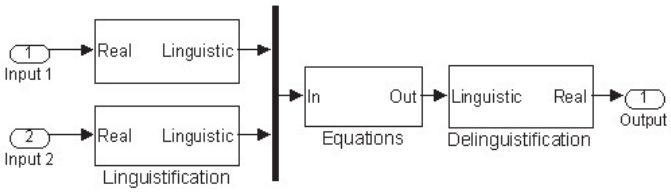


Figure 3: A LE model.

Nonlinear steady-state models can be constructed with linguistic equations, and then extended to dynamic systems using dynamic structures. Case-based systems can include both steady state and dynamic models.

2.2 Takagi-Sugeno type fuzzy systems

Takagi-Sugeno (TS) fuzzy model [4], in which the consequent is a crisp function of the antecedent variables, can be interpreted in terms of local models, see Fig. 4(a). A TS model with a common consequent structure can be understood as a global linear model with input-dependent parameters. For widely used linear functions, the standard weighted mean inference is

$$y = \frac{\sum_{i=1}^{K} \beta_i(\mathbf{x}) y_i}{\sum_{i=1}^{K} \beta_i(\mathbf{x})}, \tag{13}$$

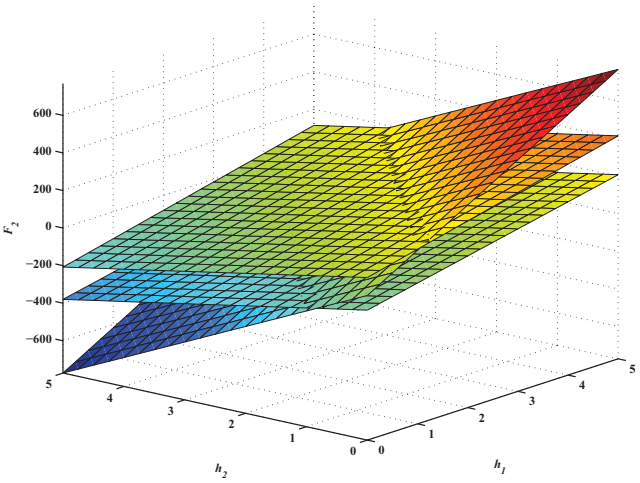
where the degree of fulfillment, $\beta_i(\mathbf{x})$, is the membership degree of the input vector $\mathbf{x}$ in the antecedent of the rule i, see Fig. 4(b). The models surface (Fig. 4(c)) is constructed from the values y_i which are calculated from the local models

$$y_i = \mathbf{a}_i^T \mathbf{x},\ i = 1, 2, \ldots, K. \tag{14}$$

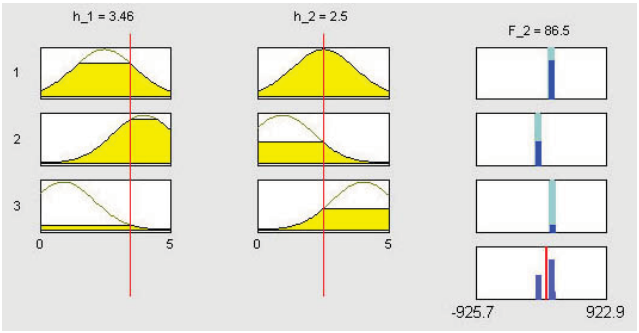
The inference method (13) introduces some undesirable properties in the border areas of the local models if the local models intersect within the border area. These drawbacks can be partly removed by using crisp transitions, $y = \max(y_1, y_2)$ and $y = -\max(-y_1, y_2)$ for a convex and a concave case, respectively. However, the slopes of the local models should differ drastically, and anyway the fuzzy inference is lost and the result is a piece-wise linear approximation. The inference (13) results are much better if the local models intersect outside the border area.

The smoothness of the model is directly dependent on the smoothness of the antecedent membership functions, e.g. the frequently used trapezoidal functions result in nonsmooth ouputs [3].

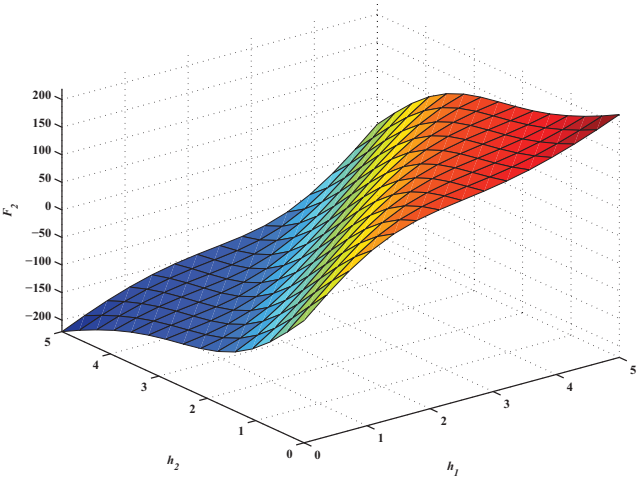
Takagi-Sugeno (TS) type fuzzy models are widely used for the identification of nonlinear systems, since the neuro-fuzzy ANFIS method [7] provides an efficient tuning method for these models. Fitting results are very good with strongly overlapping models, but process insight is lost as the individual models do not have any meaning. The ANFIS tuning increases the overlap of clusters and destroys the meanings of the individual linear models, e.g. the role of some submodels may transform into a part of a smoothing algorithm.



(a) Consequent surfaces.



(b) Fuzzy reasoning.



(c) Model surface.

Figure 4: A Takagi-Sugeno type fuzzy model.

The strong overlap may also result in a steep increase or decrease in the borders of the operating area as can be seen in Fig. 4(c). In dynamic simulation, these steep changes can be alleviated by limiting the range of the output variable, in

ISBN: 978-989-95079-6-8

addition, the operating area of rule 1 should be narrowed down from the solution provided by ANFIS tuning.

Steady-state TS models created with subtractive clustering [11] and ANFIS have proven to be very accurate in the fed-batch enzyme fermentation process [12], however dynamic simulation turned out to be too demanding [13]. Dimension reduction with clustering is necessary when the number of inputs is large. TS models based on grid partitioning require a large number of membership functions and rules which are difficult to tune in practice, e.g. models of the Kappa number in a continuous digester already resulted eight local models with three input variables when two membership functions were used for each variable [14].

The interpolation properties of the TS models can be improved by replacing the weighted mean by mechanisms which result in a piece-wise convex or concave interpolation surface, in which [3]: the gradients are bound by the gradients of the rule consequent functions, and the surface is smooth with continuous derivatives of a sufficient order. Then the model approximates the function more accurately than the surface generated by (13).

For example, the smoothing maximum function [15] connects two consequent hyperplanes, y_1 and y_2, by a smooth convex or concave surface, i.e.

$$y = y_1 + s_\gamma(y_2 - y_1), \tag{15}$$

where $s_\gamma(z)$ is a piece-wise polynomial

$$s_\gamma = \begin{cases} 0, & z \le -\gamma, \\ \frac{z+\gamma)^2}{4\gamma} & -\gamma < z < \gamma, \\ z, & z \ge \gamma \end{cases} \tag{16}$$

The interpolant starts to deviate from the consequent y_1 when $y_2 - y_1 = \gamma$. The smoothing parameter γ, which is a positive real number, is determined for each pair of adjacent rules.

Sharper borders require nonlinear consequent models, i.e. smoothing should be a part of each individual local model.

3 Multimodel LE system

A *multimodel approach* based on fuzzy LE models has been developed for combining specialised submodels [16]. The approach is aimed for systems that cannot be sufficiently described with a single set of membership definitions due to very strong nonlinearities. Additional properties can be achieved because equations and delays can also vary between different submodels. In the multimodel approach, the working area is defined by a separate working point model. The submodels are developed using the case-based modelling approach.

A multimodel system contains several submodels and a fuzzy decision system for selecting a suitable model for each situation using several working point variables. If several inputs are combined into a single working point index, the fuzzy set system is reduced to a fuzzification block (Fig. 5). Linguistic equation (LE) models have been used in several applications [17, 18, 19, 20, 21].

Linguistic Takagi-Sugeno fuzzy models (LTS) belong to this class of models, however with one limitation: the fuzzy partition is defined with same variables as the models. As *LE* models are nonlinear, the local models are also nonlinear.

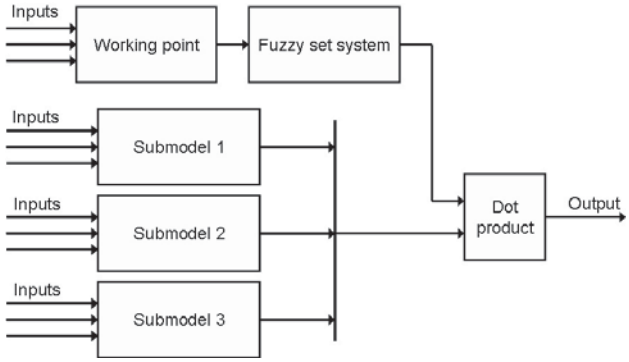


Figure 5: Multimodel LE system with fuzzy decision module.

LTS models can be developed and tuned with the same methods as the normal TS models. The only difference is that the variable values are scaled with the nonlinear scaling functions presented above.

The only difference to the normal LE model is that the equation part is handled with a fuzzy set system (Fig. 6). Nonlinear scaling is done with the same variable specific functions in each local model.

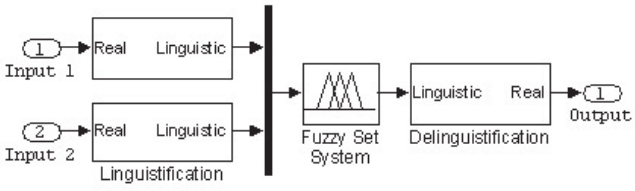


Figure 6: A fuzzy LE simulator.

4 Example: a tank system

The training environment consists of tank systems which include interactive parts where the inflow is divided between two tanks and the flow between the tanks depends on the levels of these tanks, h_1 and h_2, respectively (Fig. 7). The flow F_2 depends on the level of the tanks 1 and 2 (h_1 and h_2):

$$F_2 = c_1\sqrt{h_1 - h_2}, \tag{17}$$

$$F_3 = c_2\sqrt{h_2}. \tag{18}$$

The working point is set with the valve properties presented by coefficients c_1 and c_2.

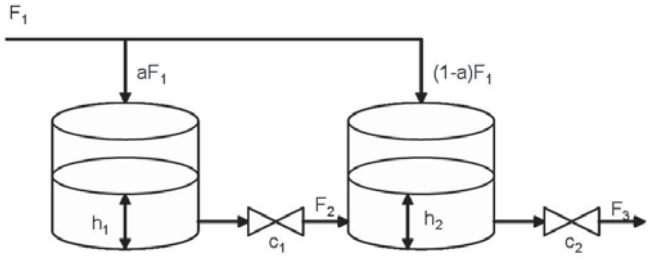
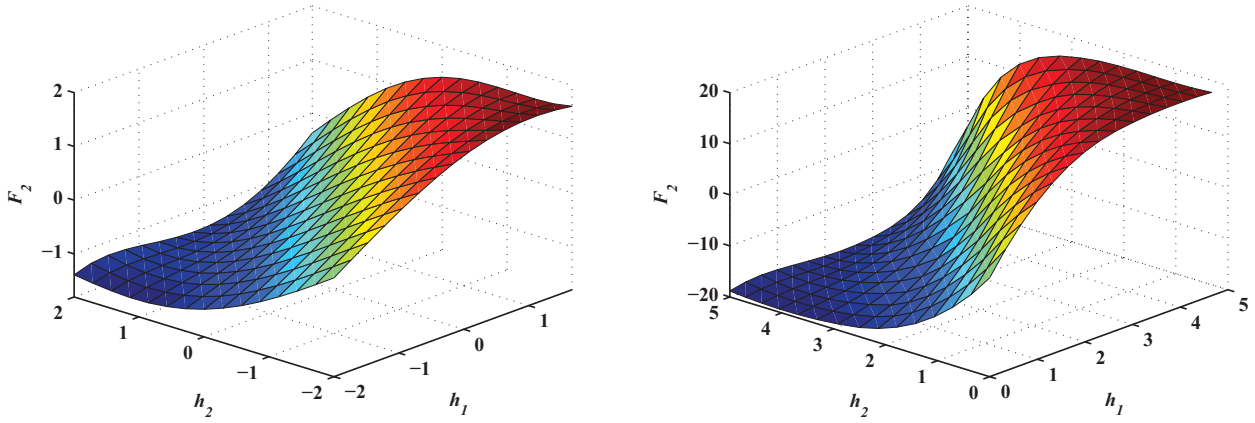


Figure 7: Interactive tank system.

ISBN: 978-989-95079-6-8



(a) The model in the linguistic range.

(b) The model in the real scale.

Figure 8: The model surface of the fuzzy LE model.

TS model. A good fitting can be achieved with a TS model

$$y_1 = 154.2h_1 - 153.9h_2 - 1.98, \quad (19)$$
$$y_2 = 57.61h_1 - 59.86h_2 - 80.08, \quad (20)$$
$$y_3 = 64.35h_1 - 60.18h_2 + 93.15. \quad (21)$$

This model shown in Fig. 4 was tuned with ANFIS resulting in a fairly extensive overlap between the submodels. Strong differences between the submodels can still be observed, but the detection of the models on the basis of their performance is not very clear. These problems were discussed in Section 2.2.

LTS model. A good fitting can be achieved with a LTS model

$$\widetilde{y_1} = 1.704\widetilde{h_1} + 1.785\widetilde{h_2}, \quad (22)$$
$$\widetilde{y_2} = 1.775\widetilde{h_1} + 3.144\widetilde{h_2}, \quad (23)$$
$$\widetilde{y_3} = 1.93\widetilde{h_1} + 5.013\widetilde{h_2}, \quad (24)$$

where the variables h_1 and h_2 are scaled with the nonlinear functions: $\widetilde{h_1} = f_1^{-1}(h_1)$ and $\widetilde{h_2} = f_2^{-1}(h_2)$. The output is first constructed in the linguistic range by using (13) and the submodels $\widetilde{y_1}$, $\widetilde{y_2}$ and $\widetilde{y_3}$. The result is then converted into a real value F_2 by (5) with the scaling parameters of the output variable: a_{out}^-, b_{out}^-, a_{out}^+, b_{out}^+ and c_{out}.

The operating areas are clearly observed, and the operation is smooth (Fig. 8(a)). The first consequent model is rather steep and operates when the levels of two tanks are close to each other (Fig. 8(b)). The second model corresponds to negative flow, i.e. $h_2 > h_1$, and the third one to positive flow, $h_1 > h_2$. No smoothing algorithms are needed, and there are no problems in the boundaries of the operating area (Fig. 8(b)). Performance evaluation also provides a clear indication of the active submodels.

5 Applications

LTS models provide good results if the same variables can be used for defining the operating areas and in the submodels. There are clear differences between the applications in this sense.

5.1 Single LE models

A single LE model proved to be a reliable solution for forecasting the Kappa number in continuous cooking with much less parameters than TS models [14]. Also, the steady-state LE model developed in an early lime kiln application provided better results than a TS model since the weaknesses of the data could be clearly observed and corrected in the LE approach [22].

A single LE model can capture nonlinear behaviour so well that the multimodel structure is not needed. The gas furnace data [23] was modelled using a single LE model, which provided very accurate results even in dynamic simulation [24]. The dynamic model of a flotation unit calculates the outlet turbidity from the properties of incoming water, chemical dosages, and previously calculated turbidity. This model is also used as an indicator of water quality [25]. The dynamic model of fluidised bed granulation consists of three interactive models: temperature, humidity, and granule size. All submodels are based on single LE models. [20]

5.2 Multimodels

Dynamic LE multimodels are used for the control design of two application: a lime kiln [10] and a solar collector field [26]. Smooth changes between the submodels were controlled by a fuzzy decision module as in Fig. 5. The lime kiln model is based on six operating areas defined by the production level and the trend of the fuel feed [16]. The interaction matrices are similar, and the differences between the submodels are introduced by scaling functions. The model has been used in a fuel quality indicator, which had an essential part in the development of a successful control system [10]. The model of a solar collector field consists of four specialised LE models, which are specified with interaction coefficients and scaling function parameters. The fuzzy decision module contains a working point model, which is based partly on different variables [24]. The simulator represents the field operation very accurately, even oscillatory conditions are handled correctly [26].

There are forecasting applications for batch cooking [19] and fed-batch fermentation [21]. In batch cooking, specific submodels are needed due to variations in the quality of the

ISBN: 978-989-95079-6-8

chips and the properties of the incoming cooking liquor. Multimodel aspects are taken into account by adapting the model to different operating conditions by selecting an appropriate speed factor depending on the H-factor and alkali level [19]. The dynamic model of fed-batch fermentation consists of three interactive models: carbon dioxide concentration, oxygen transfer rate, and dissolved oxygen concentration [21]. Transitions between the three phases, lag, exponential growth, and steady state, are defined on the basis of time, oxygen transfer rate, and glucose feed rate. The primary aim of the simulator is to detect fluctuations in the process control.

6 Conclusions

Special cases of fuzzy linguistic equation models can be understood as linguistic Takagi-Sugeno (LTS) type fuzzy models, in which nonlinear scaling is used for both inputs and outputs. The LTS models are robust solutions to applications where the same variables can be used for defining the operating areas and also in the submodels. No special smoothing algorithms are needed near the borders of the submodels.

References

[1] L. A. Zadeh. Knowledge Representation in Fuzzy Logic. *IEEE Transactions on Knowledge and Data Engineering*, 1(1):89–100, 1989.

[2] E. K. Juuso. Integration of intelligent systems in development of smart adaptive systems. *International Journal of Approximate Reasoning*, 35:307–337, 2004.

[3] R. Babuška. *Fuzzy Modeling and Identification*. Kluwer Academic Publisher, Boston, 1998.

[4] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, 1985.

[5] W. Pedrycz. An identification algorithm in fuzzy relational systems. *Fuzzy Sets and Systems*, 13:153–167, 1984.

[6] R. Babuška and H. Verbruggen. Neuro-fuzzy methods for nonlinear system identification. *Annual Reviews in Control*, 27:73–85, 2003.

[7] J.-S. R. Jang. Anfis: Adaptive-network-based fuzzy inference systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3):665–685, 1993.

[8] P. P. Angelov and D. P. Filev. An approach to online identification of Takagi-Sugeno fuzzy models. *IEEE Transactions on Systems, Man, & Cybernetics*, 34(1):484–498, 2004.

[9] T. A. Johansen and R. Babuška. Multiobjective identification of Takagi-Sugeno fuzzy models. *IEEE Transactions on Fuzzy Systems*, 11(6):847–860, 2003.

[10] E. K. Juuso. Fuzzy control in process industry: The linguistic equation approach. In H. B. Verbruggen, H.-J. Zimmermann, and R. Babuska, editors, *Fuzzy Algorithms for Control, International Series in Intelligent Technologies*, pp. 243–300. Kluwer, Boston, 1999.

[11] S. Chiu. Fuzzy Model Identification Based on Clustering Estimation. *Journal of Intelligent & Fuzzy Systems*, 2(3):Sept., 1994.

[12] U. Saarela, K. Leiviskä, E. Juuso, and A. Kosola. Modelling of a fed-batch enzyme fermentation process. In *IFAC International Conference on Intelligent Control Systems and Signal Processing. Faro, Portugal*. IFAC, 2003.

[13] E. K. Juuso. Dynamic simulation of a fed-batch enzyme fermentation process. In *Proceedings of SIMS 2005, 46th Conference on Simulation and Modeling, Trondheim, Norway*, pp. 117–124. Tapir Academic Press, Trondheim, 2005.

[14] K. Leiviskä, E. Juuso, and A. Isokangas. Intelligent modelling of continuous pulp cooking. In K. Leiviskä, editor, *Industrial Applications of Soft Computing*, pp. 147–158. Springer, Heidelberg, 2001.

[15] I. Zhang. A smoothing-out technique for min-max optimization. *Math. Prog.*, 19:61–77, 1980.

[16] E. K. Juuso. Intelligent dynamic simulation of a lime kiln with linguistic equations. In *ESM'99: Modelling and Simulation: A tool for the Next Millenium, 13th European Simulation Multiconference, Warsaw, Poland*, pp. 395–400, Delft, The Netherlands, 1999. SCS.

[17] E. K. Juuso. Modelling and simulation with intelligent methods. In *White paper of the Virtual Institute for Simulation (SimServ): www.sim-serv.com*. Sim-Serv, 2004. 17 pp.

[18] E. K. Juuso. Applications of smart adaptive system in pulp and paper industry. In *Proceedings of Eunite 2004 - European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems, Aachen, Germany*, pp. 21–33. Wissenschaftsverlag Mainz, Aachen, 2004.

[19] E. K. Juuso. Forecasting batch cooking results with intelligent dynamic simulation. In *Proceedings of the 6th EUROSIM Congress on Modelling and Simulation, Ljubljana, Slovenia*, volume 2, page 8 pp. University of Ljubljana, Ljubljana, Slovenia, 2007.

[20] E. K. Juuso. Intelligent modelling of a fluidised bed granulator used in production of pharmaceuticals. In *Conference Proceedings of SIMS 2007 - The 48th Scandinavian Conference on Simulation and Modeling, Göteborg (Särö)*, pp. 101–108. Linkping University Electronic Press, Linkping, Sweden, 2007.

[21] E. K. Juuso. Intelligent dynamic simulation of a fed-batch enzyme fermentation process. In *Tenth International Conference on Computer Modelling and Simulation, EUROSIM/UKSim, Cambridge, UK*, pp. 301–306. The Institute of Electrical and Electronics Engineers IEEE, 2008.

[22] E. K Juuso, T. Ahola, and K. Leiviskä. Fuzzy modelling of a rotary lime kiln using linguistic equations and neuro-fuzzy methods. In *Proceedings of the 3rd IFAC Symposium on Intelligent Components and Instruments for Control Applications - SICICA'97, Annecy, France*, pp. 579–584, 1997.

[23] G. E. P. Box and G. M. Jenkins. *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco, 1970.

[24] E. K. Juuso. Intelligent systems design with linguistic equations. In *9th Workshop Fuzzy Control des GMA-FA 5.22 am 4/5.11.1999, Dortmund, Deutschland. Forschungsberict Nr. 0449*, pp. 177–196, Dortmund, 1999. Universitt Dortmund, Facultt fur Electrotechnik.

[25] I. Joensuu, M. Piironen, and E. Juuso. Dynamic simulator for dosing of water treatment chemicals. In *Proceedings of European Symposium on Computer Aided Process Engineering-15 (Escape-15), Barcelona, Spain, Computer-aided chemical engineering, 20A*, pp. 301–306. Elsevier, Amsterdam, 2005.

[26] E. K. Juuso. Intelligent dynamic simulation of a solar collector field. In *Simulation in Industry, 15th European Simulation Symposium ESS 2003*, pp. 443–449. SCS, Gruner Druck, Erlangen, Germany, 2003.

ISBN: 978-989-95079-6-8

Knowledge Discovery in the Prediction of Bankruptcy

R. J. Almeida[1] S. Vieira[2] V. Milea[1] U. Kaymak[1] J. M. C. Sousa[2]

1.Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
2.Technical University of Lisbon, Instituto Superior Técnico, Dept. Mechanical Engineering/CIS-IDMEC
Av. Rovisco Pais 1, 1049-001 Lisbon
Email: {rjalmeida,kaymak,milea}@ese.eur.nl, {susana,jsousa}@dem.ist.utl.pt

Abstract— *Knowledge discovery in databases (KDD) is the process of discovering interesting knowledge from large amounts of data. However, real-world datasets have problems such as incompleteness, redundancy, inconsistency, noise, etc. All these problems affect the performance of data mining algorithms. Thus, preprocessing techniques are essential in allowing knowledge to be extracted from data. This work presents a real world application of knowledge discovery in databases, with the objective of prediction of bankruptcy. For this task fuzzy classification models based on fuzzy clustering are used, which are developed solely from numerical data. This data set has missing values, extreme values and also presents a much smaller bankruptcy class than the not bankruptcy class, which makes it a challenging problem in the scope of KDD.*

Keywords— Knowledge discovery in databases, feature selection, missing data, noisy data, prediction of bankruptcy, fuzzy classification.

1 Introduction

With the increase of economic globalization and evolution of information technology, financial data are being generated and accumulated at an exponential rate. It is used to keep track of companies, business performance, monitor market changes, and support financial decision-making. This rapidly growing volume of data triggered the need for automated approaches that allow effective and efficient utilization of massive financial data to support companies and individuals in strategic planning and investment decision-making.

Knowledge discovery can contribute to solving business problems in finance by finding patterns, causalities, and correlations in business information and market prices that are not immediately apparent to managers because the volume of data is too large or is generated too quickly to be screened by experts. Knowledge discovery has already been applied to a number of financial applications, including development of trading models, investment selection, loan assessment, portfolio optimization, fraud detection and bankruptcy prediction, amongst others. The prediction of bankruptcy has been been previously investigated in terms of the likelihood of success for the introduction of fuzzy systems for decision support [1]

The prediction of corporate failure or bankruptcy has been characterized as one of the most important problems facing business and government [2]. It also is a problem that affects the economy of every country. The number of failing firms is important for the economy of a country and it can be considered as an index of the development and robustness of the economy. The high individual, economic, and social costs encountered in corporate failures or bankruptcies make this problem very important to parties such as auditors, management, government policy makers, and investors [3].

There is a long history of research attempting to develop bankruptcy prediction models based on financial variables and other indicators of financial distress, using a wide variety of techniques. The pioneer in predicting business failure ratios is considered to be [4]. The predictive accuracy of the initial approaches has varied from around 65% [5] to around 90% [6]. Higher predictive accuracy is often achieved by using samples concentrated in a few industries, using samples with widely varying bankruptcy/non-bankruptcy company sizes or making inappropriate assumptions about real world bankruptcy/non-bankruptcy frequencies.

The data set used in this work, concerning the bankruptcy, is different from all of the above mentioned, so no direct comparison of results can be made. Also, other works never mentioned that data sets had missing values and extreme values, as in this work.

In this work we present a full KDD process, applied in a real-world problem: the prediction of bankruptcy. The data set used is quite challenging as it has missing values and extreme values. It also presents a much smaller bankruptcy class than the not bankruptcy class. Several possibilities were tested in each step of the KDD process, such as data preparation, feature selection and fuzzy classification, and we discuss them briefly although we give more focus on the best results obtained. Note that the use of fuzzy systems for classification, besides building a numeric prediction model also represents the model behaviour in terms of linguistic rules, making it possible to interpret, which is an important final step in the KDD process.

The outline of the paper is as follows. In Section 2 we briefly present the KDD steps used to obtain a bankruptcy classifier. In Section 3 we present techniques used for data preparation. Feature selection is presented in Section 4 and Section 5 describes the procedures used to derive Takagi-Sugeno fuzzy models by means of fuzzy clustering. The data used in the work and the results are presented in Section 6. Finally, conclusions and future work are given in Section 7.

2 KDD process

The search for knowledge in large data sets, with the use of different hypothesis spaces, is the central and necessary phase within the discovery process. A large number of methods have been developed that handle many search tasks, but hypotheses

ISBN: 978-989-95079-6-8

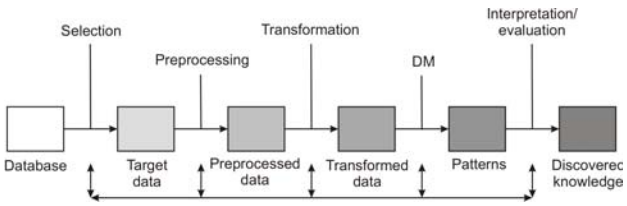


Figure 1: Phases in the KDD process, adapted from [8]

inference and verification is only a part of the whole process of knowledge discovery. As any other process, it has its environment, its phases, and runs under certain assumptions and constraints. The process undertakes many phases, namely [7]:

1. Definition and analysis of the problem;
2. Understanding and preparation of data;
3. Setup of the search for knowledge;
4. The actual search for knowledge;
5. Interpreting mined patterns;
6. Deployment and practical evaluation of the solutions.

The KDD process, activities and phases, is shown in Figure 1.

Compared to the traditional manual analysis, KDD provides a much higher degree of system autonomy, especially in processing large hypotheses spaces. However, at the current state of the art, a human analyst still makes many decisions in the course of a discovery process.

The KDD process starts from specification of a given problem and data understanding, and ends with actionable conclusions from the discovered knowledge. The output of DM is, in general, a set of patterns, some of which possibly represent discovered knowledge. In the next sections we will briefly explain each step of the KDD process.

3 Data Preparation

Real-world databases are highly susceptible to noisy, missing, and inconsistent data due to their typically huge size. Thus, data preparation consumes most of the time needed to mine data [9], and can substantially improve the overall quality of the patterns mined and/or the time required for the actual mining [10].

This section is a brief overview of some of the concepts regarding data preparation to yield the best possible model. For more details refer to [9, 10].

3.1 Missing Data

When applying data analysis methods to real problems, we often find that the data sets contain many missing elements. There are two forms of randomly missing data: missing completely at random (MCAR) and missing at random (MAR) [11]. Missing values MCAR, behave like a random sample and their probability does not depend on the observed data or the unobserved data [12, 13]. MAR exists when missing values are not randomly distributed across all observations but are randomly distributed within one or more subsamples.

A possible approach to deal with the missing values is to discard all incomplete data, and then execute the data analysis method on the remaining data. However, if missing values are frequent, the data set size may be considerably reduced, yielding unreliable or distorted results. A way to minimize this extreme data reduction problem is presented in [14]. If values are missing completely at random they can be imputed. Widely used imputation methods use the variables mean, median or the most probable value as a replacement [11, 10].

3.2 Noisy Data

Noise is considered to be a random error or variance in a measured variable [10]. A specific type of noise, which the data in this study has, is extreme values, which in some cases can be labelled as outliers. In this work numerically sensitive mining algorithms are used, and it is recommended to normalize the individual variable distributions [15].

Outliers are defined as a large deviation from the mean value of the rest of the data. Usually outliers can be thrown out of the data set, as they bias the analytical results. Distribution normalization deals with the problems presented by valid outliers. An outlier is valid if it represents an accurate measurement and still falls well outside the range of the majority of values. These should not be discarded.

The distribution normalization can be achieved using transformations. In cases where the data have strong asymmetry, many outliers or batches at different levels with widely differing spreads a power transformation may alleviate this problem, without violating the necessary transformation properties [16].

4 Feature Selection

One of the great challenges in classification is selecting the important input variables from all possible input variables. Classification problems involve a large number of potential inputs. The number of inputs actually used by the model must be reduced to the necessary minimum, especially when dealing with fuzzy models that are, presumably, nonlinear and contain many parameters. Therefore, it is necessary to select carefully the variables that are relevant for the feature class.

Feature selection is a process that chooses a subset of M features from the original set of n features ($M \leq n$), so that the feature space is optimally reduced according to a certain criteria.

Even when a good criteria exists for model selection, there is no guarantee that a model based on a given set of variables is optimal unless all possible combinations of variables have been explored. The problem is known to be NP-hard [17]. Hence, finding an optimal solution requires building a model for each possible combination of input variables, which becomes computationally prohibitive for problems involving even a moderate number of candidate input variables.

Feature selection algorithms, essentially divide into wrappers and filters [18]. In this work we use wrappers that make use of specific learning algorithms to evaluate variables in the context of the learning problem, rather than independently. Wrappers share strengths and weaknesses of the learning algorithms and have the advantage of using the actual hypothesis accuracy as a measure of subset quality. Furthermore, wrapper methods do tend to outperform filter methods [18].

ISBN: 978-989-95079-6-8

The variable selection procedure can be used with various performance criteria for model selection. In real-world databases sometimes one of the classes is more difficult to classify then the others. This can happen, for instance, when one of the classes is much bigger than the other or the interest of the problem is a specific class. To cope with this problem, we use a criterion, that assigns specific weights to each class in the model evaluation of the feature selection algorithm [19].

In this work we compare the well known sequential forward selection (SFS) and the sequential backward elimination (SBE) against the newly proposed ant feature selection algorithm (AFS). The SFS and SBE search algorithms may not be the best search methods, nor guarantee an optimal solution, but they are popular because they are simple, fast, provide a very reasonable solution and are much more efficient than exhaustive search.

Sequential forward selection and sequential backward elimination were first used in the context of feature selection for pattern classification. SFS was first used in [20] and was later used in [21]. It has also been used to determine input variables for fuzzy models in [22] and [23]. SBE was first used in [20] and in [24] it was used with simple linear models to provide a first-round elimination of the input variables for a fuzzy model.

4.1 Ant Feature Selection

Ant algorithms were first proposed by Dorigo [25] as a multi-agent approach to difficult combinatorial optimization problems, such as traveling salesman problem, quadratic assignment problem or supply chain management [26]. Here we present an implementation of ACO applied to feature selection, where the best number of features is determined automatically.

In this approach, two objectives are considered: minimizing the number of features and minimizing the error classification. Two cooperative ant colonies optimize each objective. The first colony determines the number (cardinality) of features and the second selects the features based on the cardinality given by the first colony. Thus, two pheromone matrices and two different heuristics are used. The heuristic value is computed using the Fisher discriminant criterion for feature selection [27], which ranks the features giving them a given relative importance.

The best number of features is called *features cardinality* N_f. The determination of the *features cardinality* is addressed in the first colony sharing the same minimization cost function J^τ with the second colony, which in this case aggregates both the maximization of the classification accuracy and the minimization of the features cardinality. Hence, the first colony determines the size of the subsets of the ants in the second colony, and the second colony selects the features that will be part of the subsets.

The objective function of this optimization algorithm aggregate both criteria, the minimization of the classification error rate and the minimization of the features cardinality:

$$J^\tau = w_1 \frac{N_e^\tau}{N_n} + w_2 \frac{N_f^\tau}{n} \tag{1}$$

where $\tau = 1, \ldots, g$, g being the number of ants, w_1 and w_2 are weights, N_n is the number of used data samples, n is the total number of features, N_e is the number of errors produced by the solution and N_f is the features cardinality.

To evaluate the classification error, a fuzzy classifier is built for each solution following the procedure described in Section 5. This approach was presented in [28].

5 Fuzzy Classification

In this section we outline the basics of the adopted fuzzy reasoning scheme for pattern classification problems. Let us consider a n-dimensional classification problem for which N patterns $\vec{x}_p = (x_p^1, \ldots, x_p^n)$, $p = 1, 2, \ldots, N$ are given from κ classes $C_1, C_2, \ldots, C_\kappa$. The task of a pattern classifier is to assign a given pattern $\vec{x}$ to one of the κ possible classes based on its features values. Thus, a classification task can be represented as a mapping $\psi : X \subset \Re^n \rightarrow \{0,1\}^\kappa$ where $\psi(\vec{x}) = \vec{c} = (c_1, ..., c_\kappa)$ such that $c_k = 1$ and $c_j = 0 (j = 1, \ldots, \kappa, j \neq k)$.

Assuming that there is an arbitrary ordering of the classes, one way to solve this classification problem is to consider classifiers with a continuous output, e.g., a Takagi-Sugeno affine system [29]. The output of an affine Takagi-Sugeno fuzzy rule is

$$y_k = \vec{a}\vec{x} + b_k \tag{2}$$

where y_k is the output for rule k, $\vec{a}_k$ is a parameter vector and b_k is a scalar offset. In classification problems the output should be a discrete value corresponding to one of the classes to be identified. So a threshold T_l can be used on the output y_k, to decide which class that it belongs to, as $x_k \in C_l$ if $y_k \in T_l$.

To form the fuzzy system model from the data set with N data samples, given by $X = [\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N]^T$, $Y = [y1, y2, \ldots, yN]^T$ where each data sample has a dimension of n ($N >> n$), the structure is first determined and afterwards the parameters of the structure are identified. The number of rules characterizes the structure of a fuzzy system. Fuzzy clustering in the Cartesian product-space $X \times Y$ is applied to partition the training data. The partitions correspond to the characteristic regions where the systems behavior is approximated by local linear models in the multidimensional space. Given the training data X_T and the number of clusters K, a suitable clustering algorithm is applied.

The fuzzy clustering algorithms used in this work are based on the optimization of an objective function. In particular we use the fuzzy c-means (FCM) [15], the Gustafson-Kessel (GK) [30], the possibilistic c-means (PCM) [31], fuzzy possibilistic c-means (FPCM) [32] and the recent possibilistic fuzzy c-means (PFCM) [33].

The FCM functional uses a probabilistic constraint which states that the sum of membership degrees must equal one [15]. Problems arise in situations, where the total membership of a data point to all the clusters does not equal one, as in the presence of outliers. Clearly, one would like the memberships for representative feature points to be as high as possible, while unrepresentative points should have low membership in all clusters. The PCM objective function relaxes this constraint [31].

Gustafson-Kessel extended the standard fuzzy c-means algorithm by employing an adaptive distance norm, in order to detect clusters of different geometrical shapes in one data set [30]. As the GK algorithm is based on an adaptive distance

ISBN: 978-989-95079-6-8

measure, it is less sensitive to scaling (normalization, standardization) of the data.

Fuzzy-possibilistic c-means, simultaneously produces both memberships and possibilities. FPCM tries to solve the noise sensitivity defect of FCM, and also overcomes the coincident clusters problem of PCM. Note that FCM and FPCM will not generate the same membership values, even if both algorithms are started with the same initialization [32].

FPCM imposes a constraint on the typicality values. PFCM relaxes this constraint but retains the column constraint on the membership values. The PFCM functional has two constants that define the relative importance of fuzzy membership and typicality values.

Memberships and typicalities are both considered important for correct interpretation of data substructure. When the objective is to classify a data point, membership may be a better choice as it is natural to assign a point to that cluster whose representative vector is closest to the data point. On the other hand, when seeking the clusters, i.e., while estimating the centroids, typicality is an important means for alleviating the undesirable effects of outliers.

Depending on the clustering algorithm used in this work, a fuzzy partition matrix $U = [\mu_{ik}]$ and/or the typicality matrix $T = [t_{ik}]$ will be obtained. The fuzzy sets in the antecedent of the rules are identified by means of the matrix U and T which have dimensions $[N \times K]$. One dimensional fuzzy sets A_{ij} are obtained from the multidimensional fuzzy sets by projections onto the space of the input variables x_j. This is expressed by the point-wise projection operator of the form $\mu_{A_{ij}}(x_{jk}) = \text{proj}_j(\mu_{ik})$ The antecedent membership functions can now be obtained from the fuzzy partition matrix or from the typicality matrix. The point-wise defined fuzzy sets A_{ij} are then approximated by appropriate parametric functions. The consequent parameters for each rule are obtained by means of linear least square estimation, which concludes the identification of the classification system.

6 KDD Applied to Prediction of Bankruptcy

The data set used in this work contains data from 1817 companies, each one described by 52 features (including the class feature), containing financial, behavioral and qualitative features (as perceived by the account manager).

Some of the features contain extreme values (EV) as their value is 10000 larger then the other values. In this case, the extreme values are believed to be different values for different type and size of companies, containing valuable information.

The data set contains about 10% of missing values. Only 18.38% (334) of the companies have all the features complete and there are only 5 features without missing values, one of those features being the status of the company.

Each company in the data set has two possible status (classes): status 0 (bankrupt) and status 1 (not bankrupt). The distribution of the classes is uneven: only 4.3% (78) of the companies have the status 0, and the remaining 95.7% (1739) have a status 1. This distribution skewness is common in bankruptcy data.

The KDD steps taken in this work to obtain a compact fuzzy classification model are the following:

1. Manual selection of relevant data from the available data.
2. Preprocessing of the data to deal with extreme values, using a power transformation followed by a linear transformation.
3. Preprocessing of the data to deal with missing values, by replacement of missing values using the most probable value.
4. Searching and selecting the relevant features using search algorithms.
5. Obtaining a fuzzy classification model by using only the selected features in the previous step.

The following computational protocols were used: $\varepsilon = 0.0001$, maximum number of iterations 100, fuzzy exponent $m = 2$ and $\eta = 2$, and the Euclidean norm is used. For both PCM and PFCM we first run FCM to termination. All trials terminated with the convergence criteria after a few iterations.

6.1 Data Preprocessing

When analysing the data we found that seven features can be discarded because they contain company descriptive features that may not be relevant to this research.

Extreme Values Linear transformations of re-expressed data present little additional difficulty in interpretation. Since power transformations are monotonic for positive data values, we transform the data so that it has only positive values, maintaining the missing values, using the linear transformation,

$$z_{jk} = z^*_{jk} + |\min z_j| \tag{3}$$

where the asterisk denotes the unscaled data. After this linear transformation, which does not alter the shape of the data, we apply a power transformation $T(x) = \log(x)$, because this transformation alters the distribution of the data and bring the extreme values to a value closer to the other values [16]. We chose a matched transformation, presented as:

$$z^1 = a \; + \; bT(x), \tag{4}$$

where a and b are chosen, by using a point x_0 and require that:

$$z_0 = a \; + \; bT(x) = x_0, \tag{5}$$

and, furthermore, that the derivative of z with respect to x, evaluated at x_0, to be 1. That is,

$$\left.\frac{dz}{dz}\right|_{x_0} = \frac{d[a + bT(x)]}{dx} = b \left.\frac{dT(x)}{dx}\right|_{x_0} = 1. \tag{6}$$

This method relies on the linearity of the transformation near the center of the range x. For sake of simplicity, we chose x_0 as the median value of each feature. Other point could have been chosen. Furthermore, we normalized the obtained values of the matched transformation so that all the features are contained between the interval $[0, 1]$, obtaining the data set z^2.

Missing Values Almost every company has missing values for their features, and so discarding the companies with missing values cannot be considered a feasible approach for this research. Filling in the missing value during the data preprocess, was used. After the data transformation, we imputed the missing values using probable value, calculated using the Expectation-Maximization (EM) algorithm [34].

ISBN: 978-989-95079-6-8

6.2 Fuzzy Classification

One of the most important advantages of the, rather complex, transformation described, can be seen in Table 1. This table shows the obtained accuracy obtained with the data before the transformation, using 6 clusters, and the obtained results after the transformation with only 3 clusters. Since the class 0 is so small, we used 4-fold cross validation. As can be seen the results obtained with the data before transformation are more disperse. This does not happen with the transformed data. The GK algorithm could not be used in the original data because of the extreme values. For this reason we decided not to included it in this comparison. Furthermore the converge of the clustering algorithms is improved and less number of clusters are needed to obtain good results.

Table 1: Accuracy obtained for bankruptcy data with 6 clusters (Raw data) and 3 clusters (Processed data).

	Raw data			Processed data		
Alg.Cl.	Max.	Mean	Min.	Max.	Mean	Min.
FCM	0.928	0.885	0.800	0.972	0.946	0.919
PCM	0.925	0.894	0.733	0.969	0.945	0.919
FPCM-U	0.924	0.877	0.800	0.953	0.944	0.939
PFCM-U	0.926	0.894	0.733	0.955	0.941	0.927
FPCM-T	0.925	0.891	0.800	0.939	0.939	0.939
PFCM-T	0.925	0.891	0.800	0.939	0.925	0.919

6.3 Feature Selection

After the fuzzy classification models with all the features were extracted, we look to further simplify the model, using feature selection. We applied each algorithm of feature selection (SFS,SBE and AFS), using the weighted accuracy for each feature class present [19], and all of the clustering algorithms in study (GK,FCM, FFCA, PCM, FPCM and PFCM), to both the data normalized and not normalized.

In this case, if no weights were used, the algorithms would choose the features that would maximize the class no bankruptcy as this is the dominant class. Usually after one or two iterations the process would stop. In this case, without the use of the weights the results are highly biased towards the bigger class.

The features which were selected throughout our tests varied between 2 with the sequential forward search and 37 with the sequential backward search. This result was expected as different types of features selection algorithms choose features in a different manner, and we tested a number of different clustering algorithms. Good results were obtained using the ants feature selection. The number of features chosen was 15, 3 rules were derived and the obtained accuracy was 78.9% for companies that are bankrupt and 94.9% for companies that are not bankrupt. The fuzzy clustering algorithm used was PCM.

The use of fuzzy systems for classification, besides building a numeric prediction model also represents the model behaviour in terms of linguistic rules, which is very natural for human to understand. Interpretability is considered to be the main advantage of fuzzy systems over other non-fuzzy alternatives like statistical models or neural networks.

The simplest classification model obtained with the data normalized, only has 2 features and 3 rules and was derived using the PCM algorithm. The obtained accuracy is 57.9% for the companies that are bankrupt, whereas the percentage of companies that are not bankrupt is 97.7%. If we compare the possibility of interpretability of this model against the model derived in [3], that had 70 rules and 35 features then it can be considered that these results are good. Also bare in mind that this data set has missing values, extreme values and one class which is much smaller than the other.

The rule base model is relatively simple, as only 3 rules and 2 features are used, the understanding of the consequents of the model is not a simple task. The obtained rules, for this simple model are:

1. **If** $PROFIAT$ is Low **and** $CFEQ$ is Low **then** $y_1 = 0.35PROFIAT + 0.16CFEQ - 0.16$
2. **If** $PROFIAT$ is $Medium$ **and** $CFEQ$ is $Medium$ **then** $y_2 = -24.52PROFIAT - 18.65CFEQ + 39.12$
3. **If** $PROFIAT$ is $High$ **and** $CFEQ$ is $High$ **then** $y_3 = 5.54PROFIAT + 8.49CFEQ - 11.85$

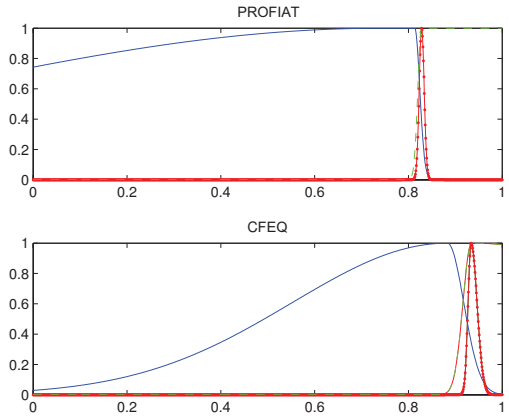


Figure 2: Membership Functions bankruptcy Model. solid - Low, dotted - Medium, dashed - High.

The obtained membership functions are shown in Fig. 2. The features we have used are the profit after tax (PROFIAT), and cash flow to equity (CFEQ). Note that the former is the company's net operating after tax profit for investors, while the latter is the cash that can be paid to the equity shareholders after the company expenses. It is interesting to note that the membership functions are skewed towards the higher values, which indicates that the difference between company defaulting or not is quite small. According to our simple model, a small value on either of these features is an indication that the company is performing badly and it is likely to default. It is interesting to note that the *medium* membership function are located in high values. Therefore, even companies with high values of both features can still default.

7 Conclusions

In this work we present a real world application of knowledge discovery for prediction of bankruptcy using a databases that has noisy, missing, and inconsistent data. We present for each step of the KDD process several possibilities that we tested in order to obtain good fuzzy classification models. Before fuzzy models could be extracted from the data, it is necessary to represent the real-world objects of interest in the data

ISBN: 978-989-95079-6-8

in a way that this specific method can access the data. With data preparation, although time consuming, it was possible to derive compact fuzzy models with only a few features that predicted the bankruptcy with an high accuracy rate.

Acknowledgments

This work is partially supported by the Portuguese Government and FEDER under the programs: Programa de financiamento Plurianual das Unidades de I&D da FCT (POCTI-SFA-10-46-IDMEC) and by the FCT grant SFRH/25381/2005, Fundação para a Ciência e a Tecnologia, Ministério do Ensino Superior, da Ciência e da Tecnologia, Portugal.

References

[1] M. Setnes, U. Kaymak, and H.R. van Nauta Lemke. Fuzzy target selection in direct marketing. *IEEE Transactions on Fuzzy Systems*, 7(1):368–369, 1999.

[2] D.E. O'Leary. Using neural networks to predict corporate failure. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 7(3):187–197, 1998.

[3] J. Stefanowski and S. Wilk. Evaluating business credit risk by means of approach-integrating decision rules and case-based learning. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 10(2):97–114, June 2001.

[4] E. I. Altman. Financial ratios, discriminate analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4):589–609, 1968.

[5] D.H. Lindsay and A. Campbell. A chaos approach to bankruptcy prediction. *Journal of Applied Business Research*, 12(4):1–9, 1996.

[6] T.B. Bell, G.S. Ribar, and J. Verchio. Neural nets versus logistic regression: a comparison of each models ability to predict commercial bank failures. In *Proceedings of the 1990 Deloitte and Touche University of Kansas Symposium on Auditing Problems*, pages 29–54, 1990.

[7] G. Piatetsky-Shapiro U. Fayyad and P. Smyth. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34, 1996.

[8] P. Smyth U.M. Fayyad, G. Piatetsky-Shapiro and R. Uthurusamy. *Advances in Knowledge Discovery and Data Mining*. MIT Press, Cambridge, MA, USA, 1996.

[9] D. Pyle. *Data Preparation for Data Mining*. Morgan Kaufmann, San Francisco, CA, 1999.

[10] J. Han and M. Kamber. *Data Mining Concepts and Techniques*. Morgan Kaufmann Publishers Inc, San Francisco, CA, 2000.

[11] H. Timm, C. Döring, and R. Kruse. Differentiated treatment of missing values in fuzzy clustering. In Taner Bilgiç, Bernard De Baets, and Okyay Kaynak, editors, *Proceedings of the 10th International Fuzzy Systems Association World Congress (IFSA 2003), Lecture Notes in Artifical Intelligence*, volume 2715.

[12] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data, Second Edition*. Wiley and Sons, New York, 2002.

[13] J. L. Schafer. *Analysis of Incomplete Multivariate Data*. Chapman & Hall, London, 1997.

[14] M. Setnes and U. Kaymak. Fuzzy modeling of a client preference from large data sets: an application to target selection in direct marketing. *IEEE Transactions on Fuzzy Systems*, 9(1):153–163, 2001.

[15] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[16] M.A. Stoto J.D. Emerson. Transforming data. In F. Mosteller D.C. Hoaglin and J.W. Tukey, editors, *Understanding Robust and Exploratory Data Analysis*, chapter 4. Wiley-Interscience, New York, London, Sydney, Toronto, 2000.

[17] E. Amaldi and V. Kann. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science*, 209(1–2):237–260, 1998.

[18] R. Kohavi and G.H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.

[19] C. A. Silva R.J. Almeida and J.M.C. Sousa. A new feature selection criterion for fuzzy classification. In *Proc. of the 2006 IEEE International Conference on Fuzzy Systems, WCCI 06*, pages Session Classification, pp. 1513–1520, Vancouver, Canada, July 2006.

[20] T. Marill and D.M. Green. On the effectiveness of receptors in recognition systems. *IEEE Transactions on Information Theory*, 9(1):11–17, 1963.

[21] A.W. Whitney. A direct method of nonparametric measurement selection. *IEEE Transactions on Computers*, 20(9):1100–1103, 1971.

[22] M. Sugeno and G.T. Kang. Structure identification of fuzzy model. *Fuzzy Sets and Systems*, 28(1):15–33, 1988.

[23] M. Sugeno and T. Yasukawa. A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, 1(1):7–31, February 1993.

[24] H. Watanabe K. Tanaka, M. Sano. Modeling and control of carbon monoxide concentration using a neurofuzzy technique. *IEEE Transactions on Fuzzy Systems*, 3(3):271–279, 1992.

[25] Marco Dorigo, M. Birattari, and T. Stützle. Ant colony optimization. *IEEE Computational Intelligence Magazine*, 1(4):28–39, 2006.

[26] C. A. Silva, J. M. C. Sousa, T. A. Runkler, and J. M. G. Sá da Costa. Distributed optimization of a logistic system and its suppliers using ant colony optimization. *International Journal of Systems Science*, 37(8):503–512, 2006.

[27] R. O. Duda, P. E. Hart, and D. G Stork. *Pattern Classification, Second Edition*. Wiley–Interscience Publication, 2001.

[28] S. M. Vieira, J. M. C. Sousa, and T. A. Runkler. Fuzzy classification in ant feature selection. In *Proc. of 2008 IEEE World Congress on Computational Intelligence, WCCI 2008*, pages 1763–1769, Hong Kong, China, June 2008.

[29] T. Takagi and M. Sugeno. Fuzzy identification of systems and its aplication to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15(1):116–132, 1985.

[30] D. Gustfson and W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *Proceedings of IEEE Conference on Decision and Control, CDC*, pages 761–766, San Diego, CA, USA, 1979.

[31] R. Krishnapuram and J.M. Keller. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110, May 1993.

[32] N.R Pal, K. Pal, and J.C Bezdek. A mixed c-means clustering model. In *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, volume 1, pages 11–21, Barcelona, Catalonia, July 1997.

[33] J.M. Keller N.R. Pal, K. Pal and J.C.Bezdek. A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions of Fuzzy Systems*, 13(4):517–530, August 2005.

[34] T. Schneider. Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate*, 14:853–871, October 2001.

Generalizations of Fuzzy C-Means Algorithm to Granular Feature Spaces, based on Underlying Metrics: Issues and Related Works

Vasile Georgescu

Faculty of Economics and Business Administration, University of Craiova
Craiova, Romania
Email: vgeo@central.ucv.ro, v_geo@yahoo.com

Abstract—*This paper considers dissimilarity measures and clustering techniques for two special cases of set-defined objects: fuzzy granules and subsequence time series. To deal with clustering of such kind of objects, we propose two implementations that generalize the Fuzzy C-Means algorithm to granular feature spaces. Granular computing is a paradigm oriented towards capturing and processing meaningful pieces of information, the so-called information granules. In a granular feature space, such as a space populated with p-dimensional fuzzy granules, we are concerning with both granular data samples and granular centroids (center of clusters). In order to accommodate clustering algorithms to work in a granular environment we have to choose and/or define appropriate metrics and descriptors. Either a crisp distance between granules or the defuzzified value of a fuzzy distance has to be chosen. On the other hand, subsequence time series clustering requires a generalization of Fuzzy C-Means algorithm in a similar way. It involves a set-defined centroid and appropriate dissimilarity measures to determine the degree to which time sequences are different from their centroid. Furthermore, we discuss related work in granular clustering and subsequence time series clustering.*

Keywords—Granular clustering, Metrics in granular feature spaces, Fuzzy C-Means clustering, Agglomerative granular clustering algorithms, Subsequence time series clustering.

1 *P*-dimensional fuzzy granules: representation and cardinality measure

In what follows, the attention will be restricted to the class Φ^p of normal fuzzy convex granules on $\Re^p$, whose α-level sets are nonempty compact convex sets for all $\alpha > 0$. Each fuzzy granule $M \in \Phi^p$ is uniquely characterized by its support function $s_M(u,\alpha) = \sup\{ < u, x > | \, x \in M^\alpha \}$, $u \in S^{p-1}, \alpha \in (0,1]$, where S^{p-1} is the (p-1)-dimensional unit sphere of $\Re^p$ (i.e. $\|u\| = 1$) and $< \cdot, \cdot >$ is the inner product in $\Re^p$. The support function is a mapping from the class of fuzzy sets Φ^p into the space of functions $L\left(S^{p-1} \times [0,1]\right)$, which preserves addition and multiplication with non-negative scalars.

A p-dimensional fuzzy granule $A = A^1 \times \cdots \times A^p$ is defined on the product space $X = X_1 \times \cdots \times X_p$. Representation and cardinality measure for A can be given either in terms of membership functions:

$$\mu_A : \Re^p \to [0,1],$$
$$\mu_A(x) = \min\left(\mu_{A^1}(x_1), \ldots, \mu_{A^p}(x_p) \right), \quad \forall\, x \in X$$

$$|A| = \int_{\xi \in Supp(A)} \mu_A(x)\, dx =$$
$$= \int_{x_1 \in Supp(A^1)} \cdots \int_{x_p \in Supp(A^p)} \min\left(\mu_{A^1}(x_1), \ldots, \mu_{A^p}(x_p)\right) dx_1 \ldots dx_p$$

or in terms of α – level intervals:

$$A(\alpha) = A^1(\alpha) \times \cdots \times A^p(\alpha); \qquad \alpha \in [0,1]$$

$$|A| = \int_0^1 |A(\alpha)|\, d\alpha = \int_0^1 \prod_{j=1}^{p} \left|A^j(\alpha)\right| \, d\alpha.$$

In particular, let A^j be an LR-fuzzy set. We have:

$$A^j(\alpha) = \left[A^{jL}(\alpha),\, A^{jR}(\alpha)\right]_{LR}$$

$$|A| = \int_0^1 \prod_{j=1}^{p} \left| A^{jR}(\alpha) - A^{jL}(\alpha) \right| d\alpha.$$

The latest integral can be evaluated numerically by means of a quadrature formula (e.g. the adaptive Simpson quadrature).

2 Metrics in granular feature spaces

2.1 A few examples

We focus on distances in granular feature spaces populated with fuzzy granules. There is an enormous literature that covers this topic (at least in case of one-dimensional fuzzy spaces). There is also a large diversity of approaches, among which we can distinguish:

a) Membership focused distances (vertical)
b) Spatially focused distances (horizontal)
c) Mix of spatial and membership distances (tolerance)
d) Feature distances (low or high dimensional representations)
e) Morphological (mixed focus)

Most of these metrics are defined as crisp distances. As a typical example, we can consider an L_2-metric on the space of normal compact convex fuzzy sets Φ^p using the L_2-metric on the Hilbert space of square-integrable functions

ISBN: 978-989-95079-6-8

$L_2\left(S^{p-1}\times[0,1]\right)$. This space is equipped with the inner product

$$< M, N >_\lambda = p\cdot \int_{[0,1]} \int_{S^{p-1}} s_M(u,\alpha)\ s_N(u,\alpha)\ \mu(du)\ \lambda(d\alpha).$$

where μ is normalized Lebesgue measure on S^{p-1} (i.e. $\mu(S^{p-1})=1$) and λ is normalized Lebesgue measure on $[0,1]$. Now, the corresponding L_2-metric results of the following form:

$$\delta_2^{(\lambda)}(M,N)=\|s_M-s_N\|_2 = \left(p\cdot \int_{[0,1]}\int_{S^{p-1}} |\, s_M(u,\alpha)-s_N(u,\alpha)\,|^2 \mu(du)\,\lambda(d\alpha)\right)^{1/2}.$$

Set-defined distances (such as fuzzy distances) have been also proposed ([1]).

For example, a *fuzzy distance* of two non-empty one-dimensional fuzzy sets A and B is defined as the fuzzy set $D_f(A,B)=\left(\Re^+,\ \mu_{D_f}(A,B)\right)$ with the membership function

$$\mu_{D_f(A,B)}(y)=\begin{cases}\sup\{\alpha\in[0,1]\mid A_\alpha, B_\alpha\neq\varnothing;\ \rho(A_\alpha,B_\alpha)\le y\} \\ \quad for\ y\le \lim\limits_{\beta\to\min\{h(M),h(N)\}_-}\rho(A_\beta,B_\beta) \\ 0 \qquad\qquad otherwise\end{cases}$$

A generalization of a fuzzy distance to the *p*-dimensional feature space is easiest if we define fuzzy granules as a Cartesian product of fuzzy sets, where the expression for the fuzzy distance can be simplified using distances in each dimension. A simplification can be obtained by choosing some particular kinds of fuzzy sets.

2.2 *A fuzzy distance between p-dimensional fuzzy granules*

In [5] we introduced a fuzzy distance between *p*-dimensional fuzzy granules defined as a Cartesian product of LR-fuzzy sets. Let us consider a p-dimensional feature space $X=X_1\times\cdots\times X_p$ and a configuration of n objects, each *one described by a p-dimensional fuzzy granule. Two arbitrary objects u and v are imprecisely located in the feature space by means of two fuzzy granules:* $A_u=A_{u1}\times\cdots\times A_{up}$ *and* $A_v=A_{v1}\times\cdots\times A_{vp}$. *Assume the components* A_{uk} *and* A_{vk} $(k=1,\ldots,p)$ *belonging to the* k^{th} *dimension of X are LR-fuzzy sets, i.e.* $A_{uk}(\alpha)=(A^L_{uk}(\alpha),\ A^R_{uk}(\alpha))$ *and* $A_{vk}(\alpha)=(A^L_{vk}(\alpha),\ A^R_{vk}(\alpha))$. *Therefore, the following inequalities hold:*

$$A^L_{uk}(\alpha)\le x_{uk}(\alpha)\le A^R_{uk}(\alpha);\quad A^L_{vk}(\alpha)\le x_{vk}(\alpha)\le A^R_{vk}(\alpha).$$

By subtracting $x_{vk}(\alpha)$ from $x_{uk}(\alpha)$, we obtain

$$A^L_{uk}(\alpha)-A^R_{vk}(\alpha)\le x_{uk}(\alpha)-x_{vk}(\alpha)\le A^R_{uk}(\alpha)-A^L_{vk}(\alpha)$$

and then, using well-known properties such as $\min\limits_{w\le x\le z}|\,x\,|=\max(0,w,-z)$ and $\max\limits_{w\le x\le z}|\,x\,|=\max(w,-z)$, respectively, the range of $|x_{uk}(\alpha)-x_{vk}(\alpha)|$, for any α in $[0,1]$, can be written as follows:

$$|x_{uk}(\alpha)-x_{vk}(\alpha)|\in \Big[\max\Big(0,\ A^L_{uk}(\alpha)-A^R_{vk}(\alpha),\ A^L_{vk}(\alpha)-A^R_{uk}(\alpha)\Big), \max\Big(A^R_{uk}(\alpha)-A^L_{vk}(\alpha),\ A^R_{vk}(\alpha)-A^L_{uk}(\alpha)\Big)\Big].$$

Now, the identity

$$\max(w,z)=\frac{1}{2}\left(w+z+|\,w-z\,|\right),\ \forall w,z\in\Re,$$

allows us to derive the left and right $\alpha-$bounds of a granular distance, namely the fuzzy distance. We have:

$$d^{\min}_{uv}(\alpha)=\frac{1}{4}\Bigg(\sum_{k=1}^{p}\Big[A^L_{uk}(\alpha)-A^R_{vk}(\alpha)+A^L_{vk}(\alpha)-A^R_{uk}(\alpha)+ +|\,A^L_{uk}(\alpha)-A^R_{vk}(\alpha)-A^L_{vk}(\alpha)+A^R_{uk}(\alpha)\,|+ +\Big|A^L_{uk}(\alpha)-A^R_{vk}(\alpha)+A^L_{vk}(\alpha)-A^R_{uk}(\alpha)+ +|\,A^L_{uk}(\alpha)-A^R_{vk}(\alpha)-A^L_{vk}(\alpha)+A^R_{uk}(\alpha)\,|\Big|\Big]^2\Bigg)^{\frac{1}{2}}$$

$$d^{\max}_{uv}(\alpha)=\frac{1}{2}\Bigg(\sum_{k=1}^{p}\Big[A^R_{uk}(\alpha)-A^L_{vk}(\alpha)+A^R_{vk}(\alpha)-A^L_{uk}(\alpha)+ +|\,A^R_{uk}(\alpha)-A^L_{vk}(\alpha)-A^R_{vk}(\alpha)+A^L_{uk}(\alpha)\,|\Big]^2\Bigg)^{\frac{1}{2}}.$$

Finally, the fuzzy distance between the two p-dimensional fuzzy granules A_u and A_v results in the form of a LR-fuzzy set:

$$d_{uv}(\alpha)=\left(d^{\min}_{uv}(\alpha),\ d^{\max}_{uv}(\alpha)\right)=\left(d^L_{uv}(\alpha),\ d^R_{uv}(\alpha)\right).$$

A point-wise distance can then be obtained by defuzzifying the fuzzy distance d_{uv}, e.g. by computing its centroid. An alternative way may be that of averaging the α-level intervals $[d^L_{uv}(\alpha),\ d^R_{uv}(\alpha)]$ over $[0,1]$:

$$Defuzz(d_{uv})=E\left[d_{uv}(\alpha)\right]=\frac{1}{2}\int_{[0,1]}\left|d^R_{uv}(\alpha)+d^L_{uv}(\alpha)\right|\lambda(d\alpha).$$

where λ is a normalized Lebesgue measure on $[0,1]$, e.g.

$$\lambda([0,1])=\int_0^1 w(\alpha)\,d\alpha=1,\ \text{where}\ w(\alpha)=2\alpha.$$

Remark: The fuzzy distance between two p-dimensional fuzzy granules A_u and A_v with empty intersection, whose components A_{uk} and A_{vk} are trapezoidal fuzzy sets for all $k=1,\ldots,p$, is still a trapezoidal fuzzy set. This is because, in

ISBN: 978-989-95079-6-8

such a case, $d_{uv}^{L}(\alpha)$ is an increasing function that interpolates linearly between $d_{uv}^{L}(0)$ and $d_{uv}^{L}(1)$ whereas $d_{uv}^{R}(\alpha)$ is a decreasing function that interpolates linearly between $d_{uv}^{R}(1)$ and $d_{uv}^{R}(0)$ (see Fig. 1).

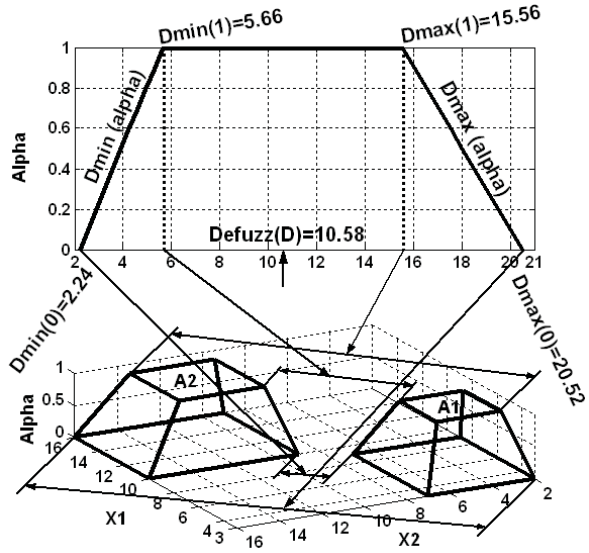


Figure 1: Two-dimensional fuzzy granules, the fuzzy distance between them and its defuzzified value.

The fuzzy distance between pairs of fuzzy granules with non-empty intersection, is linear-shaped, but not necessarily trapezoidal-shaped (see Fig. 2).

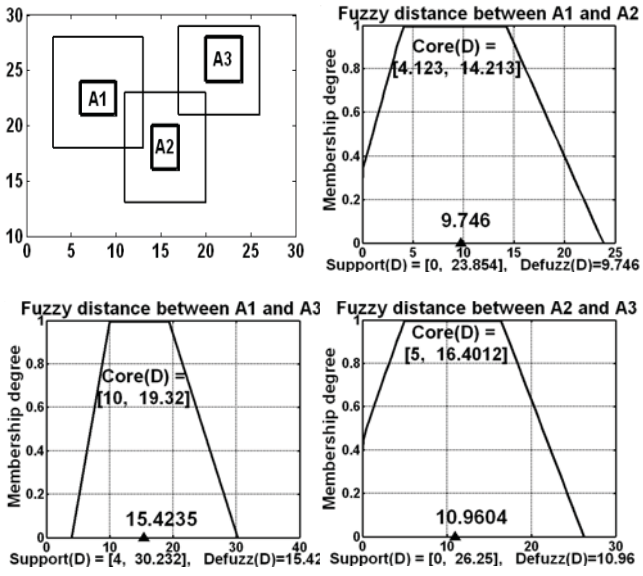


Figure 2: Two-dimensional fuzzy granules with non-empty intersection and the pair-wise fuzzy distances.

Obviously, our formal definition of fuzzy distance introduced above applies similarly to fuzzy granules of any finite dimension p. Fig. 3 illustrates the case of three-dimensional fuzzy granules.

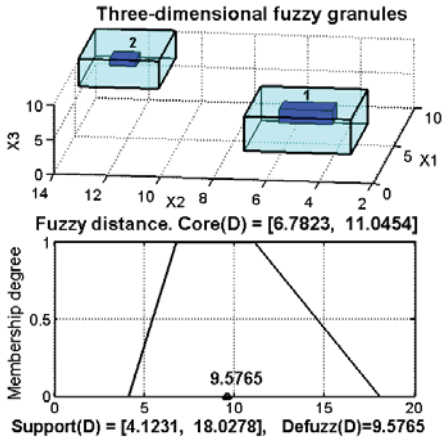


Figure 3: Three-dimensional fuzzy granules, the fuzzy distance and its defuzzified value.

3 Granular Clustering

3.1 A generalization of fuzzy c-means algorithm to granular feature spaces

We propose a generalization of Fuzzy C-Means Algorithm to fuzzy feature spaces, which leads to adopting a well defined distance between p-dimensional fuzzy granules (either a crisp distance, or the defuzzified value of the fuzzy distance defined above).

The Fuzzy C-means clustering algorithm is based on the minimization of an objective function called *C-means functional*:

$$J(X;U,V)=\sum_{i=1}^{c}\sum_{k=1}^{N}(u_{ik})^{m}\left\|x_k-v_i\right\|_A^2$$

where

$$V=(v_1,\ldots,v_c),\qquad v_i\in\Re^n$$

is a vector of *cluster prototypes* (centers), which have to be determined, and

$$D_{ikA}^2=\left\|x_k-v_i\right\|_A^2=(x_k-v_i)'A(x_k-v_i)$$

is a squared inner-product distance norm.

In a granular feature space, such as a space populated with p-dimensional fuzzy granules, x_k are granular samples and v_i are granular centroids.

Remark. The granular centroid of the set $\{A_u\}_{u=1,\ldots,n}$ of LR-type p-dimensional fuzzy granules is still an LR-type p-dimensional fuzzy granule.

Fig. 4 shows a set $\{A_1, A_2, A_3\}$ of 3-dimensional fuzzy granules, their granular centroid G, as well as the three pair-wise granular centroids G_{12}, G_{13} and G_{23}, respectively.

Furthermore, the distance D_{ikA}^2 must be replaced with a well defined square distance between p-dimensional fuzzy granules. Since the fuzzy distance introduced above have been defined as a square distance, we can use the square value of the fuzzy distance in order to replace D_{ikA}^2.

ISBN: 978-989-95079-6-8

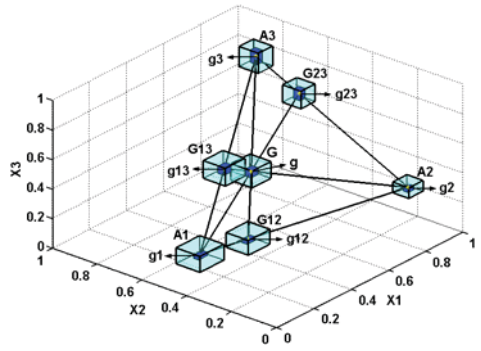


Figure 4. Granular and point-wise weighted centroids of 3D fuzzy granules (the granular version of median theorem).

For testing the granular clustering algorithm we started from a set of crisp data (Fig. 5).

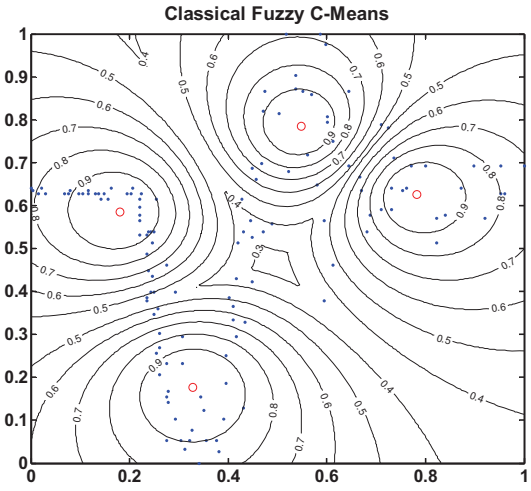


Figure 5: Fuzzy C-means clustering for crisp data.

Then, we transformed these initial data into two-dimensional fuzzy granules, i.e. as a Cartesian product of trapezoidal shaped fuzzy sets on each dimension. Fig. 6 shows 4 fuzzy clusters populated with two-dimensional fuzzy granules

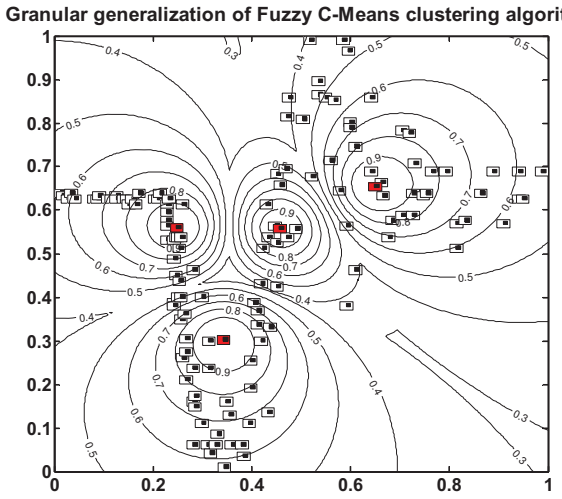


Figure 6: Granular fuzzy C-means clustering.

3.2 *Related work*

In ([4], [5]) we proposed an agglomerative granular clustering method that generalizes the method in [12], originally proposed by Pedrycz and Bargiela. Since resorting to a different granulation framework (fuzzy set theory instead of interval mathematics) involves a refinement of the metric formalism, the fuzzy distance has been used for allowing fuzzy granules (rather than hyperboxes) to be processed during the cluster growing process. The main step of the iterative process is finding the two closest information granules in order to aggregate them into a more comprehensive one. Let $C = agg(A, B)$ be the resulting granule. In terms of p-dimensional LR-fuzzy granules, the aggregation process can be carried out as follows:

$$C = C_1 \times \cdots \times C_p$$

with $C_i^{\alpha} = \left(x_{C_i}^L - \ell_{C_i} \cdot L^{-1}(\alpha),\ x_{C_i}^R + r_{C_i} \cdot R^{-1}(\alpha)\right)$.

The core and the support of C_i is obtained as

$$C_i^1 = \left(x_{C_i}^L, x_{C_i}^R\right) = \left(\min\left(x_{A_i}^L, x_{B_i}^L\right),\ \max\left(x_{A_i}^R, x_{B_i}^R\right)\right)$$

$$C_i^0 = \left(x_{C_i}^L - \ell_{C_i}, x_{C_i}^R + r_{C_i}\right) =$$

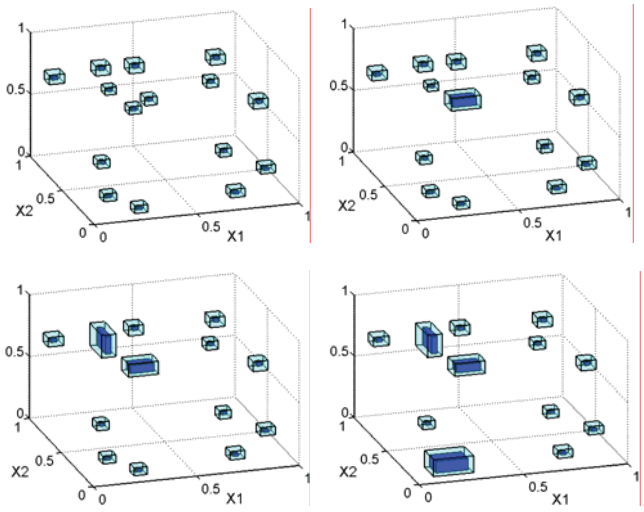
$$= \left(\min(x_{A_i}^L - \ell_{A_i}, x_{B_i}^L - \ell_{B_i}, \max(x_{A_i}^R + r_{A_i}, x_{B_i}^R + r_{B_i})\right).$$

The compatibility measure guiding the search for the two closest fuzzy granules can now be defined as

$$compat(A, B) = 1 - Defuzz\left(d(A, B)\right) \cdot e^{-\beta|C|}$$

where $|C| = \int_0^1 \left|C^{\alpha}\right| d\alpha = \int_0^1 \prod_{i=1}^{p} \left|C_i^{\alpha}\right| d\alpha$ is the cardinality of C and β is a tuning coefficient.

Maximizing the compatibility measure means that the pair of candidate fuzzy granules to be clustered should not only be close enough (i.e., the distance between them should be small), but the resulting granule should be compact (i.e., its expansion along every direction must be well-balanced). Fig. 7 shows the sequence of cluster growing over the granular clustering process.



ISBN: 978-989-95079-6-8

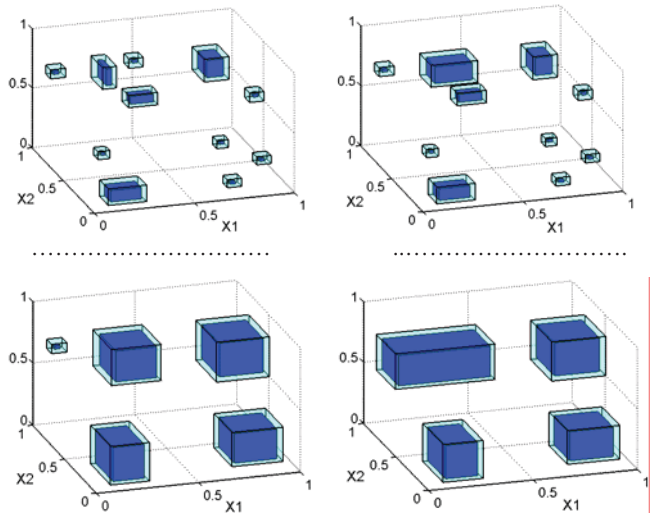


Figure 7: The sequence of cluster growing over the granular clustering process.

4 Time Series Clustering

4.1 Time series (dis)similarity

Roughly speaking, a granule is a set-defined object. In these terms, a subsequence time series can be viewed as a special kind of granule.

Let $S = y_m, \ldots, y_{m+w-1}$ be a subsequence with length w of time series $Y = y_1, \ldots, y_n$, where $1 \le m \le n - w + 1$. Subsequences will be represented as vectors in a w-dimensional vector space.

Clustering of subsequence time series requires the definition of a similarity or of a distance measure. Examples of dissimilarity measures are based on Euclidean norms, piecewise linear approximations, dynamic time warping, longest common subsequences, and probabilistic similarity models.

The usage of the Euclidean distance is subject to the constraint that both time subsequences are of the same length w. Thus we can define the dissimilarity between sequences S_1 and S_2 as $L_p(S_1, S_2)$, that is, the distance between the two $w-$dimensional vectors measured by the L_p norm (when p = 2, this reduces to the familiar Euclidean distance).

Some major disadvantages of the Euclidean distance are as follows: it does not allow for different *baselines* in the time sequences; it is very sensitive to *phase shifts* in time; it does not allow for *acceleration and deceleration* along the time axis.

Other specialized distance measures have been described for time series clustering, such as Dynamic Time Warping, DTW, and Longest Common Subsequence Similarity, LCSS.

Dynamic Time Warping (DTW) is an extensively used technique in speech recognition and allows acceleration–deceleration of signals along the time dimension.

Following [11], let us consider two sequences (of possibly different lengths) $Q = \{q_1, \ldots, q_n\}$ and $C = \{c_1, \ldots, , c_m\}$.

To align two sequences using DTW, we construct an n-by-m matrix where the (i^{th}, j^{th}) element of the matrix contains the distance $d(q_i , c_j)$ between the two points q_i and c_j (i.e. $d(q_i, c_j) = (q_i - c_j)^2$). Each matrix element (i, j) corresponds to the alignment between the points q_i and c_j. A warping path W is a contiguous set of matrix elements that defines a mapping between Q and C. The k^{th} element of W is defined as $w_k = (i, j)_k$, so we have:

$$W = w_1, w_2, \ldots, w_k, \ldots, w_K, \qquad \max(m,n) \le K < m+n-1.$$

The warping path is typically subject to several constraints.

- Boundary conditions: $w_1 = (1, 1)$ and $w_K = (m, n)$.
- Continuity: Given $w_k = (a, b)$, then $w_{k-1} = (a', b')$, where $a - a' \le 1$ and $b - b' \le 1$. This restricts the allowable steps in the warping path to adjacent cells.
- Monotonicity: Given $w_k = (a, b)$, then $w_{k-1} = (a', b')$, where $a - a' \ge 0$ and $b - b' \ge 0$. This forces the points in W to be monotonically spaced in time.

There are exponentially many warping paths that satisfy the above conditions, however we are interested only in the path which minimizes the warping cost:

$$DTW(Q, C) = \min\left(\sqrt{\sum_{i=1}^{K} w_k}\right) \Big/ K .$$

The K in the denominator is used to compensate for the fact that warping paths may have different lengths.

This path can be found efficiently using dynamic programming to evaluate the following recurrence, which defines the cumulative distance $\gamma(i, j)$ as the distance $d(i, j)$ found in the current cell and the minimum of the cumulative distances of the adjacent elements:

$$\gamma(i, j) = d(q_i, c_j) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\}.$$

The Euclidean distance between two sequences can be seen as a special case of DTW where the k^{th} element of W is constrained such that $w_k = (i, j)_k$, $i = j = k$.

The warping path is also constrained in a global sense by limiting how far it may stray from the diagonal. The subset of the matrix that the warping path is allowed to visit is called the warping window. The two most common constraints in the literature are the Sakoe-Chiba band and the Itakura parallelogram. We can view a global or local constraint as constraining the indices of the warping path $w_k = (i, j)_k$, such that $j - r \le i \le j + r$, where r is a term defining the allowed range of warping, for a given point in a sequence. In the case of the Sakoe-Chiba band, r is independent of i; for the Itakura parallelogram, r is a function of i.

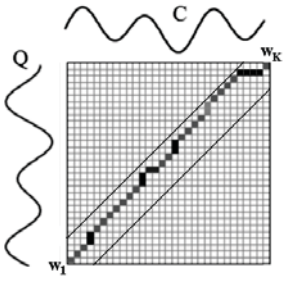


Figure 8: Optimal warping path with the Sakoe-Chiba band as global constraints.

ISBN: 978-989-95079-6-8

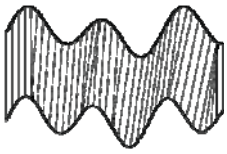


Figure 9: Aligning two time sequences using DTW.

4.2 Non-overlapping subsequence time series clustering

Our implementation of Fuzzy C-Means algorithm for non-overlapping subsequence time series clustering is based essentially on using the DTW distance, which is proved to be largely superior to the Euclidian distance.

We also report a real-world application of this method for clustering non-overlapping subsequences of length 30 from Bucharest Stock Exchange Bet Index. The dataset contains 2210 daily values. Fig. 10 shows ten clusters, which consist of several subsequence time series, grouped around a cluster centroid.

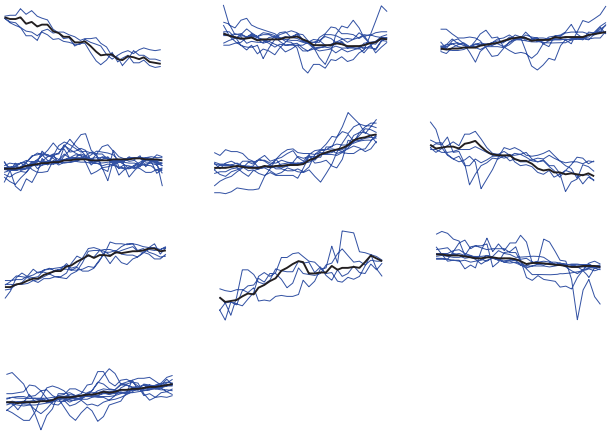


Figure 10: Non-overlapping subsequence time series of length 30: clusters and centroids obtained by fuzzy c-means.

4.3 Overlapping subsequence time series clustering

Subsequence Time Series (STS) clustering typically employs a clustering technique to the subsequences of a time series generated using a sliding window technique (Fig. 11). Although it is very popular, Keogh et al. [9] reported, for the first time, a surprising anomaly: cluster centers obtained using STS clustering closely resemble "sine waves", irrespective of the nature of original time series itself, and therefore, the results of STS clustering are meaningless.

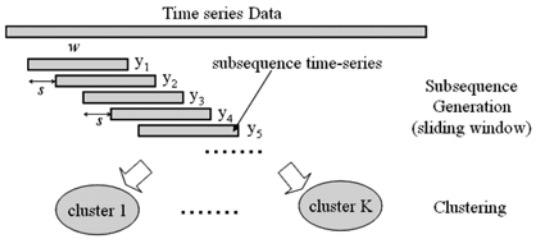


Figure 11: STS clustering using a sliding window technique.

According to the Keogh's criticism, superposition of slightly shifted subsequences causes the generation of sine waves. In order to explicitly exclude superposition, a motif-based clustering algorithm has been proposed in [10]. Another solution to avoid sine wave has been reported in [2], which consists of aligning the phase at the frequency with maximum spectrum power with respect to each time series subsequence, and then, to apply a clustering algorithm to them. Although these approaches are promising, the need of identifying solid mathematical foundations behind STS clustering steel remains.

References

[1] I. Bloch. On fuzzy distances and their use in image processing under imprecision. *Pattern Recognition*, 32:1873–1895, 1999.

[2] R. Fujimaki, S. Hirose and T. Nakata. Theoretical analysis of subsequence time-series clustering from a frequency-analysis viewpoint. *SIAM International Conference on Data Mining*, 506-517, 2008.

[3] V. Georgescu. Granular vs. point-wise metrics and statistics to accommodate machine learning algorithms in granular feature spaces, *Fuzzy Economic Review,* vol.XII, No.2, 45-74, 2007.

[4] V. Georgescu. A Generalization of Symbolic Data Analysis Allowing the Processing of Fuzzy Granules, *Lecture Notes in Artificial Intelligence,* vol. 3131, 215-226, 2004.

[5] V. Georgescu. A Formal Framework Allowing to Deal with Granular Clustering not only in Terms of Hyperboxes, but also in terms of Fuzzy Granules, *Proceedings of MS'2004-France*, Lyon, July 5–7, 2004.

[6] V. Georgescu. On the Foundations of Granular Computing Paradigm, *Fuzzy Economic Review,* vol.VIII, No.2, 73-105, 2003.

[7] V. Georgescu. Multivariate Fuzzy-termed Data Analysis: Issues and Methods, *Fuzzy Economic Review,* vol.VI, No.1, 19-48, 2001.

[8] D. Gunopulos and G. Das. Time series similarity measures and time series indexing. *ACM SIGMOD Conference*, Santa Barbara, CA., 2001.

[9] E. Keogh, J. Lin and W. Truppel. Clustering of time series subsequences is meaningless: implications for previous and future research. *Proceedings of the 3rd IEEE International Conference on Data* Mining, 115–122, 2003.

[10] E. Keogh, B. Chiu and S. Lonardi. Probabilistic discovery of time series motifs. *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.

[11] E. Keogh and C. A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7: 358-386, 2005.

[12] W. Pedrycz and A. Bargiela Granular Clustering: a Granular Signature of Data, *IEEE Transactions on Systems, Man and Cybernetic.,* vol. 32, No.2, 212-224, 2002.

[13] W. Pedrycz. *Knowledge-Based Clustering. From Data to Information Granules*, John Wiley & Sons, 2005.

[14] W. Pedrycz. (ed.). *Granular Computing: An Emerging Paradigm*, Physica-Verlag, 2001.

[15] L. A. Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, 90: 111-117, 1997.

ISBN: 978-989-95079-6-8

Enhancing OLAP Querying with the aid of H-IFS

Panagiotis Chountas[1] Ermir Rogova[1] Krassimir T. Atanassov[2]

[1]DKMG – SECS, University of Westminster, London, HA1 3TP, UK
[2]CLBME, Bulgarian Academy of Sciences, Bl. 105, Sofia-1113, BULGARIA
Email: chountp@wmin.ac.uk, krat@argo.bas.bg

Abstract— *Over the past years we have witnessed an increasing interest in expressing user or domain preferences or knowledge inside database queries. First, it appeared to be desirable property of a query system to offer more expressive query languages that can be more faithful to what a user intends to say. Second, a classical query in the sense of relational paradigm may also have a restricted answer or sometimes an empty set of answers, while a relaxed version of the query enhanced with background or domain knowledge might be matched by some items in the database.*

Keywords— Query answering, OLAP, IFS, multidimensional modelling.

1 Introduction

The need for flexible systems to manage value uncertainty has been the focus for database researchers [1], mainly at theoretical level and in the context of the relational model.

At the same time OLAP technology required [2] the extension of the relational systems with the inclusion of the data-cube and operators to operate over it. Alternatively, new models [3] were proposed to support OLAP based querying on top of multidimensional views. Both approaches support the organisation of data around several axes of analysis. In OLAP based systems, when it comes to the model level, support for value uncertainty will be required at the fact level as well at the level of dimensions with the support of non-rigid hierarchies [4]. Still [4] considers that facts and dimensions as in [5], [6] represent structural information [7].

Current research issues for OLAP systems can be summarised as follows

- Flexible models are required to support value uncertainty at fact level as well as at the dimension level with the provision of non rigid dimensions
- Flexibility should not be limited at the structural level. It should be allowed also at the query level. Users should be allowed to synthesise their own model of dimensions for analysis purposes based on existing structure. Dimensions may be based in either rigid or non-rigid hierarchies.

Recently in OLAP systems a need has been identified for enhancing the query scope with the aid of "kind-of" relation that describe knowledge as well as ordering of the elements of a domain or a hierarchical universe.

However, in our context, the terms of the hierarchy [8], and the relations between terms are not fuzzy and do not represent [9] only kind-of relations. These observations led us to introduce the concept of closure of the H-IFS which is a developed form defined on the whole hierarchy.

The rest of the paper is organised as follows; In Section two we define the basic properties of Intuitionistic Fuzzy sets and H-IFS. In Section 3 the semantics of the H-IFS cubic representation [14], [15] in contrast to the basic multidimensional-cubic structures are presented we also define the extended SQL-OLAP aggregators. In Section 4 we present the main concepts involved in the designing and implementation the 'IF-Oracle' OLAP utility and also demonstrate the potential of 'IF-Oracle' utility. Finally we conclude and compare our proposal with other school of thoughts.

2 Intuitionistic Fuzzy Sets A-IFS

2.1 Foundations

Each element of an Intuitionistic fuzzy [10, 11] set has degrees of membership or truth (μ) and non-membership or falsity (ν), which don't sum up to 1.0 thus leaving a degree of hesitation margin (π).

As opposed to the classical definition of a fuzzy set given by

A′= $\widetilde{A} = \{< x, \mu_A(x) > | x \in X\}$ where $\mu_A(x) \in [0,1]$ is the membership function of the fuzzy set A′, an Intuitionistic fuzzy set A is given by

$$A = \{< x, \mu_A(x), \nu_A(x) > | x \in X\}$$
$$\mu_A : X \to [0,1] \text{ and } \nu_A : X \to [0,1]$$

Such that

$$0 \le \mu_A(x) + \nu_A(x) \le 1$$

and $\mu_A : X \to [0,1], \nu_A : X \to [0,1]$ denote a degree of membership and a degree of non-membership of X∈A, respectively. Obviously, each fuzzy set may be represented by the following Intuitionistic fuzzy set;

$$A = \{< x, \mu_A(x), 1 - \mu_A(x) > | x \in X\}$$

For each Intuitionistic fuzzy set in X, we will call $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$, Intuitionistic fuzzy index of $x \in A$ which expresses a lack of knowledge of whether x belongs to A or not. For each $x \in A$ $0 \le \pi_A(x) \le 1$.

Definition 1. Let A and B be two Intuitionistic fuzzy sets defined on a domain X. A is included in B (denoted A ⊆ B) if and only if their membership functions and non-membership functions satisfy the condition:

ISBN: 978-989-95079-6-8

$(\forall X \in X)\ (\mu_A(x) \leq \mu_B(x)\ \&\ \nu_A(x) \geq \nu_B(x))$

Two scalar measures are classically used in classical fuzzy pattern matching to evaluate the compatibility between an ill-known datum and a flexible query, known as

- a possibility degree of matching, **Π(Q/ D)**
- a necessity degree of matching, **N(Q/ D)**

Definition 2. Let Q and D be two Intuitionistic fuzzy sets defined on a domain X and representing, respectively, a flexible query and an ill-known datum.

The possibility degree of matching between Q and D, denoted Π(Q/D), is an "optimistic" degree of overlapping that measures the maximum compatibility between Q and D, and is defined by:

$$\Pi(Q/D) = \left\langle \sup_{x\in X} \min(1-\nu_Q(x), \nu_Q(x)), \inf_{x\in X} \max(1-\nu_D(x), \nu_D(x)) \right\rangle$$

The necessity degree of matching between Q and D, denoted N(Q/D), is a "pessimistic" degree of inclusion that estimates the extent to which it is certain that D is compatible with Q, and is defined by:

$$N(Q/D) = \left\langle \inf_{x\in X} \max(\mu_Q(x), 1-\mu_Q(x)), \sup_{x\in X} \min(\mu_D(x), 1-\mu_D(x)) \right\rangle$$

The problem occurring from defining Intuitionistic fuzzy sets based on the "kind-of" relation is that two different Intuitionistic fuzzy sets on the same hierarchy do not necessarily have the same definition domain, which means they cannot be compared using the classic comparison operations **Π**(Q/ D), N(Q/ D).

2.2 From IFS to H-IFS

The definition domains of the hierarchical fuzzy sets [12, 13] that we propose below are subsets of hierarchies composed of elements partially ordered by the "kind of" relation. An element l_i is more general than an element l_j (denoted $l_i \sim l_j$), if l_i is a predecessor of l_j in the partial order induced by the "kind-of" relation of the hierarchy. An example of such a hierarchy is given in 'Figure 1'.

Definition 3. Let F be a H-IFS defined on a subset D of the elements of a hierarchy L. It degree is denoted as <μ, ν>. The closure of F, denoted clos(F), is a H-IFS defined on the whole set of elements of L and its degree $<\mu, \nu>_{clos(F)}$ is defined as follows.

For each element **l** of L, let $S_L = \{l_1, \ldots, l_n\}$ be the set of the smallest super-elements in D.

If S_L is not empty,

$$<\mu, \nu>_{clos(F)}(S_L) = <\max_{1\leq i\leq n} \mu(L_i), \min_{1\leq i\leq n} \nu(L_i)>$$

else

$$<\mu, \nu>_{clos(F)}(S_L) = <0,0>$$

In other words, the closure of a H-IFS F is built according to the following rules. For each element l_1 of L:

- If l_I belongs to F, then l_I keeps the same degree in the closure of F (case where $S_L = \{l_I\}$).
- If l_I has a unique smallest super-element l_I in F, then the degree associated with l_I is propagated to L in the closure of F, $S_L = \{l_I\}$ with $l_I > l_I$)

If L has several smallest super-elements $\{l_1, \ldots, l_n\}$ in F, with different degrees, a choice has to be made concerning the degree that will be associated with l_I in the closure. The proposition put forward in definition 3, consists of choosing the maximum degree of validity μ and minimum degree of non validity v associated with $\{l_1, \ldots, l_n\}$. We refer to as the *Optimistic strategy*.

The proposition put forward in definition 3, consists of choosing the maximum degree of validity μ and minimum degree of non validity v associated with $\{l_1, \ldots, l_n\}$. We refer to as the *Optimistic strategy*.

We can also utilise a *Pessimistic strategy* which consists of choosing the minimum degree of validity μ and maximum degree of non validity v associated with $\{l_1, \ldots, l_n\}$.Alternatively, an *Average strategy* could be utilised, which consists of calculating the IF-Average and applying it to the degrees of validity μ and non-validity v.

It has been observed that two different H-IFSs, defined on the same hierarchy, can have the same closure, as in the following example.

The H-IFSs Q={Wine<1,0>, Red Wine<0.7,0.1>, Brown Wine<1,0>, White Wine <0.4,0.3>} and R={Wine<1,0>, Red Wine<0.7,0.1>, Brown Wine<1,0>, Pinot Noir <1.0,0.0>} have the same closure, represented 'Figure 1' below.

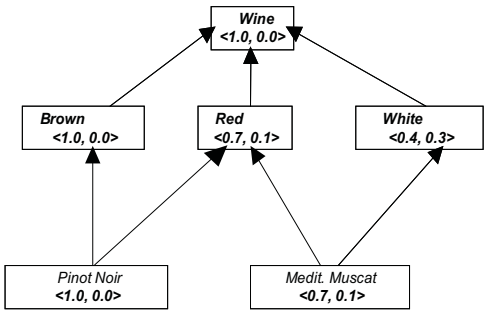


Figure 1: Common closure of the H-IFS's Q and R

Such H-IFS's form equivalence classes with respect to their closures.

Definition 4. Let F be a hierarchical fuzzy set, with dom(F) = $\{l_1, \ldots, l_n\}$, and F_{-k} the H-IFS resulting from the restriction of F to the domain dom(F) \ $\{l_k\}$. l_k is deducible in F if

$$<\mu, \nu>clos_{(F-k)}(l_k) = <\mu, \nu>clos_{(F)}(l_k)$$

As a first intuition, it can be said that removing a derivable element from a hierarchical fuzzy set allows one to eliminate redundant information. But, an element being derivable in F does not necessarily mean that removing it from F will have no consequence on the closure: removing k from F will not impact the degree associated with k itself in the closure, but it may impact the degrees of the sub-elements of k in the closure.

For instance, if the element Brown Wine is derivable in Q, according to *definition 4*, removing Brown Wine <1,0> from Q would not modify the degree of Brown Wine itself in the resulting closure, but it could modify the degree of its

ISBN: 978-989-95079-6-8

sub-element Pinot Noir. Thus, Brown Wine <1,0> cannot be derived or removed. This remark leads us to the following definition of a minimal hierarchical fuzzy set.

Definition 5. In a given equivalence class (that is, for a given closure C), a hierarchical fuzzy set is said to be **minimal** if its closure is C and if none of the elements of its domain is derivable**.**

Obtaining the Minimal H-IFS

Step 1: Assign Min-H-IFS ← ∅. Establish an order so that the sub-elements $\{l_1,\ldots,l_n\}$ of the hierarchy L are examined after its super-elements.

Step 2: Let l_1 be the first element & $(l_1)/\langle\mu, \nu\rangle \neq (l_1)/\langle 0, 0\rangle$ then add l_1 to Min-H-IFS and

$$\langle\mu, \nu\rangle_{clos(Min\text{-}HIFS)}\ (l_1)= (l_1)/\langle\mu, \nu\rangle.$$

Step 3: Let us assume that K elements of the hierarchy L satisfy the condition $\langle\mu, \nu\rangle_{clos(Min\text{-}HIFS)}\ (l_i)=(l_i)/\langle\mu, \nu\rangle$.
In this case the Min-H-IFS do not change. Otherwise go to next element l_{k+1} and execute Step 4.

Step 4: The $l_{k+1}/\langle\mu_{k+1}, \nu_{k+1}\rangle$ associated with l_{k+1}.
In this case l_{k+1} is added to Min-H-IFS with the corresponding $\langle\mu_{k+1}, \nu_{k+1}\rangle$.

Step 5: Repeat steps three and four until $clos_{(Min\text{-}HIFS)}=C$.

For instance S_1 and S_2 are **minimal** (none of their elements is derivable). They cannot be reduced further.

S_1= Wine<1,0>

S_2={Wine<1,0>, Red Wine<0.7,0.1>, Pinot Noir<1,0>, White Wine <0.4, 0.3>}

3 Representing H-IFS Inside Cubes

In this section we present the semantics of the H-IFS cubic representation [14], [15] in contrast to the basic multidimensional-cubic structures. The basic cubic operators are extended and enhanced with the aid of Intuitionistic Fuzzy Logic.

Since the emergence of the OLAP technology [16] different proposals have been made to give support to different types of data and application purposes. One of these is to extend the relational model (ROLAP) to support the structures and operations typical of OLAP. Further approaches [17], are based on extended relational systems to represent data-cubes and operate over them.

Nowadays, information and knowledge-based systems need to manage imprecision in the data and more flexible structures are needed to represent the analysis domain. New models have appeared to manage incomplete datacube [18], imprecision in the facts and the definition of fact using different levels in the dimensions.

Nevertheless, these models continue to use inflexible hierarchies thus making it difficult to merge reconcilable data from different sources with some incompatibilities in their schemata. These incompatibilities arise due to different perceptions-views about a particular modelling reality.

In addressing the problem of representing flexible hierarchies we propose a new multidimensional model that is able to treat with imprecision over conceptual hierarchies based on Intuitionistic Fuzzy logic. The use of conceptual hierarchies enables us to:

- define the structures of a dimension in a more perceptive way to the final user, thus allowing a more perceptive use of the system.
- query information from different sources or even use information or preferences given by experts to improve the description of hierarchies, thereby getting more knowledgeable query results. We outline a unique way for incorporating "H-IFS" relations, or conceptual imprecise hierarchies as dimensions with respect to the model proposed in [19].

3.1 Semantics of the IF-Cube in contrast to Crisp Cube

In this section we review the semantics of Multidimensional modelling and Intuitionistic Fuzzy Logic and based on these we propose a unique concept named as Intuitionistic Fuzzy Cube (IF-Cube). The IF-Cube is the basis for the representation of flexible hierarchies and thus flexible facts.

3.1.1 Overview of the Cube Model

A logical model that influences the database design and the query engines is the *multidimensional-cubic* view of data in the warehouse. In a multidimensional data model, there is a set of *numeric measures* that are the objects of analysis. Examples of such measures are sales, budget, etc. Each of the numeric measures depends on a set of *dimensions,* which provide the context for the measure. The attributes of a dimension may be related via a hierarchy of relationships. In the above example, the product name is related to its category and the industry attribute through a hierarchical relationship, see Fig. 2

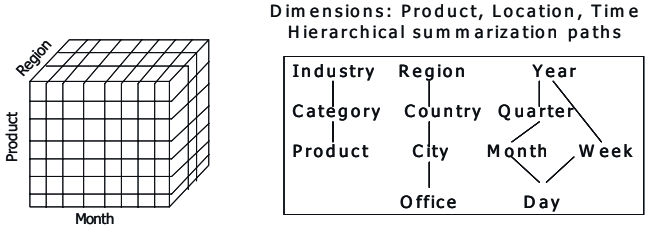


Figure 2: Cube_Sales, Rigid Hierarchies for Product, Location, Time Dimensions

According to [19] a cube structure is defined as a 4-tuple, <D, M, A, f> where the four components indicate the characteristics of the cube. These characteristics are:
a set of n dimensions $D = \{d_1, d_2, \ldots, d_n\}$ where each d_i is a dimension name, extracted from a domain $dom_{dim(i)}$. A set of k measures $M = \{m_1, m_2, \ldots, m_k\}$ where each m_i is a measure name, extracted from a domain $dom_{measure(i)}$. The set of dimension names and measures names are disjoint; i.e., $D \cap M = 0$. A set of t attributes $A = \{a_1, a_2, \ldots, a_t\}$ where each a_i is an attribute name, extracted from a domain $dom_{attr(i)}$. A one-to-many mapping $f : D \rightarrow A$, i.e. there exists, corresponding to each dimension, a set of attributes.

3.1.2 Semantics of the IF-Cube

ISBN: 978-989-95079-6-8

In contrast, an **IF-Cube** is an abstract structure that serves as the foundation for the multidimensional data cube model. Cube C is defined as a five-tuple *(D, l, F, O, H)* where:

- *D* is a set of dimensions
- *l* is a set of levels $l_1, ..., l_n$,
- A dimension $d_i = (l \leq O, l_\perp, l_\top)$ $dom(d_i)$ where
- $l=l_i\ i=1...n.$ l_i is a set of values and $l_i \cap l_j = \{\}$
- $\leq O$ is a partial order between the elements of *l.*
- To identify the level *l* of a dimension, as part of a hierarchy we note it as *dl.* $l_\perp$: base level $l_\top$: top levelfor each pair of levels l_i and l_j we have the relation :

$\mu_{ij} : l_i \times l_j \rightarrow [0,1]$ $\quad v_{ij} : l_i \times l_j \rightarrow [0,1]$ $0 < \mu_{ij} + v_{ij} < 1$

- *F* is a set of fact instances with schema $F = \{<x, \mu_F(x), v_F(x)> | x \in X\}$, where $x=<att_1, ...,att_n>$ is an ordered tuple belonging to a given universe *X,* $\mu_F(x)$ *and* $v_F(x)$ are the degree of membership and non-membership of *x* in the fact table *F* respectively.
- *H* is an object type history that corresponds to a cubic structure*(l, F, O, H')* which allows us to trace back the evolution of a cubic structure after performing a set of operators i.e. aggregation.

The example below provides a sample imprecise cube (*D, l, F, O, H) i.e. sales* and a conceptual non-rigid hierarchy product with reference to wine consisting of $l_i, ..., l_n$ levels with respective levels of membership and non membership < μ_{ij} v_{ij}, >.

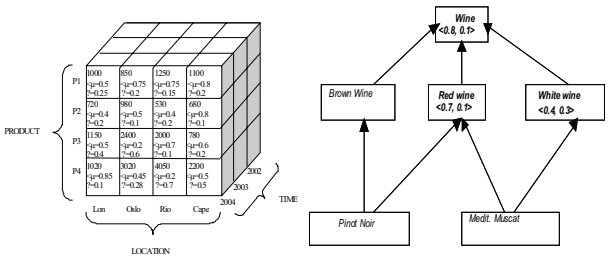


Figure 3: *Imprecise Cube Sales, IF Hierarchy 'Wine*

3.1.3 Cube Aggregation

Aggregation (A): An aggregation operator *A* is a function *A(G)* where $G = \{<x, \mu_F(x), v_F(x)> | x \in X\}$ where $x=<att_1, ...,att_n>$ is an ordered tuple belonging to a given universe.

X, $\{att_1, ..., att_n\}$ is the set of attributes of the elements of *X,* $\mu_F(x)$ *and* $v_F(x)$ are the degree of membership and non-membership of *x.*

The result is a bag of the type $\{<x', \mu_F(x'), v_F(x')> | x' \in X\}$. To this extent, the bag is a group of elements that can be duplicated and each one has a degree of μ and ν.

Input: C_i = *(D, l, F, O, H)* and the function *A(G)*

Output: C_o = (D, l_o, F_o, O_o, H_o)

The definition of the extended group operators allows us to define the extended group operators ***Drill up (Δ), and Drill Down (Ω).***

Drill up (Δ): The result of applying Drill up over dimension d_i at level dl_r using the aggregation operator A over a datacube $C_i = (D_i, l_i, F_i, O, H_i)$ is another datacube $C_o = (D_o, l_o, F_o, O, H_o)$.

Input: $C_i = (D_i, l_i, F_i, O, H_i)$

Output: $C_o = (D_o, l_o, F_o, O, H_o)$

An object of type history is a recursive structure. The structured history of the datacube allows us to keep all the information when applying *Drill up* and get it all back when *Drill Down* is performed. To be able to apply the operation of *Drill up* we need to make use of the IF_{SUM} aggregation operator.

Drill Down (Ω): This operator performs the opposite function of the *Drill up* operator. It is used to Drill Down from the higher levels of the hierarchy with a greater degree of generalization, to the leaves with the greater degree of precision. The result of applying *Drill Down* over a datacube C_i = *(D, l, F, O, H)* having H=*(l', D', A', H')* is another datacube C_o= *(D', l', F', O, H').*

Input: C_i=*(D, l, F, O, H)*

Output C_o=*(D',l',F',O,H')* where *F' → set of fact instances defined by operator A.*

The defined IF OLAP Cube model allows us to:

- accommodate imprecise facts
- utilise *conceptual hierarchies* used for aggregation purposes in the cases of roll-up and Drill Down operations.
- offer a unique feature such as keeping track of the history when we move between different levels of a hierarchical order.

In the next section we demonstrate the usefulness of the H-IFS notion and the extended aggregation operators for extending the query capabilities of Oracle10g. We developed an ad-hoc utility 'IF-Oracle' implemented on top of Oracle10g that allow us to:

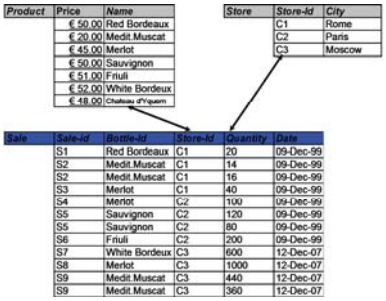
- Define an H-IFS hierarchy
- Incorporate hierarchical knowledge in the form of H-IFS as part of the standard OLAP queries.
- Enhance the scope of query answers against the Oracle10g standard query answers.

4 Embedding IF Cubes in Oracle10g

IF-Oracle has been developed using Visual Studio.Net as an ad-hoc utility that is attached to and enhances Oracle10g DBMS query capabilities.

For demonstrating the functionality of IF-Oracle let us consider a sample multidimensional model, Fig.4 in the form of a star schema that describes sales of Vitis Vinifera type wines.

ISBN: 978-989-95079-6-8



Product	Price	Name
	€ 50,00	Red Bordeaux
	€ 20,00	Medit.Muscat
	€ 45,00	Merlot
	€ 50,00	Sauvignon
	€ 51,00	Friuli
	€ 52,00	White Bordeux
	€ 48,00	Chateau d'Yquem

Store	Store-Id	City
	C1	Rome
	C2	Paris
	C3	Moscow

Sale	Sale-Id	Bottle-Id	Store-Id	Quantity	Date
	S1	Red Bordeaux	C1	20	09-Dec-99
	S2	Medit.Muscat	C1	14	09-Dec-99
	S2	Medit.Muscat	C1	16	09-Dec-99
	S3	Merlot	C1	40	09-Dec-99
	S4	Merlot	C2	100	09-Dec-99
	S5	Sauvignon	C2	120	09-Dec-99
	S5	Sauvignon	C2	80	09-Dec-99
	S6	Friuli	C2	200	09-Dec-99
	S7	White Bordeux	C3	600	12-Dec-07
	S8	Merlot	C3	1000	12-Dec-07
	S9	Medit.Muscat	C3	440	12-Dec-07
	S9	Medit.Muscat	C3	360	12-Dec-07

Figure 4**:** Sample of a Star Schema

After forming the structure and storing it as a concept relation in Oracle10g, we perform the calculation of the hierarchical closure of the H-IFS and its weights.
The user now has the choice of selecting three different strategies: ***Optimistic, Pessimistic*** or ***Average*** as defined on section 2.2.
Let's assume that the user's interest lays on finding information about Red wines. Fig.5 below shows the hierarchy after weights have been estimated and assigned reflecting the user's intent.

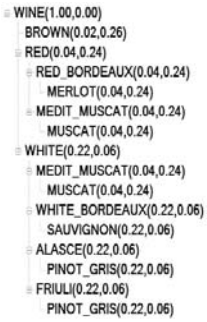


Figure 5: Vitis Vinifera sub-hierarchy view with weights

We can observe that the principle of the H-IFS closure *(see definition 3)* has been preserved when propagating the degree of validity μ and non-validity ν from super-elements to sub-elements by using the optimistic strategy.
The degree of validity and non-validity <μ, ν > are calculated as follows:

$$\mu = \frac{|c_l|}{|c_{l-1}|} \qquad \nu = \frac{|\neg c_l|}{|c_{l-1}|}$$

Where c_l corresponds to those elements from the fact table that absolutely satisfy the selection criteria with reference to a node in the hierarchy. C_{l-1} represents the elements children elements of that selection on a lower level that satisfy the selection condition to some extent. It is obvious that $\pi = 1-(\mu+\nu)$.
After adding the hierarchy into the repository and automatically calculating the weights for the requested nodes, the user can utilize the ad-hoc interface for execution of queries either in standard SQL or make use of the enhanced Select clause and features that IF-Oracle provides.

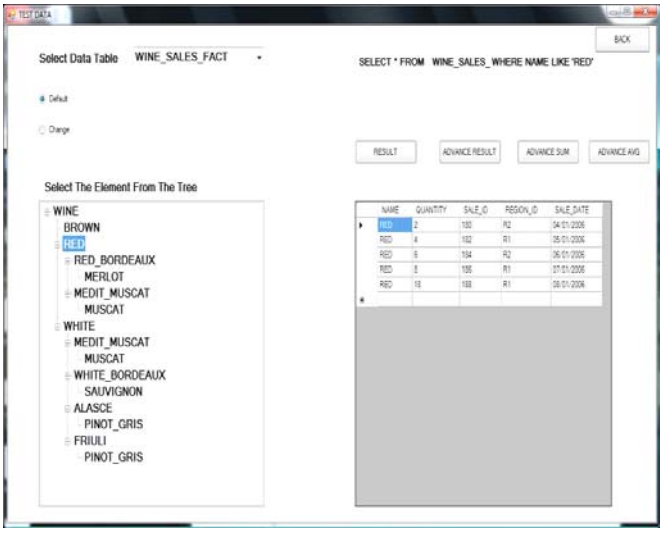


Figure 6**:** Standard SQL output for "Red" wine

Fig. 6 shows the results of a user request for "Red" wine executed in standard SQL provided by Oracle10g.
In contrast, Fig. 7 shows the output after executing the same query, but this time using the IF-Oracle utility.

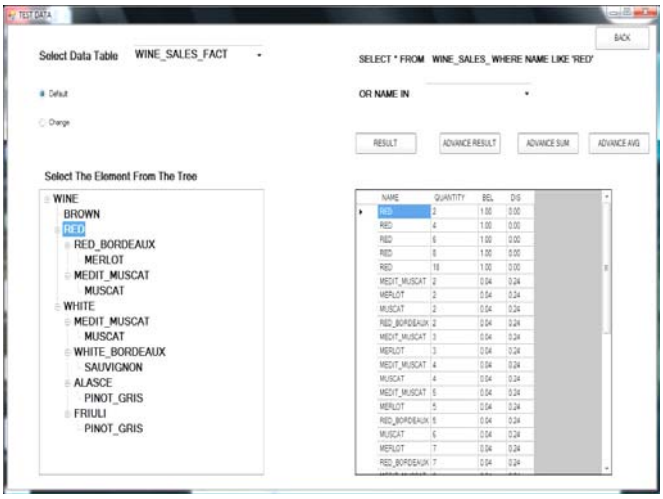


Figure 7: Enhanced SQL output for "Red" wine

By comparing the two figures, one can observe that IF-Oracle produces a knowledge-based answer instead of mindlessly matching the records against the word "Red".
The results show that IF-Oracle not only retrieves sales of "Red" bottles, but also sales of bottles that are classified as red wines by the knowledge represented in the H-IFS hierarchy as "Merlot", "Red Bordeaux", "Medit. Muscat", etc. with indicative degrees of <μ, ν> relevant to the user's preference.
At this point one may decide to further enhance the query capabilities of the IF-Oracle utility by allowing versions of hierarchies. In such case similarities [20] and dissimilarities [21] between different versions should be reflected in the query results, or in the common acceptable DWH schema.

ISBN: 978-989-95079-6-8

5 Conclusions & Comparisons

In this paper, we focus on integrating hierarchical users preferences or intent in OLAP queries with the aim on enhancing the OLAP scope and in return to get richer answer, closer to user requests. We provide a means of using background knowledge to re-engineer query processing and answering with the aid of H-IFS and Intuitionistic Fuzzy relational representation.

The hierarchical links defined on the basis of the H-IFS closure are representing hierarchical knowledge in different forms. A cubic OLAP model presented that allows dimensions to be defined as H-IFS.

In comparing our framework with the bipolar querying school of thought as presented in [22], [23], and [24] the reader may take into account the following:

- The bipolar querying school of thought is the main vehicle for allowing users to express their preferences as part of a query formulation. Allowing thus queries to come up with more knowledgeable answers.

However our framework is resolving a different problem, related specifically to OLAP in the following ways:

- It allows users to import knowledge from external sources in order to redefine the axis of analysis and eventually the cubic structure for different business scenarios.
- The IF-Oracle utility can operate using either cubic structures-MOLAP or relations-ROLAP.
- The IF-Oracle utility assumes that stored data are precise. However it can also cope with imprecise data. The importance here is not on the goodness of data, is mainly on how well the stored data do fit in different scenarios or changing dimensions.

Future research efforts will concentrate on incorporating knowledge arriving from external sources either semi structured or unstructured, considering the web as such source.

References

[1] S. Abiteboul, R. Agrawal, P. Bernstein, M. Carey, S. Ceri, B. Croft, D. Dewitt, M. Franklin, H. Garcia Molina, D. Gawlick, J.Gray, L. Haas, A. Halevy, J. Hellerstein, Y. Ioannidis, M. Kersten, M. Pazzani, M. Lesk, D. Maier, J. Naughton, H. Schek, T. Sellis, A. Silberschatz, M. Stonebraker, R. Snodgrass, J. Ullman, G. Weikum, J. Widom, S. Zdonik, "The Lowell Database Research Self Assessment", Communications of the ACM, 2005, Volume 48, Number 5, pp 111-118.

[2] T.B. Pedersen, C.S. Jensen, Multidimensional "Data Modeling for Complex Data", In Proc. of 15th ICDE, IEEE Computer Society, 1999, pp.336-345

[3] J. Gray, A. Bosworth, A. Layman, H. Pirahesh, "Data cube: a relational aggregation operator generalizing group-by, cross-tabs and subtotals", Journal of Data Mining and Knowledge Discovery, Springer Verlag, Vol.1, No1, 1997, pp 29--53.

[4] M. Delgado, C. Molina, D. Sanchez, A. Vila, L. Rodriguez-Ariza, "A fuzzy multidimensional model for supporting imprecision in OLAP", In Proc. of FUZZ-IEEE, 2004, vol.3, pp 1331- 1336

[5] P. Chamoni, S. Stock, "Temporal Structures in Datawarehousing", In Proc. of the DaWaK'99, ACM Portal, 1999 pp. 353 - 358

[6] T.B. Pedersen, C.S. Jensen, C.E. Dyreson, "A foundation for capturing and querying complex multidimensional data", In Information Systems, ACM Portal, Vol. 26, No. 5, 2001, pp. ages: 383 - 423

[7] A.O. Mendelzon, A. Vaisman, "Temporal Queries in OLAP", In Proc. of the 26th VLDB'00, 2000. ACM Portal, pp. 242 -253

[8] E. Rogova, P. Chountas, K. Atanassov, "Flexible hierarchies and fuzzy knowledge-based OLAP". – FSKD 2007, IEEE Computer Society Press Vol.2, pp. 7-11

[9] P. Chountas, "On Intuitionistic fuzzy sets over universes with hierarchical structures" Notes on Intuitionistic Fuzzy Sets, Vol. 13, No 1, pp 52-56

[10] K. Attanasov, "Intuitionistic Fuzzy Sets: Theory and Applications", Springer-Verlag, 1999

[11] K. Atanassov , "Remarks on the Intuitionistic fuzzy sets" – Fuzzy Sets and Systems, Vol. 51, 1992, No 1, 117-118

[12] B. Kolev, P. Chountas, E. Rogova, K. Atanassov "Representation of Value Im-perfection with the Aid of Background Knowledge: H-IFS" Studies in Computational Intelligence , Springer Verlag, Vol.109, pp 473-494.

[13] P. Chountas, K Atanassov, E. Rogova, "H-IFS: Modelling & Querying over Hierarchical Universes", International Conference On Information Processing And Management of Uncertainty-IPMU'2008, ISBN 978-84-612-3061-7, 2008, pp 1628-1634.

[14] E. Rogova, P. Chountas, "On Imprecision Intuitionistic Fuzzy Sets & OLAP – The Case for KNOLAP", IFSA'07, Springer-Verlag GmBh , Theoretical advances and application of fuzzy logic and soft computing, 2007, pp. 11-20

[15] R. Kimball, "The Data Warehouse Toolkit". New York: John Wiley & Sons, 1996.

[16] S. Chaudhuri, U. Dayal, V.Ganti, "Database Technology for Decision Support Systems". In: Computer, Vol. 34, p. 48-55, 2001

[17] M. Jarke et al. "Fundamentals of data warehouses" Springer, London, 2002

[18] C. Dyreson, "Information retrieval from an incomplete data cube", VLDB, Morgan Kaufman Publishers, 1996, pp. 532-543.

[19] H. Thomas & A. Datta, "A Conceptual Model and Algebra for On-Line Analytical Processing in Decision Support Databases", Information Systems Research 12: 2001, pp.83-102,

[20] E. Szmidt, J. Kacprzyk "Distances between Intuitionistic fuzzy sets, Fuzzy Sets and Systems", Vol. 114, No.3, 2000, pp505–518.

[21] E. Szmidt, J. Kacprzyk "A new concept of similarity measure for Intuitionistic fuzzy sets and its use in group decision making" LNAI 3558, 2005, pp272–282

[22] H. Prade, D. Dubois. "Handling bipolar queries in fuzzy information processing". In: Fuzzy information processing in databases. José Galindo (Eds.), Vol. 1, Information Science Reference, Hershey, PA. , 2008, pp. 97-114.

[23] S. Zadrozny "Bipolar queries revisited" In: V. Torra, et al. (Eds.): Modeling Decisions for Artificial Intelligence, LNCS, Springer Verlag Vol. 3558, 2005 pp. 387-398.

[24] S. Zadrozny J. Kacprzyk "Bipolar queries and queries with preference" Proc. 17th International Conference DEXA'06, 2006 pp. 415-419.

ISBN: 978-989-95079-6-8

Modeling Investor Optimism with Fuzzy Connectives

Milan Lovric, Rui J. Almeida, Uzay Kaymak[1] Jaap Spronk[2]

[1] Erasmus School of Economics, Erasmus University Rotterdam
Rotterdam, The Netherlands
[2] Rotterdam School of Management, Erasmus University Rotterdam
Rotterdam, The Netherlands
Email: {lovric,almeida,kaymak}@ese.eur.nl, jspronk@rsm.nl

Abstract— *Optimism or pessimism of investors is one of the important characteristics that determine the investment behavior in financial markets. In this paper, we propose a model of investor optimism based on a fuzzy connective. The advantage of the proposed approach is that the influence of different levels of optimism can be studied by varying a single parameter. We implement our model in an artificial financial market based on the LLS model. We find that more optimistic investors create more pronounced booms and crashes in the market, when compared to the unbiased efficient market believers of the original model. In the case of extreme optimism, the optimistic investors end up dominating the market, while in the case of extreme pessimism, the market reduces to the benchmark model of rational informed investors.*

Keywords— artificial financial market, agent-based modeling, behavioral finance, fuzzy aggregation, investor sentiment, optimism index.

1 Introduction

Artificial financial markets are models for studying the link between individual investor behavior and financial market dynamics. They are often computational models of financial markets, and are usually comprised of a number of heterogeneous and boundedly rational agents, which interact through some trading mechanism, while possibly learning and evolving. These models are built for the purpose of studying agents' behavior, price discovery mechanisms, the influence of market microstructure or the reproduction of the stylized facts of real-world financial time-series (e.g. fat tails of return distributions and volatility clustering).

A similar bottom-up approach has been utilized in agent-based computational economics (ACE) - the computational study of economies modeled as evolving systems of autonomous interacting agents [22]. A methodology analogous to agent-based modeling also comes from the physical sciences as the Macroscopic Simulation - a tool for studying complex systems by simulating many interacting microscopic elements [16]. A number of reviews of studies with artificial financial markets are available, e.g. [15] and [10].

Since agent-based models can easily accommodate complex learning behavior, asymmetric information, heterogeneous preferences, and ad hoc heuristics [4], such simulations are particularly suitable to test and generate various behavioral hypothesis. The idea of individual investors who are prone to biases in judgment, who are frame-dependent and use various heuristics, which might lead to anomalies on the market level, has been explored within the field of behavioral finance. Behavioral finance is the branch of finance which applies knowledge from psychology and sociology to discover and explain phenomena inconsistent with the paradigm of the expected utility of wealth and narrowly defined rational behavior [7]. A number of surveys and books on behavioral finance and behavioral economics topics can be found, for example [20], [8], and [2]. This complementarity of behavioral finance research and the agent-based methodology has been recognized in the literature (e.g. [15], [5]). Some of the early studies that pursue the idea of explicit accounting for behavioral theories in agent-based financial market simulations are [21] and [9]. In [21] the focus is on overconfidence and loss aversion, while [9] study the social interaction between investors.

One of the key characteristics that govern investor behavior is the optimism or pessimism of the investors. The link between asset valuation and investor sentiment has been the subject of considerable debate in the finance, and has been studied in the context of mispricing (departures from the fundamentals) [3], the limits of arbitrage [17], as well as the underreaction and overreaction of stock prices [1]. Two methodological approaches can be found in the finance literature. One is concerned with finding adequate proxies for the aggregate investor sentiment, and using them in statistical analysis to explain the variation of stock prices and the occurrences of mispricing, such as bubbles and crashes. The other one is a bottom-up approach that aims at modeling individual investor optimism and pessimism by using the insights from psychological theories. For these theories, it is important to have a flexible framework that can be adapted to capture the complexity of human decision making behavior.

In fuzzy decision theory, a wide range of connectives (aggregation operators) has been proposed and studied in order to model the flexibility of human decision making. In this sense, the use of fuzzy connectives for modeling elements of behavioral finance is promising, since the wide range of behaviors documented in the behavioral finance literature necessitates the use of a flexible framework for aggregating information. In this paper, we make a step in this direction by proposing a model of investor optimism based on fuzzy aggregation.

In probabilistic decision theory, such as the Prospect Theory [12, 23] and Rank-Dependent Utility Theory [19], optimism and pessimism are modeled using the probability weighting function. If, for example, the decision under risk is considered, a decision problem is presented using risky prospects, i.e. a set of possible outcomes and their probabilities. Because of the probability weighting, the decision weights asso-

ISBN: 978-989-95079-6-8

ciated with the outcomes are not equal to their probabilities (as would be in the case of Expected Value Theory or Expected Utility Theory). To model optimism we would need to specify and parameterize such a probability weighting function that gives more decision weight to good outcomes and less decision weight to bad outcomes. However, an empirically observed probability weighting function is usually S-shaped, which means that when dealing with such prospects, people are at the same time optimistic about the best outcomes, pessimistic about the worst outcomes, and insensitive to middle outcomes [25].

A decision maker's optimism or pessimism has also been studied within a fuzzy decision making setting. Various fuzzy connectives studied in this context have parameters that denote explicitly the optimism or pessimism degree of a decision maker. Apart from the well-known Hurwicz operator [11], the grade of compensation in Zimmermann–Zysno operator [26] can also be interpreted as an index of optimism. All these operators view the decision as a mixture of conjunctive and disjunctive behavior, and the degree of optimism determines which aggregation type dominates and to which degree. Another optimism–pessimism index was proposed in [24], where the parameter of the generalized averaging operator [6] is interpreted as the decision maker's characteristic degree of optimism. This is an intuitive way of modeling degree of optimism, since optimism is now modeled as the disposition of the decision maker to believe or give importance to positive events compared to his/her disposition to consider negative events [14]. An application of this operator in the risk management of power networks has been considered in [13].

In this paper, we propose a model of investor optimism based on the generalized averaging operator. The advantage of the proposed approach is that the influence of different levels of optimism can be studied by varying a single parameter. We study the effects of investor optimism in an artificial financial market based on the Levy, Levy, Solomon (LLS) model [16]. In previous publications, investor psychological biases such as overconfidence have also been studied by using this model [18].

The outline of the paper is as follows. Section 2 explains the basics of the LLS model in which we study the investor optimism and pessimism. Section 3 describes the setup of the experiments we have conducted. Section 4 presents the results of the simulations. Section 5 concludes the paper and discusses possible extensions for the future research.

2 Model Description

The proposed model of investor optimism is based on the LLS microscopic simulation model [16] with a small homogeneous subpopulation of efficient market believers (EMBs) as described in [16]. LLS model is a well-known and early econophysics model, rooted in a utility maximization framework. Variants of the model have been published in a number of articles and a book, and the model has also been critically evaluated in [27].

2.1 Asset Classes

As in the original LLS model, there are two investments alternatives: a risky stock (or market index) and a risk-free asset (bond). This is in line with many of the agent-based artificial financial markets, which typically do not deal with portfolio selection in multi-asset environments. The risky asset pays at the beginning of each period a dividend which follows a multiplicative random walk according to

$$\tilde{D}_{t+1} = D_t(1+\tilde{z}), \tag{1}$$

where $\tilde{z}$ is a random variable distributed uniformly in the interval $[z_1, z_2]$. The bond pays interest with a rate of r_f.

2.2 Agent Behavior

Many early agent-based artificial financial markets were based on a small number of relatively simple strategies. Such markets have been labeled as *few-type models* [15]. Typically, strategies (or agents who employ them) could be divided into two groups: *fundamental* (based on a perceived fundamental value) and *technical* (based on the past prices, e.g. some form of trend extrapolation). *Zero-intelligence* framework in which agents trade randomly, might be useful for to studying the influence of market microstructure, and sometimes a small number of such agents is included into a few type model to provide liquidity for other agents.

LLS model follows a standard framework where preferences (and risk attitude) are captured by an agent's utility function, and the objective is the maximization of expected utility. But even in such a framework there are many possibilities for the functional form of the utility, which differ in descriptive validity and analytical tractability. When empirical support is taken into account, most evidence suggests DARA (Decreasing Absolute Risk Aversion) and CRRA (Constant Relative Risk Aversion), which motivates the choice of power (myopic) utility function in [16]

$$U(W) = \frac{W^{1-\alpha}}{1-\alpha}. \tag{2}$$

LLS model contains two types of investors: (1) Rational Informed Investors (RII) and (2) Efficient Market Believers (EMB).

2.2.1 RII investors

RII investors know the dividend process, and therefore can estimate fundamental value as the discounted stream of future dividend, according to the Gordon model

$$P^f_{t+1} = \frac{D_t(1+\tilde{z})(1+g)}{k-g}, \tag{3}$$

where k is the discount factor of the expected rate of return demanded by the market for the stock, and g is the expected growth rate of the dividend. RII investors assume that the price will converge to the fundamental value in the next period. In each period RII investor i chooses the proportion of wealth to invest in stocks and bonds so that he or she maximizes the expected utility of wealth in the next period, given by the following equation from [16]:

$$\begin{aligned} EU(\tilde{W}^i_{t+1}) &= \frac{(W^i_h)^{1-\alpha}}{(1-\alpha)(2-\alpha)} \frac{1}{(z_2-z_1)} \left(\frac{k-g}{k+1}\right) \frac{P_h}{xD_t} \\ &\times \left\{ \left[(1-x)(1+r_f) + \frac{x}{P_h}\left(\frac{k+1}{k-g}\right) D_t(1+z_2)\right]^{(2-\alpha)} \right. \\ &\left. - \left[(1-x)(1+r_f) + \frac{x}{P_h}\left(\frac{k+1}{k-g}\right) D_t(1+z_1)\right]^{(2-\alpha)} \right\}. \end{aligned} \tag{4}$$

ISBN: 978-989-95079-6-8

Based on the optimal proportion, they determine the number of stocks demanded by multiplying this optimal proportion with their wealth. Since all RII investors are assumed to have the same degree or risk aversion (parameter α), they will all have the same optimal proportion x. The actual number of demanded shares might differ only if investors differ in their wealth. However, as in the experiments of [16] we assume that they all start with the same initial wealth.

2.2.2 EMB investors

EMB investors believe that the price accurately reflects the fundamental value. However, since they do not know the dividend process, they use *ex post* distribution of stock returns to estimate the *ex ante* distribution. EMB investor i uses a rolling window of size m^i, and is in the original model [16] said to be *unbiased* if, in absence of additional information, he or she assigns the same probability to each of the past m^i return observations [16]. Hence, the original, unbiased EMBs assume that returns come from a discrete uniform distribution

$$\Pr^i(\tilde{R}_{t+1} = R_{t-j}) = \frac{1}{m^i}, \text{ for } j = 1, ..., m^i. \tag{5}$$

The expected utility of EMB investor i is given by [16]

$$EU(\tilde{W}^i_{t+1}) = \frac{(W^i_h)^{1-\alpha}}{(1-\alpha)} \sum_{j=1}^{m^i} \Pr^i(\tilde{R}_{t+1} = R_{t-j}) \times \left[(1-x)(1+r_f) + xR_{t-j}\right]^{(1-\alpha)}. \tag{6}$$

In accordance with the LLS model [16], for all EMB investors an investor specific noise is added to the optimal investment proportion x^* (that maximizes the expected utility) in order to account for various departures from rational optimal behavior ($\tilde{\varepsilon}^i$ is truncated so that $0 \leq x^i \leq 1$, imposing the constraint of no borrowing and no short-selling), i.e.

$$x^i = x^{*i} + \tilde{\varepsilon}^i. \tag{7}$$

2.2.3 Sentiment EMBs

In this paper we create a new EMB type, called the *sentiment EMBs* by using a fuzzy set connective. *Sentiment EMBs* use generalized aggregation operator to estimate future returns, using the rolling window of size m^i. The prediction of the next period return for each investor i is given by

$$\tilde{R}_{t+1} = \left(\frac{1}{m^i} \sum_{j=1}^{m^i} (R_{t-j})^s \right)^{1/s}. \tag{8}$$

The higher the parameter s, the higher the estimate of the return (more closer to the maximum value from the sample), and vice versa. In such a way, we use parameter s to capture the phenomena of investor optimism and pessimism.

In our experiments we consider a several special cases of the generalized mean:

- $s \to -\infty$, the minimum of the sample;
- $s = -1$, the harmonic mean;
- $s \to 0$, the geometric mean;
- $s = 1$, the arithmetic mean;
- $s = 2$, the quadratic mean;
- $s \to \infty$, the maximum of the sample.

Since there is only one value for the expected return, instead of a probability distribution, the expected utility of sentiment EMB investor i is given by

$$EU(\tilde{W}^i_{t+1}) = \frac{(W^i_h)^{1-\alpha}}{(1-\alpha)} \left[(1-x)(1+r_f) + x\tilde{R}_{t+1}\right]^{(1-\alpha)}. \tag{9}$$

The investors will maximize this expected utility if in each period they invest all their wealth either in the stock or in the bond, depending on the actual comparison between the expected return on the stock $\tilde{R}_{t+1}$ and the return on the riskless bond $(1 + r_f)$.

2.3 Market Mechanism

LeBaron in [15] describes four types of market mechanisms used in agent-based artificial financial markets. In this paper, as in the original LLS model, we use clearing by *temporary market equilibrium*. RII and EMB investors determine optimal proportion in the stock so as to maximize the expected utility of their wealth in the next period. However, expected utility is the function of the future price, which is in the current period unknown. Investors therefore need to determine optimal proportions, and respective demands for shares, for various hypothetical prices. The equilibrium price P_t is set to that hypothetical price for which the total demand of all investors in the market equals the total number of outstanding shares, according to

$$\sum_i N^i_h(P_t) = \sum_i \frac{x^i_h(P_t) W^i_h(P_t)}{P_t} = N. \tag{10}$$

Table 1: Parametrization

Symbol	Value	Explanation
M	950	Number of RII investors
M_2	50	Number of EMB investors
m	10	Memory length of EMB investors
α	1.5	Risk aversion parameter
N	10000	Number of shares
r_f	0.01	Riskless interest rate
k	0.04	Required rate of return on stock
z_1	-0.07	Maximal one-period dividend decrease
z_2	0.10	Maximal one-period dividend growth
g	0.015	Average dividend growth rate

3 Experiments with investor optimism

In the benchmark model where only RII investors are present in the market, there is no trade, the log prices follow random walk, and there is no excess volatility of the market price [16]. In the experiment with a small fraction of homogeneous (with respect to memory length) and unbiased EMB investors (of the original model), the market dynamics show semi-predictable (unrealistic) booms and crashes, with substantial trading in the

ISBN: 978-989-95079-6-8

market and excess volatility [16]. This experimental setup of [16] is also the basis for the experiments in this paper.

In our new model we conduct six experiments for six different levels of optimism of EMB investors, that correspond to the special cases of the parameter s. In each experiment the market consists of 95% RII investors and 5% EMB investors, with the parametrization given in Table 1. We run 100 independent 1000-period-long simulations, with different initial seeds of the random number generators. The results in the Table 2 are averaged over these 100 simulations.

Table 2: Results

	$s=-\infty$	$s=-1$	$s=0$
$\sigma(P)$	6.0249	12.8370	17.8668
$\sigma(P^f)$	5.7159	5.7159	5.7159
excess volatility %	5.41	124.59	212.58
mean volume p.p. %	0.48	9.04	6.40
	$s=1$	$s=2$	$s=\infty$
$\sigma(P)$	27.4739	28.8751	25.0327
$\sigma(P^f)$	5.7159	5.7159	5.7159
excess volatility %	380.66	405.18	337.95
mean volume p.p. %	2.82	1.18	0.12

4 Results

Fig. 1 shows a typical price dynamics from the first experiment with pessimistic EMB investors. The market price closely follows the fundamental price which is driven by the random dividend process. Hence, this experiment resembles the benchmark model in which there are only RII investors in the market. Pessimistic investors predict next period return with the minimum return in the sample of past returns. The minimum return is almost always bellow the risk-less return, so the optimal investment for pessimistic EMB investors is to invest everything in bond. The actual investment proportion will slightly vary due to the error term in (7). Only in rare occasions when there is a series of returns higher than the risk-less return, the EMB investors will invest in the risky asset. The results in Table 2 show that for this experiment the volatility of the market price is similar to the volatility of the fundamental price, which means that there is a low excess volatility. The relative mean volume per period shows that there is very little trading in the market, i.e. from period to period the investors do not change much their portfolio holdings.

Fig. 2 shows the price development for the second experiment with slightly more optimistic investors that predict future return using the harmonic mean. The results of this experiment qualitatively and quantitatively resemble the results of the original model with a small fraction of unbiased EMB investors (which predict future returns using a uniform discrete distribution over the observed returns). The market exhibits cyclical booms and crashes to the fundamental value. According to Table 2, the market is more volatile, and there is also more trading. This exchange of risky assets between RII and EMB investors occurs mostly when the booms begin and when they crash.

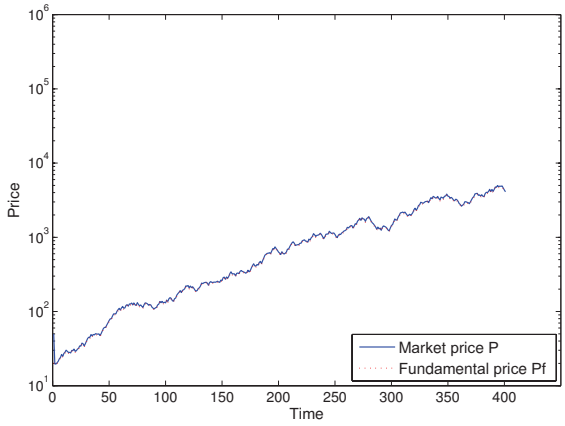


Figure 1: Price dynamics with 95% RII and 5% minimum sentiment EMB.

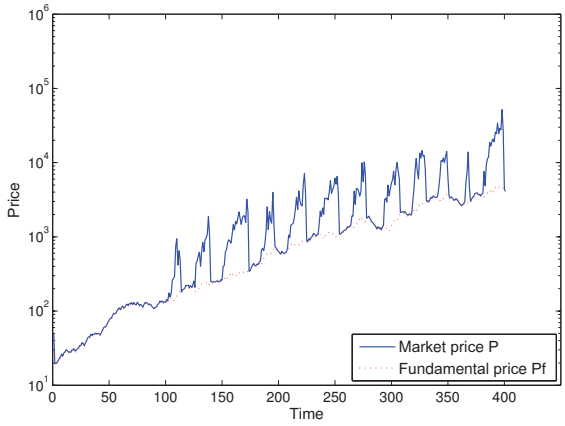


Figure 2: Price dynamics with 95% RII and 5% harmonic sentiment EMB.

Fig. 3, Fig. 4, and Fig. 5 depict the market dynamics when EMB investors are even more optimistic. As the index of optimism increases the market shows more extreme (longer lasting) booms, followed by very sharp crashes. During these bubbles the EMB investors aggressively invest in the risky asset, while the RII investor divest expecting that the overvalued asset would fall to its fundamental value. The crash occurs when there is a series of low returns, due to low dividend realizations, so the EMB investors suddenly shift toward a risk-less asset. However, as soon as a better return is realized, EMB investors invest in the risky asset and a new boom starts. From Table 2 it is also evident that the more optimistic EMB investors are, the more volatile market price is. However, the trading is reduced because the booms are longer lasting, i.e. the cycles of booms and crashes appear less frequently.

In the case of full optimism, there is an ongoing market bubble, as shown in Fig. 6. The market does not crash because the maximum return in the rolling window of past returns is always above the risk-less return, so the EMB investors are always highly invested in the risky asset. The trading in this experiment is even more reduced, but the volatility of the market price is also somewhat reduced. The reason for the latter is that the crash does not occur within the experiments.

ISBN: 978-989-95079-6-8

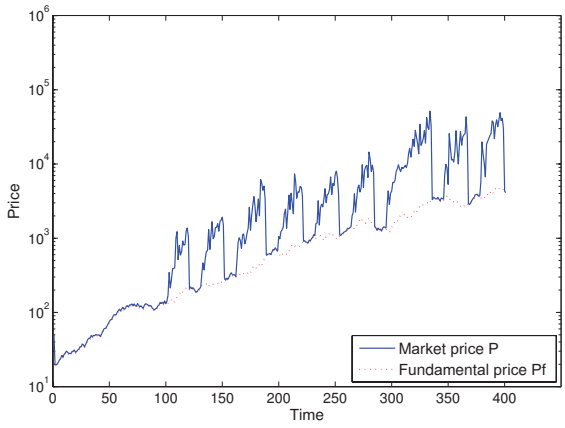


Figure 3: Price dynamics with 95% RII and 5% geometric sentiment EMB.

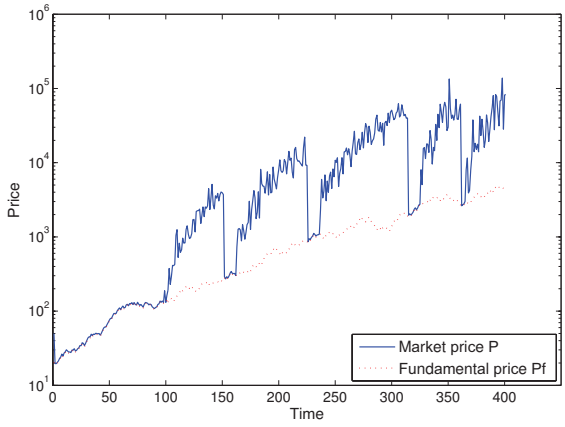


Figure 4: Price dynamics with 95% RII and 5% arithmetic sentiment EMB.

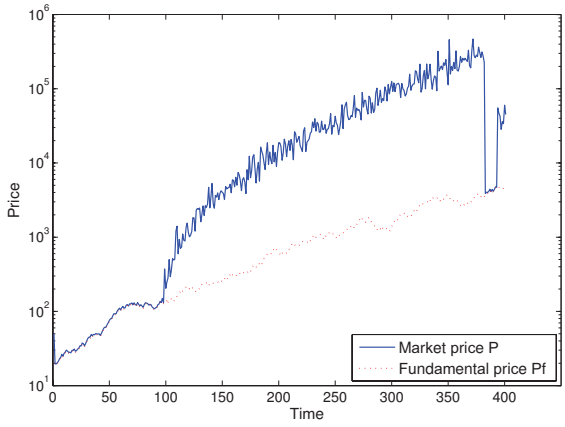


Figure 5: Price dynamics with 95% RII and 5% quadratic sentiment EMB.

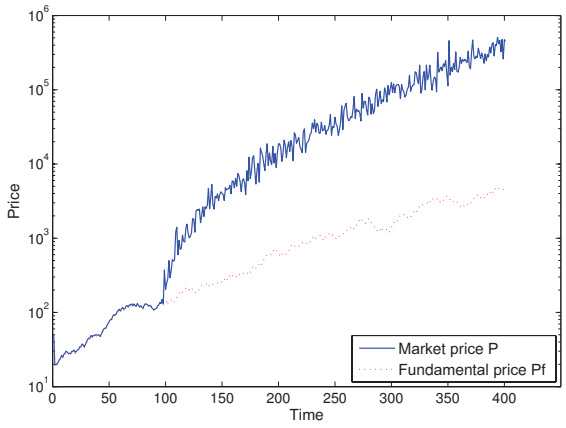


Figure 6: Price dynamics with 95% RII and 5% maximum sentiment EMB.

Fig. 7 shows the development of the relative wealth of RII investors over time. At the beginning, RII investors posses 95% of all the wealth in the market. In the case of extremely pessimistic EMB investors, RII investors end up asymptotically dominating the market. This is because the LLS market is a growing market, and only RII investors are investing in the risky asset and exploiting that growth. Conversely, in the case of extreme optimism, EMB investors are highly invested in the stock, and eventually dominate the market. In non-extreme cases of optimism, both types of investor coexist in the market.

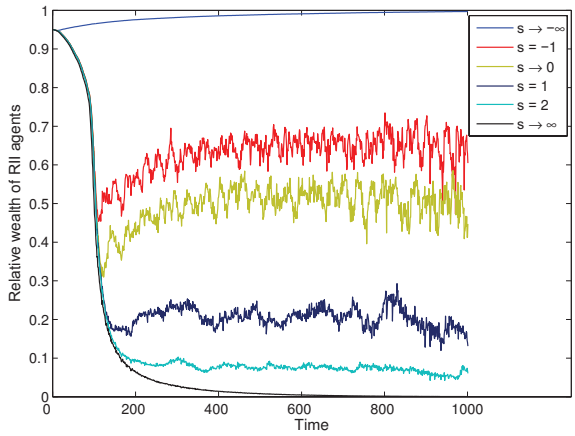


Figure 7: Relative wealth dynamics of RII against sentiment EMBs with various levels of optimism.

5 Discussion and conclusions

In this paper we have used a fuzzy connective to study investor optimism in the modified LLS model of the stock market [16]. We show how changes in the formation of expectations by EMB investors can have a marked impact on the price dynamics. The levels of investor optimism are related to the occurrences of market booms and crashes, as well as measures of excess volatility and trading volume.

ISBN: 978-989-95079-6-8

Since our current experiments focus only on the case of homogeneous EMB investors, we would like to conduct our next experiments in the case of heterogeneous EMB investors with various memory lengths. We expect that the choice of memory length has a great impact on the occurrence of booms and crashes, particularly in extreme cases of optimism, because the larger that window of past returns is, the less likely is that all the returns are bellow or above the risk-less return. In future research, we will extend our analysis to the interplay of investors with various degrees of optimism within the same market. Furthermore, we would like to implement an updating mechanism by which the level of investor optimism changes based on the past performance.

As we have used the same model to study investor overconfidence [18], a distinct although related behavioral phenomenon, it would also be interesting to study both phenomena at the same time. The overconfidence in the model [18] refers to the peakedness of the return distribution around the mean of return observations, while optimism in this model determines how that mean is chosen (ranging from the minimum observation to the maximum observation in the sample of past returns).

This paper demonstrates the advantage of using a fuzzy connective for modeling investor optimism, as we were able to control investor optimism by varying only a single parameter. The results of our experiments show that this parameter was a valid choice for an index of optimism in the context of financial markets. In future research other fuzzy set connectives could be investigated for agent decision making.

References

[1] N. Barberis, A. Shleifer, and R. Vishny. A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343, 1998.

[2] N. C. Barberis and R. H. Thaler. A survey of behavioral finance. In G. M. Constantinides, M. Harris, and R. M. Stulz, editors, *Handbook of the Economics of Finance*, volume 1, pages 1053–1128. Elsevier, 1 edition, 2003.

[3] G. W. Brown and M. T. Cliff. Investor sentiment and asset valuation. *Journal of Business*, 78(2):405–440, 2005.

[4] N. T. Chan, B. LeBaron, and T. Poggio. Agent-based models of financial markets: A comparison with experimental markets. MIT Artificial Markets Project, 1999.

[5] S.-H. Chen and C.-C. Liao. Behavioral finance and agent-based computational finance: toward an integrating framework. In *Proceedings of 7th Joint Conference on Information Sciences*, pages 1235–1238, North Carolina, USA, September 2003.

[6] H. Dyckhoff and W. Pedrycz. Generalized means as model of compensative connectives. *Fuzzy Sets and Systems*, 14(2):143–154, 1984.

[7] G. M. Frankfurter and E. G. McGoun. Market efficiency or behavioral finance: The nature of the debate. *The Journal of Psychology and Financial Markets*, 1(4):200–210, 2000.

[8] J. Goldberg and R. von Nitzsch. *Behavioral Finance*. John Wiley and Sons Ltd, 2001.

[9] A. O. I. Hoffmann, W. Jager, and E. J. H. V. Eije. Social simulation of stock markets: Taking it to the next level. *Journal of Artificial Societies and Social Simulation*, 10(2), 2007.

[10] C. H. Hommes. Heterogeneous agent models in economics and finance. In L. Tesfatsion and K. L. Judd, editors, *Handbook of Computational Economics*, volume 2, chapter 23, pages 1109–1186. Elsevier, 1 edition, 2006.

[11] L. Hurwicz. A class of criteria for decision-making under ignorance. Cowles Commission Discussion Paper Statistics No. 356, Cowles Foundation for Research in Economics, Yale University, Feb. 1951.

[12] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

[13] U. Kaymak, R. Babuska, H. R. van Nauta Lemke, and G. Honderd. A fuzzy logic system for steady-state security analysis of power networks. *Journal of Intelligent and Fuzzy Systems*, 6(3):345–359, 1998.

[14] U. Kaymak and H. R. van Nauta Lemke. A sensitivity analysis approach to introducting weight factors into fuzzy multicriteria decision making. *Fuzzy Sets and Systems*, 97(2):169–182, 1998.

[15] B. LeBaron. Agent-based computational finance. In L. Tesfatsion and K. L. Judd, editors, *Handbook of Computational Economics*, volume 2, chapter 24, pages 1187–1233. Elsevier, 1 edition, 2006.

[16] M. Levy, H. Levy, and S. Solomon. *Microscopic Simulation of Financial Markets: From Investor Behavior to Market Phenomena*. Academic Press, 2000.

[17] J. B. D. Long, A. Shleifer, L. H. Summers, and R. J. Waldmann. Noise trader risk in financial markets. *Journal of Political Economy*, 98(4):703–738, 1990.

[18] M. Lovric, U. Kaymak, and J. Spronk. Overconfident investors in the LLS agent-based artificial financial market. In *Proceedings of the IEEE Symposium on Computational Intelligence for Financial Engineering (CIFEr)*, pages 58–65, Nashville, USA, March–April 2009.

[19] J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3(4):323–343, 1982.

[20] H. Shefrin. *Beyond Greed and Fear - Understanding Behavioral Finance and the Psychology of Investing*. Harvard Business School Press, 2000.

[21] H. Takahashi and T. Terano. Agent-based approach to investor's behavior and asset price fluctuation in financial markets. *Journal of Artificial Societies and Social Simulation*, 6(3), 2003.

[22] L. Tesfatsion and K. Judd. Agent-based computational economics: A constructive approach to economic theory. In L. Tesfatsion and K. L. Judd, editors, *Handbook of Computational Economics*, volume 2, chapter 16, pages 831–880. Elsevier, 1 edition, 2006.

[23] A. Tversky and D. Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

[24] H. R. van Nauta Lemke, J. G. Dijkman, H. van Haeringen, and M. Pleeging. A characteristic optimism factor in fuzzy decision-making. In *Proceedings IFAC Symposium on Fuzzy Information, Knowledge Representation and Decision Analysis*, pages 283–288, Marseille, France, July 1983.

[25] H. U. Weber. From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin*, 115(2):228–242, 1994.

[26] H.-J. Zimmermann and P. Zysno. Latent connectives in human decision making. *Fuzzy Sets and Systems*, 4:37–51, 1980.

[27] E. Zschischang and T. Lux. Some new results on the Levy, Levy and Solomon microscopic stock market model. *Physica A*, 291:563–573, 2001.

ISBN: 978-989-95079-6-8

Tail Point Density Estimation Using Probabilistic Fuzzy Systems

R. J. Almeida U. Kaymak

Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
Email: {rjalmeida,kaymak}@ese.eur.nl

Abstract— *Value at Risk (VaR) is a popular measure for quantifying the market risk that a financial institution faces into a single number. Due to the complexity of financial markets, the risks associated with a portfolio may vary over time. For accurate VaR estimation, it is necessary to have flexible methods that adapt to the underlying data distribution. In this paper, we consider VaR estimation by using probabilistic fuzzy systems (PFS). Contrary to previous publications, our focus is on the modeling of the tail points of the distribution of returns. We study two approaches to designing probabilistic fuzzy VaR models that take into account the extreme values of the data and compare their performance with the performance of a GARCH model. It is found that the VaR estimation process is simplified and improved by our proposed method.*

Keywords— Density estimation, extreme values, fuzzy histograms, probabilistic fuzzy systems, value-at-risk.

1 Introduction

Due to the volatile nature of the financial markets risk management is an important activity for financial institutions that operate in these markets. As a result of risk management, activities are undertaken to reduce the possibility of failure to an acceptable range. These activities may include portfolio adjustment, hedging or insurance [1, 2]. Nowadays, the financial sector operates under strict guidelines, which have been imposed through international agreements, partly due to various financial failures that have happened in 1990's. For example, due to the Basel Agreement, the financial institutes must have well documented procedures to manage the market risk, the credit risk and the operational risk that they are exposed to [3].

Managing risk is strongly dependent on the information available. When the amount of information grows beyond a specific level, there is a need for a concise representation of the risk a company or institution is facing. Due to the complex nature of financial markets in which many parties exchange information and interact through trading, the overall risk for a company is influenced by many internal and external factors. It is customary for management to classify different types of risk and develop models for dealing with each type of risk in order to keep the risk management problem tractable. One of the different types of risk that a financial institution has to deal with is the market risk, which is the exposure to the uncertain market value of a portfolio [4]. Value-at-risk (VaR) is a way to quantify the market risk. It is a single number for the senior management to express and summarise the total market risk of a portfolio with financial assets. Value at Risk measures the worst expected loss over a given horizon under normal market conditions at a given confidence level. Due to regulations, large banks must nowadays base their market risk capital requirements on the VaR estimate [3]. This drives the continued research into newer and better VaR models.

Different types of approaches have been proposed for VaR estimation, such as simulation and parametric approaches. In the parametric approaches, the risk is quantified in terms of volatility, which is expressed as the standard deviation σ of the portfolio. The simplest models of volatility assume that it does not vary over time, while more advanced models acknowledge that volatility varies dynamically over time. The dynamic aspect of volatility could be modelled in various ways. A popular model where volatility changes dynamically in time is the GARCH (Generalised Auto Regressive Heteroscedasticity) model [5]. For the GARCH $(1, 1)$ model, which is used quite often in practice, the variance is estimated using a first-order autoregressive model of the squared returns.

The disadvantage of the parametric approach is that, due to the complexity of financial markets, the data usually do not follow the parametric distributions that are assumed to underly the data generating process. For example, the returns are typically non-Gaussian, they have fat tails and volatility clustering is often observed in financial markets see *e.g.* [6, 7]. Therefore, flexible modelling approaches such as non-parametric modelling or semi-parametric modelling are needed in which the models can adapt themselves into the underlying actual data distribution. In this context, neural network models for VaR estimation have been studied by various researchers [8, 9, 10] as well as fuzzy set models [11, 12]. Probabilistic fuzzy systems [13] were also used to model VaR by estimating the whole density function, by following the distribution of the data and distributing more membership functions around the origin.

In this paper we focus on tail point density estimation using probabilistic fuzzy systems, to correctly estimate VaR. For this purpose we consider two approaches for estimating the conditional parameters of the PFS model, by giving more relevance to the extreme values, and comparing their performance in obtaining value-at-risk models. A Mamdami-type probabilistic fuzzy system [14] is used for this purpose. The location of the antecedent membership functions is determined by using fuzzy clustering and maximum likelihood parameter estimation is used for determining the probability parameters of the PFS. The validity of the VaR models is evaluated by using a statistical back-testing method based on failure rates.

The outline of the paper is as follows. In Section 2 we discuss the basics of probabilistic fuzzy systems and the concept of fuzzy histograms. In Section 3 we give a brief introduction to VaR modelling and VaR models. In Section 4 we present two methods for estimating the PFS parameters that give more relevance to the tail points of the distribution, where the VaR

ISBN: 978-989-95079-6-8

is estimated. The experimental setup for the empirical study using six different assets are given in Section 5, while the results are reported in Section 6. Finally, conclusions and future work are given in Section 7.

2 Probabilistic Fuzzy Systems

Probabilistic fuzzy systems (PFS) are based on the concept of the probability of a fuzzy event, as defined by Zadeh [15]. It is assumed that the input space is a subset of $\mathbb{R}^n$ and that the rule consequents are defined on a finite domain $Y \in \mathbb{R}$. The PFS consists of a set of rules whose antecedents are fuzzy conditions and whose consequents are probability distributions. In this study, we consider Mamdani PFS in which the rules have the following form [14].

$$\begin{aligned} \text{Rule } R_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then} \\ y \text{ is } C_{q1} \text{ with } \Pr(C_{q1}|A_q) \text{ and} \\ y \text{ is } C_{q2} \text{ with } \Pr(C_{q2}|A_q) \text{ and } \ldots \text{ and} \\ y \text{ is } C_{qN} \text{ with } \Pr(C_{qN}|A_q). \end{aligned} \quad (1)$$

Hence, a Mamdani PFS is a generalisation of a Mamdani fuzzy system in which the deterministic fuzzy rules are replaced with probabilistic fuzzy rules. These rules specify a probability distribution over a collection of fuzzy sets that partition the output domain. The rules of a PFS also express linguistic information and they can be used to explain the model behavior by a set of linguistic rules. This way, the system deals both with linguistic uncertainty as well as probabilistic uncertainty. The interpretation of the probabilistic fuzzy rules is as follows. Given the occurrence of a (multidimensional) antecedent fuzzy event A_q, which is a conjunction of the fuzzy conditions defined on input variables, each of the consequent fuzzy events C_j is likely to occur. The selection of which of the consequent fuzzy events occurs, is done proportionally to the conditional probabilities $\Pr(C_j|A_q)$, $(j = 1, 2, \ldots, N)$. This applies for all the rules R_q, $q = 1, 2, \ldots, Q$. Note that two conditional probabilities $\Pr(C_j|A_q)$ and $\Pr(C_j|A_{q'})$ will be different, in general.

Let

$$\beta_q(\mathbf{x}) = \frac{\mu_{A_q}(\mathbf{x})}{\sum_{q'=1}^{Q} \mu_{A_{q'}}(\mathbf{x})} \quad (2)$$

be the normalised degree of fulfillment of rule R_q, where μ_{A_q} is the degree of fulfillment of rule R_q. When $\mathbf{x}$ is n-dimensional, μ_{A_q} is determined as a conjunction of the individual memberships in the antecedents computed by a suitable t-norm, *i.e.* ,

$$\mu_{A_q}(\mathbf{x}) = \mu_{A_{q_1}}(x_1) \circ \cdots \circ \mu_{A_{q_n}}(x_n), \quad (3)$$

where x_n is the n-th components of $\mathbf{x}$ and $\circ$ denotes a t-norm. Then, it can be shown that the output of the above Mamdani PFS is a conditional probability density function if an additive reasoning scheme is used with multiplicative aggregation of the rule antecedents [16]. The conditional probability of the output given an input vector $\mathbf{x}$ can be computed as

$$f(y|\mathbf{x}) = \sum_{j=1}^{N} \frac{\sum_{q=1}^{Q} \beta_q(\mathbf{x}) \Pr(C_j|A_q) \mu_{C_j}(y)}{\int_{-\infty}^{\infty} \mu_{C_j}(y) dy}, \quad (4)$$

assuming that the output space is well-formed, *i.e.* the output membership values satisfy

$$\sum_{j=1}^{N} \mu_{C_j}(y) = 1, \qquad \forall y \in Y. \quad (5)$$

It is also possible to compute the crisp output of the probabilistic fuzzy system by taking the conditional expectation of the output according to

$$E(y|\mathbf{x}) = \int_{-\infty}^{\infty} y f(y|\mathbf{x}) dy. \quad (6)$$

However, we are not interested in the expected output of the system in this paper. We are primarily interested in the output of the PFS as a fuzzy histogram, by using (4).

Assuming Y is fuzzily partitioned in a set of N fuzzy classes C_j described by membership functions $\mu_{C_j}(y)$, then the (fuzzy) column $f_j(y)$ for fuzzy class C_j can be estimated according to

$$f_j(y) = \frac{\Pr(C_j)\mu_{C_j}(y)}{\int_{-\infty}^{\infty} \mu_{C_j}(y) dy}, \quad (7)$$

The numerator in (7) describes a probability weighted with membership function $\mu_{C_j}(y)$. The denominator of (7) is a scaling factor representing the fuzzified size of class C_j (which in the one-dimensional continuous case, equals the *fuzzy length* of the interval C_q). The complete pdf $f(y)$ is again approximated by a summation of the functions $f_j(y)$:

$$f(y) \approx f_{app}(y) = \sum_{j=1}^{N} f_j(y) = \sum_{j=1}^{N} \frac{\Pr(C_j)\mu_{C_j}(x)}{\int_{-\infty}^{\infty} \mu_{C_j}(y) dy}. \quad (8)$$

Due to the overlap of the fuzzy sets, fuzzy histograms approximate probability distributions better, in practice. In Fig. 1 a representation of this phenomenon is shown, where a normal probability density function is approximated using both a crisp and a fuzzy histogram. In both cases, seven classes have been used.

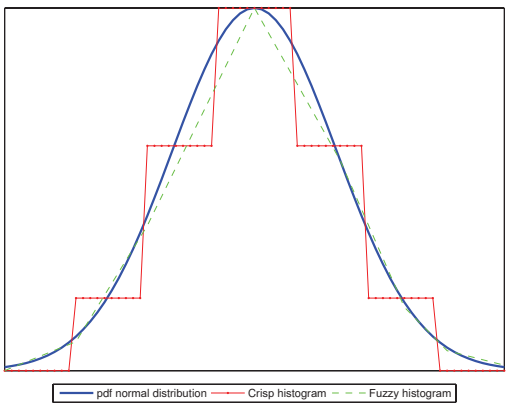


Figure 1: Fuzzy Histogram

3 VaR Estimation

In this section, we discuss value-at-risk estimation as an application example for probabilistic fuzzy systems. Value-at-risk is a single number for the senior management to express and

ISBN: 978-989-95079-6-8

summarise the market risk of a portfolio of financial assets. The VaR value of a portfolio is always calculated over a time horizon h at a significance level c. It indicates the maximum loss that a portfolio of assets will suffer over a horizon of h (days) with a confidence of c. An overview of the mainstream value at risk estimation methods can be found in [17]. Several methods are also discussed in [18]. Various building blocks of VaR measurement, methods for model validation as well as the differences between the parametric and nonparametric estimation approaches are discussed in [3].

3.1 Value at risk

Assume that a portfolio has value W_t at time t. Let r denote the one period percentage return of the portfolio. If $f(r)$ is the probability density function of the returns, define r_v such that

$$1 - c = \int_{-\infty}^{r_v} f(r)dr. \tag{9}$$

The value at risk V_t of the portfolio at time t is then defined as $V_t = -r_v W_t$. Assuming that the returns are distributed normally, the key step in the value at risk estimation can be formulated as determining the standard deviation σ of the returns distribution. This is also called *volatility estimation*.

Probabilistic fuzzy systems have been successfully applied in estimating the whole density distribution. In [13], good results were obtained by scaling the consequent membership functions, which were distributed with more membership functions are around the origin. The distribution of consequent functions alters the approximation of the density functions. Since VaR is calculated at the negative endpoint of the distribution function, it is logical to try to model the tail points more precisely. In Section 4 two approaches are presented that focus on the modeling of the tail points.

3.2 Model validation

Model validation is the process of checking whether a model performs adequately, and can be done in various ways. In this paper, we consider *exception based back testing*. Kupiec has developed a statistical test for assessing the validity of a VaR model [19]. Kupiec confidence regions are defined by the tail point of the log-likelihood ratio LR_{uc}

$$\begin{aligned} LR_{uc} &= -2\ln\left[c^{T-N}(1-c)^N\right] \\ &\quad +2\ln\left\{\left[1-\left(\frac{N}{T}\right)\right]^{T-N}\left(\frac{N}{T}\right)^N\right\}. \end{aligned} \tag{10}$$

In (10), N is the number of exceptions and T is the total number of observations. This ratio is shown to be asymptotically χ^2-distributed, with 1 degree of freedom, under the null hypothesis that the VaR model is valid [19]. Note that the Kupiec test statistic is two sided. Hence, the model is rejected both when there are too few exceptions, (the model is too conservative), as well as when there are too many exceptions, (the model underestimates the volatility). In this paper, we apply the Kupiec test with 95% confidence to assess the validity of the VaR models.

4 Parameter Learning for Tail Points

The parameters of the probabilistic fuzzy systems consist of the number of rules in the system, the parameters of the membership functions (*i.e.* number, type, location, etc.) and the probability parameters $\Pr(C_j|A_q)$ of the stochastic mapping between the antecedent and the consequents. The identification of all the parameters of the PFS simultaneously can be very time consuming and it suffers from the problem of multiple local minima. Thus, we use process knowledge to establish values of a subset of parameters. The other parameters are then optimised given the values of this subset of parameters.

In this work we determine the parameters of the antecedent membership functions by using a fuzzy clustering heuristic, that uses the fuzzy c-means algorithm [20] on the product space of the antecedent variables, and the spreads of the membership functions are derived from the distribution of the data, using a fuzzy covariance matrix, as proposed in [13]. Since the output membership functions must satisfy (5), it is convenient to use triangular membership functions for the output partition. These membership functions are combined with shouldered membership functions at the edges of the domain, as depicted in Fig. 2(a), to ensure that the domain is always covered by the fuzzy partition, no matter how extreme the values may be. The distribution of the membership functions can be uniform over the universe of discourse, or it can be varying. Two possible partitions for the output space are shown in Fig. 2. More membership functions placed around the edges of the universe of discourse, allows to better capture the variability in this region and to better model the tail points of the distribution. As one moves towards the edges of the universe of discourse, the membership functions have smaller supports and the separation between them decreases. Note that the triangular membership functions in this partition are not symmetric.

Assuming that the membership functions in the rule antecedents and the rule consequents have been defined, the optimal probability parameters $\Pr(C_j|A_q)$ can now be determined by using maximum likelihood parameter estimation, in which the log-likelihood function

$$J = \sum_{k=1}^{K} \ln\left(\Pr(y_k|\mathbf{x}_k)\right) \tag{11}$$

is maximised where K is the number of samples in the data set [21]. In (11), it is assumed that the samples in the data set are independent of one another. A suitable initialisation for iterative optimisation for maximum likelihood estimation is given by direct estimation from the data by using

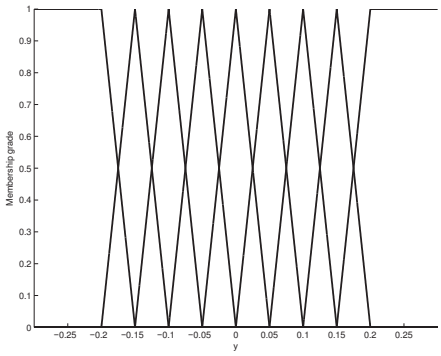
$$\Pr(C_j|A_q) = \frac{\sum_{k=1}^{K}\mu_{C_j}(y_k)\mu_{A_q}(\mathbf{x}_k)}{\sum_{k=1}^{K}\mu_{A_q}(\mathbf{x}_k)}. \tag{12}$$

During the parameter estimation of the optimal probability parameters $\Pr(C_j|A_q)$, more relevance can be given to the extremes of the universe of discourse. This can be accomplished by using an appropriate weight function $w(y_k)$ on the maximum likelihood parameter estimation, dependent on the distribution of the variable, that will give more weight to the tail points, in the form:

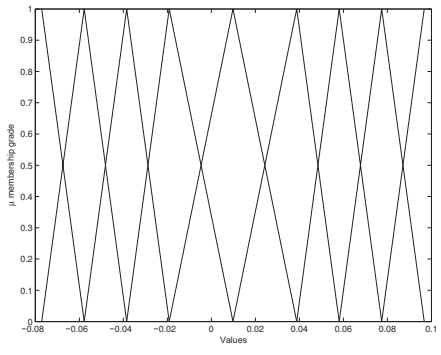
$$J = \sum_{k=1}^{K} \ln\left(w(y_k)\Pr(y_k|\mathbf{x}_k)\right). \tag{13}$$

The most simple kind of functions that can be used is an

ISBN: 978-989-95079-6-8



(a) Equally spaced.



(b) More sets towards the edges.

Figure 2: Two possible partitions for the output space.

affine function, such as

$$w(y_k) = \begin{cases} (\min_j(y_j)/2)^{-1} y_k & \text{if } y_k \leq 0, \\ (\max_j(y_j))^{-1} y_k & \text{if } y_k > 0. \end{cases} \quad (14)$$

Note that in this function more weight is given to the negative side, to better model the negative tail points and possibly obtain better VaR estimation.

5 Experimental Setup

We have studied the performance of the histogram-based probabilistic fuzzy systems to estimate VaR for six different stocks: KPN, ABN AMRO, JiaLing, BaoShan, COSCO and Merchant Bank. The performance of PFS models has been compared with the performance of the GARCH models.

In this paper, we consider value at risk models for one period ahead. In other words, the horizon over which the value at risk is computed is one day. Extensive models could be identified for the multiple day case (i.e. h-day horizon), but usually one suffices by using the simple $\sqrt{(h)}$ scaling, where the h-day value at risk is taken as $\sqrt{(h)}$ times the one-day value at risk.

The probabilistic fuzzy models that we consider use the returns r_t at period t to predict the distribution of the returns at period $t+1$. In our models, we have used nine antecedent membership functions and nine consequent membership functions. Hence, the fuzzy system had nine rules. The input space was partitioned using the FCM algorithm with nine clusters. In such a system, there are 81 probability parameters $\Pr(C_j|A_q)$ (nine for each rule). Since FCM has the tendency to place more clusters in regions covered with more data, there are more antecedent membership functions in the centre, where more samples are available.

The output space was partitioned with triangular membership functions. These triangular membership functions are combined with shouldered membership functions at the edges of the domain, to ensure that the domain is always covered by the fuzzy partitions, no matter how extreme the returns may be on a particular day.

Given the fuzzy membership functions whose parameters are determined as above, the conditional probability parameters for the PFS are determined by using maximum likelihood estimation. We used uniformly distributed triangular membership functions over the universe of discourse, combined with (13), and we name this model PFS$_{UD}$. For the case where we used the maximum likelihood parameter estimation as stated in (11), we used varying size of triangular membership functions over the universe of discourse, and we name this model PFS$_V$. Both approaches could be applied together. Nonetheless, in this work, we are interested in studying different ways of modelling the tail points of the probability density functions obtained with PFS, and thus we study them separately. Given the conditional probability distribution of one period returns, the value at risk of the portfolio is obtained by using (9).

The steps necessary for computing the one-period value-at-risk of a portfolio can now be summarised as follows for PFS$_1$ models.

1. Collect the price series regarding the portfolio and compute the one-period returns. Create training and validation data sets.
2. Determine antecedent membership functions by applying fuzzy c-means clustering to compute the locations of the membership functions and use cluster covariance to obtain the spreads of the membership functions.
3. Select the number of consequent membership functions and form a triangular partition, equally spaced in the case of PFS$_{UD}$ and with more sets towards the edges on PFS$_V$.
4. Given the definitions of the antecedent and the consequent membership functions, determine the optimal probability parameters of the PFS by maximising (13) for the PFS$_{UD}$ and (11) for the PFS$_V$.
5. Using the test set, compute the estimated conditional probability distribution function for the one-period returns for each observation in the test set.
6. Given the conditional probability distribution functions, compute the VaR by using (9).
7. Validate the model by using exception based back-testing as explained in Section 3.2.

6 Results

This section reports the application results for the proposed approaches to the stocks in study, for the validation data sets. Table 1 shows the obtained results of the exception-based back testing for the best GARCH and probabilistic fuzzy models.

ISBN: 978-989-95079-6-8

This table shows the number of exceptions that have occurred in the validation data for different levels of the confidence parameter c. The bold face numbers indicate that the model is not rejected according to the test statistic. The non-rejection region for the Kupiec test statistic is also shown.

Table 1: Failure rates for back testing

Asset	c	PFS_{UD}	PFS_V	GARCH	Non-Rejection Region
	95%	**26**	**30**	**19**	$16 < N < 36$
ABN	97.5%	**15**	**13**	**13**	$6 < N < 20$
	99%	**5**	**6**	**9**	$1 < N < 10$
	95%	**29**	**22**	11	$16 < N < 36$
KPN	97.5%	**14**	**15**	**8**	$6 < N < 20$
	99%	**10**	**9**	**4**	$1 < N < 10$
	95%	**30**	**35**	**22**	$16 < N < 36$
JiaLing	97.5%	**11**	**14**	**14**	$6 < N < 20$
	99%	**6**	**6**	6	$1 < N < 10$
	95%	**24**	**20**	12	$16 < N < 36$
BaoShan	97.5%	**14**	**13**	**8**	$6 < N < 20$
	99%	**9**	**4**	**6**	$1 < N < 10$
	95%	**26**	**25**	14	$16 < N < 36$
COSCO	97.5%	**9**	**9**	**11**	$6 < N < 20$
	99%	**5**	**6**	**5**	$1 < N < 10$
	95%	**27**	**27**	10	$16 < N < 36$
Merchant	97.5%	**15**	**14**	5	$6 < N < 20$
	99%	**6**	**3**	**4**	$1 < N < 10$

As can be seen in Table 1 the GARCH models are rejected for some data sets, while the PFS models are accepted for all data sets. This implies the presence of extreme values in the returns series that are not captured by the GARCH model [7]. Note that for KPN, the number of exceptions obtained for $c = 99\%$ is on the limit indicated by the Kupiec test.

Table 2 shows the initial probability parameters obtained with (12). It can be seen that all probability variables are positive according to this estimation. Table 3 and Table 4 shows the optimal probability parameters obtained after maximum likelihood estimation for PFS_{UD} and PFS_V, respectively.

Table 2: Initial probability parameters for ABN AMRO model.

	Consequent								
Rule	1	2	3	4	5	6	7	8	9
1	0.1003	0.1333	0.2066	0.0567	0.0441	0.1215	0.1012	0.1218	0.1143
2	0.0565	0.1351	0.1659	0.0792	0.0744	0.0972	0.1617	0.1489	0.0811
3	0.0459	0.1679	0.1495	0.1300	0.1024	0.0773	0.1463	0.1359	0.0448
4	0.0544	0.1683	0.1700	0.1105	0.0647	0.0802	0.1454	0.1579	0.0468
5	0.0516	0.1578	0.1800	0.1119	0.0692	0.0770	0.1547	0.1472	0.0506
6	0.0563	0.1648	0.1760	0.0956	0.0655	0.1206	0.1340	0.1227	0.0646
7	0.0700	0.1877	0.1659	0.0748	0.0405	0.0901	0.1147	0.2002	0.0562
8	0.0529	0.1625	0.1626	0.1122	0.0999	0.0930	0.1313	0.1286	0.0570
9	0.0539	0.1729	0.1624	0.1132	0.0746	0.0772	0.1476	0.1505	0.0476

Note that some of the probability parameters are now zero. As tables Table 3 and Table 4 show, more probability mass is centered around the middle consequents. A higher probability mass around the center may indicate that the extreme returns are much less frequent, but this is not always the case. The returns seem to revert to the average after a positive or negative extreme value, and the volatility seems to be rather different for the average returns and the extremes [16]. Hence these results capture the tail behavior in the returns distribution. When we compare these results with the ones obtained when distributing more membership functions are around the origin [13], we can see that the tails of the return distribution are more pronounced in the presented work. We believe that this is due to the fact the in the presented work, both proposed approaches better model and capture the extreme values of the data.

Table 3: Probability parameters for PFS_{UD} ABN AMRO model after optimisation.

	Consequent								
Rule	1	2	3	4	5	6	7	8	9
1	0	0	0.0397	0.2675	0.6928	0	0	0	0
2	0	0	0.1049	0.3454	0.5497	0	0	0	0
3	0	0.0090	0	0.3466	0.6225	0.0219	0	0	0
4	0.0562	0	0.0700	0.5111	0.1847	0.1224	0	0.0556	0
5	0	0.0121	0.0074	0.4865	0.4797	0	0.0142	0	0
6	0	0.0357	0	0.4496	0.4550	0.0085	0.0513	0	0
7	0.0145	0.0472	0	0.4219	0.4774	0.0226	0.0163	0	0
8	0	0	0	0.3355	0.5736	0	0.0908	0	0
9	0	0.0441	0	0.1853	0.6198	0.1286	0	0	0.0222

Table 4: Probability parameters for PFS_V ABN AMRO model after optimisation.

	Consequent								
Rule	1	2	3	4	5	6	7	8	9
1	0	0.0123	0.0326	0.2897	0.6654	0	0	0	0
2	0	0.0310	0	0.4509	0.4618	0	0.0563	0	0
3	0	0.0435	0	0.2054	0.6082	0.1208	0	0	0.0222
4	0.0596	0	0.0719	0.5362	0.1819	0.0909	0	0.0596	0
5	0	0.0122	0	0.3315	0.6317	0.0245	0	0	0
6	0	0	0.1068	0.4104	0.4828	0	0	0	0
7	0	0.0070	0.0244	0.4559	0.4989	0	0.0138	0	0
8	0.0229	0	0.0096	0.4270	0.5231	0.0071	0.0104	0	0
9	0	0	0	0.3320	0.5878	0	0.0801	0	0

The obtained probability density functions for ABN-AMRO using the PFS_V are shown in Fig. 3. As can be seen, for some cases the distribution does not resembles a normal distribution. Specifically, the cases were the distribution exhibits fat tails, correspond to high volatility data.

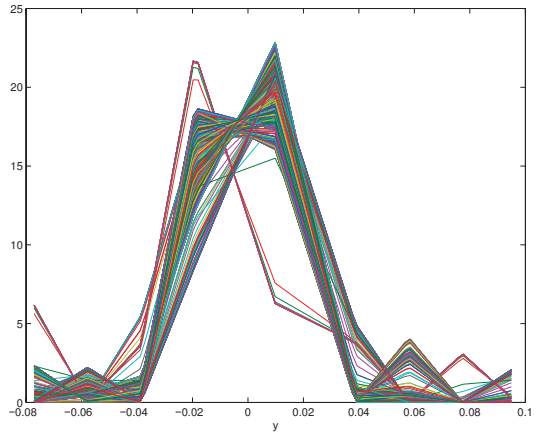


Figure 3: Obtained pdf ABN-AMRO using PFS_V

It is also interesting to consider how the VaR values estimated by the PFS compare to the values estimated by the GARCH models. Table 5 shows the sum of the differences between the VaR estimated and the actual losses in the periods where the VaR estimation is smaller than the actual losses, i.e., when exceptions occur. As can be seen in Table 5, the ex-

ISBN: 978-989-95079-6-8

Table 5: VaR Exceptions

Asset	c	PFS_{UD}	PFS_V	GARCH
ABN	95%	0.0339	0.0333	0.0446
	97.5%	0.0430	0.0452	0.0510
	99%	0.0586	0.0558	0.0587
KPN	95%	0.0591	0.0596	0.0887
	97.5%	0.0744	0.0724	0.1015
	99%	0.0902	0.0881	0.1166
JiaLing	95%	0.0301	0.0288	0.0405
	97.5%	0.0371	0.0348	0.0463
	99%	0.0469	0.0434	0.0533
BaoShan	95%	0.0439	0.0244	0.0350
	97.5%	0.0478	0.0314	0.0400
	99%	0.0531	0.0463	0.0460
COSCO	95%	0.0327	0.0330	0.0499
	97.5%	0.0428	0.0429	0.0571
	99%	0.0561	0.0570	0.0656
Merchant	95%	0.0264	0.0266	0.0362
	97.5%	0.0312	0.0318	0.0413
	99%	0.0371	0.0388	0.0475

pected losses are in most cases smaller in the PFS models, with the exception of BAOSHAN. In other cases where the GARCH model has smaller total expected losses than the PFS model, the GARCH model leads to a smaller number of exceptions, which may indicate a conservative model.

7 Conclusions

We have proposed two approaches for determining the model parameters of probabilistic fuzzy systems for value at risk modelling, where both approaches give more relevance to the tail points of the distribution of values. In one of the approaches we distributed more membership functions around the edges of the universe of discourse and used maximum likelihood to estimate the conditional parameters of the model. In the other approach we distributed the membership functions uniformly over the universe of discourse and used a weighted maximum likelihood to estimate the conditional parameters of the model.

The performance of the proposed models has been compared to the VaR estimation by using the popular GARCH (1,1) volatility estimation. It is found that PFS models are not rejected by back testing, while GARCH models are sometimes rejected. Furthermore, we show how VaR estimation is improved by using these approaches. This shows the added flexibility that comes through the use of the probabilistic fuzzy models, enabling them to adapt to the properties of the data and of the problem at study.

References

[1] R. A. Brealey and S. C. Myers. *Principles of Corporate Finance*. McGraw Hill, New York, 7th edition edition, 2001.

[2] J. Hull. *Options, Futures and Other Derivatives*. Prentice Hall, Upper Saddle River, 4 edition, 2000.

[3] Philippe Jorion. *Value at Risk : the new benchmark for managing financial risk*. McGraw–Hill, New York, 2 edition, 2001.

[4] Glyn A. Holton. *Value at Risk: thoery and practice*. Academic Press, Amsterdam, The Netherlands, 2003.

[5] T. Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[6] Rama Cont. Volatility clustering in financial markets: empirical facts and agent-based models. In G. Teyssiére and A. P. Kirman, editors, *Long Memory in Economics*, pages 289–310. Springer, Berlin, 2007.

[7] N.S. Balke and T.B. Fomby. Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. *Journal of Applied Econometrics*, 9:181–200, 1994.

[8] J. W. Taylor. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting*, 19(4):299–311, July 2000.

[9] N. Chapados and Y. Bengio. Cost functions and model combination for var-based asset allocation using neural networks. *IEEE Transactions on Neural Networks*, 12(4):890–906, July 2001.

[10] M. Jiang, Z. Q. Meng, and Q. Y. Hu. A neural network model on solving multiobjective conditional value-at-risk. In F. Yin, J. Wang, and C. Guo, editors, *Advances in Neural Networks*, number 3174 in Lecture Notes in Computer Science, pages 1000–1006. Springer, Berlin, 2004.

[11] Zdeněk Zmeškal. Value at risk methodology under soft conditions approach (fuzzy-stochastic approach). *European Journal of Operational Research*, 161:337–347, 2005.

[12] Umberto Cherubini and Giovanni Della Lunga. Fuzzy value-at-risk: accounting for market liquidity. *Economic Notes by Banca Monte dei Paschi di Siena SpA*, 30(2):293–312, 2001.

[13] Du Xu and Uzay Kaymak. Value-at-risk estimation by using probabilistic fuzzy systems. In *Proceedings of the 2008 World Congress on Computational Intelligence*, pages 2109–2116, Hong-Kong, June 2008.

[14] Uzay Kaymak, Willem-Max van den Bergh, and Jan van den Berg. A fuzzy additive reasoning scheme for probabilistic Mamdani fuzzy systems. In *Proceedings of the 2003 IEEE International Conference on Fuzzy Systems*, volume 1, pages 331–336, St. Louis, USA, May 2003.

[15] L. A. Zadeh. Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23(2):421–427, 1968.

[16] Jan van den Berg, Uzay Kaymak, and Willem-Max van den Bergh. Financial markets analysis by using a probabilistic fuzzy modelling approach. *International Journal of Approximate Reasoning*, 35:291–305, 2004.

[17] Darrell Duffie and Jun Pan. An overview of value at risk. *Journal of Derivatives*, 4(3):7–49, Spring 1997.

[18] Z. Wiener. Introduction to VaR (value-at-risk). In D. Galai, D. Ruthenberg, M. Sarnat, and B. Schreiber, editors, *Risk Management and Regulation in Banking*, pages 47–63. Kluwer Academic Publishers, Boston, 1999.

[19] Paul H. Kupiec. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 2:73–84, 1995.

[20] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[21] L. Waltman, U. Kaymak, and J. van den Berg. Maximum likelihood parameter estimation in probabilistic fuzzy classifiers. In *Proceedings of the 14th Annual IEEE International Conference on Fuzzy Systems*, pages 1098–1103, Reno, Nevada, USA, May 2005.

ISBN: 978-989-95079-6-8

Optimization of an Oil Production System using Neural Networks and Genetic Algorithms

Guillermo Jimenez de la C[1,2] Jose A. Ruz-Hernandez[1] Evgen Shelomov[1]

Ruben Salazar M.[1,2]

1 Universidad Autonoma del Carmen, Av. 56 # 4 esq.
Av. Concordia, Col. Aviacion, C.P. 24180, Cd. del Carmen, Campeche, Mexico.
2 Instituto Mexicano del Petroleo, Av. Periferica Norte # 75 esq. Calle 35B, Col. San Agustin del Palmar,
C.P. 24118, Cd. del Carmen, Campeche, Mexico.
Email: gjc@mexico.com, jruz@pampano.unacar.mx, eshelomov@pampano.unacar.mx, rsalazar@imp.mx

Abstract— *This paper proposes an optimization strategy which is based on neural networks and genetic algorithms to calculate the optimal values of gas injection rate and oil rate for oil production system. Two cases are analyzed: a) A single well production system and b) A production system composed by two gaslifted wells. For both cases an objective function is maximized to reduce production cost. The proposed strategy shows the ability of the neural networks to approximate the behavior of an oil production system and the genetic algorithms to solve optimization problems when a mathematical model is not available.*

Keywords— Genetic algorithms, injection gaslift, neural network, optimization, oil production system, perceptron multilayer.

1 Introduction

The daily operation of an oil and gas production system, have many decisions, which affect the volumes produced and the cost of production oil. These decisions are taken at different levels in the organization, but eventually they will reach the physical production system [1]. Fig. 1 gives an overview of a physical gas lift production system. For such oil production systems, the decisions are related to find the lift gas rate for each well giving the maximum total oil production rate at very instance of time.

An objective function is a single-valued and well-defined mathematical function mapping the values of the decision variables into a performance measure. Examples of such performance measures are the total oil production rate, net present value (profit), or the recovery of the reservoir. To improve the performance of the production system, a question to be answered is: what decisions are better to maximize or minimize the objective function?. In the process of making good decisions, information about the production system is used. This information may include the physical properties such as pipe diameters and lengths, or it may include measurements from the production system.

To support making good decisions, well models may be used to develop the production plans. Typically, well test are performed to determine the gas to oil ratio, water cut, and production rates of each individual well. Well test are performed by routing a well to a dedicated separator. This separator will separate three phases, and a rate transmitter is connected to the outlet for each phase. The well model is update using the measurements taken during a test. Fluid sampling may be used to obtain the fluid composition including the water cut.

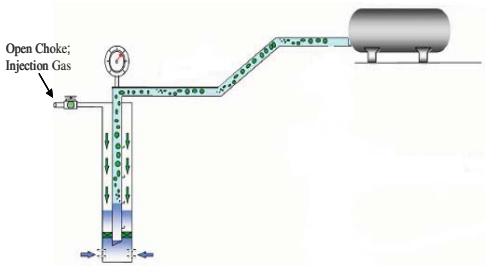


Figure 1: Gas lift production system of a single well.

The objective of gas lift is to increase oil production or allow nonrateing wells to rate by reducing the hydrostatic head of the fluid column in the well [2]. By injecting gas into the tubing, the density of the wellbore fluid decrease; thus, the pressure-drop component resulting from gravity is reduced. However, the gas lift also gives a larger pressure-drop component resulting from friction, giving some optimal gas lift rate for the well. Usually, the available lift-gas for a group of wells is less than the sum of the individual optimum lift-gas rates for each well. The gas-lift optimization problem is established to find the lift gas rates for each well giving the maximum total oil production rate subject to a gas lift processing capacity constraint and possibly other operational and processing constraints.

In this paper, a model-based optimization via neural networks and genetic algorithms is developed and used to calculate the optimal values of gas injection rate and oil rate of a gas lift production system. Two cases were analyzed: a) A single well production system and b) A production system composed by two gas lifted wells. For both cases maximize the objective function to reduce production cost. The proposed strategy shows the ability of the neural networks to approximate the behavior of an oil production system and to solve optimization problems when a mathematical model is not available.

This paper presents a methodology of hybrid computational intelligent using neural networks and genetic algorithm. Others related references for example; genetic algorithms for neural network training on transputers, neural network weight selection using genetic algorithms, studies on the speed of convergence of neural network training using genetic algorithm, automatic generation of neural networks

ISBN: 978-989-95079-6-8

with parameter setting based on genetic algorithms, evolutionary algorithms for neural network design and training, others.

2 Optimization strategy based on a neural network and genetic algorithms

In order to solve the gas-lift optimization problem, an optimization procedure based on a neural network and genetic algorithms was developed. The strategy selected is based on three components as illustrated in Fig. 2. The first one is a neural network which is used to approximate the gas lift performance curve, the second one uses an objective function to satisfy a performance index and the third one is used to solve the optimization problem via genetic algorithms.

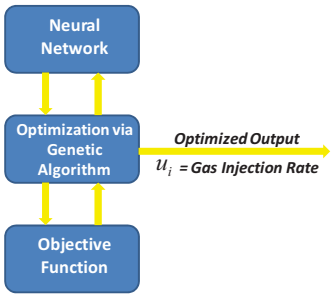


Figure 2: Strategy of optimization based on a neural network and genetic algorithms.

2.1 Neural Network

A Multi-Layer Perceptron (MLP) is selected. The use of this kind of networks to approximate functions and carry out identification process goes back to more than one decade [3]. The following equation is used to determine the structure of a MLP with a single hidden or intermediate layer, a neuron in the output layer with function of linear activation and M hidden neurons is used;

$$F(x_1,...,x_P)=\sum_{i=1}^{M}\alpha_i g(\sum_{j=1}^{p} w_{ij}x_j-\theta_i) \qquad (1)$$

The expression to define the neural network used in our strategy is given by

$$Q_i = g[u_i(i-1)] \qquad (2)$$

where:

Q_i : is the estimated produced oil rate (STB/day).

u_i : is the Gas lift rate injected into the well (MMscf/day)

g: function activation

In this application, the approximation procedure is done using a neural network multilayer perceptron with three layers. The hidden layer has neurons using sigmoid activation function and the output layer has a unique neuron with linear activation function. Different MLPs are trained by means of Levenberg-Marquardt Algorithm, which uses the criterion of middle square error to update the neural network weights. The corresponding MLP is displayed in Fig. 3.

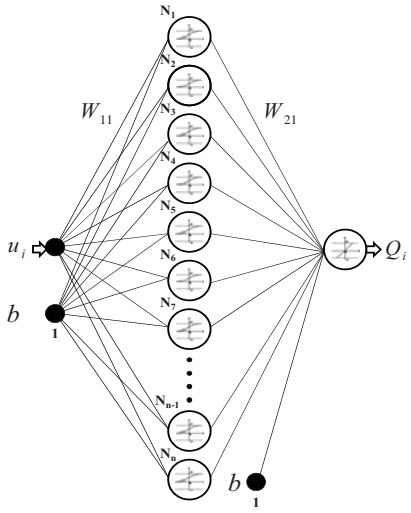


Figure 3: Neural network MLP.

2.2 Objective function

The expression to define the objective function used in our strategy of optimization is given by:

$$J(u)=\sum_{i=1}^{N}\alpha_i Q_i(u_i)-\beta_i u_i \qquad (3)$$

where:

α_i : Production costs by 1 STB/day produced oil.

β_i : Supply costs by 1 MMscf/day injected gas.

N : amount wells.

α_i and β_i are required to balance units

[STB/day]/[MMscf/day]= STBD/MMscfd

The objective function given in (3) can give a net gain, relating adequately processes numerical comparison of the output of the desired product and quantity of flow of injection gas.

2.3 Optimization via genetic algorithms

In order to maximize the objective function, given in (3), the following simple genetic algorithm is applied [4].

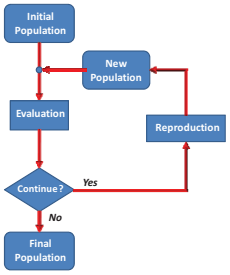


Figure 4: Simple genetic algorithm pseudocode diagram.

ISBN: 978-989-95079-6-8

The properties of genetic algorithms which are applied are the following ones

- Type of genetic algorithm: Simple genetic algorithm [4].
- Amount of unit in population: 100 chromosomes
- Amount of units that contain "N" genes, to find the optimal value: 200 generations.
- Probability of mutation= 0.001.
- In order to select the units the match method is used.

In this application we take into account that the values of injection gas must satisfy the following conditions.

$$u_i = \{0.00, 0.01, 0.002, \ldots, 5.99, 6\} \quad (4)$$

The previous condition represents the precision considered when the injection gas is sampled.

Other conditions are:

- The rate of gas injection will be a value in the following interval (MMscf/day):

$$0.00 \le u_i \le 6.00 \quad (5)$$

- There is not mathematical model of the process; however, experimental data can be obtained from a well simulator.
- A neural network model of the process is available, which can be used to approximate its behavior and to construct the objective function.

Two cases are simulated; the first one corresponds to a single well production system, then the equation (3) can be rewritten like:

$$Max[J(u)] = \alpha_1 Q_1 - \beta_1 u_1 \quad (6)$$

and the second one considers a production system composed by two wells, then the equation (3) can be rewritten like:

$$Max[J(u)] = \alpha_1 Q_1 + \alpha_2 Q_2 - \beta_1 u_1 - \beta_2 u_2 \quad (7)$$

3 Results and Discussions

3.1 First case: Produced oil by a single well

The corresponding oil production system is illustrated in Fig. 1. For this case, two single wells are considered. The first one has a pressure in the well head (P_{wh}) equal to 14 kg/cm^2 and the second one has 12 Kg /cm^2. A simulation program is used to collect data and to train neural networks using neural networks toolbox of Matlab.

The best trained neural networks are described in Table 1 and Table 2. Other parameters like the static pressure in the reservoir, lengths and diameters of the tubing and lines, chokes and others pipe components are also considered. Figure 5 illustrates the simulation for a well with P_{wh}= 14 kg/cm^2.

Table 1: Neural network architecture and obtained errors with P_{wh} = 14 Kg/cm^2.

Number of neurons		error	error (%)
Intermedia layer	Output layer		
30	1	0.1365	0.002

Table 2: Neural network architecture and obtained errors with P_{wh} = 12 Kg/cm^2.

Number of neurons		error	error (%)
Intermedia layer	Output layer		
30	1	0.0119	0.0004

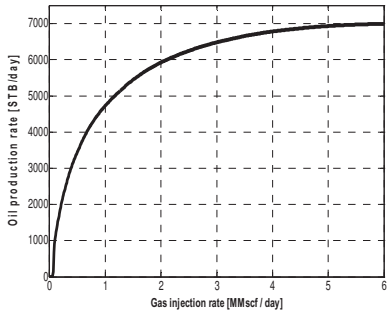


Figure 5: Estimated production curve oil production rate vs gas injection rate, P_{wh} = 14 Kg/cm2.

Figure 6 displays the obtained results when the gain of production for this well with P_{wh}=14 Kg/cm^2 is maximized. Furthermore,α_1=*28*USD/STBD and β_1=*5250*USD/MMscfd are considered in equation (6); α_1 *and* β_1 estimates cost for this year.

The obtained results in this case are contained in Table 3.

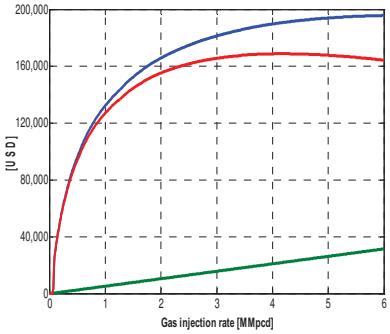


Figure 6: Optimization results; income of production (blue), costs by supply of injection gas (green) and objective function (Optimal gain-red).

ISBN: 978-989-95079-6-8

Table 3: Gain of production results

Optimal gain production (USD)	Income optimal production (USD)	Costs by supply of injection gas (USD)	Obtained optimal gas injection rate (MMscf / day)
168,840	190,680	21,840	4.16

Figure 7 illustrates the simulation results for a well when P_{wh}=12 kg/cm^2 is used.

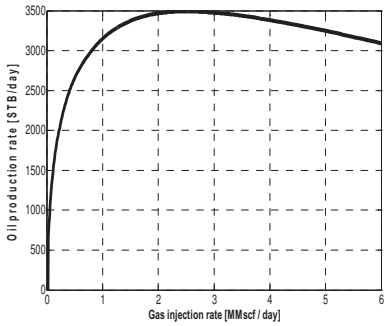


Figure 7: Estimated production curve oil production rate vs gas injection rate, P_{wh} = 12 Kg/cm2.

Figure 8 displays the obtained results via optimization procedure to maximize the gain of production for this well when P_{wh}=12 Kg/cm2 is used. As considered above, α_1=*28*USD/STBD and β_1=*5250*USD/MMscfd are used in equation (6).

The obtained results in this case are contained in Table 4.

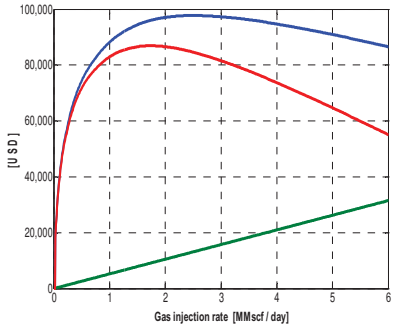


Figure 8: Optimization results; income of production (blue), costs by supply of injection gas (green) and objective function (Optimal gain-red).

Table 4: Gain of production results

Optimal gain production (USD)	Income optimal production (USD)	Costs by supply of injection gas (USD)	Obtained optimal gas injection rate (MMscf / day)
86,921	96,004	9,082	1.73

3.2 Second case: Produced oil rate by a production system based on two wells.

In Fig. 9 is shown an oil production system of two wells. Production system made up of wells of 12 kg/cm^2 and 14 Kg /cm^2. A simulation program is used to collect data and to train neural networks using neural networks toolbox of Matlab. The best trained neural networks is described in Table 5.

Table 5: Neural network architecture and obtained errors, two wells.

Number of neurons		error	error (%)
Intermedia layer	Output layer		
30	1	-0.3702	-0.0037

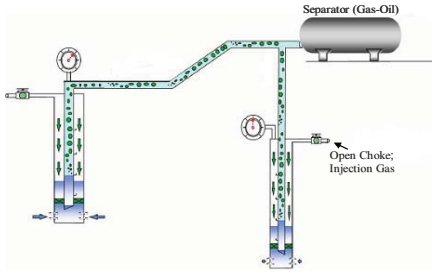


Figure 9: Gas lift production system composed by two wells.

3.2.1 Sub case 1. $u_i = u_1 = u_2$

The Fig.10 illustrates the simulation for a system composed by two wells (Black).

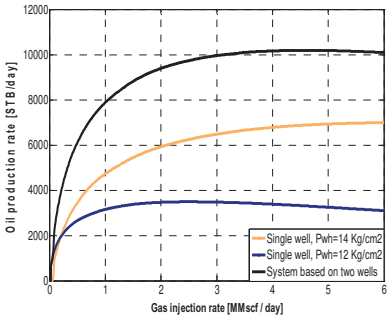


Figure 10: Estimated production curve oil production rate vs gas injection rate. System composed by two wells.

Figure 11 displays the simulation results of maximizing of the gain of production for two wells. Assume that α_1= α_2=*28*USD/STBD and β_1=β*2=5250*USD/MMscfd are considered in equation (7). Furthermore, objective function is subject to the next constrains:

$$0.00 \le u_i \le 6.00 \text{ and } u_2 = u_1 .$$

The obtained results in this subcase are contained in Table 6.

ISBN: 978-989-95079-6-8

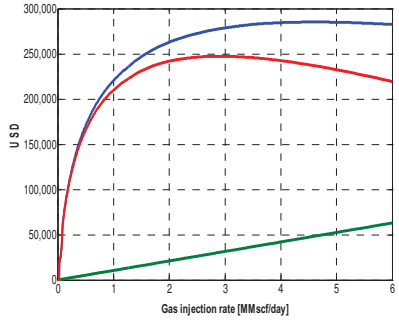


Figure 11: Optimization results; income of production (blue), costs by supply of injection gas (green) and objective function (Optimal gain-red).

Table 6: Gain of production results

Optimal gain production (USD)	Income optimal production (USD)	Costs by supply of injection gas (USD)	Obtained optimal gas injection rate (MMscf / day)	
247,260	277,550	30,240	2.88	2.88

3.2.2 Sub case 2. $u_1+u_2=K$

The Fig. 12 illustrates the simulation for a system composed by two wells (Black).

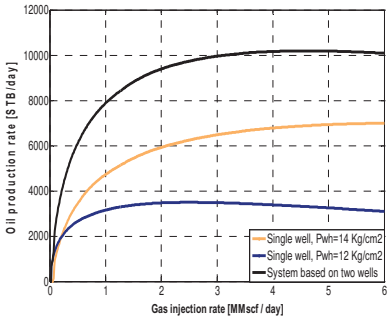


Figure 12: Estimated production curve oil production rate vs gas injection rate. System composed by two wells.

Figure13 presents the results of maximizing of the gain of production for two wells. α_1 *and* β_1 considered in objective function subject to constrains :

$$0.00 \le u_i \le 6.00$$

and

$$u_2 + u_1 = K$$

where $K = 4$. Furthermore, $\alpha_1=\alpha_2=28$USD/STBD and $\beta_1=\beta_2=5250$USD/MMscfd are considered in equation (7).

The obtained results in this subcase are contained in table 7.

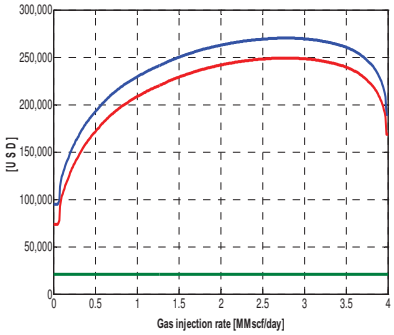


Figure 13: Optimization results; income of production (blue), costs by supply of injection gas (green) and objective function (Optimal gain-red).

Table 7: Gain of production results

Optimal gain production (USD)	Income optimal production (USD)	Costs by supply of injection gas (USD)	Obtained optimal gas injection rate (MMscf / day)	
249,360	270,360	21,000	2.79	1.21

4 Conclusions

The obtained results shows that neural networks and genetic algorithms are useful to optimize the costs and gains in an oil production systems. The implementation of these methodologies to be applied in petroleum industry allows to increase the gains, to reduce the costs, save natural elements and to extend the life of wells.

Acknowledgements

The authors thank support of Universidad Autonoma del Carmen, Mexico on project PR/131/2008.

References

[1] H.P. Bieker, O. Slupphaug, and T. A. Johansen, "Real-Time Production Optimization of Oil and Gas Production Systems: A Technology Survey," *SPE Production & Operations Journal*, vol. 22, no. 4, pp. 382-391, November 2007.

[2] A. Vasper, "Auto, Natural, or In-Situ Gas-Lift Systems Explained," *SPE Production & Operations Journal*, vol. 23, no. 1, pp. 75-80, February 2008.

[3] J. Nocedal, and S.J. Wright, *Numerical Optimization*. New York: Springer, 1999. K Narendra, and K. Parthasarathy, "Identification and control of dynamical systems using neural networks". IEEE Transactions on Neural Networks, vol.1, no 1. pp. 4-27, 1990.

[4] D. Goldberg, (1989). Genetic Algorithms in Search, Optimization, and Machine Learning (Addison-Wesley, Readings, MA).

[5] W.Y. Fang, and K.K. Lo, "A Generalized Well-Management Scheme for Reservoir Simulation," *SPE Reservoir Engineering Journal*, vol. 11, no. 2, pp. 116-120, 1996.

[6] E.P. Kanu, J. Mach, and K.E. Brown. "Economic Approach to Oil Production and Gas Allocation in Continuous Gas Lift," *Journal of Petroleum Technology*, vol. 33, no. 10, pp. 1887-1892, 1981.

ISBN: 978-989-95079-6-8

[7] K Hornik, M. Stinchcombe, and White H. “Multilayer Feedforward Networks are Universal Aproximators”, Neural Networks, vol. 2, no. 5, 1989.

[8] M. Norgaard, O. Ravn; Poulsen, N. K., y L. K., Hansen, “Neural Networks for Modelling and Control of Dynamic Systems”, Springer-Verlag, London, 2000.

[9] Prosper. Petroleum Experts Ltd. User’s Guide IPM V5.3Build #173 Jun 6, 2006.

[10] J. A. Ruz-Hernandez., D. A. Suarez, E. Shelomov. and A. Villavicencio, “Predictive Control Based on an Auto-Regressive Neuro-Fuzzy Model Applied to the Steam Generator Startup Process at a Fossil Power Plant”, Revista de Computación y Sistemas, vol. 6, no. 3, CIC-IPN, ISSN 1405 5546, Mexico, pp.204-212, 2003.

[11] E. N. Sánchez, G A. Alanís, “Redes Neuronales, Conceptos Fundamentales y Aplicaciones al Control Automático”, Prentice Hall, 2006.

[12] J. M. Corchado, F. Díaz. “Redes Neuronales Artificiales, un enfoque Practico”. Universidad de Vigo. ISBN. 848158145-3.

[13] G. Cybenko, “Aproximation by superpositions of sigmoidal function”, Math. Control, Sygnals and System, vol. 2, nº 4, 1989.

ISBN: 978-989-95079-6-8

Rule Based Fuzzy Cognitive Maps in Socio-Economic Systems

Joao Paulo Carvalho[1,2] Jose A. B. Tome [1,2]

1 INESC-ID - Instituto de Engenharia de Sistemas e Computadores

2 IST – Instituto Superior Tecnico, TULisbon
Lisboa, Portugal
Email: {joao.carvalho, jose.tome}@inesc-id.pt

Abstract *— This paper shows how Rule Based Fuzzy Cognitive Maps can be useful to model qualitative socio-economic systems, by discussing and presenting a macroeconomic model that, although developed eight years ago and based on simple qualitative rules, described and predicted the current economic crisis when the policy behavior of the banking system regarding tax-rates was assumed similar to the one that ended up occurring.*

Keywords — Qualitative Modelling of Dynamic Systems, Rule Based Fuzzy Cognitive Maps, Socio-Economic Systems.

1 Introduction♠

Economic models have traditionally been based on mathematics. Econometry, the quantitative science of modelling the economy, focus on creating models to help explain and predict variables of interest in economics. However the most common econometric models are usually very imprecise and are not usually valid but on a very short term. This can easily be seen on the regular predictions made to most macroeconomic indicators: most yearly predictions made by governments, economic entities or independent experts, must usually get corrected every trimester due to inaccuracies in the models used to predict their values. This is essentially due to the fact that most econometric models tend to ignore the existence of the feedback loops that make any alteration in any component of the model to potentially be propagated until that component is affected by its own previous change on a relatively short term.

The more precise models that try to address this issue are usually based on differential equations [13, 20]. However, due to the dimension of these systems (very high number of variables involved), these models demand a strong knowledge in mathematics, and a huge amount of time to be developed. However, even when these kinds of models are possible, they usually tend to ignore that economy is a social science, and therefore is subject to qualitative uncertainties due to human and social factors that are not easily captured using strict quantitative mathematic models.

♠ This work was supported in part by the FCT - Portuguese Foundation for Science and Technology under project PTDC/MAR/66231/2006

The use of approaches that include both the existence of feedback cycles and the intrinsic qualitative social nature associated with economy, should lead to the implementation of more accurate models.

Dynamic Cognitive Maps (DCM), where Rule Based Fuzzy Cognitive Maps (RB-FCM) are included, are a qualitative approach to modeling and simulating the Dynamics of Qualitative Systems (like, for instance, Social, Economical or Political Systems) [2,3,4,5,6,7,8,9,10,11]. RB-FCM were developed as a tool that can be used by non-engineers and/or non-mathematicians and eliminates the need for complex mathematical knowledge when modeling qualitative dynamic systems. In this paper one shows that RB-FCM can be used to model socio-economical systems using as an example a model that was originally designed eight years ago, but that can explain and simulate the current world economic situation.

2 Dynamic Cognitive Maps

The term Dynamic Cognitive Maps has been recently used to describe techniques that allow simulating the evolution of cognitive maps through time. Axelrod [1] work on cognitive maps (CM) introduced a way to represent real-world qualitative systems that could be analyzed using several methods and tools. However, those tools only provided a way to identify the most important structural elements of the CM. Complete, efficient and practical mechanisms to analyze and predict the evolution of data in CM were not available for years due to several reasons. System Dynamics tools like those developed by J.W.Forrester [12] could have provided the solution, but since in CM numerical data may be uncertain or hard to come by, and the formulation of a mathematical model may be difficult, costly or even impossible, then efforts to introduce knowledge on these systems should rely on natural language arguments in the absence of formal models. Fuzzy Cognitive Maps (FCM), as introduced by Kosko [15, 16, 17], were developed as a qualitative alternative approach to system dynamics. However, although very efficient and simple to use, FCM are causal maps (a subset of cognitive maps that only allow basic symmetric and monotonic causal relations between concepts)[9], and, in most applications, a

ISBN: 978-989-95079-6-8

FCM is indeed a man-trained Neural Network that is not fuzzy in a traditional sense and does not exploit usual fuzzy capabilities. They do not share the properties of other fuzzy systems and the causal maps end up being quantitative matrixes without any qualitative knowledge.

Several evolutions and extensions have been added to the original FCM model, but none addresses so many FCM issues as RB-FCM. RB-FCM were developed as a tool that models and simulates real world qualitative system dynamics while trying to avoid the limitations of those approaches. The following sub sections resume some features of RB-FCM that are useful to the comprehension of this paper.

2.1 Rule Based Fuzzy Cognitive Maps

RB-FCM allow a representation of the dynamics of complex real-world qualitative systems with feedback, and the simulation of events and their influence in the system. They can be represented as fuzzy directed graphs with feedback, and are composed of fuzzy nodes (Concepts), and fuzzy links (Relations). RB-FCM are true cognitive maps since are not limited to the representation of causal relations. Concepts are fuzzy variables described by linguistic terms, and Relations are defined with fuzzy rule bases.

RB-FCM are essentially iterative fuzzy rule based systems where we added fuzzy mechanisms to deal with feedback, introduced timing mechanisms [8] and new ways to deal with uncertainty propagation, and were we defined several kinds of Concept relations (Causal, Inference, Alternatives, Probabilistic, Opposition, Conjunction, etc. [2,5]) to cope with the complexity and diversity of the dynamic qualitative systems we are trying to model. Among new contributions brought by RB-FCM, there is a new fuzzy operation – the Fuzzy Carry Accumulation - [4,7], which is essential to model the mechanisms of qualitative causal relations (FCR – Fuzzy Causal Relations) while maintaining the simplicity and versatility of FCM.

There are 2 main classes of Concepts: Levels , that represent the absolute values of system entities (e.g., LInflation is Good); and Variations, that represent the change in value of a system entity in a given amount of time (e.g., VInflation increased very much). By allowing the definition of both the absolute value of a concept and its change through time, RB-FCM have the means to properly model the dynamics of a system.

2.1.1 Expressing Time in Dynamic Cognitive Maps

Time is probably the most essential factor when modeling a dynamic system. However, most DCM approaches seem to ignore this fact. In order to maintain consistency in the process of modeling the dynamics of a qualitative system, it is necessary to develop and introduce timing control mechanisms. To allow the representation of time flow, delays, and the inhibition of certain relations when they have no influence on a given instant, changes were made to the engine of RB-FCM. More details regarding RB-FCM time mechanisms can be found in [8,11]

3 A Qualitative Macro Economic Model as an Example of DCM Modeling in Socio-Economic Systems

In this section one presents a model that was developed eight years ago. The primary goal when this problem was approached was to show the capabilities and ease of use of RB-FCM to model the dynamics of qualitative real world systems. Even if the final model is rather complex and does not contains apparent flaws, it is not, and was never intended to be, a complete model, since it wasn't developed by economic experts (even though some were consulted). However, as it can be seen in the obtained results, the model exhibits a behaviour that is able to describe the current economic crisis and the reasons that lead to it.

Classic cognitive mapping techniques [1] were used as the first step to obtain the model: the concepts and relations were extracted from a short column published in Portuguese newspaper Público in the year 2001 consisting on an economic expert analysis regarding "Short-term Tax Rate evolution in Europe" [18]. Throughout the text, the author introduced several concepts, supporting its theories while explaining the relations between concepts using qualitative knowledge. The "classic" CM obtained was much simpler than the one presented here, which was expected, since the analysis of the dynamics of a much more complex model – like the one we ended up obtaining – would require several months of work using traditional quantitative econometric approaches. The first model used only the most important concepts (the ones really necessary to a short term analysis): Tax Rates, Inflation, Consumption, Oil Price, and Food Cost. Even with such a few concepts, a realistic model becomes rather difficult to analyse due to the complexity of the relations that affect the involved concepts. However, since the goal was to show RB-FCM potential to deal with larger systems and long-term simulations, the model was evolved to be more realistic therefore including more concepts and much more relations. On this step, 13 concepts were added to the original 5 (Figure 7).

At the end of this phase of the modelling process one obtained a classic Cognitive Map – basically a graph where the nodes were the Concepts and each edge represented an existing unknown relation between a pair of Concepts.

3.1 Concept Modelling

The next step was refining the concepts to obtain a linguistic fuzzy representation for each one. This step consisted in defining the class(es) (Variation, Level) and the linguistic terms and membership functions for each concept. In dynamic systems, variations are much more important than absolute values, therefore, most concepts are Variations, some are Levels, and a few key concepts like Inflation, Tax Rate, etc., are both Variations and Levels (the Level value of these concepts is actualized according to its Variation using a special LV relation [5]).

ISBN: 978-989-95079-6-8

The linguistic terms of Levels must have a direct correspondence with the real world values. Therefore we allied common sense and expert consulting (using straight questions like "what do you consider a high value for Inflation?", and receiving answers like "around 4%") to define their membership functions. In the particular case of Levels that depend on LV relations, it was also necessary to define the real-world meaning of a certain amount of variation (e.g., a "Small" increase on inflation is around 0.3%). Figure 5 shows the linguistic terms of the Level concept LInflation.

Variation linguistic terms usually represent qualitative terms without a direct correspondence to absolute values. E.g., VInflation has 11 linguistic values ranging from "Huge Decrease" to "Huge Increase" (Figure 6). Linguistic terms of Variations can usually be represented by standard sets, which simplify and accelerate the modelling process [10].

3.2 Qualitative Modelling of a Qualitative Dynamic System

The huge advantages of using Fuzzy Rule Bases (FRB) to define qualitative relations between Concepts has been largely discussed and proved [2,3,5,9,11]. The major drawback of rule-based fuzzy inference, the combinatorial explosion of the number of rules, is avoided in RB-FCM by the use of Fuzzy Causal Relations and the Fuzzy Carry Operation [4]. Another important feature of RB-FCM is the simplicity of the process of insertion and removal of Concepts and/or Relations, which also reduces the modelling complexity of FRB [4]. Therefore one has in RB-FCM an adequate tool to model qualitative relations. However, the single fact of using linguistic rule bases to model relations does not guarantee the qualitative nature of the model. Let us see the example of Inflation modelling:

A pseudo-qualitative approach using FRB would try to closely map the widespread quantitative approaches: Inflation value is predicted by a weight averaged sum of several factors (Estimated Oil inflation, estimated Food price inflation, etc.). This method is highly dependent on the precision and validity of each factor real-world absolute value. In the proposed model, a novel approach where rules are independent from the real world absolute values was used. The model is based on a qualitative definition of inflation: Economics theory states that economic growth depends on inflation – without inflation there is no growth; In fact, the worst economic crisis (30´s for instance) are associated with deflation; Therefore, it is desirable and expected that all factors that affect inflation have a certain cost increase – If all factors suffer a normal increase, then the inflation will maintain its normal and desired value. Therefore, one can state the following qualitative relation for each of those *n* factors:

"If $factor_n$ has a normal increase, then Inflation will maintain"

This statement is part of the fuzzy rule base of a causal relation. Since fuzzy causal effects are accumulative and their effect is a variation in the value of the consequent, then if all factors that cause inflation have the normally expected increase, Inflation will not vary. If some factors increase more than expected and the others maintain their value then inflation will somehow increase. If a factor increase less than normal, or even decreases, then its effect is a decrease in inflation (note that the final variation of Inflation is given by the accumulation of all causal variation effects – e.g., if some pull it down a bit, and one pulls it up a lot, in the end inflation still can maintain its normal value).

It is possible to build a completely qualitative and sound causal FRB to model each factor influence of Inflation, without ever referring to absolute values. If one intends to model inflation in South America, one can maintain the rule base. All that needs to be changed are the linguistic terms of the Level Concept associated to Inflation (for instance, normal inflation would become around 8 %, and so on...). Obviously some factors are more important than others (a large increase in food might cause a large increase in Inflation, but what is considered a large increase in Oil might only cause a small increase in Inflation – average Oil price varied over 100% in the last 2 years, but other factors had a slightly above average increase, and therefore inflation had mild increase instead of a sever one…). This "relative" importance is easily modelled as a causal effect in a FRB. Table I represents an example of a causal FRB. One can also mention the fact that oil price variation has a delayed effect in inflation. RB-FCM provide mechanisms to model these kinds of timing issues [8,11]

TABLE I: FCR7 +sl Food Cost, Inflation

If Food Cost Decreases VeryMuch,	Inflation has a Large Decrease
If Food Cost Decreases Much,	Inflation has a Large Decrease
If Food Cost Decreases,	Inflation has a Large Decrease
If Food Cost Decreases Few,	Inflation Decreases
If Food Cost Decreases VFew,	Inflation Decreases
If Food Cost Maintains,	Inflation Decreases
If Food Cost Increases VFew,	Inflation has a Small Decrease
If Food Cost Increases Few,	Inflation has a Very Small Decrease
If Food Cost Increases Normally,	Inflation Maintains
If Food Cost Increases M,	Inflation has a Small Increase
If Food Cost Increases VM,	Inflation Increases

This kind of qualitative approach was used throughout the model when causal relations were involved.

As it was mentioned above, Variations usually have a standard set of linguistic terms. These allow the predefinition of certain common fuzzy causal relations (FCR). These FCR are called macros and were used to reduce the modelling effort.

The model includes other than causal relations. For instance: Oil price variation was modelled using a classic fuzzy inference rule base (FIRB) based on oil Offer/Demand (where Oil offer was decided in simulated periodic OPEP meetings); the Tax Rates were modelled considering that Banks were managed as a common business with profit in

ISBN: 978-989-95079-6-8

mind – for example, an increase in money demand would increase Tax Rates (this would be changed later (see 3.4)).

Regarding timing considerations, the system was modelled considering a one month period between iterations.

It is obviously impossible to detail every aspect of the system modelling in this paper. provides a graphic representation of the final RB-FCM model. The system consists of 18 concepts and around 400 fuzzy rules to express relations (most were automatically generated using macros). The system was described using RB-FCMsyntax (a dedicated language) – a complete description is available in [2]. Here are some guidelines regarding the description of relations in Figure 7: "FCR+" stands for a standard positive causal relation (an increase in the antecedent will cause an increase in consequent), and "FCR-" a standard negative relation (increase causes a decrease). Several "+" or "–" represent stronger effects. A "/" represents an attenuated effect. "sl" and "sr" represent biased effects (non symmetric causal relations. A "?" represents a relation which cannot be symbolically described (one must consult the FCR). A "d" represents a delay in the effect. FIR stands for Fuzzy Inference Relation. The number after FCR or FIR is the label for the complete description of the rule base.

3.3 Simulation results[1,2]

The simulation of the original system provided rather interesting results. The evolution of the system through time was rather independent from the initial values and the external effects. After a certain period of time, that could vary from a few months to several years (depending on a conjugation of external factors like a war, or a severe cut in oil production), economy would end up collapsing: deflation, negative growth, 0% tax-rates. Figure 1 represents one of those cases.

Initially one could think that there was a major flaw in the model (or in the RB-FCM mechanisms), but after a discussion and analysis of the results with an economics expert, the culprit was found: the model approached the economic situation before the creation of entities that control Interest Rates (like the U.S. Federal Reserve, the European Central Bank). The lack of these entities was the main cause to economic instability until 1930's. In fact, Economics was known in the 18th and 19th century as the "Dark Science", because all theories indicated that economy was not sustainable. According to the simulation results, depression always comes after a growth period and due to an exaggerate increase in tax rates (the banks try to maximize their profit in a short period, and their greed cause an apparently avoidable crisis). Notice the similarity with the present economic crisis - one will return to this point later on. Therefore, to support the referred theory, a simple model of the European Central Bank behaviour regarding interest rates was added to the model.

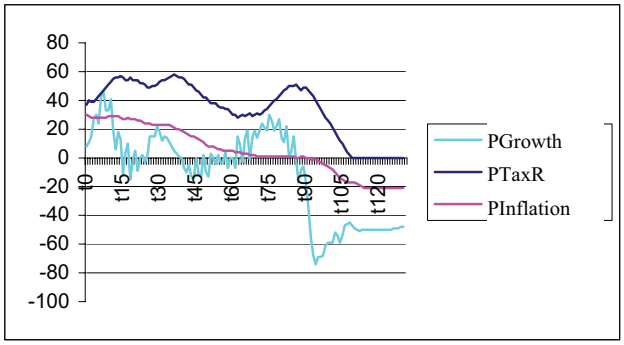


Figure 1 – Serious economic crisis: Negative growth and deflation (Predicted Growth, Predicted TaxRate, Predicted Inflation,)

3.4 Modelling European Central Bank Influence

To simulate ECB influence, a Fuzzy Inference Subsystem (FISS) – a RB-FCM block used to model the process of decision making of system entities (FISS timing mechanisms are independent of the RB-FCM) – was added to the model (Figure 2). This FISS ended up as a simple FRB with 48 rules (each with 2 antecedents)[2]. These rules were designed to inhibit the greedy bank behaviour that was identified as the cause to the unavoidable crisis.

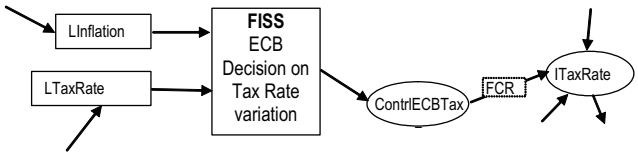


Figure 2 – FISS: ECB decision on Interest Rate variation

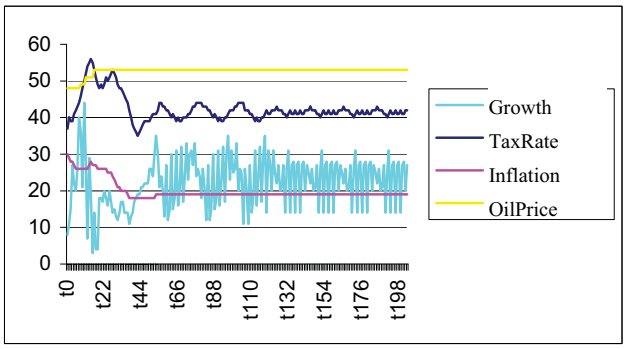


Figure 3 – Avoiding economic crisis trough ECB Interest Rate control

3.5 Complete Model Simulation

With the introduction of the ECB-FISS, system behaviour changed completely and serious crisis were avoided under normal circumstances (Figure 3). One of the most interesting results was the fact that, under normal circumstances, the economic model stabilizes around the real-world BCE predicted ideal target value for inflation (slightly below 2%)

[1] Note: This section refers to an initial version of the system that did not include concept 17

[2] All variables are in % and must be divided by 10 (e.g., 15 represents 1.5%)

ISBN: 978-989-95079-6-8

and growth (averaging slightly above 2%). Note that these values are not imposed anywhere in the model, they result from the system itself.

However, the ECB and private bank behaviour in the last two years was incredibly similar to the greedy behaviour exhibited by the model without ECB. Tax rates – see Euribor historical data [14], Figure 4 – were severely increased between 2006-08 under the pretext of controlling inflation, but, as it was found later, mostly because private banks were needing to increase their tax rates to protect themselves against prior mistakes. The variation of the Euribor tax rate + 0.8% spread rate in the last 10 years (120 month), is very much similar to what was predicted in the "greedy" model that was presented 8 years ago. As a result of that, and, as the original model predicted, we are entering a severe economic crisis as a result of that policy. Given that this is a very long term 10 year simulation (done 8 years ago), the results are incredibly more accurate than current models, that don't usually attempt to predict for longer than 2 years and usually with very inaccurate results. As a proof of this, less than 1 year ago all major economic actors were still insisting on increasing tax rates having inflation control in mind, and no major economic actor was even suspecting that deflation would be the real concern in less than 8 months.

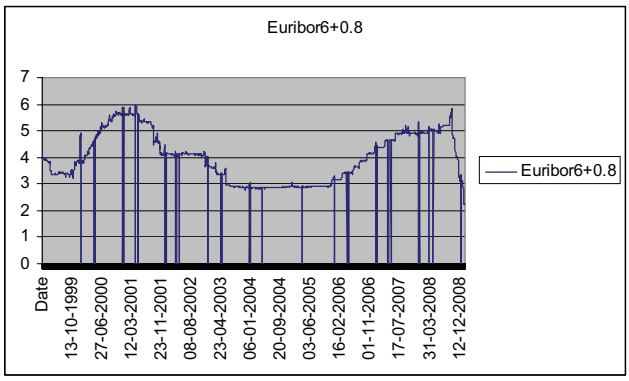


Figure 4 – Euribor6+0.8 for the last 120 months

4 Conclusions, Applications, and Future Developments

In this paper we exemplified how one can use DCM to model complex qualitative socio-economic systems, avoiding the need to use extensive and time consuming differential equation models, while obtaining very interesting and encouraging results.

By using true qualitative modelling techniques, one obtained results that look more realistic (plausible) than those obtained using quantitative approaches – where results almost never show the short term uncertainties that are so characteristic of qualitative real-world dynamic systems. In the end, the results of the presented model, that was developed eight years ago, are surprisingly realistic and could have been used to predict and avoid the current world economic crisis, even if one considers its necessary incompleteness.

5 References

[1] Axelrod, R., "The Structure of Decision: Cognitive Maps of Political Elites", Princeton University Press, 1976

[2] Carvalho, J.P., "Mapas Cognitivos Baseados em Regras Difusas: Modelação e Simulação da Dinâmica de Sistemas Qualitativos", PhD thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal, 2002

[3] Carvalho, J.P., Tomé, J.A., "Qualitative Optimization of Fuzzy Causal Rule Bases using Fuzzy Boolean Nets", Fuzzy Sets and Systems, 158(17), pp. 1931-1946, Elsevier, 2007

[4] Carvalho, J.P., Tomé,J.A., "Views on Fuzzy Sets and Systems from Different Perspectives. Philosophy and Logic, Criticisms and Applications. (Studies in Fuzziness and Soft Computing), Chapter 19 - Fuzzy Mechanisms for Qualitative Causal Relations", Edited by. Rudolf Seising, Springer, 2009

[5] Carvalho, J.P., Tomé, J.A., "Rule Based Fuzzy Cognitive Maps – Qualitative Systems Dynamics", Proceedings of the 19th International Conference of the North American Fuzzy Information Processing Society, NAFIPS2000, Atlanta, 2000

[6] Carvalho, J.P., Tomé, J.A.,"Issues on the Stability of Fuzzy Cognitive Maps and Rule-Based Fuzzy Cognitive Maps", Proceedings of the 21st International Conference of the North American Fuzzy Information Processing Society, NAFIPS2002, New Orleans, 2002

[7] Carvalho, J.P., Tomé,J.A., "Fuzzy Mechanisms For Causal Relations", Proceedings of the Eighth International Fuzzy Systems Association World Congress, IFSA'99, Taiwan, 1999

[8] Carvalho, J.P., Tomé,J.A.,"Rule Based Fuzzy Cognitive Maps - Expressing Time in Qualitative System Dynamics", Proceedings of the 2001 FUZZ-IEEE, Melbourne, Australia, 2001

[9] Carvalho,J.P., Tomé, J.A., "Rule Based Fuzzy Cognitive Maps and Fuzzy Cognitive Maps - A Comparative Study", Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society, NAFIPS99, New York, 1999

[10] Carvalho,J.P., Tomé, J.A., "Using Interpolated Linguistic Term to Express Uncertainty in Rule Based Fuzzy Cognitive Maps", Proceedings of the 22nd International Conference of the North American Fuzzy Information Processing Society, NAFIPS2003, Chicago, 2003

[11] Carvalho, J.P., Wise,L., Murta, A., Mesquita, M., "Issues on Dynamic Cognitive Map Modelling of Purse-seine Fishing Skippers Behavior", WCCI2008 - 2008 IEEE World Congress on Computational Intelligence, Hong-Kong , pp. 1503-1510 , 2008

[12] Forrester, J.W., several papers available online at http://sysdyn.mit.edu/sd-intro/home.html

[13] Halanay, A., Samuel, J., "Differential Equations, Discrete Systems and Control: Economic Models (Mathematical Modelling: Theory and Applications)", Kluwer Academic, 1997

[14] http://www.euribor.org/html/content/euribor_data.html

[15] Kosko, B., "Fuzzy Cognitive Maps", International Journal of Man-Machine Studies, 1986

[16] Kosko, B., "Fuzzy Engineering", Prentice-Hall International Editions, 1997

[17] Kosko, B., "Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence", Prentice-Hall International Editions, 1992

[18] Neves, A., "Taxas de Juro podem descer já este mês", Público 2001/08/10

ISBN: 978-989-95079-6-8

[19] Zadeh, L., "Fuzzy Sets and Applications: Selected Papers, Wiley-Interscience", 1987

[20] Zhang, W.-B., "Differential Equations, Bifurcations and Chaos in Economics", World Scientific, 2005

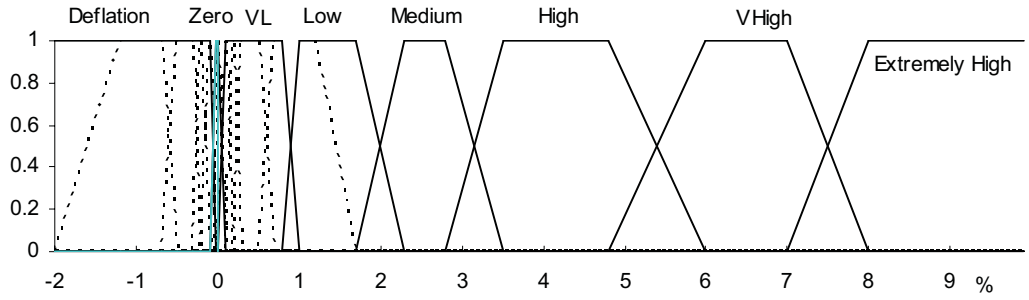


Figure 5 – Level "Inflation" linguistic terms. Dotted linguistic terms represent the variation degrees of Inflation {Huge_Decrease, Large_Decrease, Decrease, Small_Decrease, VS_Decrease, Maintain, VS_Increase,…., Huge_Increase}

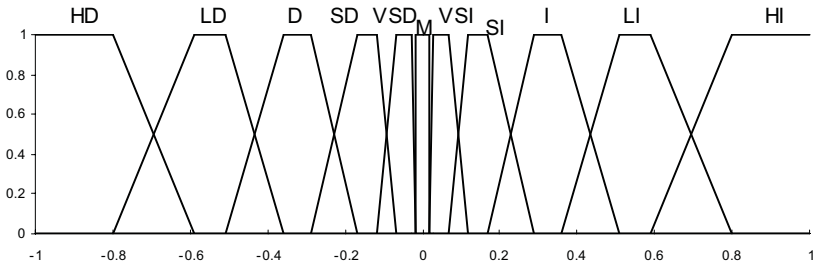


Figure 6 – Variation "Inflation" linguistic terms: {Huge Decrease, Large Decrease,…,Huge Increase}. x-scale values are normalized values. There is no relation no real world values

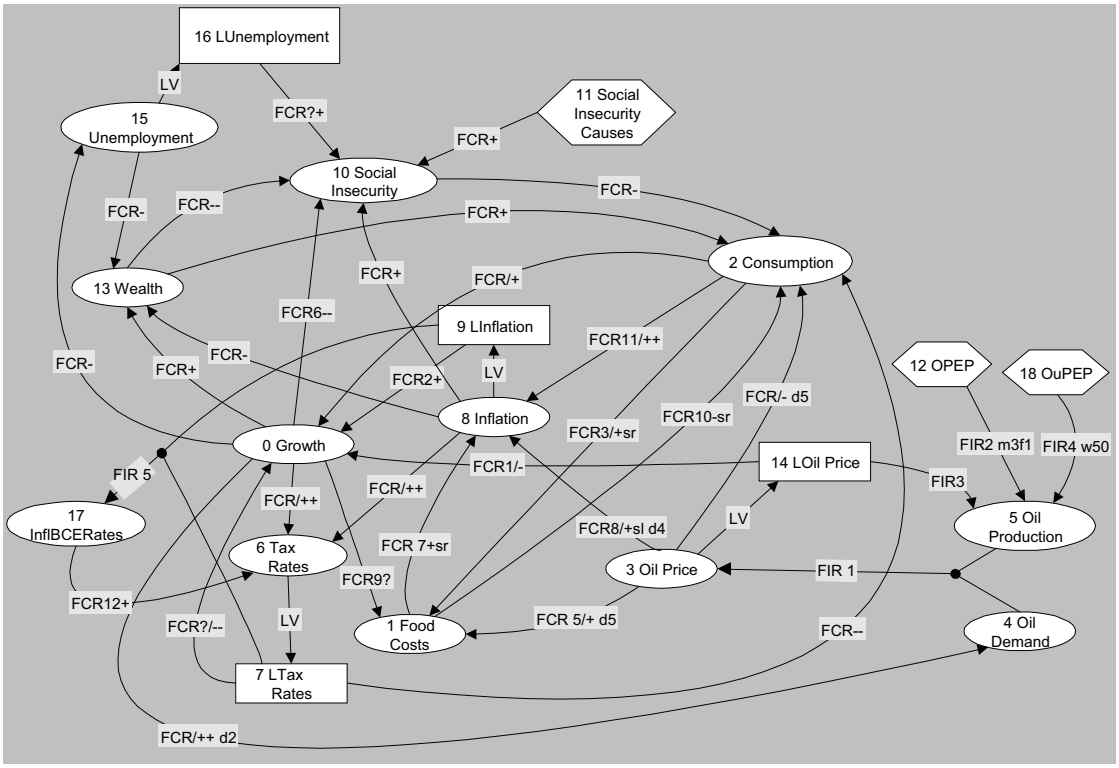


Figure 7 – RB-FCM: A qualitative model of economy. Concept 17, representing a simple FISS was added later

ISBN: 978-989-95079-6-8

Operational Semantics for a Fuzzy Logic Programming System with Defaults and Constructive Answers

Hannes Strass Susana Munoz-Hernandez Victor Pablos Ceruelo

Technical University of Madrid (Spain) *
E-mail: hannes.strass@alumnos.upm.es, {susana, vpablos}@fi.upm.es

***Abstract*— *In this paper we present the operational semantics of RFuzzy, a fuzzy Logic Programming framework that represents thruth values using real numbers from the unit interval. RFuzzy provides some useful extensions: default values to represent missing information, and typed terms to intuitively restrict predicate domains. Together, they allow the system to give constructive answers in addition to truth values. RFuzzy does not confine to a particular Fuzzy Logic, but aims at being as general as possible by using the notion of aggregation operators.*

Keywords*— *Operational Semantics, Logic Programming Application

1 Introduction

For many real-world problems, crisp knowledge representation is not perfectly adequate. Information that we handle might be imprecise, uncertain, or even both. Classical two-valued logic cannot easily represent these qualitative aspects of information. To address this issue, multiple frameworks for incorporating uncertainty in logic have been developed over the years: fuzzy set theory, probability theory, multi-valued logic, or possibilistic logic; to mention only some.

From the point of view of practical tools to support this reasoning the field is not so rich. Logic programming is however a perfect candidate for the implementation of these tools because it is traditionally used for problem solving and knowledge representation.

1.1 Fuzzy Logic approaches

The result of introducing Fuzzy Logic into Logic Programming has been the development of several fuzzy systems over Prolog. These systems replace the inference mechanism, SLD-resolution, of Prolog with a fuzzy variant that is able to handle partial truth [1]. Most of these systems implement the fuzzy resolution introduced by Lee in [2], examples being the Prolog-Elf system [3], the Fril Prolog system [4] and the F-Prolog language [5]. However, there is no common method for fuzzifying Prolog as has been noted in [6].

1.2 Fuzzy Prolog

One of the most promising fuzzy tools for Prolog was the "Fuzzy Prolog" system [7, 8]. This approach is more general than others in some respects:

1. A truth value is a finite union of sub-intervals on $[0, 1]$.
2. A truth value is propagated through the rules by means of an *aggregation operator*. The definition of *aggregation operator* is general.
3. Crisp and fuzzy reasoning are consistently combined in a Prolog compiler [9].

Fuzzy Prolog adds fuzziness to a Prolog compiler using CLP($\mathcal{R}$) instead of implementing a new fuzzy resolution as other former fuzzy Prologs do. So, it uses Prolog's built-in inference mechanism, and the constraints and their operations provided by CLP($\mathcal{R}$) to handle the concept of partial truth. It represents intervals as constraints over real numbers and *aggregation operators* as operations with these constraints.

There are other proposals, e.g. in [10], that provide an interpretation of truth values as intervals, but Fuzzy Prolog proposed to generalise this concept to unions of intervals for the first time.

1.3 Multi-adjoint logic

Over the last few years several papers have been published by Medina et al. [11, 12, 13] about multi-adjoint programming. The theoretical model described in these works led to the development of FLOPER [14], another Fuzzy Logic Programming system. It has a Logic-Programming-inspired syntax and provides free choice of aggregation operators and credibility of rules just as RFuzzy does. There are however some things that FLOPER cannot do: (1) deal with missing information (which RFuzzy does by default truth value declarations), (2) type atoms and predicates to give constructive answers, and (3) provide syntactic sugar to express truth value functions.

1.4 Motivation

The generality of the approach pursued in [7, 8] turned out to make many users feel uncomfortable: Fuzzy Prolog is rather expressive, so it is not always clear how knowledge should be represented. Furthermore, interpreting the output that comes as a sequence of constraints maybe possible for a human but is very hard to do for a computer program – especially basing a decision upon it may not be straightforward.

To address these issues, we propose the RFuzzy framework. It is considerably simpler to use than the above-mentioned Fuzzy Prolog, but still contains many of its nice features. In RFuzzy, truth values will be represented by real numbers from the unit interval $[0, 1]$. This simplifies making modelling decisions and interpreting the output of the system. We still use the general concept of aggregation operators to be able to model different Fuzzy Logics. The rules of RFuzzy programs will have attached a *credibility* value to them: it allows the author

*This work is partially supported by the project DESAFIOS - TIN 2006-15660-C02-02 from the Spanish Ministry of Education and Science, by the Spanish Ministry of Science and Innovation Research Staff Training Program - BES-2008-008320 and by the project PROMESAS - S-0505/TIC/0407 from the Madrid Regional Government.

ISBN: 978-989-95079-6-8

of the rule to express how much they confide in the relation expressed by the rule. In addition, RFuzzy offers features that are very useful for knowledge representation, namely default values and types.

The rest of the paper is organised as follows: Section 2 introduces the abstract syntax of RFuzzy. In Section 3 we present an operational semantics for RFuzzy and illustrate it with an example. The last but one section shortly sketches the key points of the implementation and Section 5 concludes.

2 RFuzzy Syntax

We will use a signature Σ of function symbols and a set of variables V to "build" the *term universe* $\mathrm{TU}_{\Sigma,V}$ (whose elements are the *terms*). It is the minimal set such that each variable is a term and terms are closed under Σ-operations. In particular, constant symbols are terms.

Similarly, with use a signature Π of predicate symbols to define the *term base* $\mathrm{TB}_{\Pi,\Sigma,V}$ (whose elements are called *atoms*). Atoms are predicates whose arguments are elements of $\mathrm{TU}_{\Sigma,V}$. Atoms and terms are called *ground* if they do not contain variables. As usual, the *Herbrand universe* H is the set of all ground terms, and the *Herbrand base* B is the set of all atoms with arguments from the Herbrand universe.

To combine truth values in the set of real truth values $[0,1]$, we will make use of *aggregation operators*. A function $\hat{F} : [0,1]^n \to [0,1]$ is called an aggregation operator if it verifies $\hat{F}(0,\ldots,0) = 0$ and $\hat{F}(1,\ldots,1) = 1$. We will use the signature Ω to denote the set of used operator symbols F and $\hat{\Omega}$ to denote the set of their associated aggregation operators $\hat{F}$. An n-ary aggregation operator is called *monotonic in the i-th argument*, if additionally $x \leq x'$ implies $\hat{F}(x_1,\ldots,x_{i-1},x,x_{i+1},\ldots,x_n) \leq \hat{F}(x_1,\ldots,x_{i-1},x',x_{i+1},\ldots,x_n)$. An aggregation operator is called *monotonic* if it is monotonic in all arguments.

Immediate examples for aggregation operators that come to mind are typical examples of t-norms and t-conorms: minimum $\min(a,b)$, maximum $\max(a,b)$, product $a \cdot b$, and probabilistic sum $a + b - a \cdot b$.

The above general definition of aggregation operators subsumes however all kinds of minimum, maximum or mean operators.

Definition. Let Ω be an aggregation operator signature, Π a predicate signature, Σ a term signature, and V a set of variables.

A *fuzzy clause* is written as

$$A \xleftarrow{c,F_c}_{F} B_1,\ldots,B_n$$

where $A \in \mathrm{TB}_{\Pi,\Sigma,V}$ is called the head, $B_1,\ldots,B_n \in \mathrm{TB}_{\Pi,\Sigma,V}$ is called the body, $c \in [0,1]$ is the credibility value, and $F_c \in \Omega^{(2)}$ and $F \in \Omega^{(n)}$ are aggregation operator symbols (for the credibility value and the body resp.)

A *fuzzy fact* is a special case of a clause where $n = 0$, $c = 1$, F_c is the usual multiplication of real numbers "$\cdot$" and $F = v \in [0,1]$. It is written as $A \leftarrow v$.

A *fuzzy query* is a pair $\langle A, v\rangle$, where $A \in \mathrm{TB}_{\Pi,\Sigma,V}$ and v is either a "new" variable that represents the initially unknown truth value of A or it is a concrete value $v \in [0,1]$ that is asked to be the truth value of A. ⌟

Intuitively, a clause can be read as a special case of an implication: we combine the truth values of the body atoms with the aggregation operator associated to the clause to yield the truth value for the head atom. For this truth value calculation we are completely free in the choice of an operator.

Example. Consider the following clause, that models to what extent cities can be deemed good travel destinations – the quality of the destination depends on the weather and the availability of sights:

$$\text{good-destination(X)} \xleftarrow{1.0,\cdot} \text{nice-weather(X), many-sights(X).}$$

The credibility value of the rule is 1.0, which means that we have no doubt about this relationship. The aggregation operator used here in both cases is the product "$\cdot$". We enrich the knowledge base with facts about some cities and their continents:

$$\begin{gathered}
\text{nice-weather(madrid)} \leftarrow 0.8,\\
\text{nice-weather(istanbul)} \leftarrow 0.7,\\
\text{nice-weather(moscow)} \leftarrow 0.2,\\
\text{many-sights(madrid)} \leftarrow 0.6,\\
\text{many-sights(istanbul)} \leftarrow 0.7,\\
\text{many-sights(sydney)} \leftarrow 0.6,\\
\text{city-continent(madrid, europe)} \leftarrow 1.0,\\
\text{city-continent(moscow, europe)} \leftarrow 1.0,\\
\text{city-continent(sydney, australia)} \leftarrow 1.0,\\
\text{city-continent(istanbul, europe)} \leftarrow 0.5,\\
\text{city-continent(istanbul, asia)} \leftarrow 0.5.
\end{gathered}$$

Some queries to this program could ask if Madrid is a good destination, $\langle \text{good-destination(madrid)}, \mathsf{v}\rangle$. Another query could ask if Istanbul is the perfect destination, $\langle \text{good-destination(istanbul)}, 1.0\rangle$.The result of the first query will be the real value 0.48 and the second one will fail. It can be seen that no information about the weather in Sydney or sights in Moscow is available although these cities are "mentioned". ◇

In the above example, the knowledge that we represented using fuzzy clauses and facts was not only vague but moreover incomplete. As this is rather the norm than the exception, we would like to have a mechanism that can handle non-present information.

In standard logic programming, the closed-world assumption is employed, i.e. the knowledge base is not only assumed to be sound but moreover to be complete. Everything that can not be derived from the knowledge is assumed to be false. This could be easily modelled in this framework by assuming the truth value 0 as "default" truth value, so to speak. Yet we want to pursue a slightly more general approach: arbitrary default truth values will be explicitly stated for each predicate. We even allow the definition of different default truth values for different arguments of a predicate. This is formalised as follows.

Definition. A *default value declaration* for a predicate $p \in \Pi^{(n)}$ is written as $\texttt{default}(p(X_1,\ldots,X_n)) =$

ISBN: 978-989-95079-6-8

$[\delta_1$ if $\varphi_1, \ldots, \delta_m$ if $\varphi_m]$ where $\delta_i \in [0,1]$ for all i. The φ_i are first-order formulas restricted to terms from $\mathrm{TU}_{\Sigma,\{X_1,\ldots,X_n\}}$, the predicates $=$ and $\neq$, the symbol true, and the junctors $\wedge$ and $\vee$ in their usual meaning. ⌟

Example (continued). Let us add the following default value declarations to the knowledge base and thus close the mentioned gaps.

$$\texttt{default}(\mathsf{nice\text{-}weather(X)}) = 0.5,$$
$$\texttt{default}(\mathsf{many\text{-}sights(X)}) = 0.2,$$
$$\texttt{default}(\mathsf{good\text{-}destination(X)}) = 0.3$$

They could be interpreted as: when visiting an arbitrary city of which nothing further is known, it is likely that you have nice weather but you will less likely find many sights. Irrespective of this, it will only to a small extent be a good travel destination.

To model the fact that a city is not on a continent unless stated otherwise, we add another default value declaration for city-continent: $\texttt{default}(\mathsf{city\text{-}continent(X, Y)}) = 0.0$. Notice that in this example $m = 1$ and $\varphi_1 = \mathsf{true}$ for all the default value declarations. ◇

The default values allow our knowledge base to answer arbitrary questions about predicates that occur in it. But will the answers always make sense? To stay in the above example, if we ask a question like "What is the truth value of nice-weather(australia)?" we will get the answer "0.5" which does not make too much sense since Australia is not a city, but a continent.

To address this issue, we introduce types into the language. Types can be viewed as inherent properties of terms – each term can have zero or more types. We use them to restrict the domains of predicates.

Definition. A *term type declaration* assigns a type $\tau \in \mathcal{T}$ to a term $t \in H$ and is written as $t : \tau$. A *predicate type declaration* assigns a type $(\tau_1, \ldots, \tau_n) \in \mathcal{T}^n$ to a predicate $p \in \Pi^n$ and is written as $p : (\tau_1, \ldots, \tau_n)$, where τ_i is the type of p's i-th argument. ⌟

Example (continued). Using the set of types $\mathcal{T} = \{\mathsf{City}, \mathsf{Continent}\}$, we add some term type declarations to our knowledge base:

$$\mathsf{madrid} : \mathsf{City}, \mathsf{istanbul} : \mathsf{City},$$
$$\mathsf{sydney} : \mathsf{City}, \mathsf{moscow} : \mathsf{City};$$
$$\mathsf{africa} : \mathsf{Continent}, \mathsf{america} : \mathsf{Continent},$$
$$\mathsf{antarctica} : \mathsf{Continent},$$
$$\mathsf{asia} : \mathsf{Continent}, \mathsf{europe} : \mathsf{Continent}.$$

We also type the predicates in the obvious way:

$$\mathsf{nice\text{-}weather} : (\mathsf{City}),$$
$$\mathsf{many\text{-}sights} : (\mathsf{City}),$$
$$\mathsf{good\text{-}destination} : (\mathsf{City}),$$
$$\mathsf{city\text{-}continent} : (\mathsf{City}, \mathsf{Continent}).$$

◇

For a ground atom $A = p(t_1, \ldots, t_n) \in B$ we say that it is *well-typed with respect to* T iff $p : (\tau_1, \ldots, \tau_n) \in T$ implies $\tau_i \in \mathfrak{t}_T(t_i)$ for all i.

For a ground clause $A \xleftarrow{c,F_c}_F B_1, \ldots, B_n$ we say that it is well-typed w.r.t. T iff all B_i are well-typed for $1 \leq i \leq n$ implies that A is well-typed (i.e. if the clause preserves well-typing). We say that a non-ground clause is well-typed iff all its ground instances are well-typed.

Example (continued). With respect to the given type declarations, city-continent(moscow, antarctica) is well-typed whileas city-continent(asia, europe) is not. ◇

A *fuzzy logic program* P is a triple $P = (R, D, T)$ where R is a set of fuzzy clauses, D is a set of default value declarations, and T is a set of type declarations.

From now on, when speaking about programs, we will implicitly assume the signature Σ to consist of all function symbols occurring in P, the signature Π to consist of all the predicate symbols occurring in the program, the set $\mathcal{T}$ to consist of all types occurring in type declarations in T, and the signature Ω of all the aggregation operator symbols. For Ω we will furthermore require that all operators from $\hat{\Omega}$ be monotonic.

Lastly, we introduce the important notion of a "well-defined" program.

Definition. A fuzzy logic program $P = (R, D, T)$ is called *well-defined* iff

- for each predicate symbol p/n occurring in R, there exist both a predicate type declaration and a default value declaration;
- all clauses in R are well-typed;
- for each default value declaration $\texttt{default}(p(X_1, \ldots, X_n)) = [\delta_1$ if $\varphi_1, \ldots, \delta_m$ if $\varphi_m]$, the formulas φ_i are pairwise contradictory and $\varphi_1 \vee \cdots \vee \varphi_m$ is a tautology, i.e. exactly one default truth value applies to each element of p/n's domain.

⌟

3 Operational Semantics

The possibility to define default truth values for predicates offers us a great deal of flexibility and expressivity. But it also has its drawbacks: reasoning with defaults is inherently non-monotonic – we might have to withdraw some conclusions that have been made in an earlier stage of execution. To capture this formally, we attach to each truth value an attribute that indicates how this value has been concluded. There are 3 different cases of how a truth value can be determined:

- exclusively by application of program facts and clauses, represented by the symbol ▼ denoting the attribute value *safe*,
- by indirect use of default values, represented by the symbol ♦ denoting the attribute value *unsafe (mixed)*, or
- directly via a default value declaration, represented by the symbol ▲ denoting the attribute value *unsafe (pure)*.

We need to be able to compare the attributes (in order to be able to prefer one conclusion over another) and to combine them to keep track of default value usage in the course of

ISBN: 978-989-95079-6-8

computation. This is formalised by setting the ordering $<_a$ on truth value attributes such that $\blacktriangle <_a \blacklozenge <_a \blacktriangledown$.

The operator $\circ : \{\blacktriangle, \blacklozenge, \blacktriangledown\} \times \{\blacktriangle, \blacklozenge, \blacktriangledown\} \to \{\blacktriangle, \blacklozenge, \blacktriangledown\}$ is then defined as:

$$x \circ y := \begin{cases} \blacktriangledown & \text{if } x = y = \blacktriangledown \\ \blacktriangle & \text{if } x = y = \blacktriangle \\ \blacklozenge & \text{otherwise} \end{cases}$$

The operator $\circ$ is designed to keep track of attributes during computation: only when two "safe" truth values are combined, the result is known to be "safe", in all other cases it is "unsafe". It should be noted that "$\circ$" is monotonic.

The truth values that we use in the description of the semantics will be real values $v \in [0, 1]$ with an attribute (i.e. a $z \in \{\blacktriangle, \blacklozenge, \blacktriangledown\}$) attached to it. We will write them as zv. The ordering $\preccurlyeq$ on the truth values will be the lexicographic product of $<_a$, the ordering on the attributes, and the standard ordering $<$ of the real numbers. The set of truth values is thus totally ordered as follows:

$$\bot \prec \blacktriangle 0 \prec \cdots \prec \blacktriangle 1 \prec \blacklozenge 0 \prec \cdots \prec \blacklozenge 1 \prec \blacktriangledown 0 \prec \cdots \prec \blacktriangledown 1.$$

A *valuation* $\sigma : V \to B$ is an assignment of ground terms to variables. Each valuation σ uniquely constitutes a mapping $\hat{\sigma} : \mathrm{TU}_{\Sigma,V} \to B$ that is defined in the obvious way.

A *fuzzy Herbrand interpretation* (or short, *interpretation*) of a fuzzy logic program is a mapping $I : B \to \mathbb{T}$ that assigns truth values to ground atoms.

The *domain* of an interpretation is the set of all atoms to which a "proper" truth value is assigned: $\mathsf{Dom}(I) := \{A \mid A \in B, I(A) \succ \bot\}$.

For two interpretations I and J, we say I *is less than or equal to* J, written $I \sqsubseteq J$, if $I \sqsubseteq J$ iff $I(A) \preccurlyeq J(A)$ for all $A \in B$.

Accordingly, the infimum (or intersection) and supremum (or union) of interpretations are, for all $A \in B$, defined as $(I \sqcap J)(A) := \min(I(A), J(A))$ and $(I \sqcup J)(A) := \max(I(A), J(A))$.

The pair $(\mathcal{I}_P, \sqsubseteq)$ of the set of all interpretations of a given program with the interpretation ordering forms a complete lattice. This follows readily from the fact that the underlying truth value set $\mathbb{T}$ forms a complete lattice with the truth value ordering $\preccurlyeq$.

Definition. [Model] Let $P = (R, D, T)$ be a fuzzy logic program.

For a clause $r \in R$ we say that I *is a model of the clause* r and write

$$I \Vdash A \xleftarrow{c,F_c}_F B_1, \ldots, B_n$$

iff for all valuations σ, we have: if $I(\sigma(B_i)) = z_i v_i \succ \bot$ for all i, then $I(\sigma(A)) \succcurlyeq z'v'$ where $z' = z_1 \circ \cdots \circ z_n$ and $v' = \hat{F}_c(c, \hat{F}(v_1, \ldots, v_n)))$.

For a default value declaration $d \in D$ we say that I *is a model of the default value declaration* d and write

$$I \Vdash \mathtt{default}(p(X_1, \ldots, X_n)) = [\delta_1 \text{ if } \varphi_1, \ldots, \delta_m \text{ if } \varphi_m]$$

iff for all valuations σ, we have: if $\sigma(p(X_1, \ldots, X_n))$ is well-typed (w.r.t. T), then there exists an $1 \le j \le m$ such that $\sigma(\varphi_j)$ holds and $I(\sigma(p(X_1, \ldots, X_n))) \succcurlyeq \blacktriangle \delta_j$.

We write $I \Vdash R$ if $I \Vdash r$ for all $r \in R$ and similarly $I \Vdash D$ if $I \Vdash d$ for all $d \in D$.

Finally, we say that I *is a model of the program* P and write $I \Vdash P$ iff $I \Vdash R$ and $I \Vdash D$. ⌟

The operational semantics will be formalized by a transition relation that operates on (possibly only partially instantiated) computation trees. Here, we will not need to keep track of default value attributes $\{\blacktriangle, \blacklozenge, \blacktriangledown\}$ explicitly, it will be encoded into the computations.

Definition. Let Ω be a signature of aggregation operator symbols and W a set of variables with $W \cap V = \varnothing$.

A *computation node* is a pair $\langle A, e \rangle$, where $A \in \mathrm{TB}_{\Pi,\Sigma,V}$ and e is a term over $[0, 1]$ and W with function symbols from Ω. We say that a computation node is *ground* if e does not contain variables. A computation node is called *final* if $e \in [0, 1]$. A final computation node will be indicated as $\underline{\langle A, e \rangle}$.

We distinguish two different types of computation nodes: C-nodes, that correspond to applications of program clauses, and D-nodes, that correspond to applications of default value declarations.

A *computation tree* is a directed acyclic graph whose nodes are computation nodes and where any pair of nodes has a unique (undirected) path connecting them. We call a computation tree ground or final if all its nodes are ground or final respectively.

For a given computation tree t we define the *tree attribute*

$$z_t = \begin{cases} \blacktriangledown & \text{if } t \text{ contains no D-node} \\ \blacklozenge & \text{if } t \text{ contains both C- and D-nodes} \\ \blacktriangle & \text{if } t \text{ contains only D-nodes} \end{cases}$$

⌟

Computation nodes are essentially generalizations of queries that keep track of aggregation operator usage.

Computation trees as defined here should not be confused with the usual notion of SLD-trees. While SLD-trees describe the whole search space for a given query and thus give rise to different derivations and different answers, computation trees describe just a state in a single computation.

The computation steps that we perform on computation trees will be modelled by a relation between computation trees.

Definition. [Transition relation] For a given fuzzy logic program $P = (R, D, T)$, the transition relation $\rightarrowtriangle$ is characterized by the following transition rules:

Clause: $t/\left[\boxed{\langle A', v \rangle}\right] \rightarrowtriangle$

$$t/\left[\boxed{\langle A', v \rangle} \quad / \quad \begin{array}{c} \mathsf{C}\langle A', F_c(c, F(v_1, \ldots, v_n)) \rangle \\ \diagup \quad \diagdown \\ \langle B_1, v_1 \rangle \ \cdots \ \langle B_n, v_n \rangle \end{array}\right] \mu$$

If there is a (variable disjoint instance of a) program clause $A \xleftarrow{c,F_c}_F B_1, \ldots, B_n \in R$ and $\mu = \mathrm{mgu}(A', A)$. (Take a non-final leaf node and add child nodes according to a program clause; apply the most general unifier of

ISBN: 978-989-95079-6-8

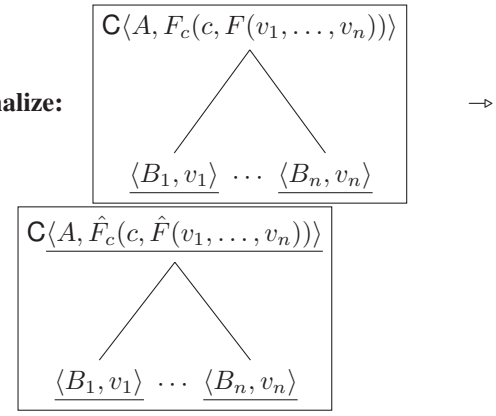
the node atom and the clause head to all the atoms in the tree.)

Note that we immediately finalize a node when applying this rule for a fuzzy fact.

Default: $\boxed{t\,[\langle A, x\rangle]} \quad \rightarrowtriangle \quad \boxed{t\left[\langle A, x\rangle / \mathsf{D}\underline{\langle A, \delta_j\rangle}\right]\mu}$

If A does not match with any program clause head, there is a default value declaration $\texttt{default}(p(X_1, \ldots, X_n)) = [\delta_1 \text{ if } \varphi_1, \ldots, \delta_m \text{ if } \varphi_m] \in D$, μ is a substitution such that $p(X_1, \ldots, X_n)\mu = A\mu$ is a well-typed ground atom, and there exists a $1 \leq j \leq m$ such that $\varphi_j\mu$ holds. (Apply a default value declaration to a non-final leaf node thus finalizing it.)

Finalize:



(Take a non-final node whose children are all final and replace its truth expression by the corresponding truth value.)

⌟

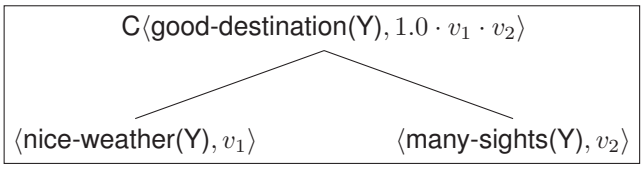
Here, the notation $t[A]$ means "the tree t that contains the node A somewhere". Likewise, $t[A/B]$ is to be read as "the tree t where the node A has been replaced by the node B".

Asking the query $\langle A, v\rangle$ corresponds to applying the transition rules to the initial computation tree $\langle A, v\rangle$. The computation ends *successfully* if a final computation tree is created, the truth value of the instantiated query can then be read off the root node. We will illustrate this with an example computation.

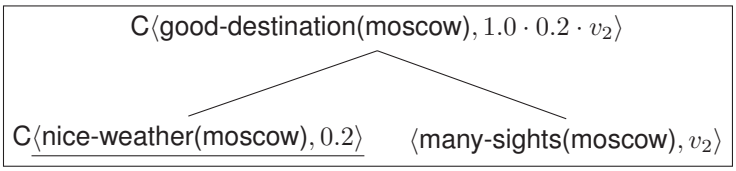
Example (continued). We start with the tree

$$\boxed{\langle \text{good-destination(Y)}, v\rangle}.$$

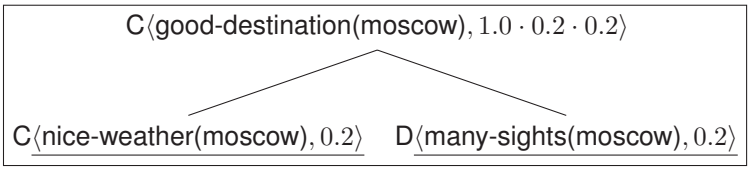
Applying the **Clause**-transition to the initial tree with the program clause good-destination(X) $\xleftarrow{1.0,\cdot}$ nice-weather(X), many-sights(X) yields



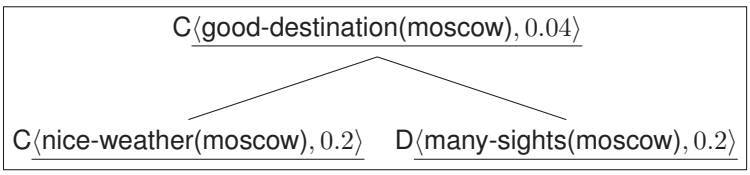
Now we apply **Clause** to the left child with nice-weather(moscow) ← 0.2:



Since there exists no clause whose head matches many-sights(moscow), we apply the **Default**-rule for many-sights to the right child.



In the last step, we finalise the root node.



The calculated truth value for good-destination(moscow) is thus 0.04. ◇

The actual operational semantics is now given by the truth values that can be derived in the defined transition system. This "canonical model" can be seen as a generalisation of the success set of a program.

Definition. Let P be a well-defined fuzzy logic program. The canonical model of P for $A \in B$ is defined as follows:

$$\mathsf{cm}(P) := \left\{ A \mapsto z_t v \;\middle|\; \begin{array}{l}\text{there exists a computation starting}\\ \text{with } \langle A, w\rangle \text{ and ending with a fi-}\\ \text{nal computation tree } t \text{ with root node}\\ \underline{\langle A, v\rangle}\end{array} \right\}$$

⌟

It can be verified that the canonical model $\mathsf{cm}(P)$ is indeed a model of P.

4 Implementation

RFuzzy is implemented as a package of the Ciao Prolog System [15]. It consists essentially of a set of rules that translate the RFuzzy Syntax to ANSI Prolog using the expansion of code of the packages in Ciao Prolog. The predicates of the program that have been declared as fuzzy get an additional argument that makes the truth value explicit. The resulting program can then be interpreted and executed "as usual".

Example (continued). We have no space to describe the implementation syntax used at RFuzzy, but it can be easily deduced from the implementation of our running example (not all the clauses are shown here).

```
nice_weather(madrid) value 0.8.
nice_weather(moscow) value 0.2.
```

ISBN: 978-989-95079-6-8

```
many_sights(madrid) value 0.6.
many_sights(sydney) value 0.6.

good_destination(X) cred (prod,1.0):~ prod
        nice_weather(X),
        many_sights(X).
```

The default value declarations are also very similar to the abstract syntax.

```
:- default(nice_weather/1, 0.5).
:- default(many_sights/1, 0.2).
:- default(good_destination/1, 0.3).
```

We use crisp predicates to represent types.

```
city(madrid).
city(moscow).
city(sydney).

:- set_prop(nice_weather/1) => city/1.
:- set_prop(many_sights/1) => city/1.
:- set_prop(good_destination/1) => city/1.
```

To ask queries to the system, we add a variable that is going to be instantiated with the truth value.

```
?- good_destination(moscow, V).

V = 0.04 ?
yes
```

But we cannot only ask for truth values of fully instantiated atoms. The real power of RFuzzy lies in the ability to provide constructive answers. For example, if we want to know "What is the best travel destination according to the knowledge base?" we just ask the following query. It looks for a destination `D` with a truth value `V` for which no destination with a higher truth value `V1` exists.

```
?- good_destination(D, V),
   \+ (good_destination(_, V1),
       V1 > V).

D = madrid,
V = 0.48 ?

yes
```

As we see, the system returns "Madrid" as best destination and thus answers a question with an object rather than a truth value. ◇

5 Conclusions and Future Work

We presented the operational semantics of the RFuzzy framework for Fuzzy Logic Programming and showed some features of the implementation [1] via an example. We finally remark that a least model semantics and a least fixpoint semantics for RFuzzy also have been defined and proven equivalent to the operational semantics shown here.

[1] A complete release of the implementation is available at `http://babel.ls.fi.upm.es/software/rfuzzy`

References

[1] P. Vojtas. Fuzzy logic programming. *Fuzzy Sets and Systems*, 124(1):361–370, 2001.

[2] R. C. T. Lee. Fuzzy Logic and the resolution principle. *Journal of the Association for Computing Machinery*, 19(1):119–129, 1972.

[3] M. Ishizuka and N. Kanai. Prolog-ELF incorporating fuzzy Logic. In *IJCAI*, pages 701–703, 1985.

[4] J. F. Baldwin, T. P. Martin, and B. W. Pilsworth. *Fril: Fuzzy and Evidential Reasoning in Artificial Intelligence*. John Wiley & Sons, 1995.

[5] D. Li and D. Liu. *A Fuzzy Prolog Database System*. John Wiley & Sons, New York, 1990.

[6] Z. Shen, L. Ding, and M. Mukaidono. Fuzzy resolution principle. In *Proc. of 18th International Symposium on Multiple-valued Logic*, volume 5, 1989.

[7] C. Vaucheret, S. Guadarrama, and S. Munoz-Hernandez. Fuzzy prolog: A simple general implementation using clp(r). In P.J. Stuckey, editor, *Int. Conf. in Logic Programming, ICLP 2002*, number 2401 in LNCS, page 469, Copenhagen, Denmark, July/August 2002. Springer-Verlag.

[8] S. Guadarrama, S. Munoz-Hernandez, and C. Vaucheret. Fuzzy Prolog: A new approach using soft constraints propagation. *Fuzzy Sets and Systems, FSS*, 144(1):127–150, 2004. ISSN 0165-0114.

[9] S. Munoz-Hernandez, C. Vaucheret, and S. Guadarrama. Combining crisp and fuzzy Logic in a prolog compiler. In J. J. Moreno-Navarro and J. Mariño, editors, *Joint Conf. on Declarative Programming: APPIA-GULP-PRODE 2002*, pages 23–38, Madrid, Spain, September 2002.

[10] H. T. Nguyen and E. A. Walker. *A first Course in Fuzzy Logic*. Chapman & Hall/Crc, 2000.

[11] J. Medina, M. Ojeda-Aciego, and P. Votjas. Multi-adjoint Logic Programming with continuous semantics. In *LPNMR*, volume 2173 of *LNCS*, pages 351–364, Boston, MA (USA), 2001. Springer-Verlag.

[12] J. Medina, M. Ojeda-Aciego, and P. Votjas. A procedural semantics for multi-adjoint Logic Programming. In *EPIA*, volume 2258 of *LNCS*, pages 290–297, Boston, MA (USA), 2001. Springer-Verlag.

[13] J. Medina, M. Ojeda-Aciego, and P. Votjas. A completeness theorem for multi-adjoint Logic Programming. In *International Fuzzy Systems Conference*, pages 1031–1034. IEEE, 2001.

[14] P.J. Morcillo and G. Moreno. Floper, a fuzzy logic programming environment for research. In *Proceedings of the Spanish Conference on Programming and Computer Languages, PROLE 2008*, Gijn, Spain, 2008.

[15] M. Hermenegildo, F. Bueno, D. Cabeza, M. Carro, M. García de la Banda, P. López-García, and G. Puebla. The CIAO Multi-Dialect Compiler and System: An Experimentation Workbench for Future (C)LP Systems. In *Parallelism and Implementation of Logic and Constraint Logic Programming*, pages 65–85. Nova Science, Commack, NY, USA, April 1999.

ISBN: 978-989-95079-6-8

Applying Bacterial Memetic Algorithm for Training Feedforward and Fuzzy Flip-Flop based Neural Networks

László Gál[1,2] János Botzheim[3,4] László T. Kóczy[1,4] Antonio E. Ruano[5]

1 Institute of Information Technology and Electrical Engineering,
Széchenyi István University, Győr, Hungary
2 Department of Technology, Informatics and Economy,
University of West Hungary Szombathely, Hungary
3 Department of Automation Széchenyi István University Győr, Hungary
4 Department of Telecommunication and Media Informatics
Budapest University of Technology and Economics, Budapest, Hungary
5 Centre for Intelligent Systems, FCT, University of Algarve, Portugal
Email: gallaci@ttmk.nyme.hu, {botzheim, koczy}@{sze, tmit.bme}.hu, aruano@ualg.pt

Abstract—*In our previous work we proposed some extensions of the Levenberg-Marquardt algorithm; the Bacterial Memetic Algorithm and the Bacterial Memetic Algorithm with Modified Operator Execution Order for fuzzy rule base extraction from input-output data. Furthermore, we have investigated fuzzy flip-flop based feedforward neural networks. In this paper we introduce the adaptation of the Bacterial Memetic Algorithm with Modified Operator Execution Order for training feedforward and fuzzy flip-flop based neural networks. We found that training these types of neural networks with the adaptation of the method we had used to train fuzzy rule bases had advantages over the conventional earlier methods.*

Keywords— Bacterial Memetic Algorithm, Fuzzy Flip-Flop, Levenberg-Marquardt method, Neural Network.

1 Introduction

Bacterial type evolutionary algorithms are inspired by the biological bacterial cell model [1,2]. The *Bacterial Memetic Algorithm* (BMA) is a recent method for fuzzy rule base extraction from input-output data for a certain system [7]. We have investigated its properties intensely and found some points where its performance in the fuzzy rule base identification could be improved. The recent bacterial type algorithms we proposed were named *Bacterial Memetic Algorithm with Modified Operator Execution Order* (BMAM), *Improved Bacterial Memetic Algorithm* (IBMA) and *Modified Bacterial Memetic Algorithm* (MBMA) [3,4]. They are both memetic algorithms and apply alternatively global and local search for identifying fuzzy rule bases from input-output data automatically when no human expert to define the rules is available.

Neural Networks belong to the Soft Computing area like Fuzzy Systems and Evolutionary Computing. They can be used for modeling a certain system where input-output data pairs exist. The neural networks are inspired by biological phenomena: the brain itself and other parts of the neural system.

Fuzzy Flip-Flops are extended forms of the binary flip-flops that are widely used in digital technics [5]. They use fuzzy logic operations instead of Boolean logic ones and require fuzzy inputs, furthermore they produce fuzzy outputs instead of digital values.

Our previous works were developing the *Bacterial Memetic Algorithm* applied for *fuzzy rule base identification* (FRBI) and investigating various types of *Fuzzy Flip-Flops* (F^3) used in *feedforward neural networks* (FFNN) as replacements of the neurons [6]. We trained the *Fuzzy Flip-Flop based Neural Networks* (FNN) with the *Levenberg-Marquardt* (LM) based training method as it is a widely used and accepted one. However, we faced the same problems with the LM based feedforward neural network training as in the fuzzy rule base identification. Therefore we have adopted the *Bacterial Memetic Algorithm with Modified Operator Execution Order* (BMAM) for training Neural Networks. Our goal was to improve the learning capabilities of feedforward neural networks with a bacterial type evolutionary approach.

In this paper we propose the adaptation of the BMAM for training feedforward neural networks, and we study and evaluate the respective results. From another aspect another paper was proposed here where we report on the findings of our investigations of the properties of different types of FNNs trained with BMAM [14].

2 Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM)

2.1 Bacterial Memetic Algorithm (BMA)

Bacterial Memetic Algorithm (BMA) is a very recent approach used for fuzzy rule base identification (FRBI) [7]. It combines global and local search. For the global search it uses bacterial type evolutionary approach and for the local search the *Levenberg-Marquardt* method is deployed. Previous work confirmed that the *Pseudo-Bacterial Genetic Algorithm* (PBGA) and the *Bacterial Evolutionary Algorithm* (BEA) were rather more successful in this area than the conventional genetic algorithms [1,2].

2.1.1 Bacterial mutation

PBGA is a special kind of *Genetic Algorithm* (GA) [8], it introduces a new "genetic" operation called *bacterial mutation*. For the algorithm, the first step is to determine how the problem can be encoded in a *bacterium* (*chromosome*). In case of modelling fuzzy systems the

ISBN: 978-989-95079-6-8

parameters of the model – all the breakpoints of the rule base – have to be encoded in the *chromosome*. The next step is to generate initial *bacteria* randomly. Then an optimization process is started utilizing mainly the *bacterial mutation*, in order to refine the model parameters.

The *bacterial mutation* operation tries to improve the parts of the *chromosomes*. Therefore each individual (*bacterium)* is selected one by one and a number of copies of the selected individual (*clones*) are created. Then the same part or parts are randomly chosen from all clones and it (they) is (are) mutated (except one single clone that remains unchanged). Mutation means to replace the part with a random value in a specified range.

After the mutations all the clones are evaluated (SSE, MSE, BIC) and the best clone is selected whose mutated part or parts are transferred to the other clones. Theoretically, this operation copies just a few parameters from one clone to the other clones (*gene transfer*), but in practice, the other clones will not differ from the best clone at the end at all. So, this operation can be done with discharging all the clones except the best one and then cloning further the best clone.

After the selection of the best clone and transferring its mutated part or parts to the other clones the above procedure is repeated until all the parts are mutated exactly once. The final best clone is remaining in the population and all the other clones are destroyed. The *bacterial mutation* cycle is done on the other individuals in the population e.g. in a parallel processing way.

At the end of the *complete bacterial mutation cycle* a new generation of bacteria is arisen.

The *Bacterial Evolutionary Algorithm* (BEA) is based on the PBGA supported by a new genetic operation called *gene transfer* operation. This operation can play an important role in the FRBI process as it establishes relationships among the individuals of the population (useful in somewhat changing environment) and is able to increase or decrease the number of the rules in a fuzzy rule base (useful in determining the appropriate size of the fuzzy rule base). Because this behaviour is not exploited in our investigations when training neural networks this operation is not described in detail.

2.1.2 The Levenberg-Marquardt method (LM)

The *Levenberg-Marquardt* (LM) method [9] is a gradient based iterative procedure. It is used for least squares curve fitting for a given set of empirical data, minimizing the *sum of squared error* function (SSE). It can be used for fuzzy rule extraction alone, but it generates only locally optimal rule base in the neighbourhood of the initial rules.

The *Error Back Propagation* algorithm (EBP or BP) was a great improvement in neural network research, but it has weak convergence rate. The LM algorithm is more complex and requires more computational effort than the BP, but it has much better convergence rate properties. Therefore the LM algorithm is one of the most popular training functions for feedforward back propagation networks.

2.1.3 Bacterial Memetic Algorithm (BMA)

Memetic algorithms combine evolutionary and local search methods [10]. The evolutionary part is able to find the global optimum region, but is not suitable to find the accurate minimum in practice. The gradient based part is able to reach the accurate optimum, but is very sensitive to the initial position in the search space and is unable to avoid the local optimum. Combining global and local search is expected to be beneficial.

Bacterial Memetic Algorithm (BMA) combines the *Bacterial Evolutionary Algorithm* (BEA) and the *Levenberg-Marquardt* (LM) method. In the past we used it for fuzzy rule base extraction, among others.

The BMA integrates its two components, the BEA and the LM method in the following way:

1. *Bacterial Mutation* operation for each individual,
2. *Levenberg-Marquardt* method for each individual,
3. *Gene Transfer* operation for a partial population.

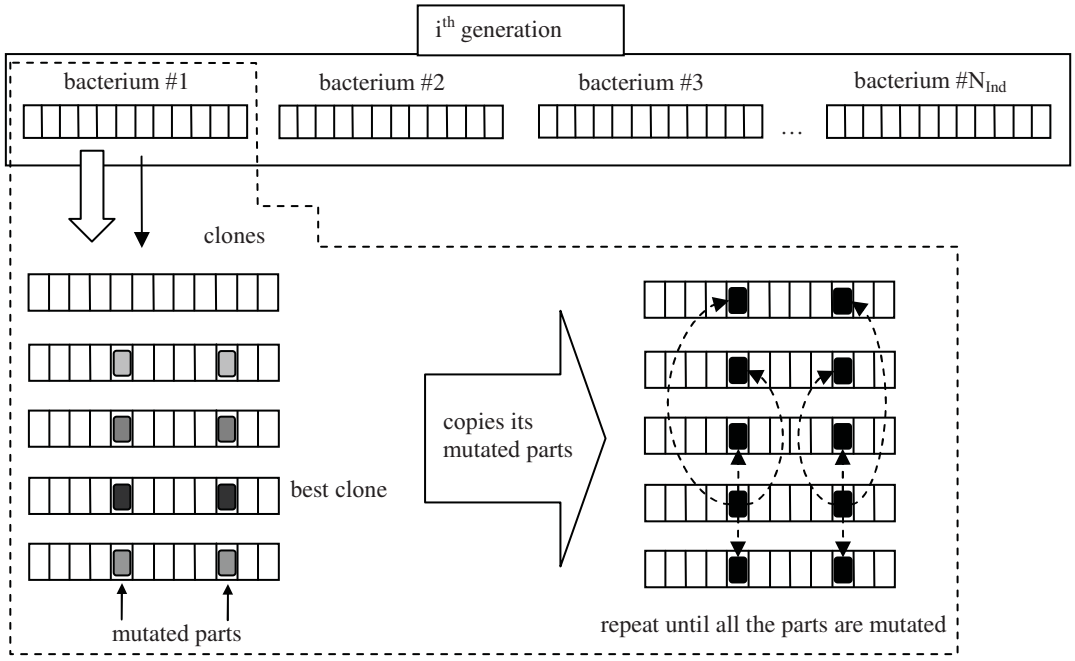


Figure 1: Bacterial mutation (one individual)

ISBN: 978-989-95079-6-8

This way the LM method is nested into the BEA, so that local search is done for every global search cycle.

2.2 Bacterial Memetic Algorithm with Modified Operator Execution Order (BMAM)

Although BMA provides a very good speed of convegrenve towards the optimal model parameters there are some points of the algorithm where the performance could be increased. We proposed new techniques to improve its performance. Some of them contain modifications that are not useful in training FFNNs (handling the *knot order violation* that can occur in applying LM for FRBI) (IBMA, MBMA) [3]. Another improvement to BMA is the *Bacterial Memetic Algorithm with Modified Operator Execution Order* [4] which exploits the *Levenberg-Marquardt* method more efficiently.

Instead of applying the LM cycle *after* the *bacterial mutation* as a separate step, the modified algorithm executes several LM cycles *during* the *bacterial mutation* after *each mutational* step.

The *bacterial mutation* operation changes one or more parameters of the modeled system randomly, and then it checks whether the model obtained by this way performs better than the previous models or the models that have been changed concurrently this way in the other clones. The mutation test cycle is repeated until all the parameters of the model have gone through the bacterial mutation.

In the mutational cycle it is possible to gain a temporary model that has an instantaneous fitness value that is worse than the one in the previous or the concurrent models. However, it is potentially better than those, because it is located in such a region of the search space that has a better local optimum than the other models do. In accordance to this, if some Levenberg-Marquardt iterations are executed after *each* bacterial mutational step, the test step is able to choose some potentially valued clones that could be lost otherwise.

In the *Bacterial Memetic Algorithm with Modified Operator Execution Order*, after *each mutational step* of *every single bacterial mutation iteration* several LM iterations are done. Several tests have shown it is enough to run just 3 to 5 of LM iterations per mutation to improve the performance of the whole algorithm. The usual test phase of the *bacterial mutation* operation follows after the LM iterations. After the complete *modified bacterial mutation* follows the LM method that is used in the original BMA, where more, e.g. 10 iterational steps, are done with all the individuals of the population towards reaching the local optimum. After all this the *gene transfer* operation is executed if needed.

3 Fuzzy flip-flops (F^3)

Fuzzy flip-flops are extended forms of binary flip-flops used in the conventional digital technics. We have dealt with the fuzzy extensions (complements) of the binary J-K and D flip-flops. Various types of fuzzy flip-flops are implemented and tested (set, reset type and the general type using the unified equation; J-K, D and Choi type D; based on minmax, algebraic, Yager, Dombi, Hamacher and Frank t-norms and co-norms, resp.) [11]. Because of an interesting property some fuzzy flip-flops can be used for implementing a sigmoid like transfer function and so constructing Multilayer Perceptron Neural Networks.

In our previous works we studied the behavior of various type fuzzy flip-flops, illustrating their characteristics by their respective graphs. We proposed also the concept of *fuzzy flip-flop based neural networks* and investigated their function approximation capabilities [6, 12].

4 Fuzzy flip-flop based feedforward neural networks (FNN)

In our team extensive research was done with the leadership of R. Lovassy in the field of fuzzy flip-flops. As we mentioned it before, various fuzzy norms can be used for building *fuzzy flip-flop based neural networks* (FNNs). The basic idea was to substitute the fuzzy flip-flops with sigmoidal transfer function instead of traditional neurons. The flip-flops are based on various norms, consequently, their transfer functions have different slopes. Fixing the value of the present state Q (in the characteristical equation), often we obtained "good" enough sigmoidal transfer function character [6]. First of all, to train this kind of neural network with a usual training method BP or LM the derivatives of these transfer functions have to be also calculated. Then the FNN can be used and trained in the usual way. We found that the FNNs we created had good approximation properties. [12].

5 Training feedforward neural networks

With an appropriate transfer function and its derivative the *Error Back Propagation* algorithm (BP) can be used for training feedforward neural networks (FFNN). However, it has weak convergence rates.

The LM algorithm is more complex and requires more computational effort than the BP one, but it has much better convergence rates. The LM algorithm is one of the most popular training methods for feedforward neural networks despite of its higher memory requirements and higher complexity.

The training of the FFNNs begins with the random generation of initial weights and biases. Then the training method selected is applied. An update vector is generated that has to be applied for the vector that contains the weights and biases.

When using BP or LM based training methods one faces the drawback of these local searchers described in the next section.

6 Using BMAM in training FFNNs

Although the LM method based training of the neural networks works much more efficiently than the BP based one it has all drawbacks of the local search methods. The training is very sensitive to the (parameter's) initial position of the search space. An inconveniently generated random parameter set with the initial weights and biases determines a hardly trainable neural network with a weak performance at the end of the LM method based training procedure. This is because the LM method is a local searcher and thus it is unable to avoid the local minima.

We decided to apply bacterial type evolutionary algorithms because they proved to be rather successful in our previous

ISBN: 978-989-95079-6-8

works, better than the other evolutionary approaches. We preferred the *Bacterial Memetic Algorithm with Modified Operator Execution Order* because it converged faster than the original BMA. (And contained not only FRBI related improvements, like IBMA and MBMA do.)

We did not implement here the *gene transfer* operation because it was not useful with the neural network training we have done (in not changing environments, there was no need to change the structure of the NN or the number of the neurons).

The detailed steps of the BMAM used for NN training are described below:

1. Create the initial population – neural networks with two hidden layers – and initialize the neural network's input, layer weights and biases randomly as before a usual LM training procedure. Each individual contains the weights and biases – the parameters of the model – encoded in the chromosome. In a 1-4-3-1 NN the number of the parameters to be encoded are 2*4+4*3+2*3+1 = 27 parameters per individual.
2. Apply the *modified bacterial mutation* for each individual.
    a. Each individual is selected one by one.
    b. N_{Clones} copies of the selected individual are created ("*clones*").
    c. Choose the same part or parts randomly from the clones and mutate it (except one single clone that remains unchanged during this mutation cycle).
    d. Run some conventional LM method based NN training iterations (3-5 epochs).
    e. Select the best clone (simulate and evaluate the NNs) and transfer *all* of its parts to the other clones.
    f. Repeat the part selection-mutation-*LM*-selection-transfer cycle until all the parts are mutated, improved and tested.
    g. The best individual is remaining in the population, all other clones are deleted.
    h. This process is repeated until all the individuals have gone through the *modified bacterial mutation*.
3. Apply the LM method based NN training to each individual (e.g. 10 epochs per individual per generation)
4. Repeat the procedure above from the *modified bacterial mutation* step until a certain termination criterion is satisfied (e.g. maximum number of generations = 20 generation).

The experimental setup was:

- General PC (2GHz), Windows XP, Matlab
- Test function: $f(x) = \frac{\sin(c_1 \cdot x) \cdot \sin(c_2 \cdot x)}{c_3} + c_4$
- $c_1 = 0.2, c_2 = 0.07, c_3 = 2, c_4 = 0.5$
- Number of individuals in the population: 5
- Number of clones: 5.

We tested the new training algorithm in two ways.

6.1 Test 1

In the first test group we applied the new BMAM based NN training. We created a 1-4-3-1 feedforward neural network with the usual sigmoid transfer function and with selected fuzzy flip-flop based neurons. Our goal was to investigate the improvement of BMAM based training over the conventional LM based one so four transfer functions were selected: sigmoid (tansig), Dombi Fuzzy D Flip-Flop (Dombi DF^3), Frank Fuzzy D Flip-Flop (Frank DF^3) and Frank Choi-type Fuzzy D Flip-Flop (Frank CDF^3).

Our goal was here to train a NN that is hard to be trained [12]. The number of neurons (4 and 3 in the hidden layers) was relatively low. It makes possible to recognise the performance improvement (MSE) better, to see that the model complexity may be reduced with the better training, and to avoid overfitting.

We ran 30-30 trainings for each case mentioned above. Then the maximum, minimum, median and mean MSE values calculated.

In our previous work we chose the median to characterize the trainability of the FFF based NNs because in case of training with the LM based way there were several unsuccessful trainings where the final model was unusable and produced too high MSE. That is why we had to analyse 30-100 runs. The best value (minimum MSE) was more or less randomly good so it could not be used as a reliable value for indicate the trainability. In case of the mean value a single one unsuccessful training deteriorates many very successful training results. With using the median this random extreme values could be avoided.

Table 1 and 2 contains the results of these tests. One can see that using BMAM result much more better quality models (lower MSE). Using BMAM results lover maximum MSE values than median or mean values of the LM based training respectively. Furthermore the median and mean MSE values of the BMAM based training are very close to the minimum values of the LM based training respectively.

Table 1: MSE values of LM based training

LM based	Max	Min	Median	Mean
Tansig	0.06452	1.2×10^{-5}	0.00712	0.01891
Dombi DF^3	0.11962	0.04263	0.05732	0.06045
Frank DF^3	0.06644	0.00459	0.04642	0.04159
Frank CDF^3	0.06645	0.00593	0.05486	0.04697

Table 2: MSE values of BMAM based training

BMAM	Max	Min	Median	Mean
Tansig	0.00180	3.14×10^{-7}	2.38×10^{-5}	0.00034
Dombi DF^3	0.03994	0.02342	0.03362	0.03294
Frank DF^3	0.01091	0.00187	0.00680	0.00716
Frank CDF^3	0.02519	0.00460	0.00833	0.00933

Figure 2 to 9 show MSE histograms of 30-30 runs with various transfer functions and training methods. One can see that if using BMAM there is no need to run 30 or 100 complete training cycles to gain an excellent quality model because every BMAM trained model have very low MSE value.

ISBN: 978-989-95079-6-8

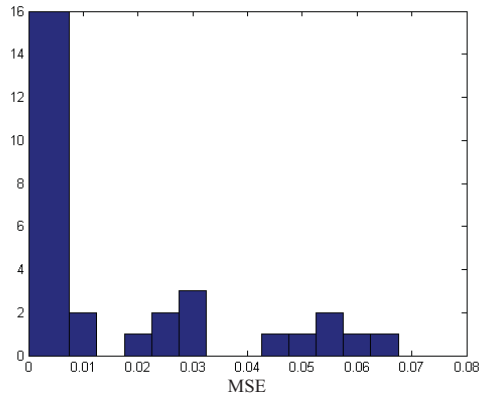


Figure 2: LM – Tansig NN histogram

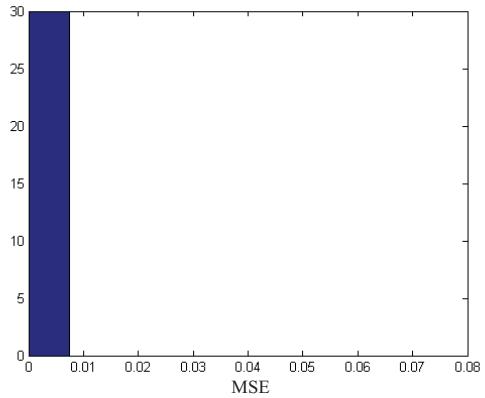


Figure 3: BMAM – Tansig NN histogram

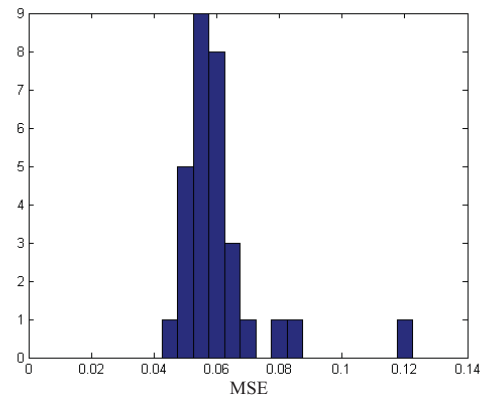


Figure 4: LM – Dombi DF^3 histogram

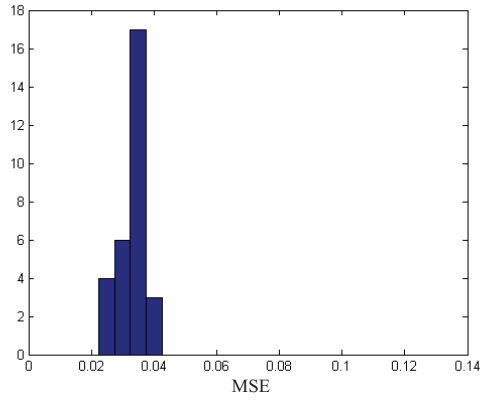


Figure 5: BMAM – Dombi DF^3 NN histogram

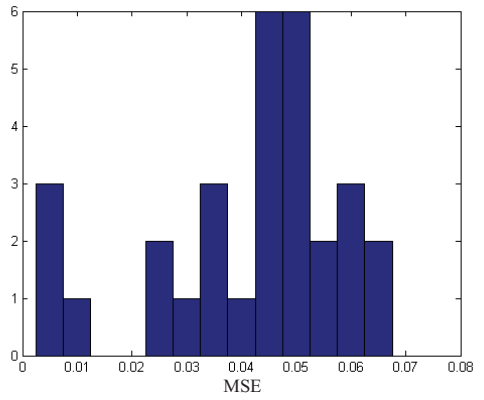


Figure 6: LM – Frank DF^3 NN histogram

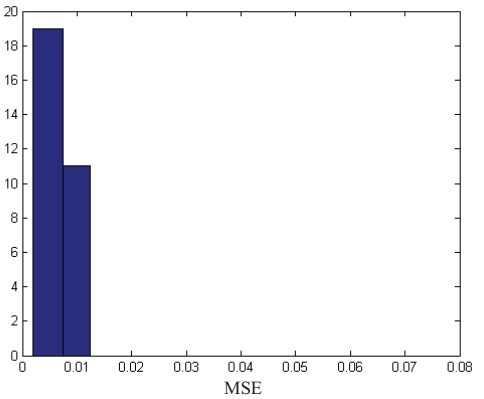


Figure 7: LM – Frank DF^3 histogram

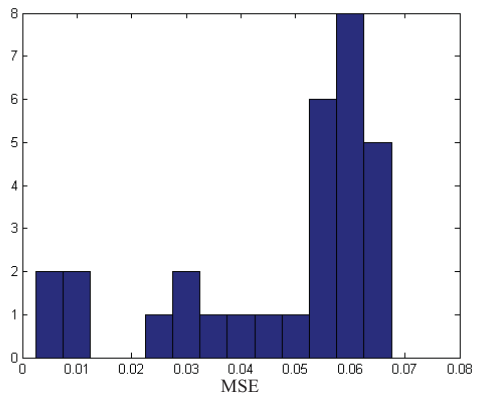


Figure 8: LM – Frank CDF^3 NN histogram

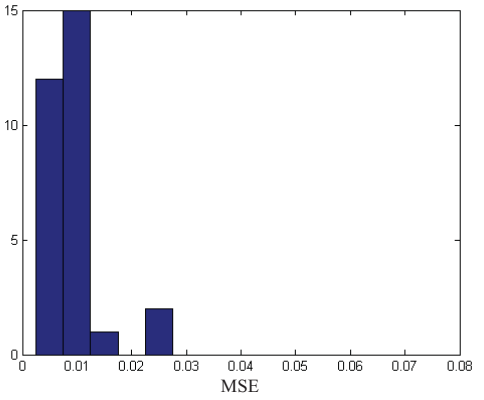


Figure 9: BMAM – Frank CDF^3 NN histogram

ISBN: 978-989-95079-6-8

6.2 Test 2

In the second group of tests we utilized the BMAM to identify quasi optimal parameter values of various types of fuzzy flip-flops used in FNNs. 1-8-8-1 neural networks were created because the good trainability was much more important than the lower complexity of the model here. This way the optimal parameter values are easier to identify.

Therefore we enhanced the capability of the BMAM based training method in a manner that the parameter (internally fixed Q values) of the fuzzy flip-flop used in Fuzzy Flip-Flop Neural Networks (FNN) can be encoded into the *chromosome*. This way it participates in the *bacterial mutation* cycle so the quasi optimal value of this parameter can be identified at the end of the BMAM based training.

Because this parameter is not affected by the LM training we applied two different versions of the *bacterial mutation* especially for this parameter. The first one is the original *bacterial mutation* (generate random values in the range of [0, 1]), while the second one increments or decrements the current fixed Q value with a very fine random step.

Table 2 shows the expected ranges and the quasi-optimal internally fixed Q values of several FNNs identified by the BMAM training method. The expected ranges were derived from our previous work [12].

Table 2: Expected ranges and fixed Q values by BMAM

Type of FNN	Expected range	Fixed Q value identified by BMAM
Algebraic JK FF	0 – 0.4	0.25
Algebraic D FF	~0.1,~0.5, ~0.9	0.91
Algebraic C D FF	<0.15, 0.4 – 0.6, >0.85	0.53
Dombi D FF	<0.1 or >0.9	0.924
Frank D FF	0.25 – 0.45	0.31

Further investigations will be focused on using BMAM based training method to identify of the other variable parameters of Yager, Dombi, Hamacher, Frank norms based FNNs.

7 Conclusions

In this paper we introduced the adaptation of *the Bacterial Memetic Algorithm with Modified Operator Execution Order* for *training feedforward neural networks*, especially *neural networks* built from *Fuzzy Flip-Fops* (F^3s).

We applied this new approach to training neural networks and *fuzzy flip-flop based neural networks*. Our goal was to get a quasi-optimal result with only a single one or a very low number of training sequences whose error does not exceed (or very rarely exceeds) an acceptable level. Despite the usual tradeoffs between the complexity and accuracy [13] this way there is no need to run a few hundred of training cycles to get an acceptable model.

Our tests have shown that BMAM used for training FFNNs and fuzzy flip-flop based FFNNs is a very successful tool. Although it requires more computational effort than the conventional training methods it produces a higher quality model (so the complexity of the model can be reduced) with only one training cycle.

Furthermore we enhanced the capability of the BMAM based training method in a manner that the parameter or parameters of the fuzzy flip-flop used in *Fuzzy Flip-Flop based Neural Networks* (FNN) can be encoded into the chromosome. This way it participates in the *bacterial mutation* cycle so the quasi optimal values of these parameters can be identified at the end of the BMAM based training.

Acknowledgment

This paper was supported by the Széchenyi University Main Research Direction Grant 2009, National Scientific Research Fund Grant OTKA T048832 and K75711, SEK Scientific Grant 2009, and the National Office for Research and Technology.

References

[1] Nawa, N. E., Hashiyama, T., Furuhashi, T. and Uchikawa, Y., *A study on fuzzy rules discovery using pseudo-bacterial genetic algorithm with adaptive operator*, Proceedings of IEEE Int. Conf. on Evolutionary Computation, ICEC'97, 1997.

[2] Nawa, N. E. and Furuhashi, T., *Fuzzy Systems Parameters Discovery by Bacterial Evolutionary Algorithms*, IEEE Transactions on Fuzzy Systems 7, 1999, pp. 608-616.

[3] Gál, L., Botzheim, J. and Kóczy, L. T., *Improvements to the Bacterial Memetic Algorithm used for Fuzzy Rule Base Extraction*, Computational Intelligence for Measurement Systems and Applications, CIMSA 2008, Istanbul, Turkey, 2008, pp. 38-43.

[4] Gál, L., Botzheim, J. and Kóczy, L. T., *Modified Bacterial Memetic Algorithm used for Fuzzy Rule Base Extraction*, 5th International Conference on Soft Computing as Transdisciplinary Science and Technology, CSTST 2008, Paris, France, 2008.

[5] K. Ozawa, K. Hirota and L. T. Kóczy, *Fuzzy flip-flop*, In: M. J. Patyra, D. M. Mlynek, eds., Fuzzy Logic. Implementation and Applications, Wiley, Chichester, 1996, pp. 197-236.

[6] R. Lovassy, L. T. Kóczy and L. Gál, *Multilayer Perceptron Implemented by Fuzzy Flip-Flops*, IEEE World Congress on Computational Intelligence, WCCI 2008, Hong Kong, pp. 1683-1688.

[7] Botzheim, J., Cabrita, C., Kóczy, L. T. and Ruano, A. E. , *Fuzzy rule extraction by bacterial memetic algorithm*, IFSA 2005, Beijing, China, 2005, pp.1563-1568.

[8] Holland, J. H., *Adaptation in Nature and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, The MIT Press, Cambridge, MA, 1992.

[9] Marquardt, D., *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, SIAM J. Appl. Math., 11, 1963, pp. 431-441.

[10] Moscato, P., *On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms*, Technical Report Caltech Concurrent Computation Program, Report. 826, California Institute of Technology, Pasadena, California, USA,1989.

[11] R. Lovassy, L. T. Kóczy and L. Gál, *Analyzing Fuzzy Flip-Flops Based on Various Fuzzy Operations*, Acta Technica Jaurinensis Series Intelligentia Computatorica vol. 1, no. 3, 2008, pp. 447-465.

[12] R. Lovassy, L. T. Kóczy and L. Gál, *Function Approximation Capability of a Novel Fuzzy Flip-Flop Based Neural Network*, IJCNN 2009 Atlanta - accepted.

[13] L. T. Kóczy and A. Zorat, *Fuzzy systems and approximation*, Fuzzy Sets and Systems, Vol.85, pp. 203-222, 1995.

[14] R. Lovassy, L. T. Kóczy and L. Gál, *Optimizing Fuzzy Flip-Flop Based Neural Networks by Bacterial Memetic Algorithm*, IFSA/EUSFLAT 2009, Lisbon, Portugal, 2009 - accepted

ISBN: 978-989-95079-6-8

Interactive Fuzzy Modeling by Evolutionary Multiobjective Optimization with User Preference

Yusuke Nojima, Hisao Ishibuchi

Department of Computer Science and Intelligent Systems, Osaka Prefecture University
1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan
Email: {nojima, hisaoi}@cs.osakafu-u.ac.jp

Abstract— *One of the new trends in genetic fuzzy systems (GFS) is the use of evolutionary multiobjective optimization (EMO) algorithms. This is because EMO algorithms can easily handle two conflicting objectives (i.e., accuracy maximization and complexity minimization) when we design accurate and compact fuzzy rule-based systems from numerical data. Since the main advantage of fuzzy rule-based systems compared with other non-linear ones is their linguistic interpretability, the design of fuzzy rule-based systems can be viewed as linguistic data mining from numerical data. From the data mining point of view, the required knowledge strongly depends on its user. That is, the interpretability of fuzzy rule-based systems should be evaluated by taking into account the user's preference. Although there exist a number of interpretability measures in the literature, users usually do not know which measure represents their preference beforehand. In this paper, we propose interactive fuzzy modeling by evolutionary multiobjective optimization with user's preference. User's preference is represented by several satisfaction level functions which can be interactively modified by the user. The user's preference is used as one of multiple objectives in an EMO algorithm. As a case study, we apply our approach to real world time-series data of land price movements in Japan and demonstrate a user interface of our approach.*

Keywords— evolutionary multiobjective optimization, fuzzy modelling, interactive evolutionary computation, user preference.

1 Introduction

There are two major goals in the design of fuzzy rule-based systems: accuracy maximization and complexity minimization. Since the mid-1990s, a large number of approaches have been proposed for improving the accuracy of fuzzy rule-based systems while maintaining their interpretability [1-22]. Genetic algorithms have been frequently used under the name of genetic fuzzy systems (GFS) [23-25]. We can easily handle two conflicting objectives: accuracy maximization and complexity minimization by the weighted sum of them or multiobjective formulations using evolutionary multiobjective optimization (EMO) algorithms [26-28]. One of the hottest issues in GFS is how to measure the interpretability of fuzzy rule-based systems [29-34]. A number of interpretability measures have been already proposed and implemented in GFS. Interpretability is, however, very subjective for users. Let us assume the following two rule sets.

[Fuzzy rule-based system A]
If x_1 is *big* then y is 10,
If x_3 is *small* then y is 5,
If x_2 is *very small* then y is 2,
If x_3 is *big* then y is 11.

[Fuzzy rule-based system B]
If x_1 is *big* and x_2 is *big* and x_3 is *small* then $y=x_1+5x_2+9$,
If x_1 is *very big* and x_2 is *very small* then $y = 2x_2+2$,
If x_1 is *very small* and x_2 is *small* and x_3 is *big* then y=12.

If we assume that a rule set with a small number of rules is interpretable, the fuzzy rule-based system B is more interpretable. However, the fuzzy rule-based system A seems to be more interpretable with respect to the rule length and the rule type. From this observation, we can say that the interpretability is totally dependent on user's preference (see [35] for more detailed discussions). There is another problem on interpretability. Users usually know which is interpretable for them only after comparing among some alternative fuzzy rule-based systems. That is, an interactive optimization process must be needed for the users.

In our former studies [36, 37], we have proposed an interactive genetic fuzzy rule selection method for pattern classification problems. A preference function is used as one objective function in an EMO algorithm. The preference function is composed of several satisfaction level functions. The reason why we used several satisfaction level functions is that users don't know appropriate criteria and their priorities among them beforehand. These satisfaction level functions are interactively modified during the evolution under the framework of interactive evolutionary computation [38-41]. In this paper, we apply this idea to a fuzzy modeling problem. We deal with time-series data of land price movements and demonstrate the prototype of a user interface.

This paper is organized as follows: Section 2 explains fuzzy modeling and its interactive optimization process. Section 3 explains the tackled problem and demonstrates the effectiveness of our method with a prototype of our user interface. Section 4 concludes this paper.

2 Interactive Fuzzy Modeling

2.1 Fuzzy Modeling

In this paper, for an n-input and single-output nonlinear function $y = y(\mathbf{x})$, we use the following fuzzy if-then rules:

$$\text{Rule } R_k : \text{If } x_1 \text{ is } A_{k1} \text{ and } ... \text{ and } x_n \text{ is } A_{kn} \text{ then } y \text{ is } B_k, \quad k = 1, 2, ..., N, \tag{1}$$

where x_i is the i-th input variable of an n-dimensional input vector $\mathbf{x} = (x_1, \ldots, x_n)$, y is an output variable, k is a rule index, A_{ki} is an antecedent linguistic label (e.g., *small* and *large*) for

ISBN: 978-989-95079-6-8

x_i, B_k is a consequent linguistic value, and N is the total number of fuzzy if-then rules.

The following fuzzy reasoning method has been frequently used in fuzzy rule-based systems since its first proposal in a neuro-fuzzy system [42]:

$$\hat{y}(\mathbf{x}) = \frac{\sum_{k=1}^{N} \mu_k(\mathbf{x}) \cdot b_k}{\sum_{k=1}^{N} \mu_k(\mathbf{x})}, \quad (2)$$

where $\mu_k(\mathbf{x})$ is the compatibility grade of the fuzzy if-then rule R_k with the input vector $\mathbf{x}$, and b_k is a representative real number of the consequent linguistic value B_k. The compatibility grade $\mu_k(\mathbf{x})$ is usually calculated by the product operation as

$$\mu_k(\mathbf{x}) = \mu_{k1}(x_1) \times \cdots \times \mu_{kn}(x_n), \quad (3)$$

where $\mu_{ki}(\cdot)$ is the membership function of the antecedent linguistic value A_{ki}. The representative real number b_k can be viewed as a result of the defuzzification of the consequent linguistic value B_k.

The fuzzy reasoning method in (2) can be viewed as a simplified version of the Takagi-Sugeno (TS) model where a linear function is used in the consequent part of each fuzzy if-then rule. The simplified fuzzy reasoning method in (2) has several advantages. For example, its reasoning mechanism is very simple, and it is suitable for gradient-based learning algorithms.

Since we use multiple granularities of fuzzy sets for an input vector, sometime the effect of specific rules becomes lower due to general rules. For giving specific rules more weight, we use an idea of inclusion-based fuzzy reasoning [43]. We extend the inclusion relation $R_A \subset R_B$ in [43]. When only the two rules R_k and R_q with the relation $|\mathbf{A}_k| > |\mathbf{A}_q|$, are compatible with the input vector $\mathbf{x}$, the specific rule R_q is mainly used in fuzzy reasoning. That is, the weight of the general rule R_k is discounted. Our idea is to determine the amount of the discount for R_k using the compatibility grade $\mu_q(\mathbf{x})$ of the specific rule R_q with the input vector $\mathbf{x}$. More specifically, the weight of R_k is defined as $(1-\mu_q(\mathbf{x}))$. When the specific rule R_q is fully compatible with the input vector $\mathbf{x}$, the weight of the general rule R_k is zero. This means that R_k has no effect on the calculation of the estimated output value $\hat{y}(\mathbf{x})$. On the other hand, when the compatibility grade of R_q with $\mathbf{x}$ is very small, the amount of the discount for R_k is also very small. In this case, R_k has almost the same weight as R_q. Since the general rule R_k may include multiple rules, its weight is defined as

$$w(R_k, \mathbf{x}) = \prod_{\substack{q=1 \\ |\mathbf{A}_q| < |\mathbf{A}_k|}}^{N} (1 - \mu_q(\mathbf{x})). \quad (4)$$

When there are no compatible fuzzy if-then rule smaller than R_k, $w(R_k, \mathbf{x})$ is specified as $w(R_k, \mathbf{x}) = 1$ because the weight of R_k should not be discounted in this case. It should be noted that the weight of each rule depends on the compatibility grades of other rules with the input vector $\mathbf{x}$. This means that the weight is context-dependent. Different weights are assigned to the same rule for different input vectors. Moreover, the same rule may have different weights for the same input vector in different rule bases because the weight of each rule depends on other rules.

Using the rule weight $w(R_k, \mathbf{x})$ of each fuzzy if-then rule R_k, our inclusion-based fuzzy reasoning method is written as

$$\hat{y}(\mathbf{x}) = \frac{\sum_{k=1}^{N} w(R_k, \mathbf{x}) \cdot \mu_k(\mathbf{x}) \cdot b_k}{\sum_{k=1}^{N} w(R_k, \mathbf{x}) \cdot \mu_k(\mathbf{x})}. \quad (5)$$

2.2 Multiobjective Fuzzy Rule Selection for Modeling

We use a simple two-stage method for designing rule sets. In the first phase, a large number of candidate rules are generated from the possible combinations of membership functions. The consequent real number is specified as the weighted average of output values of compatible input-output pairs as

$$b_k = \frac{\sum_{p=1}^{m} \mu_{\mathbf{A}_k}(\mathbf{x}_p) \cdot y_p}{\sum_{p=1}^{m} \mu_{\mathbf{A}_k}(\mathbf{x}_p)}, \quad (6)$$

where $\mu_{\mathbf{A}_k}(\mathbf{x}_p)$ is the compatibility grade of the input vector $\mathbf{x}_p$ with the antecedent part $\mathbf{A}_k$ of the linguistic rule R_k.

In the second phase of our rule selection, a number of fuzzy rule sets are selected by a multiobjective genetic algorithm. Any subset S of the N candidate rules can be represented by a binary string of length N as

$$S = s_1 s_2 \cdots s_N, \quad (7)$$

where s_k=1 and s_k=0 represent the inclusion of the k-th candidate rule R_k in S and the exclusion of R_k from S.

Each fuzzy rule set S is evaluated by the three objectives:

$f_1(S)$: the total square error by S,

$f_2(S)$: the number of selected fuzzy rules in S,

$f_3(S)$: the overall user preference for S.

The first and second objectives have been frequently used and correspond to accuracy maximization and complexity minimization, respectively. The first objective is calculated by,

$$f_1(S) = \sum_{p=1}^{m} (y_p - \hat{y}_p(\mathbf{x}_p))^2. \quad (8)$$

The third objective $f_3(S)$ is the newly proposed objective in this paper. We explain it in the next subsection. The problem formulation of multiobjective genetic fuzzy rule selection is written as

Minimize $f_1(S)$ and $f_2(S)$, and maximize $f_3(S)$. (9)

ISBN: 978-989-95079-6-8

We use NSGA-II of Deb et al. [27] to search for a number of non-dominated fuzzy rule-based systems with respect to these three objectives. In this paper, uniform crossover and biased bit-flip mutation are used in NSGA-II. The biased mutation is that a larger probability is assigned to the mutation from 1 to 0 than that from 0 to 1.

Figure 1 shows the whole procedure of the proposed method. We specify an interval (i.e., the number of generations) for internal evaluations. During this interval, the satisfaction level functions are not changed. After the interval, the user checks some of non-dominated rule sets and modifies the satisfaction level functions. Then another internal evaluation process starts. By repeating this interactive process, the user can specify the own preference and find the rule set with the high user preference value.

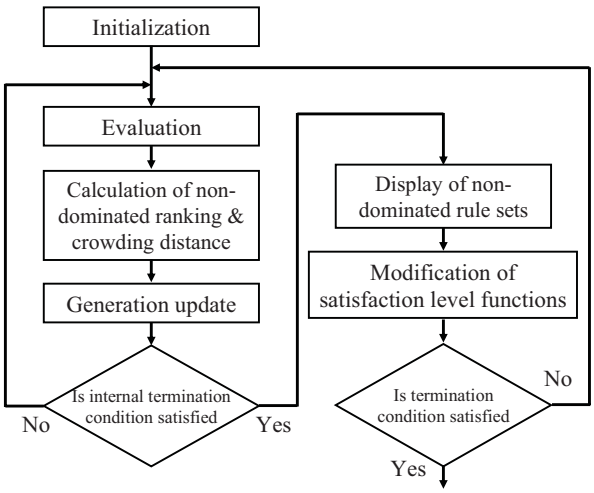


Figure 1: The whole procedure of the proposed method.

2.3 *Preference Function*

In our former study [36, 37], we have proposed a preference function composed of several satisfaction level functions. The inputs for the satisfaction level functions are criteria on the accuracy and interpretability of fuzzy rule sets. Each satisfaction level function is represented by a trapezoidal function in Fig. 2. $g^r(S)$ is the value of r-th criterion for the rule set S. $u^r(g^r(S))$ is the output of r-th satisfaction level function. Users specify the preference and priority for each criterion by moving the point B in Fig. 2. That is, the u_x and u_y of the point B mean the preference and priority for the criterion, respectively. The u_x can be also regarded as the maximum criterion level.

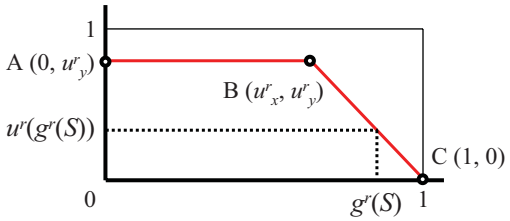


Figure 2: A trapezoidal function for representing satisfaction level functions of each criterion.

The satisfaction level function can be viewed as the requirement level. During evolution, users can modify the satisfaction level function according to the temporally obtained non-dominated rule sets.

The third objective function $f_3(S)$ for an overall user's preference is calculated by

$$f_3(S) = \sum_{r=1}^{N_C} u^r\left(g^r(S)\right), \tag{10}$$

where N_c is the number of criteria.

3 Case Study

3.1 *Problem Description*

We apply our proposed method to a simple fuzzy modelling problem for time-series data. The data we used in this paper is the land price movements of the three major metropolitan areas in Japan available from Ministry of Land, Infrastructure, Transport and Tourism webpage (http://www.mlit.go.jp/index_e.html). The data includes the land price movements from 1980 to 2000. In this period, the bubble economy was a big problem: the increased demand for office buildings in city centres due to internationalization and informatization.

The data is composed of 63 pairs of two inputs (i.e., *area* and *year*) and one output (*change of land price*). For simplicity, we normalized the input attributes into [0, 1]x[0, 1] space. We used seven categorical values (all possible combinations) for *area* attribute. For year attribute, we used 48 fuzzy membership functions shown in Fig. 4 and *don't care* condition. Each peak of triangular membership functions corresponds to one of years. These partitions could be understandable (e.g., 90's, Mid of 80's, around 1988).

From this data, 343 fuzzy if-then rules were generated as candidate rules. Thus, the search space is 2^{343}.

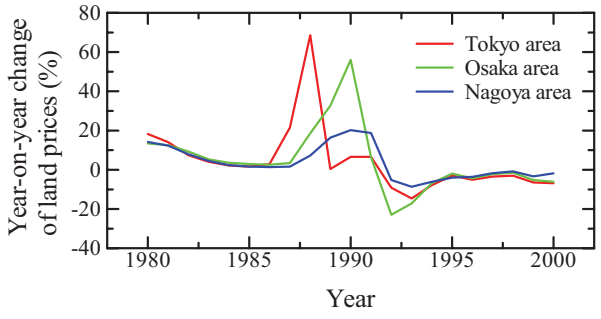


Figure 3: Year-on-year change of land prices of three major metropolitan areas in Japan from 1980 to 2000.

3.2 *Criteria for Representing User's Preference*

There are a lot of interpretability measures in the literature. In this case study, we used four simple criteria like:

- Maximum square error by S,
- Overlap among antecedent sets in S,
- Total square error by S,
- The number of fuzzy rules in S.

ISBN: 978-989-95079-6-8

The maximum square error is calculated by

$$g^1(S) = \max_{p=1,\ldots,m} (y_p - \hat{y}_p(\mathbf{x}_p))^2 . \quad (11)$$

When a user gives a high priority to this criterion, the larger changes could be fitted by a fuzzy rule set.

We normalized each value of four criteria within the valid ranges based on the distribution values in the pre-simulation without user's preference. The reason why we used *the total square error* and *the number of fuzzy rules* as the criteria for user's preference is to reduce the search space based on the user's preference.

The choice of interpretability measures is future research issues. The correlation among measures must be examined.

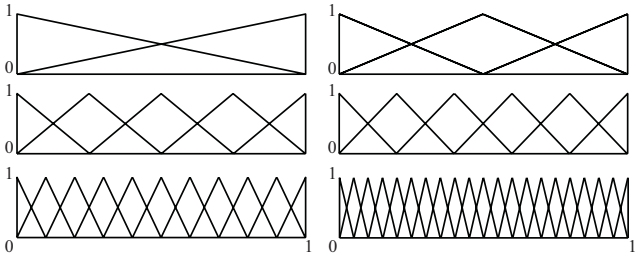


Figure 4: Fuzzy partitions with different granularities for *Year* attribute.

3.3 Prototype of User Interface

Figure 5 shows the prototype of our user interface. There are two windows. The left one represents the actual land price movement and the inferred land price movement calculated by the chosen fuzzy rule set.

At the middle of the left window, there are two graphs for representing non-dominated rule sets in terms of "the total square error and the number of rules" and "the total square error and the user preference value". Red open plots represent non-dominated rule sets. Blue closed plot means the chosen rule set. The above inferred land price movement corresponds to the chosen rule set. Users can choose one of the non-dominated rule sets by clicking any plot in the graphs. The right CUI window shows the rules in the chosen rule set.

At the bottom of the left window, there are four satisfaction level functions. Users can change the shapes of these functions by moving each point B in Fig. 2.

The button "Evolve" is a trigger to start internal evaluations. In this paper, we specified the number of generations for internal evaluations as 100.

3.4 Some Results

When a user specified the satisfaction level functions for the maximum error and the number of rules as in Fig. 6, the user obtained a rule set with a small number of rules. The rule set seems to represent the original characteristics of the data. From Table 1, we can see that there are some general rules and specific rules in the fuzzy rule set. The value in parentheses in Table 1 represents the range of 0.5-level set of used membership function. For example, "1990 [1]" means the smallest partition in which the peak is 1990.

Table 1: Obtained rule set in Fig. 6.

Area	Year	Change %
Osaka	1990 [1]	56.1
Osaka	1992 [1]	-22.9
Tokyo	1988 [1]	68.6
Tokyo & Osaka	1993 [1]	-15.9
Osaka & Nagoya	1990 [2]	28.4
Tokyo & Osaka & Nagoya	1988 [4]	17.2
Tokyo & Osaka & Nagoya	-	4.3

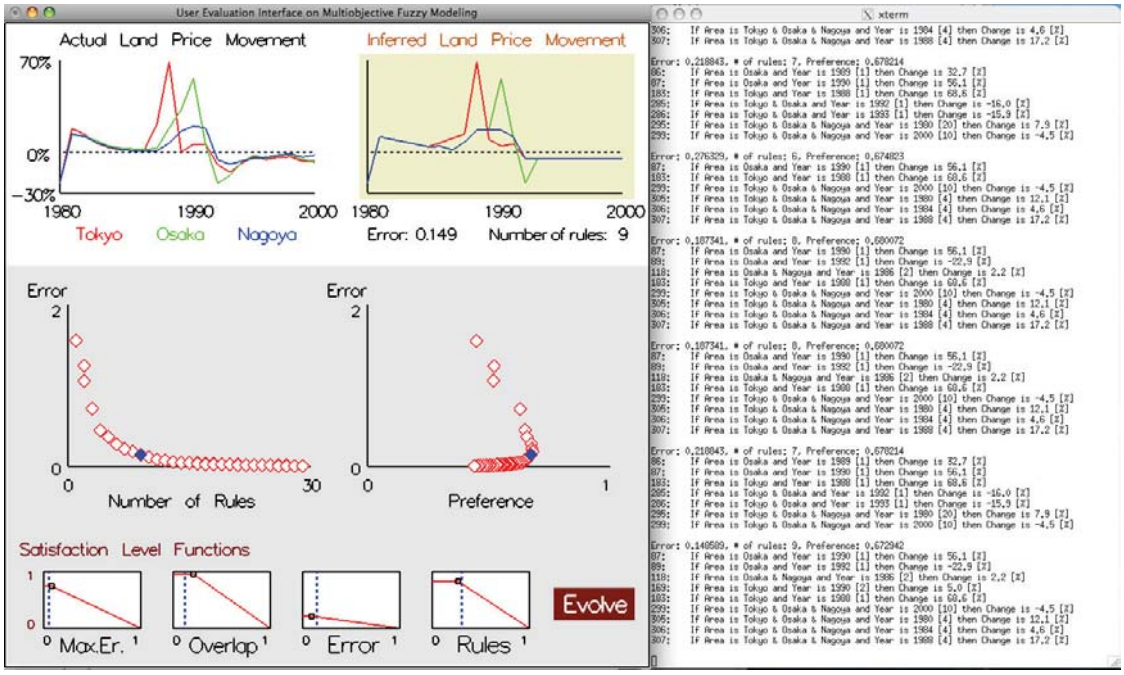


Figure 5: Prototype of user interface for interactive fuzzy modelling.

ISBN: 978-989-95079-6-8

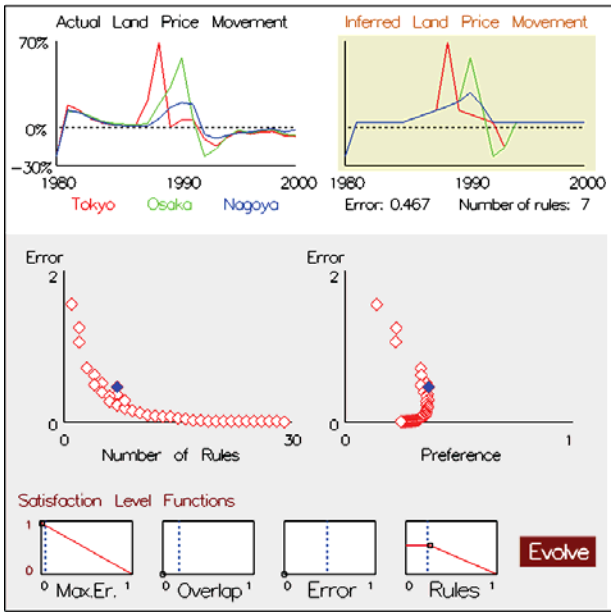


Figure 6: Example 1.

When a user gave a high priority to "Overlap among antecedent sets in S" as in Fig. 7, the user obtained a rule set with only one general rule and six specific rules (see Table 2).

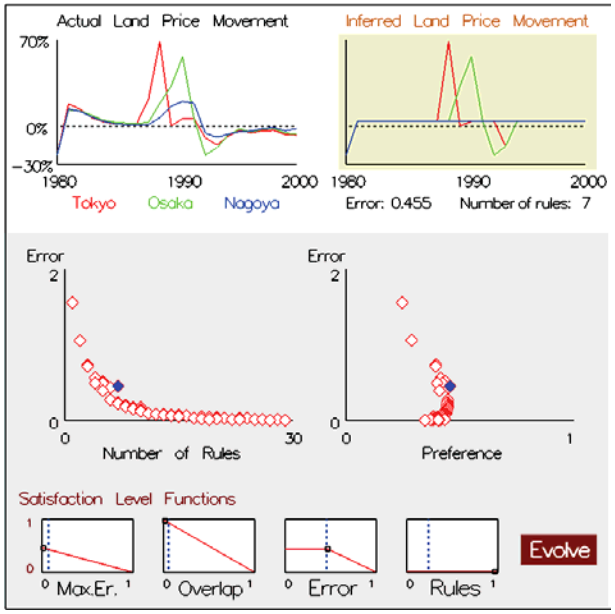


Figure 7: Example 2.

Table 2: Obtained rule set in Fig. 7.

Area	Year	Change %
Osaka	1989 [1]	32.7
Osaka	1990 [1]	56.1
Osaka	1992 [1]	-22.9
Tokyo	1988 [1]	68.6
Tokyo	1989 [1]	0.4
Tokyo & Osaka	1993 [1]	-15.9
Tokyo & Osaka & Nagoya	-	4.3

When a user gave high priorities to accuracy criteria (i.e., maximum error and total square error), a very accurate rule set was obtained as in Fig. 8.

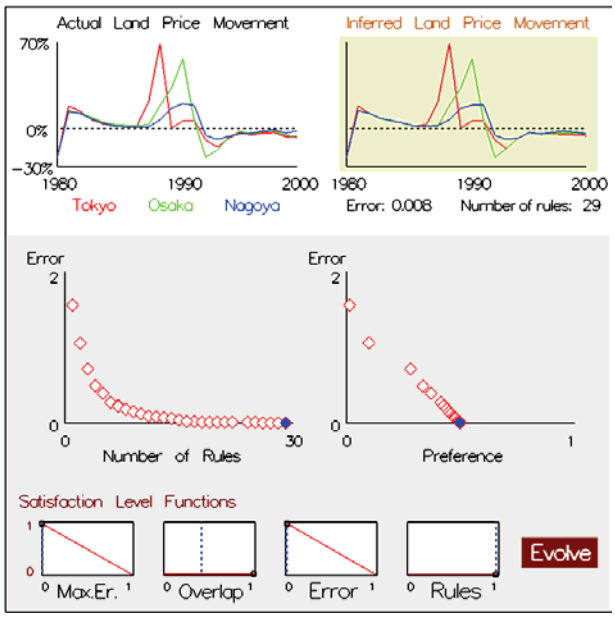


Figure 8: Example 3.

4 Conclusions

In this paper, we incorporated user's preference into multiobjective fuzzy modeling. We proposed a preference function composed of four satisfaction level functions. We utilized this preference function as an additional objective in an EMO algorithm. Through a case study, we demonstrated that a user can interactively specify satisfaction level functions during the evolution. We also showed that the user can obtain an accurate and interpretable fuzzy rule-based system based on his/her own preference.

In our case study, we intuitively selected four criteria to represent user's preference. Further studies are needed to choose appropriate criteria. We also have other interesting research issues to be discussed in future studies such as the visualization of multi-dimensional data and the minimization of human user's fatigue caused by the interaction with our system. The latter includes automated preference modeling.

Acknowledgment

This work partially supported by Grand-in-Aid for Young Scientists (B): KAKENHI (18700228).

References

[1] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms, *Fuzzy Sets and Systems* 65 (2/3) 237-253, 1994.

[2] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, Selecting fuzzy if-then rules for classification problems using genetic algorithms, *IEEE Trans. on Fuzzy Systems* 3 (3) 260-270, 1995.

[3] H. Ishibuchi, T. Murata, and I. B. Turksen, Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems, *Fuzzy Sets and Systems* 89 (2) 135-150, 1997.

ISBN: 978-989-95079-6-8

[4] M. Setnes, R. Babuska, and B. Verbruggen, Rule-based modeling: Precision and transparency, *IEEE Trans. on Systems, Man, and Cybernetics - Part C* 28 (1) 165-169, 1998.

[5] J. Yen and L. Wang, G. W. Gillespie, Improving the interpretability of TSK fuzzy models by combining global learning and local learning, *IEEE Trans. on Fuzzy Systems* 6 (4) 530-537, 1998.

[6] D. Nauck and R. Kruse, Obtaining interpretable fuzzy classification rules from medical data, *Artificial Intelligence in Medicine* 16 (2) 149-169, 1999.

[7] Y. Jin, Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement, *IEEE Trans. on Fuzzy Systems* 8 (2) 212-221, 2000.

[8] M. Setnes and H. Roubos, GA-based modeling and classification: Complexity and performance, *IEEE Trans. on Fuzzy Systems* 8 (5) 509-522, 2000.

[9] H. Ishibuchi, T. Nakashima, and T. Murata, Three-objective genetics-based machine learning for linguistic rule extraction, *Information Sciences* 136 (1-4) 109-133, 2001.

[10] L. Castillo, A. Gonzalez, and R. Perez, Including a simplicity criterion in the selection of the best rule in a genetic fuzzy learning algorithm, *Fuzzy Sets and Systems* 120 (2) 309-321, 2001.

[11] H. Roubos and M. Setnes, Compact and transparent fuzzy models and classifiers through iterative complexity reduction, *IEEE Trans. on Fuzzy Systems* 9 (4) 516-524, 2001.

[12] J. Abonyi, J. A. Roubos, and F. Szeifert, Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization, *International Journal of Approximate Reasoning* 32 (1) 1-21, 2003.

[13] H. Ishibuchi and T. Yamamoto, Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining, *Fuzzy Sets and Systems* 141 (1) 59-88, 2004.

[14] J. Casillas, O. Cordon, M. J. del Jesus, and F. Herrera, Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction, *IEEE Trans. on Fuzzy Systems* 13 (1) 13-29, 2005.

[15] R. Alcala, J. Alcala-Fdez, J. Casillas, O. Cordon, and F. Herrera, Hybrid learning models to get the interpretability-accuracy trade-off in fuzzy modeling, *Soft Computing* 10 (9) 717-734, 2006.

[16] P. Pulkkinen and H. Koivisto, Identification of interpretable and accurate fuzzy classifiers and function estimators with hybrid methods, *Applied Soft Computing* 7 (2) 520-533, 2007.

[17] H. Ishibuchi and Y. Nojima, Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, *International Journal of Approximate Reasoning* 44 (1) 4-31, 2007.

[18] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni, A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems, *Soft Computing* 11 (11) 1013-1031, 2007.

[19] A. F. Gomez-Skarmeta, F. Jimenez, and G. Sanchez, Improving interpretability in approximative fuzzy models via multiobjective evolutionary algorithms, *International Journal of Intelligent Systems* 22 (9) 943-969, 2007.

[20] J. Gonzalez, I. Rojas, H. Pomares, L. J. Herrera, A. Guillen, J. M. Palomares, and F. Rojas, Improving the accuracy while preserving the interpretability of fuzzy function approximators by means of multi-objective evolutionary algorithms, *International Journal of Approximate Reasoning* 44 (1) 32-44, 2007

[21] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu, Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index, *Soft Computing* 13 (5) 437-449, 2009.

[22] L. Sanchez, J. Otero, and I. Couso, Obtaining linguistic fuzzy rule-based regression models from imprecise data with multiobjective genetic algorithms, *Soft Computing* 13 (5) 467-479, 2009.

[23] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems*, World Scientific, 2001.

[24] O. Cordon, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, Ten years of genetic fuzzy systems: Current framework and new trends, *Fuzzy Sets and Systems* 141 (1) 5-31, 2004.

[25] F. Herrera, Genetic fuzzy systems: Taxonomy, Current research trends and prospects, *Evolutionary Intelligence* 1, 27-46, 2008.

[26] K. Deb, *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, Chichester, 2001.

[27] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. on Evolutionary Computation* 6 (2) 182-197, 2002.

[28] C. A. C. Coello, D. A. van Veldhuizen, and B. B. Lamont, *Evolutionary algorithms for solving multi-objective problems*. Kluwer Academic Publishers, Boston, 2002.

[29] S. Guillaume, Designing fuzzy inference systems from data: An interpretability-oriented review, *IEEE Trans. on Fuzzy Systems* 9 (3) 426-443, 2001.

[30] R. Mikut, J. Jakel, and L. Groll, Interpretability issues in data-based learning of fuzzy systems, *Fuzzy Sets and Systems* 150 (2) 179-197, 2005.

[31] C. Mencar, G. Castellano, and A. M. Fanelli, On the role of interpretability in fuzzy data mining, *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 15 (5) 521-537, 2007.

[32] C. Mencar and A. M. Fanelli, Interpretability constraints for fuzzy information granulation, *Information Sciences* 178 (24) 4585-4618, 2008.

[33] S. M. Zhou and J. Q. Gan, Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling, *Fuzzy Sets and Systems* 159 (23) 3091-3131, 2008.

[34] J. M. Alonso, L. Magdalena, and S. Guillaume, HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism, *International Journal of Intelligent Systems* 23 (7) 761-794, 2008.

[35] H. Ishibuchi and Y. Nojima, Discussions on interpretability of fuzzy systems using simple examples, *Proc. of 2009 IFSA World Congress and 2009 EUSFLAT conference* (to appear).

[36] Y. Nojima and H. Ishibuchi, Incorporation of user preference into multiobjective genetic fuzzy rule selection for pattern classification problems, *Proc. of 14th International Symposium on Artificial Life and Robotics 2009*, 186-189, 2009.

[37] Y. Nojima and H. Ishibuchi, Interactive genetic fuzzy rule selection through evolutionary multiobjective optimization with user preference, *Proc. of 2009 IEEE Symposium on Computational Intelligence in Multicriteria Decision-Making*, 141-148, 2009.

[38] H. Takagi, Interactive evolutionary computation: fusion of the capabilities of EC optimization an human evaluation, *Proceedings of the IEEE*, 81 (9) 1275-1296, 2001.

[39] S. B. Cho, Towards creative evolutionary systems with interactive genetic algorithm. *Applied Intelligence* 16: 129-138, 2002.

[40] A. M. Brintrup, J. Ramsden, H. Takagi and A. Tiwari, Ergonomic chair design by fusing qualitative and quantitative criteria using interactive genetic algorithms. *IEEE Trans. on Evolutionary Computation* 12 (3) 343-354, 2008.

[41] N. Kubota, Y. Nojima, F. Kojima, and T. Fukuda, Multiple fuzzy state-value functions for human evaluation through interactive trajectory planning of a partner robot, *Soft Computing* 10 (10), 891-901, 2006.

[42] H. Ichihashi, Iterative fuzzy modeling and a hierarchical network, *Proc. of 4th IFSA world congress*, 49-52, 1991.

[43] H. Ishibuchi, T. Yamamoto, and T. Nakashima, An approach to fuzzy default reasoning for function approximation, *Soft Computing* 10 (9), 850-864, 2006.

ISBN: 978-989-95079-6-8

Incremental possibilistic approach for online clustering and classification

Moamar Sayed Mouchaweh, Bernard Riera

Université de Reims Champagne-Ardenne, Centre de Recherche en STIC (URCA-CReSTIC)
Moulin de la Housse, BP 1039, 51687 Reims Cedex, France
Email: {moamar.sayed-mouchaweh,bernard.riera}@univ-reims.fr

Abstract— *In this paper, we propose to develop the supervised classification method Fuzzy Pattern Matching to be in addition a non supervised one. The goal is to monitor dynamic systems with a limited prior knowledge about their functioning. The detection of the occurrence of new states as well as the reinforcement of the estimation of their membership functions are performed online thanks to the combination of supervised and non supervised classification modes. No information in advance about the shape of classes or their number is required to achieve this detection and estimation reinforcement.*

Keywords— Classification, clustering, Sequential learning, Fuzzy Pattern Matching.

1 Introduction

Pattern Recognition (PR) [7] is the study of how machines can learn from experience to make sound decisions about the classes of patterns of interest. PR involves two stages: preprocessing and classification. The preprocessing includes feature extraction [1] and selection [12]. The classification stage is a mapping of a pattern from the feature space into the decision one. The latter is defined by a set of predefined classes. This mapping is done using a classifier. The latter is a method or algorithm which generates a class membership function in order to classify unlabelled incoming patterns into one of the predefined classes. Depending on the information available for classifier training, one can distinguish between supervised [10] and unsupervised [2] learning. In the first case, called also classification, there exists a set of patterns with their class assignment or label, called learning set. The goal of supervised learning is to learn a set of membership functions that allows the classification of new patterns into one of existing classes. The problem of unsupervised learning, also called clustering, arises if clusters' memberships of available patterns, and perhaps even the number of clusters, are unknown. In such cases, a classifier is learned based on similar properties of patterns. Hence, the clustering aims to partition a given set of patterns into clusters based on their similarity.

Semi supervised learning techniques use small or limited labelled patterns to estimate the classes' membership functions and the unlabelled ones to detect the occurrence of new classes and refine their membership functions estimation. Examples of semi supervised methods can be found in [4, 9, 11, 13] and the references therein. These methods are based either on the use of the Expectation Maximisation algorithm for maximum likelihood based parameters estimation [9], on the integration of an incremental algorithm for the update of classifiers' parameters [13], on the optimisation of an objective or learnable distance function [4] or on a classifier retraining to integrate new labelled points [11]. The popularity of these methods can be attributed to the fact that new information can be incorporated resulting better estimation of classes' membership functions and thus more prediction accuracies thanks to the unlabelled patterns. However, the representativeness of the labelled data is of crucial importance especially for small ratios of labelled to unlabelled patterns [9]. This is due to the fact that the clustering is guided by the labelled patterns.

One of the applications of PR methods is the monitoring of dynamic systems. System states, normal or faulty, are characterized by classes in the feature space. The performance of statistical PR methods depends on the prior knowledge, or learning set, about system behavior. The number of available learning patterns is often limited and small compared to the dimension of the feature space. Thus, it becomes hard to estimate the class membership function leading to a large variance in parameter estimates and thus higher classification error rates. Moreover, the behavior of a dynamic system can assume different operating states in the course of time. The learning set cannot contain patterns about all these states especially the faulty ones. Thus, the occurrence of these missing states must be anticipated online and integrated in the data base. In this paper, we propose a solution for these problems. This solution is based on the use of the supervised classification method Fuzzy Pattern Matching [3]. This method presents the advantage to process data with a low and constant classification time according to the size of data base. We propose to develop FPM as an unsupervised classification method. The goal is to combine the supervised and unsupervised learning strategies within a single algorithm leading to a semi supervised version of FPM.

The paper is organized as follows. Firstly, the functioning of FPM is illustrated. In the next section, the proposed solution to perform supervised and unsupervised classification using FPM is detailed. The performance of semi supervised FPM is illustrated and tested using some simulated examples.

ISBN: 978-989-95079-6-8

2 Supervised Fuzzy Pattern Matching

2.1 Learning phase

In the learning phase, the probability histograms are constructed for each class according to each attribute. The number of bins h for a histogram is experimentally determined. This number has an important influence on the performances of FPM [14]. The histogram upper and lower borders can be determined either as the maximal and minimal learning data coordinates or by experts. The height of each bin $b_k^j, k \in \{1,2,..,h\}$, according to each attribute j is the number of learning patterns $n_{ib_k^j}$ of the class C_i located in this bin. The probability distribution $\{ p_i^j(y_{b_k^j}), k \in \{1,2,..,h\}, j \in \{1,2,..,d\}\}$, of the class C_i, $i \in \{1,2,..,c\}$, according to the attribute j is obtained by dividing the height of each bin by the total number N_i of learning patterns belonging to the same class C_i. These probabilities are assigned to the bins' centres $y_{b_k^j}, k \in \{1,2,..,h\}$:

$$p_i^j(y_{b_k^j}) = \frac{n_{ib_k^j}}{N_i} \quad (1)$$

The Probability Density Function (PDF) is obtained by a linear linking between bins heights centres.

In order to take into account the uncertainty and the imprecision contained in the data, the probability distribution is converted into possibility one $\left\{\pi_i^j(y_{b_k^j}), k \in \{1,2,..,h\}, j \in \{1,2,..,d\}\right\}$. The conversion is performed using the transformation of Dubois and Prade [6] defined by:

$$\pi_i^j(y_{b_k^j}) = \sum_{f=1}^{h} \min(p_i^j(y_{b_k^j}), p_i^j(y_{b_f^j})) \quad (2)$$

A linear linking between bins heights centres converts the distribution of possibilities into density one Π_i^j. This operation is repeated for all the attributes of each class.

2.2 Classification Phase

The membership function for each class C_i according to each attribute j is considered to be numerically equivalent to the possibility distribution [15]. Thus, the classification of a new pattern x, whose values of the different attributes are x^1, .., x^j, .., x^d, is made in two steps :

Determination of the possibility membership value $\pi_i^j(x^j)$ of x^j to each class C_i according to the attribute j by a projection on the corresponding possibility density Π_i^j,

Merging all the possibility values $\pi_i^1(x^1)$, $\pi_i^2(x^2)$,.., $\pi_i^d(x^d)$ concerning the class C_i, into a single one by the aggregation operator "minimum" :

$$\pi_i(x) = \min(\pi_i^1(x^1), \pi_i^2(x^2),.., \pi_i^d(x^d)) \quad (3)$$

The result π_i of this fusion corresponds to the global possibility value that x belongs to the class C_i. Finally, x is assigned to the class for which it has the maximum membership value.

3 Semi Supervised Fuzzy Pattern Matching

The proposed semi supervised FPM has an agglomerative characteristic. Thus, it does not require any prior information about the number of classes. The classes' membership functions are constructed sequentially with the patterns' arrival. According to the ratio $r = \frac{L}{UL + L}$ of the number L of labelled points to the one UL of unlabelled points, the proposed method can be totally supervised, $r = 1$, or totally unsupervised, $r = 0$. Let $r_i = \frac{L_i}{UL_i + L_i}$ be the ratio of labelled points L_i belonging to the class C_i to the unlabelled ones UL_i which will be assigned to C_i. In the case that $0 \le r_i < 1$, the benefit of semi supervised FPM is to enhance the quality of class's membership estimation thanks to the incorporation of the unlabelled points in this class. This enhancement is performed online thanks to the use of an incremental approach as we can see later. While if $r_i = 0$, the benefit of semi supervised FPM is to detect this new class and to learn its membership function online. Thus, semi supervised FPM presents benefits in both classification and clustering.

In the case of $r_i = 0$, the first incoming unlabelled pattern is considered as the point prototype of a new class and its possibilistic membership function according to each attribute is computed as in supervised FPM based on this only pattern. The next unlabelled pattern is either classified in this created class, if it has a membership value according to this class, or considered as a point prototype of a new class. After the classification of each new pattern, the membership function of the corresponding class is updated online using an incremental algorithm. Due to the initialisation, created classes may need to be merged. This merging is performed using a similarity measure. The functioning of semi supervised FPM involves the following two steps.

3.1 Classes detection and local adaptation step

Let $x = (x^1, x^2,.., x^d) \in IR^d$ be a given pattern vector in a feature space constituted of d parameters or attributes. There is no learning set containing labelled patterns, nor a prior information about classes' probability density shape or their number. Each attribute is divided into equal intervals defining the bins of the histogram according to this attribute. This histogram is used to estimate the conditional probability density for the class that x is driven from. Let $X_{\min}^j$ and $X_{\max}^j$ be respectively the lower and upper borders of the histogram according to the attribute j. These borders can be determined by expert as the minimal and maximal values that an attribute can reach. Let h be the number of histogram's bins, then each bin according to the attribute j has the width :

ISBN: 978-989-95079-6-8

$$\Delta^j = \frac{X^j_{\max} - X^j_{\min}}{h},\ j \in \{1,2,..,d\} \quad (4)$$

Thus the limits of these bins are defined as follows :

$$b^j_1 = [X^j_{\min}, X^j_{\min} + \Delta^j],\ b^j_2 = [X^j_{\min} + \Delta^j, X^j_{\min} + 2\Delta^j],$$
$$.., b^j_h = [X^j_{\min} + (h-1)\Delta^j, X^j_{\max}],\ j \in \{1,2,..,d\} \quad (5)$$

The classes detection and local adaptation step involves two strategies : detection of new classes and local adaptation of their membership functions. The local adaptation strategy is based on an update of classes' possibility densities after the classification of each new pattern so that classifier can follow online gradual temporal, or local, changes of classes' membership functions. This online update requires a recursive representation of classes' possibility densities. However the incremental updating cannot detect abrupt changes as changes in the number of clusters. This abrupt change is followed up by the detection strategy which is based on the fact that each new rejected pattern by all the learned classes is considered as a point prototype of a new class. The detection strategy is a mechanism for adjusting the number of clusters online, which is incremented after the detection of each new cluster or class.

A) Detection of new classes strategy

The first rejected pattern x according to all the known c classes is considered as the point prototype of the first new class: $C_c \leftarrow x, c \leftarrow c+1$. The PDF is obtained as in supervised FPM. If x is located in the bin $b^j_k, k \in \{1,2,..,h\}$, then the probability histogram of C_c according to the attribute j is : $p^j_c = \{p^j_{c1} = 0, p^j_{c2} = 0, .., p^j_{ck} = 1, .., p^j_{ch} = 0\}$. The possibility histogram will then be computed using (2). Since there is just one pattern, the possibility histogram is equal to the probability one. The possibility density of the class C_c is obtained by a linear linking between the centre of the bin b^j_k , which has the height 1, and the ones of its left b^j_{k-1} and right b^j_{k+1} neighbours, which have both at present the height 0. Generally, if $C = \{C_1, C_2, .., C_c\}$ is the set of learned classes at present, x a new pattern which is rejected by all the learned classes. The detection strategy is defined as follows :

$$\pi_i(x) = 0, \forall i \in \{1,2,..,c\} \Rightarrow c \leftarrow c+1,$$
$$C_c = \{x\}, \pi_c = \{\pi^1_c, .., \pi^j_c, .., \pi^d_c\} \quad (6)$$

B) Local adaptation strategy

For a next pattern x', the membership value to each class C_i, $\forall i \in \{1,2,..,c\}$, will be obtained by a projection on its possibility density Π^j_i according to each attribute j and then merging the values according to all the attributes using the aggregation operator "minimum" as in supervised FPM. If the membership value $\pi_i(x')$ of x' to the class C_i is different of zero, then this pattern will be assigned to the class C_i and the possibility densities of this class according to each attribute will be incrementally updated. To establish an incremental update of possibility densities, let $p^j_i = \{p^j_{i1}, p^j_{i2}, .., p^j_{ik}, .., p^j_{ih}\}$ and $\pi^j_i = \{\pi^j_{i1}, \pi^j_{i2}, .., \pi^j_{ik}, .., \pi^j_{ih}\}$ define respectively the probability and possibility histograms of the class C_i according to the attribute j. Let $p'^j_i = \{p'^j_{i1}, p'^j_{i2}, .., p'^j_{ik}, .., p'^j_{ih}\}$ and $\pi'^j_i = \{\pi'^j_{i1}, \pi'^j_{i2}, .., \pi'^j_{ik}, .., \pi'^j_{ih}\}$ define respectively the updated probability and possibility histograms of the class C_i according to the attribute j after the assignment of x' to the class C_i. Let suppose for the simplicity that : $p^j_{ih} < p^j_{i(h-1)} < .. < p^j_{i1}$, then these new probabilities can be computed incrementally by [14] :

$$x'^j \in b^j_k, \forall k \in \{1,..,h\} \Rightarrow p'^j_{ik} = p^j_{ik} * \frac{N_i}{N_i+1} + \frac{1}{N_i+1},$$
$$p'^j_{iz} = p^j_{iz} * \frac{N_i}{N_i+1}, \forall z \in \{1,..,h\}, z \neq k \quad (7)$$

Then the new possibilities can be computed using Dubois and Prade transformation defined by (2). Thus, the local adaptation strategy is defined as follows :

$$\pi_i(x') = \max_{z \in \{1,..,c\}} (\pi_z(x')) \Rightarrow C_i \leftarrow \{C_i, x'\},$$
$$\pi'_i = \{\pi'^1_i, \pi'^2_i, .., \pi'^j_i, .., \pi'^d_i\} \quad (8)$$

3.2 Classes merging step

The occurrence order of incoming patterns influences the final constructed clusters. This entails the possibility to obtain several different partitions or number of clusters. Thus, several clusters can represent the same class. These clusters must be merged into one cluster to obtain one partition and one membership function. This fusion can be done either by expert or by a merging measure. The later measures the overlap or closeness between constructed clusters. There are different measures for merging clusters in the literature. Most of them are based on a similarity measure between clusters, which takes into account either the degree of overlapping of clusters or the distance between clusters' centres. The clusters overlapping degree is based on the number of ambiguous patterns, belonging to several clusters, and their membership values to these clusters. If the number of these ambiguous patterns is large enough and their membership values to several clusters are high then these clusters cannot be considered as heterogeneous anymore and must be merged. An interesting similarity criterion which takes into account at the same time the number of ambiguous patterns as well as their membership values is defined by (Frigui *el al.* 1996) :

$$\delta_{iz} = 1 - \frac{\sum_{x \in C_i \text{ or } x \in C_z} |\pi_i(x) - \pi_z(x)|}{\sum_{x \in C_i} \pi_i(x) + \sum_{x \in C_z} \pi_z(x)} \quad (9)$$

δ_{iz} is the fuzzy similarity measure between the classes C_i and C_z. More this measure is close to one, more the two clusters are overlapped. We adopt this measure for the merging step of semi supervised FPM.

ISBN: 978-989-95079-6-8

The merging criterion can be applied offline or online. In the first case, several iterations of clustering with a variable number of clusters start with a high over-specified number (an upper bound). Then, the clusters number is reduced gradually until an appropriate number is found. In each iteration, similar clusters are merged and this procedure is repeated until no more clusters can be merged and finished with the optimal number of clusters. In dynamic applications, the merging operation must be done online because the whole unlabelled patterns are not available *a priori*. Thus, the merging criterion can be tested either after the classification of each pattern or in each time window. The problem of the testing after each classification pattern is the calculation complexity which depends on the cardinality of clusters to be merged. However, the clusters are merged online when this measure reaches a predefined threshold. While in the other case, the merging can be delayed according to the size of the time window, but the calculation complexity is less than the one of the first case. We propose to update the fuzzy similarity measure within a time window. This update requires the calculation of the new membership values for all the patterns of all the classes inside which new patterns have been assigned. Indeed, if a new incoming pattern x is assigned to the cluster C_i, then the fuzzy similarity measure must be computed only between C_i and the other clusters C_z, $z \in \{1,..,c\}, z \neq i$.

When two clusters are merged, their membership functions must also be merged. We propose to merge the membership functions online using an incremental approach. Let $p_i^j = \{p_{i1}^j, p_{i2}^j,.., p_{ik}^j,.., p_{ih}^j\}$, $p_z^j = \{p_{z1}^j, p_{z2}^j,.., p_{zk}^j,.., p_{zh}^j\}$ be respectively the probability distributions of the two clusters C_i and C_z to be merged. Each bin probability is computed as (1) : $p_{ik}^j = \frac{n_{ib_k^j}}{N_i}$, $p_{zk}^j = \frac{n_{zb_k^j}}{N_z}, \forall k \in \{1,..,h\}$. The probability of the bin b_k^j after the merging of the patterns of the two clusters C_i and C_z is equal to :

$$p_{izk}^j = \frac{n_{ib_k^j} + n_{zb_k^j}}{N_i + N_z} \tag{10}$$

Where $n_{ib_k^j}$ and $n_{zb_k^j}$ are respectively the number of patterns of the classes C_i and C_z located in the bin b_k^j according to the attribute j. N_i and N_z are respectively the number of patterns of the classes C_i and C_z. We can rewrite (10) as follows:

$$p_{izk}^j = \frac{n_{ib_k^j}}{N_i} * \frac{N_i}{N_i + N_z} + \frac{n_{zb_k^j}}{N_z} * \frac{N_z}{N_i + N_z} = p_{ik}^j * \frac{N_i}{N_i + N_z} + p_{zk}^j * \frac{N_z}{N_i + N_z} \tag{11}$$

Using (11), the probability distribution of the class after merging is obtained incrementally according to each attribute. Based on (2), the corresponding possibility distributions defining the membership functions can be obtained.

4 Experimental Results

Figure 1 presents a simulated data base of two classes in a feature space of two attributes. The two classes are of different sizes, the first has 200 patterns and the second has 100 patterns. The distribution of each class has two independent normal variables with different standard deviations and means. Anyway, the two classes are not overlapped.

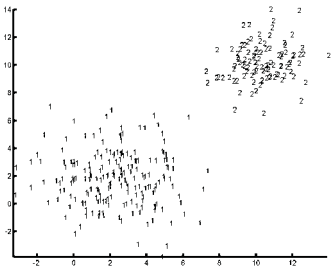


Figure 1: Example of two classes with different sizes, standard deviations and means.

We apply the proposed semi supervised FPM by selecting the patterns with a complete random order. The goal is to test the robustness of our algorithm against the initialisation problem. The experience is repeated several times with a different random pattern's occurrences at each time. We start the experience with a time window which is equal to the size of the data set, i.e., the merging step is applied after the reception of all the available patterns in the data base. Semi supervised FPM detects between 2 to 5 clusters according to the initialisation. Figure 2 shows the result of our algorithm in the case of the detection of 5 clusters.

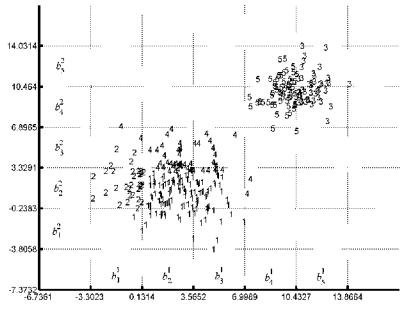


Figure 2: Results of the detection and local adaptation step of semi supervised FPM for the example of Figure 1.

The fuzzy similarity measures between the classes in this case are :

$\delta_{12} = 0.23$, $\delta_{13} = 0$, $\delta_{14} = 0.69$, $\delta_{15} = 0$, $\delta_{23} = 0$,
$\delta_{24} = 0.42$, $\delta_{34} = 0.04$, $\delta_{35} = 0.52$, $\delta_{45} = 0.04$

The merging threshold $\lambda_{\min} = 0.23$ is sufficient to merge the classes C_1, C_2 and C_4 into one cluster and C_3 with C_5 into one another cluster in order to obtain at the end the two necessary clusters. Based on the experimentation of several time windows' sizes, a time window of 30 patterns is sufficient to

ISBN: 978-989-95079-6-8

well estimate the correct number of clusters. Anyway, the width of the time window depends of the application: its dynamic, and the initialisation. If this width is too large, then the clusters merging will be delayed, while if it is too small the fuzzy similarity measure value may not be enough to validate the merging.

4.1 Classes of non convex shape

Supervised FPM works well if the classes are separated by at least one attribute. This is due to the fact that the classification of a pattern by FPM is based on a selection of one attribute. Another consequence of the selection of one attribute is that FPM does not respect the shape of classes if this shape is not convex, which is the case of the majority of real applications. Indeed, FPM provides always rectangular membership level curves for all the classes. Figure 3 presents a case for which the classes are not separated by at least one attribute. In addition, the classes C_1 and C_3 are of non convex shape. Figure 4 presents the results of the application of semi supervised FPM on this data, with a histogram containing 8 bins. Indeed, a higher number of histogram's bins is necessary when the classes are not separated by at least one attribute. A threshold equal to 0.14 is the required one for the merging measure to obtain the three classes.

Figure 5 shows the membership level curves obtained for this data set after the merging of clusters. Here the application of (11) leads to obtain one membership function for each class according to each attribute equivalent to the one resulting by the application of supervised FPM on each class after merging. We can see that these curves do not respect the shape non convex of classes C_1 and C_3. This, as we said before, is due to the classification decision based on the selection of one attribute. Inspired of the multi-prototypes approach [5] used in the literature to respect the shape of classes, we propose to merge the membership functions of the classes as follows :

$$\pi_{iz} = \min(1, \pi_i + \pi_z) \quad (12)$$

Where π_{iz} is the membership function after the fusion of classes C_i and C_z.

Using (12) means that each class is composed of several subclasses. Each subclass keeps its membership function. The application of (12) provides the membership level curves of Figure 6. We can see that these curves respect the classes shape. In addition, no supplementary computation is required to obtain this fusion. When the number of bins h increases, the number of subclasses increases also. This leads to obtain membership level curves which respect more precisely the classes' shape. However, increasing too much h entails the appearance of some membership peaks in the centre of classes.

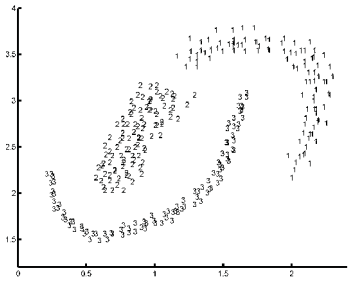


Figure 3: Classes which are not separated by at least one attribute and their shape is not convex (classes C_1 and C_3).

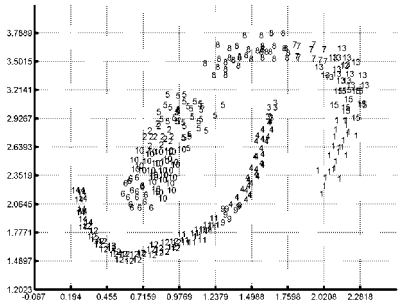


Figure 4: Clusters obtained by the application of semi supervised FPM on the data set of Figure 3.

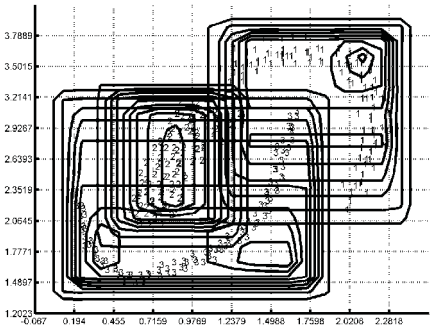


Figure 5: Membership level curves for the example of Figure 3. The membership functions of classes are obtained using (11).

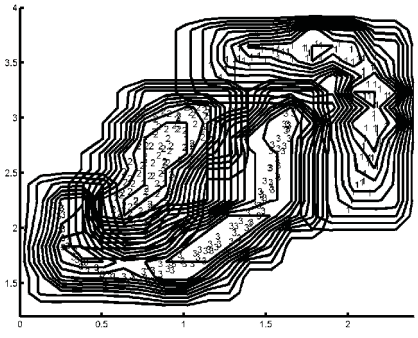


Figure 6: Membership level curves for the example of Figure 3 using (12).

ISBN: 978-989-95079-6-8

4.2 Overlapped classes

In the case of overlapped classes, the similarity measure becomes useless since the classes are not well separated. In addition, some points can be misclassified according to the ratio of labelled points to the total number of points in data sets. To allow testing the performance of semi supervised FPM in the case of overlapped classes, we take the following artificial 2-dimensioal dataset (Figure 7) available at http://www.stats.ox.ac.uk/pub/PRNN/. It is a normal mixtures data set. The training data consists of two classes with 125 patterns in each class. Each of the two classes has bimodal distribution. The testing set is an independent set of 1000 patterns drawn from the same distribution. The reported misclassification error is based on this testing set. The patterns were selected randomly which means that labelled patterns about some classes may not be presented at all. The experience is repeated 50 times to take into account the effect of initialisation. Table 1 presents the obtained results using semi supervised FPM. We can see that when *r* increases the misclassification error decreases thanks to the existence in advances of labelled patterns about some classes.

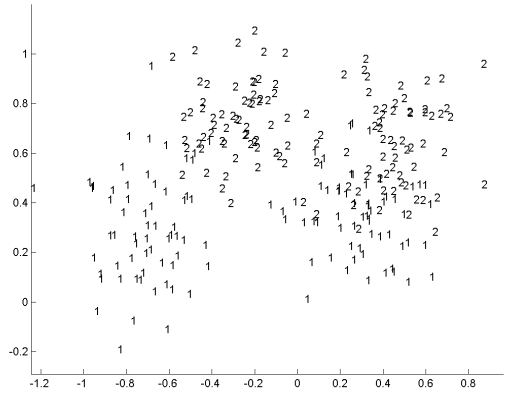


Figure 7: Normal mixtures data set.

Table 1: Misclassification Rate (*MR*) and its STandard Deviation (*STD*) in % according to different values of *r* for the normal mixture data set.

r %	0%	10%	20%	30%	40%
MR %	13.49	11.89	9.91	9.39	9.19
STD %	12.52	3.71	1.27	1.03	0.62
50%	60%	70%	80%	90%	100%
8.84	8.61	8.43	8.26	8.16	8.10
0.39	0.32	0.30	0.27	0.2	0

5 Conclusions

In this paper, the supervised classification method Fuzzy Pattern Matching (FPM) is developed to be also an unsupervised classification one. The goal is to obtain a semi supervised classification method adapted to dynamic systems for which a limited prior knowledge is available. Since the unsupervised learning technique is not based on a distance measure, the proposed algorithm will not favour the smaller sized clusters. In addition, it can start with no prior information. Finally, the membership functions can be adapted to elongated and non convex clusters.

We are developing FPM to be operant in the case of non stationary data. Indeed, in many practical situations, the environment changes. A learning data set used to construct the membership functions will be no more valid after a certain time. Thus, the classification method must be able to forget the information which is no more valid or representative of classes and adapt the membership functions based only on the recent and useful one.

Acknowledgements

This work is integrated in the regional project MOSYP (Performances Measurements and Optimization of Production Systems). The authors thank the region Champagne-Ardennes within this project (CPER ICOS).

References

[1] Benediktsson, J. A., J. R. Sveinsson, Feature extraction for multisource data classification with artificial neural networks, Int. journal of remote sensing 18(4) (1997) 727-740.

[2] Bezdek, J. C., N. R. Pal, Some new indexes of cluster validity, IEEE Tran. on Systems, Man and Cybernetics 28(3) (1998) 301-315.

[3] Cayrol, M., H. Farreny, H. Prade, Fuzzy Pattern Matching, Kybernetes 11, (1982) 103-116.

[4] Demiriz, K. P. B., M. J. Embrechts, Semi-supervised clustering using genetic algorithms, Artificial Neural Networks in Engineering, (1999) 809–814.

[5] Devillez, A., M. Sayed Mouchaweh, P. Billaudel, A process monitoring module based on fuzzy logic and pattern recognition, International Journal of Approximate Reasoning 37 (2004) 43-70.

[6] Dubois, D., H. Prade, On possibility/probability transformations, Fuzzy Logic (1993) 103-112.

[7] Duda, R. O., P. E. Hart, D. E. Stork, Pattern Classification second edition, Wiley, New York, (2001).

[8] Frigui, H., R. Krishnapuram, Robust algorithm for automatic extraction of an unknown number of clusters from noisy data, Pattern Recognition Letters 17 (1996) 1223-1232.

[9] Gabrys, B., L. Petrakieva, Combining labelled and unlabelled data in the design of pattern classification systems, International Journal of Approximate Reasoning 35(3), (2004) 251-273.

[10] Jain, A. K., P. Devijver, J. Kittler, Advances in statistical pattern recognition, Pattern Recognition Theory and Applications, Springer-Verlag, Berlin, Heidelberg, (1987) 1-19.

[11] Kate, R. J., R. J. Mooney, Semi-Supervised Learning for Semantic Parsing using Support Vector Machines, Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Rochester, NY, (2007) 81-84.

[12] Liu, H., L. Yu, Toward Integrating Feature Selection Algorithms for Classification and Clustering, IEEE Trans. on knowledge and data engineering 17 (4) (2005).

[13] Marsili-Libelli, S., Adaptive Fuzzy Monitoring and Fault Detection.Int. Journal of COMADEM 1(3) (1998) 31-37.

[14] Sayed Mouchaweh, M., A. Devillez, G. V. Lecolier, P. Billaudel, Incremental learning in real time using fuzzy pattern matching, Fuzzy Sets and Systems 132(1) (2002) 49-62.

[15] Zadeh, L. A., Fuzzy sets as a basis for a theory of possibility, Fuzzy sets and systems 1(3-2) (1978).

ISBN: 978-989-95079-6-8

Optimistic Fuzzy Weighted Average

Sylvie Galichet, Reda Boukezzoula[1]

1. Polytech'Savoie, LISTIC, Université de Savoie, Domaine Universitaire, BP 80439,
74944 Annecy le Vieux Cedex, France
Email: {sylvie.galichet, reda.boukezzoula}@univ-savoie.fr

Abstract *– The fuzzy weighted average (FWA) is used in many engineering problems where aggregation of fuzzy information is dealt with. In this framework, many algorithms have been proposed to compute efficiently the FWA according to Zadeh's extension principle. However, due to fuzzy interval calculus, the exact solution presents a characteristic that may be viewed as an important drawback. Indeed, replacing each individual by the fuzzy weighted average in the assessment of the population score leads to a result different from the original one. The presented work is an attempt to propose an optimistic counterpart for the FWA that eliminates the mentioned characteristic, possibly undesirable. Actually, the optimistic FWA is computed by using a modified division operator. When existing, the latter is the inverse operator of the fuzzy multiplication. Contrary to the conventional FWA, the optimistic one can be computed by a sequence of elementary arithmetic operations.*

Keywords *– Fuzzy arithmetic, Fuzzy weighted average, Gradual real numbers*

1 Introduction

The average is probably the easiest and the most widespread solution to aggregate information. The average is defined as the single value which all the individuals of a population should have so that their total is unchanged. When the sample contains several times the same individuals, it is possible to reformulate the average as a weighted average by introducing coefficients related to each individual. So a weighted average is an average in which a weight w_i is assigned to each quantity x_i to be averaged. The weights determine the relative importance of each quantity on the average. Considering N numbers x_i with associated weights w_i, the weighted average y is expressed as follows:

$$y = f(x_1, x_2, ..., x_N, w_1, w_2, ..., w_N) = \frac{x_1 w_1 + x_2 w_2 + ... + x_N w_N}{w_1 + w_2 + ... + w_N} \quad (1)$$

In an imprecise environment where information are poorly defined, it may be appropriate to represent scores and weighting coefficients by fuzzy numbers $\tilde{x}_i$ and $\tilde{w}_i$. In this case, the weighted average becomes a fuzzy number $\tilde{y}$ as well, i.e. the fuzzy weighted average (FWA).

If we follow the traditional fuzzy set theory, the fuzzy average $\tilde{y} = f(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_N, \tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_N)$ is obtained according to the extension principle that is:

$$\mu_{\tilde{y}}(y) = \sup_{\substack{x_1, .., x_N, w_1, .., w_N \ / \\ y = f(x_1, .., x_N, w_1, .., w_N)}} \min(\mu_{\tilde{x}_1}(x_1), .., \mu_{\tilde{x}_N}(x_N), \mu_{\tilde{w}_1}(w_1), .., \mu_{\tilde{w}_N}(w_N)) \quad (2)$$

Many authors proposed computational algorithms providing a discrete but exact solution of (2). All methods are based on the α-cut representation of fuzzy sets and interval analysis as initially suggested by Dong and Wong [5]. In interval computation, efficient algorithms exist for specific classes of functions, especially for the class of fractionally-linear functions [12]. The weighted average is an example of such functions and the problem of computing exactly the FWA is thus practically solvable. Liou and Wang [14] were the first to observe that since the x_i appear only in the numerator of (1), only the smallest values of the x_i are used to find the smallest value of (1), and only the largest values of the x_i are used to find the largest value of (1). This decomposition is used to reduce the algorithmic complexity in recent proposed approaches ([4], [8], [9], [12], [13], [15]).

Even if the above mentioned iterative procedures determine the exact solution for the FWA and so cope with multiple appearance of variables in the expression of the function f in (1), there is no known closed-form formula for computing $\tilde{y}$. Moreover, there is no available algorithm that could be directly implemented using elementary arithmetic operations between fuzzy operands. This aspect is studied in [16] where the authors are interested in designing a symbolic engine that could transform any given function into a sequence of elementary operations for which fuzzy interval computation would achieve exact result (without overestimation). Unfortunately, it was proven by Nguyen *et al.* [16] that operations with one or two fuzzy operands are not sufficient to describe generic functions on fuzzy sets.

In addition to the difficulty of obtaining the exact solution, the fuzzy weighted average has an important drawback directly related to fuzzy interval calculus. Indeed, replacing each individual by the fuzzy weighted average in the assessment of the population score leads to a result different from the original one. Actually, when the fuzzy average is viewed as expressing a requirement of maximal tolerance on a variable y which is itself the result of a computation involving quantities x_1, x_2, ..., x_N whose values are implicitly constrained by this calculation, it is desirable to solve the fuzzy

ISBN: 978-989-95079-6-8

equation:

$$B \otimes \tilde{y} = A \tag{3}$$

with $B = \tilde{w}_1 \oplus \tilde{w}_2 \; .. \oplus .. \; \tilde{w}_N$ (4)

and $A = (\tilde{w}_1 \otimes \tilde{x}_1) \oplus (\tilde{w}_2 \otimes \tilde{x}_2) \oplus \ldots \oplus (\tilde{w}_N \otimes \tilde{x}_N)$

for determining $\tilde{y}$. For example, such a procedure would be probably well suited to the case where the FWA is used for aggregating rule contributions in a Sugeno-like fuzzy system that would deal with fuzzy inputs.

In this context of «optimistic fuzzy interval calculus» [6], standard fuzzy operations can not be used directly for computing $\tilde{y}$. It means that a modified division denoted $\ominus_{\approx}$ must be searched for so that computing $\tilde{y}$ as $A \ominus_{\approx} B$ guarantees the satisfaction of the equality constraint (3). In the chosen context, both fuzzy quantities A and B are computed with the same weights $\tilde{w}_i$ according to (4) but the computation of $\tilde{y}$ does not handle explicitly weight interaction. The proposed approach is thus different from constrained fuzzy arithmetics as dealt with by Klir in [10] whose objective is an appropriate handling of interactive variables.

The motivation of this paper is twofold. Indeed, the presented work is an attempt:

- to propose an optimistic counterpart for the fuzzy weighted average,
- to compute the latter by a sequence of elementary operations.

This paper is organized as follows. Section 2 introduces fuzzy intervals with emphasis on their profile representation and their combination by arithmetic operators. In section 3, a new modified division operator is proposed for directly solving fuzzy equations. Then, section 4 is devoted to the implementation of the optimistic fuzzy weighted average with the presentation of two examples taken from the literature.

2 Fuzzy arithmetic operators

2.1 Fuzzy Intervals and Profiles

Let us consider a unimodal fuzzy interval A with kernel value K_A and support $S_A = [S_A^-, S_A^+]$ where S_A^- and S_A^+ are respectively the lower and upper bounds, of the interval S_A.

In order to specify the fuzzy interval shape, two additional functions are used to link the support with the kernel value. These functions, called left and right profiles, respectively denoted A^- and A^+, are defined from the membership function μ_A by:

$$\begin{aligned} A^-(\lambda) &= \text{Inf } \{x \mid \mu_A(x) \geq \lambda; x \geq S_A^-\} \\ A^+(\lambda) &= \text{Sup } \{x \mid \mu_A(x) \geq \lambda; x \leq S_A^+\} \end{aligned} \tag{5}$$

where $\lambda \in [0, 1]$. The profile A^- (resp. A^+) is an increasing (resp. decreasing) mapping that corresponds to the left (resp. right) part of the fuzzy interval A. It can be easily stated that:

$$K_A = A^-(1) = A^+(1), \tag{6}$$

$$S_{A=} [A^-(0), A^+(0)]. \tag{7}$$

Finally, the fuzzy interval A is univoquely defined by its left and right profiles. Thus, in the same way that the conventional interval S_A is denoted $[S_A^-, S_A^+]$, the fuzzy interval A will be denoted $[A^-, A^+]$. Equivalently, A can be viewed as the family of nested intervals $A(\lambda) = [A^-(\lambda), A^+(\lambda)]$, $\lambda \in [0, 1]$, when an explicit formulation with respect to λ is preferred.

For simplicity, the following additional notations are used in the remaining part of the paper. Given the interval $S = [S^-, S^+]$, its midpoint $M(S)$, its radius $R(S)$ and its relative extent $Rex(S)$ are respectively defined by:

$$M(S) = (S^+ + S^-) / 2, \tag{8}$$

$$R(S) = (S^+ - S^-) / 2, \tag{9}$$

$$Rex(S) = R(S) / M(S). \tag{10}$$

2.2 Arithmetic operators

Let $A = [A^-, A^+]$ and $B = [B^-, B^+]$ be two fuzzy intervals, the classical four arithmetic operations are expressed by:

$$A \oplus B = [A^- + B^-, A^+ + B^+] \tag{11}$$

$$A \ominus B = [A^- - B^+, A^+ - B^-] \tag{12}$$

$$A \otimes B = [\min Z, \max Z] \quad \text{where } Z = \{A^-B^-, A^-B^+, A^+B^-, A^+B^+\} \tag{13}$$

$$A \ominus B = [A^-, A^+] \otimes [1/B^+, 1/B^-] \quad \text{for } B \text{ such that } 0 \notin S_B \tag{14}$$

In order to cope with the twofold objective of solving equation (3) for determining the optimistic fuzzy average $\tilde{y}$ and of computing the fuzzy solution by a sequence of elementary operations, it is quite natural to search for a representation of $\tilde{y}$ in the form:

$$\tilde{y} = A \ominus B \tag{15}$$

where A and B are defined according to equation (4).

However, using the usual division and multiplication operators given by (14) and (13), it can be easily stated that substituting $\tilde{y}$ computed according to (15) into (3) gives a result more imprecise than the original A. Figure 1 illustrates this characteristic using two triangular fuzzy numbers A and B, with $K_A = 2.4$, $S_{A=} [0.8, 5.8]$ and $K_B = 1$, $S_{B=} [0.4, 1.9]$.

At best $A \subseteq B \otimes (A \ominus B)$ which means that the desired equality is generally not achieved. This problem is related to the lack of inverses in the calculus of fuzzy quantities.

Thus, a way around overestimation problems must be searched for outside standard arithmetic operations. One may think of using fuzzy arithmetic with requisite equality constraints as proposed in [10]. Klir's idea consists in doing fuzzy arithmetic with constraints dictated by the context of the problem. In practice, constraints are achieved by a-priori knowing that the α-cuts from two variables are the same. The approach is thus efficient for avoiding overestimation due to the occurrence of interactive variables. However, even when using constrained arithmetic, the calculus of fuzzy quantities is still pessimistic about the precision.

In the context of optimistic fuzzy calculus, we propose to use the modified division operator presented in next section.

ISBN: 978-989-95079-6-8

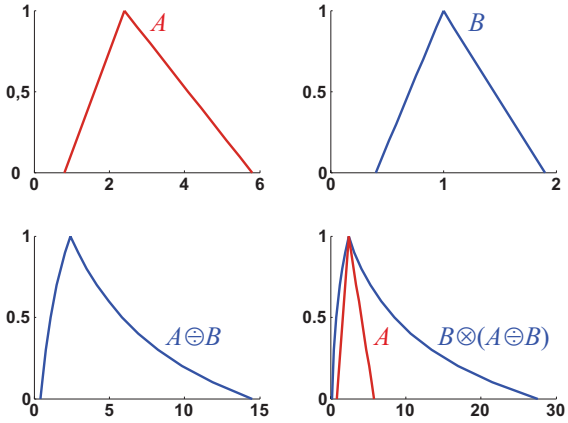


Figure 1: Fuzzy division and multiplication

3 The modified $\ominus_{\approx}$ Operator

3.1 Definition

Let $A = [A^-, A^+]$ and $B = [B^-, B^+]$ be two unimodal fuzzy intervals with $0 \notin S_B$. The result of the modified division $A \ominus_{\approx} B$ is defined from two gradual real numbers [7], denoted Φ^- and Φ^+, defined by their assignment functions from [0, 1] to the reals, as :

$$\Phi^-(\lambda) = \frac{Num^-(\lambda)}{Den^-(\lambda)} = \frac{M(A(\lambda)) - R(A(\lambda)) \cdot sign(M(S_B))}{M(B(\lambda)) - R(B(\lambda)) \cdot sign(Num^-(\lambda))} \quad (16)$$

and

$$\Phi^+(\lambda) = \frac{Num^+(\lambda)}{Den^+(\lambda)} = \frac{M(A(\lambda)) + R(A(\lambda)) \cdot sign(M(S_B))}{M(B(\lambda)) + R(B(\lambda)) \cdot sign(Num^+(\lambda))} \quad (17)$$

It is clear that, given A and B, both gradual numbers Φ^- and Φ^+ can always be computed. It is thus always possible to define $A \ominus_{\approx} B$ as the ordered pair (Φ^-, Φ^+) considering Φ^- as a left profile and Φ^+ as a right profile. Such entities are studied in [11] where it is proposed to add an extra feature, called "orientation", to pairs of profiles so as to build "ordered fuzzy numbers".

It is shown in [2] that the modified division $\ominus_{\approx}$ is the inverse operator of $\otimes$, that is $B \otimes (A \ominus_{\approx} B) = (A \ominus_{\approx} B) \otimes B = A$. Further details on the way equations (16) and (17) were established can be found in [3].

There is no guarantee that the ordered pair (Φ^-, Φ^+) is a gradual interval [7]. In other words, computations of Φ^- and Φ^+ according to (16) and (17) do not guarantee that $\forall \lambda \in [0, 1], \Phi^-(\lambda) \le \Phi^+(\lambda)$. It is however possible to determine a necessary and sufficient condition on operands A and B, under which $A \ominus_{\approx} B = (\Phi^-, \Phi^+)$ is the gradual interval $[\Phi^-, \Phi^+]$.

For unimodal operands, requiring that $\Phi^- \le \Phi^+$ at the kernel and support levels, that is :

$$\Phi^-(1) = \Phi^+(1) = K_\Phi$$
$$\Phi^-(0) \le K_\Phi \le \Phi^+(0)$$

leads to the following condition on A and B:

$$\frac{1 - |Rex(S_A)|}{1 - S \cdot |Rex(S_B)|} \le \frac{\delta K}{\delta M} \le \frac{1 + |Rex(S_A)|}{1 + |Rex(S_B)|} \quad (18)$$

where $\delta K = K_A / K_B$, $\delta M = M(S_A) / M(S_B)$ and $S = sign(1 - |Rex(S_A)|)$. Extending (18) to all λ-cuts leads to a necessary and sufficient condition under which $[\Phi^-, \Phi^+]$ is a gradual interval. Without any additional monotonicity assumption, the obtained interval is not always a fuzzy interval in the sense that λ-cuts are not necessarily nested.

Next subsection illustrates different behaviors of the pair (Φ^-, Φ^+).

3.2 Modified division of fuzzy triangular sets

Using the modified division operator for computing the division $A \ominus_{\approx} B$ where A and B are the triangular fuzzy subsets dealt with previously in figure 1, the fuzzy subset plotted in figure 2 is obtained. It can be stated that the latter is less imprecise than the standard division $A \ominus B$. As desired, the computed result is the exact solution of the fuzzy equation $B \otimes X = A$. It induces that X is implicitly less imprecise than A which is clearly verified in figure 2.

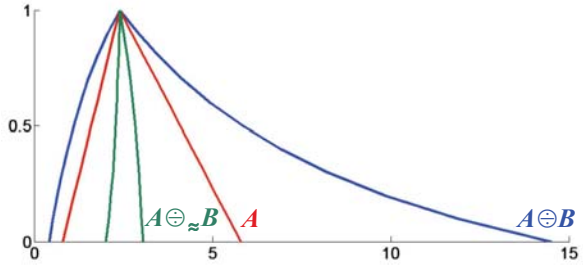


Figure 2: Modified fuzzy division

One may try to compute $1 \ominus_{\approx} A$ in order to determine the inverse of A according to $\ominus_{\approx}$ where 1 is the degenerated fuzzy subset associated with the crisp value 1. Considering the fuzzy subset A defined previously in figure 1, the result of the computation using equations (16) and (17) is given in figure 3. It appears clearly that (Φ^-, Φ^+) is not a gradual interval since the left and right profiles are exchanged, that is: $\Phi^+(\lambda) < \Phi^-(\lambda)$ for $\lambda \neq 1$.

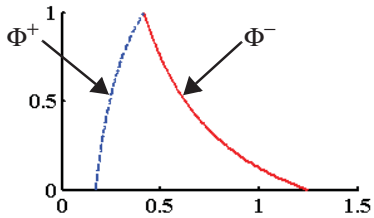


Figure 3: Computation of Φ^- and Φ^+ for $1 \ominus_{\approx} A$

It would have been possible to detect the «non-inversibility» of A prior to calculus according to the violation of condition (18).

The «non-inversibility» statement can be generalized to any

ISBN: 978-989-95079-6-8

non degenerated fuzzy subset. Indeed, let us suppose a fuzzy unimodal interval A, $0 \notin S_A$, such that $1 \ominus_{\approx} A$ exists. In this case, condition (18) has to be verified. As $K_1 = M(S_1) = 1$ and $R(S_1) = Rex(S_1) = 0$, condition (18) becomes:

$$\frac{1}{1-\left|Rex(S_A)\right|} \leq \frac{M(S_A)}{K_A} \leq \frac{1}{1+\left|Rex(S_A)\right|} \qquad (19)$$

As $0 \notin S_A$, $M(S_A)$ and K_A have the same sign. Let us first consider that $M(S_A)$ and K_A are both positive. Then, inequality (19) is rewritten as follows:

$$\frac{M(S_A)}{M(S_A)-R(S_A)} \leq \frac{M(S_A)}{K_A} \leq \frac{M(S_A)}{M(S_A)+R(S_A)} \qquad (20)$$

otherwise expressed as:

$$S_A^+ = M(S_A)+R(S_A) \leq K_A \leq M(S_A)-R(S_A) = S_A^- \qquad (21)$$

The single case where condition (21) holds, i.e. A is invertible, is for A being a crisp positive number. The case when $M(S_A)$ and K_A are both negative leads to a similar conclusion.

3.3 Modified division for computing FWA

In the context of computing weighted average, weights are usually supposed to be positive numbers. Extending to fuzzy weighted average, fuzzy positive weights $\tilde{w}_i$ are also assumed. Thus, computing the FWA $\tilde{y}$ as $A \ominus_{\approx} B$ with B being the sum of the weights, i.e. $B = \tilde{w}_1 \oplus \tilde{w}_2 \,..\oplus..\, \tilde{w}_N$, one can restrict the closed form of the modified division to the case of B positive. Equations (16) and (17) are then rewritten as $A \ominus_{\approx} B = [\Phi^-, \Phi^+]$, with

$$\Phi^-(\lambda) = \frac{A^-(\lambda)}{M(B(\lambda)) - R(B(\lambda)) \cdot sign(A^-(\lambda))} \qquad (22)$$

and

$$\Phi^+(\lambda) = \frac{A^+(\lambda)}{M(B(\lambda)) + R(B(\lambda)) \cdot sign(A^+(\lambda))}. \qquad (23)$$

4 Examples

4.1 Two-term average

To illustrate the proposed method for computing the fuzzy weighted average, a two-term example discussed in [9] is considered. Hence, we focus on the computation of:

$$\tilde{y} = (\tilde{x}_1 \otimes \tilde{w}_1 \oplus \tilde{x}_2 \otimes \tilde{w}_2) \ominus_{\approx} (\tilde{w}_1 \oplus \tilde{w}_2) \qquad (24)$$

where fuzzy triangular scores $\tilde{x}_1$, $\tilde{x}_2$ and fuzzy triangular weights $\tilde{w}_1$, $\tilde{w}_2$ are illustrated in figure 4. Using the profile representation, it follows:

$\tilde{x}_1 = [\lambda, 2-\lambda]$,
$\tilde{x}_2 = [\lambda+2, 4-\lambda]$,
$\tilde{w}_1 = [0.3\,\lambda, 0.9-0.6\lambda]$,
$\tilde{w}_2 = [0.3\lambda+0.4, 1-0.3\lambda]$.

The computation of $\tilde{y}$ according to (24) leads to the fuzzy result plotted in figure 5. As both fuzzy variables $\tilde{x}_1$ and $\tilde{x}_2$ are positive, according to (22) and (23) the developed forms of the left and right profiles of $\tilde{y}$ are given by :

$$\tilde{y}^- = \frac{\tilde{x}_1^- \cdot \tilde{w}_1^- + \tilde{x}_2^- \cdot \tilde{w}_2^-}{\tilde{w}_1^- + \tilde{w}_2^-} \qquad (25)$$

$$\tilde{y}^+ = \frac{\tilde{x}_1^+ \cdot \tilde{w}_1^+ + \tilde{x}_2^+ \cdot \tilde{w}_2^+}{\tilde{w}_1^+ + \tilde{w}_2^+} \qquad (26)$$

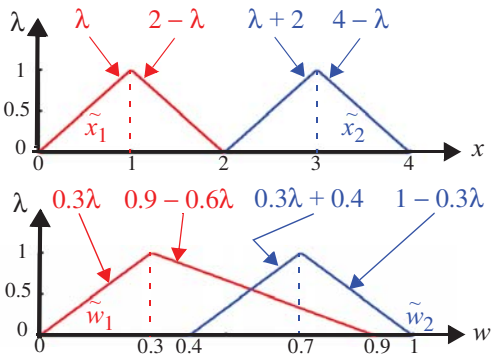


Figure 4: Fuzzy quantities and weights

Althrough the λ parametrization has been omitted in equations (25) and (26), all involved profiles are fonctions from the unit interval [0, 1] to the real line. It means that the addition, multiplication and division operations are function operations.

It can be observed that the computed optimistic FWA is a gradual interval. However, that one is not a fuzzy interval since its left profile is not an increasing function with respect to λ.

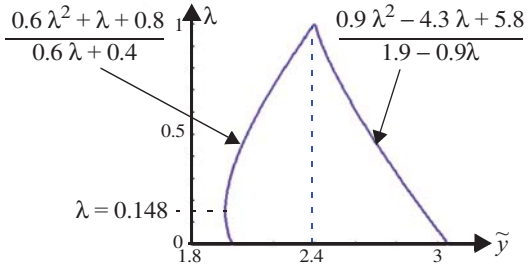


Figure 5: Optimistic fuzzy weighted average

For comparison purpose, figure 6 regroups the exact FWA according to Zadeh's extension principle with its lower and upper approximations, respectively obtained with the modified division operator and the conventional one. At the kernel level (λ = 1), all approaches give the same result, i.e. the weighted average for precise weights and numbers. At the other levels, the following inclusion is obtained:

$$\mathrm{FWA}_{\mathrm{optimistic}} \subseteq \mathrm{FWA}_{\mathrm{exact}} \subseteq \mathrm{FWA}_{\mathrm{arithmetic}}. \qquad (27)$$

In the very simple example under consideration, it can be observed that the $\tilde{x}_i$ variables are sorted. Then, according to [13], the analytical closed form of the exact FWA is expressed by:

$$\tilde{y}^- = \frac{\tilde{x}_1^- \cdot \tilde{w}_1^+ + \tilde{x}_2^- \cdot \tilde{w}_2^-}{\tilde{w}_1^+ + \tilde{w}_2^-} \qquad (28)$$

$$\tilde{y}^+ = \frac{\tilde{x}_1^+ \cdot \tilde{w}_1^- + \tilde{x}_2^+ \cdot \tilde{w}_2^+}{\tilde{w}_1^- + \tilde{w}_2^+} \qquad (29)$$

ISBN: 978-989-95079-6-8

However, in the general case it may not be possible to order the $\tilde{x}_i$ fuzzy variables and numerical methods have to be applied for discretized λ values. In this context, mathematical programming techniques may be preferred ([9], [8]). Methodologies based on the handling of gradual numbers are also sensitive to the ordering problem. It is thus suggested in [7] to divide the λ-range in several parts for which specific orders are defined.

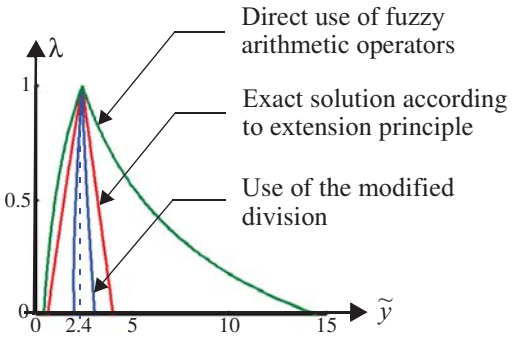


Figure 6: Fuzzy weighted average

4.2 Five-term example

This example taken from [13] consists in computing a five-term FWA, with the fuzzy triangular numbers $\tilde{x}_i$ and $\tilde{w}_i$, i = 1, ..., 5, given in table 1 and illustrated in figure 7.

Table 1: Fuzzy numbers and fuzzy weights

$[\, S_{\tilde{x}}^{-} \;\; K_{\tilde{x}} \;\; S_{\tilde{x}}^{+} \,]$	$[\, S_{\tilde{w}}^{-} \;\; K_{\tilde{w}} \;\; S_{\tilde{w}}^{+} \,]$
[1 2 3]	[1 2 5]
[2 5 7]	[2 2.5 3]
[6 8 9]	[4 7 9]
[7 9 10]	[3 4 7]
[10 11 12]	[2 3 4]

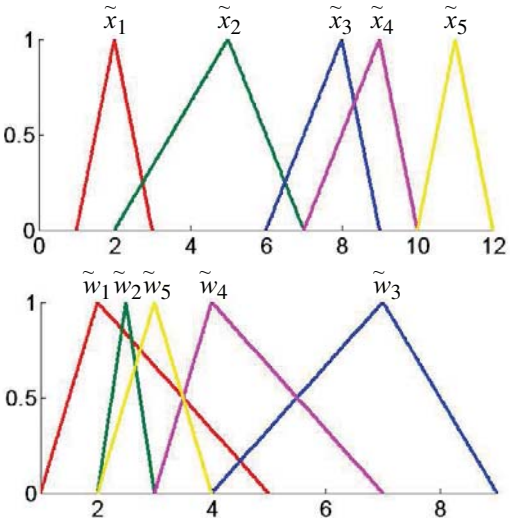


Figure 7: Fuzzy numbers and fuzzy weights

Using the modified division operator to compute the FWA, the result plotted in figure 8 is obtained. It can be observed that the optimistic FWA is here a fuzzy number, i.e. both left and right profiles are monotonic. As expected, the inclusion property (27) is valid.

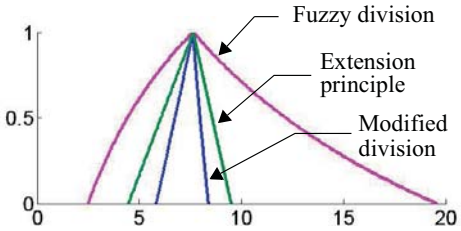


Figure 8: Five-term FWA

5 Conclusion

Based on the use of a modified fuzzy division operator, an optimistic counterpart of the usual fuzzy weighted average was proposed. The modified division is the inverse operation of the fuzzy multiplication. Consequently, the computed optimistic fuzzy average can replaced all individuals without modifying the weighted sum of the population. Moreover, the computation can be achieved by a sequence of elementary arithmetic operations. Unfortunately, the optimistic average may not be a fuzzy number but only an interval of gradual numbers. Actually, the modified division definition only guarantees that the result can be expressed in the form of an ordered pair of gradual numbers without further requirement on the profile monotonicity.

Two simple examples from the fuzzy literature have been used for illustration. More complicated and realistic cases must be further tested. For example, the proposed optimistic average may be used in the context of determining cluster centers for linguistic fuzzy C-means ([1]). Another possible use may be found in the aggregation of Sugeno-like rule consequents. More conceptual works have also to be developed in order to correctly position the optimistic FWA with respect to the exact one.

References

[1] S. Auephanwiriyakul, J. M. Keller, "Analysis and Efficient Implementation of a Linguistic Fuzzy C-Means", IEEE Trans. on Fuzzy Systems, Vol. 10, no. 5, pp. 563-582, 2002.

[2] R. Boukezzoula, S. Galichet, L. Foulloy, "Inverse arithmetic operators for fuzzy intervals", EUSFLAT'07, Ostrava, 2007.

[3] R. Boukezzoula, L. Foulloy, S. Galichet, "Inverse controller design for interval fuzzy systems, IEEE Trans. on fuzzy systems, Vol. 14, no. 1, pp. 111-124, 2006.

[4] P.T. Chang, K.C. Hung, K.P. Lin, C.H. Chang, "A comparison of discrete algorithms for fuzzy weighted average", IEEE Trans. on Fuzzy Systems, Vol. 14, no. 5, pp. 663-675, 2006.

[5] W.M. Dong, F.S. Wong, "Fuzzy weighted averages and implementation of the extension principle", Fuzzy Sets

ISBN: 978-989-95079-6-8

and Systems, no. 21, pp. 183-199, 1987.

[6] D. Dubois, E. Kerre, R. Mesiar, H. Prade, "Fuzzy interval analysis", in Fundamentals of Fuzzy Sets, The handbooks of fuzzy sets series, Eds. D. Dubois and H. Prade, 2000.

[7] J. Fortin, D. Dubois, H. Fargier, "Gradual numbers and their application to fuzzy interval analysis", IEEE Trans. on Fuzzy Systems, Vol. 16, no. 2, pp. 388-402, 2008.

[8] Y.Y. Guh, C.C. Hon, E.S. Lee, "Fuzzy weighted average: the linear programming approach via Charnes and Cooper's rule", Fuzzy Sets and Systems, no. 117, pp. 157-160, 2001.

[9] C. Kao, S.T. Liu, "Fractional programming approach to fuzzy weighted average", Fuzzy Sets and Systems, no. 120, pp. 435-444, 2001.

[10] G.J. Klir, "Fuzzy arithmetic with requisite constraints", Fuzzy Sets and Systems, no. 91, pp. 165-175, 1997.

[11] W. Kosinski, "On fuzzy number calculus", Int. J. Appl. Math. Comput. Sci., Vol. 16, no. 1, pp. 51-57, 2006.

[12] R. Lea, V. Kreinovich, R. Trejo, "Optimal interval enclosures for fractionally-linear functions, and their application to intelligent control", Reliable Computing, Vol. 2, no. 3, pp. 265-286, 1996.

[13] D.H. Lee, D. Park, "An efficient algorithm for fuzzy weighted average", Fuzzy Sets and Systems, no. 87, pp. 39-45, 1997.

[14] T.-S. Liou, M.J.J. Wang, "Fuzzy weighted average: an improved algorithm", Fuzzy Sets and Systems, no. 49, pp. 307–315, 1992.

[15] F. Liu, J. Mendel, "Aggregation Using the Fuzzy Weighted Average as Computed by the Karnik–Mendel Algorithms", IEEE Trans. on Fuzzy Systems, Vol. 16, no. 1, pp. 1-12, 2008.

[16] H.T. Nguyen, V. Kreinovich, V. Nesterov, M. Nakamura, "On hardware support for interval computations and soft computing: Theorems", IEEE Trans. on Fuzzy Systems, Vol. 5, no. 1, pp. 108-127, 1997.

ISBN: 978-989-95079-6-8

Evolving Inverse Fuzzy Models for Uncalibrated Visual Servoing in 3D Workspace

P. J. S. Gonçalves[1,2] J.M.C. Sousa[2] J.R. Caldas Pinto[2]

1.Instituto Politécnico de Castelo Branco, Escola Superior de Tecnologia
Av. Empresário, 6000-767 Castelo Branco, Portugal
2.IDMEC/IST,Technical University of Lisbon (TU Lisbon)
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
Email: pgoncalves@est.ipcb.pt, {j.sousa,jcpinto}@dem.ist.utl.pt

Abstract— *In this paper, evolving inverse fuzzy models obtained online for uncalibrated visual servoing in 3D workspace, are developed and validated in a six degrees of freedom robotic manipulator. This approach will recursively update the inverse fuzzy model based only on measurements at a given time instant. The uncalibrated approach does not require calibrated kinematic and camera models, as needed in classical visual servoing to obtain the Jacobian. Experimental results obtained in a PUMA robot performing eye-to-hand visual servoing in 3D workspace are used to demonstrate the validity of the proposed approach, when compared to the previous developed off-line learning.*

Keywords— Visual Servoing, Fuzzy Control, Robotics, Evolving Fuzzy Systems.

1 Introduction

Industrial robotic manipulators use sensor-based control to perform tasks, either in structured or unstructured environments. Vision sensors provide a vast amount of information on the environment in which robots move. Thus, vision is essential for robots working in unstructured environments. This type of sensors definitely enlarges the potential applications of the actual robotic manipulators. Visual servoing can be defined as a method to control dynamic systems using the information provided by visual sensors. In this paper, the control of a robot manipulator end-effector using a single camera looking to the robot (eye-to-hand) is addressed [1].

Early approaches for visual servoing are based on a model already known, i.e. the robot-camera model, from which the relation between the features and the robot kinematics is analytically obtained [2]. Apart from the stated approaches where the robot-camera model is already known, it can also be estimated [3, 2]. This type of systems, *Uncalibrated Visual Servoing*, can deal with unknown robot kinematics and/or unknown camera parameters. By using this type of robot-camera models, the system becomes independent of robot type, camera type or even camera location.

In this paper, the robot-camera model estimation, from motions in the 3D workspace, by learning is addressed both online and off-line using fuzzy techniques to obtain a controller capable of controlling the system. An inverse fuzzy model is used to derive the inverse robot-camera model, i.e. the Jacobian, in order to compute the joints and end-effector velocities in a straightforward manner. The inverse fuzzy model cab be applied directly as a controller, which is a simple way to implement a controller in real-time. Note that this feature is very important in robotic systems.

From the modeling techniques based on soft computing, fuzzy modeling is one of the most appealing. As the robotic manipulator together with the visual system is a nonlinear system, which is only partly known, it is advantageous to use fuzzy modeling as a way to derive a model (or an inverse model as in this case) based only on measurements. In this paper will be applied and compared two approaches, off-line [2], and on-line [4, 5]. The first approach have already proven to perform well in visual servoing systems. It's main drawback is that the learning must be performed off-line. Since the environment in which the robot operates can vary importantly, new models should be derived very often, which take time and also the robot must stop the task that is performing. That is why an on-line modeling of the visual servoing system is preferable. This approach was not yet applied to control robotic manipulators using vision and is the objective of the present paper.

The paper is organized as follows. Section 2 describes briefly the concept of visual servoing and presents the uncalibrated visual servoing approach. Section 3 presents very briefly on-line and off-line fuzzy modeling, and discusses the problem of identifying inverse fuzzy models directly from data measurements. Section 4 describes the experimental setup and presents the obtained results, where the identified inverse fuzzy models are discussed. Finally, Section 5 presents the conclusions and the future work.

2 Visual Servoing

Machine Vision and Robotics can be used together to control a robotic manipulator. This type of control, which is defined as Visual Servoing, uses visual information from the work environment to control a robotic manipulator performing a given task. In the following sub-sections classical and uncalibrated visual servoing are presented, along with their main goals.

2.1 Classical Visual Servoing

Within the visual servoing framework, there are three main categories related with the information obtained from the image features:

- *image-based visual servoing* [1, 6], which uses direct information from the object in the image, i.e. image features,

ISBN: 978-989-95079-6-8

- *position-based visual servoing* [1], which uses 3D information of the object from the image(s), e.g. the 3D coordinates of the object or the rotation and translation between the camera and object frames, obtained using a CAD model of the object to perform an on-line pose estimation,
- *hybrid visual servoing* [7], which combines the first two approaches and is a possible solution to some drawbacks of image-based and position-based methods [8].

In the following, the classical approaches to modeling and control of position-based visual servoing are presented.

2.1.1 Modeling the Position-Based Visual Servoing System

In this paper position-based visual servoing with 3D features [9] is used in an eye-to-hand system [1], where the camera is fixed and looking the robot and object. The 3D image features, s are 3D points of the object in the camera frame, p. The kinematic modeling of the transformation between the image features velocities, $\dot{s}$, and the joints velocities $\dot{q}$ is defined as follows [9]:

$$\dot{s} = \begin{bmatrix} -I_3 & S(p) \end{bmatrix} \cdot {}^{c}W_e \cdot {}^{e}J_R \cdot \dot{q} = J \cdot \dot{q}\,, \qquad (1)$$

where I_3 is the 3×3 identity matrix, $S(p)$ is the skew-symmetric matrix of the 3D point p, ${}^{c}W_e$ is defined as the transformation between the camera and end-effector frames velocities, and ${}^{e}J_R$ is the robot Jacobian. The 3D point is obtained from the captured image using a pose estimation algorithm [9].

2.2 Uncalibrated Visual Servoing

To derive an accurate Jacobian, J, a perfect modeling of the camera, the chosen image features, the position of the camera related to the world, and the depth of the target related to the camera frame must be accurately determined. Even when a perfect model of the Jacobian is available, it can contain singularities, which hampers the application of a control law. Remind that the Jacobian must be inverted to send the camera velocity to the robot inner control loop. When the Jacobian is singular, the control cannot be correctly performed.

There are visual servoing systems that obviate the calibration step and estimate the robot-camera model either online or offline. The robot-camera model may be estimated:

- analytically, using nonlinear least square optimization [10], and
- by learning or training, using fuzzy membership functions and neural networks [11, 2].

In addition, the control system may estimate an image Jacobian and use the known robot model, or a coupled robot-camera Jacobian may be estimated.

To overcome the difficulties regarding the Jacobian, a new type of differential relationship between the features and camera velocities was proposed in [11]. This approach estimates the variation of the image features, when an increment in the camera position is given, by using a relation G. This relation is divided into G_1 which relates the position of the camera and the image features, and F which relates their respective variation:

$$s + \delta s = G(q + \delta\, q) = G_1(q) + F(q, \delta\, q). \qquad (2)$$

Considering only the variations in (2):

$$\delta s = F(q, \delta\, q), \qquad (3)$$

the inverse function F^{-1} is given by:

$$\delta q = F^{-1}(\delta s, q)\,, \qquad (4)$$

and it states that the joint variation depends on the image features variation and the previous position of the robot manipulator. Equation (4) can be discretized as

$$\delta q(k) = F_k^{-1}(\delta s(k+1), q(k)). \qquad (5)$$

In image-based visual servoing, the goal is to obtain a joint velocity, $\delta q(k)$, capable of driving the robot according to a desired image feature position, $s(k+1)$, with an also desired image feature error, $\delta s(k+1)$, from any position in the joint spaces. This goal can be accomplished by modeling the inverse function F_k^{-1}, using inverse fuzzy modeling as proposed in this paper and presented in Section 3. This new approach to image-based visual servoing allows to overcome the problems stated previously regarding the Jacobian and the calibration of the robot-camera model.

3 Inverse Fuzzy Modeling

3.1 Off-Line Fuzzy Modeling

Fuzzy modeling often follows the approach of encoding expert knowledge expressed in a verbal form in a collection of if–then rules. Parameters in this structure can be adapted using input-output data. When no prior knowledge about the system is available, a fuzzy model can be constructed entirely on the basis of system measurements. In the following, we consider data-driven modeling based on fuzzy clustering [12, 13].

We consider rule-based models of the Takagi-Sugeno (TS) type. TS models consist of fuzzy rules describing a local input-output relation, typically in an affine form:

$$\begin{aligned} R_i : \textbf{If } & x_1 \text{ is } A_{i1} \textbf{ and } \ldots \textbf{and } x_n \text{ is } A_{in} \\ \textbf{then } & y_i = \mathbf{a}_i \mathbf{x} + b_i\,, \quad i = 1, 2, \ldots, K. \end{aligned} \qquad (6)$$

Here R_i is the ith rule, $\mathbf{x} = [x_1, \ldots, x_n]^T$ are the antecedent variables, $A_{i1}, \ldots, A_{in}$ are fuzzy sets defined in the antecedent space, and y_i is the rule output variable. K denotes the number of rules in the rule base, and the aggregated output of the model, $\hat{y}$, is calculated by taking the weighted average of the rule consequents:

$$\hat{y} = \frac{\sum_{i=1}^{K} \beta_i y_i}{\sum_{i=1}^{K} \beta_i}\,, \qquad (7)$$

where β_i is the degree of activation of the ith rule: $\beta_i = \Pi_{j=1}^{n} \mu_{A_{ij}}(x_j)$, $i = 1, \ldots, K$, and $A_{ij}(x_j) : \mathbb{R} \rightarrow [0, 1]$ is the membership function of the fuzzy set A_{ij} in the antecedent of R_i.

To identify the model in (6), the regression matrix X and an output vector $\mathbf{y}$ are constructed from the available data: $\mathrm{X}^T = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$, $\mathbf{y}^T = [y_1, \ldots, y_N]$, where $N \gg n$ is the number of samples used for identification. The number of rules, K, the antecedent fuzzy sets, A_{ij}, and the consequent parameters, $\mathbf{a}_i, b_i$ are determined by means of fuzzy clustering

ISBN: 978-989-95079-6-8

in the product space of the inputs and the outputs [13]. Hence, the data set Z to be clustered is composed from X and $\mathbf{y}$: $\mathrm{Z}^T = [\mathrm{X}, \mathbf{y}]$. Given Z and an estimated number of clusters K, the Gustafson-Kessel fuzzy clustering algorithm [14] is applied to compute the fuzzy partition matrix U.

The fuzzy sets in the antecedent of the rules are obtained from the partition matrix U, whose ikth element $\mu_{ik} \in [0, 1]$ is the membership degree of the data object $\mathbf{z}_k$ in cluster i. One-dimensional fuzzy sets A_{ij} are obtained from the multi-dimensional fuzzy sets defined point-wise in the ith row of the partition matrix by projections onto the space of the input variables x_j. The point-wise defined fuzzy sets A_{ij} are approximated by suitable parametric functions in order to compute $\mu_{A_{ij}}(x_j)$ for any value of x_j.

The consequent parameters for each rule are obtained as a weighted ordinary least-square estimate. Let $\theta_i^T = [\mathbf{a}_i^T; b_i]$, let X_e denote the matrix $[\mathrm{X}; \mathbf{1}]$ and let W_i denote a diagonal matrix in $\mathbb{R}^{N \times N}$ having the degree of activation, $\beta_i(\mathbf{x}_k)$, as its kth diagonal element. Assuming that the columns of X_e are linearly independent and $\beta_i(\mathbf{x}_k) > 0$ for $1 \leq k \leq N$, the weighted least-squares solution of $\mathbf{y} = \mathrm{X}_e\theta + \epsilon$ becomes

$$\theta_i = \left[\mathrm{X}_e^T \mathrm{W}_i \mathrm{X}_e\right]^{-1} \mathrm{X}_e^T \mathrm{W}_i \mathbf{y}\,. \tag{8}$$

3.2 On-Line Fuzzy Modeling

The model obtained from the techniques presented in the previous section is assumed to be fixed, since it is learned in off-line mode. Recently attention is focused in on-line learning [4], where in a first phase input-output data is partitioned using unsupervised clustering methods and in a second phase, parameter identification is performed using a supervised learning method.

In On-Line Fuzzy Modeling and according to [4], also rule-based models of the TS type, are considered. Typically in the affine form described in (6), where the input-output data is acquired continuously. The new data, arriving at some time instant, can bring new information from the system, which could indicate a change in its dynamics. This information may change an existing rule, by changing the spread of the membership functions, or even introduce a new one. To achieve this, the algorithm must be able to judge the informative potential and the importance of the new data.

In the following are briefly presented the several steps of the algorithm used for on-line fuzzy modeling, proposed in [4], evolving fuzzy systems. The first step is based on the subtractive clustering algorithm [15], where the input-output data is partitioned. The procedure used must be initialized, i.e. the focal point of the first cluster is equal to the first data point and its potential is equal to one. Starting from the first data point, the potential of the next data point is calculated recursively using a Cauchy type function of first order:

$$P_k(z_k) = \frac{1}{1 + \frac{1}{k-1}\sum_{i=1}^{k-1}\sum_{j=1}^{n+1}(d_{ik}^j)^2}, \quad k = 2, 3, \ldots \tag{9}$$

where $P_k(z_k)$ denotes the potential of the data point z_k calculated at time k; $d_{ik}^j = z_i^j - z_k^j$, denotes projection of the distance between two data points (z_i^j and z_k^j) on the axis z^j.

When a new data point arrives it also influences the potential of the already defined center of the clusters ($z_i^*, i = 1, 2, ..., K$). A recursive formula for the update of the cluster centers potential is defined in [4]:

$$P_k(z_i^*) = \frac{(k-1)P_{(k-1)}(z_i^*)}{k - 2 + P_{(k-1)}(z_i^*) + P_{(k-1)}(z_i^*) + \sum_{j=1}^{n+1}(d_{ik}^j)^2},$$

where $P_k(z_i^*)$ is the potential at time k of the cluster center, related to the rule i.

The next step of the algorithm is to compare the potential of the actual data point to the potentials of the existing cluster centers. According to the two approaches tested in this paper [4, 5] a crucial part of the evolving fuzzy systems is the rule creation and modification. The first approach is called Evolving Takagi-Sugeno (eTS) and the second eXtended Takagi-Sugeno (xTS).

In eTS, if the potential of a new data point is higher than the potential of the existing cluster centers, then the new data point is accepted as a new cluster center and a new rule is formed. If in addition to the previous condition the new data point is close to an old cluster center, the old cluster center is replaced. This last condition is defined by:

$$\frac{P_k(Z_k)}{\underset{i=1}{\overset{K}{max}} P_k(z_i^*)} - \frac{\delta_{min}}{r} \geq 1, \tag{10}$$

where $r \in [0.3; 0.5]$ is the spread of the antecedent [4], and δ_{min} is the shortest distance between the new candidate Z_k and all the existing cluster centers z_i^*.

In xTS, if the potential of a new data point is higher than the maximum potential of all the existing clusters centers or is lower than the minimum potential of all the existing clusters centers, then the new data point is accepted as a new cluster center and a new rule is formed. In addition to the previous condition, if the next condition is true then the old cluster is replaced because the candidate is close to an old cluster center:

$$\underset{i=1}{\overset{K}{min}} \|z_k - z_i^*\|_j \geq \frac{r_j^i}{2}, \tag{11}$$

where r_j^i is the spread that is not fixed, like in the previous approach. In this approach the spread is adaptive and can vary in each element j of every cluster (rule) i, making possible to obtain hyper-ellipsoidal clusters. The recursive formula to obtain the adaptive spread is as follows:

$$r_{jk}^p = \rho \cdot r_{j(k-1)}^p + 0.5\sqrt{D_{jk}^p}; p = \underset{i=1}{\overset{K}{argmin}} \|z_k - z_i^*\|, \tag{12}$$

where D_{jk}^p is the local scatter and $\rho = 0.75$, both defined in [5].

For both approaches the fuzzy sets in the antecedent of the rules are gaussian, with the form:

$$\mu_{ij} = e^{-r\|x_j - x_{ij}^*\|^2}, \tag{13}$$

The consequents of the fuzzy rules are obtained using the global parameter estimation procedure based on the weighted recursive least squares, presented in [4].

3.3 Inverse modeling

For the robotic application in this paper, the inverse model is identified using input-output data from the inputs $\dot{q}(k)$ and

ISBN: 978-989-95079-6-8

Figure 1: The eye-to-hand experimental setup.

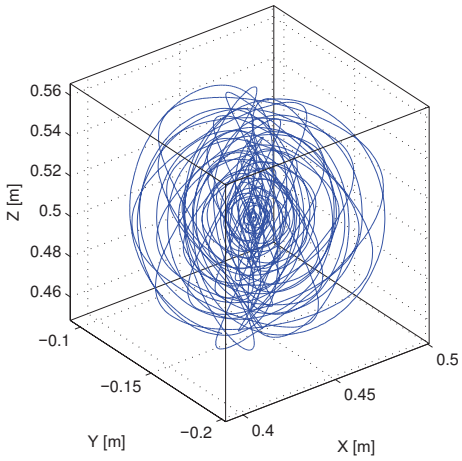


Figure 2: Robot end-effector 3D spiral trajectory.

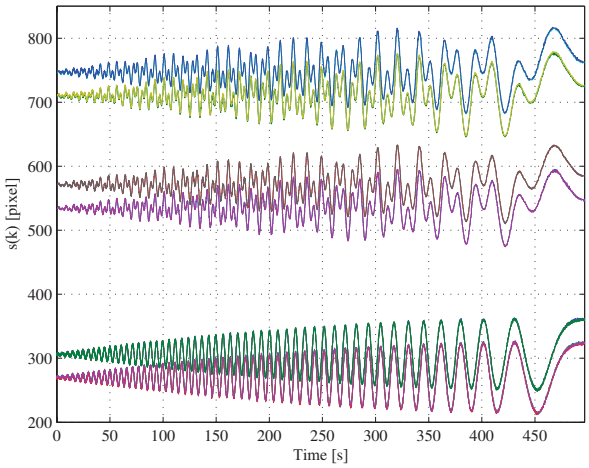


Figure 3: Input data for inverse model identification, image features $s(k)$.

the outputs $\delta s(k+1)$, following the procedure described in [2]. In this paper, the approach presented in [3] to obtain the 3D training set, was used. Note that we are interested in the identification of an inverse model as in (5).

4 Results

4.1 Experimental Setup

A Puma 560 Robotic Manipulator and a Vector CCi4 camera, in eye-to-hand configuration, were used to demonstrate the validity of the approach proposed in this paper. The experimental setup is presented in Fig. 1. The visual control algorithms were implemented in real-time using MatLab Simulink, and the xPC Target toolbox. The robot inner loop velocity control, performs at 1 KHz and the visual loop control at 12.5 Hz. A planar object is rigidly attached to the robot end-effector and is described with eight points, which centroids were used as the sixteen image features. In this paper, the robot moves in its 3D workspace (moving joints 1, 2 and 3). To maintain the image features in the camera field of view during the servoing, the planar target is set to be perpendicular to the camera optical axis by moving joints 4 and 5, as shown in Fig. 1.

4.2 Inverse Modeling Results

Following the work in [3], to obtain the identification data, the robot swept the 3D workspace in the camera field of view in a 3D spiral path, starting in the spiral center, Fig. 2, which allows to obtain the model for the end-effector position.

The variables needed for identification, $\delta q(k)$ and $\delta s(k+1)$, are obtained from the spiral when setting the desired position to the spirals center, by:

$$\delta q(k) = \frac{q^* - q(k)}{\Delta t} \tag{14}$$

$$\delta s(k+1) = s^* - s(k+1) \tag{15}$$

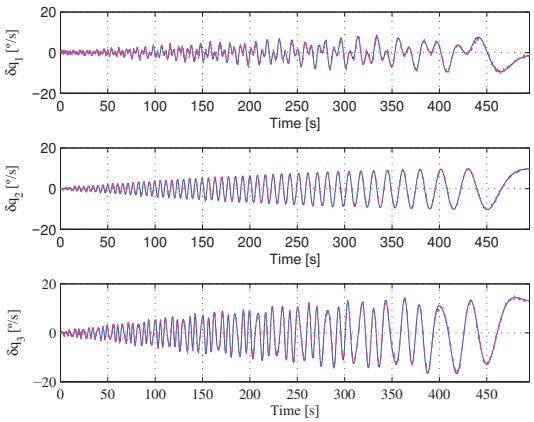
This allows to cover a wide range of values for $\delta q(k)$ and $\delta s(k+1)$. The 3D spiral also allows to control the precision of the model, by increasing or decreasing its parameters. Moving the first, the second and third joints of the PUMA Robotic Manipulator, the end-effector position follows the desired 3D spiral. The end-effector orientation is set to maintain the target parallel to the image plane.

The inverse fuzzy model was identified from the 3D spiral trajectory data, i.e. the joint velocities (see Fig. 4) and the sixteen image features obtained (see Fig. 3) using (14).

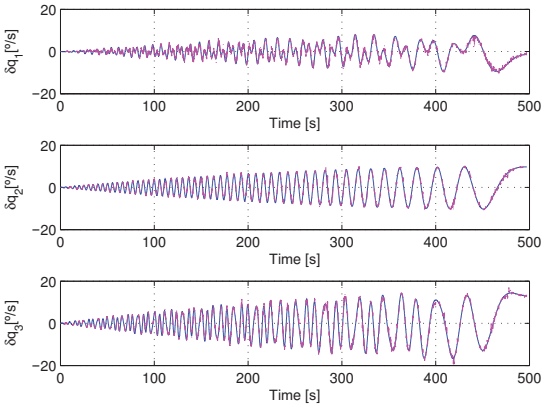
Note that to identify the inverse model, one cannot simply feed the inputs as outputs and the outputs as inputs. Since the inverse model (5) is a non-causal model, the output of the inverse model must be shifted one step, see [16].

For the identification of joints 1, 2 and 3, the obtained joint velocities $\dot{q}(k)$, are shown in Fig. 4, solid lines. Note that three fuzzy models are identified, one for each joint velocity. In the plots it is hard to see the difference between the real output data and the output of the inverse fuzzy model, because they are very similar, output of the models in dash-dotted lines. In order to measure modeling accuracy, this paper uses the Variance Accounted For (VAF) [2].

ISBN: 978-989-95079-6-8



(a) Comparison of the output data of the off-line inverse fuzzy model, $\dot{q}(k)$, joints 1 to 3. Solid – real output data, and dash-dotted – output of the off-line inverse fuzzy model.



(b) Comparison of the output data of the on-line inverse fuzzy model (xTs), $\dot{q}(k)$, joints 1 to 3. Solid – real output data, and dash-dotted – output of the on-line inverse fuzzy model.

Figure 4: Output data for inverse model identification, joints 1 to 3.

In Table 1 are presented the VAF's for the off-line modeling (oTS) and for the two on-line modeling approaches (eTS and xTS). The best result is for the off-line case as expected, although the on-line case also performs very well. The best result for on-line modeling is presented by xTS but it costs more rules, 110, and consequently more computational time to update all the cluster centers and rules.

Note that the number of clusters is pre-defined to four in oTS and that the spread $r = 0.4$ for eTS. In xTS the spread is adaptive.

5 Conclusions and Future Works

This paper introduces a novel contribution to eye-to-hand visual servoing, based on on-line fuzzy modeling to obtain an uncalibrated visual servoing system. Two methods for on-line fuzzy modeling (eTS) and (xTS) have showed excellent results when compared with the results obtained for the off-line approach. This proves the validity of the proposed approach

Table 1: Results of the inverse fuzzy model, obtained for each joint with the three different approaches.

	Joint 1		Joint 2		Joint 3	
	VAF	Rules	VAF	Rules	VAF	Rules
oTS	96.4%	4	98.5%	4	98.5%	4
eTS	95.6%	50	98.4%	50	98.5%	50
xTS	95.6%	110	98.4%	110	98.6%	110

to uncalibrated visual servoing. With the robot-camera model being able to adapt on-line to changes in the environment, the robot can now operate on different conditions without the need for re-learning off-line the model or perform any kind of calibration.

As expected the off-line modeling produces better results when compared to both on-line approaches, because it takes into account all the data points. In on-line modeling, only the actual data point and the defined cluster centers are used to update the model, introducing or replacing rules. That is why the on-line approach performs not as good as the off-line approach. However we are convicted that the differences in the models, will not compromise the use of the model to control the robot.

As future work, the proposed evolving inverse fuzzy model approach will be extended to the tri-dimensional robot workspace and also will be tested to control the PUMA Robotic Manipulator for the straight line trajectory presented in Fig. 2, in order to verify the results already achieved for the inverse model obtained off-line. The influence of using the eTS or xTS models will also be studied, i.e. the trade-off between accuracy (VAF) and computational effort (number of rules), when controlling the robot.

Acknowledgements

This work was funded by FCT trough "Programa POCI2010, Unidade 46", subsidized by FEDER.

References

[1] S. Hutchinson, G. Hager, and P. Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12(5):651–670, 1996.

[2] P.J. Sequeira Gonçalves, L.F. Mendonça, J.M. Sousa, and J.R. Caldas Pinto. Uncalibrated eye-to-hand visual servoing using inverse fuzzy models. *IEEE Transactions on Fuzzy Systems*, 16(2):341–353, 2008.

[3] P.J. Sequeira Gonçalves, A. Paris, C. Christo, J.M.C. Sousa, and J.R. Caldas Pinto. Uncalibrated visual servoing in 3d workspace. *Lecture Notes in Computer Science*, 4142:225–236, 2006.

[4] P. Angelov and D. Filev. An approach to online identification of takagi-sugeno fuzzy models. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34:484–498, 2004.

[5] P. Angelov and Xiaowei Zhou. Evolving fuzzy systems from data streams in real-time. In *Proceedings of the International Symposium on Evolving Fuzzy Systems*, pages 29–35, Ambelside, Lake District, UK, 2006.

[6] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326, 1992.

ISBN: 978-989-95079-6-8

[7] E. Malis, F. Chaumette, and S. Boudet. 2 $\frac{1}{2}$ d visual servoing. *IEEE Transactions on Robotics and Automation*, 15(2):238–250, 1999.

[8] F. Chaumette. Potential problems of stability and convergence in image-based and position-based visual servoing. *The Confluence of Vision and Control, D. Kriegman, G . Hager, A.S. Morse (eds.), LNCIS Series*, (1998):66–78, 2000.

[9] E. Cervera and P. Martinet. Combining pixel and depth information in image-based visual servoing. In *Proceedings of the Ninth International Conference on Advanced Robotics*, pages 445–450, Tokyo, Japan, 1999.

[10] J. Peipmeier, G. McMurray, and H. Lipkin. Uncalibrated dynamic visual servoing. *IEEE Trans. on Robotics and Automation*, 20(1):143–147, February 2004.

[11] I.H. Suh and T.W. Kim. Fuzzy membership function based neural networks with applications to the visual servoing of robot manipulators. *IEEE Transactions on Fuzzy Systems*, 2(3):203–220, 1994.

[12] R. Babuška. *Fuzzy Modeling for Control*. Kluwer Academic Publishers, Boston, 1998.

[13] J.M. Sousa and U. Kaymak. *Fuzzy Decision Making in Modeling and Control*. World Scientific Pub. Co., Singapore, 2002.

[14] D. E. Gustafson and W. C. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *Proceedings IEEE CDC*, pages 761–766, San Diego, USA, 1979.

[15] S. L. Chiu. Fuzzy model identification based on cluster estimation. *Journal of Intelligent Fuzzy Systems*, 2:267–278, 1994.

[16] J.M. Sousa, C. Silva, and J. Sá da Costa. Fuzzy active noise modeling and control. *International Journal of Approximate Reasoning*, 33:51–70, April 2003.

ISBN: 978-989-95079-6-8

On the reliability of the color gradient vector argument approach.

Belén Prados-Suarez[1] Pilar Sobrevilla[2] Eduard Montseny[3] Santiago Romaní[4]

1.Dept. of Software Engineering, University of Granada
Granada, Spain.
2.MAII Department, Technical University of Catalonia
Barcelona, Spain.
3.ESAII Department, Technical University of Catalonia
Barcelona, Spain.
4.Computer Engineering and Mathematics Department, Rovira I Virgili University
Tarragona, Spain.
Email: belenps@ugr.es, {pilar.sobrevilla,eduard.montseny}@upc.edu, santiago.romani@urv.net

Abstract— *Today many algorithms for monochrome images have been developed based on the information provided by the gradient vector. However, these algorithms can not be applied to color images without an accurate approximation of the color gradient vector argument. Moreover, there is a great number of applications that require to know with the maximum accuracy as possible the direction of the color variation in color images. This paper presents a new breakthrough in our work on the approximation of the argument of the color gradient vector. Taking into consideration that Hue and Saturation provide very useful basis for judging color uniformity, and that in the Smith's HSI color space there is a close relationship between chromaticity and how humans perceive colors, in our proposal the argument of the Chromaticity gradient vector is approached based on the Hue and Saturation information. Moreover, we present a fuzzy measure of the quality or reliability of the proposed approach that, considering the inherent vagueness of the chromaticity values of the image pixels, provides us with a credibility-accuracy degree of the obtained values.*

Keywords— Color gradient, Color image processing, Gradient argument, Gradient reliability, Gradient vector.

1 Introduction

Edge detection is one of the most important tasks in digital image processing and machine vision systems. Generally speaking, an edge implies the existence of abrupt changes or discontinuities in some visual property as light intensity, texture or color.

The advantage of color edge detection schemes over grayscale approaches is easily demonstrated by considering the fact that those edges that exist at the boundary between regions of different colors cannot be detected in grayscale images if there is no change in intensity. Among the existing approaches to detect color borders we can find active contour based methods [1], which use different approaches to perform the deformation process to dynamically adapt the active contour; multi-scale techniques [2], based on comparing the evolution of the borders through different scales; or morphological algorithms [3], that locate borders by computing the difference between a dilation and an erosion. However these methods just localize the maximum color variations or approach their magnitude, but not their direction.

A class of methods which not only localize the maximum color variations (edge localization) but also consider the direction of this variation (edge orientation) is the class of gradient based techniques.

The first methods proposed to compute the color gradient vector were mostly direct applications of the early intensity gradient methods. According to Lucchesse et al [4] these methods can be divided in those techniques that embed the variations of all color channels in a single measure [5, 6], and those that compute the gradient in each channel and then combine them according to certain criteria [7, 8]. A comprehensive analysis of color edge detectors can be found in [9].

Despite the importance and usefulness of the information provided by the argument of the gradient vector for many applications (texture characterization and analysis, 3D reconstruction, vehicle tracking and guidance,...), few techniques explicitly present and approach edge orientation. Among them we can find: proposals that approach the argument as the arc-tangent of the quotient between the vertical and horizontal partial derivatives [10]; Tensor based proposals [11] wherein the direction of the global color gradient is obtained from the tensor gradient components of multi-images regarded as vector fields; and Matrix based proposals [12] that approach the color gradient modulus as the highest eigenvalue in a given matrix, and its direction is given by the eigenvector associated to that eigenvalue.

The majority of the methods presented are focused on the RGB space. However, its lack of distinction between chromatic and achromatic information makes it unsuitable for obtaining edges in an image. A comparison of color edge detectors across several color spaces is presented in [13].

In this paper we present our ongoing work on the argument of the color gradient vector, focused on the Smith's HSI color space [14], which we began in [15]. In our proposal the argument of the *Chromaticity* gradient vector is approached based on the Hue and Saturation information, and we provide a fuzzy reliability degree of the obtained values for each edge pixel. This is quite useful, not only to evaluate the accuracy of our approach, but also to be used in applications that require to know the precision degree with which they are working. This is the case of astronomical images processing [16], the selection of optimal imaging planes in MRI [17] , or even the design of dichroic mirrors for LCDs [18]. In these applications

ISBN: 978-989-95079-6-8

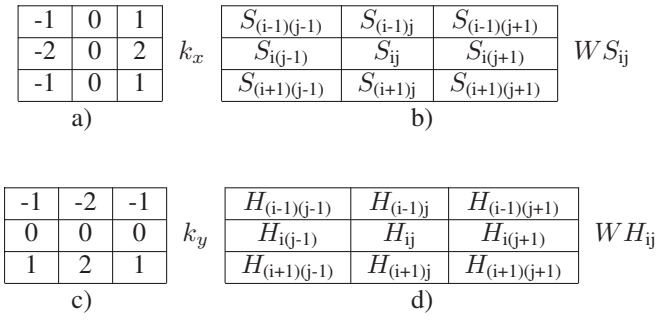


Figure 1: a) and c) Sobel Operators for the x and y directions, respectively. b) and d) 3x3 Saturation and Hue components windows of the source image, respectively.

an accurate contour extraction is necessary, and this is not possible without a good approach to the argument of the gradient vector, which also deals with the problems of imprecise contours.

Considering it, the paper is organized as follows. First, in section 2, we generalize the Sobel operators to compute the Hue and Saturation partial derivatives along the x and y directions. Then we present, in section 3, the way we combine the chromatic information to approach the partial derivatives; it will be used as directional components of the chromaticity gradient vector for approaching its argument. In section 4, we present the process and experiments carried out for obtaining a measure to evaluate the reliability of the color gradient vector argument approached. Finally we present some conclusions.

2 Proposed approximation of the Hue and Saturation partial derivatives

From the existing transformations to convert RGB components into perceptual ones, we have chosen Smith's model [14] based on its high independence between the three components, and its easier and faster calculation compared to other nonlinear models. This way we can consider color as a 1D+2D signal by decomposing it into its achromatic (I) and chromatic (H and S) components, what will provide us with a most useful basis for judging color uniformity.

In [15], having into consideration that Sobel masks (see Figure 1-a and 1-c) have slightly superior noise-suppression characteristics [13] than other operators, as Prewitt one, and that Sobel operator allows convolving separately the Hue and Saturation channels of the image, we presented an approximation to the partial derivatives of the chromatic color components extending this operator to approximate the variation of the Hue and Saturation components along the x and y directions.

2.1 Saturation First Order Derivatives

Given a pixel p_{ij} of the source image I, we will note by S_{ij} the magnitude of its Saturation, and WS_{ij} the window containing the Saturation values of the pixels falling into the 3x3 raster window centered at the pixel (see Figure 1-b).

Due to the fact that the areas of highest saturation gradient are where the saturation of the image changes rapidly over a few pixels, and are thus likely to represent edges, we can compute the Saturation partial derivatives, $\partial Sij/\partial x$ and $\partial Sij/\partial y$, by convolving kernels k_x and k_y with the WS_{ij} window , what provides us with the values given at equations (1) and (2).

2.2 Hue First Order Derivatives

Similarly to the Saturation case, we will note by H_{ij} the Hue magnitude of pixel p_{ij}, and WH_{ij} the window containing the Hue values of the pixels falling into the 3x3 raster window centered at that pixel (see Figure 1-d). However, while Euclidean distance is appropriate to compare the saturation values of two given colors, it is not the case of the Hue component. This is due to the ownership of Hue circularity, that is observed when this component is represented in a two-dimensional space. To overcome this problem, and consider the direction (the sign) of the difference between two Hue values we defined a Hue directed distance as follows:

Definition 2.1 *Let H_1 and H_2 be the hue values of pixels p_{i_1,j_1} and p_{i_2,j_2}, respectively. Then the* directed Hue distance *is defined as:*

$$\widehat{d}(H_1,H_2)=\begin{cases} H_2-H_1 & if\ \ |H_2-H_1|\leq 128 \\ H_2-H_1-255 & if\ \ H_2-H_1>128 \\ H_2-H_1+255 & if\ \ H_2-H_1<-128 \end{cases} \tag{3}$$

Then, convolution of kernels k_x and k_y are applied on the directed distances (3) obtained from the Hue values of WH_{ij}, instead of applying directly the Sobel operator over it.

As a result the Hue partial derivatives in both directions, $\partial Hij/\partial x$ and $\partial Hij/\partial y$, are given by equations (4) and (5).

3 Chromatic Gradient Vector Argument Approach

Once the Hue and Saturation partial derivatives have been approximated, next step consists in combining these values for obtaining the directional components of the *Chromatic Gradient Vector -CGV-*, or *Chromatic Partial Derivatives -CPD-*, which will be proportional to the directional components of the chromatic gradient vector, according to a given proportionality factor. In our case this problem has been figured out by mixing and merging previously obtained partial derivatives approximations according to next process.

1. First we consider two reference systems, one for the *x* direction, RS_x, and other for the *y* direction, RS_y, whose axes are the partial derivatives of *H* and *S* in the corresponding directions, ie: $RS_x=\{\partial H/\partial x, \partial S/\partial x\}$ and $RS_y=\{\partial H/\partial y, \partial S/\partial y\}$. On these systems we represent the values of the approaches obtained at the previous step, what provides us with two chromaticity vectors, one in the *x* direction, $\overrightarrow{Chr_{ij}x}=(\partial H_{ij}/\partial x, \partial S_{ij}/\partial x)$, and other in the *y* direction, $\overrightarrow{Chr_{ij}y}=(\partial H_{ij}/\partial y, \partial S_{ij}/\partial y)$ (line-dot vectors of Figure 2-a and 2-b). We also obtain de modules of these vectors, $\|\overrightarrow{Chr_{ij}x}\|$ and $\|\overrightarrow{Chr_{ij}y}\|$, which we call the *Directional Chromaticity Variations*.

2. Because the vectors obtained are represented into two different reference systems, to be able to mix and merge the information they contain is necessary to get a reference system wherein both vectors can be represented.

ISBN: 978-989-95079-6-8

$$\partial S_{ij}/\partial x = \frac{k_x * WS_{ij}}{8} = \frac{(S_{(i-1)(j+1)} - S_{(i-1)(j-1)}) + 2(S_{(i)(j+1)} - S_{(i)(j-1)}) + (S_{(i+1)(j+1)} - S_{(i+1)(j-1)})}{8} \quad (1)$$

$$\partial S_{ij}/\partial y = \frac{k_y * WS_{ij}}{8} = \frac{(S_{(i+1)(j-1)} - S_{(i-1)(j-1)}) + 2(S_{(i+1)(j)} - S_{(i-1)(j)}) + (S_{(i+1)(j+1)} - S_{(i-1)(j+1)})}{8} \quad (2)$$

$$\partial H_{ij}/\partial x = \frac{\widehat{d}(H_{(i-1)(j+1)}, H_{(i-1)(j-1)}) + 2\widehat{d}(H_{(i)(j+1)}, H_{(i)(j-1)}) + \widehat{d}(H_{(i+1)(j+1)}H_{(i+1)(j-1)})}{8} \quad (4)$$

$$\partial H_{ij}/\partial y = \frac{\widehat{d}(H_{(i+1)(j-1)}, H_{(i-1)(j-1)}) + 2\widehat{d}(H_{(i+1)(j)}, H_{(i-1)(j)}) + \widehat{d}(H_{(i+1)(j+1)}H_{(i-1)(j+1)})}{8} \quad (5)$$

To do it, due to axes $\partial H/\partial x = \partial S/\partial x$ and $\partial H/\partial y = \partial S/\partial y$ point to the higher potentials of *H* and *S*, and that both components grow equally within these axes, our proposal is to project the chromaticity vectors into them. So, our approach to the above mentioned *CPD* vectors, $\overrightarrow{\partial C_{ij}/\partial x}$ and $\overrightarrow{\partial C_{ij}/\partial y}$, is given by equations (6) and (7), where $\overrightarrow{u} = (\cos(45^o), \sin(45^o))$. These new vectors, that appear as dotted arrows in Figures 2-a and 2-b, will be used at next section to approach of the *CGVA*.

$$\overrightarrow{\partial C_{ij}/\partial x} = proj_{\overrightarrow{u}}\overrightarrow{Chr_{ij}x} \quad (6)$$

$$\overrightarrow{\partial C_{ij}/\partial y} = proj_{\overrightarrow{u}}\overrightarrow{Chr_{ij}y} \quad (7)$$

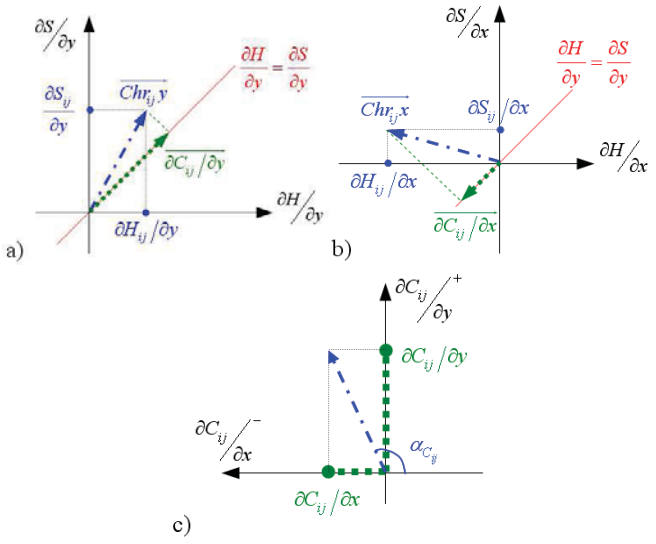


Figure 2: a, b: Calculation of the *CPD* vectors in each direction. c: Approach to the *CGVA*.

3. Applying the usual calculation of a projection, as we are interested in considering the argument of the *CPD* vectors, but considering their direction, the components of the *CGVA* approach are given using the scalar product, as depicted by equations (8) and (9).

$$\partial C_{ij}/\partial x = (\partial H_{ij}/\partial x, \partial S_{ij}/\partial x) \cdot (\sqrt{2}/2, \sqrt{2}/2) \quad (8)$$

$$\partial C_{ij}/\partial y = (\partial H_{ij}/\partial y, \partial S_{ij}/\partial y) \cdot (\sqrt{2}/2, \sqrt{2}/2) \quad (9)$$

Then, translating these values to the coordinate system of axes $\partial C_{ij}/\partial x$ and $\partial C_{ij}/\partial y$ we get the approach to the *Chromatic Gradient Vector*, and the argument of this vector, $\alpha_{C_{ij}} = \arctan(\frac{\partial C_{ij}/\partial y}{\partial C_{ij}/\partial x})$, provides us with the *CGVA* approach, as represented by Figure 2-c. This argument provides us with a direction that is perpendicular to the contour and so it is 0 for a vertical edge and increases for edges moving anti-clockwise of it.

4 Reliability degree of the color gradient argument approach

To assess the performance (goodness) of the proposed approach, we have defined a measure of reliability that must consider the sources of uncertainty of the factors having influence on the results obtained when the approach is applied on an image. So, for obtaining the reliability measure we have developed a set of experiments that allow us to study the relationship among: the error given by the approach with regard to the actual value of the argument, the *Chromatic Partial Derivatives*, and the *Directional Chromaticity Variations*. To carry out our experiments we have designed a set of images as explained in the next section.

4.1 *Obtaining the set of synthetic images*

To perform our experiments we built a set of images, consisting of a square with a circle in the center. The process for obtaining the images has been carried out as follows:

1. First, after dividing the RGB cube into 512 boxels, we select a point of each one of the boxes, so obtaining a set of 512 colors distributed almost homogeneously.

2. Then we generated a set of images using all the possible combinations of colors for background and foreground. For each image if a pixel is completely included in the background or in the circle, it is assigned the corresponding color. If the pixel belongs to background and foreground, after evaluating the area of the pixel belonging to each region, we assign to the pixel a color that is the weighted average of the two colors.

3. Finally, after smoothing the images using a 3x3 window to create a gradient in the boundary of the circles, the transformation from RGB to HSI is carried out.

4.2 *Experimental design*

For all the pixels in the borders of the circumferences, we have computed the argument of the *Chromaticity Gradient Vector* using the proposed approach. Since we know the position of the point and the center of the circle, we know the real value of the argument for each border pixel ($\arg_{ij}$). Having both

ISBN: 978-989-95079-6-8

values we can accurately evaluate the error at pixel (i, j) as $E^{ij} = |\arg_{ij} - \arg \alpha_{C_{ij}}|$.

In our experiments for obtaining the reliability degree we have studied the relationship among the aforementioned error, the *Directional Chromaticity Variations* ($\|\overrightarrow{Chr_{ij}x}\|$, and $\|\overrightarrow{Chr_{ij}y}\|$), and the smaller angles of the *Chromaticity Vectors* ($\overrightarrow{Chr_{ij}x}$, and $\overrightarrow{Chr_{ij}y}$) with the lines $\partial S/\partial x = -\partial H/\partial x$ and $\partial S/\partial y = -\partial H/\partial y$. These angles are given by:

$$E(\alpha_x^{ij}) = \begin{cases} |\arg \alpha_x^{ij} - 135| & \text{if } 45 < \arg \alpha_x^{ij} < 225 \\ |\arg \alpha_x^{ij} - 315| & else \end{cases} \quad (10)$$

$$E(\alpha_y^{ij}) = \begin{cases} |\arg \alpha_y^{ij} - 135| & \text{if } 45 < \arg \alpha_y^{ij} < 225 \\ |\arg \alpha_y^{ij} - 315| & else \end{cases} \quad (11)$$

Where α_x^{ij} and α_y^{ij} are the angles of the vectors $\overrightarrow{Chr_{ij}x}$ and $\overrightarrow{Chr_{ij}y}$ with the lines $\partial S/\partial x = \partial H/\partial x$ and $\partial S/\partial y = \partial H/\partial y$, respectively.

For each pixel (i, j) in the border of the circles we have obtained the values of E^{ij}, $\|\overrightarrow{Chr_{ij}x}\|$, $\|\overrightarrow{Chr_{ij}y}\|$, $E(\alpha_x^{ij})$, and $E(\alpha_y^{ij})$. Then we have analyzed and discussed the relationship between the average error and the values obtained for the four parameters considered by pairs.

4.3 Experimental analysis

For analyzing the obtained results we have represented in $3-D$ graphics the error values E^{ij} for each point (i, j) of the planes: $(E(\alpha_x^{ij}), E(\alpha_y^{ij}))$, $(\|\overrightarrow{Chr_{ij}x}\|, \|\overrightarrow{Chr_{ij}y}\|)$, and the planes $(\|\overrightarrow{Chr_{ij}x}\|, E(\alpha_x^{ij}))$, $(\|\overrightarrow{Chr_{ij}x}\|, E(\alpha_y^{ij}))$, $(\|\overrightarrow{Chr_{ij}y}\|, E(\alpha_y^{ij}))$, and $(\|\overrightarrow{Chr_{ij}y}\|, E(\alpha_x^{ij}))$. In these representations the gray level is proportional to the value of the average error for each point. So, the higher the gray level the lowest the error is.

As can be observed at the left image of Figure 3, when we consider the differences of the arguments, the values of the errors are small or large, and errors distribution indicates a great dependency among the error and the values of $E(\alpha_x^{ij})$, $E(\alpha_y^{ij})$.

When we consider the values of the *Directional Chromaticity Variations* (right image of Figure 3) the values of the errors are small and medium. However, the relationship among the error values and those of the parameters is very small, what is due to for almost each pair $(\|\overrightarrow{Chr_{ij}x}\|, \|\overrightarrow{Chr_{ij}y}\|)$ the errors are small.

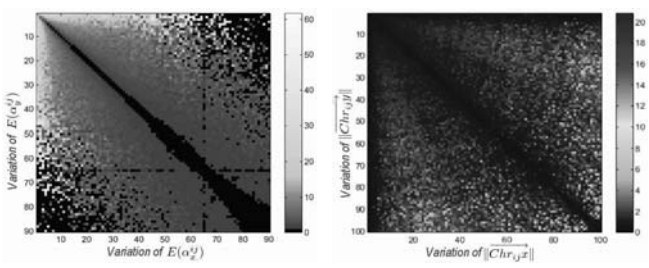


Figure 3: Representation of error magnitude for $E(\alpha_x^{ij})$ vs $E(\alpha_y^{ij})$ in the left image, and for $\|\overrightarrow{Chr_{ij}x}\|$ vs $\|\overrightarrow{Chr_{ij}y}\|$ in the right image.

In the case of the plane $(\|\overrightarrow{Chr_{ij}y}\|, E(\alpha_x^{ij}))$, and similarly for $(\|\overrightarrow{Chr_{ij}x}\|, E(\alpha_y^{ij}))$, the errors are mainly small. The medium and large errors that occur are highly concentrated, leaving almost the entire plane to the small errors (see left image of Figure 4), what implies a correlation among the error and the values of the parameters.

Finally, having a look at right image of Figure 4 it can be observed that in the case of the plane $(\|\overrightarrow{Chr_{ij}y}\|, E(\alpha_y^{ij}))$ appear small and big error values. In this case, while small errors are highly concentrated large errors are widely distributed, which indicates a certain relation between the error and the parameter values. A similar behavior is observed in the case of the plane $(\|\overrightarrow{Chr_{ij}x}\|, E(\alpha_x^{ij}))$

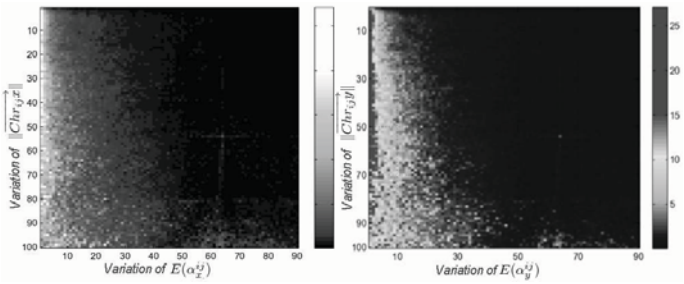


Figure 4: Left image represents the error magnitude for $\|\overrightarrow{Chr_{ij}^{x}}\|$ vs $E(\alpha_x^{ij})$.Right image represents the error magnitude for $\|\overrightarrow{Chr_{ij}^{y}}\|$ vs $E(\alpha_y^{ij})$.

As a consequence of previous analysis it can be deduced that the most important information for obtaining the reliability of the proposed approach is given by considering the values of $E(\alpha_x^{ij})$ vs $E(\alpha_y^{ij})$, and the less relevant information is provided by the values of $\|\overrightarrow{Chr_{ij}x}\|$ vs $\|\overrightarrow{Chr_{ij}y}\|$. On the other hand some information is provided when we consider the pairs $(E(\alpha_x^{ij}), \|\overrightarrow{Chr_{ij}y}\|)$ and $(E(\alpha_x^{ij}), \|\overrightarrow{Chr_{ij}x}\|)$.

4.4 Experimental results and Reliability degree

Based on the results obtained at previous step, we have considered the *Reliability Labels*: $Great$ (G), $Medium$ (M), $Small$ (S), and *Very Small* (VS). Then a reliability degree has been associated to each point of the considered plains, according to the error at each point, by means of the membership functions depicted at Figure 5.

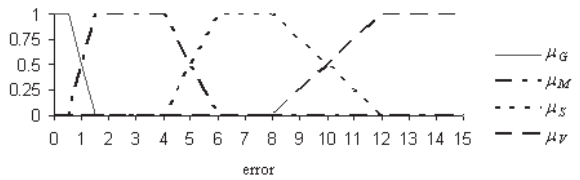


Figure 5: Reliability membership functions.

These membership functions have been obtained through a heuristic analysis of the results. The border values between each of the two central functions have been selected so that the error value of the 75%, the 85% and the 90% of the pixels are lower than such value.

The graphs obtained for each one of the cases considered at the previous step leads to the appearance of some lines. For each case, these lines act as boundaries between the different reliability degrees of the points in the corresponding plain.

ISBN: 978-989-95079-6-8

So, these lines will allow us to obtain the *Reliability Functions* of the proposed approach. Due to lack of space here only present the results obtained for the case of the $E(\alpha_x^{ij})$ vs $E(\alpha_y^{ij})$. For this case Figure 6 depicts the result of using the reliability degrees. Having a look at this image is it clear that appear some lines delimiting the different reliability regions, from $Great$ (dark grey level) to *Very Small* (very clear grey level). The lines obtained in this case are given by equations (12) to (22). Afterwards, these lines are used for obtaining the corresponding *Reliability Degrees* that, in this case are depicted by equations (23) to (26).

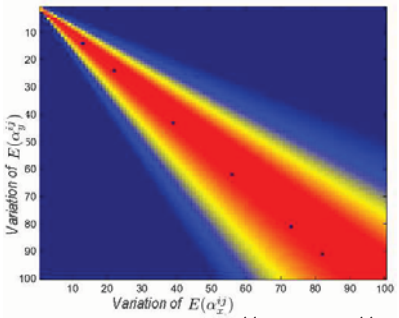


Figure 6: Graph obtained for $E(\alpha_x^{ij})$ vs $E(\alpha_y^{ij})$ using the reliability membership functions.

$$r_0(i,j) : E_x^{ij} = E_y^{ij} \tag{12}$$

$$r_1(i,j) : 25 \cdot E_x^{ij} - 27,5 \cdot E_y^{ij} + 6 = 0 \tag{13}$$

$$r_2(i,j) : 16 \cdot E_x^{ij} - 29 \cdot E_y^{ij} + 13 = 0 \tag{14}$$

$$r_3(i,j) : 10,6 \cdot E_x^{ij} - 29,1 \cdot E_y^{ij} + 16,65 = 0 \tag{15}$$

$$r_4(i,j) : 4,75 \cdot E_x^{ij} - 29,25 \cdot E_y^{ij} + 18,375 = 0 \tag{16}$$

$$r_5(i,j) : 2,5 \cdot E_x^{ij} - 29,5 \cdot E_y^{ij} + 13,5 = 0 \tag{17}$$

$$r_6(i,j) : 27,5 \cdot E_x^{ij} - 25 \cdot E_y^{ij} - 6 = 0 \tag{18}$$

$$r_7(i,j) : 29 \cdot E_x^{ij} - 16 \cdot E_y^{ij} - 13 = 0 \tag{19}$$

$$r_8(i,j) : 29,1 \cdot E_x^{ij} - 10,6 \cdot E_y^{ij} - 16,65 = 0 \tag{20}$$

$$r_9(i,j) : 29,25 \cdot E_x^{ij} - 4,75 \cdot E_y^{ij} - 18,375 = 0 \tag{21}$$

$$r_{10}(i,j) : 29,5 \cdot E_x^{ij} - 2,5 \cdot E_y^{ij} - 13,5 = 0 \tag{22}$$

Finally, for obtaining the *Reliability Degree* of the *Chromatic Gradient Vector Argument Approach* value for each point, the previous reliabilities for each case has to be aggregated. To do it, considering that all the individual values have influence in the final reliability degree, we can consider a *T-Norm* to perform the aggregation. Among the great variety of available *T-Norm* we propose to use the *Minimum* due to: its simplicity, fast computation, and that it is the greatest *T-Norm*.

5 Conclusions

In this paper we have presented an approach of the chromatic component for the argument of the gradient vector. After computing the hue and saturation partial derivatives on each direction, this information has been merged and combined through projections for obtaining the directional components of the chromaticity gradient vector used to approach the argument.

We have also presented a study of the error variability produced by the proposed approach with regard to the four main parameters used within the proposed approach. This will allow us to improve the approximation, as well as obtaining an approach for the module of the gradient vector.

A very important point to highlight is that, as a result of the analysis of the error variability, we obtained a value that provides the credibility of the value provided by the approximation. This is very important because it will allow to use the value of the gradient vector argument for each point knowing its reliability degree.

Acknowledgment

This paper has been partially supported by the projects P07-TIC-03175, P06-TIC-01570 and TIN2007-68063.

References

[1] A. Maalouf, A. Maalouf, P. Carre, B. Augereau, and C. Fernandez-Maloigne. Foveal wavelet-based color active contour. In P. Carre, editor, *Image Processing, 2007. ICIP 2007. IEEE Int. Conf. on*, volume 1, pages I – 245–I – 248, 2007.

[2] I. Vanhamel, I. Vanhamel, I. Pratikakis, and H. Sahli. Multiscale gradient watersheds of color images. *Image Processing, IEEE Transactions on*, 12(6):617–626, 2003.

[3] A.N. Evans and A.N. Evans. Morphological gradient operators for colour images. In *Image Processing, 2004. ICIP '04. 2004 Int. Conf. on*, volume 5, pages 3089–3092 Vol. 5, 2004.

[4] L. Lucchese and S. K. Mitra. Color image segmentation: A state-of-art survey. In *Proc. Indian Nat. Sci. Acad. (INSA-A)*, volume 67-A, pages 207–221, 2001.

[5] M. Chapron. A color edge detector based on statistical rupture tests. In *Image Processing, 2000. Proc. 2000 Int. Conf. on*, volume 2, pages 820–823 vol.2, 2000.

[6] M. A. Ruzon and C. Tomasi. Color edge detection with the compass operator. *Proc. IEEE Conf. Computer Vision and Pattern Recognition.*, 2:160–166, 1999.

[7] M. Heddley. Segmentation of color images using spatial and color space information. *J. Electron. Imag.*, 1(4):374–380, 1992.

[8] Jianping Fan, Jianping Fan, D.K.Y. Yau, A.K. Elmagarmid, and W.G. Aref. Automatic image segmentation by integrating color-edge extraction and seeded region growing. *Image Processing, IEEE Transactions on*, 10(10):1454–1466, 2001.

[9] A. Koschan and M. Abidi. Detection and classification of edges in color images. *Signal Processing Magazine, IEEE*, 22(1):64–73, 2005.

[10] T. Carron and P. Lambert. Color edge detector using jointly hue, saturation and intensity. In *Image Processing, 1994. Proc. ICIP-94., IEEE Int. Conf.*, volume 3, pages 977–981, 1994.

[11] Silvano Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, 33(1):116–125, 1986.

[12] Anwander-A, Neyran-B, and Baskurt-A. Multiscale color gradient for image segmentation. *Traitement du Signal*, 18(2):129–42, 2001.

[13] S. Wesolkowski, S. Wesolkowski, M.E. Jernigan, and R.D. Dony. Comparison of color image edge detectors in multiple color spaces. In M.E. Jernigan, editor, *Image Processing, 2000. Proc. 2000 Int. Conf. on*, volume 2, pages 796–799 vol.2, 2000.

ISBN: 978-989-95079-6-8

$$\mu_G(E_x^{ij},E_y^{ij})=\begin{cases}0 & \text{if } r_8(i,j)<0 \text{ or } r_2(i,j)>0\\ 1 & \text{if } r_6(i,j)>0 \text{ and } r_1(i,j)<0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_7(i,j))}{d((E_x^{ij},E_y^{ij}),r_7^{ij})+d((E_x^{ij},E_y^{ij}),r_6(i,j))} & \text{if } r_6(i,j)<0 \text{ and } r_7(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_2(i,j))}{d((E_x^{ij},E_y^{ij}),r_2^{ij})+d((E_x^{ij},E_y^{ij}),r_1(i,j))} & \text{else}\end{cases} \tag{23}$$

$$\mu_M(E_x^{ij},E_y^{ij})=\begin{cases}0 & \text{if } r_8(i,j)<0 \text{ or } r_3(i,j)>0\\ 1 & \text{if } (r_6(i,j)<0 \text{ and } r_7(i,j)>0)\\ & \textit{or } (r_1(i,j)>0 \text{ and } r_2(i,j)<0)\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_0(i,j))}{d((E_x^{ij},E_y^{ij}),r_0^{ij})+d((E_x^{ij},E_y^{ij}),r_6(i,j))} & \text{if } r_0(i,j)>0 \text{ and } r_6(i,j)<0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_0(i,j))}{d((E_x^{ij},E_y^{ij}),r_0^{ij})+d((E_x^{ij},E_y^{ij}),r_1(i,j))} & \text{if } r_0(i,j)<0 \text{ and } r_1(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_8(i,j))}{d((E_x^{ij},E_y^{ij}),r_8^{ij})+d((E_x^{ij},E_y^{ij}),r_7(i,j))} & \text{if } r_7(i,j)<0 \text{ and } r_8(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_3(i,j))}{d((E_x^{ij},E_y^{ij}),r_3^{ij})+d((E_x^{ij},E_y^{ij}),r_2(i,j))} & \text{else}\end{cases} \tag{24}$$

$$\mu_S(E_x^{ij},E_y^{ij})=\begin{cases}0 & \text{if } (r_{10}(i,j)<0 \text{ or } r_5(i,j)>0)\\ & \textit{or } (r_7(i,j)>0 \text{ and } r_2(i,j)<0)\\ 1 & \text{if } (r_8(i,j)<0 \text{ and } r_9(i,j)>0)\\ & \textit{or } (r_3(i,j)>0 \text{ and } r_4(i,j)<0)\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_7(i,j))}{d((E_x^{ij},E_y^{ij}),r_7^{ij})+d((E_x^{ij},E_y^{ij}),r_8(i,j))} & \text{if } r_7(i,j)<0 \text{ and } r_8(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_2(i,j))}{d((E_x^{ij},E_y^{ij}),r_2^{ij})+d((E_x^{ij},E_y^{ij}),r_3(i,j))} & \text{if } r_2(i,j)>0 \text{ and } r_3(i,j)<0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_{10}(i,j))}{d((E_x^{ij},E_y^{ij}),r_{10}^{ij})+d((E_x^{ij},E_y^{ij}),r_9(i,j))} & \text{if } r_9(i,j)<0 \text{ and } r_{10}(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_5(i,j))}{d((E_x^{ij},E_y^{ij}),r_5^{ij})+d((E_x^{ij},E_y^{ij}),r_4(i,j))} & \text{else}\end{cases} \tag{25}$$

$$\mu_{VS}(E_x^{ij},E_y^{ij})=\begin{cases}0 & \text{if } r_9(i,j)>0 \text{ and } r_4(i,j)><0\\ 1 & \text{if } r_{10}(i,j)>0 \text{ and } r_5(i,j)<0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_9(i,j))}{d((E_x^{ij},E_y^{ij}),r_9^{ij})+d((E_x^{ij},E_y^{ij}),r_{810}(i,j))} & \text{if } r_9(i,j)<0 \text{ and } r_{10}(i,j)>0\\ \dfrac{d((E_x^{ij},E_y^{ij}),r_4(i,j))}{d((E_x^{ij},E_y^{ij}),r_4^{ij})+d((E_x^{ij},E_y^{ij}),r_5(i,j))} & \text{else}\end{cases} \tag{26}$$

[14] A.R. Smith. Color gamut transform pairs. *Computer Graphics*, 12:12–19, 1978.

[15] P. Sobrevilla B. Prados-Suarez, E. Montseny and S. Romanís. An approach to color gradient vector argument considering data uncertainty. *in Procc Joint Eusflat-LFA Conf.*, 2:705–708, 2000.

[16] Changbom-Park and Yun-Young-Choi. Morphology segregation of galaxies in color-color gradient space. *Astrophysical Journal, Letters*, 635(1):L29–32, 2005.

[17] Levine RA Okada RD Brady TJ. Dinsmore RE, Wismer GL. Magnetic resonance imaging of the heart: positioning and gradient angle selection for optimal imaging planes. *American journal of roentgenology, radium therapy and nuclear medicine.*, 143(6):1135–42, 1984.

[18] GU P. GUAN A. LU W., ZHENG Z. Gradient angle design of dichroic mirrors used in lcd projection display. In *Proc Int Disp Workshops*, volume 11, pages 1697–1699, 2004.

ISBN: 978-989-95079-6-8

Evolutionary Robot Vision and Fuzzy Evaluation for Natural Communication of Partner Robots

Akihiro Yorita Naoyuki Kubota

Graduate School of System Design, Tokyo Metropolitan University,
Hino, Tokyo 191-0065, Japan
e-mail:yorita-akihiro@sd.tmu.ac.jp, kubota@tmu.ac.jp

Abstract— *This paper proposes a method of evolutionary robot vision based on a steady-state genetic algorithm and fuzzy evaluation. In order to improve the communication capability of human-friendly partner robots, the perception of human face should be performed as correctly as possible. First, we discuss the concept of evolutionary robot vision in dynamic environments. Next, we propose growing neural gas for preprocessing as a bottom-up processing, and steady-state genetic algorithm for template matching in human face recognition as a top-down processing. In order to improve the performance of the human face recognition, we use fuzzy evaluation for evaluating the degree of human face. Finally, we show several experimental results and discuss the effectiveness of the proposed method.*

Keywords— Face Recognition, Evolutionary Computation, Robot Vision, Fuzzy Theory, Partner Robots

1 Introduction

Natural communication and interactions among people and robots have been widely researched [1,2]. Furthermore, various types of human-friendly robots such as pet robots, amusement robots, and partner robots have been developed to communicate with people [3,4]. However, there are still many problems to realize natural communication between people and robots. For example, it is difficult for a robot to identify the speaking person and the place interacting with the person. As a result, the content of utterances from the robot might not be suitable for the person and place.

Relevance theory is helpful to discuss the natural communication between a human and robot [21]. Relevance theory is based on a definition of relevance and two principles of the relevance. One is a cognitive principle that human cognition is geared to the maximization of relevance. The other is a communicative principle that utterances create expectations of optimal relevance. The central claim of relevance theory is that the expectations of relevance raised by an utterance are precise enough, and predictable enough to guide the hearer towards the speaker's meaning. Furthermore, in relevance theory, human thought is shared between two people rather than transmitted. Each person has his or her own cognitive environment. An important role of utterances, facial direction, pointing behaviors, and gestures is to make the hearer pay attention to a specific target object or person. As a result, the cognitive environment of the hearer can be enlarged by the utterances or gestures. The cognitive environment shared between two people is called a mutual cognitive environment. Based on the above discussion, a robot also should have such a cognitive environment, and the robot should keep updating the cognitive environment according to the current perception through the interaction with a human in order to realize the natural communication.

Furthermore, the utterance capability of a robot can be applied for preventing dementia of elderly people. Robotic conversation can activate the brain of such elderly people and can improve their concentration and memory abilities. Nursing care for the elderly people can be expected to keep their health by having conversations with robots. However, it is very difficult to continue the meaningful and attractive conversations with robots. Therefore, such a robot requires adaptive perceptual systems to communicate with a human flexibly, and adaptive action systems to learn human behaviors. To realize the learning through interaction with people, we must consider a total architecture of the cognitive development. The cognitive development for robots has been discussed in the fields such as cognitive robotics and embodied cognitive science [8-10]. In the previous research of cognitive robotics, many researchers have proposed the learning methods for the achievement of joint attention, imitative learning, linguistic acquisition from the viewpoints of babies and infants [6,10]. On the other hand, we focus on the refinement of associative memory by using symbolic information used for utterances and patterns based on visual information through interaction with people as cognitive development of robots. We proposed the concept of structured learning and discussed the importance of total architecture of the learning mechanism [30]. However, we did not discuss the performance of the human detection so much. In the previous works, we proposed a simple method of people tracking based on the combination of skin color and hair color, and we have a problem of misdetection of people by objects with similar color combination in the background image. In this paper, we propose a method for detecting a human face based on evolutionary computation and fuzzy evaluation in order to improve the performance of people tracking. The both of evolutionary computation and fuzzy theory are useful and practical in the search under the environment including noise. We proposed the concept of evolutionary robot vision based on the analogy between visual perception and evolutionary search [32]. We apply the concept of evolutionary robot vision and fuzzy evaluation for people tracking.

The paper is organized as follow. In the section 2, we explain the concept of evolutionary robot vision. Next, in the section 3 we propose a growing neural gas for color extraction and a steady-state genetic algorithm with fuzzy inference for face recognition. Section 4 shows several experimental results and discuss the effectiveness of the proposed method.

2 Evolutionary Robot Vision

2.1 Partner robots

We developed two types of partner robots; a mobile PC called MOBiMac and a human-like robot called Hubot in order to realize the social communication with a human [25,26] (Fig.1). Each robot has two CPUs and many sensors such as CCD camera, microphone, and ultrasonic

ISBN: 978-989-95079-6-8

sensors. Therefore, the robots can conduct image processing, voice recognition, target tracing, collision avoidance, map building, and imitative learning.

We have applied steady-state genetic algorithm (SSGA), spiking neural networks (SNN), self-organizing map (SOM), and others for human detection, motion extraction, gesture recognition, and shape recognition based on image processings [25-28]. However, the image processing takes much computational time and cost. Therefore, we discuss the applicability of fuzzy theory and evolutionary computation in robot vision.

2.2. Active Vision for Robots

Computer vision is a research stream on image processing, image understanding, image recovery and others on a computer [9]. The quality of image depends strongly on the lighting condition related with a camera system. Active vision is often used to improve the robustness and flexibility and to eliminate the ill-posed conditions based on the control of camera systems [9]. Robot vision is deeply related with active vision, because a problem on the perception and action of a robot must be solved at the same time. A robot takes actions to perceive the environment when the robot does not know the environment much. Therefore, the robot vision is based on the time-series of image processing, not the processings on a single image. Various technologies for image processings are required for realizing the robot vision, *e.g.*, color processing, target detection, template matching, shape recognition, motion extraction, and optical flow. Recently, evolutionary computation has been applied to improve the performance of image processing. We also have discussed the applicability of the evolutionary computation in robot vision [32-34]. In fact, we proposed a method of people detection, people tracking and gesture recognition.

2.3 Evolutionary Computation for Robot Vision

Evolutionary computation (EC) is a field of simulating evolution on a computer. Evolutionary optimization methods are fundamentally iterative generation and alternation processes of multiple candidate solutions. The optimization is done by the multi-point search operating on a set of individuals, which is called a population. First, we discuss the role of evolutionary search in dynamic environments. Figure 2 shows the temporal patterns of spatial changes in dynamic environments where the vertical axis indicates the state of environmental conditions represented as a value. If the search speed of EC is faster than the changing speed of the environmental conditions, EC can obtain feasible solutions in the facing environmental conditions. However, EC should be adaptive to the facing environmental conditions if the environmental changes can be observed. If the environmental change is very slow and the change is small (Fig.2 (a)), the mutation range should be large according to the amount of the environmental change. Basically, this kind of change can be considered as some noise in a stationary environment. The environment of Fig.2 (b) includes big changes, but the environmental condition is stationary after a big change. If the big change can be observed, most of candidate solutions should be replaced with randomly generated candidate solutions. The environment of Fig.2 (c) is changing non-stationarily with both features of Figs.2 (a) and (b). In general, the change of visual images corresponds to the environment of Fig.2 (c). The visual image of a mobile robot changes according to both the dynamics of environmental changes and the robotic motion. Therefore,

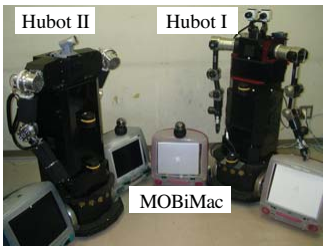


Figure 1: Partner robots; MOBiMac and Hubot

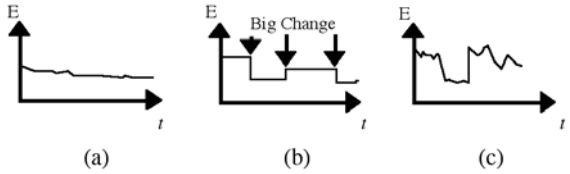
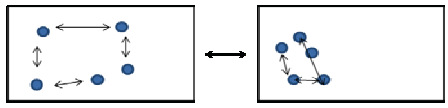
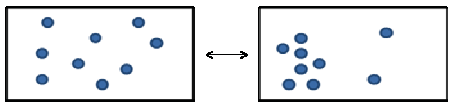


Figure 2: Patterns of changes in dynamic environments

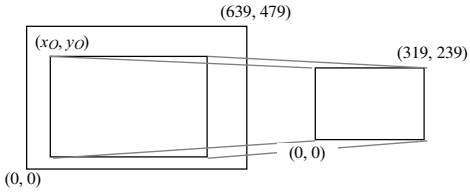


(a) Movement of attentive points in the visual search



(b) Distribution of candidate solutions in evolutionary search

Figure 3: Comparison of visual search and evolutionary search



(a) Original Image (b) Attention range

Figure 4: Preprocessing for human face detection

we discuss visual perception based on active robot vision and EC from the viewpoint of human visual perception.

Figure 3 shows the comparison between the visual search and evolutionary search. The left and right figures show the distribution of search points in the focused search and distributed search, respectively. The visual search controls the searching area based on the fast movement of attentive point. The region of interest (ROI) for the information extraction is deeply related with the search of geometrical features included in the visual target. The search results are reflected to the next visual search. On the other hand, the evolutionary search controls the searching area based on the selection pressure to candidate solutions. If the selection pressure is high, the candidate solutions are centralized toward the better candidate solutions. Otherwise, candidate solutions are globally distributed in the search space. The next search points in the evolutionary search are generated by crossover and mutation. The search in ROI is mainly performed by mutation and local search, while a new search point for ROI is generated by crossover. The degree of interest is calculated by the fitness value. If the fitness value is high, the focused search should be performed. We

ISBN: 978-989-95079-6-8

apply a steady-state genetic algorithm (SSGA) to realize the continuous and real-time search for the robot vision like human visual perception in a dynamic environment, because SSGA can easily obtain feasible solutions through environmental changes with low computational cost.

2.4 Image Processing for Pattern Recognition

Various types of pattern matching methods such as template matching, cellular neural network, neocognitron, and dynamic programming (DP) matching, have been applied for the human detection in image processing. In general, pattern matching is composed of two steps of target detection and target recognition. The aim of target detection is to extract a target candidate from an image (segmentation), and the aim of the target recognition is to identify the target from classification templates. Furthermore, we can discuss the pattern matching from the viewpoints of the bottom-up processing and top-down processing. The candidate detection is considered as the bottom-up processing based on the color distribution in segments of an image, while the target recognition is considered as top-down processing based on the similarity between the extracted target candidate and classification templates. The synthetic combination of bottom-up processing and top-down processing realizes the efficient and effective search. In this paper, we apply growing neural gas and SSGA for bottom-up processing and top-down processing, respectively.

3. Human Tracking based on Evolutionary Robot Vision

3.1 Growing Neural Gas for Bottom-up Preprocessing

An image is a set of pixels with color information. The segmentation into target candidates based on color distribution is very important to reduce the computational cost of direct template matching. Therefore, we apply an unsupervised clustering method for the segmentation based on color distribution in an image. Growing neural gas (GNG) is a competitive learning network as one of variants of SOM [22] used as a clustering method.

The learning algorithm of GNG is shown as follows.

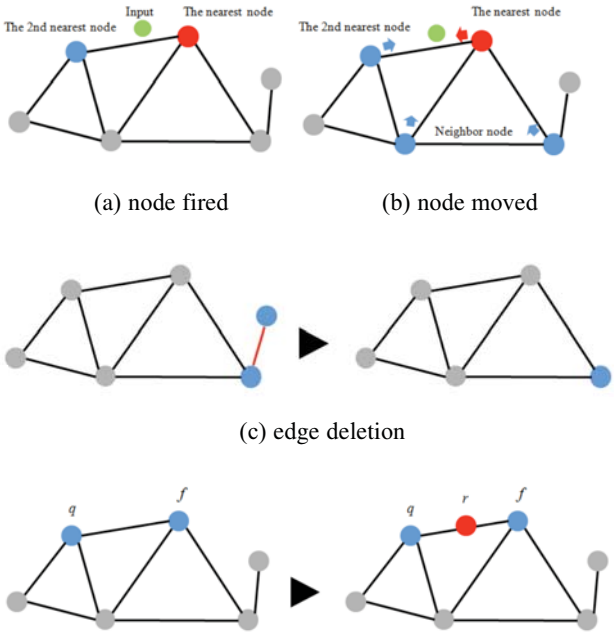
w_i : the nth dimensional vector of a node ($w_i \in \mathbf{R}^n$)
A : a set of nodes
N_i : a set of nodes connected to the ith node
c : a set of edges
$a_{i,j}$: the age of the edge between ith and jth node

Step 0. Generate two units at random position, w_{C1}, w_{C2} in **Rn**. Initialize the connection set.
Step 1. Generate at random an input data v according to $p(v)$.
Step 2. Determine the nearest unit s_1 and the second-nearest unit s_2 by

$$s_1 = \arg\min_{i \in A} \|v - w_i\| \quad (1)$$

$$s_2 = \arg\min_{i \in A \setminus \{s_1\}} \|v - w_i\| \quad (2)$$

where v is composed of the position (x, y) and color information (R, G, B) of color pixel on the image(see Fig.5(a)).
Step 3. If a connection between s_1 and s_2 does not yet exist , create it. Set the age of the connection between s_1 and s_2 to zero.



(a) node fired (b) node moved

(c) edge deletion

(d) node addition

Figure 5: How to learn GNG nodes and edge

Original Image GNG for Human shape recognition

Figure 6: An example of GNG

$$a_{s_1,s_2} = 0 \quad (3)$$

Step 4. Add the squared distance between the input signal and the winner to a local error variable E_{s1}(see Fig.5(b)).

$$E_{s_1} \leftarrow E_{s_1} + \|v - w_{s_1}\|^2 \quad (4)$$

Step 5. Adapt the reference vectors of the winner and its direct topological neighbors by the learning rate η_1^G and η_2^G, respectively.

$$w_{s_1} \leftarrow w_{s_1} + \eta_1^G \cdot (v - w_{s_1}) \quad (5)$$

$$w_j \leftarrow w_j + \eta_2^G \cdot (v - w_j) \quad if\ c_{s_1,j} = 1 \quad (6)$$

Step 6. Increment the age of all edges emanating from s_1.

$$a_{s_1,j} \leftarrow a_{s_1,j} + 1 \quad if\ c_{s_1,j} = 1 \quad (7)$$

Step 7. Remove edges with the age larger than $a_{\max}$. If units have no more emanating edges after this, remove those units(see Fig.5(c)).
Step 8. If the number of input signals generated so far is an integer multiple of a parameter λ, insert a new unit as follows(see Fig.5(d)).
Step 8-1. Determine the unit q with the maximum accumulated error.

$$q = \arg\max_{i \in A} E_i \quad (8)$$

Step 8-2. Determine the unit f with the maximum accumulated error among the neighbors of q.

$$f = \arg\max_{c \in N_q} E_c \quad (9)$$

Step 8-3. Add a new unit r to the network and interpolate its reference vector from q and f.

ISBN: 978-989-95079-6-8

$$w_r = 0.5 \cdot \left(w_q + w_f \right) \quad (10)$$

Step 8-4. Insert edges connecting the new unit r with units q and f, and remove the original edge between q and f.

Step 8-5. Decrease the error variables of q and f by the discount rate α.

$$E_q \leftarrow (1-\alpha)E_q \text{ and } E_f \leftarrow (1-\alpha)E_f \quad (11)$$

Step 8-6. Interpolate the error variable of r from q and f

$$E_r = 0.5 \cdot \left(E_q + E_f \right) \quad (12)$$

Step 9 Decrease the error variables of all units

$$E_c \leftarrow (1-\beta)E_c, \ c \in A \quad (13)$$

Step 10 If a termination condition (e.g., some performance measure) is not yet fulfilled continue with step 2.

Figure 6 shows an example of the learning of GNG. In this way, the color distribution can be extracted from the image by using GNG.

3.2 People Detection and Tracking as Top-down Processing

SSGA simulates a continuous model of the generation, which eliminates and generates a few individuals in a generation (iteration) [13]. The genotype is represented by $g_{i,j}$ (i=1,2,..., G, j=1,2,..., M) and fitness value is represented by f_i. One iteration is composed of selection, crossover, and mutation. The worst candidate solution is eliminated ("Delete least fitness" selection strategy), and is replaced with the candidate solution generated by the crossover and the mutation.

We use the elitist crossover and adaptive mutation [15]. The elitist crossover randomly selects one individual and generates an individual by combining genetic information from the selected individual and the best individual with the crossover probability. If the crossover probability is satisfied, the elitist crossover is performed. Otherwise, a simple crossover is performed between two randomly selected individuals. Next, the following adaptive mutation is performed to the generated individual,

$$g_{i,j} \leftarrow g_{i,j} + \left(\alpha_j \cdot \frac{f_{\max} - f_i}{f_{\max} - f_{\min}} + \beta_j \right) \cdot N(0,1) \quad (14)$$

where f_i is the fitness value of the ith individual, f_{max} and f_{min} are the maximum and minimum of fitness values in the population; $N(0,1)$ indicates a normal random variable with a mean of zero and a variance of one; α_j and β_j are the coefficients ($0<\alpha_j<1.0$) and offset ($\beta_j>0$), respectively. In the adaptive mutation, the variance of the normal random number is relatively changed according to the fitness values of the population in case of maximization problems.

The robot must recognize a human face from complex background speedily. Therefore, we use SSGA for human detection as one of search methods. The human face candidate positions based on human skin and hair colors are extracted by SSGA with template matching. Figure 7 shows a candidate solution of a template used for detecting a human face. A template is composed of numerical parameters of $g_{i,1}^H$, $g_{i,2}^H$, $g_{i,3}^H$, and $g_{i,4}^H$. The number of individuals is G^H. The initial population of SSGA for human detection at the discrete time step t is updated by using the reference vector of GNG in addition to the candidate solutions obtained at the previous time step t-1. The fitness value of the ith individual is calculated by the following equation,

$$f_i^H = C_{Skin}^H + C_{Hair}^H + \eta_1^H \cdot C_{Skin}^H \cdot C_{Hair}^H - \eta_2^H \cdot C_{Other}^H \quad (15)$$

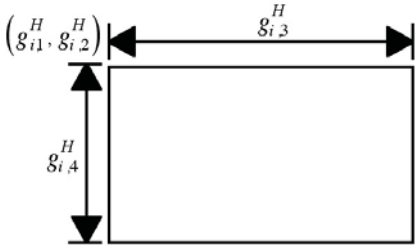
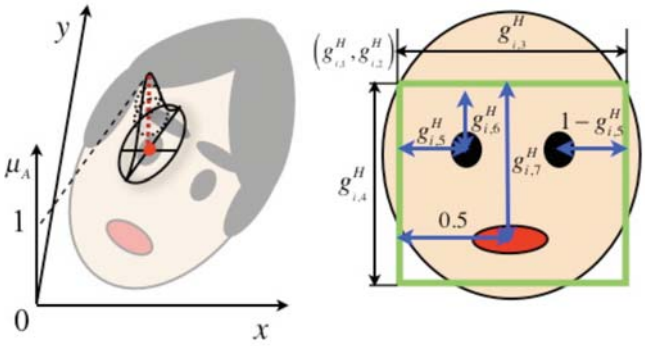


Figure 7: A template used for human detection in SSGA



(a) Fuzzy Evaluation based on Gaussian membership (b) The position of facial landmarks

Figure 8: A template used for face recognition

where C_{Skin}^H, C_{Hair}^H and C_{Other}^H indicate the numbers of pixels of the colors corresponding to human skin, human hair, and other colors, respectively; η_1^H and η_2^H are the coefficients (η_1^H, $\eta_2^H>0$). Therefore, this problem results in the maximization problem. The iteration of SSGA is repeated until the termination condition is satisfied. Here SSGA for the human detection is called SSGA-H.

Since SSGA-H extracts the area of skin colors and hair colors in the human detection, various objects except humans might be detected. Therefore, the human tracking is performed according to the time series position of the ith human candidate ($g_{i,1}^H$, $g_{i,2}^H$) obtained by SSGA-H. The position of the jth human candidate in the human tracking ($X_{k,1}$, $X_{k,2}$) is updated by the nearest human candidate position within the tracking range. In addition, the width and height of the human candidate for the human tracking ($X_{k,3}$, $X_{k,4}$) are updated by the size of the detected human ($g_{i,3}^H$, $g_{i,4}^H$). The update is performed as follows (j=1,2,3,4);

$$X_{k,j} = (1-\lambda)X_{k,j} + \lambda \cdot x_{i,j} \quad (16)$$

Furthermore, the time counter for the reliability of human tracking is used. If the human candidate position in the human tracking is performed, the time counter is incremented. Otherwise, the time counter is decremented. If the time counter is larger than the threshold (HT), the human count is started. Sometimes, several human candidates are close each other, because several human candidates in a single human can be generated by the human detection. Therefore, the removal processing is performed when human candidates are coexisting within the tracking range.

3.3 Fuzzy Evaluation for Face Recognition

The human detection based on color distribution sometime extracts something with similar color distribution. In order to improve the performance of human face recognition, we

ISBN: 978-989-95079-6-8

Table 1: Initial values used in Gaussian membership function

	$a_{h,1}$	$a_{h,2}$	$b_{h,1}$	$b_{h,2}$
Right eye	0.2	0.2	0.4	0.2
Left eye	0.8	0.2	0.4	0.2
Mouth	0.5	0.6	0.4	0.37

Table 2: Search range of $g_{i,5}^H, g_{i,6}^H, g_{i,7}^H$

	$g_{i,5}^H$	$g_{i,6}^H$	$g_{i,7}^H$
min	0.1	0.15	0.7
max	0.35	0.35	0.9

can use the color information of facial landmarks. The human detection by SSGA can be considered as a coarse search, and the extraction of facial landmarks can be considered as a fine search. We apply fuzzy evaluation for face recognition based on the position of facial landmarks (Fig.8 (a)). We use the position of eyes and mouth. We use

$$\mu_{A(h,i,j)} = \begin{cases} \exp\left(-\frac{(x-a_{h,1})^2}{2b_{h,1}^{\ 2}} - \frac{(y-a_{h,2})^2}{2b_{h,2}^{\ 2}} \right) & if\ p(i,j) = c_h \\ 0 & otherwise \end{cases} \tag{17}$$

where (x,y) is the normalized position of the pixel (i,j) on the image; $p(i,j)$ is the color ID of a the pixel (i,j) on the image; c_h is the color ID of the hth facial landmark is the center of a facial landmark; $(a_{h,1},a_{h,2})$ and $(b_{h,1},b_{h,2})$ are the normalized position and size of the hth facial landmark in the template candidate extracted by SSGA. Therefore, this value is high if the color pixel corresponding to the facial landmark is near with the center of the facial landmark. Therefore, we can evaluate the degree of existing each facial landmark as follows;

$$f_{Land,h} = \sum_{(i,j)\in g_k} \mu_{A(h,i,j)} \tag{18}$$

where g_k is the template of the kth candidate solution in SSGA; h is facial the landmark ID. Furthermore, we can evaluate the degree of face as follows;

$$f_{Face} = \prod_{h=1}^{H} f_{Land,h} \tag{19}$$

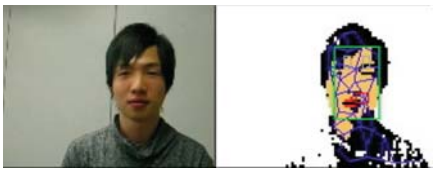
where H is the number of facial landmarks. We use right eye (h=1), left eye (h=2) and mouth (h=3) for the evaluation (H=3). Figure 8 (b) shows the positions of facial landmarks where $(g_{i,5}^H, g_{i,6}^H)$ is the position of the right eye and $(0.5, g_{i,7}^H)$ is the position of the mouth. Therefore, $(a_{1,1},a_{1,2}) = (g_{i,5}^H, g_{i,6}^H)$, $(a_{2,1},a_{2,2}) = (1-g_{i,5}^H, g_{i,6}^H)$, and $(a_{3,1},a_{3,2}) = (0.5, g_{i,7}^H)$. However, the position of each facial landmark is peculiar to a person. Therefore, the position of membership function corresponding to each facial landmark should be updated according to the detected person. As a result, the number of parameters for human detection is 7.

4 Experimental results

This section shows experimental results of human recognition of a partner robot. The maximal number of nodes in GNG is 50. The population size of SSGA is 50. The number of generations (iterations) SSGA in each frame is 200 including the initial evaluations after the frame of image is updated. This value is relatively small comparing with that of the search by standard GA, but the search by SSGA is a time series of continuous search in a dynamic environment including a small change. Table 1 shows the initial values of shape information used in fuzzy evaluation, and Table 2 shows the search range of face parameters used in SSGA.

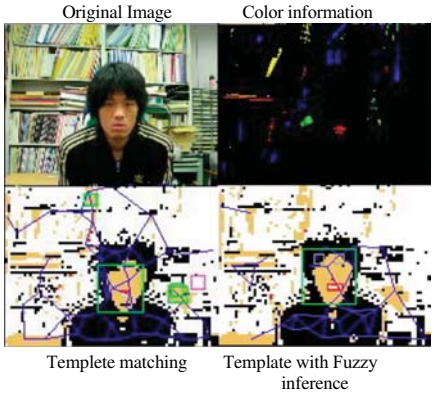
Figure 9 shows experimental results of the proposed method. First of all, we conducted experiments on face recognition of two different people (Case 1 and Case 2). Figure 9.(a) and (b) show snapshots of original image (left), and human-object detection results by SSGA (right) where a

(a) Case 1



(b) Case 2

(c) Case 3



(d) Case 4

Figure 9: Experimental results of face recognition

Table 3: The values of human face obtained in experimental results of Cases 1 and 2

	$g_{i,5}^H$	$g_{i,6}^H$	$g_{i,7}^H$
case 1	0.31	0.19	0.77
case 2	0.26	0.27	0.78

ISBN: 978-989-95079-6-8

green box indicates the extracted human face. Table 3 shows the values of human face obtained in experimental results of Cases 1 and 2. In these results, the proposed method can extract human face with identifying the position of facial landmarks correctly.

Next, we conducted experiments on face recognition where the image includes skin color of object (Case 3) and the image includes the complicated background (Case 4). Figure 9 (c) shows experimental results of the proposed method in Case 3. The proposed method can extract a human face, although the person wears a helmet with having a skin-color book. Figure 9 (d) shows experimental results of the proposed method in Case 4; original image (upper-left), color information (upper-right), template matching without fuzzy evaluation (lower-left), and template matching with fuzzy evaluation (lower-right). The proposed method succeeds to extract a human face correctly. On the other hand, the template matching without fuzzy evaluation fails to extract human face, because some books located in the left of the person on the background image are recognized as a person. In this way, the proposed method can extract a human face in various environmental conditions.

5 Conclusions

This paper proposed a method of human face detection based on the template matching with a steady-state genetic algorithm and fuzzy evaluation. The experimental results show the effectiveness of the proposed method. Membership functions are very useful to evaluate candidate solutions including noisy data. The essence of the proposed method in the flexibility of the search by combining steady-state genetic algorithm and fuzzy evaluation.

As a future work, we will develop a method of human face detection in case of rotation of human face. Furthermore, we apply the proposed method to the associative learning between the perceptual information and symbolic information peculiar with the interacting person.

REFERENCES

[1] J.Cassell, Embodied Conversational Agents: Representation and Intelligence in User Interface, AI Magazine, Winter Vol.22, No.3, pp.67-83, 2001.

[2] Y.Nakauchi, R.Simmons, A Social Robot that Stands in Line, Journal of Autonomous Robots, Vol. 12, No. 3, pp.313-324, 2002.

[3] M.Imai, T.Ono, H.Ishiguro: Physical Relation and Expression: Joint Attention for Human-Robot Interaction, IEEE Transactions on Industrial Electronics, Vol.50, No.4, pp.636-643, 2003.

[4] H.Ishiguro, M.Shiomi, T.Kanda, D.Eaton, and N.Hagita, Field Experiment in a Science Museum with communication robots and a ubiquitous sensor network, Proc. of Workshop on Network Robot System at ICRA2005, 2005.

[5] J.A.Anderson, E.Rosenfeld,. *Neurocomputing*, The MIT Press, Cambridge, Massachusetts, US , 1988.

[6]. David Marr, Vision, W. H. Freeman, San Francisco, 1982.

[7] C.-H. Chen, L.-F. Pau, P.-S-P. Wang, *Handbook of Pattern Recognition & Computer Vision*, World Scientific Publication Co. Pte. Ltd, 1993.

[8] K.Fukushima Neural network model restoring partly occluded patterns, *Knowledge-Based Intelligent Information and Engineering Systems, (7th International Conference, KES 2003)*, Part II, eds: V. Palade, R. J. Howlett, L. Jain, Berlin - Heidelberg: Springer-Verlag, pp. 131-138, 2003.

[9] H.H.Bulthoff, S..W. Lee, T.A. Poggio, C. Wallraven , *Biologically Motivated Computer Vision*, Springer-Verlag, 2002.

[10] D.Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Massachusetts: Addison Wesley Publishing Company Inc, 1989.

[11] J.H.Holland, Adaptation in Natural and Artificial Systems , First edition, Massachusetts, The MIT Press, 1992.

[12] D.Fogel, Evolutionary Computation, New York: IEEE Press, 1995.

[13] G.Syswerda, G. A study of reproduction in generational and steady-state genetic algorithms, *Genetic Algorithms*, Morgan Kaufmann Publishers Inc., San Mateo, 1991.

[14] S.Watanabe, T.Hiroyasu, M.Miki, LCGA : Local Cultivation Genetic Algorithm for Multi-Objective Optimization Problems, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO2002), pp.702, 2002.

[15] M.Lozano, F.Herrera, N.Krasnogor, D.Molina, Real-Coded Memetic Algorithms with Crossover Hill-Climbing, Evolutionary Computation, Vol. 12, No. 3, Pages 273-302, MIT Press, Cambridge, Massachusetts, US, 2004.

[16] J.Branke, Evolutionary Optimization in Dynamic Environments, Volume 3 of the Book Series on Genetic Algorithms and Evolutionary Computation, Kluwer Academic Pub Published, 2001.

[17] R. C. Eberhart, J. Kennedy, Y. Shi, Swarm Intelligence, Morgan Kaufmann Publisher, San Francisco, 2001.

[18] E. Bonabeau, M. Dorigo, G. Theraulaz, Swarm Intelligence: From Natural to Artificial Systems, Oxford University Press, New York, 1999.

[19] M. Dorigo and T. Stutzle, Ant Colony Optimization. Bradford Books, The MIT Press, 2004.

[20] Indra A.Sulistijono and N.Kubota, Human Head Tracking Based on Particle Swarm Optimization and Genetic Algorithm, Journal of Advanced Computational Intelligence and intelligent Informatics, Vol.11 No.6, Jul. 2007, pp.681-687, 2007

[21] D.Sperber, D.Wilson, Relevance - Communication and Cognition, Blackwell Publishing Ltd., 1995.

[22] B. Fritzke. A growing neural gas network learns topologies. In G. Tesauro, D. S. Touretzky, and T. K. Leen, editors, Advances in Neural Information Processing Systems 7, pages 625-632. MIT Press, Cambridge MA, 1995.

[23] M.Hynar, M.Burda, J.Sarmanova: Unsupervised clustering with growing self-organizing neural network -- a comparison with non-neural approach. DATESO pp.58-68, 2005

[24] B. Fritzke. Growing cell structures - a self-organizing network for unsupervised and supervised learning. Technical Report ICSTR-93-026, International Computer Science Institute, Berkeley, May 93.

[25] N.Kubota, K.Nishida, Cooperative Perceptual Systems for Partner Robots Based on Sensor Network, International Journal of Computer Science and Network Security, Volume 6, Number 11, pp.19-28, 2006.

[26] N.Kubota, Visual Perception and Reproduction for Imitative Learning of A Partner Robot, WSEAS Transaction on Signal Procesing, Issue 5, Vol. 2, pp. 726-731, 2006.

[27] N.Kubota, Computational Intelligence for Structured Learning of A Partner Robot Based on Imitation, *Information Science*, No. 171, pp.403-429, 2005.

[28] N.Kubota; Y.Nojima, F.Kojima, T.Fukuda,. Multiple Fuzzy State-Value Functions for Human Evaluation through Interactive Trajectory Planning of a Partner Robot, *Soft Computing,* Vol.10, No.10 pp.891-901, 2006.

[29] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld,"Face Recognition: A Literature Survey", ACM Computing Survey, December Issue, pp.399-458, 2003.

[30] N.Kubota, A.Yorita, Structured Learning for Partner Robots based on Natural Communication, Proc. (CD-ROM) of 2008 IEEE Conference on Soft Computing in Industrial Applications (SMCia), pp. 303-308, 2008.

[31] N.Kubota, A.Yorita, M.Satomi, Visual Tracking of A Partner Robot through Interaction with People, Proc. of Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems (SCIS & ISIS 2008), pp. 1032-1037, 2008.

[32] N.Kubota, Y.Tomioka, Evolutionary Robot Vision for Human Tracking of Partner Robots in Ambient Intelligence, Proc (CD-ROM) of Congress of Evolutionary Computation, pp. 1491-1496, 2007.

[33] N.Kubota, T.Ohinata, S.Wakisaka, H.Liu, Extraction of Facial Landmarks for a Partner Robot Based on Evolutionary Computation, Proc. of Computation Intelligence, Robotics and Autonomous Systems (CIRAS 2007), pp. 44-49, 2007.

[34] Indra A.Sulistijono, and N.Kubota, Evolutionary Robot Vision for Multiple Human Heads Tracking of A Partner Robot, Proceeding of 16th IEEE International Conference on Robot & Human Interactive Communication (Ro-Man2007), pp. 463-468, 2007.

ISBN: 978-989-95079-6-8

Combining Wavelets and Computational Intelligence Methods with applications on Multi-class Classification datasets.

Carlos C. Bracho[1]

1 Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada
Email: carloscb@ece.ualberta.ca

Abstract—*In this paper, we propose a novel algorithm for wavelet feature extraction as input to a supervised Multi-Class Classifier to improve classification performance. In particular, to select the best wavelets coefficient features, we first compute the energy-based variance distribution from wavelets coefficients at different subbands as well as the entropy-based fuzzy measures associated with the training instances. Once we get these entropy-based fuzzy measures associated with the different subsets of wavelets subbands, we apply the Möbius Transform to these entropy-based fuzzy measures to extract the Multivariate Mutual Information associated with the different subsets of wavelets subbands. The goal of these measures is twofold: assign weights (based on the wavelets information content) to all subsets of wavelets subbands and extract the independent (in terms of the multivariate mutual information) subsets of wavelets subbands. In our case, the optimal subsets of wavelets subbands as wavelets features vectors to train a Bayesian Network Model are those which provide a multivariate mutual information equal to zero. Experimental results with the multi-class SRBCT cancer dataset, show that our proposed approach achieves lower classification error in comparison with other methods proposed in the literature.*

Keywords— fuzzy measures, mutual information, wavelets.

1 Introduction

In recent years, Wavelets-based classification have gained a great relevance in different kinds of applications such as image processing [6][7], time series[8][2], bioinformatics[9][12], and pattern recognition[10][11]. The early research in wavelets-based classification was focused mainly on extracting energy values from the wavelet-subbands decomposition and using them for classification. However, in pattern recognition tasks such as classification, it is well known that proper feature selection help to improve the classification performance. Hence, a process which consists on removing irrelevant wavelets features as well as selecting a subset of wavelets from the wavelet-subbands decomposition has had a paramount importance in Wavelets-based Classification tasks.

In this paper, we propose a novel algorithm that uses entropy-based fuzzy measures and multivariate mutual information–based tests to identify the suitable subsets of wavelets features as the correct input to a supervised K2-based Bayesian Network to improve classification performance. At present time, Bayesian Networks [13[14] have became one of the most important technologies in the area of applied artificial intelligence. Roughly speaking, Bayesian Networks are graphical structures that model the probabilistic cause-effect relationships among several related variables. Particularly, the K2 Algorithm[15] is a greedy algorithm that learns the structure of Bayesian Networks from data in an efficient way given a prior order of the variables (in improper order of data will provide poor classification performance). In our case, we are proposing entropy-based fuzzy measures, its Möbius representation and multivariate mutual information to provide not only the correct order but also the correct wavelets features to a K2-based Multi-Class Bayesian Network for classifying a multi-class cancer dataset.

The remaining of this paper is organized as follows: in Section 2, we provide background on Wavelets, Bayesian Networks, Fuzzy Measures, Entropy-based Fuzzy Measures and Fuzzy Measures-based Multivariate Mutual Information. In Section 3, we detail our multi-class classification approach. In Section 4 we present the experimental results obtained using a cancer dataset with our proposed approach. Finally, in Section 5, we offer a summary and discussion of future work.

2 Background

2.1 The Wavelet Transform

The Wavelet Transform (WT)[2] is a natural context for analyzing the time-varying properties of most real-world signals. In a narrow sense, WT decomposes a signal into different time scales (wavelet- subbands decomposition) to detect features of its short-term as well as its long-term dynamics. Essentially, this transform is defined in two spaces[2]: continuous and discrete. In the continuous space, the Continuous Wavelet Transform CWT is defined by means of two basic functions (known as the mother and the father wavelet function, respectively). The father wavelet function $\varphi(t)$ depending of time t is used to capture the smooth part of the data. By contrast, the mother wavelet function $\psi(t)$ is stretched and translated to detect the rough components of the data. These two functions, to be considered as wavelet functions, must satisfy certain properties [2].

Complementary to the CWT described above, the Discrete Wavelet Transform (DWT) can be defined from an engineering point of view by means of cascading a set of couple of low and high filters [2]. These couples of filters are known in the literature as the wavelet and scaling filters. Together, this set of filters must fulfill the Quadrature Mirror Filter (QMF) property to obtain a perfect reconstruction pass-band filter.

Of particular interest in this paper, is the application of a modified version of DWT known as the Maximal Overlap

ISBN: 978-989-95079-6-8

Wavelet Transform (MODWT)[2], to provide wavelets features to a K2-based Bayesian Network. Roughly speaking, MODWT is a rescaled version of the wavelet and scaling filter used in the common DWT but with the following advantages:

- Wavelet coefficients are aligned with the original signal values.
- MODWT provides a suitable format for the traditional tabular framework used in machine learning approaches.
- MODWT provides more redundancy at each wavelet-subband.

Definition 1
Let *j*, *t* be the decomposition level (wavelet subband), and time index, respectively, and X_t a real-valued uniformly sampled signal whose size is *N*. The j^{th} level MODWT wavelet and scaling coefficient $W_{j,t}$ and $V_{j,t}$, respectively, are defined by convolving the impulse response of the wavelet and scaling filters with the signal X_t in the following way [2]:

$$W_{j,t} = \sum_l h_{j,l} * X_{t-l \bmod N} \quad (1)$$

$$V_{j,t} = \sum_l g_{j,l} * X_{t-l \bmod N} \quad (2)$$

Where *l* is the length of both the wavelet filter impulse response *h* and the scaling impulse response *g*. In this paper, MODWT was implemented by means of a Pass-Band Wavelet Filter (using the WMTSA matlab toolbox [2]). In particular, we use a Daubechies Extremal Phase Filter with *l=6* (DB6). One fundamental property of this filter is that it transforms a signal in terms of difference of several order of averages at different wavelet-subbands.

2.2.1 MODWT Energy-based variance distribution
Another important property of WT is its ability to preserve the energy of a signal. Roughly speaking, the WT decomposes the total energy of a signal at different wavelet-subbands defining an Energy Density Distribution (EDD). In particular, this EDD is computed by means of wavelets and scaling coefficients of MODWT expressed as follows[2]:

$$\sum_t X_t^2 = \sum_j \sum_{n=0}^{N_j-1} W^2{}_{j,t} + \sum_{n=0}^{j-1} V^2{}_n \quad (3)$$

A closely related concept to the wavelet-based energy decomposition is the wavelet variance which can be associated with a probability distribution and used in this paper for purposes of computing the entropy-based fuzzy measures. Mathematically, this wavelet variance is defined as follows [2]:

$$\sigma_X^2 = \frac{1}{N} \sum_{j=1}^{MaxLevel} \left\| W_j \right\|^2 + \frac{1}{N} \left\| V_{MaxLevel} \right\|^2 - \overline{X}^2 \quad (4)$$

Where:

N: Number of observations in the signal.
$MaxLevel$: Maximum level of decomposition .

W_j : Vector of wavelet coefficients defined at subband j .

$V_{MaxLevel}$: Vector of scaling coefficients defined at subband j .

$\overline{X}$: Sample mean of signal X_t.

2.2 Bayesian Networks
Bayesian Networks [16][13] have become one of the most important technologies in the area of applied artificial intelligence. They have shown to offer reliable methods for prediction, decision making, classification, and data mining in different areas such as medicine, image processing, marketing, banking, finance, etc. Roughly speaking, Bayesian Networks are graphical structures that model the probabilistic cause-effect relationships among several related variables. Mathematically, a Bayesian Network is modelled by means of a Directed Acyclic Graph (representing the above cause-effect relationships) and a set of Conditional Probabilities (one conditional probability for each node given its parent set in the graph).

Definition 2
A Bayesian Network $B = (\chi, G, P)$ consists of :

- A Direct Acyclic Graph $G = (V, E)$ with nodes $V = (V_1, V_2, .., V_n)$.
- A set of discrete random variables χ modeling the nodes of G.
- A set of conditional probability distributions $P(X_v \mid X_{parent(v)})$ for each random variable $X_v \in \chi$ where :

$$P(\chi) = {}_{v \in V} P(X_v \mid X_{pa(v)}) \quad (5)$$

The last property specifies a joint probability distribution over X known as the Bayesian Network chain rule.

ISBN: 978-989-95079-6-8

2.2.1 K2-based Bayesian Network learning

Generally speaking, the structure learning process under the K2 algorithm is made by means of an optimization process in which a quality measure Q of a Bayesian Network structure given a training dataset X is maximized; that is:

$$\max Q(B_S \mid X) \tag{6}$$

K2 is a greedy search algorithm that learns from data presenting a good performance when a prior order of the data exists. This order refers to have first in a sequence the parents and then the children. Assuming that the attributes have an order, the algorithm starts setting the parent of node x_1 to the empty set $\varnothing$. Then, the algorithm visists each subsequent node following the order in the sequence order adding $Parents_i$ to the set of parents of node attribute x_i only if the inclusion of the $Parents_i$ to the node attribute x_i maximize the network-structure posterior probability distribution. At the same time, it includes the relationship between the parent and the children in question by adding arcs between them.

2.2.2 Fuzzy Measures

Fuzzy measures are mainly known in the context of Fuzzy Integrals[3][5]. Grabisch[4], presents the usage of fuzzy measures and fuzzy integral in supervised classification and feature extraction problems. In these papers the author proposes an optimization approach which is based on a heuristic least mean squares (HLMS) algorithm, to compute the fuzzy measures as well as the Choquet Integral in classification tasks obtaining remarkable results.

Definition 3

Let the triple $K = (\Omega, A, g)$ a fuzzy measure space containing a family A of subsets defined on Ω. A set-valued function $g : A \to 0,1]$ is called a fuzzy measure if:

$$g(\varnothing) = 0 \tag{7}$$

$$g(S) = 1 \tag{8}$$

$$A_i \subset A_j \to g(A_i) \le g(A_j). \forall A_i, A_j \in S \tag{9}$$

Any set-valued function $g : A \to [0,1]$ can be uniquely expressed in terms of its MöbiusTransform represented by:

$$g(T) = \sum_{S \in T} m^g(S), \ \forall T \subseteq A, \tag{10}$$

where the set function $m^g : A \to [0,1]$ is called the Möbius Inversion Transform of g and is given by[17]:

$$m^g(S) = \sum_{T \subseteq S} (-1)^{|S|-|T|} g(T), \ \forall S \subseteq A, \tag{11}$$

2.2.2.1Entropy-based fuzzy measures

In comparison with the Grabish approach, in this paper we are interested in applying and extending a novel unsupervised approach to compute entropy-based fuzzy measures proposed by Kojadinovic[1]. The main characteristic of this approach is the replacement of the notion of attributes importance with the notion of information content in attributes by means of the entropy and the mutual information concepts.

Definition 5

Let 2^S denote the power set of $S = \{W_1, W_2, .., W_m\}$. An entropy-based fuzzy measure is a set-valued function $g_H : 2^S \to [0,1]$ defined as follows[1]:

$$g_H(S) = \begin{cases} 0 & S = \varnothing \\ (-1)^{|S|+1} H(P_{(\overleftarrow{W}_1,..)}) & S = \{\overleftarrow{W}_1,.\} \end{cases} \tag{12}$$

where :

$H(P_{(\overleftarrow{W}_1,..)})$: Entropy associated with the joint probability distribution $P_{(\overleftarrow{W}_1,..)}$). (In our case W_m means the maximum wavelet-subband, $P_{(\overleftarrow{W}_1,..)}$ refers to the joint probability distribution among the different subset of wavelet-subbands, and $H(P_{(\overleftarrow{W}_1,..)})$ to their associated entropy).

Definition 6

Any entropy-based fuzzy measure $g^H : 2^S \to [0,1]$ can be estimated by:

$$\tilde{g}^H = \frac{H(P_{(W_{i1},..,W_{is})})}{H(P_{(W_{i1},..,W_{iM})})} \tag{13}$$

Definition 7

Any entropy-based fuzzy measure $g^H : 2^S \to [0,1]$ can be uniquely expressed in terms of its Möbius Transform represented by:

$$g^H(S) = \begin{cases} 0 & S = \varnothing \\ (-1)^{|S|+1} I(\overleftarrow{W}_1;..) & S = \{\overleftarrow{W}_1;..\} \end{cases} \tag{14}$$

where :

I : *Multivariate Mutual Information*

ISBN: 978-989-95079-6-8

Proposition 1
A set of wavelets subbands is called independent if its Möbius Transform associated with its entropy-based fuzzy measure is equal to zero.

Proof. From equation 7, clearly $(-1)^{|S|+1} \neq 0$ for all S. Hence, $I(\overleftarrow{X}_1;..\overleftarrow{X}_r) = 0$ implies $g^H(S) = 0$ ■.

3 Proposed Approach

The proposed approach aims to identify the suitable wavelet information as the correct input for the the Bayesian Network. By contrast with other wavelet feature extraction approaches reported in the literature (in which most of them deal at the granularity of wavelets coefficients), this approach presents a novel approach which deals with sets of wavelets-subbands using entropy-based fuzzy measures, its Möbius Transform and multivariate mutual information. To compute the entropy-based fuzzy measures associated with each set of wavelets-subbands, we first consider the energy-based variance distribution at different wavelet-subbands (this Energy Distribution is discretized to fulfill the monotonicity property of fuzzy measure) as random vectors. This set of energy-based variance distributions are used to estimate the entropy-based fuzzy measures associated with the different sets of wavelets subbands. Once we have defined the set of entropy-based fuzzy measures, we apply the Möbius Inversion Transform to these entropybased fuzzy measures and then we select those sets of wavelets subbands whose Möbius Inversion Transform is equal to zero (i.e. set of independent wavelets-subbands in all cancer classes whose multivariate mutual information is equal to zero). This set of wavelets subbands are the wavelet-training and wavelet-testing to our supervised Bayesian Network Model (see Algorithm I and II).

Algorithm I: Training Phase

1.- Data Preprocessing withWavelet Transform

- Apply Daubechies (DB6) Extremal Phase Filter to the training dataset.

2.- Energy-based Variance Distribution Estimation and discretization

- Compute equations (3) and (4).

3.- Estimate entropy-based Fuzzy Measures

- Compute the Wavelet-subbands Join Probability Distributions with the maximum likelihood estimator[1].
- Compute the entropy associated with the above Wavelet-subbands Joint Probability Distributions.
- Compute entropy-based fuzzy measures using equation (13).

4.- Apply the Möbius Transform to the entropy-based Fuzzy Measures obtained in step 3 to obtain their associated Multivariate Mutual Information

- Compute equations (14).

5.- Wavelet Feature Vector Extraction to train Bayesian Network.

- Select sets of wavelet-subbands whose Möbius Transform is equal (or aprox.) to zero (Proposition 1).
- Concatenate wavelets coefficients of different wavelet-subbands that fulfill the above condition. These wavelets features are the input to train a Bayesian Network. (If exists several sets of wavelet-subbands that fulfill the above condition, you might select the subset with less wavelet coefficients)

End Algorithm I

Algorithm II: Testing Phase

1.- Data Preprocessing withWavelet Transform

- Apply Daubechies (DB6) Extremal Phase Filter to the testing dataset.

2.- Select the same set of wavelet-subbands obtained in step 5 of algorithm I.

- Apply 10-fold cross validation to verify performance of the trained Bayesian Network with the selected wavelet coefficients.

End Algorithm II

4 Experimental results

Given the great interest in the scientific community in exploring DNA microarray which in general consist of thousands of genes, to evaluate the performance of our proposed approach, we use a highly and noisy multidimensional multi-class microarray dataset. In particular, we use a cancer microarray dataset known as SRBCT[9] consisting of 2,308 genes and 63 samples from four types of cancers: neuroblastoma (NB), rhabdomyosacoma (RMS), Burkitt lymphomas (BL) and Ewing family of tumors(EWS). In our case, we consider a total of 32 instances for the training phase partitioned uniformly with respect to each kind of cancer (8 samples for each type of cancer) and 31 instances as the testing dataset (eight samples from each type of cancer except one with seven instances as testing corresponding to the Burkitt lymphomas).

ISBN: 978-989-95079-6-8

In the training phase after applying the first four steps, we found that the wavelet-subband 5 (containing 2308 wavelet coefficients) provide us a multivariate mutual information equal to zero fulfilling the condition established in proposition 1. Once we apply step 5 of algorithm 1, we use a 10-fold cross-validation method and consider an input SRBCT-training dataset to the Multi- Class Bayesian Network (which is implemented. in the data mining environment known as weka [19]) in the following way:

$$T = \left\{ \left(\overleftarrow{W}_i, y_i \right) : \overleftarrow{W}_i \in \Re^{2308} \right\} \tag{14}$$

where:

$$\overleftarrow{W}_i \in \Re^{2308} \tag{15}$$

$$y_i \in \{NB, RMS, BL, EWS\} \tag{16}$$

Once trained, we test our Bayesian Network classifier obtaining the following results reported in table I (in comparison with other results reported in the literature, in this table we can see that there are two methods providing 100% of accuracy partitioning the multi-class problem as a collection of N binary sub-problems. In our case, we solve the multi-class problem without using this artefact):

Table I. Classification Performance

Method	Accuracy %
Our Method	**92%**
KNN	30%
Naive Bayes	60%
lvs1 SVM/Wavelet[9]	100%
SVM-OVA	80%
Decision Tree	75%

5 Conclusions

We have reported experimental results using a novel wavelet-computational intelligence approach for improving classification of multi-class datasets. In particular, we propose a novel approach for extracting entropy-based fuzzy measures from the wavelet-subband decomposition as well as their multivariate mutual information (by means of the Möbius Inversion Transform). With this information, we select the most suitable wavelets (based on the concept of independence given in terms of multivariate mutual information) to train and evaluate a multiclass Bayesian Network to classify cancer types obtaining satisfactory results. In conclusion, the propose approach provides a general framework to deal with data which are inherently noisy and high dimensional such as microarray gene expression data. Future work in this research will proceed in refining and validating the proposed approach in other applications.

Acknowledgment

This work was partially supported by CONACYT Mexico. The author also likes to thank all the reviewers of this paper for their valuable comments.

References

[1] I. Kojadinovic, Unsupervised aggregation of commensurate correlated attributes by means of the Choquet Integral and Entropy Functionals. *International Journal of Intelligent Systems*. Vol 23, pp. 128-154. 2008.

[2] Donald B. Percival, A. T.Walden,*Wavelet Method for Time Series Analysis*. Cambridge Series in Statistical and Probabilistics Mathematics. 2000.

[3] G. Choquet, Theory of Capacities. *Annales de l'Institut Fourier*, Vol 5, pp. 131-295. 1953.

[4] M. Grabisch*, Fuzzy integral for classification and feature extraction.* Fuzzy measures and integrals Theory and applications. Physica Verlag, pp. 135-148. 2000.

[5] G. Beliakov, *Aggregation Functions: A guide for Practitioners*. Springer Studies in Fuzziness and Soft Computing. 2007.

[6] K. Huang and S. Aviyente. Information-theoretic wavelet packet subband selection for texture classification.*Signal Processing* 86, 1410-1420, 2006.

[7] M. Shoshany.Wavelet decomposition for reducing flux density effects on hyperspectral classification. *IEEE Geoscience and Remote Sensing Letters*, v6, No. 1, 38-41, January 2009.

[8] T. Qian. *Wavelet Analysis and Applications*. Birkhauser, First Edition, January 2007.

[9] S. Prabakaran, R. Sahu, S. Verma. Classification of multi class dataset using wavelet power spectrum. *Data Mining Knowledge Discovery*, v15, 297-319, 2007.

[10] D. Li,W. Pedrycz, N. Pizzi. FuzzyWavelet packet based feature extraction method and its application to biomedical signal classification. *IEEE Transactions on Biomedical Eng*, v52-6, 1132- 1139, June 2005.

[11] T. Celik, and T. Tjahjadi. Multiscale texture classification using dual-tree complex wavelet transform. *Pattern Recognition Letters*, v30-3, 331-339, February 2009.

[12] C. Aggarwal, and P. Bradley. On the use of Wavelet Decomposition for String Decomposition. *Data Mining Knowledge Discovery*, v10-2, 117-139, 2005.

[13] U. Kjaerulff, and A. Madsen. *Bayesian Networks and Influence Diagrams*. Information Science and Statistics, Springer, 2008.

[14] X. Chen, G. Anantha, X. Lin. Improving Bayesian Network Structure Learning with Mutual Information-Based Node Ordering in the K2 Algorithm. *IEEE Transactions on Knowledge and Data Eng*, v20-5, 1-13, May, 2008.

[15] G. Cooper and E. Herskovits. A Bayesian Method for the induction of Probabilistic Networks from Data. *Machine Learning*, v9, 309-347, 1992.

[16] R. Neapolitan and X. Jian. *Probabilistics methods for financial and marketing informatics*. Morgan and Kauffman publishers, 2007.

[17] H. Nguyen and E. Walker. *A first course in Fuzzy logic*. Chapman & Hall/CRC , 2006.

[18] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006.

[19] I. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, Morgan Kauffman, 2005.

ISBN: 978-989-95079-6-8

Poincaré recurrence theorem in MV-algebras

Beloslav Riečan
Faculty of Natural Sciences, Matej Bel University
Tajovského 40, 974 01 Banská Bystrica, Slovakia
Mathematical Institute of Slovak Acad. of Sciences
Štefánikova 49, 814 73 Bratislava, Slovakia
Email: riecan@fpv.umb.sk

Abstract— *The classical Poincaré weak recurrence theorem states that for any probability space $(\Omega, \mathcal{S}, P)$, any P-measure preserving transformation T, and any $A \in \mathcal{S}$, almost all points of A return to A. In the present paper the Poincaré theorem is proved when the σ-algebra $\mathcal{S}$ is replaced by any σ-complete MV-algebra.*

Keywords— Measure preserving transformation, Poincaré recurrence theorem, σ-complete MV-algebra.

1 Introduction

Let $(\Omega, \mathcal{S}, P)$ be a probability space, i.e. $\Omega \neq \emptyset$, $\mathcal{S}$ is a σ-algebra of subsets of Ω (i.e. $\Omega \in \mathcal{S}; A \in \mathcal{S} \Longrightarrow \Omega \setminus A \in \mathcal{S}$, and $A_n \in \mathcal{S}$ $(n = 1, 2, ...) \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{S}$), and $P : \mathcal{S} \to [0, 1]$ is such that $P(\Omega) = 1$, and $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$, whenever $A_n \in \mathcal{S}$ $(n = 1, 2, ...)$ and $A_n \cap A_m = \emptyset$ $(n \neq m)$. Let $T : \Omega \to \Omega$ be such that $A \in \mathcal{S}$ implies $T^{-1}(A) \in \mathcal{S}$, and

$$P(T^{-1}(A)) = P(A), \tag{1}$$

whenever $A \in \mathcal{S}$; such transformations T are called measure-preserving. The Poincaré recurrence theorem states that almost every point $x \in A$ will return to A, i.e.

$$P(A \setminus \bigcup_{n=1}^{\infty} T^{-n}(A)) = 0, \tag{2}$$

whenever $A \in \mathcal{S}$. There is also the stronger variant of the Poincaré theorem: almost all points of A will return to A infinitely many times. It means that for any $x \in A \setminus B$ (where $P(B) = 0$) and any $k \in N$ there exists $n \geq k$ such that $T^n(x) \in A$:

$$P(A \setminus \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} T^{-1}(A)) = 0. \tag{3}$$

The strong recurrence theorem has been proved in various connections (see [7] for Boolean algebras or [3] for topological spaces). In [8] the theorem has been proved for a special class of MV-algebras (σ-complete weakly σ-distributive MV algebras with product). In the present paper we prove the weak version for arbitrary MV-algebras.

2 MV-algebras

The notion of an MV-algebra has been introduced by Chang [1] as an algebraic system

$$(M, \oplus, \odot, \neg, u, 0)$$

where $\oplus, \odot$ are binary operations, $\neg$ is a unary oparation and $0, u$ are fixed elements. By the Mundici theorem ([4]) every MV-algebra corresponds to a unique unital l-group G (lattice ordered group with a distinguished order unit u),

$$\begin{aligned} M = [0, u] &= \{x \in G; 0 \leq x \leq u\}, \\ a \oplus b &= (a + b) \wedge u, \\ a \odot b &= (a + b - u) \vee 0, \\ \neg a &= u - a. \end{aligned} \tag{4}$$

Analogously as in quantum structures (see [2]), instead of a probability masure, a state $m : M \to [0, 1]$ can be defined.

2.1. Definition. A *state* on an MV-algebra $(M, \oplus, \odot, \neg, u, 0)$ is a mapping $m : M \to [0, 1]$ satisfying the following conditions:

(i) $m(u) = 1$;

(ii) m is additive,
i.e. $a \odot b = 0 \Longrightarrow m(a \oplus b) = m(a) + m(b)$;

(iii) m is continuous, i.e. $a_n \nearrow a \Longrightarrow m(a_n) \nearrow m(a)$.

2.2. Proposition. Any state $m : M \to [0, 1]$ is strongly additive, i.e. it satisfies the following implication $a_1 + a_2 + ... + a_n \leq u \Longrightarrow m(a_1 + a_2 + ... + a_n) = \sum_{i=1}^{n} m(a_i)$.

Proof. It can be proved by induction.

The following proposition is evident.

2.3. Proposition. A mapping $m : M \to [0, 1]$ is a state if and only if the following conditions are satisfied:

(i) $m(u) = 1$;

(ii) if (a_n) is a sequence of elements of M such that $a_1 + a_2 + ... + a_n \leq u$ for any $n \in N$, then

$$m(\bigvee_{n=1}^{\infty} \sum_{i=1}^{n} a_i) = \sum_{i=1}^{\infty} m(a_i). \tag{5}$$

3 Main result

3.1. Definition. Let M be a σ-complete MV-algebra, and $m : M \to [0, 1]$ be a state. By an *m-preserving transformation* of M we understand a mapping $\tau : M \to M$ satisfying the following conditions:

(i) $\tau(u) = u, \tau(0) = 0$;

ISBN: 978-989-95079-6-81

(ii) $\tau(a \odot b) = \tau(a) \odot \tau(b)$;

(iii) $a \leq b \Longrightarrow \tau(b - a) = \tau(b) - \tau(a)$;

(iv) $\tau(a \wedge b) = \tau(a) \wedge \tau(b)$;

(v) $\tau(\bigvee_{n=1}^{\infty} a_n) = \bigvee_{n=1}^{\infty} \tau(a_n)$;

(vi) $m(\tau(a)) = m(a)$.

3.2. Definition. $a \setminus b = a \odot (\neg b)$.

3.3. Theorem. Let M be a σ-complete MV-algebra, $m : M \to [0, 1]$ be a state. Then for any $a \in M$

$$m(a \setminus \bigvee_{i=1}^{\infty} \tau^i(a)) = 0. \tag{6}$$

Proof. Put $b_j = \bigvee_{i=j}^{\infty} \tau^i(a), b = a \setminus \bigvee_{i=1}^{\infty} \tau^i(a) = (a - \bigvee_{i=1}^{\infty} \tau^i(a)) \vee 0 = (a - b_1) \vee 0$.

By induction it can be proved that

$$b + \tau(b) + ... + \tau^n(b) \leq \bigvee_{j=0}^{n+1} \bigvee_{i=j+1}^{n+1} (\tau^j(a) - b_i) \vee 0.$$

Since

$$(\tau^i(a) - b_i) \vee 0 \leq \tau^i(a) \leq u$$

we obtain

$$b + \tau(b) + ... + \tau^n(b) \leq u.$$

By the strong additivity property

$$m(\bigvee_{i=0}^{\infty} \tau^i(b)) = \sum_{i=0}^{\infty} m(\tau^i(b)) = \sum_{i=0}^{\infty} m(b),$$

whence

$$m(b) = 0.$$

3.4. Example.

Consider a probability space $(\Omega, \mathcal{S}, \mu)$ and the MV-algebra $\mathcal{M} = \{f : \Omega \to [0, 1]; f \text{ is } \mathcal{S} \text{ - measurable } \}$,

where

$$f \oplus g = \min(f + g, 1),$$
$$f \odot g = \max(f + g - 1, 0),$$
$$\neg f = 1 - f,$$

$u = 1_\Omega, 0 = 0_\Omega$. Moreover, define $m : \mathcal{M} \to [0, 1]$ by the formula

$$m(f) = \int_\Omega f d\mu,$$

and

$$\tau(f) = f \circ T$$

where T is measure preserving map. By Theorem 3.3 we obtain that

$$\int_\Omega (f - \bigvee_{i=1}^{\infty} f \circ T^i) \vee 0) d\mu = 0,$$

hence μ-almost everywhere

$$(f - \bigvee_{i=1}^{\infty} f \circ T^i) \vee 0 = 0,$$

or equivalently

$$f \leq \bigvee_{i=1}^{\infty} f \circ T^i$$

μ-almost everywhere.

Acknowledgment

The paper was supported by grant VEGA 1/2002/05 and grant APVV LPP-0046-06.

References

[1] C.C. Chang. Algebraic analysis of many valued logics. *Trans. Amer. Math. Soc.* 88: 467 - 490, 1958.

[2] A. Dvurečenskij and S. Pulmannová. *New Trends in Quantum Structures*. Kluwer, Dordrecht, 2000.

[3] P. Maličký P. Category version of the Poincaré recurrence theorem. *Topology and its Applications* 154:2709 - 2713, 2007.

[4] D. Mundici. Interpretation of AFC^*-algebras in Łukasiewicz sentential calculus. *J. Funct. Anal.* 65: 15 - 63, 1986.

[5] M.G. Nadkarni. *Basic Ergodic Theory*. Birkhäuser Verlag, Basel, 1998.

[6] H. Poincaré. *Les méthodes nouvelles de la mécanique classique céléste*, vol. 3. Gauthiers - Villars, Paris, 1899.

[7] B. Riečan. A note on the Poincaré recurrence theorem on Boolean rings. (In Russian.) *Mat. - fyz. čas.* 15: 13 - 22, 1965.

[8] B. Riečan. Poincaré recurrence theorem in MV-algebras. *Math. Slovaca*, submitted.

ISBN: 978-989-95079-6-82

Notes on Q-probabilities on intuitionistic fuzzy events

Magdaléna Renčová
Faculty of Natural Sciences, Matej Bel University
Tajovského 40, 074 01 Banská Bystrica, Slovakia
E-mail: rencova@fpv.umb.sk

Abstract— *Following [9] some properties of Q-probability and Q-states are studied. Representation theorem of Q-probabilities and Q-states, the existence of the joint observable and The central limit theorem are proved.*

Keywords— Q-probability, Q-state, representation theorem, intuitionistic fuzzy events, joint Q-observable, Central limit theorem.

1 Introduction

Although there are different opinions about intuitionistic fuzzy events, the following definitions are accepted generally ([1], [5]). Let $(\Omega, \mathcal{S})$ be a measurable space. By an intuitionistic fuzzy event ([5]) we mean any pair

$$\mathbf{A} = (\mu_A, \nu_A)$$

of $\mathcal{S}-$measurable functions, such that $\mu_A, \nu_A : \Omega \to [0,1]$ and $\mu_A + \nu_A \leq 1$.

The function μ_A is the membership function and the function ν_A is the non-membership function. The family $\mathcal{F}$ of all intuitionistic fuzzy events is ordered in the following way:

$$\mathbf{A} \leq \mathbf{B} \Leftrightarrow \mu_A \leq \mu_B, \nu_A \geq \nu_B.$$

Evidently

$$\mathbf{A} \wedge \mathbf{B} = (\mu_A \wedge \mu_B, \nu_A \vee \nu_B),$$
$$\mathbf{A} \vee \mathbf{B} = (\mu_A \vee \mu_B, \nu_A \wedge \nu_B).$$

It is easy to see that $\mathbf{A_n} \nearrow \mathbf{A}$ if and only if $\mu_{A_n} \nearrow \mu_A$ and $\nu_{A_n} \searrow \nu_A$.

The notion of intuitionistic fuzzy event is a natural generalization of the notion of a fuzzy event. Given a fuzzy event μ_A, the pair $(\mu_A, 1-\mu_A)$ is an intuitionistic fuzzy event, so intuitionistic fuzzy events can be seen as generalizations of fuzzy events. Hence we want to define probability on intuitionistic fuzzy events generalizing probability on fuzzy events. And actually, two constructions were proposed independently by Gregorzewski [5] and Gerstenkorn [4], both based on the Łukasiewicz operations

$$a \oplus b = \min(a+b, 1),$$

$$a \odot b = \max(a+b-1, 0).$$

Operations $\oplus, \odot$ on $[0,1]^2$ (not necessarily Łukasiewicz operations) can be naturally extended to intuitionistic fuzzy events in the following way

$$\mathbf{A} \oplus \mathbf{B} = (\mu_A \oplus \mu_B, \nu_A \odot \nu_B),$$

$$\mathbf{A} \odot \mathbf{B} = (\mu_A \odot \mu_B, \nu_A \oplus \nu_B),$$

where $\mathbf{A} = (\mu_A, \nu_A)$ and $\mathbf{B} = (\mu_B, \nu_B)$.

If $\mu : \Omega \to [0,1]$ is a fuzzy set, then $(\mu, 1-\mu)$ is an IF set corresponding to this fuzzy set. Similarly as in the classical case, in the fuzzy case and in the quantum case, a probability (or state) has been introduced as a mapping $m : \mathcal{F} \to [0,1]$ being continuous, additive and satisfying some boundary conditions. Here the main difference is the additivity which is now of the following form

$$m(\mathbf{A}) + m(\mathbf{B}) = m(\mathbf{A} \oplus \mathbf{B}) + m(\mathbf{A} \odot \mathbf{B}).$$

There exists a general representation theorem for IF-probability. If $(\Omega, \mathcal{S}, P)$ is a probability space, then to any Łukasiewicz state $m : \mathcal{F} \to [0,1]$ there exists $\alpha \in [0,1]$ such that

$$m(\mathbf{A}) = (1-\alpha) \int_\Omega \mu_A dP + \alpha(1 - \int_\Omega \nu_A dP)$$

for any $\mathbf{A} \in \mathcal{F}$ (see [2]). Of course, the constructions (see [4] [5]) can be obtained as a very special case.

Generally, there are infinitely many possibilities how to define additivity

$$m(\mathbf{A}) + m(\mathbf{B}) = m(\mathcal{S}(\mathbf{A}, \mathbf{B})) + m(\mathcal{T}(\mathbf{A}, \mathbf{B}))$$

where

$$\mathcal{S}(\mathbf{A}, \mathbf{B}) = (S(\mu_A, \mu_B), T(\nu_A, \nu_B)),$$
$$\mathcal{T}(\mathbf{A}, \mathbf{B}) = (T(\mu_A, \mu_B), S(\nu_A, \nu_B))$$
$$S, T : [0,1]^2 \to [0,1]$$

being such binary operations (T is a t-norm and S is dual t-conorm [6]), that

$$S(u,v) + T(1-u, 1-v) \leq 1.$$

The Kolmogorov probability theory has 3 fundamental notions: probability, random variable and expectation. In our fuzzy case, an analogous situation occurs.

Throughout this paper we consider the following operations with intuitionistic fuzzy events

$$A \oplus_Q B = ((\mu_A{}^n + \mu_B{}^n)^{\frac{1}{n}} \wedge 1; 1 - ((1-\nu_A)^n + (1-\nu_B)^n)^{\frac{1}{n}} \wedge 1),$$

$$A \odot B = (\mu_A, \nu_A) \odot (\mu_B, \nu_B) = ((\mu_A + \mu_B - 1) \vee 0; (\nu_A + \nu_B) \wedge 1).$$

Remark 1.1 *The operation $\oplus_Q$ was introduced by Yager [6], the operation $\odot$ is Łukasiewicz operation. This is a special case of operations studied in [9], where $\varphi(u) = u^n$, $n \in N$ is fixed for each $u \in [0,1]$. Special case $n = 2$ is studied in [2].*

We are not able to embed the family $\mathcal{F}$ with these operations into an MV-algebra. Of course, we are able to prove probability representation theorems, to construct the joint observable and prove such fundamental theorems as central limit theorem or laws of large numbers.

ISBN: 978-989-95079-6-83

2 Q-probability and Q-observables

Definition 2.1 *Let $\mathcal{F}$ be the family of all intuitionistic fuzzy events, $\mathcal{J}$ be the family of all compact subintervals of the unit interval $[0,1]$. Q-probability is any mapping $\mathcal{P}: \mathcal{F} \to \mathcal{J}$ satisfying the following conditions:*

(i) $\mathcal{P}((\mathbf{1},\mathbf{0})) = [1,1], \mathcal{P}((\mathbf{0},\mathbf{1})) = [0,0]$;

(ii) $\mathbf{A} \odot \mathbf{B} = (0,1) \Rightarrow \mathcal{P}(\mathbf{A} \oplus_Q \mathbf{B}) = \mathcal{P}(\mathbf{A}) + \mathcal{P}(\mathbf{B})$;

(iii) $\mathbf{A_n} \nearrow \mathbf{A} \Rightarrow \mathcal{P}(\mathbf{A_n}) \nearrow \mathcal{P}(\mathbf{A})$.
(Here $[a_n, b_n] \nearrow [a,b]$, if $a_n \nearrow a, b_n \nearrow b$.)

Remark 2.2 *If A is a crisp set, then $\mu_A = \xi_A, \nu_A = \xi_B$, Theorem 2.7 implies that*

$$\mathcal{P}^{sharp}(A) = (1-\alpha)p(A) + \alpha r(A)$$

$$\mathcal{P}^{sharp}(A') = (1-\alpha)p(A') + \alpha r(A')$$

hence $\mathcal{P}^{sharp}(A) + \mathcal{P}^{sharp}(A') = (1-\alpha)p(\Omega) + \alpha r(\Omega) = 1$. It follows that there is a large class of examples extending the classical definition. It makes possible to construct different models describing some real processes.

Definition 2.3 *A mapping $m: \mathcal{F} \to [0,1]$ is called a Q-state, if the following conditions are satisfied:*

(i) m((1,0))=1, m((0,1))=0;

(ii) $\mathbf{A} \odot \mathbf{B} = (0,1) \Rightarrow m(\mathbf{A} \oplus_Q \mathbf{B}) = m(\mathbf{A}) + m(\mathbf{B})$;

(iii) $\mathbf{A_n} \nearrow \mathbf{A} \Rightarrow m(\mathbf{A_n}) \nearrow m(\mathbf{A})$.

Example 2.4 *Let $(\Omega, \mathcal{S}, p)$ be a probability space, then a natural example of Q-state is a function $m: \mathcal{F} \to [0,1]$ defined by the following*

$$m((\mu_A, \nu_A)) = \int_\Omega {\mu_A}^n dp,$$

where $n \in N$ is fixed natural number.

Let us suppose, that $\mathcal{P}$ maps $\mathcal{F}$ to $\mathcal{J}$. We will present this mapping with functions $\mathcal{P}^\flat, \mathcal{P}^\sharp: \mathcal{F} \to [0,1]$ in the following manner $\mathcal{P}(\mathbf{A}) = [\mathcal{P}^\flat(\mathbf{A}), \mathcal{P}^\sharp(\mathbf{A})], \mathbf{A} \in \mathcal{F}$. Shorter notation is used further on is $\mathcal{P} = [\mathcal{P}^\flat, \mathcal{P}^\sharp]$.

Theorem 2.5 *$\mathcal{P}: \mathcal{F} \to \mathcal{J}$, is a Q-probability if and only if $\mathcal{P}^\flat, \mathcal{P}^\sharp: \mathcal{F} \to [0,1]$ are Q-states.*

Proof Let us suppose that $\mathcal{P}$ is an Q- probability, then since

$$[1,1] = \mathcal{P}((\mathbf{1},\mathbf{0})) = [\mathcal{P}^\flat((\mathbf{1},\mathbf{0})), \mathcal{P}^\sharp((\mathbf{1},\mathbf{0}))],$$

we have $1 = \mathcal{P}^\flat((\mathbf{1},\mathbf{0}))$ and $1 = \mathcal{P}^\sharp((\mathbf{1},\mathbf{0}))$.
Further let $\mathbf{A} \odot \mathbf{B} = (\mathbf{0},\mathbf{1})$. Then

$$[\mathcal{P}^\flat(\mathbf{A}) + \mathcal{P}^\flat(\mathbf{B}), \mathcal{P}^\sharp(\mathbf{A}) + \mathcal{P}^\sharp(\mathbf{B})] = [\mathcal{P}^\flat(\mathbf{A}), \mathcal{P}^\sharp(\mathbf{A})] + [\mathcal{P}^\flat(\mathbf{B}), \mathcal{P}^\sharp(\mathbf{B})] = \mathcal{P}(\mathbf{A}) + \mathcal{P}(\mathbf{B}) = \mathcal{P}(\mathbf{A} \oplus_Q \mathbf{B}) = [\mathcal{P}^\flat(\mathbf{A} \oplus_Q \mathbf{B}), \mathcal{P}^\sharp(\mathbf{A} \oplus_Q \mathbf{B})],$$

hence

$$\mathcal{P}^\flat(\mathbf{A}) + \mathcal{P}^\flat(\mathbf{B}) = \mathcal{P}^\flat(\mathbf{A} \oplus_Q \mathbf{B})$$

and

$$\mathcal{P}^\sharp(\mathbf{A}) + \mathcal{P}^\sharp(\mathbf{B}) = \mathcal{P}^\sharp(\mathbf{A} \oplus_Q \mathbf{B}).$$

Finally

$$\mathbf{A_n} \nearrow \mathbf{A} \text{ implies } [\mathcal{P}^\flat(\mathbf{A_n}), \mathcal{P}^\sharp(\mathbf{A_n})] = \mathcal{P}(\mathbf{A_n}) \nearrow \mathcal{P}(\mathbf{A}),$$

hence

$$\mathcal{P}^\flat(\mathbf{A_n}) \nearrow \mathcal{P}^\flat(\mathbf{A}) \text{ and } \mathcal{P}^\sharp(\mathbf{A_n}) \nearrow \mathcal{P}^\sharp(\mathbf{A}).$$

The opposite implication can be proved similarly. □

Let us find the representation theorems for Q-states and Q-probabilities. We are able to find this representation only for representable Q-probabilities.

Definition 2.6 *Q-probability $\mathcal{P}_0 = [\mathcal{P}_0^\flat, \mathcal{P}_0^\sharp]: \mathcal{F} \to \mathcal{J}$ is representable, if there exist functions $f, g: R^2 \to R$ and probabilities $p, r: \mathcal{S} \to [0,1]$, such that*

$$\mathcal{P}_0((\mu_A, \nu_A)) =$$

$$= [f(\int_\Omega (\mu_A)^n dp, \int_\Omega (1-\nu_A)^n dr), g(\int_\Omega (\mu_A)^n dp, \int_\Omega (1-\nu_A)^n dr)].$$

Definition 2.7 *Q-state $m_0: \mathcal{F} \to [0,1]$ is representable, if there exist a function $f: R^2 \to R$ and the probabilities $p, r: \mathcal{S} \to [0,1]$ such that*

$$m_0((\mu_A, \nu_A)) = f\left(\int_\Omega (\mu_A)^n dp, \int_\Omega (1-\nu_A)^n dr\right).$$

Theorem 2.8 Representation Theorem
Let $m_0: \mathcal{F} \to [0,1]$ be a representable Q-state. Then there exist $\alpha \in [0,1]$ and probabilities $p, r: \mathcal{F} \to [0,1]$ such that for each $A \in \mathcal{F}$

$$m_0((\mu_A, \nu_A)) = (1-\alpha) \int_\Omega {\mu_A}^n dp + \alpha(\int_\Omega (1-\nu_A)^n dr),$$

where $n \in N$ is fixed.

Firstly, let us prove the following lemma.

Lemma 2.9 *Let $f: [0,1]^2 \to R$ be an additive and continuous function, then f is linear.*

Proof Let $f: [0,1]^2 \to R$, we show, that for each $A \in [0,1]^2$ and each $\alpha \in R$ such that

$$f(\alpha A) = \alpha f(A).$$

Consider cases:

(I) $\alpha \in N$, then $f(\alpha A) = f(\underbrace{A + ... + A}_{\alpha}) = \alpha f(A)$.

(II) $\alpha \in Q^+$, so $\exists p, q \in Z, (p,q) = 1, \frac{p}{q} > 0$.
We have $f(A) = \underbrace{f\left(\frac{1}{q}A\right) + ... + f\left(\frac{1}{q}A\right)}_{q}$,
where $A \in [0,1]^2$, so $\frac{1}{q}A \in [0,1]^2$.
Then $f(A) = qf\left(\frac{1}{q}A\right)$ and so $f\left(\frac{1}{q}A\right) = \frac{1}{q}f(A)$.
Let us take $A \in [0,1]^2$ such that $\frac{p}{q}A \in [0,1]^2$,
then $\frac{1}{q}A \in [0,1]^2$ and

ISBN: 978-989-95079-6-84

$$f(\tfrac{p}{q}A) = \underbrace{f\left(\frac{1}{q}A\right) + ... + f\left(\frac{1}{q}A\right)}_{p} =$$
$$= pf\left(\tfrac{1}{q}A\right) = \tfrac{p}{q}f(A).$$

(III) Since f is continuous and $f(rA) = rf(A)$ for $r \in Q^+$, such that $A, rA \in [0,1]^2$, by approximating any real $x \in R$ by rational $(x_n)_{n=1}^{\infty}$ we get

$$f(xA) = f(\lim_{n\to\infty} x_n A) = \lim_{n\to\infty} f(x_n A) =$$
$$= \lim_{n\to\infty} x_n f(A) = xf(A).$$

Thus we proved that f is linear on [0,1]. □

Proof of Theorem 2.7
Let m_0 be a representable Q-state, we are looking for a formula for function f from Definition 2.6. From property (i) of Q-state (Definition 2.2) we get

$$m_0((\mathbf{1},\mathbf{0})) = f(\int_\Omega 1^n dp, \int_\Omega (1-0)^n dr) = f(1,1),$$

because of that

$$\int_\Omega 1 dp = \int_\Omega 1 dr = 1$$

we get

$$m_0((\mathbf{1},\mathbf{0})) = f(1,1) = 1.$$

Analogously $m_0((\mathbf{0},\mathbf{1})) = f(0,0) = 0$.
What about additivity? Assume that $A \odot B = (\mathbf{0},\mathbf{1})$, so

$$\mu_A + \mu_B \leq 1, \nu_A + \nu_B \geq 1.$$

We get

$$m_0(A \oplus_Q B) =$$
$$= f(\int_\Omega (\sqrt[n]{((\mu_A)^n + (\mu_B)^n)})^n dp,$$
$$\int_\Omega (1 - (1 - \sqrt[n]{((1-\nu_A)^n + (1-\nu_B)^n)}))^n dr) =$$
$$= f(\int_\Omega (\mu_A)^n + (\mu_B)^n dp, \int_\Omega (1-\nu_A)^n + (1-\nu_B)^n dr).$$

Analogously

$$m_0(A) + m_0(B) =$$
$$f(\int_\Omega (\mu_A)^n dp, \int_\Omega (1-\nu_A)^n dr) + f(\int_\Omega (\mu_B)^n dp, \int_\Omega (1-\nu_B)^n dr).$$

Let us denote by

$$\int_\Omega \mu_A dp = u_1, \int_\Omega \mu_B dp = v_1,$$

$$\int_\Omega (1-\nu_A) dr = u_2, \int_\Omega (1-\nu_B) dr = v_2,$$

we get

$$m_0(A \oplus_Q B) = f(u_1 + v_1, u_2 + v_2),$$
$$m_0(A) + m_0(B) = f(u_1, u_2) + f(v_1, v_2).$$

Since the property (ii) holds, we have

$$f(u_1 + v_1, u_2 + v_2) = f(u_1, u_2) + f(v_1, v_2),$$

and that is why the function f is linear ($f(x+y) = f(x) + f(y)$ holds). This equality holds by the previous Lemma. Finaly f is continuous by the (iii) property of Definition 2.2. □

Theorem 2.10 Representation Theorem of Q-probabilities *If $\mathcal{P}_0$ is a representable Q-probability, then there exist real numbers $\alpha, \beta \in [0,1]$ and probability mesures $p, r_1, r_2 : \mathcal{S} \to [0,1]$ such that $\alpha r_1 \leq \beta r_2$ that for each $A = (\mu_A, \nu_A) \in \mathcal{F}$ there holds*

$$\mathcal{P}_0((\mu_A,\nu_A)) = [(1-\alpha)\int_\Omega (\mu_A)^n dp + \alpha \int_\Omega (1 -$$
$$\nu_A)^n dr_1, (1-\beta)\int_\Omega (\mu_A)^n dp + \beta \int_\Omega (1-\nu_A)^n dr_2].$$

Proof Let $\mathcal{P}_0(A) = [\mathcal{P}_0^\flat(A), \mathcal{P}_0^\sharp(A)]$ be representable Q-probability. Following Q-states $\mathcal{P}_0^\flat, \mathcal{P}_0^\sharp$ could be written by previous formulas. □

3 Q-observables and p-joint Q-observables

First, let us denote Borelian sets by $\mathcal{B}(R)$.

Definition 3.1 *A mapping $x : \mathcal{B}(R) \to \mathcal{F}$ is called a Q-observable, if the following conditions are satisfied:*

(i) $x(R) = (1,0), x(\emptyset) = (0,1)$;

(ii) *if* $A \cap B = \emptyset$ *then* $x(A) \odot x(B) = (\mathbf{0},\mathbf{1})$, *and* $x(A \cup B) = x(A) \oplus_Q x(B)$;

(iii) $A_n \nearrow A \Rightarrow x(A_n) \nearrow x(A)$.

Theorem 3.2 *Let $x : \mathcal{B}(R) \to \mathcal{F}$ be an Q-observable, $\mathcal{P} = [\mathcal{P}^\flat, \mathcal{P}^\sharp] : \mathcal{F} \to \mathcal{J}$ be an Q-probability. Then the functions $\mathcal{P}^\flat \circ x : \mathcal{B}(R) \to [0,1], \mathcal{P}^\sharp \circ x : \mathcal{B}(R) \to [0,1]$, are probability measures.*

Proof The proof is straightforward. □

Theorem 3.3 *Let $x : \mathcal{B}(R) \to \mathcal{F}$ be an Q-observable, $x(A) = (x^\flat(A), 1 - x^\sharp(A)); \omega \in \Omega$. Then the functions $p_\omega^\flat, p_\omega^\sharp : \mathcal{B}(R) \to [0,1]$ defined by*

$$p_\omega^\flat(A) = (x^\flat(A)(\omega))^n;$$

$$p_\omega^\sharp(A) = (x^\sharp(A)(\omega))^n$$

are probability measures.

Proof Use instead of $\varphi(u) = u^n$ in Theorem 2.7 in [9]. □

ISBN: 978-989-95079-6-85

Definition 3.4 *Let $x, y : \mathcal{B}(R) \to \mathcal{F}$ be Q-observables. By the p-joint Q-observable h of x, y we understand a mapping $h : \mathcal{B}(R^2) \to \mathcal{F}$ satisfying the following conditions*

(i) $h(R^2) = (\mathbf{1}, \mathbf{0}); h(\emptyset) = (\mathbf{0}, \mathbf{1})$;

(ii) *if* $A \cap B = \emptyset$ *then* $h(A) \odot h(B) = (\mathbf{0}, \mathbf{1})$ *and* $h(A \cup B) = h(A) \oplus_Q h(B)$;

(iii) $A_n \nearrow A \Rightarrow h(A_n) \nearrow h(A)$;

(iv) $h(C \times D) = x(C).y(D)$ *for any* $C, D \in \mathcal{B}(R)$. *here* $(\mu_C, \nu_C).(\mu_D, \nu_D) = (\mu_C.\mu_D, 1 - (1 - \nu_C).(1 - \nu_D))$

Remark 3.5 *Analogously we can extend Definition 3.4 for finite collection of Q-observables.*

Theorem 3.6 *To any Q-observables $x, y : \mathcal{B}(R) \to \mathcal{F}$ there exists their p-joint Q-observable $h : \mathcal{B}(R^2) \to \mathcal{F}$.*

Proof Use instead of $\varphi(u) = u^n$ in Theorem 2.9 in [9]. □

4 Application of Q-observables

Let us mention one version of Central limit theorem: let $(\xi_i)_{i=1}^{\infty}$ be a sequence of independent, equally distributed, square integrable random variables,

$$E(\xi_i) = a, \sigma^2(\xi_i) = \sigma^2 \text{ for all } i \in N.$$

Then for any $t \in R$ there holds

$$\lim_{n\to\infty} p\left(\left\{\omega; \frac{\overline{\zeta_n}(\omega) - a}{\sigma}\sqrt{n} < t\right\}\right) = \Phi(t).$$

Here $\overline{\zeta_n} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$ and $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du$.
Now we are going to formulate an analogous assertion for Q-observables.
First, we shall mention some useful definitions:

Definition 4.1 *For any Q-probability $\mathcal{P} = [\mathcal{P}^\flat, \mathcal{P}^\sharp] : \mathcal{F} \to \mathcal{J}$ and any Q-observable $x : \mathcal{B}(R) \to \mathcal{F}$ we define the expected values by*

$$E_\flat(x) = \int_R t d\mathcal{P}_x^\flat(t);$$

$$E_\sharp(x) = \int_R t d\mathcal{P}_x^\sharp(t)$$

and the variances by

$$\sigma_\flat^2(x) = \int_R (t - E_\flat(x))^2 d\mathcal{P}_x^\flat(t);$$

$$\sigma_\sharp^2(x) = \int_R (t - E_\sharp(x))^2 d\mathcal{P}_x^\flat(t),$$

where $\mathcal{P}_x^\flat = \mathcal{P}^\flat \circ x, \mathcal{P}_x^\sharp = \mathcal{P}^\sharp \circ x$, assuming that the integrals exist.

Assume $T = (\xi_1, ..., \xi_n) : \Omega^n \to R^n$ is a random vector and $g : R^n \to R$ is a Borel measurable function (e.g. $g(u_1, ..., u_n) = \frac{1}{n}\sum_{i=1}^{n} u_i$). Then

$$g(\xi_1, ..., \xi_n) = g \circ T : \Omega^n \to R^n,$$

is a transformation of T. Hence we get the following formula

$$(g \circ T)^{-1}(A) = T^{-1}(g^{-1}(A))$$

for any $A \in \mathcal{B}(R)$. The formula justifies the following definition.

Definition 4.2 *Let $g_n : R^n \to R$ be a Borel function, $x_1, ..., x_n : \mathcal{B}(R) \to \mathcal{F}$ be Q-observables, $h_n : \mathcal{B}(R^n) \to \mathcal{F}$ their joint observable. Then the g_n-transformation of h_n is a Q-observable $y_n : \mathcal{B}(R) \to \mathcal{F}$ given by $y_n(A) = h_n(g_n^{-1}(A))$ for any $A \in \mathcal{B}(R)$.*

Definition 4.3 *Let $(x_n)_{n=1}^{\infty}$ be a sequence of Q-observables, $(h_n)_{n=1}^{\infty}$ be a sequence of the joint Q-observables $h_n : \mathcal{B}(R^n) \to \mathcal{F}$ of $x_1, x_2, ..., x_n$ (for $n \in N$), $m : \mathcal{F} \to [0, 1]$ be a Q-state. The sequence $(x_n)_{n=1}^{\infty}$ is independent (with respect to m), if for any $n \in N$ and any $C_1, C_2, ..., C_n \in \mathcal{B}(R)$ there holds*

$$m(h_n(C_1 \times C_2 \times ... \times C_n)) = m(x_1(C_1))...m(x_n(C_n)).$$

Definition 4.4 *A sequence $(x_n)_{n=1}^{\infty}$ of Q-observables is equally distributed, if $m(x_n(A)) = m(x_1(A))$ for any $n \in N$ and $A \in \mathcal{B}(R)$.*

Theorem 4.5 (Central limit theorem)
Let $(x_n)_{n=1}^{\infty}$ be a sequence of independent, equally distributed, square integrable Q-observables, where $E_\flat(x_n) = a^\flat, (E_\sharp(x_n) = a^\sharp)$ $\sigma_\flat^2(x_n) = \sigma_\flat^2$, $(\sigma_\sharp^2(x_n) = \sigma_\sharp^2)$ for each $n \in N$. Then for any $t \in R$ there the following holds

$$\lim_{n\to\infty} \mathcal{P}^\flat\left(\tfrac{x_1+...+x_n-na^\flat}{\sigma_\flat\sqrt{n}}((-\infty, t))\right) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du,$$

$$(\lim_{n\to\infty} \mathcal{P}^\sharp\left(\tfrac{x_1+...+x_n-na^\sharp}{\sigma_\sharp\sqrt{n}}((-\infty, t))\right) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du).$$

Proof Use Theorem 4.1 in [9]. □

5 Conclusion

Generalizing some notions proposed in [2] we constructed a $Q-$probability theory. The theory includes some known results ($n = 1, n = 2$), and also it opens the door for some other applications. We have proved some representation theorems. As an open question and an inspiration for the future research remains the problem of conditional probabilities for this framework.

Acknowledgement

The paper was supported by grant VEGA 1/0539/08.

ISBN: 978-989-95079-6-86

References

[1] Atanassov, K.: *Intuitionistic Fuzzy Sets: Theory and Applications.* Physica Verlag, New York, 1999.

[2] Atanassov, K., Riečan, B.: *On two new types of probability on IF-events.*In: Proc. of the First International Workshop on Intuitionistic Fuzzy Sets, Generalized Nets and Knowledge Engineering, Warsawa, 2008, pp. 11-22.

[3] Čunderlíková - Lendelová, K., Riečan, B.: *The probability theory on B-structures.* In: Proc.of the First International Workshop on Intuitionistic Fuzzy Sets, Generalized Nets and Knowledge Engineering, Warsawa, 2008, pp. 33-60.

[4] Gerstenkorn, T., Manko, J.: *Probabilities of intuitionistic fuzzy events.* In: Issues in Inteligent Systems: Paradigms (O.Hrzniewicz et al. eds.). EXIT, Warsawa, 63-58.

[5] Grzegorzewski, P., Mrowka, E.: *Probability of intuitionistic fuzzy events.* In: Soft Methods in Probablity, Statistics and Data Analysis. (P.Grzegorzewski et al. eds.). Physica Verlag, New York 2002, 105-115.

[6] Klement, E.P., Mesiar, R., Pap, E.:*Triangular Norms.* Trends in Logic, Vol. 8, Kluwer Academic Publishers, Boston/London/Dordrecht, 2000, 406 pp.

[7] Lendelová, K.: *Strong law of large numbers for IF-events.* Proc. Eleventh Int. Conf. IPMU, Paris 2006, 2363-2366.

[8] Renčová, M., Riečan, B.: *Probability on IF - sets: an elementary approach.* In: First International Workshop on Intuitionistic Fuzzy Sets, Generalized Nets and Knowledge Engineering. London : University of Westminster, 8 - 17, 2006.

[9] Renčová, M.: On the $\varphi-$probability and $\varphi-$observables. Submitted to Fuzzy Sets and Systems. 2008.

[10] Riečan, B.: *General form of M-probabilities on IF-events.* In: Proc. of IPMU, Spain, 1675-1677, 2008.

[11] Riečan, B.: *On a problem of Radko Mesiar: general form of IF probabilities.* Fuzzy Sets and Systems 152, 2006, 1485 - 1490.

[12] Riečan, B.: *Probability theory on intuitionistic fuzzy sets.* In: A volume in honor of Daniele Mundici´s 60^{th} birthday. Lecture Notes in Computer Science, 2007.

ISBN: 978-989-95079-6-87

A new bridge between fuzzy sets and statistics

Pedro Terán

Departamento de Estadística e I.O. y D.M.
Universidad de Oviedo
E-33071 Gijón (Spain)
Email: teranpedro@uniovi.es

Abstract— *Our basic question is as follows: given a, maybe multivariate, probability distribution, how to formalize the idea that some points are more central than others? We present an account of the notion of centrality which is based on fuzzy events and is valid for single distributions and for families of distributions (including imprecise probability models).*

This unifying framework is natural and subsumes many known concepts like the following (not an exhaustive list): (a) univariate location estimators like the mean, the median and the mode, (b) the interquartile interval and the Lorenz curve of a random variable, (c) several generalized medians, trimmed regions and statistical depth functions from multivariate analysis, (d) most known location estimators for random sets, (e) the probability mass function of a discrete random variable and the coverage function of a random closed set, (f) the Choquet integral with respect to an infinitely alternating or infinitely monotone capacity.

Keywords— Centrality, Fuzzy event, Multivariate analysis, Random set, Statistical depth.

1 Introduction

All points are central, but some points are more central than others.[1]

Our aim is to show how a simple idea with a natural interpretation within fuzzy set theory allows to define, in a unified way, many notions from statistics. These include well-known location estimators for random variables; some multivariate generalizations which grew into the theory of statistical depth functions in the last fifteen years; trimming methods based on them; means and medians of random sets; and even Choquet integrals. In fact, the notion applies as soon as we have a model to which we are able to associate a family of probability distributions.

The main tool is the notion of a fuzzy event and its probability [25]. A *fuzzy event* in a measurable space Ω is a measurable mapping $A : \Omega \to [0, 1]$. The membership $A(\omega)$ denotes the degree to which event A occurs when $\omega \in \Omega$ occurs. A probability measure on Ω extends to all fuzzy events by the natural formula

$$P(A) = \int A \mathrm{d}P.$$

The structure of this communication is as follows. Section 2 presents the fuzzy notion of centrality of a point in a probability distribution. Section 3 extends it to families of distributions. Sections 4, 5, 6 and 7 exhibit specific examples covered by our framework. Finally, Section 8 closes the paper with some final remarks.

[1] Paraphrasing a famous sentence from Orwell's 'Animal farm'.

2 Centrality and location estimators based on fuzzy events

What does it mean to speak about the 'centrality' of a point in a data set or a probability distribution? Although it is informally clear that the mean, the median and the mode are three traditional answers to the problem of pinpointing where the 'center' of a distribution is, a formal definition of centrality is far more elusive. In this section, an intuitively appealing view of centrality based on fuzzy set theory is presented.

We begin by choosing a family $\mathcal{A}$ of fuzzy events, called *reference events*. Roughly speaking, a point x is considered *central* if the probability of a reference event is high whenever it contains x. That is, *x is central if the probable reference events are implied by the occurrence of x*, whereas x is not central if it belongs to unlikely reference events. Note that this definition of centrality is relative to the chosen family of reference events.

This notion is reminiscent of the mode as the most likely point, and actually the mode is what appears when the reference events are chosen to be the (crisp) points.

It is possible to derive a crisp definition of centrality from the ideas above: we could say that a point is central if

$$P(A) \geq A(x) \quad \forall A \in \mathcal{A}.$$

But although this might capture the notion that the value of $P(A)$ is large insofar as the degree to which x belongs to A is large, a gradual restriction seems even more faithful to the idea. Thus, we say that a point is *central to degree at least* α (with respect to the family $\mathcal{A}$) if

$$P(A) \geq \alpha \cdot A(x) \quad \forall A \in \mathcal{A}.$$

The set of all the central points to degree at least α is the α-cut of a fuzzy set. That allows us to define the fuzzy set C given by

$$C(x) = \sup\{\alpha \in (0, 1] \mid \forall A \in \mathcal{A}, P(A) \geq \alpha A(x)\},$$

whose membership function indicates the degree of centrality of each point in the probability distribution P.

This definition can be rewritten in the language of fuzzy logic, as follows.

Proposition 1. *Denote Goguen's fuzzy implication by I. Then,*

$$C(x) = \inf_{A \in \mathcal{A}} I(A(x), P(A)).$$

Since $P(A)$ can be seen as the degree of truth of the proposition 'A is probable' in an adequate fuzzy logic (see e.g. [7]), the fuzzy set C captures the intuitive idea that its elements are those such that *'for any reference event A, if $x \in A$ then A is probable'*.

In principle, the Goguen implication can be replaced by a different implication. In this paper, we content ourselves with showing that many statistical objects arise from this choice (whose reason is historical, as explained in the final section), without analyzing other possible choices. However, note that changing the implication may lead to obtaining the same examples with a mere modification of the shape of the reference events.

From a statistical point of view, those points with the maximal centrality can serve as location estimators which might be called *maximal centrality estimators*. In general, there may be more than one maximally central point (like e.g. several modes), so all the maximally central points form the *central region*. Points with centrality at least α form the *α-central region*.

3 Extension to families of distributions

As shown in Section 4, the notion of centrality presented above is directly related to some traditional location estimators, as well as to several multivariate generalizations which have been introduced and studied mostly in the last fifteen years. However, to account for some of those cases and specially to cover more general types of models like random sets, it is necessary to extend it to study the centrality of a point in a *family* of distributions.

When the model consists of more than one single distribution, the underlying distribution generating the data is known or assumed to belong to a specified family. Some situations covered by this more general setting are the following.

(a) A parametric or non-parametric model.

(b) Upper and lower probabilities.

(c) A Choquet capacity.

(d) A random set.

(e) A credal set.

(f) A neighbourhood of a probability.

(g) Other families defined in terms of the distribution of a random variable.

Given a random variable ξ, instances of (g) include the family $\{P_{g \cdot \xi} \mid g \text{ is } [0,1]\text{-valued}\}$ used to define the zonoid [11] of ξ; and the family $\{Q \leq \alpha^{-1} P_\xi\}$ used to define trimmed regions in [4]). In such examples, a family of distributions is constructed whose central region contains interesting information about ξ, although maybe no longer related to centrality itself. For instance, the lift zonoid of ξ, a multivariate generalization of the Lorenz curve, is the central region of a family of distributions defined from the distribution of ξ.

In such a that situation, for statistical purposes a point should be considered central in the family if it is potentially a central point of the true distribution, which could be any member of the family. The rationale is that a model defined by a family of distributions is more imprecise than pinpointing a single distribution, so the compatibility of a point being central with the information available should be larger [2].

That suggests the following definition: given a family $\mathcal{P}$ of probability distributions, the fuzzy set C of central points is defined by

$$C(x) = \sup\{\alpha \in (0,1] \mid \forall A \in \mathcal{A}, \sup_{P \in \mathcal{P}} P(A) \geq \alpha A(x)\}.$$

The rest of the definitions (α-central region, etc.) are modified analogously.

Observe that the location information in a family of distributions is summarized using the same object as for a single distribution. Thus, the complexity of that summary is the same as for a single distribution (although calculating C may be computationally more expensive). That implies that statistical procedures based on C, devised for single distributions, will extend immediately to the situations above, particularly imprecise probability models (Choquet capacities, random sets).

As will be shown later, in the cases when C coincides with a statistical depth function such procedures already exist.

Let us finally note that some of the results below may require, for technical reasons, that the family $\mathcal{P}$ considered is not an arbitrary one but compact with respect to the topology of weak convergence (convergence in distribution).

4 Univariate examples

As already mentioned, the three basic examples of location estimators arise as special cases of maximally central points of a distribution.

Theorem 2. *Let ξ be a random variable. For appropriate choices of $\mathcal{A}$ and $\alpha \in (0,1]$, the α-central region C_α is:*

(1) The expectation of ξ, if it exists (otherwise, the corresponding α-central region is empty).

(2) The median of ξ.

(3) The mode of ξ, if ξ is discrete.

The necessary choices of α in Theorem 2 are: $\alpha = 1$ in (1); $\alpha = .5$ in (2); α the height of C in (3). In fact, in all the three cases α coincides with the height of the fuzzy set C. In subsequent examples, for ease of presentation we omit the choice of α in each case.

[2] In any case, it is also interesting to consider the dual approach: to which degree a point is guaranteedly central in *all* the distributions of the model. That was pointed out to me by Didier Dubois.

As to the other ingredient, $\mathcal{A}$, one may ask whether the families of reference events needed to retrieve these classical estimators may be very complicated. In fact, they are quite simple:

(1) $\{A_y(x) = |x-y|/(1+|x-y|)\}_{y\in\mathbf{R}}$ for the expectation.

(2) $\{A_y^+(x) = I_{[y,\infty)}(x), A_y^-(x) = I_{(-\infty,y]}(x)\}_{y\in\mathbf{R}}$ for the median.

(3) $\{A_y(x) = I_{\{y\}}(x)\}_{y\in\mathbf{R}}$ for the mode.

In (2) and (3), I_B denotes the indicator function of a crisp set B.

Subsequent examples use the same choices of $\mathcal{A}$ or suitable multivariate generalizations. Again to avoid burdensome presentation, we omit the specific choice of $\mathcal{A}$ in each case.

Interestingly, (1) is associated to a new characterization of the expectation of a random variable in terms of the bounded metric $d(x, y) = |x - y|/(1 + |x - y|)$.

Proposition 3. *Let ξ be an integrable random variable. Then, its expectation is the unique real number x such that, for any number y,*

$$d(x, y) \leq Ed(\xi, y);$$

namely $E\xi$ is identified by the property that every number is closer in d-distance to $E\xi$ than it is (in mean) to ξ.

Proof (outline of the main ideas). There exists a characterization of the mean, due to Shafik Doss [6], as the unique real number x such that, for any y,

$$\|x - y\| \leq E\|\xi - y\|.$$

It suffices to show that both properties are equivalent. The easy part, with the help of the Jensen inequality, is to show that a number satisfying the condition in the statement must also fulfil Doss's property (and so our condition is formally stronger and only the mean could satisfy it).

To prove that the mean has indeed the property in the statement we need to go against Jensen's inequality, so things get more involved. The key is to take a divergent sequence of values for y and show that, when y is very far from x, Jensen's inequality is almost an equality, so in the limit we can 'reverse' it. □

Actually, this geometric characterization is valid in any separable Banach space, if we just replace the Euclidean distance by the norm.

A couple more univariate examples are as follows. These are no longer related to location estimation.

Proposition 4. *Let ξ be a random variable. For appropriate choices of $\mathcal{A}$ and $\alpha \in (0, 1]$, the α-central region C_α is:*

(1) The interquartile interval of ξ.

(2) The Lorenz curve of ξ (more exactly, the curve together with the region below it, its hypograph).

Both the interquartile interval and the median are examples of the interval between the p-th quantile and the $(1 - p)$-th quantile (for $p = .25$ and $p = .5$). Modulating the value of α (precisely, $\alpha = .5 - p$), we can obtain any such interval for $0 \leq p \leq .5$. For $p = 0$, that is the interval between the minimum and maximum values of ξ.

In this connection, let us remark that two known tools in exploratory data analysis, the box-plot and its multivariate generalization (the bag-plot), are plotted using only information subsumed by our framework.

5 Multivariate examples

In the multivariate case, the absence of a natural order makes it impossible for a point to have all the usual properties of the univariate median. That has led to the definition of many *depth functions* [14]; often the point maximizing the depth function is called a *generalized median* or a *maximal depth estimator*.

Depth functions aim at ordering, in a center-outward sense, the points in a data set or, more generally, those of $\mathbf{R}^d$ with respect to a given distribution. A typical application of depth (among many others) is outlier detection: sample points with very small depth are trimmed out as outliers.

The notion of depth is still somewhat vague, and its semantics is specially unclear for multimodal or very asymmetric distributions. Zuo and Serfling [26] underlined four desirable properties of depth functions, which had already been studied in particular cases:

(a) Depth is maximal at a center of symmetry of the distribution, if the latter exists.

(b) Depth decreases along any ray departing from the center.

(c) Depth vanishes as the distance to the center goes to infinity.

(d) The depth function should be affinely equivariant (so that conclusions do not depend on the chosen coordinate system).

It must be emphasized that, often, accepted notions of statistical depth fail some of those properties (e.g. L^p-depth).

From any depth function, an *α-trimmed region* can be constructed which is formed by those points having depth at least α. Plotting the contours of the α-trimmed regions constitutes an easy device for multivariate exploratory analysis.

To give the reader a grasp of how depth functions work, let us just mention some ways to calculate depth.

Convex hull peeling depth of a point in a data sample is calculated as follows. Take the convex hull of the sample. The points outside have depth 0. Now the sample points in the boundary of the convex hull are 'peeled off'; points left outside the cloud by doing so have depth

proportional to 1. The sample points in the boundary of the convex hull of the remaining points are peeled off, and points left outside the cloud are given depth proportional to 2. The procedure goes on until the innermost points have been peeled off.

Simplicial depth of a point is calculated as the probability that it belongs to the random simplex whose vertices are independent and identically distributed copies of the variable.

Similarly, expected convex hull depth is inversely proportional to the number k of copies of the variable which are needed to ensure that the point is in the Aumann expectation of the sample $\{\xi_1, \cdots, \xi_k\}$.

Finally, for halfspace depth one considers the probability that the variable lies in a halfspace, and calculates the infimum of such values over those halfspaces whose boundary contains the point.

Some statistical depth functions in the literature are fuzzy sets of central points, as the following result shows.

Theorem 5. *Let ξ be a random vector. For appropriate choices of $\mathcal{A}$, $\mathcal{P}$ and $\alpha \in (0, 1]$, the centrality $C(x)$ is:*

(1) The convex hull peeling depth of x (Huber [9], 1972; Barnett [2], 1976).

(2) The simplicial depth of x (Liu [15], 1990).

(3) The majority depth of x (Liu–Singh [16], 1993).

(4) The probability of the event $\{\xi = x\}$, if ξ is discrete.

The univariate examples presented above extend to the multivariate setting as follows (including the central regions associated to further notions of depth).

Theorem 6. *Let ξ be a multivariate random vector. For appropriate choices of $\mathcal{A}$, $\mathcal{P}$ and $\alpha \in (0, 1]$, the α-central region C_α is:*

(1) The expectation of ξ, if it exists.

(2) The generalized median of ξ according to any of the depth functions (1-3) above.

(3) The α-trimmed region of ξ according to any of the depth functions (1-3) above.

(4) The generalized median and the α-trimmed region of ξ according to the halfspace depth (Tukey [23], 1975).

(5) The generalized median and the α-trimmed region of ξ according to the zonoid depth (Mosler [17], 2002).

(6) The generalized median and the α-trimmed region of ξ according to the symmetric order depth (Cascos–López-Díaz [5], 2004).

(7) The generalized median and the α-trimmed region of ξ according to the expected convex hull depth (Cascos [3], 2007).

(8) The mode of ξ, if ξ is discrete.

(9) The zonoid of ξ (see e.g. [11, 17]).

(10) The lift zonoid of ξ (Koshevoy–Mosler [12], 1998).

The lift zonoid generalizes the Lorenz curve and so serves to study multivariate concentration and inequality. It is also remarkable that it characterizes the distribution of ξ.

6 Multivalued examples

When a random set is considered, the statistical model takes the following form. The underlying random variable ξ is not directly observed; only a larger set of values X containing ξ is observed. Then, we know that ξ is such that $\xi \in X$ almost surely, namely ξ is a *selection* of X. In that case, $\mathcal{P}$ is the family of all selections of X.

Most location estimators for random sets are unified as special cases of central or α-central regions.

Theorem 7. *Let X be a random compact set and $\mathcal{P}$ the family of all its selections. Then, for appropriate choices of $\mathcal{A}$ and $\alpha \in (0, 1]$, the α-central region C_α is:*

(1) The Aumann expectation of X (Kudō [13], 1954; Aumann [1], 1965) if Ω is non-atomic or X is convex (in general, the convex hull of the Aumann expectation).

(2) The Vorob'ev expectation of X (Vorob'ev [24], 1984).

(3) The Herer expectation of X (Herer [8], 1990).

(4) The radius-vector expectation of X (Stoyan–Stoyan [18], 1994).

(5) The Vorob'ev median of X (Stoyan–Stoyan [18], 1994).

Another notion covered by our framework is the *coverage function* of a random set X, given by $p_X(x) = P(x \in X)$.

Proposition 8. *Let X be a random closed set. Then, p_X is both an integral depth function in the sense of [22] and a fuzzy set of central points, corresponding to taking $\mathcal{A}$ to be the family of all crisp points and $\mathcal{P}$ the family of all selections of X.*

The corresponding central region is formed by all the *fixed points* of the random set.

7 Choquet capacities

Let ν be a *Choquet capacity* in the sense of [10], namely a function from the Borel σ-algebra of $\mathbf{R}^d$ to $[0, 1]$ with the following properties:

a) $\nu(\emptyset) = 0$, $\nu(\Omega) = 1$, $\nu(A) \leq \nu(B)$ if $A \subset B$,

b) $\nu(C_n) \searrow \nu(C)$ if $C_n \searrow C$ and C_n, C are closed,

c) $\nu(A_n) \nearrow \nu(A)$ if $A_n \nearrow A$.

A more general definition, in which closed sets are replaced by compact sets, is possible. A capacity is called *2-alternating* if

$$\nu(A \cup B) + \nu(A \cap B) \leq \nu(A) + \nu(B),$$

and *infinitely alternating* if

$$\nu(\bigcap_{i=1}^{n} A_i) \leq \sum_{\emptyset \neq I \subset \{1,\ldots,n\}} (-1)^{\mathrm{card}(I)+1} \nu(\bigcup_{i \in I} A_i)$$

whenever $A_1, \ldots, A_n$ are Borel sets.

The dual capacity to ν is given by $\overline{\nu}(A) = 1 - \nu(A^c)$. The dual to a 2-alternating or infinitely alternating capacity has the property called 2-monotonicity or infinite monotonicity, respectively. The *Choquet integral* of a random variable ξ with respect to ν is

$$\int_{(C)} \xi \mathrm{d}\nu = \int_0^\infty \nu(\{\xi \geq t\})\mathrm{d}t + \int_{-\infty}^0 [\nu(\{\xi \geq t\}) - 1]\mathrm{d}t,$$

which exists provided at least one of the two improper Riemann integrals is finite. In that case, ξ is called ν-integrable.

Our notion of centrality can be applied to capacities by taking $\mathcal{P}$ to be the core of ν, namely the family of all probability distributions $P \leq \nu$ dominated by ν. Observe that Zadeh's definition of the probability of a fuzzy event extends immediately to capacities by setting $\nu(A) = \int_{(C)} A \mathrm{d}\nu$.

Proposition 9. *Let ν be a 2-alternating Choquet capacity and let I be Goguen's fuzzy implication. Then, taking $\mathcal{P}$ to be the core of ν results in*

$$C(x) = \inf_{A \in \mathcal{A}} I(A(x), \nu(A)).$$

The Choquet integrals with respect to ν and $\overline{\nu}$ are sometimes called the upper and lower Choquet integrals. They can be retrieved in our framework.

Theorem 10. *Let ν be an infinitely alternating Choquet capacity and $\mathcal{P}$ its core. Let ξ be a ν-integrable random variable. Then, for an appropriate choice of $\mathcal{A}$, the central region C_1 is the interval between $\int_{(C)} \xi \mathrm{d}\overline{\nu}$ and $\int_{(C)} \xi \mathrm{d}\nu$.*

As a consequence, we obtain the following geometric characterizations of the Choquet integral.

Proposition 11. *Let ν be a Choquet capacity, and let ξ be a ν-integrable random variable. If ν is infinitely alternating, then the Choquet integral $\int_{(C)} \xi \mathrm{d}\nu$ is the largest value x satisfying either of the following equivalent properties:*

(1) $|a - x| \leq \int_{(C)} |a - \xi| \mathrm{d}\nu$ *for all* $a \in \mathbf{R}$,

(2) $d(a, x) \leq \int_{(C)} d(a, x) \mathrm{d}\nu$ *for all* $a \in \mathbf{R}$.

Analogously, if ν is infinitely monotone, then $\int_{(C)} \xi \mathrm{d}\nu$ is the smallest value x satisfying either property.

8 Concluding remarks

1. The proofs of the results presented in this contribution use pieces of information from several submitted papers, mostly [22, 20] but also [19, 21], which are not explicitly about fuzzy sets. The proofs will be presented in a forthcoming paper. The communication format allows us to present the ideas in a convenient sequence from particular to general, although the logical path of their derivation is more involved.

2. There is an interesting connection between defuzzification and depth-based statistical estimation. Indeed, defuzzification (understood as obtaining a single point, not a crisp set) of a fuzzy set of central points appears very close to location estimation. The two main methods of depth-based estimation in the literature are generalized medians and the depth-weighted estimators given by

$$DWE = \int xD(x)\mathrm{d}x / \int D(x)\mathrm{d}x,$$

see e.g. [27]), which correspond to the maximum and centroid methods of defuzzification.

3. The choice of the Goguen implication, rather than any other fuzzy implication, is rooted in earlier works which inspired our own. The genealogy is as follows. Koshevoy and Mosler [11] defined the *zonoid trimmed regions*

$$\{E[\xi \cdot g(\xi)] \mid g : \mathbf{R}^d \to [0, \alpha^{-1}] \text{ is measurable}\}$$

formed by the expectations of ξ with respect to all probability measures Q such that $\frac{\mathrm{d}Q}{\mathrm{d}P} \leq \alpha^{-1}$. Cascos generalized that idea to an abstract definition of *integral trimmed regions*

$$D^\alpha_{\mathcal{F}}(x) = \{x \mid \exists Q \leq \alpha^{-1} P \ : \ \forall f \in \mathcal{F} \ f(x) \leq \int f \mathrm{d}Q\}$$

where $\mathcal{F}$ is a family of integrable functions) which were studied in his Ph.D. thesis and the paper [4]. One already finds $P \geq \alpha Q$ explicitly there, although as a comparison between probabilities. Integral trimmed regions in the sense of [4] receive a natural interpretation in connection to stochastic orders. They coincide with integral trimmed regions in the sense of [22] (no longer interpretable from stochastic orders) if $\mathcal{F}$ is formed by upper semicontinuous functions. In turn, those are a special case of integral central regions [22], which are very similar to central regions in this paper (although, the notions of integral depth in [22] and fuzzy centrality in this paper seem to be more far apart).

4. It is natural to ask whether fuzzy centrality and/or statistical depth are in turn related to Zadeh's proposal that probability theory should include a notion of the *usual value* of a variable.

Acknowledgment

Research has been partially carried out while the author was with the *Universidad de Zaragoza*, and partially funded by Spain's *Ministerio de Educación y Ciencia* (MTM2005-02254 and TSI2005-02511), *Ministerio de Ciencia e Innovación* (MTM2008-01519) and Aragón's *Gobierno* (PM2007-034).

References

[1] R. J. Aumann (1965). Integrals of set-valued functions. *J. Math. Anal. Appl.* 12:1–12.

[2] V. Barnett (1976). The ordering of multivariate data. *J. Royal Statist. Soc. A* 139:318–352 (with discussion).

[3] I. Cascos (2007). The expected convex hull trimmed regions of a sample. *Comput. Statist.* 22:557–569.

[4] I. Cascos, M. López-Díaz (2005). Integral trimmed regions. *J. Multivar. Anal.* 96:404–424.

[5] I. Cascos-Fernández, M. López-Díaz (2004). The zonoid hull of a multivariate probability distribution and its trimmed regions. In: M. López-Díaz, M. Á. Gil, P. Grzegorzewski, O. Hryniewicz, J. Lawry, editors. *Soft methodology and random information systems*, 423–430. Springer, Berlin.

[6] S. Doss (1949). Sur la moyenne d'un élément aléatoire dans un espace distancié [*On the mean of a random element in a metric space*, in French]. *Bull. Sci. Math.* 73:1–26.

[7] P. Hájek, L. Godo, F. Esteva (1995). Fuzzy logic and probability. *Proc. 11th Annual Conf. Artif. Intelligence*, 237–244.

[8] W. Herer (1990). Mathematical expectation and martingales of random subsets of a metric space. *Probab. Math. Statist.* 11:291–304.

[9] P. J. Huber (1972). Robust statistics: a review. *Ann. Math. Statist.* 43:1041–1067.

[10] P. J. Huber, V. Strassen (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.* 1:251–263.

[11] G. Koshevoy, K. Mosler (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.* 25:1998–2017.

[12] G. Koshevoy, K. Mosler (1998). Lift zonoids, random convex hulls and the variability of random vectors. *Bernoulli* 4:377–399.

[13] H. Kudō (1954). Dependent experiments and sufficient statistics. *Nat. Sci. Rep. Ochanomizu Univ.* 4:151–163.

[14] R. Liu, R. Serfling, D. Souvaine, editors (2006). *Data depth: robust multivariate analysis, computational geometry and applications*. American Mathematical Society, Providence.

[15] R. Y. Liu (1990). On a notion of data depth based on random simplices. *Ann. Statist.* 18:405–414.

[16] R. Y. Liu, K. Singh (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* 88:252–260.

[17] K. Mosler (2002). *Multivariate dispersion, central regions and depth. The lift zonoid approach*. Lect. Notes Statist. 165. Springer, Berlin.

[18] D. Stoyan, H. Stoyan (1994). *Fractals, random shapes and point fields*. Wiley, Chichester.

[19] P. Terán. On the equivalence of Aumann and Herer expectations of random sets. *Test* 17:505–514.

[20] P. Terán. Classical location estimators for random sets in the light of statistical depth. *Submitted for publication*.

[21] P. Terán. Herer and Aumann expectations of random sets. *Submitted for publication*.

[22] P. Terán. Integral central regions and integral depth. *Submitted for publication*.

[23] J. W. Tukey (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, 523–531. Canad. Math. Congress, Montreal.

[24] O. Y. Vorob'ev (1984). *Mean-measure modelling* [in Russian]. Nauka, Moscow.

[25] L. A. Zadeh (1968). Probability measures of fuzzy events. *J. Math. Anal. Appl.* 23:421–427.

[26] Y. Zuo, R. Serfling (2000). General notions of statistical depth function. *Ann. Statist.* 28:461–482.

[27] Y. Zuo, H. Cui, X. He (2004). On the Stahel-Donoho estimator and depth-weighted means of multivariate data. *Ann. Statist.* 32:167–188.

L-Prime Spectrum of Modules

Bayram Ali Ersoy[1] and Dan Ralescu[2]
1 Yildiz Technical University, Department of mathematics, Davutpasa Kampusu, Esenler, Istanbul, TURKEY.
2 University of Cincinnati, Cincinnati, OH 45221-0025, USA.
E-mail : ersoya@yildiz.edu.tr, dan.ralescu@uc.edu

Abstract: *We investigate the Zariski Topology on the L-prime spectrum of modules consisting of the collection of all prime L-submodules and prove some useful results.*

Keywords: L- ideal, L-submodule, L-prime submodule, Special L-submodule, L-Irreducible submodule, L-Prime Spectrum, Zariski topology, Generic point.

1 Introduction

Throughout this paper *R* is a commutative ring with identity and *M* is a unitary *R*-module. The prime spectrum *Spec(R)* and the topological space obtained by introducing Zariski topology on the set of prime ideals of a commutative ring with identity play an important role in the fields of commutative algebra, algebraic geometry and lattice theory. Also, recently the notion of prime submodules and Zariski topology on *Spec(M)*, the set of all prime submodules of a module *M* over a commutative ring with identity *R*, are studied by many authors (for example see [1,2], [3]) that is in [3] a module with Zariski topology is called *top module* and it is shown that every multiplication module is a topmodule. As it is well known [4], introduced the notion of a fuzzy subset μ of a nonempty set *X* as a function from *X* to unit real interval $I=[0,1]$. [5] replaced I by a complete lattice *L* in the definition of fuzzy sets and introduced the notion of *L*-fuzzy sets. the notion of fuzzy groups was introduced by Rosenfeld [6]); and fuzzy sub-modules of *M* over *R* were first introduced by Negoita and Ralescu [7]. İn [8] Pan studied fuzzy finitely generated modules and fuzzy quotient modules (also see [9]). In the recent five years a remarkable amount of work has been done on fuzzy ideals in general and prime fuzzy ideals in particular, and some interesting topological properties of the spectrum of fuzzy prime ideals of a ring are obtained (see [10], [11], [12], [13], [14], [15] and [16,17].

Finally R.Ameri and R. Mahjoob 2007 studied Zariski topology on $L-Spec(M)$ defining all prime submodules of M. They investigated some basic properties of prime L-submodules and characterize the prime L-submodules of M. They established the relationship between primeless and L-primeless for a given module via the role of the lattice L. Finally they investigated the Zariski topology on $L-Spec(M)$. They showed that for L-top modules Zariski topology on $L-Spec(M)$ exists. In this paper we contınue the paper [18] R.Ameri and R. Mahjoob 2007. $L(M)$ defines all $L-$submodules of M. We define L-special submodule of M to construct a topology on L-spec(M). Then we examine the relationship between with the L-special submodule and L-irreducible submodule. L_T_p module is defined to construct a topology. We introduce that If $L(M)$ is L_T_p module then L-Spec(M) is a T_0 space. We show that L-Spec(M) is a T_1 space if and only if for every $\mu\in\text{L-Spec(M)}$ is maximal. The topological space $(\text{L-Spec(M)}=\aleph,T)$ where M is a Noetherian $R-$module is compact. We show that where Y is a closed subset of $\aleph$, If $J(Y)$ is L-prime submodule of M then Y is irreducible. And if $\eta\in Y$ is a generic point of Y then Y is irreducible.

2 Preliminaries.

Throughout of this paper, by *R* we mean a commutative ring with identity, and *M* is a unital *R*-module and *L* denotes a complete lattice. By an *L*-subset μ of a non-empty set *X*, we mean a function μ from *X* to *L* and if *L* = [0, 1], then μ is called a fuzzy subset of *X*. *LX* denotes the set of all *L*-subsets of *X*.

Definition 2.1: A fuzzy ideal of R is a fuzzy subset of R such that;
i- $\mu(x-y)\geq\min(\mu(x),\mu(y)) \quad \forall x,y\in R$
ii- $\mu(xy)\geq\mu(x) \quad \forall x,y\in R$
The set of all *L*-ideals of *R* is denoted by *L I (R)*.

Definition 2.2: If μ is a fuzzy subset of M, then for any $t\in[0,1]$ the set $\mu_t=\{x\in M \mid \mu(x)\geq t\}$ is called a level subset of M with respect to μ.

Definition 2.3([19]): Let μ and σ be a fuzzy ideal of R Then $\forall x\in R$ there exists $y,z\in R$ s.t.
$\mu.\sigma(x,y)=\sup_{x=yz}\{\min(\mu(y),\sigma(z))\}$.

Definition 2.4: Let R be a ring and $\mu\in LI(R)$. A non constant μ is called L-prime ideal for every $\eta,\beta\in LI(R)$, $\eta\beta\subseteq\mu$ implies that $\eta\subseteq\mu$ or $\beta\subseteq\mu$.

By L-Spec(R) we mean the set of all prime *L*-ideals of *R*.

Definition 2.5: A fuzzy submodule of M is a fuzzy subset of M such that
1- $\mu(0)=1$

2- $\mu(rx) \geq \mu(x) \quad \forall r \in R$ and $\forall x \in M$

3- $\mu(x+y) \geq \min(\mu(x), \mu(y)) \quad \forall x, y \in M$

L-(M) denotes the set of all *L*-submodules of M .

Theorem 2.6: μ is a fuzzy submodule of M if and only if μ_t is a submodule of M for all $t \in \operatorname{Im}\mu$.

Definition 2.7 ([20])**:** For $\mu, \nu \in L^M$ and $\eta \in L^R$, define $\mu : \eta \in L^M$ as follows: $\mu : \eta = \bigcup \{ \nu \in L^M \mid \eta\nu \subseteq \mu \}$.

Definition 2.8: Let M be an R module and $\mu \in L(M)$. A non constant μ is called L-prime submodule for every $\eta \in LI(R)$ and $\beta \in L(M)$, $\eta\beta \subseteq \mu$ implies that $\eta \subseteq \mu$ or $\beta \subseteq \mu : 1_M$.

By L-Spec(M) we mean the set of all *L*- prime submodules of M .

3 L-prime Spectrum on Modules

Example3.1:

$$\mu(x,y) = \begin{cases} 1 & \text{if } (x,y) = (0,0) \\ 1/3 & \text{if } (x,y) \in (\mathbb{Z},0) - \{(0,0)\} \\ 0 & \text{otherwise} \end{cases}$$ is L-submodule of $(\mathbb{Z},0)$,

$$\sigma(x,y) = \begin{cases} 1 & \text{if } (x,y) = (0,0) \\ 1/2 & \text{if } (x,y) \in (0,\mathbb{Z}) - \{(0,0)\} \\ 0 & \text{otherwise} \end{cases}$$ is L-submodule of $(0,\mathbb{Z})$ and

$$\beta(x,y) = \begin{cases} 1 & \text{if } (x,y) = (0,0) \\ 0 & \text{otherwise} \end{cases}$$ is L-prime submodule of $(\mathbb{Z},\mathbb{Z})$. $\mu \not\subset \beta$ and $\nu \not\subset \beta$ but $\mu \cap \nu \subseteq \beta$. We show that $\eta \cap \nu \subseteq \mu \Rightarrow \eta \subseteq \mu \text{ or } \nu \subseteq \mu$ is not always true for L-prime submodules in general; Every prime submodule is not L-Special submodule.

Definition 3.2: A submodule μ of M is called L-Special submodule, if for any η, ν L-submodule of M such that $\eta \cap \nu \subseteq \mu \Rightarrow \eta \subseteq \mu \text{ or } \nu \subseteq \mu.$

Example 3.3: Every prime L-ideal of M is L-Special submodule of R module R .

Definition 3.4: Let μ be a L-subset of M . $V(\mu)$ is a subset of all L-prime submodules of M defined by $V(\mu) = \{ \eta \mid \eta \text{ is } L-\text{prime submodule of } M,\ \mu \subseteq \eta \}$

For any L-submodule μ of M , it is easy to show that $V(1_M) = \varnothing$, $V(0_M) = L - Spec(M)$, $\bigcap_{i \in I} V(\mu_i) = V(\sum_{i \in I} \mu_i)$ and $V(\eta) \cup V(\nu) \subseteq V(\eta \cap \nu)$.

Proposition 3.5: Let M be an R module. For any L-submodule η, ν of M , $V(\eta \cap \nu) = V(\eta) \cup V(\nu)$ if and only if every L-prime ideal μ of M is L-Special submodule.

Proof: Suppose that for all $\eta, \nu \in L(M)$, $V(\eta \cap \nu) = V(\eta) \cup V(\nu).$

Let $\eta \cap \nu \subseteq \mu$ for all L-prime ideal μ of M .

$$\eta \cap \nu \subseteq \mu \Rightarrow \mu \in V(\eta \cap \nu) = V(\eta) \cup V(\nu)$$
$$\Rightarrow \mu \in V(\eta) \text{ or } \mu \in V(\nu)$$
$$\Rightarrow \eta \subseteq \mu \text{ or } \nu \subseteq \mu.$$

Conversely, let any every L-prime ideal μ of M be L-Special submodule. Then

$$\mu \in V(\eta \cap \nu) \Leftrightarrow \eta \cap \nu \subseteq \mu$$
$$\Leftrightarrow \eta \subseteq \mu \text{ or } \nu \subseteq \mu$$
$$\Leftrightarrow \mu \in V(\eta) \cup V(\nu).$$

Definition 3.6: A L-submodule μ of M is called L-irreducible submodule, if $\eta \cap \nu = \mu \Rightarrow \eta = \mu$ or $\nu = \mu$ for any η, ν L-submodule of M .

Proposition 3.7: Every L-special submodule μ of M is L-irreducible module.

Proof: Let $\eta \cap \nu = \mu.$ Then $\eta \subseteq \mu \text{ or } \nu \subseteq \mu$. And also $\mu \subseteq \eta \text{ or } \mu \subseteq \nu$ then $\eta = \mu \text{ or } \nu = \mu.$

Proposition 3.8: Let μ be L-submodule of M , ν be prime L-submodule of M . If $[\eta : \mu] \subseteq [\nu : \mu]$ implies $\eta \subseteq \nu$ for each η L-submodule of M , then ν is L-special submodule of M.

Proof: Let $\eta \cap \mu \subseteq \nu.$ Since $[\nu : 1_M]$ is L-prime ideal by R. Ameri ([1]) theorem 3.6

$$[\eta \cap \mu : 1_M] \subseteq [\nu : 1_M] \Rightarrow [\eta : 1_M] \cap [\mu : 1_M] \subseteq [\nu : 1_M]$$
$$\Rightarrow [\eta : 1_M] \subseteq [\nu : 1_M] \text{ or } [\mu : 1_M] \subseteq [\nu : 1_M]$$
$$\Rightarrow \eta \subseteq \nu \text{ or } \mu \subseteq \nu.$$

Definition 3.9: An L-submodule of μ of M is called T' L-submodule, if

i) $(\eta \cap \nu)\mu = \eta\mu \cap \nu\mu \qquad \forall \eta, \nu \in$ LI(R)

ii) $[\eta:\mu]\subseteq[\nu:\mu]\Rightarrow\eta\subseteq\nu \quad \forall\eta,\nu\in L(M)$.

Theorem 3.10: Let μ be $T^{'}$ L-submodule of M. Then η is L-special submodule if and only if for $(\eta\subseteq\mu)$ $[\eta:\mu]$ is a special L-ideal.

Proof: Let $\nu,\beta\in LI(R)$ and $\mu,\eta\in L(M)$ such that $\nu\cap\beta\subseteq[\eta:\mu]$. Then since
$[\eta:\mu]\mu\subseteq\eta$,
$(\nu\cap\beta)\mu\subseteq[\eta:\mu]\mu\Rightarrow\nu\mu\cap\beta\mu\subseteq\eta$. Since η is L-special submodule, $\nu\mu\subseteq\eta \ \ or\ \beta\mu\subseteq\eta$. Then $\nu\subseteq[\eta:\mu]$ or $\beta\subseteq[\eta:\mu]$.

Conversely, let $\eta\cap\nu\subseteq\mu$ for all $\nu,\eta\in LI(R)$ and $\mu,\beta\in L(M)$ Then $[\eta\cap\nu:\beta]\subseteq[\mu:\beta]\Rightarrow[\eta:\beta]\cap[\nu:\beta]\subseteq[\mu:\beta]$. Therefore for $T^{'}$ L-submodule then $[\eta:\beta]\subseteq[\mu:\beta]\ or\ [\nu:\beta]\subseteq[\mu:\beta]$. Then $\eta\subseteq\mu$ or $\nu\subseteq\mu$.

For any L-submodule μ of M, $V(\mu)$ denotes the set of all prime submodules of M containing μ, that is; $V(\mu)=\{\eta\in\aleph|\ \mu\subseteq\eta\}$. It is clear that; $V(1_M)=\{\ \}$ and $V(0_M)=\aleph$. Also we can show that for any family of submodules of $\{\mu_i\}_{i\in I}$ of M and $\mu,\nu\in L(M)$, $\bigcap_{i\in I}V(\mu_i)=V(\sum_{i\in I}\mu_i)$, $V(\mu)\bigcup V(\nu)\subseteq V(\mu\bigcap\nu)$. Therefore if $\zeta(M)$ denotes the collection of all subsets $V(\mu)$ of $L-Spec(M)$ then $\zeta(M)$ contains empty set and itself. Also $\zeta(M)$ closed under arbitrary intersection. But $\zeta(M)$ is not closed under finite union. Therefore on the set of all prime submodules Zariski topology does not exist since third rule of Zariski topology is not satisfied, i.e; $V(\mu\cap\nu)\neq V(\mu)\cup V(\nu)$. If this inequality replaces with equality these modules can be called as L_T_p modules.

Definition 3.11: Let μ,ν be L-submodule of M. $L(M)$ is called L_T_p module if and only if $V(\mu\cap\nu)=V(\mu)\cup V(\nu)$.

Lemma 3.12: $L(M)$ is L_T_p module if and only if every L-prime submodule μ is L-special submodule of M.

Proof: From proposition 2.5.

Lemma 3.13: Let μ,β be an L-submodule of M. If $(\alpha\subseteq\mu)$ such that $\langle\alpha\rangle=\beta$, then $V(\alpha)=V(\beta)$.

Proof:Let μ,β be an L-submodule of M and $\langle\alpha\rangle=\beta$. Since $\alpha\subseteq\beta$, $V(\beta)\subseteq V(\alpha)$.
Conversely Let $\mu\in V(\alpha)$. Then $\alpha\subseteq\mu$. This implies that $\langle\alpha\rangle\subseteq\langle\mu\rangle$ and so $\beta\subseteq\langle\mu\rangle$. Since μ is L-prime submodule of M, then $\beta\subseteq\mu$. Therefore $\mu\in V(\beta)$. This completes the proof.

Definition 3.14: $L(M)$ is distrubutive if $\eta,\nu,\mu\in L(M)$
i) $(\eta+\nu)\cap\mu=(\eta\cap\mu)+(\nu\cap\mu)$
ii) $(\eta\cap\nu)+\mu=(\eta+\mu)\cap(\nu+\mu)$.

Definition 3.15: Let Y be a subset of $L(M)$. $J(Y)$ is the intersection of all prime submodules which belongs to Y.

Definition 3.16: Let μ be a L-subset of M.
$p.rad(\mu)=\cap\{\eta|\ \mu\subseteq\eta,\ \eta \text{ is prime L-submodule}\}$ is called a radical of μ.

Definition 3.17: Let M be an R module.
$\aleph$=L-Spec(M)=$\{\mu|\mu \text{ is prime L-submodule of M}\}$ is called L-prime spectrum of M.

Definition 3.18: $V(x)=\{\eta\in\aleph|\ \eta(x)=1\}$ for all $x\in M$.

If we could not find any L-prime submodule ν which contains L-prime submodule μ, then $p.rad(\mu)=1_M$.

Theorem 3.19: Let $L(M)$ be L_T_p module with $p.rad(\mu)=\mu$. Then $L(M)$ is distrubutive.

Proof: $\eta,\nu,\mu\in L(M)$.

$$
\begin{aligned}
(\eta + \nu) \cap \mu &= p.rad((\eta + \nu) \cap \mu) \\
&= J(V((\eta + \nu) \cap \mu)) \\
&= J(V(\eta + \nu) \cup V(\mu)) \\
&= J(V(\eta \cup \nu) \cup V(\mu)) \\
&= J(V(\eta \cup \nu) \cup V(\mu)) \\
&= J((V(\eta) \cap V(\nu)) \cup V(\mu)) \\
&= J((V(\eta) \cup V(\mu)) \cap (V(\nu) \cup V(\mu))) \\
&= J((V(\eta \cap \mu) \cap (V(\nu \cap \mu))) \\
&= J((V(\eta \cap \mu) \cup (\nu \cap \mu))) \\
&= J((V(\eta \cap \mu) + (\nu \cap \mu))) \\
&= p.rad((\eta \cap \mu) + (\nu \cap \mu)) \\
&= ((\eta \cap \mu) + (\nu \cap \mu))
\end{aligned}
$$

Therefore $L(M)$ is distrubutive.

Theorem 3.20: Let $L(M)$ be L_T_p module, μ be L-submodule of M, and Y be a subset of L-Spec(M). Then

i) $V(\mu)$ is closed in L-Spec(M) and $J(Y)$ is an L-submodule of M equal to $p.rad(J(Y))$.

ii) $V(J(Y))$ is the closure of Y in L-Spec(M).

Proof: i) It is clear that $V(\mu)$ is closed in L-Spec(M) and $J(Y)$ is an L-submodule of M. Finally

$$
\begin{aligned}
p.rad(J(Y)) &= p.rad(\cap\{\eta \,|\, \eta \text{ is prime L-submodule in } Y\} \\
&= \cap\{\eta \,|\, \eta \in Y\} \qquad \textbf{ii} \\
&= J(Y).
\end{aligned}
$$

) Let $V(\mu)$ be closed set containing Y. That is $Y \subseteq V(\mu)$. Consequently, $V(J(Y)) \subseteq V(\mu)$. Since $Y \subseteq V(J(Y))$, then $V(J(Y))$ is the smallest closed subset of L-Spec(M). Thus $\overline{Y} = V(J(Y))$.

Proposition 3.21: Let Y be a subset of $L - Spec(M)$. Then

i) $J(V(J(Y))) = p.rad(J(Y)) = J(Y)$

ii) $V(J(V(Y))) = V(p.rad(Y)) = V(Y)$.

Corollary 3.22: For every family $\{Y_i\}_{i \in \Lambda}$ be a closed subsets of L-Spec(M), $J(\cap_{i \in \Lambda} Y_i) = p.rad(\sum J(Y_i))$.

Proof: Since $\{Y_i\}_{i \in \Lambda}$ be a closed subsets of L-Spec(M), then every $Y_i = V(Y_i)$ for each $i \in \Lambda$. So

$$
\begin{aligned}
J(\cap_{i \in \Lambda} Y_i) &= J(\cap_{i \in \Lambda} V(J(Y_i))) \\
&= J(V(\cup_{i \in \Lambda} J(Y_i))) \\
&= J(V(\sum J(Y_i))) \\
&= p.rad(\sum J(Y_i)) \text{ by proposition 2.20.}
\end{aligned}
$$

Corollary 3.23: Let $\mu \in Y \subseteq$ L-Spec(M). Then $\overline{\{\mu\}}$ the closure of μ is the set $V(\mu)$. We say that $\{\mu\}$ is the closed point of L-Spec(M) if and only if μ is maximal submodule in L-Spec(M).

Proof: Let $Y = \{\mu\}$ then $V(J(Y)) = \overline{\{\mu\}}$ by theorem 2.19 and $V(J(Y)) = V(J(\mu)) = V(\mu)$.

$\{\mu\}$ is the closed, that is $\{\mu\} = \overline{\{\mu\}}$. This implies $V(J(\mu)) = \mu$. So μ is maximal submodule in L-Spec(M).

Conversely now suppose that μ is maximal in L-Spec(M), then $V(\mu) = \{\mu\} = Y$ and so $Y = \overline{Y}$. Hence Y is closed.

Proposition 3.24: If $L(M)$ is L_T_p module and $\mu, \nu \in$ L-Spec(M). Then $\mu \in \overline{\{\nu\}}$ if and only if $\mu \subseteq \nu$.

Proposition 3.25: Let $L(M)$ be L_T_p module and L-Spec(M) is a T_0 space.

Proof: Let $\mu, \nu \in$ L-Spec(M) be two distinct points. We have two cases;

i) $\mu \in \overline{\{\nu\}} \Rightarrow \nu \subseteq \mu$. Since $\mu \neq \nu$, $\nu \notin \overline{\{\mu\}}$, then $\nu \in \overline{\{\mu\}}^c$. Therefore $\overline{\{\mu\}}^c$ is an open set which contains ν but not μ.

ii) $\nu \in \overline{\{\mu\}} \Rightarrow \mu \subseteq \nu$. Since $\mu \neq \nu$, $\mu \notin \overline{\{\nu\}}$, then $\mu \in \overline{\{\nu\}}^c$. Therefore $\overline{\{\nu\}}^c$ is an open set which contains μ but not ν.

Proposition 3.26: L-Spec(M) is a T_1 space if and only if for every $\mu \in$ L-Spec(M) is maximal.

Proof: Let $\forall \mu \in \text{L-Spec(M)}$ be maximal $\Rightarrow$ $\overline{\{\mu\}} = V(J(\mu)) = V(\mu) \Rightarrow$ Since μ is maximal, $\mu = \overline{\{\mu\}}$. This means that $\{\mu\}$ is the closed. Thus, L-Spec(M) is a T_1 space.

Conversely vice versa. (L-Spec(M) is a T_1 space. $\mu = \overline{\{\mu\}}$.)

Definition 3.27: $\aleph(\mu) = \aleph - V(\mu)$ is called the complement of $V(\mu)$ in L-Spec(M).

Proposition 3.28: Let Y be a subset of $\aleph$ and $\overline{Y}$ denote the closure of Y. Then $\overline{Y} \subseteq V(1_N)$ where $N = \bigcap_{\beta \in Y} \beta_*$.

Proof: Clearly $1_N(x) = 1 \Leftrightarrow \beta(x) = 1 \ \ \forall \beta \in Y.$ Therefore if $\eta \in Y,$ then $1_N \subseteq \eta$ and consequently $\eta \in V(1_N).$ Therefore the closed set $V(1_N)$ containing Y, contains $\overline{Y}$.

Definition 3.29: Let x be element of M, $V(x) = \{\mu \in \text{L-Spec(M)} \mid \ \mu(x)=1\}.$

Theorem 3.30: Let $f : M \to M'$ module epimorphism and $\overline{f} : \text{L-Spec(M}') \to \text{L-Spec(M)}$ be a function defined by $\overline{f}(\eta) = f^{-1}(\eta).$ Then $\overline{f}$ is continuous, injective and $\text{L-Spec(M}')$ is homeomorphic to the closed subset $V(1_{\ker f}).$ If f is an isomorphism, then $\overline{f}$ is homeomorphism.

Proof: Let μ and ν be an L-prime submodules of M'. Since f is surjective, $\overline{f}(\mu) = \overline{f}(\nu) \Rightarrow f^{-1}(\mu)(m) = f^{-1}(\nu)(m) \ \ \forall m \in M.$ Note that; R. Ameri([1]) theorem 3.14 $f^{-1}(\nu)$ is prime submodule. Then by definition, $\mu(\overline{f}(m)) = \nu(\overline{f}(m)) \ \forall m \in M.$ This implies that, $\mu = \nu.$ So $\overline{f}$ is injective.

If $V(m)$ is a basic closed set in L-Spec(M), then $\overline{f}^{-1}(V(m))$ is a basic closed set in L-Spec(N). Because

$$\begin{aligned}
\overline{f}^{-1}(V(m)) &= \left\{ \mu \in \text{L-Spec(M}') \,\middle|\, \overline{f}(\mu)(m) = 1 \right\} \\
&= \left\{ \mu \in \text{L-Spec(M}') \,\middle|\, f^{-1}(\mu)(m) = 1 \right\} \\
&= \left\{ \mu \in \text{L-Spec(M}') \,\middle|\, \mu(\overline{f}(m)) = 1 \right\} \\
&= V'(f(m)_1).
\end{aligned}$$

Hence $\overline{f}$ is continuous. Let μ and ν be an L-prime submodules of M'. $f^{-1}(\mu)$ is constant on $Kerf$. Indeed $f^{-1}(\mu)(m) = \mu(f(m)) = \mu(0) = 1$ for all $m \in Kerf$. Then $f^{-1}(\mu) \in V(1_{Kerf})$. If $\nu \in V(1_{Kerf})$ then $\nu \in \text{L-Spec(M)}$ constant on $Kerf$. Since f is an isomorphism it follows from theorem 3.5.11 that $f(\nu) \in \text{L-Spec(M}')$. This defines a function $\overline{g} : V(1_{Kerf}) \to \text{L-Spec(M}')$ where $\overline{g}(\nu) = f(\nu).$ Clearly $\overline{g} = \overline{f}^{-1}$. To prove the continuity of $\overline{g}$, $V'(f(m))$ be a closed set in $\text{L-Spec(M}')$. Then

$$\begin{aligned}
\overline{g}^{-1}(V'(f(m))) &= \overline{f}(V'(f(m)_1) \\
&= \left\{ \overline{f}(\nu) \,\middle|\, \nu(f(m)) = 1 \right\} \\
&= \left\{ f^{-1}(\nu) \,\middle|\, f^{-1}(\nu)(m) = 1 \right\} \\
&= V(m) \cap V(1_{Kerf}).
\end{aligned}$$

which is closed subset of $\text{L-Spec(M)}.$

Finally, suppose that f is an isomorphism. Then $Kerf = \{0\}$ and $V(1_{Kerf}) = V(0_M) = \text{L-Spec(M)}.$

Corollary 3.31: If N is submodule M such that $N \subseteq \bigcap_{\eta \in \aleph} \eta_*$, then L-Spec(M) and $\text{L-Spec}(M/N)$ are homeomorphic.

Proof: Let f be a natural homomorphism of M onto M/N. Then each $\eta \in \aleph$ is constant on $Kerf$.

Theorem 3.32: Let M be a Noetherian $R-$module. The topological space $(\text{L-Spec(M)} = \aleph, T)$ is compact.

Proof: If $L - \{1\}$ has no prime element, then $\aleph = \varnothing$ proof is complete. Suppose that $L - \{1\}$ have prime

elements and a be a prime element of $L-\{1\}$. Let $\{\aleph((m_i)_t)\mid i\in\Lambda, t\in K\subseteq L-\{0\}\}$ be a cover of $\aleph$ by its basic open sets. Let $\vee\{t\mid t\in K\}=b.$ Then $\{\aleph((m_i)_b)\mid i\in\Lambda\}$ also cover $\aleph$. Therefore $\aleph=\cup\{\aleph((m_i)_b)\mid i\in\Lambda\}=\aleph(\bigcup_{i\in\Lambda}(m_i)_b)).$ Hence $V(\bigcup_{i\in\Lambda}(m_i)_b))=\varnothing.$ Let N be a prime submodule of M. Define $\eta: M\to L,$ defined by $\eta(m)=\begin{cases}1 & \text{if } m\in N\\ 0 & \text{otherwise}\end{cases}$. So η is L-prime submodule then $\eta\in\{\aleph((m_i)_b)\mid i\in\Lambda\}$ and so $\bigcup_{i\in\Lambda}(m_i)_b\not\subset\eta.$ This implies that $\exists i\in\Lambda$ s.t. $(m_i)_b\not\subset\eta.$ Hence either $b>\eta(m_i)$ or b and $\eta(m_i)$ are noncomparable. In either case $m_i\notin N$. Therefore $\{m_i\mid i\in\Lambda\}$ is not contained in any proper submodule of M. $\langle\{m_1,m_2,...,m_k\}\rangle=M$ since M is Noetherian. Suppose $V(\cup_{i=1}^k(m_i)_b)\neq\varnothing$ let $\beta\in V(\cup_{i=1}^k(m_i)_b).$ Then $\cup_{i=1}^k(m_i)_b\subseteq\beta\Rightarrow\beta(m_i)\geq b$ for all $i=1,...k$. Suppose $\exists i,$ s.t. $\beta(m_i)\neq 1.$ Now $\beta(m)=\beta(\sum_{i=1}^k r_im_i)\geq\wedge_{i=1}^k\beta(m_i)\geq\wedge_{i=1}^k b=b.$ Now $\beta\in\aleph$ and so $\beta\in\{\aleph((m_i)_b)\mid i\in\Lambda\}.$ Therefore $\exists i\in\Lambda$ s.t. $\beta\in\aleph((m_j)_b).$ Then $(m_j)_b\not\subset\beta.$ Thus either $b>\beta(m_j)$ or b and $\beta(m_j)$ are noncomparable. However, $\beta(m_j)=\beta(\sum_{i=1}^k r_im_i)\geq\wedge_{i=1}^k\beta(m_i)>b.$ This is a contradiction. Hence $\beta(m_j)=1.$ This implies that $M=\beta_*.$ This is a contradiction. Therefore $V(\cup_{i=1}^k(m_i)_b)=\varnothing$. Consequently, $\{\aleph((m_i)_b)\mid i\in 1,...,k\}$ is a subcover of $\aleph$. This completes the proof.

Definition 3.33: A topological space T is called irreducible if for any decomposition $T=A_1\cup A_2$ where A_1, A_2 are closed subsets of T, then $T=A_1$ or $T=A_2$.

Theorem 3.34: Let Y be a closed subset of $\aleph$. If $J(Y)$ is L-prime submodule of M, then Y is irreducible.

Proof: Suppose that $J(Y)$ is L-prime submodule of M. Suppose $Y=Y_1\cup Y_2$, where Y_1,Y_2 are closed subsets of $\aleph$. Then $J(Y)\subseteq J(Y_1)$ and $J(Y)\subseteq J(Y_2)$. Also $J(Y)=J(Y_1\cup Y_2)=J(Y_1)\cap J(Y_2)$. Then $J(Y_1)oJ(Y_2)\subseteq J(Y_1)\cap J(Y_2)\subseteq J(Y).$ Since $J(Y)$ is L-prime submodule of M, then $J(Y_1)\subseteq J(Y)$ or $J(Y_2)\in(J(Y):1_M)$. If $J(Y_1)\subseteq J(Y)$, the proof is complete. If $J(Y_2)\in(J(Y):1_M)$, then $J(Y_2)\subseteq J(Y)$. This completes the proof.

Definition 3.35: Let Y be a closed subset of $\aleph$ and $\eta\in Y$. Then η is called a generic point of Y if $Y=\overline{\{\eta\}}$, the closure of $\eta.$

We know that ıf $Y\subseteq\aleph$ where $\aleph$ is a topological space, then Y is irreducible if and only if $\overline{Y}$ is irreducible.

Theorem 3.36: Let Y be a closed subset of $\aleph$ and $\eta\in Y$ be a generic point of Y Then Y is irreducible.

Proof: Since η is a generic point of Y, then $Y=\overline{\{\eta\}}$. Since $\{\eta\}$ is irreducible, $\overline{\{\eta\}}$ is irreducible. Therefore Y is irreducible

4 Conclusion

Letting $L-\xi(M)=\{V(\mu.1_M)\mid\mu\in LI(R)\}$. It is easy to prove that this set always induces a topology T on $L-spec(M)$. R. Ameri and R Mahjoob showed that $L-\xi(M)$ induces a topology which is called Zariski topology if and only if M is a top module. By following them we show that a topology T on $L-spec(M)$ exists if and only if $L(M)$ is L_T_p module. Under this condition topology T on $L-spec(M)$ is T_0 space. Behind the existing T_1 space, if M is a Noetherian $R-$module, then the topological space $(\text{L-Spec(M)}=\aleph,T)$ is compact.

Acknowledgement

"Bayram Ali Ersoy's work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK); Dan Ralescu's work was supported by a Taft Travel for Research Grant ".

References

[1] C.P. Lu, Prime submodules of modules, *Comm. Math. Univ*. 33 (1987) 61-69.

[2] C.P. Lu, The Zariski topology on the spectrum of a modules, *Houston Journal of Mathematics*. 25 (3) (1999) 417-432.

[3] R.L. Mccasland, M.E. Moore, P.F. Smith, On the spectrum of modules over a commutative ring. *Comm. Algebra*, 25 (1) (1997) 417-432.

[4] L.A. Zadeh, Fuzzy sets, *İnform and Control* 8 (1965) 338-353.

[5] J.A. Goguen, L-fuzzy Sets, *J. Math. Appl.* 18 (1967) 145-174.

[6]A. Rosenfeld, Fuzzy groups, *J. Math. Anal. Appl*. 35 (1971) 512-517.

[7] C.V. Negoita, D.A. Ralescu, Application of fuzzy systems analysis, Birkhauser, Basel, 1975.

[8] F.Z. Pan, Fuzzy finitely generated modules, *Fuzzy Sets and Systems*, 21 (1987) 105-113.

[9] F.I. Sidky, On radical of fuzzy submodules and primary fuzzy submodules, *Fuzzy Sets and Systems* 119 (2001) 419-425.

[10] V.N. Dixit, R. Kumar, N. Ajmal, Fuzzy ideals and fuzzy prime ideals of a ring, *Fuzzy Sets and Systems*, 44 (1991) 127-138.

[11] T.K. Mukherjee, and M.K. Sen, On fuzzy ideals of a ring I, *Fuzzy sets and systems* 21 (1987) 99-104.

[12] S.K. Bhambri, R. Kumar, P. Kumar, Fuzzy prime submodules and radical of a fuzzy submodules, *Bull. Calcutta Math. Soc*. 87 (1993) 163-168.

[13] H. Hadji-Abadi, M.M. Zahedi, Some results on fuzzy prime spectrum of a ring, *Fuzzy Sets and Systems*, 77 (1996) 235-240.

[14] R. Kumar, Fuzzy prime spectrum of a ring, *Fuzzy Sets and Systems*, 46 (1992) 147-154.

[15] R. Kumar, J.K. Kohli, Fuzzy prime spectrum of a ring II, *Fuzzy Sets and Systems*, 59 (1993) 223-230.

[16] H.V. Kumbhojkar, Spectrum of prime fuzzy ideals, *Fuzzy Sets and Systems*, 62 (1994) 101-109.

[17] H.V. Kumbhojkar, Some comments on spectrum of prime fuzzy ideals of a ring, *Fuzzy Sets and Systems*, 88 (1997) 109-114.

[18] R. Ameri, R. Mahjoob, Spectrum of prime L-submodules, *Fuzzy Sets and Systems*, 5122 (2007) .

[19] J.N. Mordeson, D.S. Malik, Fuzzy Commutative Algebra, *Worlf Scientific Publishing*, Singapore, 1998.

[20] D.S. Malik, J.N. Mordeson, L-prime spectrum of a ring, *Fuzzy Information Processing Society NAFIPS '97. 1997 Annual Meeting of the North American*, (1997) 273-278.

Author Index

ISBN: 978-989-95079-6-8